Chem. Senses 35: 675–684, 2010

doi:10.1093/chemse/bjq060
Advance Access publication June 24, 2010

Roborovskin, a Lipocalin in the Urine of the Roborovski Hamster, *Phodopus roborovskii*

Michael J. Turton[1], Duncan H.L. Robertson[1], Julia R. Smith[1], Jane L. Hurst[2] and Robert J. Beynon[1]

[1]Proteomics Group, University of Liverpool, Crown Street, Liverpool L69 3BX, UK and [2]Mammalian Behaviour & Evolution Group, University of Liverpool, Leahurst Campus, Neston CH64 7TE, UK

Correspondence to be sent to: Robert J. Beynon, Proteomics Group, University of Liverpool, Crown Street, Liverpool L69 3BX, UK. e-mail: r.beynon@liverpool.ac.uk

Accepted May 12, 2010

Abstract

Many rodents are now known to exhibit an obligate proteinuria that delivers urine-mediated chemosignals. In this paper, we explore the urinary proteins of the Roborovski hamster (*Phodopus roborovskii*). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis of urine from individual male and female Roborovski hamsters revealed 2 proteins, with approximate masses of 6 and 17 kDa, the expression pattern of which showed little variation between individuals or between sexes. Peptide mass fingerprints obtained from these 2 proteins revealed a number of features: 1) the proteins of a given mass were the same in all individuals regardless of sex, 2) the 6 kDa protein was not a fragment of the 21 kDa protein, and 3) neither protein was a fragment of a larger, conserved protein such as serum albumin. Electrospray mass spectrometry of purified protein preparations established the mass of the larger protein as invariant, at 17144 ± 2 Da in all samples. This protein has been termed roborovskin. The primary structure of roborovskin was determined by tandem mass spectrometry of peptides derived from independent and overlapping digestion with 3 proteases, supported by Edman degradation of the protein N-terminus. Roborovskin shared significant homology with olfactory-binding proteins from *Myodes glareolus* (bank vole) and with aphrodisin and submandibular protein from the golden hamster *Mesocricetus auratus*, all of which belong to the lipocalin superfamily. Lower levels of homology were also indicated between a variety of other lipocalins including the major urinary proteins from house mice and Norway rats. A model of the tertiary structure of roborovskin was constructed from the primary sequence by homology modeling. This model structure resembled other 8-stranded beta barrel lipocalins. Thus, the Roborovski hamster may demonstrate another variant of urinary lipocalin expression, as for the animals studied here, there appears to be no polymorphism in expression either between sexes or individuals.

Key words: aphrodisin, lipocalin, *Phodopus roborovskii*, Roborovski hamster, semiochemical, urinary protein

Introduction

Urine is commonly deployed by mammals as a source of semiochemicals. The complex chemical composition of urine (Albone 1984; Schwende et al. 1984) and the many different classes of molecules therein continue to provide a challenge to investigators seeking to unravel the molecular mechanisms underlying semiochemical communication. It is becoming increasingly apparent than in addition to low–molecular weight metabolites, urine can also contain substantial quantities of protein that is deliberately and specifically released to fulfill a role in chemical communication (Beynon and Hurst 2004). Urinary proteins are involatile and thus cannot function directly in a conventional model of airborne olfactory communication. All of these proteins belong to the lipocalin superfamily of proteins, the functions of which were initially thought to be the transport of hydrophobic molecules through aqueous environments (Cavaggioni and Mucignat-Caretta 2000). An increasing body of work, however, has demonstrated that the urinary proteins, which are "hardwired" into the genome of the donor animal, deliver highly complex relatively invariant signals in urine. Expression of these proteins is often sexually dimorphic, and in the major urinary proteins (MUPs), there is considerable polymorphism (Mucignat-Caretta et al. 1995; Robertson et al. 1996; McFadyen et al. 1999; McFadyen and Locke 2000;

