

RESEARCH ARTICLE

Detecting the body's reproductive hormonal brake against tissue overgrowth: Micrin/SgII-70

John E. Hart¹, Keith G. Davies^{1,2*}, Christopher R. Mundy¹, Aurelia C. Hart¹, David R. Howlett³, Russell P. Newton⁴

1 Endocrine Pharmaceuticals, Tadley, Hampshire, United Kingdom, **2** School of Life and Medical Sciences, University of Hertfordshire, Hatfield, United Kingdom, **3** Wolfson SPaRC (Sensory, Pain and Regeneration Centre), King's College London, London, United Kingdom, **4** Biochemistry Group, Institute of Life Sciences, Medical School, Swansea University, Swansea, United Kingdom

* k.davies@herts.ac.uk



Abstract

A new humoral factor has been detected, within a project aiming to disclose the body's reproductive hormonal brake against tissue overgrowth, micrin ('my-crin'). Factor purification involved fractionation of ovine and bovine ovarian follicular fluid and blood plasma and serum, with evaluation via rat bioassays *in vivo* and *in vitro*. Analysis averse, the molecular effector provided a chemical conundrum. Evidence from mass spectrometry (MS) is problematic, with spectra from MALDI-TOF, the only productive MS modality, confused by a target polypeptide exhibiting artefacts during processing, storage and MALDI set-up in terms of C-terminal truncation and water losses, along with dimerization (C-terminally truncated fragments doubled up). Evidence from chemical sequencing of amino acid (aa) residues is likewise confusing, but consistent with a spiralised depolymerisation within the Edman reaction chamber of a unitary polypeptide. Data decryption has overcome molecular intractability, supported by the results of tryptic digestion and epitope mapping using immunohistochemistry (IHC). The detected inhibitory factor relates to secretogranin II (SgII), the neurosecretory prohormone, in the form of a secreted acidic 70-aa polypeptide derivative called here SgII-70 ('sig two seventy'). The product of peptide splicing, micrin/SgII-70 is a twisted amphipathic molecule, with ends entwined (via salt bridging), the complex knotted structure confounding Edman and MALDI analyses, as well as computational molecular modelling, whilst conferring protease and heat resistance. Hexapeptide mimetics simulating both ends of the hormone together have been demonstrated in different species and settings in antiorganotrophism (tissue reduction via fewer, smaller cells) and reproductive modulation (more offspring). There are potential clinical applications for agonists in tissue overgrowth conditions such as endometriosis, PCOS, BPH and cancer, and for micrin antagonists in infertility.

OPEN ACCESS

Citation: Hart JE, Davies KG, Mundy CR, Hart AC, Howlett DR, Newton RP (2026) Detecting the body's reproductive hormonal brake against tissue overgrowth: Micrin/SgII-70. PLoS One 21(3): e0340980. <https://doi.org/10.1371/journal.pone.0340980>

Editor: Francesca Torrini, National Research Council (CNR), ITALY

Received: December 31, 2024

Accepted: December 29, 2025

Published: March 2, 2026

Copyright: © 2026 Hart et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Figshare: Supplementary Information 1: Candidates & Purifications. <https://doi.org/10.6084/m9.figshare.27110071.v2>. This project outlines the candidate protein and purification techniques utilised. Figshare: Supplementary Information 2: Assay Methods and Results. <https://doi.org/10.6084/m9.figshare.27110071.v2>

[org/10.6084/m9.figshare.30688655](https://doi.org/10.6084/m9.figshare.30688655). This project details the methodology used in exemplifying anti-organotrophism for the candidate molecule and gives results. Figshare: Supplementary Information 3: Mass Spectrometry Studies of Candidate 7500. <https://doi.org/10.6084/m9.figshare.27110284.v2>. This project presents MALDI-TOF MS data of the polypeptide Candidate 7500 and describes it as a novel proteoform of the neuroendocrine prohormone secretogranin II, dubbed sSgII-70. Figshare: Supplementary Information 4: Tryptic Digests. <https://doi.org/10.6084/m9.figshare.27110311.v2>. This project attempts to elucidate the identity of Candidate 7500/sSgII-70 using trypsinisation. Figshare: Supplementary Information 5: Molecular Modelling of sSgII-70. <https://doi.org/10.6084/m9.figshare.27110320.v2>. This project presents in silico molecular modelling of sSgII-70 directed at a conformational understanding of the molecule. Figshare: Supplementary Information 6: Extended Discussion. <https://doi.org/10.6084/m9.figshare.27110332.v2>. This project provides a discussion of possible chains of logic and associated theoretical frameworks which can be associated with and possibly explain the generated organotrophic data and MALDI analyses. Figshare: Supplementary Information 7: Arrive 2.0 Guidelines. <https://doi.org/10.6084/m9.figshare.27110371.v1>. This project provides the Arrive 2.0 Guidelines.

Funding: The early research benefitted from a SMART Feasibility award (YHF/21865/SM00) and a SMART Development award (YHF/21865/SM02) from the UK government. This funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No other specific grant has been received from any other funding agency in the public, commercial or not-for-profit sectors. The research was otherwise paid for by Endocrine Pharmaceuticals of Hampshire, UK (Company No. 03005721), which was involved in all aspects of the research.

Competing interests: JEH is founding scientist of Endocrine Pharmaceuticals of Hampshire, UK (Company No. 03005721), the commercial funder of the work, and is an employee of the company, holding shares and share options. Endocrine has a relevant patent (WO 2009/087424) and a patent application (GB2112385.6) which does not alter adherence to PLOS ONE policies on sharing data and materials. All coauthors, whose efforts were pro bono, are participants in Endocrine's share option scheme and most of those mentioned in Acknowledgements hold shares or share options in Endocrine.

Background

There is a paradoxical increase in pituitary weight in both female rats and mice in the first few weeks after ovariectomy, seemingly indicating the abolition of a net inhibitory ovarian influence, the established hypophysiotrophic effect of oestrogens notwithstanding [1]. In the male rat, castration also leads to pituitary hypertrophy, an effect that can be reversed by an aqueous testicular extract but not by a lipid testicular extract (steroid rich) or by testosterone alone [2]. Exogenous oestrogens are the most organotrophic substances known, enlarging organs visceral (e.g., heart, liver, kidneys), endocrine (e.g., pituitary, adrenal, ovaries) and reproductive (e.g., uterus) [1]. The organotrophic effects of exogenous oestrogens are not seen in hypophysectomized female rats, suggesting an indirect effect via the hypothalamic-pituitary axis [3]. The organotrophic oestrogen effect in female rats is more marked in ovariectomised animals than in intact ones [4–7]. This observation suggests gonadal inhibition, leading to the hypothesis that the gonadal inhibitory influence is stimulated by antioestrogens, accounting in part for the antiorganotrophic activity of these compounds in rodents and their therapeutic usefulness in humans [7]. The antioestrogen clomiphene, for example, reduces the size of the rat pituitary, liver, spleen, kidneys, heart, uterus and ovaries in adults whose body weights are unchanged [3]. The antioestrogenic blockade of the organotrophic effects of exogenous oestrogens is blunted by prior removal of the ovaries.

The foregoing and other pieces of evidence led to the suggestion that the body has a unitary hormonal reproductive brake against tissue overgrowth, non-steroidal in character, which brake is lifted to initiate puberty and the assumption of adult stature and whose action wanes with age, leading to a rise in cancers and the enlargement of the prostate [7]. This hypothetical hormonal factor, dubbed 'micrin' (pronounced *my-crin*; Greek, *small*, as in small organs), is conceived as being a potentially novel inhibitory effector in the hypothalamic-pituitary-gonadal component of the 'organotrophic system' of internal size regulation. The hunt for a hormone described here implicitly tests alternative hypotheses such as 'factor non-existence' and 'existing factor with newly recognized antiorganotrophic activity'. The form of the paper arises from the integration of a body of novel data (notably MS) with disclosures from prior publications and dispersed in the patent literature.

Methods

This paper is supported by seven files of supplementary information, [S1–S7 File](#). All experimental procedures, including those in vivo subject to the research checklists of [S7 File](#), were performed in compliance with applicable laws, regulations and professional standards of probity and good faith, with appropriate ethical oversight. The studies described in [S2 File](#) were conducted under United Kingdom Home Office Project Licence Nos. PPL 80/1025, PPL 80/1150, PPL 30/1487 & PPL 40/3660. As few animals were used as possible, in accordance with 'Reduction' in the Three Rs of Replacement, Reduction, Refinement. In the matter of sample identity, proprietary codes were favoured over explanatory designations to foster blind experimentation with unbiased analysis and impartial supervision. This objective was aided in any

case by project inscrutability, which countered confirmation and scepticism bias. Protocols were pre-registered to comply with UK Home Office licencing requirements and for ethics submissions but were not pre-registered in an open database, being exploratory studies intended to culminate a hormone discovery project involving prior publications, wherein are described diverse materials (with provenances) and methods (with protocols and permissions). Data integrity has been upheld throughout, with no exclusion of outliers, no withholding of null findings and with appropriate recording and archiving. Raw data are available via supplementary [S1–S6 Files](#), in the interests of transparency and the provision of a true account.

Purification

The search for the molecular explanation of an inhibitory gonadal hormonal influence on mammalian tissue masses featured bioassay guided fractionations involved ovarian follicular fluid, ovarian venous plasma and systemic blood plasma and serum mainly from ewes (*Ovis aries*), but also cattle (*Bos taurus*) and pigs (*Sus scrofa domestica*), with testing in bioassays in vivo and in vitro in the rat (*Rattus norvegicus*).

Fractionation involved spin filtration for separation and concentration plus size-exclusion chromatography, with anion exchange ('anionex') chromatography, using a positively charged resin for final purification [8]. Mass analysis involved MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionization - Time-of-Flight Mass Spectrometry, hereafter 'MALDI'; positive ion, linear mode unless otherwise specified, delayed extraction typically of 350 nsec, low mass gate usually set at 1000 Da) and SDS-PAGE (Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis, hereafter 'gel'; mainly 1D). The purification method was developed at The Babraham Institute, Cambridge, UK, hereafter 'Babraham' (see Babraham Method in [S1 File](#)). Anti-organotrophic activity was assessed in vivo by shrinkage (weight loss) of internal organs over four days in the young adult female intact rat and in vitro by inhibitory effects on the viability of rat bone marrow cells (BMCs) in primary culture ([S2 File](#) Assay Methods & Results). In the BMC assay, bone marrow in semisolid form is centrifuged from sections of femoral and tibial bone and cultured in media [9], to provide an organ-like mixed-cell population including haematopoietic stem and progenitor cells and mesenchymal stromal cells such as fibroblasts, pericytes and endothelial cells.

An internal-organ-reducing activity has been tracked through successive stages in the purification of sheep's blood [10], using the rat organometric assay in vivo for guidance ([S2 File](#)), to maximum purity represented by anionex fractions. An organotrophic pattern of 'pituitary down, adrenals up' is reported to be a micrin signature [7]. This is seen in [Fig 1](#) in rats (n=7) exposed to anionex HPLC fractions of ovarian venous plasma from ovary-intact (OV-INTACT) sheep when compared to rats receiving such fractions from the jugular vein plasma of ovariectomised (OVX) ewes. This result has been replicated more than once ([S2 File](#)). In one replication ([S2 Fig 7](#) in [S2 File](#)) the adrenal part of the micrin signature of 'pituitary down, adrenals up' reached statistical significance on an adrenal relative weight >20% above the control value (n=7, t test, P<0.05), while an apparent pituitary reduction of >8% was non-significant. Statistically, this is the reverse of the situation in [Fig 1](#), while the directionality of the organ weight changes is the same.

The surprising pattern of 'pituitary down, adrenals up', though not invariably evoked, has been recorded in other circumstances: (a) when using steroid-depleted aqueous beef testicular extract in male intact rats, in which 'prostate down' is also seen [11]; (b) in male intact rats (n=5) receiving a 0.1–0.2 M NaCl anionex FPLC fraction of ovarian venous plasma previously subjected to spin and gel filtration to provide a 10–20 kDa fraction ([S1 File](#) Babraham Method), in which the relative organ weights vs controls after a few days were pituitary (–2.3%, ns, t test), adrenals (+9.7%, ns) and prostate (–23.4%, P<0.05, significant) [10] ([S2 Fig 5](#) in [S2 File](#)); (c) in female rats (n=5) treated with clomiphene [3], a putative micrin stimulator [7]; and (d) in the same experiment, where clomiphene blunted oestrogen-induced pituitary and other organ enlargements but exaggerated an oestrogen-induced adrenal enlargement. The same directional difference is seen in young rats after ovariectomy, with pituitary and other organ weights above those of intact animals, adrenals below [12]. Antihypophysiotrophic and broadly antiorganotrophic, micrin is yet adrenotrophic.

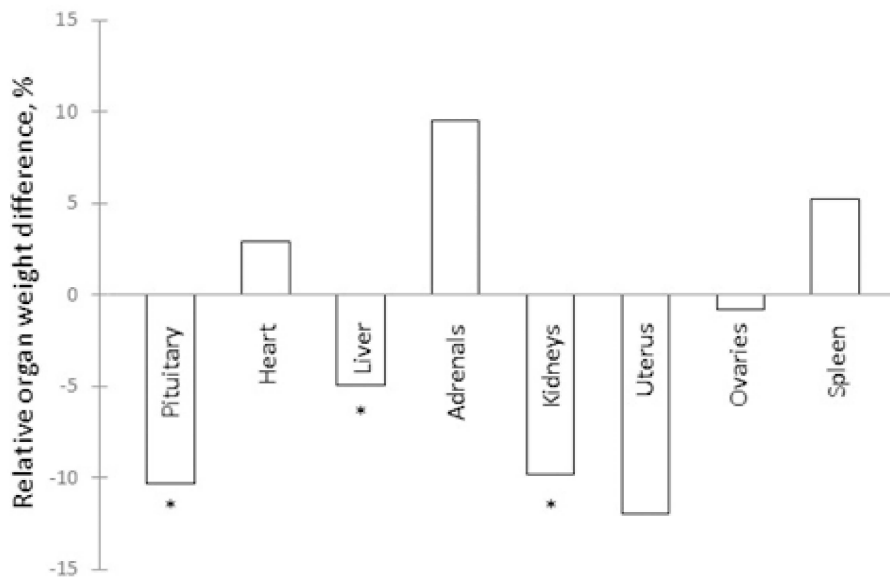


Fig 1. Organometric assay. Effects in rats in vivo of plasma anionex fractions from OV-INTACT vs OVX sheep. Post-mortem organ weights expressed as a percentage of body weight for two groups of seven intact adult female Sprague Dawley rats. The zero baseline is provided by rats receiving for four days by intraperitoneal injection a fraction (spin and gel filtered to 10-20 kDa, eluting in HPLC anionex at ~0.2 M NaCl; [S1 File](#) Babraham Method) of the jugular vein plasmas pooled from two OVX sheep. The other group of rats received the same fraction in the same amount from ovarian venous plasmas pooled from six OV-INTACT sheep. No significant body weight changes were seen to confound the picture of organ shrinkage with adrenal enlargement. * = $P < 0.05$, one-tailed t test. (N.B. OV-INTACT versus PBS controls, adrenals +8.53%, significant: [S2 File](#)).

<https://doi.org/10.1371/journal.pone.0340980.g001>

In the rat study in vivo that furnishes [Fig 1](#), gross abnormalities were absent and a histological investigation confirmed a lack of toxicity, in line with a lack of behavioural deficits. The organ mass alterations reported here and previously [[10](#)] appear to be due to cell shrinkage, reduced cell numbers by apoptosis and suppression of angiogenesis, according to the results of assays in vitro ([S2 File](#)). In summary, fewer, smaller cells.

In the standard BMC cell death assay in vitro use was made of jugular vein whole plasma from OVX sheep ($n=2$) and ovarian vein whole plasma from OV-INTACT sheep ($n=22$), the neat plasmas being putatively micrin-minus and micrin-plus, respectively. Cell survival after 48h (assessed by Alamar Blue) was normalised against one of the OVX samples as 100%, with the other at 98%. For the 22 OV-INTACT samples the mean \pm SEM was $25.7 \pm 4.7\%$ ($P < 0.0001$, two-tailed t test, significant; [S2 Fig 10](#) in [S2 File](#)). Cell shrinkage was evaluated in BMCs using Magiscan image analysis, though the results were obvious to the microscope-aided eye (images, [S2 File](#)). In the middle in cell size terms were saline controls. Larger were cells exposed to OVX plasma (putatively micrin-minus), while manifestly smaller were those given OV-INTACT plasma (putatively micrin-plus). As a proxy for cell volume, cross-sectional areas were assessed after 48h (all cells in 10 microscopic fields at x40 magnification, mean \pm SD, arbitrary units). The results were: saline control, 0.67 ± 0.38 ; plasma from OVX sheep, 0.78 ± 0.55 (+16% vs saline); and plasma from OV-INTACT, 0.47 ± 0.23 (-30% vs saline). OV-INTACT cells were 40% smaller than OVX cells in area. Wide variances precluded statistical significance (t test), but area in any case underestimates volume differences. In another study series selected anionex fractions of sheep serum produced a significant increase in apoptotic cells, as judged by Annexin V ([S2 Table 6](#) in [S2 File](#)) and a DNA fragmentation assay ([S2 Table 7](#) in [S2 File](#)), without an increase in necrotic cells, suggesting a physiological effect rather than toxicity.

From column and solvent studies used in conjunction with assays of BMC viability it was judged that the growth inhibitory agent (singular for parsimony) did not have the characteristics of a lipid or a carbohydrate [[13](#)]. (Delipidated sheep

serum samples still contained the main candidate molecule from MALDI, to be described shortly.) The molecular weight was circa 10 kDa, while charcoal reducing bioactivity in vitro probably indicated a charged molecule. A negative charge was implied by the generation of anionex fractions that shrank multiple organs in vivo (Fig 1). Peptidic character was suspected since the activity in vitro was reduced by the proteolytic enzyme trypsin and, separately, by heat, as will be described later. In terms of retention of bioactivity in vitro, plasma from blood anticoagulated with EDTA proved to be a better purification feedstock than serum or heparinised plasma (S1 File). As EDTA inhibits proteolytic enzymes via chelation of metal ions, this is consistent with micrin having a proteinaceous character. Loss of bioactivity in vivo and in vitro occurs after storage short-term (a few days), suggestive of susceptibility to protease activity or more likely pure-form lability (i.e., intrinsic instability), since lyophilization did not conquer apparent lability, instead exacerbating loss of activity after storage (S2 Fig 23 in S2 File).

Candidates

During an extensive purification campaign, a dozen candidates were identified for the sought-for tissue-mass inhibitory reproductive hormonal influence in mammals (S1 File). To summarise the bioassay-guided physicochemical fractionation campaign, two things emerged: (i) 'Candidate 7500', a group of peaks (ions) in the mass-to-charge (m/z) range 7–8000 (Figs 2–6) seen in MALDI (at Babraham, unless otherwise stated); and (ii) 'EPL001', an Edman machine N-terminal sequence of 14 aa residues, MKPLTGKVKFNNI (S1 Table 1 in S1 File; likewise obtained at Babraham), relating to an ovine blood plasma fraction containing Candidate 7500.

Counting Candidate 7500 as Candidate I and EPL001 as Candidate XI, there were eight other candidates from the physicochemical purification of sheep material, including fibrinopeptide A and androglobin. Seven of these eight were ruled out on the basis of size, contra-activity and as one-off seeming contaminants. The eighth, Candidate IV, is 'Luteal phase entity' at m/z 3450. This was later interpreted as a fragment of Candidate 7500 (see 'sSgII-70', 3000s analysis). Candidate X from the rat, NXPE family member 4 (Q5XI89, formerly FAM55D), was discounted on the basis of predicted pI. Candidate XII will be analysed later under the rubric 'Likely SgII related'. The dismissal of all but three of the twelve candidates for the inhibitory gonadal influence amounts to the discounting of a series of alternative hypotheses.

Besides the candidates described, there could in principle have been other entities present in the anionex samples undetected by MALDI, though a small molecule hunt using this MS modality detected nothing and neither did one up to m/z 25,000. Meanwhile a 'no preconceptions' investigation of anionex fractions using ESI MS showed no significant peak large or small, including the frontrunning Candidate 7500 (= Candidate I). Purification reproducibility waned in regard to this candidate, for reasons that only became clear later with the recognition of factor grand fragmentation. Candidate 7500 had not been lost. Bits fall off its C terminus, as evidenced by MALDI spectra. Factor elusiveness (i) vitiated de novo sequencing by MS/MS using trypsin proteolysis (see later in this section) and (ii) compromised the use of a polyclonal antiserum (hereafter 'antibody') raised against EPL001 (= Candidate XI), when such became available, to aid identification by subtraction from spectra.

Viable candidates remaining after the foregoing analysis are I (Candidate 7500), XI (= EPL001) and XII (= 'Likely SgII related').

Candidate 7500

Candidate 7500 is not a unitary entity, but a perplexing range of MALDI peaks of m/z 7–8000 (Figs 2–6; S3 File). In MALDI the sample is applied to a metal target plate with a low molecular weight UV absorbing organic matrix in vast excess to aid desorption and ionization. When this mixture dries it forms crystals, to which a UV laser imparts energy. The matrix heats up rapidly and sublimates, along with the sample molecules co-crystallized with it. The positive ions so produced, most singly charged, ablate into the vacuum and are accelerated in an electrical field into an electrostatic flight tube before being reflected between magnets, for separation on the basis of m/z ratio. In a Time-of-Flight (TOF) setup the

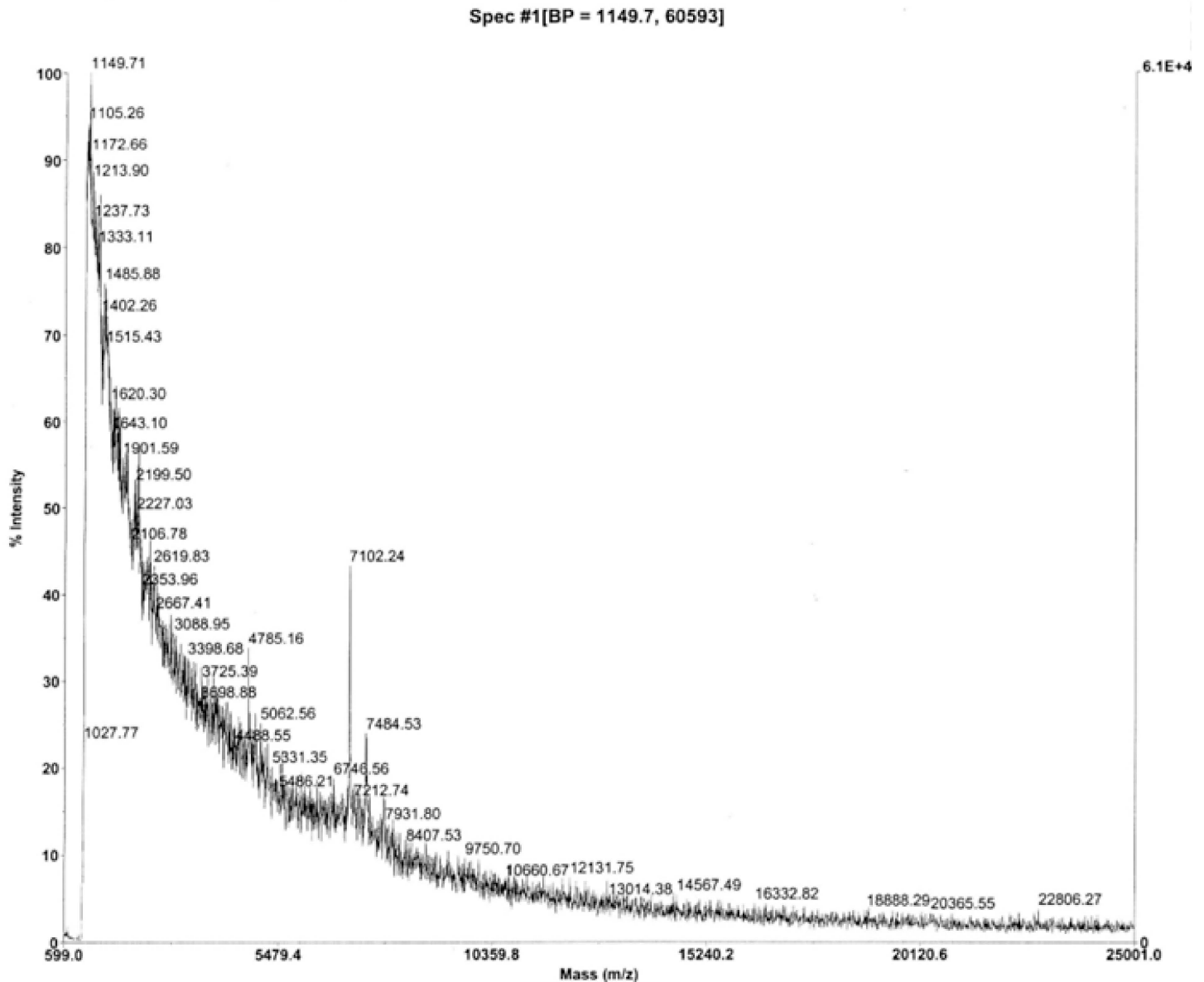


Fig 2. MALDI MS of ovine ovarian follicular fluid. Anion exchange Fraction 27 (i.e., late eluting), after spin and gel filtration. Matrix sinapinic acid, calibration against carbonic anhydrase (Babraham MS, [S3 File](#)). For Candidate 7500 MS peak analysis in regard to [Figs 2–5](#) see [Table 1](#). The peak at m/z 4785 is analysed to be a **42mer** N-terminal fragment of sSgII-70 and the 1149 base peak ('BP' in header) an **11mer**. (See [S3 Figs. 1 & 2](#) in [S3 File](#) for equivalent bovine and porcine spectra.).

<https://doi.org/10.1371/journal.pone.0340980.g002>

ions are accelerated to equal energy for mass measurement. MALDI-TOF MS offers high sensitivity, detecting proteins and peptides for example at low femtomole levels. MALDI is fairly tolerant of salts and other background molecules, simplifying sample preparation. It is regarded as a soft ionization method, which aids sensitivity, except that high laser energy had to be used in the present project to visualise exiguous material, contributing to a high noise to signal ratio. Hitting material hard and often with a laser to visualize peaks can cause sample decay and disturb calibration. The weakness of some Candidate 7500 peaks prompts concern that they represent contaminants. One potential cause of heterogeneity

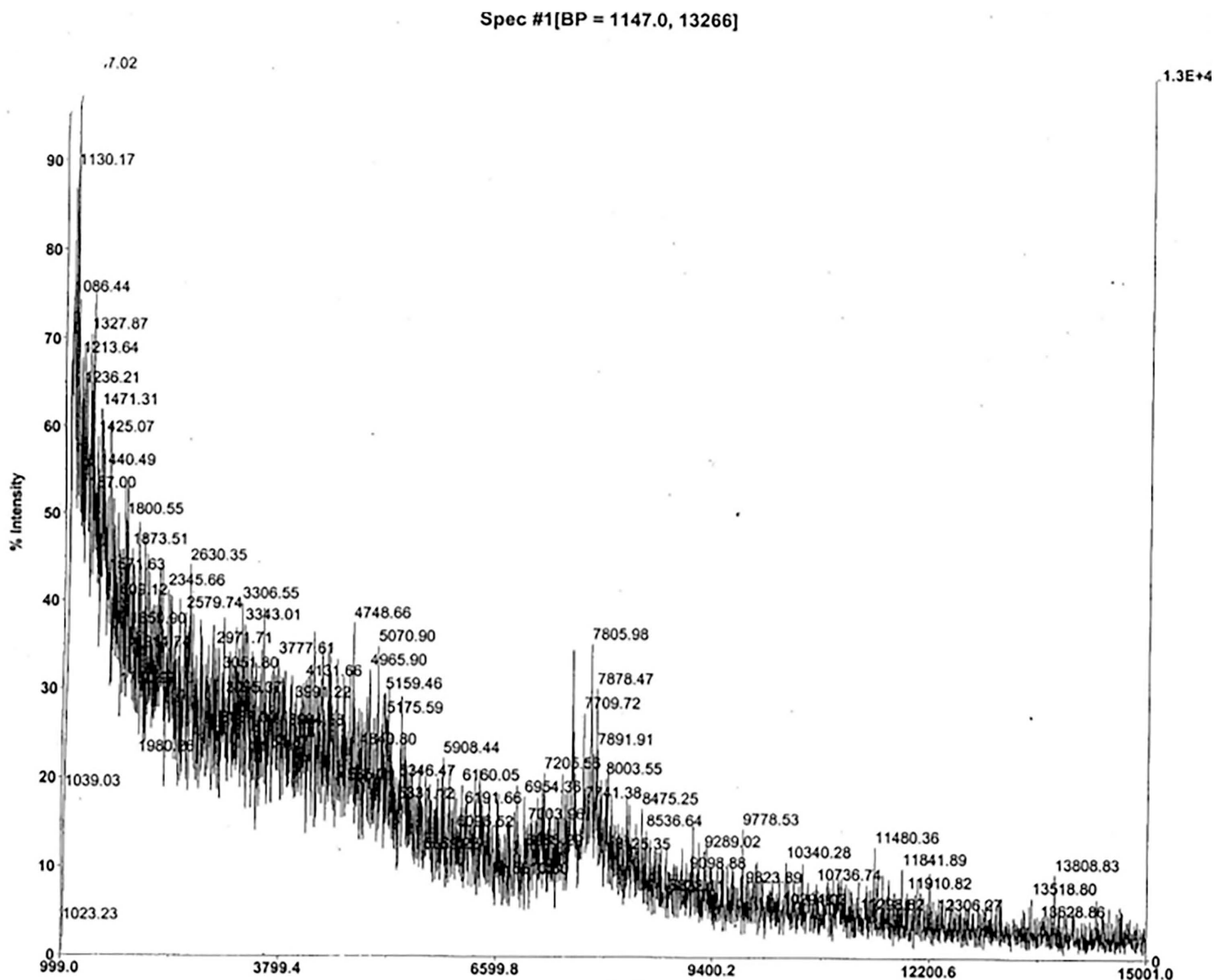


Fig 3. MALDI MS of ovine jugular vein EDTA plasma anionex Fraction 29. The Beale 4 Edman reading (xxPxxxxVxxFNxx) was obtained from a gel lacking in visible bands but excised anyway at ~7kDa. The Beale 4 aa correspond to EPL001 residues 3, 8, 11 and 12, without EPL001's 10 other aa being discernable, thus: $\text{MKPLTGVKVFNNI}_{14}$ (S1 Table 1 in [S1 File](#), SEQ ID NOs: 3 & 4). MALDI matrix and calibration as [Fig 2](#). The ion at m/z 7805 is an integer match to a predicted sSgII-70 **68mer**. A peak without mass annotation is 'c7500'. Minor peaks at 5070 and 4748 are analysed to be **45mer** and **43mer** fragments of sSgII-70, respectively. The base peak is missing from this record but is 1147 ('B.P.' in header), an sSgII-70 **11mer**.

<https://doi.org/10.1371/journal.pone.0340980.g003>

in MS is SDS, as used in electrophoresis, but Candidate 7500 showed multiple peaks even when SDS-PAGE was *not* involved (e.g., [Fig 4](#)).

Confusion is dissipated by the realisation that the multiple m/z 7–8000 peaks in MALDI probably arise from a single polypeptide subject to molecular losses in the form of C-terminal aa residues and water molecules ('dual decrementation'). The decryption key is Candidate XII, in the form of 'Likely SgII related', with an MS analysis given in full later (see 'SgII-70'). The dual decrementation scheme is shown in [Table 1](#).

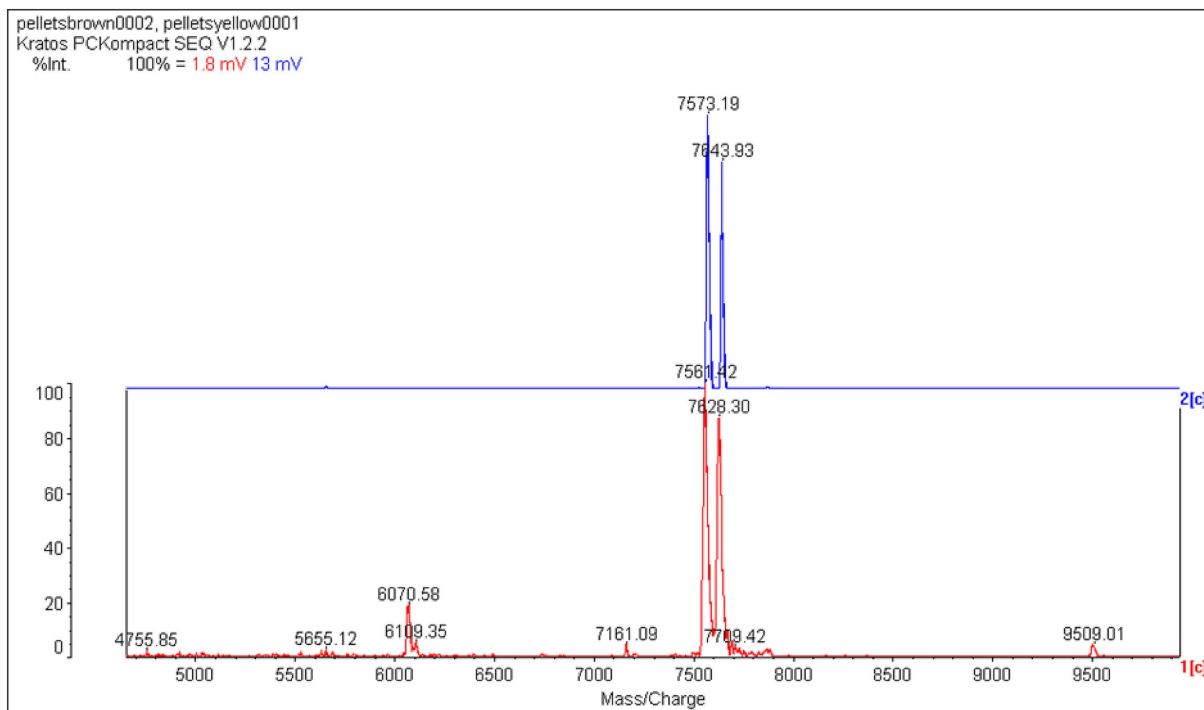


Fig 4. Precipitate. MALDI MS of a solid ‘upstream precipitate’ that formed on the 3kDa membrane during a 3–30kDa ultrafiltration of bovine ovarian follicular fluid (upper panel) and ovine jugular vein EDTA plasma (lower panel) (S1 in [S1 File](#) Sheffield Method). The sought-for mammalian Candidate 7500 flew in MS from a complex bacterially contaminated mixture added without treatment to a protein matrix. Calibration (against insulin) varied due to the use of high laser power (operator comment). The mV figures in the header (1.8=bovine, 13=ovine) are weak. Strong samples can show up to 1000 mV. The bovine double peak corresponds to an alanine difference, **67mer** vs **66mer** bSgII-70 (S3 Table 2 in [S3 File](#)). The ovine double peak corresponds to a lysine and four waters difference, **66mer** vs **65mer**, sSgII-70 ([Table 1](#)). All peaks in the lower spectra are predicted to be related to sSgII-70 (S3 Table 1 in [S3 File](#)). The item at 4755.85 is a next-integer match to an sSgII-70 predicted **42mer**. The 9509.01 item is double this, suggesting artefactual homodimerization. The lower panel pattern has been reproduced on more purified uncontaminated upstream precipitate (S3 Fig. 5 in [S3 File](#)), while the upper panel has also been replicated (S3 Fig. 4 in [S3 File](#)). MS courtesy of Carolyn Carr, University of Oxford, Oxford, UK.

<https://doi.org/10.1371/journal.pone.0340980.g004>

Candidate 7500 is eluted in anionex by low salt (0.1–0.2 M NaCl) and high salt (0.8 M NaCl). Bovine ovarian follicular fluid was subjected to the standard purification procedure of spin filtration, gel filtration and anionex (S1 in [S1 File](#) Babraham Method). The final step yielded 24 fractions (F1–24), six of which boasted *m/z* 7–8000 moieties: **F10** (7170, 7470), **F11** (7500), **F19** (7062, 7133), **F20** (7065), **F21** (7368, 7498) and **F24** (7352, 7482). The mass spectrum for Fraction 24 is shown as [Fig 6](#). No pattern is discernible in terms of *m/z* as between low-salt and high-salt forms of Candidate 7500. Concordant with acidity indicated by successful anionex separations was the failure of two laboratories to furnish MS candidates using cation exchange chromatography. Both early and late eluting anionex fractions containing Candidate 7500 reduce the viability of BMCs *in vitro*, producing a ‘double dip’ histogram, though a single late-stage dip became more characteristic over time. In one purification run involving anionex HPLC, 17 fractions of ovine plasma returned these figures in terms of BMC viability (% control), with Candidate 7500 eluting early and late and with the deepest inhibitions emphasized: 98, **72**, 79, 111, 92, 91, 81, 88, 94, 85, 93, 86, 93, **71**, **73**, 90, 77. The low-salt form of Candidate 7500 provided initial assay data (e.g., [Fig 1](#)) and MS findings, before being lost to view. Disintegration of the early form can be suspected, probably because of unappreciated effects of a change of procedure ([S1 File](#) Sheffield Method). The tenaciously bound high-salt anionex form of Candidate 7500 became the focus, providing inhibitory exemplifications *in vitro* relating for example to human breast and prostate tumour cell viability, rat cardiac hypertrophy, vascular smooth muscle growth (rat

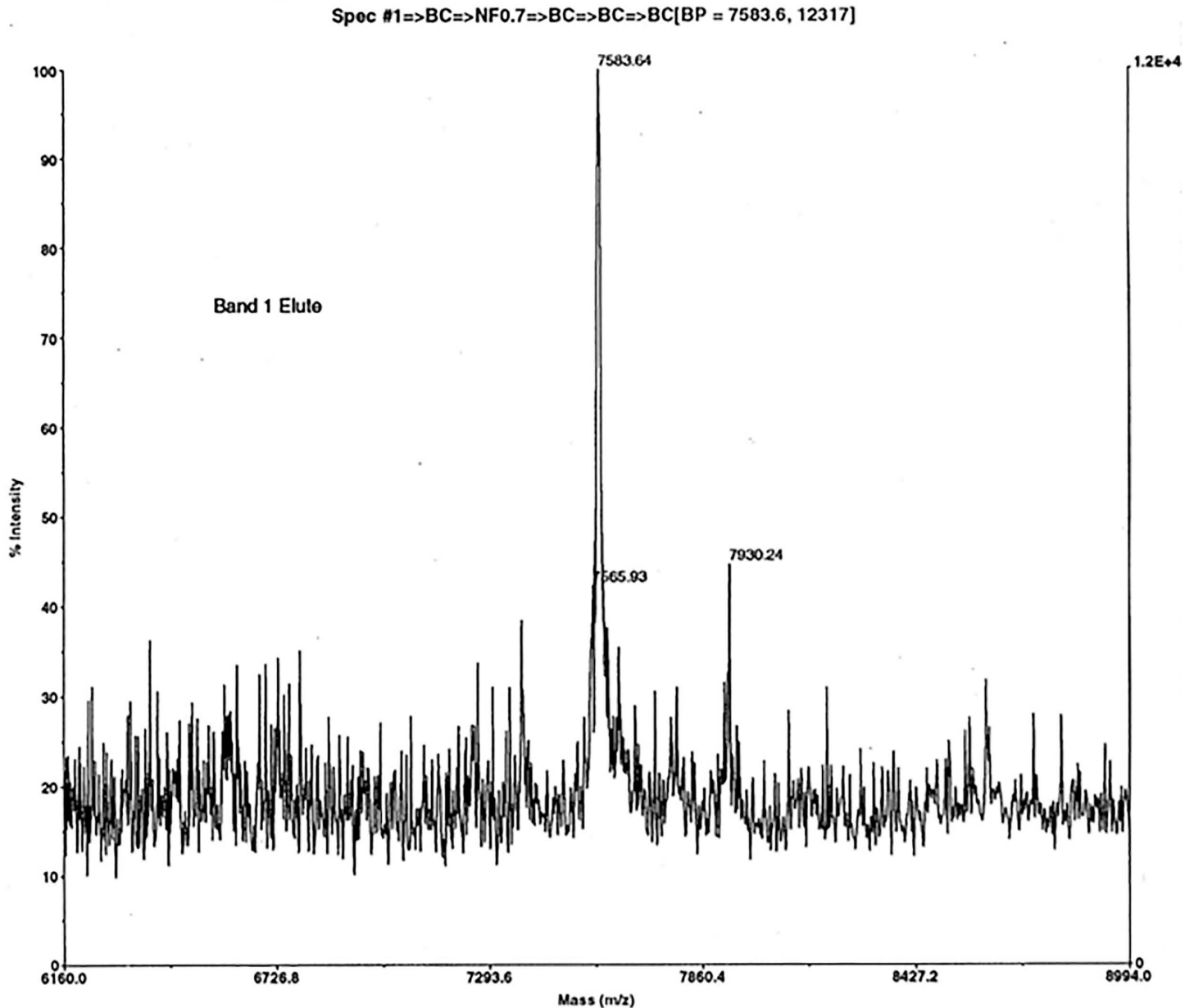


Fig 5. Precipitate with azide. MALDI MS of a solid 'upstream precipitate' that formed on the 3kDa membrane during a 3–30kDa ultrafiltration (with sodium azide) of ovine jugular vein EDTA plasma (S1 in [S1 File](#) Sheffield Method) – without the bacterial contamination of [Fig 4](#). Matrix and calibration as [Fig 2](#). Acquisition mass range 2500–25000 Da. Electroeluted SDS-PAGE Band 1 at ~7 kDa showed cell-inhibitory bioactivity in vitro. Edman sequencing of Band 1 from a prior (contaminated, azide free) purification provided the 14-residue N-terminal sequence MKPLTGKVKEFNNI (EPL001). This matched anionex Beale 4, the minimal sequence obtained from maximally purified (bacterially uncontaminated) material. Gel electroelution, as used here, provides a cleaner spectrum than solvent extraction (S3 Fig. 5 in [S3 File](#)), albeit with the same base peak, an sSgII-70 **65mer**, lacking –KANNI ([Table 1](#)).

<https://doi.org/10.1371/journal.pone.0340980.g005>

aortic stems cell numbers halved over 72h, $P < 0.05$, Bonferroni t test) and angiogenesis ([S2 File](#)). In the last-mentioned application, pooled anionex Fractions 28 & 29 of bovine ovarian follicular fluid containing Candidate 7500 inhibited venule formation in human umbilical vascular endothelial cells to a comparable degree (>50%, ANOVA, $P < 0.05$, significant: [S2 File](#)) to the well-known potent anti-angiogenic compound suramin administered at 20 μ M.

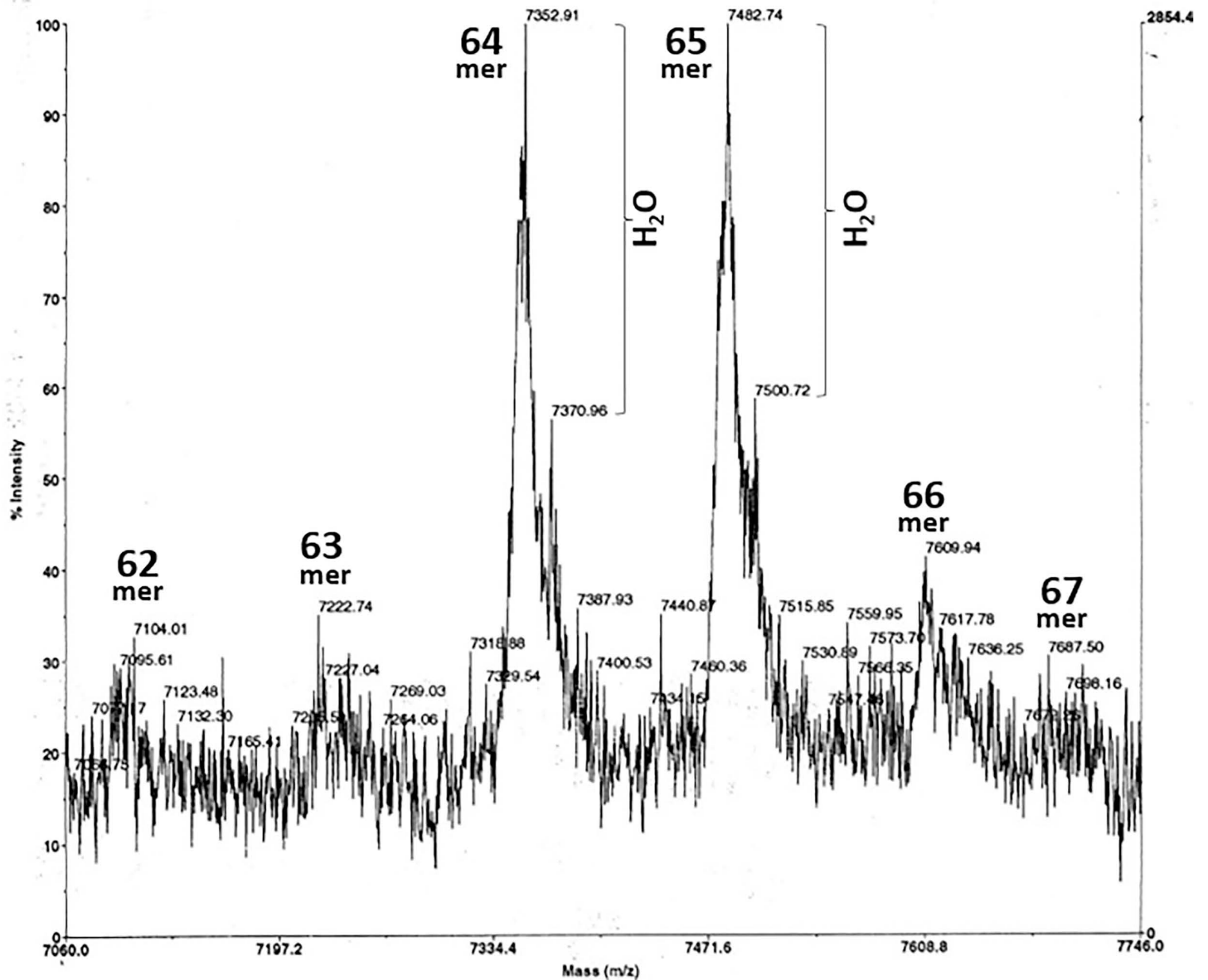


Fig 6. Bovine Candidate 7500. MALDI mass spectrum of bovine ovarian follicular fluid, analysed for variations in bSgII-70 C-terminal truncation and water differences. Anionex Fraction 24 (i.e., late eluting), after spin and gel filtration. The statistical significance of proposed matches are at the >99% confidence level (chi-squared test), as in the sheep (S3 in [S3 File](#)). Matrix and calibration as [Fig 2](#).