Hurst et al. 2001; Beynon et al. 2002; Chamero et al. 2007; Mudge et al. 2008; Hurst and Beynon 2010). The involatility of proteins provides longevity against which the decay of more volatile signals can be measured (Darwish Marie et al. 2001). Moreover, in mice, the consistent individual-specific pattern of urinary protein expression can deliver an individuality "tag" that can drive male aggression (Chamero et al. 2007), mate choice (Cheetham et al. 2007; Thom et al. 2008), and inbreeding avoidance (Sherborne et al. 2007). Finally, specific urinary proteins may drive sexual attractiveness by stimulating a response (immediate and subsequently learned) to the odor of potential mates (Roberts et al. 2010). Although the role of urinary proteins in semiochemical communication has been the focus of an increasing number of studies, these tend to be limited to the house mouse (*Mus musculus domesticus*) and the Norway rat (*Rattus norvegicus*), both of which express highly polymorphic patterns of urinary proteins. Other studies have focused on the female-specific aphrodisin in vaginal secretions of the golden hamster (*Mesocricetus auratus*) (Henzel et al. 1988; Vincent et al. 2001). A more recent study (Robertson et al. 2007) characterized an MUP from the aboriginal mouse *M. macedonicus*. In this species, there is a single urinary MUP, differing in only a few amino acids from the same proteins in *M. musculus domesticus*, a lack of polymorphism that precludes complex information coding through MUP patterns.

The Roborovski hamster (*Phodopus roborovskii*) is a dwarf hamster that inhabits the extreme desert and semidesert environments in Russia, China, Manchuria, and Mongolia. In common with other dwarf hamsters, they live in a system of subterranean tunnels and nests formed by burrowing. The extreme natural habitats of dwarf hamsters have caused them to adapt physiologically to conserve heat and water, in the latter case by developing a highly effective renal mechanism to concentrate urine and limit the volume of water lost (Natochin Iu et al. 1994). Like many other rodents, the Robrovoski hamster has poor visual acuity and relies on olfactory signals for communication. They use a range of scents to mark their environment, derived from urine, feces, vaginal secretions, and secretions from 2 specialized glands: the mid-ventral gland and the sacculus (Feoktistova and Naidenko 2006). The work presented in this study concentrates on the urinary proteins of the Roborovski hamster and has 3 distinct goals: to survey the urine of male and female Roborovski hamsters for urinary proteins, to characterize the primary structure of such proteins, and to establish the extent of structural heterogeneity and sexual dimorphism in protein expression.

Materials and methods

Animals

All *P. roborovskii* used in this study were captive bred, and although the degree of inbreeding is unknown, it is likely to be substantial. Animals were housed as breeding pairs and any offspring produced were housed in single-sex caged groups after weaning. Animals were fed Harry Hamster food (Supreme Pet Products) and fresh apple. Both food and water were supplied ad libitum.

Sample collection

Urine samples that were naturally expressed when the animal was picked up were used for initial analysis of urinary proteins by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). All further experiments used cage-deposited samples due to the low volume of naturally expressed urine (<5 μL). Deposited samples were collected from individual hamsters in cages previously cleaned rigorously by hand with dilute detergent and 4 water rinses. Animals were placed in dry cages under the above laboratory conditions for 4 h with food and water provided ad libitum. Cage-deposited samples were recovered by 3 successive 5 mL washes of Milli-Q purified water. The total protein concentration of each sample was calculated using the Coomassie Plus Protein Assay (Pierce). Urine dilution was assessed by the total creatinine concentration of each sample, measured using the Sigma Diagnostics Creatinine Assay (Sigma).

SDS-PAGE analysis

SDS-PAGE analysis of the urinary proteins of *P. roborovskii* followed the method of Laemmli (1970), using a Tris–chloride/Tris–glycine discontinuous buffer system under reducing conditions. Electrophoresis was conducted in a mini-protean system (Bio-Rad), and separated proteins were visualized with Coomassie blue stain and silver stain.

Anion exchange chromatography

Proteins from cage-deposited samples were further purified by anion exchange chromatography using a UNO-Q anion exchange column (Bio-Rad) equilibrated in 10 mM Hepes pH 8.0 and were eluted with a 0–0.5 M NaCl gradient at 1 mL/min. The absorbance of the eluate was measured at 280 nm and 0.5 mL fractions were collected. The fractions were subsequently analyzed by SDS-PAGE and electrospray ionization mass spectrometry (ESI-MS).