<https://doi.org/10.1371/journal.pone.0340980.g006>

When a late-eluting anionex fraction containing Candidate 7500 was subject to C18 reverse phase chromatography (RP-HPLC) a bimodal distribution occurred, as in the prior anionex, in regard to both the presence of Candidate 7500 and BMC inhibition *in vitro*. These bimodalisms indicate that Candidate 7500 is amphipathic. Ovine plasma anionex late-eluting Fraction 41 had MALDI peaks at m/z 2375 > 6967 > 7628. In 17 reverse phase fractions, peaks were seen only in fractions 3–7 and 11–16. In the first wave the peaks were 5400 > 7626 > 2251, in the second 5400 > 5800 > 7630 > 2255 (S3 Fig 3 in [S3 File](#)). Disintegrations of Candidate 7500 were deemed to have occurred, providing fragments that were themselves bimodally distributed. A separate study involved evaluation in two laboratories of the same anionex sample.

Table 1. MALDI MS peak analysis for ovine Candidate 7500 identified as sSgII-70. Analysis based on C-terminal truncation and water losses. Masses are for the native 70mer polypeptide sSgII-70 down to a 60mer (with predictions generated by ExPASy Compute pI/Mw [14] and double-checked via an MS/MS fragment ion calculator: see S3 File). A mean deviation of *m/z* 2.66 is delivered at a mean error (i.e., difference as a percentage of match) of 0.035%; seven integer or next-integer matches (58%); \sum Observed (n=12) 91120/ \sum Calculated Matches 91113 x100=100.01%; and chi-squared P=0.024, significant (S3 Table 5 in S3 File). An extended version of Table 1 is provided as S3 Table 1 in S3 File.

sSgII-70 mer	C-terminal residue loss	M (Da)	+1.008 MH ⁺	-18.015 (H ₂ O)	-18.015 (H ₂ O)	-18.015 (H ₂ O)	-18.015 (H ₂ O)	-18.015 (H ₂ O)	-18.015 (H ₂ O)	-18.015 (H ₂ O)	-18.015 (H ₂ O)
70	–	8175.01	8176.02	8158.00	8139.99	8121.97	8103.96	8085.94	8067.93	8049.91	8031.90
69	–1	8044.84	8045.86	8027.84	8009.83	7991.81	7973.80	7955.78	7937.77	7919.75	7901.74
68	-NI	7930.74	7931.75	7913.73	7895.72	7877.70	7859.69	7841.67	7823.66	7805.64	7787.63
67	-NNI	7816.64	7817.65	7799.63	7781.62	7763.60	7745.59	7727.57	7709.56	7691.54	7673.53
66	-ANNI	7745.56	7746.57	7728.55	7710.54	7692.52	7674.51	7656.49	7638.48	7620.46	7602.45
65	-KANNI	7617.38	7618.39	7600.37	7582.36	7564.34	7546.33	7528.31	7510.30	7492.28	7474.27
64	-YKANNI	7454.21	7455.22	7437.20	7419.19	7401.17	7383.16	7365.14	7347.13	7329.11	7311.10
63	-IYKANNI	7341.05	7342.06	7324.04	7306.03	7288.01	7270.00	7251.98	7233.97	7215.95	7197.94
62	-DIYKANNI	7225.96	7226.97	7208.95	7190.94	7172.92	7154.91	7136.89	7118.88	7100.86	7082.85
61	-DDIYKANNI	7110.87	7111.88	7093.86	7075.85	7057.83	7039.82	7021.80	7003.79	6985.77	6967.76
60	-EDDIYKANNI	6981.76	6982.77	6964.75	6946.74	6928.72	6910.71	6892.69	6874.68	6856.66	6838.65
Figure	OBSERVED		CALCULATED		DIFFERENCE		DEVIATION		% ERROR		
5	7930.24		7931.75		-1.51		1.510		0.019		
3	7878.47		7877.70		0.77		0.770		0.010		
3	7805.98		7805.64		0.34		0.340		0.004		
3	7709.72		7709.56		0.16		0.160		0.002		
4	7709.42		7709.56		-0.14		0.140		0.002		
4	7628.30		7620.46		7.84		7.840		0.103		
5	7583.64		7582.36		1.28		1.280		0.017		
5	7565.93		7564.34		1.59		1.590		0.021		
4	7561.42		7564.34		-2.92		2.920		0.039		
2	7484.53		7492.28		-7.75		7.750		0.103		
4	7161.09		7154.91		6.18		6.180		0.086		
2	7102.24		7100.86		1.38		1.380		0.019		
TOTAL	91120.98	TOTAL	91113.76	TOTAL	7.22	TOTAL	31.86		0.426		
MEAN	7593.42	MEAN	7592.81	MEAN	0.60	MEAN	2.655		0.035		
Observed vs Calculated as %			= 100.008								

<https://doi.org/10.1371/journal.pone.0340980.t001>

The focus was on ovine follicular fluid anionex late-eluting Fraction 29. This was subject to MALDI (matrix sinapinic acid, calibration against carbonic anhydrase, as for the reverse phase analysis) at Babraham. There was a main peak at 7095, medium peaks at 5454 and 4273 and a small peak at 7896. Fraction 29 was then analysed shortly afterwards at Harwell Laboratory, again using MALDI (insulin calibration). There was a main peak at 4258, with subsidiary peaks at 4387, 4306, 2946 and 2479. Peaks in the 7000s were absent. This supports the interpretation that Candidate 7500 is labile and that major subsidiary peaks represent ‘Grand Fragments’ of Candidate 7500, a view that will be developed fully later on (see ‘sSgII-70’).

A uterus-only version of the rat organometric assay *in vivo* was carried out on a 3–30kDa fraction of sheep ultrafiltered plasma subject to anionex HPLC (S1 File Sheffield Method). The zero-baseline uterine weight was provided by rats injected intraperitoneally with PBS for four days. Three groups of fractions lacking Candidate 7500 (validated by MALDI)

were tested: pooled anionex Fractions 1–11 were associated with mean relative uterine weights *below* baseline by 7%; Fractions 12–22 were *above* baseline by 6%; and Fractions 25–32 were also *above*, by 8%. None of these findings was significant (ANOVA). Fractions 23 & 24, unlike the others, contained (late-eluting) MS Candidate 7500. Pooled, these fractions were associated at termination with a uterine weight 22% *below* that of controls (n=7, P<0.002, significant; S2 Fig 8 in [S2 File](#)).

The loss of one kidney causes the other to grow bigger [15]. This is compensatory renal growth, CRG. The surviving kidney grows not to the former joint renal mass but to the former joint glomerular filtration rate. The residual kidney is under upregulatory pressure, notably from insulin-like growth factor 1 (IGF-1) [15]. As an experimental system this is a 'boosted organ' model, enabling direct administration into the organ of interest. Sensitivity is thus potentially greater than the 4-day rat organometric assay. Candidate 7500 material significantly reduced CRG in the unilaterally nephrectomised male rat (S2 Tables 3 & 4 in [S2 File](#)). Late running anionex HPLC fractions were used, containing and lacking Candidate 7500 (validated by MALDI). As much Candidate 7500 was used as was present in 360 ml of ovine pooled ovarian follicular fluid. Administration for 7 days was via an osmotic infusion mini-pump implanted into the residual kidney's pelvic region. Values for the residual kidney (i.e., post-treatment) are expressed in terms of the removed kidney (i.e., pretreatment) as % change (on the established assumption of equal starting sizes), mean \pm SEM (n=4, t test, equal variance, all findings significant except as indicated), control versus test: kidney wet weight, 41.7 \pm 2.7 vs 23.2 \pm 2.6 (P<0.005, significant); dry weight, 25.9 \pm 5.9 vs 13.5 \pm 1.1 (P<0.012, significant); protein, 65 \pm 49 vs 28 \pm 14 (P<0.2, ns); and DNA, 81 \pm 23 vs 42 \pm 15 (P<0.05, significant). Taking these four parameters together, CRG was almost halved by Candidate 7500. In a cardiovascular sense the heart is 'downstream' from the renal infusion point. The cardiac wet weight was >10% lower in the Candidate 7500 group than in controls (P<0.01, significant; S2 Table 3 and S2 Fig 8a in [S2 File](#)). No other organs were assessed.

Sheep plasma in one series was subject to the standard purification procedure of spin and gel filtration followed by anionex ([S1 File](#) Babraham Method). In six separate fractionations involving ovarian venous plasma Candidate 7500 was seen in MALDI. In contrast Candidate 7500 was *not* seen in jugular vein plasma from two OVX ewes, supporting the factor's ovarian provenance. In the search for an inhibitory *gonadal* influence, then, Candidate 7500 is fittingly 'ovary plus' and 'ovariectomy minus': OV+/OVX-. Candidate 7500 was present in the blood of hypophysectomised ewes one day and seven days post-surgery.

Candidate 7500 is higher in the follicular phase of oestrus, suggesting an appropriate reproductive relatedness ([S2 File](#)). Ovine blood serum was collected at different phases of the oestrous cycle and purified in the usual way to anionex HPLC fractions. There was a peak of Candidate 7500 in the follicular phase with a second smaller peak in late luteal. In MALDI the combined counts for Candidate 7500 in late-eluting fractions 19–21 were: follicular, ~13,140; early luteal, ~1770; late luteal, ~4570. Inhibitory activity in the BMC viability assay *in vitro* is stronger in sheep blood from the follicular phase than that from the luteal phase. Ovine samples collected in Spring were higher in Candidate 7500 and more active *in vitro* than were samples collected in Summer or Winter.

During an upscaled 3–30 kDa ultrafiltration purification, a precipitate formed on the 3 kDa filter membrane ([S1 File](#) Sheffield Method). The supernatant, designated Sheep Plasma Filtrate, was inhibitory in the BMC assay, as was the solid precipitate – dubbed 'upstream precipitate' – mixed together with cell medium. MALDI of whole upstream precipitate started by providing Candidate 7500 ([Fig 4](#)). Washing the solid caused the loss of Candidate 7500 (S1 Fig 11 in [S1 File](#)). An aqueous extract of upstream precipitate was actively inhibitory *in vitro*. The bioactivity thus followed the whereabouts of Candidate 7500, with an aqueous extract of the upstream precipitate providing the Edman sequence MKPLT/GKVKxF-NNIK/IGF/Y/DxF/VI/VI, which is denoted the 'EPL001 Extension' (S1 Table 1 in [S1 File](#)).

Modelling of bioactivity *in vitro* saw Candidate 7500 anionex fractions inhibit cardiac hypertrophy by >20%, compared with fractions lacking Candidate 7500 [18]. Measured in cultured myocytes from neonatal rats was the RNA ratio of atrial natriuretic factor, a marker of hypertrophy, to glyceraldehyde 3-phosphate dehydrogenase, a housekeeping enzyme, at

the end of a 24h exposure period. Microscopic analysis showed no extra cell death in the Candidate 7500 group but suggested a reduction in cell size. In other exemplifications *in vitro* there were fewer cells and they were smaller, with reduced cell division accompanied by cell death due to physiology (apoptosis) rather than toxicology (necrosis) [16]. One interpretation of these data is that the ovary elaborates into the circulation a cell shrinking agent, Candidate 7500.

Cell fate has been assessed in the rat bone marrow cell assay by various methods [17], including cytometry using an Annexin-V-Fluos Staining Kit (Sigma-Aldrich, Dorset, UK). Among cells exposed to sheep test plasma anionex fractions (i.e., containing Candidate 7500) pre-apoptotic and apoptotic cells comprised about 60% of the population, compared with 3–4% in controls, with both healthy and necrotic cells less prevalent in the test population. A dose-response curve for apoptosis has been obtained. This was in regard to the hypothalamic aspect of the inhibitory hormonal proposition [7]. An anion exchange fraction of aqueous extract of rat hypothalamus was evaluated via image analysis for proliferative and apoptotic action on a mixed population of rat bone marrow cells *in vitro* [17]. The late Candidate 7500 fraction was used, eluting at 0.8 M NaCl. Control cell numbers increased in numbers by more than 40% in a 24h period. Test cells receiving a single dose of the anionex fraction maintained a steady population. Among test cells exposed to a double dose of this fraction death began to occur after 5h, with 90% of the cells having undergone apoptosis by 24h, as judged morphologically. A triple dose of the fraction caused apoptosis to begin after 90 minutes. In under 4h all the cells were dead (personal communication, Ben Ferneyhough, Systems Biology Laboratory UK, Abingdon, Oxfordshire, UK). As well as being observed in primary cultures of rat bone marrow cells, apoptosis has been induced by Candidate 7500 material in a mouse bone marrow cell line (SR4987) and in a variety of human tumour cells lines (e.g., MG-63, MCF7, PC3, DU145) [18]. Candidate 7500 preparations tend to deliver stoichiometric (linear) dose-response curves rather than hormone-style sigmoidal shapes (S2 File).

Attempts were made to elucidate the aa sequence of Candidate 7500 via digestion with trypsin (porcine), which cleaves selectively on the C-terminal side of lysine and arginine residues, together with MS analysis and interrogation of peptide mass fingerprint databases. No convincing hits were seen in a campaign focussed on Candidate 7500 anionex fractions from 13 purification runs, three species (sheep, cow, pig), two source materials (ovarian follicular fluid and blood plasma), multiple MS modalities (MALDI-TOF, Delayed Extraction-MALDI-TOF, QTOF, LCQ Deca XP, ESI-QUAD-TOF and LC-MS/MS) and different online search tools such as Mascot [19] and MS-Fit [20]. A carboxypeptidase was deployed on ovine ovarian follicular fluid fractions to achieve C-terminal truncation in a MALDI study, without productive outcome. A MALDI mass spectrum of a tryptic digest is shown as Fig 7. This looks unpromising. It is the product of intense lasering: 200 shots/spectrum, as against 100 for Figs 2–3 and 50 for Fig 5, the other ‘Babraham Ovines’. There are excessive matrix peaks on the left (m/z 800s) consistent with analyte in low abundance, plus dominant obvious contaminants. Yet this spectrum yields matches within a bespoke manual analysis (see ‘sSgII-70’).

The effect of proteolysis was investigated further when sheep serum was ultrafiltered and subjected to anionex HPLC (S1 File Sheffield Method). Control fractions containing no Candidate 7500 provided a baseline viability level in the BMC assay *in vitro* over 72h, with samples in triplicate. Viability was unchanged when these fractions were exposed to porcine trypsin. Fractions containing Candidate 7500 were associated with mean cell numbers of 18.5% of the control value (i.e., 81.5% reduction in viability). Trypsinisation of these fractions resulted in cell numbers of 42.2% (i.e., 57.8% reduction in viability). Inhibitory bioactivity was thus reduced by trypsin, albeit incompletely, indicating that Candidate 7500 is proteinaceous but well defended against enzymatic attack. (Similarly heat treatment, 56°C for 1h, only caused partial deactivation, reducing cell death from 88.6% to 54.5%. Charcoal likewise partially reduces activity *in vitro*.) Ovine plasma still showed antiorganotrophic activity in the rat multi-tissue assay *in vivo* even after exposure between gel filtration and anionex to agarose-bound bacterial protease (P8790, Sigma-Aldrich, Dorset, UK). Resistance to proteolysis (and heat) is a notable feature of Candidate 7500.

Bioassay-guided fractionation provided proteinaceous Candidate 7500, for which there are multiple incidence-bioactivity correlations, notably OV+/OVX–. The totality of the evidence marshalled here indicates that Candidate 7500 accounts for the tissue-mass inhibitory reproductive hormonal influence. What is Candidate 7500’s molecular identity?

Spec #1=>BC=>BC=>NF0.7=>MC[BP = 862.3, 42824]

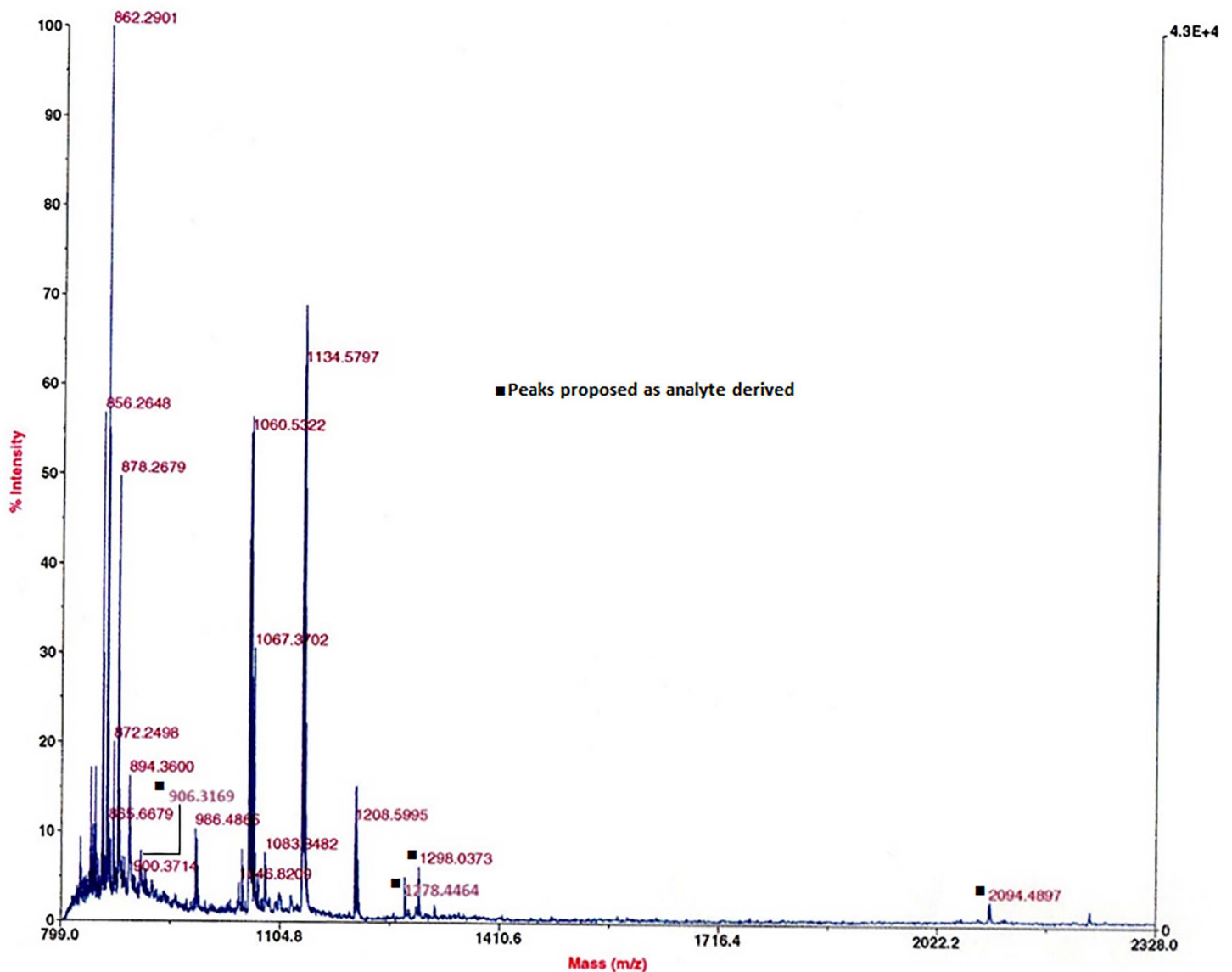


Fig 7. Tryptic digest. sSgII-70 glimpsed in a MALDI MS involving ovine ovarian follicular fluid, anionex Fraction 28 containing Candidate 7500, digested with porcine trypsin. Positive ion, reflectron mode, delayed extraction, CHCA matrix, high laser power. Prominent are matrix cluster peaks in the m/z 800s and other irrelevancies (e.g., keratin at 1060 & 1134: Mascot Contaminants [19]). Predicted fragments from a tryptic digestion in silico of sSgII-70 find minor peak matches, as per black blocks: (i) Observed m/z 906.3169, Predicted 906.0693, Difference +0.2476 (0.03% above Predicted), sequence $(3)K$ TGEKPVFK(R_{12}), no missed cleavage; (ii) Observed 1278.4464, Predicted 1278.5954, Difference -0.1490 (0.01% below Predicted), $(1)MLKTGEKPVFK(R_{12})$, one missed cleavage; (iii) Observed 1298.0373, Predicted 1294.4123, Difference +3.6250 (0.28% above Predicted), being either $(38)K$ LTGPNNQKHER(A_{50}), as the Predicted, or the MSO version of 1278 (at 1294.5948), both one missed cleavage; and (iv) Observed 2094.4897, Predicted 2091.1458, Difference +3.3439 (0.16% above Predicted), $(49)R$ ADEEQKLYTDEDDIYK(A_{67}), one missed cleavage. Items (i) and (ii) relate to a proposed reverse splice 9+61 junction at $(9)VF_{10}$. A fifth match is moot, as being at higher (but still low) intensity within the matrix ion group: Observed 872.2498, Predicted 871.9682, Difference +0.2816 (0.03% above Predicted), $(38)K$ LTGPNNQK(H_{47}), no missed cleavage. (S3 File, Digest 5.)