In-gel proteolysis

Protein bands from SDS-PAGE were digested with trypsin to generate peptides suitable for further analysis by matrix assisted laser desorption ionisation time of flight (MALDI-ToF) mass spectrometry. Excised gel plugs (ca. 1 mm^3) were destained with 50 μL of 50 mM NH_4HCO_3/50% (v/v) acetonitrile for 15 min. They were then reduced and carbamidomethylated by the addition of 50 μL of 10 mM dithiothreitol (DTT) and 50 μL of 55 mM iodoacetamide (IAA) for 20 min. Following reduction and alkylation, the gel plug was washed with 50 μL 100 mM NH_4HCO_3 for 10 min and dehydrated

with 50 μL of acetonitrile for a further 10 min. Excess acetonitrile was removed, and the dry gel plug was rehydrated with 2 μL of 6 ng/μL sequencing grade trypsin (Roche) in 50 mM NH_4HCO_3, pH 8.5. The rehydrated gel plug was then incubated at 37 °C for 5 h following which the remaining supernatant was removed and combined with peptides that were extracted from the gel plug by the addition of 9 μL of 1% (v/v) formic acid/2% (v/v) acetonitrile for 15 min.

In-solution proteolysis

Aliquots of protein (10 μg) purified from cage deposits by anion exchange chromatography were precipitated with a final concentration of 30% (w/v) trichloroacetic acid (TCA) and washed 3 times in ether. To ensure complete reduction of cysteine residues, the resulting pellet was resuspended in 100 μL of 20 mM DTT in 50 mM NH_4HCO_3 and incubated at 55 °C for 1 h. The sample was then carbamidomethylated with 100 μL of 55 mM IAA in 50 mM NH_4HCO_3 and incubated at 37 °C for 1 h in the dark. The product was again precipitated with a final concentration of 30% (w/v) TCA and resuspended in 50 mM NH_4HCO_3/pH 8.5 containing 0.4 μg/μL of sequencing grade trypsin or endopeptidase Glu-C (Roche), prior to being incubated overnight at 37 °C. Digestion with endopeptidase Lys-C (Roche) followed a similar protocol using 25 mM Tris–HCl, 1 mM ethylenediaminetetraacetic acid pH 8.5 as the digestion buffer.

MALDI-ToF mass spectrometry

Analysis of peptides from in-gel digests was undertaken using an MALDI-ToF reflectron mass spectrometer (Waters) in positive ion mode. The instrument was calibrated using a peptide mix (des-arg-bradykinin [2.4 pmol/μL, 1904.47 Da], neurotensin [2.4 pmol/μL, 1672.92 Da], adenocorticotropic hormone clip 18–39 (2.6 pmol/μL, 2465.20 Da), and oxidized insulin β-chain (30 pmol/μL, 3495.9 Da). Samples were prepared for analysis by cocrystallization with an equal volume of matrix solution (saturated α-cyano-4-hydroxy cinnamic acid in a 1:1:1:1 (v/v) solution of acetonitrile:H_2O:methanol:0.1% (v/v) trifluoroacetic acid). A 1-μL aliquot of the peptide–matrix mixture from each sample was deposited onto a 96 well MALDI target and allowed to dry at room temperature. Spectra were acquired between 1000 and 4000 m/z with the laser energy optimized to give the best signal to noise ratio for each sample. The laser firing rate was 5 Hz, and 10 spectra (collected over 2 s) were combined. The final mass spectrum was a combination of 10–15 such combined data sets, representing 100–150 individual laser shots. Each spectrum was then internally recalibrated using the trypsin autolysis peak at 2163.057 Da or by the addition of an internal calibrant, neurotensin (1672.918 Da) (120 fmol/μL). All aspects of data acquisition, processing, and machine management were controlled through the MassLynx software suite (versions 3.5 and 4.0).

Electrospray ionization mass spectrometry

ESI-MS was used in 2 modes; liquid chromatography–mass spectrometry was used for intact mass analysis, whereas tandem mass spectrometry (MS/MS) was used for peptide sequence analysis. All ESI-MS was undertaken on a Q-ToF Micro mass spectrometer (Waters) in positive ion mode. The instrument was calibrated with the product ions of a 500 fmol/μL (Glu1)-fibrinopeptide B (GluFib) solution in 50% (v/v) acetonitrile/0.1% (v/v) formic acid, infused from a syringe pump at 0.5 μL/min through a PicoTip emitter (New Objective). For intact mass analysis, the instrument was operated in ToF only mode. Samples were desalted on line with a C4 reverse phase trap and subsequently introduced into the mass spectrometer in a solution of 90% (v/v) acetonitrile/0.1% (v/v) formic acid. Raw data were gathered between 700 and 1400 m/z at a scan–interscan time of 2.4/0.1 s. These raw data were subsequently deconvoluted using the MaxEnt 1 module contained within the MassLynx 4.0 software. For MS/MS analysis of proteolytic peptides, precursor spectra were acquired between 400 and 1500 m/z at a scan–interscan time of 1.0/0.1 s. Product ion spectra were acquired between 100 and 2000 m/z at the same scan–interscan speed. Raw product ion spectra were deconvoluted using the MaxEnt 3 algorithm in the MassLynx software, with the charge state of the parent peptide determined from the isotope envelope in the precursor ion spectrum. Interpretation of product ion spectra and the determination of peptide sequences were facilitated by the PepSeq module within MassLynx 4.0. As an additional aid in the interpretation of tandem mass spectra, peptides were isotopically labeled with ^{18}O by performing proteolytic digestion in a 1:1 mix of light ($H_2[^{16}O]$) and heavy ($H_2[^{18}O]$) water. This allowed the incorporation of a 1:1 mixture of [^{16}O] and [^{18}O] atoms into the newly formed C-termini of peptides. Following MS/MS, the y-ions were readily identified as a sequence of doublets of approximately equal intensity, separated by 2 Da.

Analysis of lipocalin sequence

The deduced amino acid sequence was used in a basic alignment search tool (BLAST) search (http://blast.ncbi.nlm.nih.gov) using default parameters for protein matches. For 3D model building, the deduced Roborovski protein sequence (with a 4 amino acid ValValArgGly (VVAG) insertion) was used as an input to the Phyre (www.sbg.bio.ic.ac.uk/phyre) server to build theoretical structural models (Bennett-Lovsey et al. 2008; Kelley and Sternberg 2009). Excellent matches were obtained to a large number of lipocalins, including aphrodisin (1EJP.PDB), which was used as a template to build a homology model of the Roborovski lipocalin, using default parameters.

Results and discussion

A preliminary measure of protein output from male and female Roborovski hamsters was gained from protein and

creatinine measurements performed on urine reclaimed from cage washes. The protein:creatinine ratio gives a meaningful indication of protein output, corrected for urine concentration. This was remarkably similar in male and female Roborovski hamsters: 12.0 ± 0.8 (mean ± standard error of the mean [SEM], n = 6) in males and 13.5 ± 0.8 (mean ± SEM, n = 6) in females. These values are comparable with those obtained from inbred male laboratory mice (10.3 ± 0.8, n = 27 strains) and are higher than those obtained from female inbred laboratory mice (3.1 ± 0.3, n = 27 strains) (Cheetham et al. 2009). The sexual dimorphism of MUP expression in male and female mice is not apparent in the urinary protein output from Roborovski hamsters, which is comparable in both sexes. Samples from males and females were analyzed by 1D SDS-PAGE (Figure 1). In each case, regardless of sex, 2 major protein bands were observed with electrophoretic mobilities approximating to molecular masses of 6 and 21 kDa. The SDS-PAGE patterns were remarkably consistent between individuals and sexes, confirming the absence of sexual dimorphism and indicating little heterogeneity in protein expression. To investigate further the level of similarity between individuals and sexes and to explore the possibility that the 6 kDa band was a degradation product of the larger 21 kDa protein, small portions were removed from each band and the proteins digested with trypsin. The peptides resulting from this digestion were subsequently extracted and analyzed by MALDI-ToF mass spectrometry (Figure 2). The mass spectra from both the 21 kDa band and the 6 kDa band were strikingly similar in males and females, suggesting that in both cases the protein sequences within these bands were identical. By contrast, there was little similarity between the MALDI-ToF mass spectra derived from the 6 and 21 kDa bands. The 6 kDa band is not therefore a fragment of the larger 21 kDa band: the 6 kDa band has not been investigated further. The masses of the main trypsin peptides from both bands were also used to interrogate a nonredundant protein sequence database but no statistically significant matches were found. This precludes the possibility that the 21 or 6 kDa proteins were fragments of an abundant serum protein such as albumin, which shows a high degree of sequence conservation between species.