<https://doi.org/10.1371/journal.pone.0340980.g007>

EPL001

Candidate 7500 is a MALDI candidate from the physicochemical purification but also an SDS-PAGE candidate, being ostensibly responsible for providing a gel band at c7.5kDa ([S1 File](#)). From this band was obtained ovine aa data via chemical sequencing.

The canonical N-terminal aa sequence MKPLTGKVKEFNNI was designated EPL001, which proprietary designation also refers to a 14mer synthetic peptide of the same sequence. Derived by automated Edman degradation (Applied Biosystems Procise, Foster City, CA, US [\[21\]](#)), the EPL001 sequence was seen in whole or in part seven times in separate purification runs involving sheep material that was biologically active (i.e., antiorganotrophic in vivo and antiproliferative in vitro) and contained MALDI Candidate 7500 (S1 Table 1 in [S1 File](#)). Notable landmarks are an initial methionine (M), a position-3 proline (P), a mid-sequence valine (V) and a position-11 phenylalanine (F). Five of the seven sequences obtained have enough in common to support a mathematically validated association in the manner of a Fourier transform [\[22\]](#). Ostensibly meaningful, EPL001 is yet obscure bioinformatically, garnering no compelling hits in innumerable BLAST searches [\[17\]](#). Chemical sequencing [\[23\]](#) was intended to disclose the character of Candidate 7500, but instead provided a conundrum in its own right.

Edman sequencing and MALDI at Babraham were subjected to an internal calibration demonstrating accuracy by virtue of a contaminant, fibrinopeptide A ([S1 File](#)'s Candidate VI, discounted through contra-activity). An Edman reading, with no subsidiary findings, matched the expected database sequence to the extent of 13 of 14 N-terminal residues, while a MALDI peak from the same anionex fraction at m/z 1847 provided a match to a database prediction for ovine fibrinopeptide A of 1848Da, protonated as MH^+ to 1849 (S3 Fig 14b in [S3 File](#)).

EPL001 has proved resistant to molecular biology approaches involving the use of oligonucleotide probes and RT-PCR to find matching sequences in DNA and RNA libraries (e.g., sheep ovary and sheep hypothalamus) and the use of anti-EPL001 antibodies (see next section) to identify cDNA synthesized proteins.

Evidence relating to (ovine) EPL001 is absent from the tryptic digest campaign, which involved trypsin fragmentation and MALDI analysis ([S4 File](#)). Eleven of the thirteen tryptic digests carried out involved purified feedstock that was relevantly ovine. A digest with porcine trypsin in silico [\[24\]](#) of EPL001 (MKPLTGKVKEFNNI), at up to one missed cleavage, yields seven predicted fragments ([S4 Table 4 in S4 File](#)). In the MALDI ovine dataset there are no integer matches (i.e., unrounded whole numbers ignoring decimal fractions) to any of these and there are no next-integer hits either (i.e., next unrounded whole number up or down ignoring decimal fractions), except to a predicted fragment at 863Da (VKEFNNI). This latter match is in the form of overly prominent peaks at m/z 861/2 in nine of the eleven trypsinisations, being a known cluster of the matrix chemical CHCA, i.e., α -cyano-4-hydroxycinnamic acid. (This peak is also present in OVX material lacking Candidate 7500, supporting a claim of irrelevance.) Notable is an absence of integer or next-integer matches to the one predicted EPL001 fragment comfortably outside the matrix region and which does not match the mass of any known contaminant: MKPLTGKVK (1002Da at one missed cleavage, with the MSO variant at 1018). Concluding, there is a lack in samples of predicted tryptic fragments of EPL001.

The unproductivity of bioinformatics, molecular biology and tryptic digestion lends three-way support for the conclusion that EPL001 does not represent a true-to-life N-terminal sequence.

EPL001 has been investigated in terms of bioactivity as a synthetic 14mer peptide to which antibodies have been raised. The molecular formula of EPL001 is $C_{73}H_{129}N_{19}O_{20}S_1$, while the molecular weight is 1618.95, with a theoretical pl of 9.70 [\[14\]](#). As a synthetic peptide EPL001 surprised by being anti-organotrophic in its own right, dose-dependently, in an assay in vivo of rat compensatory renal growth after unilateral nephrectomy [\[25\]](#) and also by modulating reproduction in the nematode *Caenorhabditis elegans* [\[26\]](#). Both these activities are aspects of the inhibitory hormone concept [\[7\]](#), with the inhibition of compensatory renal growth seen with EPL001 also having been seen with Candidate 7500, as described earlier and in [S2 File](#). Crucial in CRG is the endogenous mitogen IGF-1 (P05019; UniProt [\[27\]](#) numbering throughout) [\[15\]](#). IGF's effects are evidently counteracted by EPL001 and Candidate 7500, while being potentiated by an anti-EPL001

antibody (see later). Referring to EPL001, the likelihood is low of a project-unrelated peptide having both the predicted activities of antiorganotrophism (tissue shrinkage) and reproductive modulation. Additionally, EPL001 enhanced lifespan in the nematode [26].

EPL001 has been shown to be cell-inhibitory in multiple assays in vitro. The peptide reduced cell division for example in human prostate cancer cells in cultures, indicated by the incorporation of tritiated thymidine, and reduced the proliferation of rat BMCs [22]. Human marrow derived mesenchymal stem cells in culture exposed for 14 days to EPL001 tethered to cover slips were present in ~25% lower numbers than controls [22]. Analysis using fluorescence-assisted cell sorting (FACS) indicated a lower proportion of cells in the later stages of the cell cycle, with 26.95% of control cells positive for Ki67 (nuclear cell cycle marker absent from G₀) versus 18.92% of tests. A lower average cell size is implied by these results. EPL001 thus provides fewer and smaller cells, as do ovine anionex fractions containing Candidate 7500.

Intracerebroventricular infusion of EPL001 in sheep was associated with elevated growth hormone in peripheral blood and reduced prolactin [22]. The 14mer EPL001 peptide thus shows a breadth of relevant bioactivities while presumptively only representing 20% of the sought-for proteinaceous entity of ~70 aa.

Anti-EPL001 antibodies

The raising of antibodies to the 14mer synthetic peptide EPL001 has been described previously [25], including of what became the lead antibody, a goat polyclonal antibody designated G530 [17,28]. In western blots this antibody provided bands in early and late eluting anionex fractions of serum from ovary-intact sheep, in the same incidence pattern as Candidate 7500 (S1 Fig 4 in S1 File). Candidate 7500 is OV+/OVX-; likewise the circulating endogenous antigen of an anti-EPL001 rabbit polyclonal antibody, which antigen in OVX- mode was lacking from 3–30 kDa blood plasma ultrafiltrate of OVX sheep [18]. These correlations support a connection between EPL001 and Candidate 7500.

Western blots using the G530 anti-EPL001 antibody have visualized Candidate 7500 bands in a range of biological materials, including sheep plasma, rat PC12 conditioned media and rat hypothalamic aqueous extract [17]. The synthetic antigen EPL001 is MKPLTGKVKFNNI. Dot blotting with EPL001 'Front 6' and 'Back 8' peptides has indicated that the epitope within EPL001 recognised by the anti-EPL001 goat polyclonal resides exclusively in the Back 8: K₇V₈K₉E₁₀F₁₁N₁₂N₁₃I₁₄. Such an epitope is also what is being seen by the antibody endogenously, as confirmed by preabsorption of ~7 kDa Western bands of rat hypothalamus aqueous extract with the Back 8 peptide but *not* with the Front 6, M₁K₂P₃L₄T₅G₆ [17]. The evidence from IHC peptide preabsorption studies indicates that the epitope within EPL001 of the anti-EPL001 G530 antibody is **KEFNNI** [28]. The epitope within the antibody's mammalian endogenous antigen has been deduced to be a discontinuous version of this: **KE·F·NNI**. A 14x14 sequence grid analysing EPL001 has been described [22] and will be introduced fully in the next section ('Likely SgII related'). The residues of the endogenous epitope provide the plot's southern fringe.

Immunohistochemistry (IHC) images of high specificity and endocrine relevance have been provided by anti-EPL001 antibodies [17]. Hypothalamic staining in the rat was evident in the arcuate nucleus, with individual neurons staining also in the retrochiasmatic nucleus. Staining was apparent in individual neurons in the ovine lateral and ventromedial hypothalamus and preoptic area, with heavy staining in the palisade (neuroendocrine) region of the median eminence, with axonal beading consistent with transport. Staining was not seen in either the sheep or rat pituitary. Staining was evident in human testicular germinal epithelium (also spermatogonia), without appearing in the testicular interstitial Leydig cells, whence testosterone. The rat ovary showed moderate staining in theca, granulosa, follicle cells and follicular fluid. Discrete *neuroendocrine* cells stained in intestinal and other human organs – an interpretation supported by serial-section co-localisation using the neuroendocrine marker chromogranin A. IHC findings were thus consistent with a sought-for hormonal factor. Staining of neuroendocrine relevance was also seen in embryos of the fruit fly (*Drosophila melanogaster*) [17].

Consonant with the IHC findings, ovine median eminences were used at one stage in the purification campaign [22]. An acid extract was subjected to anionex and RP-HPLC. Immunoreactivity was evinced by a single fraction using an

anti-EPL001 goat antibody. Even so, no data were obtained from Edman sequencing or MS/MS. The endogenous antigen to the anti-EPL001 antibody thus retained its mystery, yet the activity of said endogenous antigen was demonstrable in sundry systems using the antibodies. For example, in an assay *in vitro* – method in [17], with 2D video image analysis using ImageJ software – rat bone marrow cell controls increased in individual size by an average of about 11% over 24h. Test cells exposed to a 23% solution of unpurified goat polyclonal anti-EPL001 G530 antibody grew by 29% (personal communication, David Howlett, coauthor). This finding is consistent with the immunoneutralization of a cell-size-limiting antigen, presumably present in the medium. Administration of an anti-EPL001 antibody caused compensatory renal growth to overshoot in a rat assay involving unilateral nephrectomy (S2 File), also consistent with the immunoneutralisation (IN) of an endogenous inhibitor [25].

Based on an anti-EPL001 rabbit polyclonal antibody [17], a prototype ELISA (enzyme-linked immunosorbent assay) has been developed [18]. Human males had 2–3 times as much analyte in blood plasma as human females ($n=5$, t test, $P<0.05$, significant; S2 Fig 29 in S2 File).

Anti-EPL001 polyclonal antibodies have been deployed in immunoaffinity column purifications to identify the endogenous antigen without decisive outcome [22]. The identification impasse was ultimately breached by the following means: (a) antigen capture by immunoprecipitation (IP, with LC-MS) rather than immunoaffinity columns; (b) superseding automated Edman degradation with Orbitrap MS sequencing after trypsinisation; (c) consideration of a broader MW range rather than for example 3–30 or sub-10 kDa material, taking in anything that is water soluble; and (d) use of biological starting materials preserved via formalin fixation crosslinking, with antigen retrieval, since this approach worked in IHC.

‘Likely SgII related’

The sought-for antiorganotrophic hormonal influence is deemed ‘Likely secretogranin II related’. This is a triangulated result by virtue of IP/LC-MS purification, IHC and bespoke bioinformatics. SgII relatedness is the key piece in the factor identification puzzle (Fig 8). SgII is a neuroendocrine protein of the granin family that regulates the biogenesis of secretory granules and is a prohormone.

‘SgII relatedness’ arose on the basis of the G530 antibody’s use in an IP/LC-MS antigen-capture campaign directed at an aqueous extract of rat hypothalamus [17]. The rat hypothalamus is IHC+ and a neat aqueous extract is antiproliferative and proapoptotic in an assay *in vitro* involving rat BMCs, with this inhibition subject to prior immunodepletion by the anti-EPL001 antibody. For antigen capture the same antibody was bound covalently to resin beads, with control preparations preabsorbed with EPL001. Cryopreserved material (i.e., formalin *unfixed*) delivered no credible candidate. More productive was formalin fixation with antigen retrieval, simulating what had worked in IHC. Formalin (formaldehyde) reacts with primary amines on proteins to form partially reversible methylene bridge crosslinks, preserving secondary structure. Presumed factor lability had apparently been countered. A sole candidate emerged from two antigen retrieval approaches, RIPA buffer and citrate buffer. This was identified by the MS software as ‘SgIIvar’, Q8CGL8, a half-sized splice variant of rSgII (P10362, encoded by the gene *SCG2*). On the basis of preabsorption evidence from Westerns this has been interpreted as an intermediate in the intracellular processing towards secretion of an SgII proteoform, where ‘proteoform’ refers to all protein variants of a single gene including post-translational modifications and aa sequence variants. A parallel IP/LC-MS exercise involved embryos of *D. melanogaster*. Brought forth from formalin fixed and antigen retrieved material was the uncharacterised protein Q9W2X8 (Gene: *CG12644*), potentially explaining neuroendocrine staining in this species. The rSgII intermediate proteoform and fly Q9W2X8 display the same partial epitope (NNI) and a version of the rat protein’s MS ID peptide is present in the fly protein. A version of Q9W2X8’s N-terminally located MS ID peptide is lacking from the N-terminally truncated 322-residue rSgII proteoform but is present in the 619-residue full-length rSgII, in the same relative position. Fly Q9W2X8 shows detailed SgII-relevant sequence homology, notably in regard to sorting domains and dibasic residues.

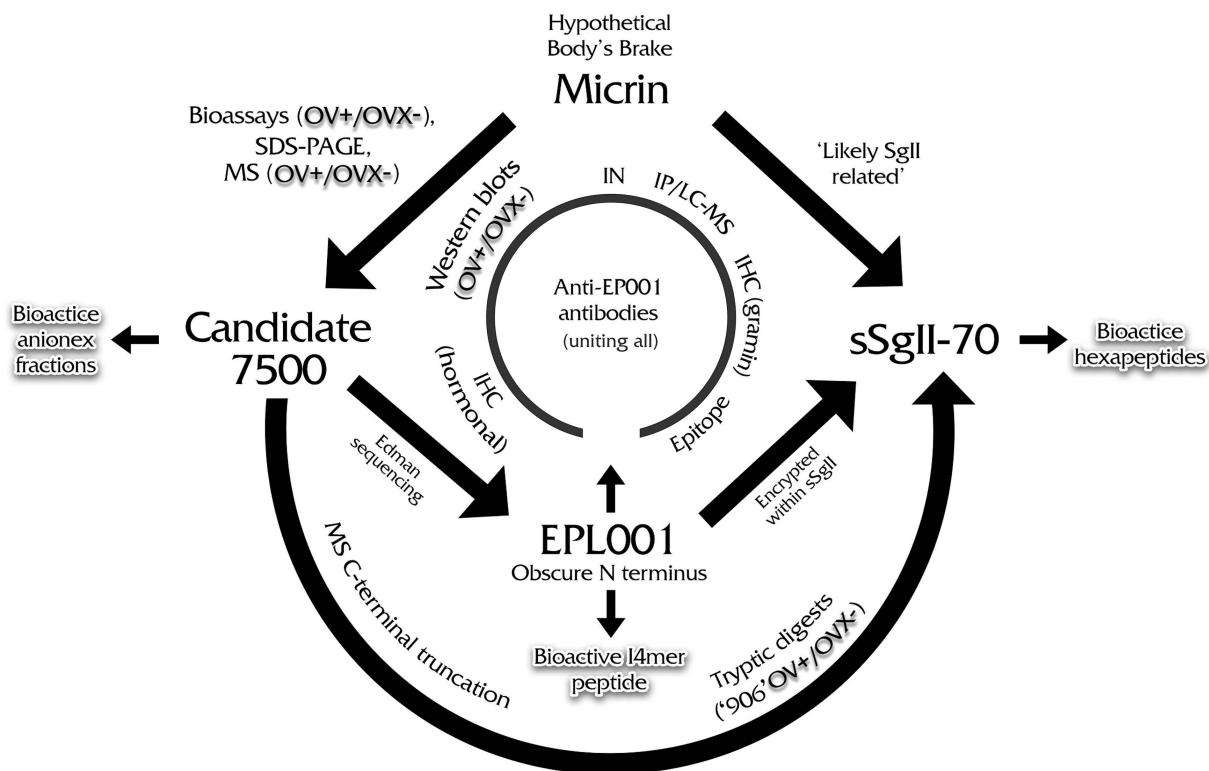


Fig 8. Puzzle. IN = immunoneutralization, IP/LC-MS = immunoprecipitation/liquid chromatography-mass spectrometry, IHC = immunohistochemistry, OV +/OVX- = ovary intact material presenting positively/ovariectomised material presenting negatively.

<https://doi.org/10.1371/journal.pone.0340980.g008>

SgII is a member of the granin family of secretory-granule proteins, which includes chromogranins A and B (CgA & CgB), SgIII and several other granin-like proteins [29]. These entities are involved in the biogenesis and operation of intracellular secretory vesicles as part of the regulated secretory pathway by which neuropeptides and protein hormones are provided for export from endocrine, neuroendocrine and neural cells. Granins are soluble, acidic, heat-stable, calcium-binding proteins that are prohormones in their own right, giving rise to (stoichiometrically) bioactive peptides [30]. SgII is encoded by gene *SCG2*, which in the rat is expressed in the brain, adrenals and testes, as judged by RNA sequencing, without significant expression elsewhere (NCBI Gene ID 24765: <https://www.ncbi.nlm.nih.gov/gene/24765>). The protein is tyrosine-sulphated by way of post-translational modification. SgII exhibits no sequence homologies with CgA & CgB and is poorly conserved compared with them. Only short peptides known as secretoneurin (SN) and EM66, proteolytically processed from the central core region of SgII, show a high sequence similarity among the vertebrates [30]. SN's activity profile includes immune and inflammatory modulation, while EM66 may be involved in modulating food intake. Hypothalamic IHC staining is suggestively similar for an anti-SN antibody [31], an anti-EM66 antibody [32] and anti-EPL001 antibodies [17], notably in regard to a plexus of stained neurons in the median eminence. EM66 visualizes additionally in the rat pituitary [33], while the antigen of the anti-EPL001 antibody does not visualize in the sheep or rat pituitary [17]. Neither peptide can be the factor sought in the present enquiry, as SN & EM66 both lack the full complement of EPL001 residues and neither can in principle be seen (in terms of epitope possession) by the main anti-EPL001 G530 antibody [28].

Granted EPL001's ovine provenance, a bioinformatic analysis has been presented based on that part of the TrEMBL database devoted to *Ovis aries* [28]. The evidence from IHC peptide preabsorption studies is consistent with the mammalian epitope of the G530 antibody used in the IP/LC-MS purification being KE·F·NNI. Of 26,443 ovine predicted proteins 1,100

proteins contain one or more NNIs. These 1,100 proteins comprise the ovine ‘NNI-ome’, with 1,181 NNIs in all. The assumption is that one of these motifs relates uniquely to the NNI in EPL001’s sequence MKPLTGKVKEF**NNI**. Within the NNI-ome’s 1,100 proteins there are 22,207 methionine residues, discounting signal sequence initiator Ms. This is the pool of candidates for the M at the start of EPL001. Ignore NNI and focus on EPL001’s other 11 residues. The Method of Exclusion can be used to eliminate all candidate Mxxxxxxx sequences bearing residues *not* present in EPL001, i.e., ARNDCQHISWY, plus M, as that is used up as the first residue. Of Mxxxxxxx sequences in the ovine NNI-ome 38 are composed exclusively of the 9 unique residues in EPL001-less-NNI, counting repeat aa as one. None has all 9. Five have 8, the rest having from 7 down to three. The five with 8 EPL001-less-NNI residues are as follows, described by similarity with other mammalian proteins: W5Q754 (titin, structural, MLKKTPLVKKG); W5PWS5 (dynein, structural, MLFVGPTGTGK); W5PP00 (ubiquitin-2, nuclear, MPKVVPTLPEG); W5PWP9 (hormonally up-regulated neu tumour-associated kinase, enzyme, MLTGTLPTVE) and W5QEU8 (secretogranin II, secretory vesicle protein, MLKTGEKPV). The stand-out candidate from amongst these five is the last, SgII, an established prohormone, the match being to SgII’s second sorting domain.

Secretogranin II has two main sorting domains that can act independently to direct the protein within the cell’s cytoplasm into secretory granules of the regulated secretory pathway [34]. In the second main sorting domain of sheep SgII there is a curiosity, as is evident from the NNI-ome bioinformatics: a sequence of aa (‘sSgII-9’) that is a 9/14 shuffled version of EPL001, with the latter as ¹MLKPLTGKVKEFNNI₁₄ and sSgII-9 as ³⁶⁷MLKTGEKPV₃₇₅ (Fig 9). Just the two residues emphasized are in register.

Missing for a full ovine match with EPL001 are residues K, F, N, N, I. The last three can be accounted for by the partial epitope ²³⁶NNI₂₃₈. This commands attention, especially since this part of sSgII, with polyanionic groups separated by hydrophobic clusters, can be suspected itself of having sorting relevance [17]. Now with a 12/14 anagram – where an anagram is defined as a word or phrase formed by rearranging the letters of another word or phrase – what of the missing K and F? Among the 43 Ks and 10 Fs in sSgII there is only one pairing: ¹⁷⁸FK₁₇₉ (Fig 9). Taking this doubleton for the sake of argument as the provider of the absent residues the contribution of ‘sSgII-14’ to this admittedly bizarre 14/14 ‘match’ is as follows: ³⁶⁷MLKTGEKPV₃₇₅...¹⁷⁸FK₁₇₉...²³⁶NNI₂₃₈. Notwithstanding the improbability of a match involving the tripartite anagram sSgII-9 + ‘sSgII-2’ + ‘sSgII-3’ being anything other than happenstance, it has been hypothesized that the EPL001

```
>tr|W5QEU8|W5QEU8_SHEEP Uncharacterized protein OS=Ovis aries GN=SCG2 PE=4 SV=1
  1 MAEAKTHWLGAVLSLIPLIFLLSEAEAAASFQFNQLLQKEPDLRLLENVQRFPSPEMIRALE 60
  61 YIEKLRQQAHKKESSPDYNPYQGVSVPLQKQKENGDLPESSRDSLSEDEWMKIIAEALRQA 120
121 ENEPQSAQKENRPYTLNSEKSFPMDDYETQQWAERKLKHMRFPPMYEENS RDNPFKR 180
181 TNEMVEEQYTPQNLATLESVFQELGKLTGPNNQKHERADEEQKLYTDEDDIYKANNIAY 240
241 EDVVGGEDWNPVEEKIESQTQEEVDRDSKENADKTEQINEEMKRSGQLGLQDEDLR KESKD 300
301 QLSDDVSKVIAYLKRLVTAAGSRGSRQNGQTGERATRLFEKPLDPQSIYQLIEISRNLQIP 360
361 PEDLIDMLKTGEKPVEPEQELEIPVEPEDISEVDLDHPDLFQNKMLSKNGYPKAPGHAVA 420
421 EAPPEGLSVEDI LNLGMESAANPKPPYFPNQYNREKVL SRLPYGPGRSKANQLPKAVWM 480
481 PDVENRQMAYENLNDKDQELGEYLARMLVKYPEIMNANPVKRVPSQGSTEDDRQGENQIE 540
541 QALKEHLSQHSSQETDRLASVSKRPLVGTTPKSDDTSNRPYLDLLEDLLVKVLEYLNQEKAEK 600
601 GREHIAKRAMENM
```

Fig 9. Sheep secretogranin II preprohormone. UniProt W5QEU8, FASTA format [27]. Includes sSgII signal sequence (M1-F30; Courel et al, 2008), which is the ‘pre’ in ‘preprohormone’. sSgII is postulated to give rise to a secreted proteoform, sSgII-70, the product of a 9+61 reverse peptide splicing (DB in an A-E modularisation, with residues emphasized). Yet SgII is self-aggregating, admitting the possibility that the 9mer D and 61mer B units of sSgII-70 derive from different SgII molecules.