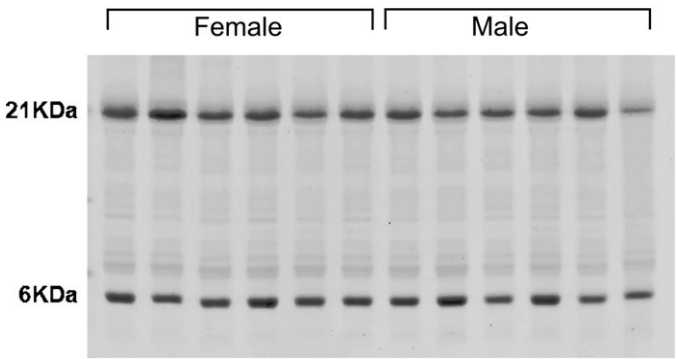

Figure 1 SDS-PAGE analysis of urinary proteins from male and female Roborovski hamsters. Urine from male and female Roborovski hamsters was collected directly or reclaimed from cage deposits as in Materials and Methods. These samples were then analyzed directly by SDS-PAGE. Molecular mass measurements were made by comparison with standard proteins that are not shown in the figure.

Concentrated cage washes from male and female Roborovski hamsters were purified and concentrated further by anion exchange chromatography (results not shown). The resulting protein preparations were subsequently desalted, and the accurate mass of the intact protein was determined by ESI-MS (Figure 3). The deconvoluted mass spectrum was dominated by a peak at 17144 Da in samples derived from male or female urine. Although this mass differs from that estimated by SDS-PAGE, this is not without precedent (Armstrong et al. 2005). The close agreement between these masses in male and female urine further confirms the lack of sexual dimorphism and structural heterogeneity.

To gain detailed information about the amino acid sequence of this protein, we adopted a strategy that combined Edman degradation, proteolysis with 3 different proteases (endopeptidase Lys-C, endopeptidase Glu-C, and trypsin), and MS/MS. The Edman degradation, performed on an electroblotted sample, was used to determine the N-terminal sequence of the mature protein, whereas MS/MS was used to determine the amino acid sequence of peptides resulting from treatment with each of the 3 proteases. The differing primary specificity of the 3 proteases produced overlapping sequences that could be assembled to yield the complete protein sequence. To assist in this process, proteolytic digestions were performed in an equimolar mixture of $H_2[^{16}O]$ and $H_2[^{18}O]$, which allowed discrimination of N-terminal (b-ion) and C-terminal (y-ion) MS/MS products. This sequence analysis is, however, subject to the caveat that MS/MS cannot distinguish between leucine and isoleucine residues, which are denoted by the letter J in this study.

The results of this analysis are summarized in Figure 4. Edman degradation revealed the first 22 amino acids (DNYAELQGEWDTLALAADNDDK). This matched the MS/MS interpretation of a 1234.04 m/z doubly charged ion $[M+2H]^{2+}$ Lys-C N-terminal peptide (L1; DNYAEJQGEWDTJAJAADNDDKJK). The amino acid sequence of L1 was used to perform a low-stringency BLAST search (blast.ncbi.nlm.nih.gov), which revealed a match (52% identity, 85% residue similarity, no gaps) to aphrodisin, a lipocalin present in the vaginal secretions of the golden hamster (SwissProt P09465). Moreover, the alignment confirmed that the N-terminus of the Roborovski protein contained the obligatory lipocalin motif GxW (Flower 2000; Flower et al. 2000). The sequence of aphrodisin was subsequently used as an additional aid in the localization and alignment of further MS/MS sequences.

An additional 35 amino acids (EEGPJRJYVREJYCNEDCSEMEVTFYVNANNQCSK) were determined from