<https://doi.org/10.1371/journal.pone.0340980.g009>

sequence is a true and meaningful reflection of sSgII structural reality, arising from the Edman machine reading *nonsequentially* [22]. In other words, EPL001 is to be found encrypted within sheep secretogranin II.

The number of anagrams of a 14mer such as EPL001 is $14! = 87.18\text{bn}$. Even though sSgII-14 is discontinuous ($9 + 2 + 3$), it is not a stray anagram of EPL001: it is a light shuffle, constrained informatively. The sSgII-14 residues are a patterned interleaving of the EPL001 residues in regard to the front half of EPL001 (1–7, emphasized) and the back half (8–14): **MLKTGEKPV** · FK · NNI. The first doubleton, ML, has the initial M of EPL001. The next two shaded doubletons, TG and KP, are exact matches for doubletons within EPL001, albeit in reverse order: **MKPLTGKVK**EFNNI. (The probability that TG & KP occur in both sSgII-9 & the EPL001 equivalent 9mer is 1 in 1640, $P = 0.0006$, significant [22].) The last of the three shuffled-sequence modules, sSgII-3, comprises a key epitope component, NNI [28].

sSgII's second sorting domain is $_{367}\text{MLKTGEKPV}_{375}$. Motifs from this present in seven available ovine Edman sequences (S1 Table 1 in [S1 File](#)) are as follows: MLK (as ML/K, with MxLK in an eighth subsidiary sequence: see [S1 File](#) Minor Sighting); TG (as TG or T/G, twice each, with MxxT in register in 'Harwell 1' of S1 Table 1 in [S1 File](#)); KP (thrice); and PV. In addition there is NNI (twice), which is present in sSgII outside its second sorting domain.

Suppose that the sSgII-14 residues MLKTGEKPV · FK · NNI comprise a deck of 14 playing cards which when shuffled yields the EPL001 sequence MKPLTGKVKEFNNI. The 14x14 sequence grid of [Fig 10](#), plotting sSgII-14 along the top and EPL001 vertically, gives the residue matches, such that the shuffled 'deck' (EPL001) has the following face values: 1, 7, 8, 2, 4, 5, 11, 9, 3, 6, 10, 12, 13, 14. The EPL001 deck is said to have a 'signature' of four because it displays that many 'rising sequences' interleaved: 1 2 3; 7 8 9 10; 4 5 6; 11 12 13 14. An arithmetic ordering (i.e., 1–14) can be achieved simply by moving the 7–10 group between 6 & 11. The four groups of cards are patterned, 3-4-3-4. The starters, 1, 7, 4 & 11, are the sSgII-14 'peaks' in the 14x14 grid, in order of decreasing height, with the other cards in each group being the associated 'slopes'. To get to the strangely meaningful EPL001, imagine that the sSgII-14 deck is cut roughly into two halves which are placed on a table and corner-flicked together, effecting what is known as a riffle shuffle. In card playing the aim is to randomize the cards to defeat predictability. The sSgII-14 deck produces the EPL001 deck after (i) a rough halving (6/8), (ii) a riffle shuffle, (iii) a true halving (7/7) and (iv) another riffle shuffle. The transformation from sSgII-70 to EPL001 is as follows, with cuts represented by vertical lines and shuffles by arrows:

$$M_1L_2K_3T_4G_5E_6 \mid K_7P_8V_9F_{10}K_{11}N_{12}N_{13}I_{14} \Rightarrow \underline{M_1K_7P_8L_2} \ V_9K_3F_{10} \mid \underline{T_4G_5K_{11}} \ E_6N_{12}N_{13}I_{14} \Rightarrow$$

$$\underline{M_1K_7P_8L_2} \ \underline{T_4G_5K_{11}} \ V_9K_3E_6F_{10}N_{12}N_{13}I_{14}$$

Adjacent shaded residues stay together throughout; adjacent underlined residues stay together through the second shuffle. The adjacencies argue against a (far-fetched) chemical correlate to a riffle shuffling, with the economy of the transformation sSgII-14 EPL001 supporting the view that the two sequences are non-randomly related. With the signature of sSgII-14 by definition one, the signature of EPL001 after a series of shuffles could be from zero (14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1) to seven, the latter in the form of doubletons: 13 14; 11 12; 9 10; 7 8; 5 6; 3 4; 1 2. A signature of four is what is typically obtained after two shuffles of a deck of 52 cards having equal independent placement probabilities, with eight shuffles and a plateau of 26 or so rising sequences representing adequate randomisation [35]. EPL001 thus has a high signature for two shuffles of a deck of 14, with its four rising sequences moreover covering all cards. What all this means is that EPL001 is far from being a randomly ordered deck. The EPL001 Edman 'card pick' from sSgII-14 takes a very particular non-random form, connoting a spiral structure.

IP/LC-MS, IHC, NNI-ome bioinformatics and sequence and shuffle analyses have vectored the project to sSgII and landed it on sSgII's second sorting domain. If EPL001 is not a shuffled sequence from an Edman-challenging sSgII proteoform what is it? The N-terminus of a novel protein which by coincidence has granin-like characteristics in terms of IHC, acidity disclosed during anionex purification, heat resistance and assay stoichiometry? Or are the seven available EPL001-related Edman sequences of [S1 File](#) in fact just stray nonsense?

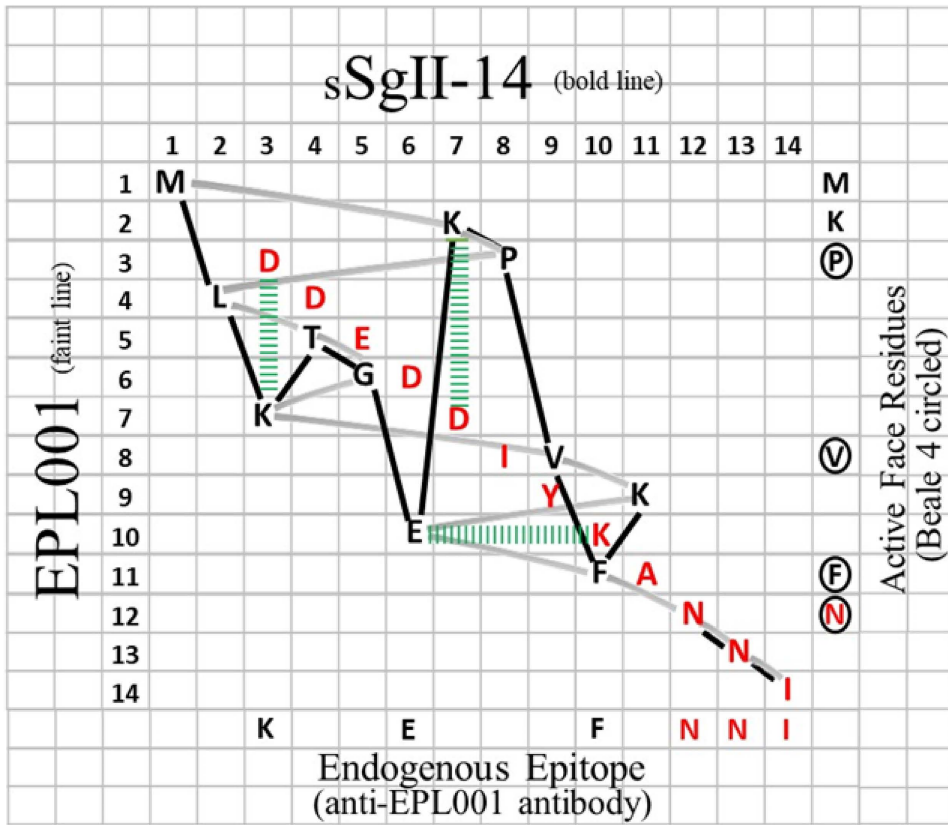


Fig 10. Grid modelling of sSgII-70. In black, N-terminal aa residues M1-K11 and in red C-terminal residues D59-I70. 'sSgII-14' (putative Edman N-terminal *input*) is plotted along the top against EPL001 (actual Edman N-terminal *output*) vertically. Following the bold line yields the aa sequence of sSgII-14 across the grid, column by column. The faint track reads out EPL001 downwards, row by row. For lysine grid placements see [22]. Green ladders represent potential salt bridges. Overlapping ends are implied.

<https://doi.org/10.1371/journal.pone.0340980.g010>

By way of location, reference can be made to SgII in the rat. The second main sorting domain in rSgII comprises residues 366–380. It is situated in the wide terrain of primary structure between, on one side, the proteolytically derived peptides SN (184–216) & EM66 (219–284) and, on the other, a peptide product called manserin (529–568) [36]. The EPL001/sSgII-14 anagram does not work in the rat or for that matter the human because of SgII-9 sequence variations: sheep MLKTGEKPV; rat MLKAGEKPN; human MLKTGEKPN. But given the *ovine* provenance of the EPL001 sequence these non-matches are immaterial.

The first of the seven available ovine Edman sequences is the 'First Sighting' (S1 Table 1 in S1 File). Derived from maximally purified (anionex) ovine ovarian follicular fluid this has the sequence MMxV?xPVG?GxFL. The First Sighting is a seeming prefigurement of EPL001 and 'likely SgII relatedness' and the identification of the protein Q9W2X8 as the fruit fly IHC antigen. There is a rough correspondence in the form of M-V-F to EPL001: MKPLTGKVKFENNI. The V in the First Sighting forms part of a PV doubleton, which is seen as such after M in the second sorting domain of sSgII (as the 9mer ³⁶⁷MxxxxxPV₃₇₅) and in the homologue thereof in the fruit fly protein Q9W2X8 (as ¹⁰⁴⁸MxxxxxxxxxPV₁₀₆₁). (The PxxxxV of EPL001 on the other hand finds an in-register match with the 'Beale 4': xxPxxxxV/LxxF/KNxx, with L and K as 'under signals': S1 Table 1 in S1 File). The fruit fly's homologue of SgII's second sorting domain is within this sequence: ¹⁰³⁸PETLDFWQREMASKAGKTRNNQPVGS₁₀₆₃. Emphasized is the anchoring doubleton ¹⁰⁴⁸MxxxxxK₁₀₅₄, which matches ³⁶⁷MxxxxxK₃₇₃ in sSgII and ₁MxxxxxK₇ in EPL001. The recognition of the importance of non-contiguous amino acid pairs

that are evolutionarily connected has propelled understanding in the field of protein modelling [37]. The M & K of sSgII's **MxxxxxK** are the main peaks in the Fig 10 grid comparing sSgII-14 and EPL001 and are analysed to be the first two residues to be read in Edman sequencing. The first and second sorting domains in the fly are self-similar in a way that is not the case in mammals. The match for the fly sorting domains is this: ^{45/1049}**AxKAGKxxxxQ**_{55/1059}, with KAG a match for the second sorting domain of rSgII.

The 14x14 sequence grid of Fig 10 compares sSgII-14 and EPL001, with sSgII-14's 1–14 numbers along the top and EPL001's 1–14 vertically. sSgII-14's initial methionine is found at the start of EPL001, providing an M in the joint top left-hand corner of the grid. In the second column for sSgII-14 is leucine, **MLKTGEKPVFKNNI**, but this appears in the fourth row to match EPL001: **MKPLTGKVKEFNNI**. Matching the residues in this fashion provides for sSgII-14 an up-and-down zigzag and for EPL001 a side-to-side zigzag, representing the stepwise Edman reading order of the sSgII target molecule. A molecular model in silico of sSgII-9 has the same general shape as the 9mer part of the sSgII-14 line on the grid (S1 Fig 3 in S1 File). EPL001 would appear to be an encoding with two aspects: of the complement of aa and of their relative positions in space. This explains why an antibody raised against EPL001 can see an sSgII proteoform.

SgII-70

It has been asserted that there is a secreted 70mer polypeptide among the proteoforms of the sheep secretogranin II gene, *SCG2* [22]. The deduced aa sequence of this 70mer polypeptide is represented in full in Fig 11. This is sSgII-70, pronounced 'sheep sig two seventy'. (The bovine equivalent is bSgII-70, rat rSgII-70, human hSgII-70.)

sSgII-70 aa sequence:

³⁶⁷₁MLKTGEKPVFKRTNEMVVEEQYTPQN**L**ATLESV**FQELG**KLTGPNNQKHERADEEQKLYT**D**DEDDIYKANNI²³⁸₇₀

Edman EPL001 (& Harwell 1) residues (S1 Table 1):

^{1 4 9 5 6 10 2 3 8 11 7}₁MLKTGEK**PVFK**₁₁ (20QY₂₁) (31SV₃₂) (37GK₃₈) ^{12 13 14}₆₈**NNI**₇₀

Endogenous IHC epitope recognised by anti-EPL001 goat polyclonal G530 antibody (Howlett et al, 2019):

⁶EK_{7 10}F ⁶⁸NNI₇₀

Active face residues (Hart et al, 2022):

₁M ⁷KPVF₁₀ ⁶⁸N

MALDI-TOF MS N-terminal grand fragments (28mer shown) & C-terminal truncation (this paper; S3):

₁MLKTGEKPVFKRTNEMVVEEQYTPQN**L**₂₈ ⁶³DIYKANNI₇₀

Tryptic fragments analysed by MALDI-TOF MS (*m/z*) (this paper; S4):

₁MLKTGEKPVFK₁₁ (1278) (1294) ₃₉LTGPNNQKHER₄₉
⁴TGEKPVFK₁₁ (906) (2091) ₅₀ADEEQKLYTDEDDIYK₆₆
 (1802) ₅₆LYTDEDDIYKANNI₇₀

Fig 11. sSgII-70 primary structure deduction. sSgII-70 superscripts are positions in the sSgII master protein (Fig 9). Helix predictions (shaded) by PSIPRED [38] uphold ChemSW [39] modelling in silico (S5 Fig. 1 in S5 File), but only the central prediction is supported by AlphaFold2 [40] (S5 Fig 10 in S5 File). EPL001 residues emphasized are anionex Beale 4. Edman superscripts denote spiralised reading order of residues within sSgII-70 to obtain EPL001. Bracketed Edman items (together with TG, unbracketed) appear thus in Harwell 1 (SEQ ID NO: 6 in S1 Table 1 in S1 File): **MKV/IT/G QY S/V G?K?**. A quadruple doubleton correspondence between Harwell 1 and sSgII-70 is unlikely to be due to chance (with Harwell 1 additionally sharing MK & TG with EPL001). Stepped-down gapped doubleton readings connote spirality.

<https://doi.org/10.1371/journal.pone.0340980.g011>

[Fig 11](#) also displays for sSgII-70 the Edman-derived EPL001 residues and the order in which they were read [17]; the IHC epitope residues recognised by the main anti-EPL001 antibody [28]; and the deduced active face of this inhibitory hormonal polypeptide [22]. MS data from the present paper are also shown.

sSgII-70 would appear to be the product of peptide splicing [28]. This is the '9+61 contention'. Peptide splicing is distinct from RNA splicing. The sSgII prohormone can be modularised as ABCDE, as per [Fig 9](#). The 9 aa module is D, the 61 aa module B. If sSgII-70 is the product of *reverse splicing* from a single molecule of sSgII, then the processed product is DB. If sSgII-70 is the product of splicing from two molecules of sSgII, A1B1C1D1E1+A2B2C2D2E2, then the possibilities are D1B2 & D2B1. (A chain arrangement of aggregated SgII molecules will be considered later.) Within the seven ovine Edman N-terminal sequences obtained (S1 Table 1 in [S1 File](#)) there is an M-F motif, seen once as $_1\text{M-F}_{10}$ and four times as $_1\text{M-F}_{11}$. This supports the 9+61 peptide splicing contention, with the 9-residue portion starting with M1 and the 61-residue portion with F10.

Further supporting the 9+61 peptide splicing contention for sSgII-70 is a re-evaluation of the evidence from tryptic digests, this time not driven by the fingerprint databases of already discovered and predicted proteins (see 'Candidate 7500'). The analysis involves subjecting the proposed 9+61 arrangement of sSgII-70 to a tryptic digest *in silico* [24] and searching for the predicted fragments in the MALDI data from the 13 trypsinisations attempted on Candidate 7500 material – and arguably finding examples, although *de novo* sequencing via MS/MS proved fruitless ([S4 File](#)). A tryptic digest spectrum is shown in [Fig 7](#) (with the peak list in [S4 File](#) as Digest 5). The feedstock for this digest was an anionex fraction of ovine ovarian follicular fluid subjected to prior spin and gel filtrations. On the face of it this spectrum is discouragingly prolific, yet discernible in the understorey are four, maybe five, fragments of sSgII-70, as per the [Fig 7](#) legend. The peaks are in series in terms of intensity and a statistical analysis (of all five hits) supports a true correlation, with chi-squared $P=0.016$, significant (S4 Table 3 in [S4 File](#)). Ovine MS candidate matches cover all trypsin-available residues in sSgII-70's proposed primary structure ([Fig 11](#)). Of 344 mass assignments in the ovine trypsinisation series a maximum of 25 (~7%) might be sSgII-70 related (S4 Table 2 in [S4 File](#)). Crucial are two integer matches to sSgII-70's proposed N terminus, at m/z 906, $_4\text{TGEKPVFK}_{11}$, and 1278, $_1\text{MLKTGEKPVFK}_{11}$. Both of these uphold the peptide spliced 9+61 junction at $_9\text{VF}_{10}$. The integer 1278 item was a singleton in the 11 ovine trypsinisations but the integer 906 signal appeared four times (= 36% prevalence), more than any item not a matrix cluster, trypsin autolysis product or keratin. The four digests producing 906 involved three different feedstocks purified in two different ways ([S1 File](#) Babraham Method & Harwell Method) and analysed by two different MALDI mass spectrometers. Notably, 906 was present in ovine ovarian venous plasma but absent from jugular vein plasma from ovariectomised controls ([S4 File](#) Digest 1), a result that is concordant with that for Candidate 7500, which is likewise OV+/OVX–. (The endogenous circulating antigen of the anti-EPL001 antibody is also OV+/OVX–, being present and absent respectively in 3–30 kDa blood plasma ultrafiltrate of OV/OVX sheep). The MS for 906-producing Digest 1 was conducted by an external provider. The peak list (S4 Fig 6 in [S4 File](#)) bears this comment: '... peaks consistent with autodigestion products of porcine trypsin, internal standards and commonly observed keratin contaminants or observed in the [gel] blank are noted'. The 906 signal was unannotated, as deemed analyte related.

The predicted mass of sSgII-70 is 8175.01 Da, using average isotopic masses [14]. The protonated version of this, MH^+ , is 8176.02, with the addition of a hydrogen ion at 1.008 Da. Such a peak has not been observed with any modality of MS with any starting material, though in MALDI a **70mer** monomer-by-deduction has been seen in a putative quadruple homodimer (2+2+2+2) in a stored sample (see later, where is discussed S3 Fig 10 in [S3 File](#)). In the [Fig 5](#) MALDI spectrum of EPL001 Gel Band 1 the main peak is at m/z 7583.64. Subtracting sSgII-70's C-terminal –KANNI by peptide bond cleavage and *two* water molecules (with each H_2O at 18.015 Da) provides a **65mer** next-integer match of 7582.36, with a difference of +1.28 and an error (difference as percentage of match) of 0.02% ([Table 1](#)). Remove a further water molecule and [Fig 5](#)'s subsidiary peak at 7565.93 finds a **65mer** next-integer match in [Table 1](#) at 7564.34, a difference of +1.59 and an error again of 0.02%. Alternatively, the same two peaks can be considered in terms of potential N-terminal losses from sSgII-70, again by peptide bond cleavage. For a match to the peak at 7583.64, subtracting MLKTG– and *three* waters

provides a **65mer** match at 7590.28, difference -6.64 at an error of 0.09% (S3 Table 2 in [S3 File](#)). The peak at 7565.93 finds a **65mer** match *one* water below, with a difference of -6.34 at an error of 0.08% (S3 Table 2 in [S3 File](#)). What is true for the two peaks in [Fig 5](#) is true for all the 7–8000 peaks in ovine [Figs 2–5](#), taken together ($n = 12$, [Table 1](#)): MALDI data modelling based on C-terminal aa residue losses provides a closer fit for the ovine data than N-terminal residue losses ([S3 File](#)). Seven of the 12 ovine peaks find integer or next-integer matches (58%) in the C-terminal truncation analysis, three (25%) in the N-terminal truncation series.

The superior data fit for *ovine* SgII-70 of C-terminal truncation over N-terminal truncation also holds true for the *bovine* series. Peaks of m/z 7–8000 ($n = 13$) are analysed from [Fig 4](#) (top panel, 3 peaks), [Fig 6](#) (8 peaks) & S3 Fig 1 in [S3 File](#) (2 peaks). There are three residue differences between the ovine and bovine SgII-70 aa sequence predictions: M16 in ovine [Fig 11](#) becomes bovine I16; N44 becomes S44; and H47 becomes R47. The predicted m/z scores are thus different throughout for the 7–8000 peaks. *C-terminal truncation* in the bovine analysis (S3 Table 3 in [S3 File](#)) delivers a mean deviation of **2.99** with a mean error at **0.04%**; two integer or next-integer matches (**15%**); $\sum \text{Observed} (n = 13) 97097 / \sum \text{Matches} 97117 \times 100 = \mathbf{99.98\%}$; and a chi-squared $P = 0.027$, significant. *N-terminal truncation* relating to the same bovine peaks (S3 Table 4 in [S3 File](#)) delivers in contrast a mean deviation of **5.13** with a mean error at **0.07%**; one next-integer match (**8%**); $\sum \text{Observed} (n = 13) 97097 / \sum \text{Matches} 97073 \times 100 = \mathbf{100.03\%}$; and a chi-squared $P = 0.057$, ns.

The MS analysis amounts to a C-terminal sequencing that upholds the concluding part of SgII-70 for both ovines and bovines as ${}_{63}\text{DIYKANNI}_{70}$. (The same conclusion holds for the pig, on the basis of a sole available spectrum, S3 Fig 2 in [S3 File](#), in which an apparent pSgII-70 **68mer** peak at m/z 7875.58 finds a predictive next-integer match at 7876.72, with *four* water losses, and an apparent **61mer** at 7088.26 finds a match at 7092.89, with *two* water losses.) Consider the upper panel of [Fig 4](#). The difference between the two peaks, both relating to SgII-70 in *bovine* ovarian follicular fluid, is equivalent to an alanine residue. The 7643.93 item relates to a **67mer** ending in $-\text{DIYKA}$; the 7573.19 item to a **66mer** ending in $-\text{DIYK}$ (S3 Table 2 in [S3 File](#)). During one purification campaign using *ovine* jugular vein plasma as feedstock in a Babraham style approach ([S1 File](#)), with 1D SDS-PAGE followed by whole gel elution, two prominent peaks were recorded using MALDI-TOF MS (Ciphergen, University of Leeds, UK; personal communication, Julian Hiscox) at m/z 7191 and 7565. There are next-integer matches to these in [Table 1](#) at 7190 (**62mer** with *two* water losses) and 7564 (**65mer** with *three* water losses).

A data fitting exercise based on C-terminal truncation and water losses works in tying MALDI observations to ‘predictions’ (meaning *matching a model*, i.e., fitting into an analytical framework, given a certain m/z observation, not *forecasting*), but is disconcerting. Molecular losses from a small protein would normally require exposure to extreme chemical or physical events or enzymatic attack, none of which is relevant in the present case. How does the heterogeneity of Candidate 7500 in MALDI reflect the situation indicated by other approaches? The aa motif NNI is the C-terminus of both EPL001 and seemingly sSgII-70. For EPL001 to be obtained by the Edman machine sSgII-70 must be present in its entirety, yet no stand-alone MS peak has been seen corresponding to an intact **70mer**, as pointed out earlier. Now consider [Fig 5](#) once more. This mass spectrum relates to Gel Band 1 which in a prior purification yielded the EPL001 Edman sequence MKPLTGKVKFNNI. The three peaks in the mass spectrum can be analysed as follows in terms of sSgII-70 in [Table 1](#): 7930.24 (**69mer** lacking $-I$) and 7583.64 & 7565.93 (**65mers** lacking $-\text{KANNI}$). It is not possible to derive the 14-residue EPL001 sequence from any of these truncated forms, in which the C-terminal NNI is diminished or lacking altogether, so it seems that C-terminal truncation is an artefactual aspect of the MALDI campaign. Sheep serum in bulk subjected to anionex and western blotting with an anti-EPL001 antibody produced mono-bands at 0.2M and 0.8M NaCl at ~ 7 kDa (S1 Fig 4 in [S1 File](#)). For this to be the case the C-terminal partial epitope $-\text{NNI}$ must be present. In other words, some amount of SgII-70 must be present in its entirety, in line with the Edman EPL001 data. The anionex purification process used was the same as in the MS investigations, but MALDI-style heterogeneity was not seen. Laddering was seen however in a western of aqueous extract of rat hypothalamus [[17](#)]. The ‘rungs’ on the ladder are too close for whole-factor multimerization to be at issue and anyway for the rungs to be visualised by the anti-EPL001 antibody the C-terminal

NNI must be present. So, MALDI-style C-terminal truncation is not relevant to the observed laddering, which has been attributed to splice variants in the rat not seen in the sheep [17]. An aqueous extract of ovine jugular vein 3–30 kDa upstream precipitate was analysed by SDS-PAGE (text after S1 Fig 10 in S1 File). Several Coomassie-stained bands above 6 kDa showed the presence in MALDI of Candidate 7500, with no or a weak correlation of MW with placement up the gel. It is concluded that MS heterogeneity in the range m/z 7–8000 relating to C-terminal truncation of Candidate 7500 is exclusively associated with the MALDI campaign.

What is to be made of the annotated peaks below m/z 7–8000 in the MALDI spectra of Figs 2–4, there being none such in Figs 5–6? These Ovine Minor Peaks can be submitted to the same sSgII-70 heterogeneity analysis as peaks in the 7000s, involving losses of C-terminal residues and water losses. This uses S3 Table 1 in S3 File, a version of Table 1 extended downwards through the mers. The results are as follows: (i) Observed m/z 6109.35 vs Match 6102.85 (**54mer** with *eight* water losses); (ii) 6070.58 vs 6064.79 (**53mer** with *three* water losses); (iii) 5655.12 vs 5655.44 (**50mer** integer match with *five* water losses); (iv) 5070.90 vs 5069.78 (**45mer** next-integer match with *three* water losses); (v) 4785.16 vs 4791.51 (**43mer** with *three* water losses); (vi) 4755.85 vs 4755.48 (**43mer** integer match with *seven* water losses); & (vii) 4748.66 vs 4749.47 (**43mer** next-integer match with *seven* water losses). In summary: $\sum \text{Observed (n=7)} / \sum \text{Matches} \times 100 = 37195.62 / 37189.32 \times 100 = 100.02\%$; chi-squared $P = 0.02$, significant (S3 Table 9 in S3 File).

Of the seven Ovine Minor Peaks just analysed four occurred in the mass spectrum in the lower panel of Fig 4. The material there was a solid 'upstream precipitate' that formed on the 3 kDa membrane during a bulk 3–30 kDa ultrafiltration of ovine jugular vein EDTA plasma (S1 File Sheffield Method). Bacterial contamination was obvious in terms of discoloration and confirmed by Edman sequencing. It appears that the only thing that flew in MS from this relatively impure sample was the sought-for sSgII-70. The use of sodium azide as an antimicrobial agent reduced but did not eliminate the upstream precipitate, which now appeared white. Separation of this uncontaminated precipitate by SDS-PAGE saw EPL001 Gel Band 1 being obtained by electroelution to provide the uncluttered mass spectrum of Fig 5. An alternative approach was to use solvent extraction of gel bands to provide upstream precipitate for MALDI (S3 Fig 5 in S3 File). As in Fig 5, the base peak (i.e., the most prominent peak, taken as 100%, in an intensity plot of relative abundance) in S3 Fig 5 in S3 File at 7579 is analysed via Table 1 to be a **65mer**, i.e. sSgII-70 lacking –KANNI, with *two* water losses. But whereas Fig 5 is uncluttered, the S3 Fig 5 in S3 File spectrum resembles the lower panel of Fig 4, except that the minor peaks are higher in relation to the base peak and there is a raggedness to the spectrum that may be due to the presence of SDS, absent from the Fig 4 preparation. There are multi-hundred stepped peaks which can be analysed by reference to S3 Table 1 in S3 File as N-terminal Grand Fragments, just like the Ovine Minor Peaks. A sub-8000 analysis is as follows: 3791 (**33mer** match 3789), 4607 (**40mer** match 4613), 5998 (**53mer** match 5992), 6741 (**59mer** integer match 6741), 7579 base peak (**65mer** match 7582) & 7924 (**69mer** match 7919): $\sum \text{Observed (n=6)} / \sum \text{Matches} \times 100 = 36640 / 36636 \times 100 = 100.01\%$.

Apparent N-terminal Grand Fragmentation of a uniform kind was seen when anionex HPLC fractions of ovine jugular vein EDTA blood plasma were analysed by MALDI-TOF MS (M-Scan, Wokingham, UK), after pooled samples were desalted using Sep-Pak C18 cartridges (S3 Fig 8 in S3 File legend). Apart from discounted Candidate VI, fibrinopeptide A, in the form of peaks in the m/z 1400s subject to MS/MS sequencing, the main finding was an array of peaks in the m/z 3000s. As an example, one MALDI spectrum (S3 Fig 8 in S3 File) is bare of significant peaks save for a cluster in the 3000s, topped by signals at m/z 3196, 3179 & 3161. In S3 Table 1 in S3 File, the downwardly extended version of Table 1, these ions find matches on the sSgII-70 **28mer** tier at 3195 (*four* water losses), 3177 (*five* water losses) & 3159 (*six* water losses). When the matrix was CHCA and the instrument was in reflectron mode sSgII-70 **28mers** were mainly seen. With sinapinic acid as matrix and the MS instrument in linear mode sSgII-70 **29mers** were mainly seen. By way of comparison, an anionex sample of ovine jugular vein EDTA plasma at Babraham provided major peaks in MALDI at 2420, being a putative **21mer** having a match in S3 Table 1 in S3 File at 2415, and 2342, being a putative **20mer** with an integer match at 2342, the former obtained with CHCA as matrix, the latter with sinapinic acid. The full analysis of the M-Scan data takes in 23 peaks in the m/z 3000s seen in 10 of 16 mass spectra (S3 Table 10 in S3 File) and involves 13

integer and next-integer hits (57%) to S3 Table 1 in [S3 File](#) values such that $\sum \text{Observed} (n=23) 74406 / \sum \text{Matches} 74402 \times 100 = 100.01\%$. A Babraham spectrum of late-eluting anionex Candidate 7500 from ovarian follicular fluid has peaks supportive of the contention that ovine MALDI peaks in the m/z 3000s are sSgII-70 N-terminal Grand Fragments: 3233 (**28mer** S3 Table 1 in [S3 File](#) match 3231), 3553 (**32mer** next-integer match 3552) & 3924 (**34mer** match 3917) (S3 Fig 11 in [S3 File](#)). Solvent extracted gel bands of upstream precipitate yielded half-sized peaks, as follows: main peak at 7572 (**65mer**), with several peaks including one at 3787 ($7572 \div 2 = 3786$, **33mer** match 3789); and main peak at 7830 (**68mer**), with a sole other peak at 3916 (**34mer** next-integer match 3917, with $7830 \div 2 = 3915$,) (see text after S1 Fig 10 in [S1 File](#), 'Extracted bands of contaminated material'). Purification Candidate IV was 'Luteal phase entity' at m/z 3450 ([S1 File](#)). It was inactive in vitro and in vivo. It is analysed to be an N-terminal Grand Fragment of sSgII-70: **30mer** match 3453.

The half-sized 'Sheep 3000' N-terminal Grand Fragments result has been replicated in the cow. A late-eluting anionex fraction of *bovine* follicular fluid was subject to MALDI analysis after five years storage, with the instrument in linear mode using sinapinic acid as the matrix (S3 Fig 9 in [S3 File](#), interpreted using S3 Table 3 in [S3 File](#)). The only significant peaks were a cluster in the m/z 3000s mainly representing putative bSgII-70 **29mer** Grand Fragments, as per the use of sinapinic acid in the sheep. $\sum \text{Observed} (n=5) 16339 / \sum \text{Matches} 16337 \times 100 = 100.01\%$, as in the sheep.

Dimerization is discernible, where 'dimer' refers to factor fragments doubled up to form *homodimers* ($\alpha\text{mer} + \alpha\text{mer}$) and *heterodimers* ($\alpha\text{mer} + \beta\text{mer}$), with the 'mer' of *dimer* meaning 'peptide' and the 'mer' of *amer* and βmer denoting residues in α or β numbers. (Dimer is NOT a reference to a **2mer**, which would be an isolated factor fragment having two aa residues, a 'two-mer'.) In [Fig 4](#) the peak at 9509 is a 'Superheavyweight', defined as an ion above the predicted MH^+ mass of sSgII-70 at m/z 8176. This ion at 9509 is twice the mass of the lowest peak, at 4755 (**43mer** S3 Table 1 in [S3 File](#) integer match 4755, with seven water losses). N-terminal Grand Fragment *homodimerization* ($\alpha\alpha$) is implied: MH^+ at 4755 + M at 4754 = 9509, integer match to Observed. A Superheavyweight in S3 Fig 5 in [S3 File](#) at 8383 is roughly the sum of the two lowest peaks, this time implying N-terminal Grand Fragment *heterodimerization* ($\alpha\beta$): 3791 (**33mer** match at 3789) + 4607 (**40mer** match at 4613) = 8398. S3 Fig 5 in [S3 File](#) is the MALDI spectrum for solvent extracted EPL001 Gel Band1 (see text after S1 Fig 10 in [S1 File](#), 'Uncontaminated gel'). Carrying on up the MW gel ladder, the next two bands boast items at 8935/8933, which are twice the size of items in the same samples at 4466/4468 (**40mer** next-integer match at 4469), potentially indicating once more N-terminal Grand Fragment homodimerization producing a Superheavyweight: MH^+ match at 4469 + M at 4468 = 8937. A conclusion is that the peaks in the lower panel of [Fig 4](#) and in the S3 Fig 5 spectrum are exclusively sSgII-70 related and artefactual, including in the matter of N-terminal Grand Fragment dimerizations producing Superheavyweights. The mers in [Fig 4](#)'s lower panel are as follows, with S3 Fig 5 in [S3 File](#) mers in brackets: – (**33**), **43** (**42**), **50/53/54** (**53**), **62** (**59**), **65/66** (**65**) & **67** (**69**). These steppings hint at patterned grand fragmentation.

The bioactivity of purified samples can be lost even after a few days of storage, indicating compromised molecularity (see shortly, slopes analysis, **7mer**). Anionex fractions of ovine plasma have been subject to analysis after five year's storage (S3 Fig 10 in [S3 File](#)). Fraction 24 (of 30) yields a picture via MALDI of sSgII-70 **19–64mers** in the form of N-terminal Grand Fragments up to items in the 7000s, together with potential homodimers ($\times 2$) and multi-homodimers ($2+2$, $2+2+2$, $2+2+2+2$). The base peak is 14615, a putative homodimer: **63mer** MH^+ monomer match 7306 + M of 7305 = 14611. Then there are peaks stepped up in mass but stepped down in intensity at 28831 (**62mer** MH^+ monomer integer match 7208 + M of 7207 $\times 3 = 28829$), 43227 (**62mer** MH^+ monomer match 7208 + M of 7207 $\times 5 = 43243$) & 64974 (**70mer** MH^+ monomer integer match 8121 + M of 8120 $\times 7 = 64961$). The analysis of this spectrum thus involves five sub-8000 non-dimeric peaks (Observed vs Matches) plus ten Superheavyweight putative homodimers (Observed) vs ten Superheavyweight-homodimer-matches-by-calculation (Matches), thus: $\sum \text{Observed} (n=15) 310596 / \sum \text{Matches} 310533 \times 100 = 100.02\%$.

Slopes on the left-hand sides of the [Figs 2–3](#) spectra and many others in [S3 File](#), often providing base peaks, call into question MALDI instrument set-up and operation. The items in these slopes are not accounted for by matrix ions, which tend to be in the m/z 800s [[41](#)], with the low mass gate mostly set anyway at 1000 Da. Slopes are seen in MALDI intensity plots of relative abundance involving Candidate 7500, though not always (e.g., [Fig 4](#)), but slopes are *not* seen in counts

plots, though the same peaks are present (S3 Fig 14a & b in [S3 File](#)). An intensity plot of an ovine plasma anionex fraction ([S1 File](#) Sheffield Method) is presented in a supplementary file (S3 Fig 13d in [S3 File](#)). Its most prevalent two peaks, at the top of the ramp, can be analysed in terms of sSgII-70: 1132 base peak (**10mer** S3 Table 1 in [S3 File](#) next-integer match 1133), 1059 (**10mer** match 1061). Peaks at 1133 and 1059 are similarly available from a separate ovine plasma anionex fractionation ([S1 File](#) Harwell Method). With a counts plot and a low mass gate set at 600 Da there are no other annotated peaks and the baseline is horizontal (S3 Fig 14a in [S3 File](#)). An attempt to bulk up analyte involved combining and concentrating eight anionex fractions of ovine blood serum containing Candidate 7500 ([S1 File](#) Babraham Method). No MALDI peaks in the m/z 7000s were evident two days later. Three peaks in the 2000s were seen instead, with a ramp on the left-hand side (S3 Fig 12 in [S3 File](#)). The 2000s are interpreted as sSgII-70 Grand Fragments: $\sum \text{Observed (n=3)} / \sum \text{Matches} 7430 / \sum \text{Matches} 7434 \times 100 = 7430/7434 \times 100 = 99.95\%$. The analysis of the highest intensity peak is as follows: 699 base peak (sSgII-70 **7mer** integer match 699, with *five* water losses). Analysing the spectra alluded to above in terms of sSgII-70: [Fig 2](#), 1027 (**10mer** match 1025) & 1149 base peak (**11mer** match 1153); [Fig 3](#), 1130 (**10mer** match 1133) & 1147 base peak (**11mer** match 1153). An aggregated analysis can be offered on the basis of dually decremented sSgII-70 of ramp peaks in six ovine spectra, all intensity plots (S3 Fig 8, S3 Fig 11 & S3 Fig 13a-d in [S3 File](#)): $\sum \text{Observed (n=35)} / \sum \text{Matches} 87503 / \sum \text{Matches} 87442 \times 100 = 100.07\%$. Concluding, the slopes on the left-hand side of *ovine* MALDI intensity plots are Candidate 7500 fragment ramps.

An aggregated analysis of *bovine* ramp peaks can be offered on the basis of dually decremented bSgII-70 in a pair of spectra, both intensity plots (S3 Fig 1 & S3 Fig 9 in [S3 File](#)): $\sum \text{Observed (n=11)} / \sum \text{Matches} 14960 / \sum \text{Matches} 14948 \times 100 = 100.08\%$.

At each step across [Table 1](#) there is the loss of a mass equivalent to neutral H₂O from an unhydrated (water-free) sSgII-70 C-terminally decremented polypeptide. Water gaps are annotated in the mass spectrum of [Fig 6](#) relating to bSgII-70. Ovine systemic plasma was subjected to spin and gel filtration, followed by anionex ([S1 File](#) Babraham Method). Fraction 20 showed a single peak in MALDI at m/z 7495 (S3 Fig 6 in [S3 File](#)), being an sSgII-70 **65mer** with *seven* water losses, according to [Table 1](#): match 7492. Reread within minutes using the same matrix, this fraction again yielded only a single peak, this time at 7550 (S3 Fig 7 in [S3 File](#)), still corresponding to a **65mer** but now with *four* water losses (match 7546). How has this difference arisen? The 7495 item was detected when the instrument was in continuous mode (i.e., with immediate lasering), the 7550 item when it was in delayed ionisation mode. A *reduction* in water losses due to a minor deferral of lasering is machine artefactuality of an intriguing kind.

The dually decremented MS data model for SgII-70, featuring C-terminal aa truncation of the **70mer** together with water losses, surprises by working for almost-intact monomers (**60mers**, [Table 1](#)) down to minor fragment base peaks, all N-terminal, as well as for Superheavyweight heterodimers ($\alpha\beta$), homodimers ($\alpha\alpha$) and multi-homodimers ($\alpha\alpha \times n$), likewise with all dimers involving dually decremented N-terminal monomer fragments.

Taking stock

Hypothesis H1 is that an inhibitory gonadal influence on tissue masses IS accounted for by the factors described (Candidate 7500, EPL001 & SgII-70). Hypothesis H2 is that an inhibitory gonadal influence on tissue masses IS NOT accounted for by the factors described (Candidate 7500, EPL001 & SgII-70).

The chain of logic behind this paper – from Candidate 7500 through the (machine-artefactual) EPL001 sequence to sSgII-70 and back to Candidate 7500 – is shown in [Fig 8](#) and available in tabular form as S6 Table 1 in [S6 File](#). Although peculiar in finding myriad MS peaks of Candidate 7500 to be related to a single bioactive polypeptide – *E pluribus unum* – and the canonical EPL001 N-terminal sequence MKPLTGKVKFNNI to be lightly mixed up within the neuroendocrine prohormone sSgII, mostly in its second sorting domain, as MLKTGEKPV-FK·NNI ([Fig 9](#)), these interpretations are upheld statistically and support H1 by delivering potential insights (S6 Table 2 in [S6 File](#)). Chief among these is the deduction of sSgII-70's active hormonal face as **M·KPVF·N** ([Fig 11](#)), with that of human and rat being **M·KPNF·N**. These realizations

have permitted the synthesis of hexapeptide mimetics displaying relevant bioactivities in the form of antiorganotrophism (based on fewer, smaller cells) and reproductive modulation [22].

Candidate 7500 is present in blood plasma from ovary-intact sheep and absent in that from OVX ewes, with the former material bioactive in vitro, the latter not. 'OV+/OVX-' is true too of the circulating antigen of an anti-EPL001 antibody and also of an MS tryptic fragment at m/z 906 ($_4$ TGEKPVFK $_{11}$) predicted from the trypsinisation in silico of sSgII-70. If the 906 item were instead OV+/OVX+ then H1 would be subject to a Baconian exclusion of hypothesis [42]. Presence of 906 in material from OVX ewes would be incompatible with its being related to a secreted hormone which circulates from a gonadal source, while having, on the basis of IHC, a conjectured paracrine hypothalamic manifestation also. (Note that oestradiol circulates from the gonads but is also made in the hypothalamus.)

No findings unequivocally discount H1, no findings unequivocally support H2. H2 explains nothing, leads nowhere and is rejected. H1 is supported by different lines of evidence – with better than triangulated support for 'OV+/OVX-' (Fig 8) and triangulated support for 'EPL001 does not exist in nature' and 'Likely SgII related' – co-identifying Candidate 7500, EPL001 & SgII-70 with each other and with the sought-for inhibitory gonadal hormonal influence on tissue masses, micrin.

Summation

The present article provides (a) MALDI data fitted within a unitary sheep and bovine 'SgII-70' model based on peptide splicing from secretogranin II and featuring artefactual molecular losses and dimerizations (factor fragments doubling up) of a kind not usually associated with MALDI investigations; (b) wayward chemical sequencing in a manner not reported before ('Edman Nonsequentialism'), apparently due to SgII-70 undergoing a spiralised depolymerisation in the Edman reaction chamber, notably providing a 14 aa N-terminal sequence denoted 'EPL001'; and (c) a polypeptide aa sequence decryption from SgII using bespoke bioinformatics and bioactivity profiling, which discloses the hormonal active face of its derivative SgII-70 and permits the production of synthetic hexapeptide mimetics thereof; with, along the way, (d) an ad hoc tryptic digest analysis upholding the SgII-70 identification, in particular via an ion at m/z 906 which is an integer match to a predicted tryptic fragment of sSgII-70, and (e) a rarefied epitope mapping exercise based on IHC, identifying the binding site endogenously of an anti-EPL001 antibody as SgII-related.

The current study – with its evidential triangulations and better, replicated results in separate laboratories by different researchers and multi-modal statistical significance – has been an exercise in classical endocrinology, with added crypt-analysis. Genetic approaches involving expression patterns, mutations, knock-outs and so on have to an extent superseded the isolation of factors by biochemical enrichment, given that fruit fly morphogens for example are present in trace amounts and assays to detect them in biological extracts tend to lack reliability [43]. Yet direct analysis of the secretome is still imperative, to characterise what is actually being made and exported.

In the hunt for a postulated tissue-mass inhibiting gonadal hormone there emerges Candidate 7500 from physicochemical purifications of ovine and bovine blood and ovarian follicular fluids. Candidate 7500, ostensibly a polypeptide, displays an enigmatic m/z 7–8000 heterogeneity in MALDI, while being undetectable in other MS modalities. Associated cell-killing activity in vitro is protease and heat resistant. With difficulty aa sequence data are extracted. The 14-residue Edman N-terminal sequence EPL001 is glimpsed in maximally purified ovine plasma anionex fractions as the Beale 4, xxPxxxx-VxxFNxx, before being apprehended more fully in less purified material as MKPLTGKVKFNNI (S1 Table 1 in S1 File) – apprehended yet in a manner uncomprehended by bioinformatics and molecular biology. Inscrutability is countered when from the use of an anti-EPL001 antibody in IP/LC-MS there emerges the concept of 'Likely SgII related', a reference to the intracellular neuroendocrine prohormone secretogranin II. This result is triangulated with the addition of evidence from IHC and bespoke bioinformatics. Micrin is SgII related, it is concluded. EPL001 is decrypted as a mixture of both ends of a 70 aa polypeptide derivative of SgII, sSgII-70. The 14x14 grid plot of Fig 10, showing Edman *output* (EPL001) vertically versus putative sSgII-70 *input* across the top, is deemed informative, by dint of delivering apparent structural insights. The deduced sequence of SgII-70 is dually decremented in an MS datafitting model involving C-terminal truncation (i.e., loss

of C-terminal aa residues as a result of peptide bond breakage) and water losses, upholding the C-terminal sequence of sSgII-70 & bSgII-70. The heterogeneity of Candidate 7500 in MALDI is further deduced to include the production from sSgII-70 of N-terminal Grand Fragments down to a **7mer** and even Superheavyweight ions in the form of fragment heterodimers ($\alpha\beta$), homodimers ($\alpha\alpha$) and multi-homodimers ($\alpha\alpha \times n$).

Bioactivity readouts of an inhibitory kind have been delivered independently by Candidate 7500 anionex fractions, EPL001 in the form of a 14mer synthetic peptide, as well as by hexapeptide mimetics of SgII-70's deduced active face (Fig 8). This is another evidential triangulation.

How is the observed heterogeneity in MS explained? MALDI is a soft ionisation technique, with laser energy imparted to target molecules via an organic matrix. It is an analytical method based on intact mass. A contrast is tandem mass spectrometry (MS/MS), which technique does indeed feature peptide bond breakages, molecular losses and the generation of fragment ions. In MS/MS peptide ions are separated in a mass analyser according to m/z then they are further fragmented and ionized during a collision-induced dissociation step. Comparable results in MALDI would involve post-source decay – of a kind never before reported with this technique and hence not favoured as an explanation here.

Seen in MALDI spectra, according to the analysis, are N-terminal **b**-ions (to use MS/MS fragmentation terminology) of SgII-70, as well as dimers thereof, all subject to the dual decrementation of C-terminal truncation and neutral water losses. With a lack of chemical, physical or enzymatic assaults on the target polypeptide, molecular decrements are perplexing, though it is possibly relevant that acidic conditions are a feature of both MALDI plate set-up and Edman degradation. In MS/MS water losses from peptides can involve the loss of carboxylic acid oxygen, side-chain hydroxyl oxygen or peptide backbone oxygen [44], occurring most prevalently from the hydroxylated side chains of S & T, with such losses also associated with G, D, Q & N [45]. Another explanation for the water decrements overall can be discounted as implausible, as encompassing paucity and plethora. This would be based on transester condensation crosslinking of a variable kind (0–8), involving the hydroxyl groups on the side chains of S, T & Y and the carboxyl groups of the side chains of D & E, releasing water. The view taken here is that SgII-70 loses C-terminal parts of itself plus water during processing and storage and due to MALDI 'plate tectonics' (choice of matrix and immediate or delayed laser), rather than as a result of MS post-source decay. It can be claimed that SgII-70 has been purified to homogeneity on the basis that in anionex fractions the only other entity identified (by m/z ratio and Edman degradation) was fibrinopeptide A, Candidate VI, which was discounted as contra-active and which was seen only in a tiny minority of spectra (e.g., S3 Fig 14b in S3 File).

Three manifestations of sSgII-70 can be discerned in the MALDI data: (i) Candidate 7500 (= peaks in the m/z 7–8000 range showing dual decrementation to provide **60–70mers**); (ii) N-terminal Grand Fragments (= fragments showing major losses from the C terminus, plus water losses), notably producing a stepped grand fragmentation pattern of **20mers**, **30mers**, **40mers**, **50mers** and **60mers**, as well as smaller base peaks; and (iii) Dimerization (= the provision of $\alpha\alpha$ fragment homodimers and $\alpha\beta$ fragment heterodimers by the doubling up of dually decremented monomer fragments, such dimerization not being seen in electrophoresis). Referring to (i) and (ii), bulked-up anionex fractions of Candidate 7500 provided a MALDI spectrum after a couple of days storage devoid of Candidate 7500 but with **19–21mer** peaks in the 2000s and a base peak at m/z 699 that was an integer match to a predicted **7mer** N-terminal fragment of sSgII-70 (S3 Fig 12 in S3 File). Yet this latter match to the sSgII-70 MS data model involves a credulity-stretching *five* water losses. This, from a **7mer** of deduced sequence MLKTGEK. Assuming the 699 integer match is to be taken at face value, then the water losses in Table 1 and its extended form, S3 Table 1 in S3 File, must come from the N terminus. Other grand fragmentations have been described in regard to the use of reverse phase chromatography after anionex and separately relating to consecutive MALDI evaluations of an anionex fraction in different laboratories ('Candidate 7500').

Referring now to (iii), Dimerization, this as an artefact in MS relates to ionisation of target molecules, generating (non-covalent) electrostatic bonds between monomers, with shared charge. Fig 4 sports both a dually decremented sSgII-70 fragment monomer (a **43mer** integer match at 4755, with *seven* water losses) and its homodimer (at an integer matched 9509 to Observation), doubly dual decremented. This is the simultaneous presence in the same spectrum of a

decremented Grand Fragment and a homodimer thereof. A five-year stored sample of ovine anionex plasma ostensibly delivered fragment homodimers in the pattern 2, 2+2, 2+2+2, 2+2+2+2 (S3 Fig 10 in [S3 File](#)). The component monomers are analysed to be dually decremented. Leaving aside the molecular alterations and the locus in the analytical process where these might occur, there is factor identification. There are integer and next-integer matches to the prediction of an SgII-70 dual decremented MS data model across the factor-fragment size range, **7–70mer**, with correspondences to the model for data groups, including dimers, in all cases at around 100%, amid the achievement of high degrees of chi-squared significance as between Observed and Matches. The model works. MALDI is deemed to have read true throughout, detecting SgII-70 faithfully, though in a manner requiring careful analysis.

Regarded as a pharmacon (bioactive molecule) SgII-70 has as a conserved pharmacophore (active face) **M·KPV/NF·N**. This has been deduced from studies in vivo and in vitro using synthetic peptides [22]. The sSgII-70 active site **M·K·PVF·N** is found in ovine derived EPL001 as follows: **MKPLTGKVKEFNNI**. This realisation accounts for (i) the relatively weak inhibitory properties of EPL001; (ii) the ineffectiveness in preventing EPL001's inhibitory activity of alanine substitution by sector (N terminus, Centre Section, C terminus); (iii) the inhibitory activity of both N-terminal and C-terminal EPL001-related hexapeptides; and (iv) the anomalous inhibitory activity in vivo and in vitro of a scrambled-sequence control peptide (EPL030), which happened to have in its sequence **PVF**, three of the Beale 4 (S1 Table 1 in [S1 File](#)), and an appropriately spaced **K·P** doubleton: **KLKMNGKNIPEVFT**. In terms of drug discovery there has been a journey from the 14mer parent compound EPL001 to lead candidate hexamer MKPVFN (EPL601). EPL001, representing only one-fifth of sSgII-70's complement of aa, is receptor active by virtue of its being an amalgam of the two ends of sSgII-70, with 56 intermediate residues (80% of the total) absent.

EPL001 cannot be found using tryptic digestion, molecular biology and bioinformatics because a source protein with this N-terminal aa sequence doesn't exist in nature. This is a triangulated negative. How in principle can an antibody raised against a synthetic linear sequence of 14 aa see a different primary sequence endogenously, with the two sequences being anagrams of one another? So, that is an antibody to ${}^1\text{MKPLTGKVKEFNNI}_{14}$ (EPL001) binding to ${}^1\text{MLKTGEKPVFK}_{11}\cdot{}^{68}\text{NNI}_{70}$ (sSgII-70), with epitopes in the form of ${}^9\text{KEFNNI}_{14}$ and ${}^7\text{KE}_{6}\cdot{}^{10}\text{F}\cdot{}^{68}\text{NNI}_{70}$, respectively [28]. This situation has arisen because the residues of the sSgII-70 polypeptide that went into the Edman machine were not read in the correct order, except for an initial M, but were instead mapped in 3D space, antibody like, to produce the EPL001 output sequence. So, an antibody raised to an EPL001-sequence peptide sees the sSgII-70 input sequence. Why would the Edman machine misread the input sequence? Because sSgII-70 depolymerises progressively in the reaction chamber, displaying residue α -amines out of order in a relatively reproducible manner, for the Edman machine to read them faithfully [22].

The endogenous epitope of the main antibody raised against the synthetic 14mer peptide EPL001 is analysed to be the SgII-related discontinuous **KE·F·NNI**, as just described. Residues at both ends of the SgII proteoform are seen simultaneously by the antibody ([Fig 11](#)). For these remote residues to be adjacent SgII-70 must be non-linear. The active face conception and successful mimesis using short synthetic peptides also implies juxtaposed ends ([Fig 11](#)). In ovine SgII-70 the deduced active face is **M·KPVF·N**, with five residues from the N terminus plus N, as in N68, from the C terminus. 'Adjacent termini' is a structural prediction for sSgII-70 of the [Fig 10](#) grid. The aa data indicate a non-linear reading path for the Edman sequencing, taking in sSgII-70's N terminus (with initial M) and its C-terminal NNI. Also suggestive of molecular end-adjacency is the 'Extension' part of the EPL001 Extension sequence, which after EPL001's C-terminal FNNI proceeds for seven further cycles (S1 Table 1 in [S1 File](#)). On the assumption that residues D, I & Y derive from the C terminus, there is an apparent shuttle between the two ends of SgII-70, with N-terminal contributions shaded and C-terminal placements underlined: **K/I G F/Y/D x F/V I/V I**. 'Both endedness' is also indicated by a mixed signal in Harwell 1, **V/I**, while 'Harwell 2' of S1 Table 1 in [S1 File](#), in its 8-reading entirety, is an interleaved sequence: **M N/F P/I/M L N V/A I T/P**.

The Edman reading paths for EPL001 and related sequences can be represented as spirals (S1 Figs 1 & 2 in [S1 File](#)). In the grid of [Fig 10](#) the EPL001 slalom zigzags and the sSgII-14 peaks and troughs likewise evoke spirality. The grid's

first EPL001 switchback sees a reading of SgII-14's P8 followed by a reading of its L2. A similar switchback is evident in all the ovine Edman sequences, but this effect is not exclusively associated with proline, as in two cases the apex of the switchback is SgII-14's V9. The 'pick' order of EPL001 Edman residues depends on α -amine availability but also steric hindrance, with the latter itself yielding to a spiral conceptualisation [22]. Spirality is connoted by the maximum-purity minimum-sequence Beale 4 (xxPxxxxV/LxxF/KNxx, S1 Table 1 in [S1 File](#)), where ,PVF_9 is a contiguous triplet motif within sSgII-70. The Beale 4 'under signals', L & K, refer to L2 & K3 within sSgII-70, as in MLKTGEKPVFK... The presence of these subsidiary signals is explicable in terms of the spiral conceptualisation just referred to. L2 should be taken second but appears fourth in EPL001 because of steric hindrance by P8. Once P8 is out of the way L2 is available to be read. Removal of V8 releases K3 into availability. The likelihood of the dual positional pairing V/L & F/K arising by chance is 1 in 167 ($P=0.006$, significant) [22]. Spirality is also implied in regard to EPL001/sSgII-14 by the riffle shuffle analysis (see 'Likely SgII related'). Spirality on a whole-molecule scale is connoted by the seven available ovine N-terminal sequences (S1 Table 1 in [S1 File](#)), with all Edman residues accounted for in terms of sSgII-70 at a coverage of 25/70 (36%), as follows, with placements on the assumption that readings D, Y & A are from the C terminus:

MLKTGEKPVFKxxxxMxxxQYxxxxxxxxxSVxxxxGKxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxDxYxANNI

In line with full-molecule spirality is the multi-hundred stepped grand fragmentation pattern detected in MALDI (e.g., [Fig 4](#), lower panel). Sorting domains within SgII are helical and amphipathic [34]. Referring to the long centre section of sSgII-70 not represented on the [Fig 10](#) grid, R12-T58, three helical regions are predicted by the PSIPRED [38] dedicated helix prediction program ([Fig 11](#)). This triple prediction is in line with the results of whole-molecule modelling with the main software used here, ChemSW's Molecular Modelling Pro Plus, version 6.22, and ChemSite, version 5.10 [39] (S5 Fig 1 in [S5 File](#)), but only the likelihood of a central helix is supported by the AI protein structure prediction software AlphaFold2 [40] (S5 Fig 10 in [S5 File](#)). In sSgII-70's second sorting domain there is a repeating pattern of hydrophobic residues (shaded) between charged residues (underlined): **MLKTGEKPVFK**. Amphipathicity of this kind presumably accounts for the bimodalism seen with sSgII-70 in anionex and RP-HPLC, which cannot be due just to a basic N-terminus and an acidic C-terminus, as bimodalism is evinced by Grand Fragments of sSgII-70 in RP-HPLC down to **19mers** lacking the C-terminus altogether (S3 Fig 3 in [S5 File](#)). Coiled coil proteins are prone to homodimerization, another feature of the current data. Summarising, there are multiple reasons for deducing tertiary structural spirality in SgII-70 in terms of both 'Terminal Spiralisation' ([Fig 10](#) grid) and 'Extra-terminal Spiralisation' (off-grid), as there are multiple separate reasons for discerning juxtaposed termini.

Treat as a black box the intracellular processing of secretogranin II, including the disassembly of a possible sorting super-domain. Into the black box goes SgII (modularised in [Fig 9](#) as ABCDE) and out of it comes the SgII-70 proteoform. Four transformations have occurred in the black box to produce SgII-70: (i) *excision* of 61mer and 9mer aa sequence modules B & D from one or two molecules of SgII; (ii) *splicing* of these sequence modules in the order 9+61 (DB) to provide a 70mer polypeptide; (iii) *juxtaposition of termini* and (iv) *spiralisation*.

On (iii), juxtaposition of termini, said termini are in any case anomalous in that the N terminus is Edman-baffling, with N-terminal fragments associated with purification bimodalism (in anionex & RP-HPLC) as well as displaying water losses in high degree and providing *all* SgII-70 ions detected in MALDI; while the C terminus contributes NNI anomalously to an EPL001 Edman sequence which is supposed to be 'N terminal'. In regard to end-adjacency a molecular hairpin has been posited, speculatively stabilised and twisted by crosslinks [46]. By way of comparison, organotrophic IGF-1 has two intramolecular disulphide bonds, insulin three. Yet there are none of the required cysteine residues in the inferred primary structure of (antiorganotrophic) SgII-70 and indeed lack of disulphide bonds is a secretogranin feature. An alternative explanation of SgII-70's behavioural oddities in MALDI and Edman could be intramolecular crosslinks between aa side chains in the form of isopeptide bonds. These are 8-backbone-atom covalent transamidation crosslinks typically between the side chains of glutamines and lysines. Such Q-K bonds are found in thyroglobulin, the precursor of thyroid hormones,

and in the clotting protein fibrin [47,48]. Protease resistant isopeptide stabilization is important in brain function, for example, and skin strength. sSgII-70 is protease and heat resistant, the latter in particular another secretogranin feature. The presence in SgII-70 of isopeptide bonds would not greatly disrupt the MS datafitting of Table 1 as at issue would be reductions for NH₃ of 17.031 instead of H₂O at 18.015Da. Such a datafitting has been attempted for two isopeptide QK bonds within sSgII-70 [46], but exhaustive molecular modelling in silico with software from ChemSW [39] has failed to deliver an image of sSgII-70 with juxtaposed termini (S5 File). Additional isopeptide bonds can be contemplated, but the aberrant ions that might be expected from a crosslinked polypeptide fragmenting in MS have not been seen. (There is also no evidence either of additive post-translational modifications, including glycan and lipid attachments.) Crosslinking is not implied by the 'linear polypeptide' tryptic digest analysis. Yes, sSgII-70 is protease and heat resistant, but the stability is evidently 'cliff edge' and therefore inconsistent with covalent crosslinking, granted apparent partial disintegration in MALDI and Edman. Covalent crosslinking does not explain Edman Nonsequentialism, according to a simulation. Using sSgII-14 (= EPL143 [22]), an E and a K were bonded to the side chains of K and E, respectively. The Edman machine classed the cleaved backbone residue connected to a substituent amino acid as eluting anomalously, registering an x in the correct position in the sequence (S2 File). *The conclusion is that there are no covalent crosslinks within SgII-70.* An elusive factor was indeed only trapped as a result of an IP/LC-MS antigen capture campaign through forced crosslinking using formalin fixation, followed by antigen retrieval.

Transforming sSgII-14 into the EPL001 Edman result requires the movement of only five sSgII-14 residues, four charged and one in a class of its own: K3⁺, K7⁺, K11⁺, E6⁻ & P8 (uniquely a proteinogenic secondary amino acid):

M^vL K₃ T G E K₇ P V_v F K₁₁ N N I

The probability of choosing one uncharged residue of any kind from sSgII-14 and the four charged residues is 10/14 x 4/13 x 3/12 x 2/11 x 1/10 x (5!/(4! x 1!)) = 5/1001 or about 1 in 200 (P=0.005, significant). The probability of choosing proline specifically as the uncharged residue, together with the four charged residues is 1/14 x 4/13 x 3/12 x 2/11 x 1/10 x (5!/(4! x 1!)) = 1/2002 or about 1 in 2000 (P=0.0005, significant). So charged residues are very much at issue. The Edman EPL001 reading seems to have arisen because of aberrant availability of free amine groups due to out-of-sequence peptide bond cleavages affecting charged residues, amounting to a multistage (non-enzymatic) depolymerisation. Charged residues seem important even by omission, with Ks and Es absent from the First Sighting & Harwell 2 (S1 Table 1 in S1 File). The theme of charged residues continues with the EPL001 Extension, where the only EPL001 residue lacking is E, with a blank x registered instead. Of 14 cycles involved in the production of the EPL001 Edman sequence, four are associated with peptide bond cleavages relating to 'anomalous lysines' [22].

Of sSgII-70's residues 25 are charged, representing over a third of the complement. There are ten basic residues, 14% of the total. Four are near the N terminus (K3, K7, K11 & R12), with the rest in the second half of the molecule (K38, K46, H47, R49, K55 & K66). The ten basic residues are outnumbered by 15 acidic Ds & Es (21% of total residues), meaning that sSgII-70 is acidic overall, confirmed by its behaviour in anionex and consistent with high solubility in water (S1 Fig 11 in S1 File). The acidic residues are distributed thus in the two halves of the molecule, with polyanionic clusters towards the C terminus: ₆ExxxxxxxxExxEExxxxxxxExxxxxE₃₅ & ₄₆ExxDEExxxxxDDEDD₆₃. The N terminus is basic, the C terminus acidic, as remarked earlier.

'Charged residues' evokes salt bridging. Such bonds are *not* predicted in the molecular modelling of a linear 70mer, the energy minimization process taking into account changes primary, secondary (hydrogen bonding) and tertiary (including ionic bonds, i.e., salt bridges). This does not preclude the possession of such bonds by SgII-70 emerging from the SgII processing black box. Salt bridges are non-covalent electrostatic bonds that arise most commonly from the anionic carboxylate (RCOO⁻) of either D or E and the cationic ammonium (RNH₃⁺) of K or the guanidinium (RNHC(NH₂)₂⁺) of R. Both the charged residues of the salt-bridge can form a hydrogen bond to the same adjacent glutamine residue, providing additional bracing stabilization in the form D/E⁺K/R [49]. The involvement of Q in the present case is suggested by the Edman

sequencing data of S1 Table 1 in [S1 File](#). A central-sequence motif in EPL001, MKPLT₅GKV₈KEFNNI, finds echoes in other sequences in the Edman series: ₅T/GKV₇, ₅TGQA₈, & ₄T/GQY₆. In the event of sSgII-70's K3 ([Fig 10](#)) being read, NNI was read later, while sSgII-70's Q20 was the start of excursions: **TGQAMEF** (where A67 seems to have been read next) and **T/GQYS/VG?K?** (where the doubletons TG·QY·SV·GK are found in the correct order within sSgII-70, furnishing enough information for a spiralsised Edman reading path to be discerned, supporting both Terminal and Extra-terminal Spiralsisation: S1 Fig 2 in [S1 File](#)). These perambulations exemplify Edman Nonsequentialism, which, to repeat, is an observation not a conjecture.

Running the C terminus of sSgII-70 up the 45° diagonal of the grid from ₆₈NNI₇₀ at bottom right aligns ₅₉DDEDD₆₃ with ₃KxxxK₇ of the N terminus, suggestively in regard to salt bridging ([Fig 10](#)). E6 and K66 also potentially align. Such interactions could account for Terminal Spiralsisation. What of Extra-terminal Spiralsisation? That part of the molecule not represented in the grid, R12-T58, cannot be an uncomplicated circllet. Indeed, mid-molecule spirality is a consensus prediction of PSIPRED [38], ChemSW [39] and AlphaFold2 [40]. Physical modelling (i.e., manual construction of a paper tape representation, with the 70 aa marked along it in the single-letter code) suggests a sole meta-level twist is likely, with two twists a lesser possibility, even three. This means that the molecule is probably minimally braided (single twist) in the manner of an ampersand, &, with a spiral region near the inward crossing and with entwined ends. SgII-70 can thus be appreciated as an amphipathic polypeptide featuring a polyacidic D/E tail looped back to a basic N terminus. This brings about electrostatic self-entanglement, with the rest of the molecule subject to a figure-of-eight spiralsisation, the knotty totality – displaying in all two or more ‘crossings’ [50,51], though SgII-70 is not a true knot topologically, having loose ends – potentially accounting for protease and heat resistance and triangulated recalcitrance in MALDI, Edman and molecular modelling in silico.

Endogenously, the prohormone SgII, like granins generally, is self-aggregating in conditions of high calcium and acidic pH, such as are found in the intragranular milieu. The main predicted aggregation sites in the sSgII sequence of [Fig 9](#) are the N-terminal ₁₈LIFLL₂₂ and the C-terminal ₅₈₅LLVKVLEYL₅₉₃ [52]. These aggregation sites are in modules A & E of [Fig 9](#). (No aggregation sites are predicted for sSgII-70, consistent with the factor circulating as a monomer, free of artefactual dual decrementation.) Molecules of sSgII can be imagined thus, using the [Fig 9](#) modularisation: A1B1C1D1E1+A2B2C2D2E2+A3B3C3D3E3+A4B4C4D4E4 etc. In a simple chain arrangement, the aggregation sites would be E1A2, E2A3, E3A4 and so on. Suppose the 9mer (module D) and 61mer (module B) of sSgII-70 come from different molecules of sSgII, a more plausible idea than reverse splicing of DB from a single molecule of SgII. The adjacent splicings to produce sSgII-70 from a chain could therefore be D1B2, D2B3, D3B4 and so on. At issue could be the processing of an aggregated spirally entwined sorting super-domain. Proteolysis and splicing from this would account for the knotty shape of sSgII-70, which would require activation energy to disentangle it to linearity. A need for activation energy is not incorporated as a consideration in either non-AI physico-chemical modelling based on energy minimization (ChemSW) [39] or AI protein modelling based on database precedents (AlphaFold2) [40]. Hence neither program discloses entangled sSgII-70, with its juxtaposed termini.

Compensatory renal growth in a rat model in vivo involving unilateral nephrectomy was significantly reduced by Candidate 7500 [16] and dose-dependently by EPL001 [25]. Crucial to this model is an increase in the bioactivity of endogenous IGF-1, the growth-promoting mitogen, most probably from the systemic circulation [15]. An anti-EPL001 antibody administered on its own caused CRG to overshoot [25]. This is consistent with the neutralisation of an endogenous inhibitor and strengthens the case for an inverse relatedness in this system between EPL001 associated inhibition and IGF-1 stimulation. An extended version of EPL001 is EPL120, having the aa sequence MKPLTGKVKFNNIKGFVI (derived from the Edman Extension of S1 Table 1 in [S1 File](#)). EPL120 showed dose-dependent inhibition of the (exogenous) IGF-1 stimulated growth in cell numbers of MCF 7 human breast cancer cells over 72h across the EPL120 dose range 0.04 nM - 4 μM [18]. Similar results have been obtained with EPL001, albeit with more variable evocation. The superior bioactivity of the EPL120 20mer over the EPL001 14mer, though they share a string of 14

residues, has been explained topologically within a concept of bimodular tri-residue attachment to a cellular receptor which is itself dimeric [22]. The forward module, MKP, of the sSgII-70 active face, **M·KPVF·N**, displays equal accessibility as represented in the two peptides, for binding to the receptor, but EPL001's rearward module, VFN, has inferior accessibility to the identically positioned VFN in EPL120. A molecular quirk, this, of adding six residues to EPL001. In the kidney research, EPL001 did not affect IGF-1 stimulated proliferation of mouse collecting duct kidney cells in vitro [25]. This indicates that when EPL001 and related peptides inhibit cell proliferation the mechanism does not involve IGF-1 receptor antagonism. It is envisaged that micrin/SgII-70, being antimitogenic and cell-size-limiting, is the (stoichiometric) inhibitory counterpart of IGF-1 within the global organ-size sensing system, via an IGF-1 independent mechanism. The mammalian data just cited are consistent with the findings in nematodes, where the equivalent of IGF-1 is DAF-2. EPL001 and related peptides prolong worm lifespan and modulate fecundity [26,53], as well as altering innate immunity [54], all DAF-2 associated aspects of metabolism.

Antiorganotrophic micrin/SgII-70 is thus conceived as the counterpart of organotrophic IGF-1. Both are polypeptides of 70 aa. The former is acidic, the latter basic, with predicted isoelectric points [14] of 4.64 for hSgII-70 and 7.76 for hIGF-1 (P05019). The bioactivity of SgII-70 is conceptualised to involve, like IGF-1, a dimerized receptor [22]. IGF-1 has two intramolecular disulphide bonds, as described earlier. Compared with IGF-1, the generality of bioactive peptides (having cell-cell influence) are smaller (6–60 aa), are processed from shorter preprohormones (50–300 aa) and have fewer crosslinks (based on 0, 2 or 4 cysteine residues) [55].

Even though Candidate 7500 appeared as part of a precipitate on a 3kDa filter membrane, the yield was low (Fig 4 legend), as multiple failed attempts at chemical sequencing attest. 'Cliff edge' lability, presumably relating to the availability of activation energy, makes SgII-70 resistant to bulk purification – a desirable goal for structural substantiation – and it is currently impossible to provide the factor in true-to-life synthetic or expressed form by dint of apparent topological complexity. Yet it has proved possible to simulate aspects of SgII-70's proposed role in the organotrophic system using stand-in molecules, crafted on the basis of deduction and ready for exploitation. Therapeutic uses for hexapeptide agonists include prostate shrinkage in BPH (S2 Fig 5 in S2 File) and reversal of tissue overgrowth in endometriosis, PCOS and cancer, with infertility among potential reproductive applications for micrin antagonists.

The suggestion that there is an inhibitor of tissue masses in the form of an 'enigmatic ovarian influence' [1] led to a theoretical conceptualisation of the body's hormonal brake against tissue overgrowth, dubbed micrin [7]. This prompted the identification of a candidate prohormone, the neuroendocrine protein secretogranin II [17], and a hormonal proteoform of that, SgII-70, enabling the development of hexapeptide mimetics [22]. Micrin/SgII-70 is conceived to be an antiorganotrophic humoral factor acting remotely in the manner of a classical hormone, with traditional ductless gland (gonadal) provenance and stimulated by antioestrogens [7]. SgII-70's granin stoichiometry is suggested to be the counting mechanism behind the regulation of tissue masses. The internal-size-regulating organotrophic system is viewed as having as a key element the hypothalamic-pituitary-gonadal axis, with adrenal input. It is hypothesized that micrin braking, gonadal and hypothalamic, is lifted at puberty to facilitate maturation and wanes with age, bringing on prostatic enlargement and cancers (S6 File). It is predicted that the endogenous hormone will reverse the pituitary hypertrophy that follows castration in the rat. This effect was attributed in the 1930s to the absence of non-steroidal testicular 'inhibin' [2], before that name attached itself to modern-day inhibin. The latter polypeptide hormone, although gonadal, acts irrelevantly at the pituitary tier to downregulate follicle-stimulating hormone, rather than at the hypothalamus to suppress the pituitary. Micrin/SgII-70 is McCullagh's inhibin of old.

Supporting information

S1 File. Candidates & Purifications. <https://doi.org/10.6084/m9.figshare.27110071.v2>. This project outlines the candidate protein and purification techniques utilised.

(DOCX)

S2 File. Assay Methods and Results. <https://doi.org/10.6084/m9.figshare.30688655>. This project details the methodology used in exemplifying antiorganotrophism for the candidate molecule and gives results.

(DOCX)

S3 File. Mass Spectrometry Studies of Candidate 7500. <https://doi.org/10.6084/m9.figshare.27110284.v2>. This project presents MALDI-TOF MS data of the polypeptide Candidate 7500 and describes it as a novel proteoform of the neuroendocrine prohormone secretogranin II, dubbed sSgII-70.

(DOCX)

S4 File. Tryptic Digests. <https://doi.org/10.6084/m9.figshare.27110311.v2>. This project attempts to elucidate the identity of Candidate 7500/sSgII-70 using trypsinisation.

(DOCX)

S5 File. Molecular Modelling of sSgII-70. <https://doi.org/10.6084/m9.figshare.27110320.v2>. This project presents in silico molecular modelling of sSgII-70 directed at a conformational understanding of the molecule.

(DOCX)

S6 File. Extended Discussion. <https://doi.org/10.6084/m9.figshare.27110332.v2>. This project provides a discussion of possible chains of logic and associated theoretical frameworks which can be associated with and possibly explain the generated organotrophic data and MALDI analyses.

(DOCX)

S7 File. Arrive 2.0 Guidelines. <https://doi.org/10.6084/m9.figshare.27110371.v1>. This project provides the Arrive 2.0 Guidelines.

(PDF)

Acknowledgments

For personal communications the following are lauded and thanked: Pat Barker, Carolyn Carr, Clare Collier, Ed Dudley, Ben Ferneyhough & Julian Hiscox. Maureen Hamon's input is appreciated on the work at Babraham, where ligand purification was pioneered by Dennis Beale. Factor hunters thereafter included Andy Scutt, Charles Dickerson & Dave Emery. Dirk Winkelhardt is thanked for tryptic digest database assistance. Gratitude is expressed to Mark Silverman for help with the riffle shuffle analysis. Thanked for graphics and reprographics is Neil Handley. JEH salutes all the researchers who have been involved in the project and the investors in Endocrine who made the work possible.

Author contributions

Conceptualization: John E. Hart, Russell P. Newton.

Data curation: John E. Hart.

Formal analysis: John E. Hart, Christopher R. Mundy, Aurelia C. Hart, David R. Howlett.

Investigation: John E. Hart, David R. Howlett.

Methodology: John E. Hart, David R. Howlett, Russell P. Newton.

Project administration: John E. Hart.

Software: Christopher R. Mundy.

Visualization: Russell P. Newton.

Writing – original draft: John E. Hart.

Writing – review & editing: Keith G. Davies, Russell P. Newton.

References

1. Hart JE. Endocrine pathology of estrogens: species differences. *Pharmacol Ther.* 1990;47(2):203–18. [https://doi.org/10.1016/0163-7258\(90\)90087-i](https://doi.org/10.1016/0163-7258(90)90087-i) PMID: [2203070](https://pubmed.ncbi.nlm.nih.gov/2203070/)
2. McCullagh DR. Dual Endocrine Activity of the Testes. *Science.* 1932;76(1957):19–20. <https://doi.org/10.1126/science.76.1957.19> PMID: [17815236](https://pubmed.ncbi.nlm.nih.gov/17815236/)
3. Hart JE. Pituitary-related weight changes affecting the liver, uterus and adrenal glands of rats treated with hexoestrol and clomiphene in high doses. *Toxicology.* 1990;61(2):185–94. [https://doi.org/10.1016/0300-483x\(90\)90019-d](https://doi.org/10.1016/0300-483x(90)90019-d) PMID: [2157304](https://pubmed.ncbi.nlm.nih.gov/2157304/)
4. Marlow HW. Properties of Ovarian Extracts. *Endocrinology.* 1939;25(5):793–7. <https://doi.org/10.1210/endo-25-5-793>
5. Matthews CS, Schwabe EL, Emery FE. The effect of continued oral administration of stilbestrol on body growth and organ weights of adult uncastrated and castrated female rats. *Growth.* 1942;6:7–22.
6. Smith OW. Increased pituitary and adrenal weight produced by estrone in intact and castrated normal and runt rats. *Endocrinology.* 1947;40(2):116–8. <https://doi.org/10.1210/endo-40-2-116> PMID: [20286610](https://pubmed.ncbi.nlm.nih.gov/20286610/)
7. Hart JE. The body has a brake: micrin is a postulated new gonadal hormone curbing tissue overgrowth and restricting reproduction. *Med Hypotheses.* 2014;83(6):775–86. <https://doi.org/10.1016/j.mehy.2014.10.009> PMID: [25456786](https://pubmed.ncbi.nlm.nih.gov/25456786/)
8. Beale D. In chromatographic and electrophoretic techniques. In: Smith I, editor. *Chromatography.* 3rd ed. London: William Heineman. 1969:189.
9. Dobson KR, Reading L, Haberey M, Marine X, Scutt A. Centrifugal isolation of bone marrow from bone: an improved method for the recovery and quantitation of bone marrow osteoprogenitor cells from rat tibiae and femuræ. *Calcif Tissue Int.* 1999;65(5):411–3. <https://doi.org/10.1007/s002239900723> PMID: [10541770](https://pubmed.ncbi.nlm.nih.gov/10541770/)
10. Hart JE. Isolated material having an antiorganotrophic effect. 1999. <https://doi.org/WO00/032208>
11. Vidgoff B, Vehrs H. Studies on the inhibitory hormone of the testes IV. effect on the pituitary, thyroid and adrenal glands of the adult male rat1. *Endocrinology.* 1940;26(4):656–61. <https://doi.org/10.1210/endo-26-4-656>
12. Freudenberger CB, Howard PM. Effects of Ovariectomy on Body Growth and Organ Weights of the Young Albino Rat. *Experimental Biology and Medicine.* 1937;36(2):144–8. <https://doi.org/10.3181/00379727-36-9148>
13. Wiles M. Endogenous mammalian factor which inhibits cellular growth in vitro. Sheffield, UK: University of Sheffield. 2002.
14. Expasy compute pl/Mw. http://web.expasy.org/compute_pi/
15. Haylor J, Hickling H, El Eter E, Moir A, Oldroyd S, Hardisty C, et al. JB3, an IGF-I receptor antagonist, inhibits early renal growth in diabetic and uninephrectomized rats. *Journal of the American Society of Nephrology.* 2000;11:2027–35.
16. Hart JE. Purified material having size and growth inhibitory effects on cells and tissues. 2000. <https://doi.org/WO01/93882>
17. Hart JE, Clarke IJ, Risbridger GP, Ferneyhough B, Vega-Hernández M. Mysterious inhibitory cell regulator investigated and found likely to be secretogranin II related. *PeerJ.* 2017;5:e3833. <https://doi.org/10.7717/peerj.3833> PMID: [29043108](https://pubmed.ncbi.nlm.nih.gov/29043108/)
18. Hart JE. Proteinaceous compounds. 2008.
19. Mascot. <https://www.matrixscience.com/>
20. MS-Fit. <https://prospector.ucsf.edu/prospector/mshome.html>
21. Applied Biosystems Procise. http://tools.thermofisher.com/content/sfs/manuals/cms_041126.pdf
22. Hart JE, Mohan S, Davies KG, Ferneyhough B, Clarke IJ, Hunt JA, et al. Hexapeptides from mammalian inhibitory hormone hunt activate and inactivate nematode reproduction. *PLoS One.* 2022;17(12):e0278049. <https://doi.org/10.1371/journal.pone.0278049> PMID: [36454864](https://pubmed.ncbi.nlm.nih.gov/36454864/)
23. Edman P, Högfeldt E, Sillén LG, Kinell P-O. Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chem Scand.* 1950;4:283–93. <https://doi.org/10.3891/acta.chem.scand.04-0283>
24. Expasy PeptideMass. https://web.expasy.org/peptide_mass/
25. Haylor JL, Parker E, Risbridger GP, Beale D, Brown BL, Dobson PRM, et al. Inhibition of compensatory renal growth by the N-terminus of a sheep-derived peptide. *Regul Pept.* 2009;152(1–3):48–53. <https://doi.org/10.1016/j.regpep.2008.10.011> PMID: [19028530](https://pubmed.ncbi.nlm.nih.gov/19028530/)
26. Davies KG, Hart JE. Fecundity and lifespan manipulations in *Caenorhabditis elegans* using exogenous peptides. *Nematol.* 2008;10(1):103–12. <https://doi.org/10.1163/156854108783360221>
27. UniProt. <http://www.uniprot.org/>
28. Howlett DR, Clarke IJ, Newton RP, Hart JE. Epitope mapping of an uncertain endogenous antigen implies secretogranin II peptide splicing. *F1000Res.* 2019;8:1732. <https://doi.org/10.12688/f1000research.20633.2> PMID: [32399184](https://pubmed.ncbi.nlm.nih.gov/32399184/)
29. Bartolomucci A, Possenti R, Mahata SK, Fischer-Colbrie R, Loh YP, Salton SRJ. The extended granin family: structure, function, and biomedical implications. *Endocr Rev.* 2011;32(6):755–97. <https://doi.org/10.1210/er.2010-0027> PMID: [21862681](https://pubmed.ncbi.nlm.nih.gov/21862681/)
30. Zhao E, Zhang D, Basak A, Trudeau VL. New insights into granin-derived peptides: evolution and endocrine roles. *Gen Comp Endocrinol.* 2009;164(2–3):161–74. <https://doi.org/10.1016/j.ygcen.2009.01.011> PMID: [19523383](https://pubmed.ncbi.nlm.nih.gov/19523383/)
31. Marksteiner J, Kirchmair R, Mahata SK, Mahata M, Fischer-Colbrie R, Hogue-Angeletti R, et al. Distribution of secretoneurin, a peptide derived from secretogranin II, in rat brain: an immunocytochemical and radioimmunological study. *Neuroscience.* 1993;54(4):923–44. [https://doi.org/10.1016/0306-4522\(93\)90585-4](https://doi.org/10.1016/0306-4522(93)90585-4) PMID: [8341425](https://pubmed.ncbi.nlm.nih.gov/8341425/)

32. Boutahricht M, Guillemot J, Montero-Hadjadje M, Barakat Y, El Ouezzani S, Alaoui A, et al. Immunohistochemical distribution of the secretogranin II-derived peptide EM66 in the rat hypothalamus: a comparative study with jerboa. *Neurosci Lett*. 2007;414(3):268–72. <https://doi.org/10.1016/j.neulet.2006.12.033> PMID: [17240057](https://pubmed.ncbi.nlm.nih.gov/17240057/)
33. Montero-Hadjadje M, Pelletier G, Yon L, Li S, Guillemot J, Magoul R, et al. Biochemical characterization and immunocytochemical localization of EM66, a novel peptide derived from secretogranin II, in the rat pituitary and adrenal glands. *J Histochem Cytochem*. 2003;51(8):1083–95. <https://doi.org/10.1177/002215540305100812> PMID: [12871990](https://pubmed.ncbi.nlm.nih.gov/12871990/)
34. Courel M, Vasquez MS, Hook VY, Mahata SK, Taupenot L. Sorting of the neuroendocrine secretory protein Secretogranin II into the regulated secretory pathway: role of N- and C-terminal alpha-helical domains. *J Biol Chem*. 2008;283(17):11807–22. <https://doi.org/10.1074/jbc.M709832200> PMID: [18299326](https://pubmed.ncbi.nlm.nih.gov/18299326/)
35. Silverman MP. Progressive Randomization of a Deck of Playing Cards: Experimental Tests and Statistical Analysis of the Riffle Shuffle. *OJS*. 2019;09(02):268–98. <https://doi.org/10.4236/ojs.2019.92020>
36. Yajima A, Ikeda M, Miyazaki K, Maeshima T, Narita N, Narita M. Manserin, a novel peptide from secretogranin II in the neuroendocrine system. *Neuroreport*. 2004;15(11):1755–9. <https://doi.org/10.1097/01.wnr.0000127634.38052.84> PMID: [15257142](https://pubmed.ncbi.nlm.nih.gov/15257142/)
37. Crow JM. Go with the fold. *Chemistry World*. 2018;22–7.
38. PSIPRED. <http://bioinf.cs.ucl.ac.uk/psipred/>
39. ChemSW. <https://www.3ds.com/products/biovia>
40. AlphaFold2. <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=kObIAo-xetgx>
41. Keller BO, Sui J, Young AB, Whittal RM. Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta*. 2008;627(1):71–81. <https://doi.org/10.1016/j.aca.2008.04.043> PMID: [18790129](https://pubmed.ncbi.nlm.nih.gov/18790129/)
42. Platt JR. Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*. 1964;146(3642):347–53. <https://doi.org/10.1126/science.146.3642.347> PMID: [17739513](https://pubmed.ncbi.nlm.nih.gov/17739513/)
43. Nüsslein-Volhard C. *Coming to Life: How Genes Drive Development*. New Haven & London: Yale University Press. 2004.
44. Ballard KD, Gaskell SJ. Dehydration of peptide [M + H]⁺ ions in the gas phase. *J Am Soc Mass Spectrom*. 1993;4(6):477–81. [https://doi.org/10.1016/1044-0305\(93\)80005-J](https://doi.org/10.1016/1044-0305(93)80005-J) PMID: [24235006](https://pubmed.ncbi.nlm.nih.gov/24235006/)
45. Sun S, Yu C, Qiao Y, Lin Y, Dong G, Liu C, et al. Deriving the probabilities of water loss and ammonia loss for amino acids from tandem mass spectra. *J Proteome Res*. 2008;7(1):202–8. <https://doi.org/10.1021/pr070479v> PMID: [18092745](https://pubmed.ncbi.nlm.nih.gov/18092745/)
46. Hart JE. Hormonal derivatives. GB Patent Application. 2021.
47. Walsh CT. *Posttranslational Modification of Proteins*. Englewood, Colorado: Roberts & Co. 2006.
48. Fuentes-Lemus E, Häggglund P, López-Alarcón C, Davies MJ. Oxidative Crosslinking of Peptides and Proteins: Mechanisms of Formation, Detection, Characterization and Quantification. *Molecules*. 2021;27(1):15. <https://doi.org/10.3390/molecules27010015> PMID: [35011250](https://pubmed.ncbi.nlm.nih.gov/35011250/)
49. Ruprecht JJ, Hellawell AM, Harding M, Crichton PG, McCoy AJ, Kunji ERS. Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. *Proc Natl Acad Sci U S A*. 2014;111(4):E426–34. <https://doi.org/10.1073/pnas.1320692111> PMID: [24474793](https://pubmed.ncbi.nlm.nih.gov/24474793/)
50. Lim NCH, Jackson SE. Molecular knots in biology and chemistry. *J Phys Condens Matter*. 2015;27(35):354101. <https://doi.org/10.1088/0953-8984/27/35/354101> PMID: [26291690](https://pubmed.ncbi.nlm.nih.gov/26291690/)
51. Michieletto D. Spot the knot. *Physics World*. 1924;29–33.
52. Tango. <http://tango.crg.es/>
53. Davies KG, Zimmerman B, Dudley E, Newton RP, Hart JE. Reproduction potentiated in nematodes (*Caenorhabditis elegans*) and guppy fish (*Poecilia reticulata*) by adding a synthetic peptide to their aqueous environment. *J Exp Biol*. 2015;218(Pt 5):778–85. <https://doi.org/10.1242/jeb.113837> PMID: [25617450](https://pubmed.ncbi.nlm.nih.gov/25617450/)
54. Davies KG, Scott J, Rowe JA, Hart JE. Peptide optimisation to reduce *Pausteria penetrans* endospore attachment to *Meloidogyne incognita*. In: London, 2008.
55. Kastin AJ. *Handbook of Biologically Active Peptides*. 2nd ed. Academic Press/Elsevier. 2013.