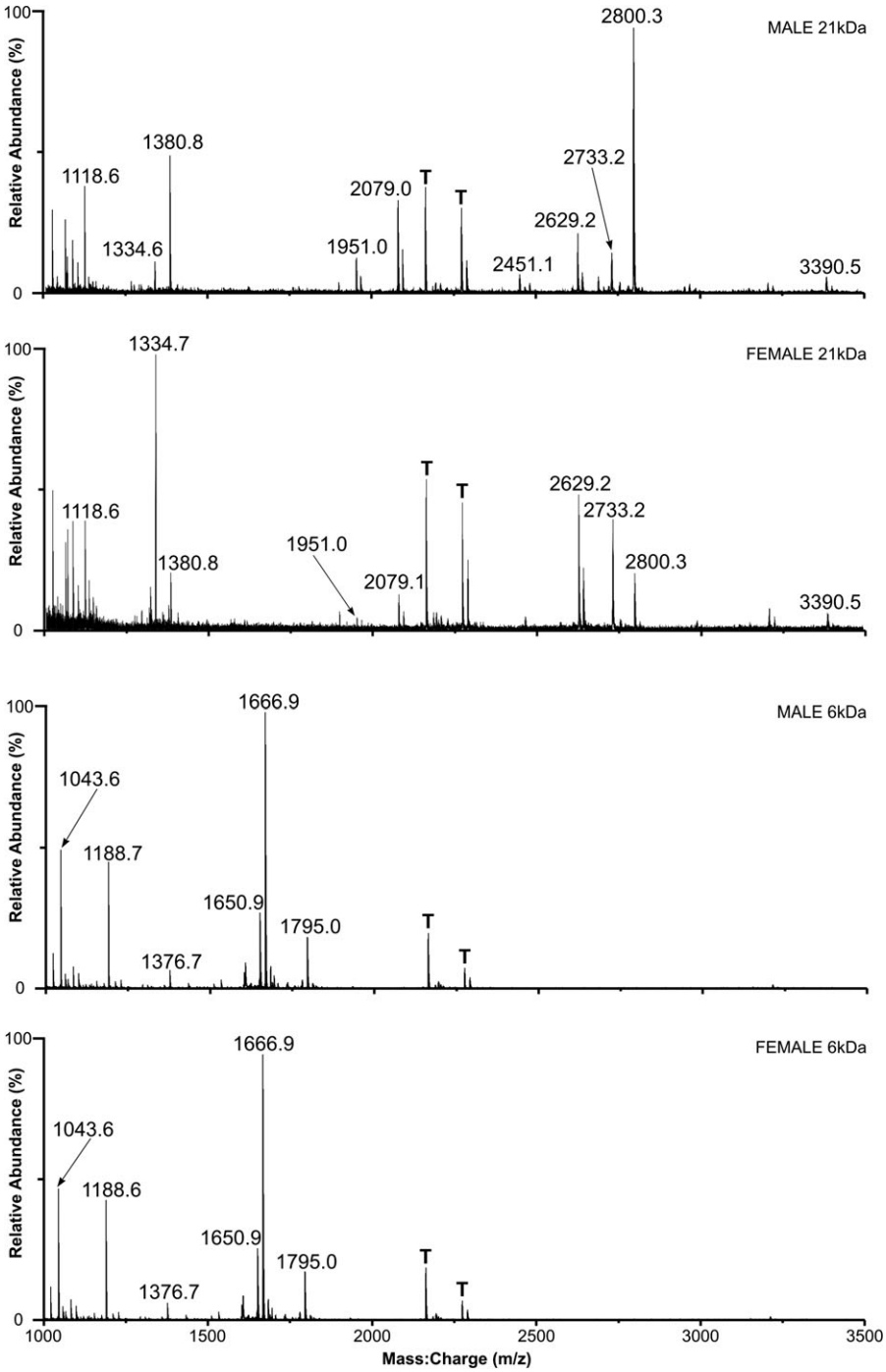

Figure 2 MALDI-ToF mass spectrometry of peptides derived from in-gel proteolytic digestion of Roborovski hamster urinary proteins. The upper 2 panels show the mass spectra obtained from the 21 kDa SDS-PAGE band from male and females, respectively. The lower 2 panels show the corresponding spectra from the 6 kDa SDS-PAGE band. Two peptide ions derived from trypsin are marked with a "T."

MS/MS of a second Lys-C peptide (L2), and these were linked to the N-terminal sequence by the MS/MS interpretation of 2 overlapping peptides: a 471.42 m/z $[M+2H]^{2+}$ trypsin fragment (T2) and a 909.6 m/z $[M+2H]^{2+}$ Glu-C peptide (G2). The peptide chain was extended by a further 18 residues (... TEVTGYKQADGTYRKTFE) by MS/MS analysis of a 906 m/z $[M+4H]^{4+}$ Glu-C peptide (G5) which overlapped the previous Lys-C fragment (L2) by 13 residues. The sequence of this Glu-C fragment was in turn overlapped by that of a further 1176 m/z $[M+3H]^{3+}$ Lys-C fragment (L3) which extended the peptide chain by a further 20 residues (GNDTFKPVYATPEQVVFTSK). The MS/MS spectrum

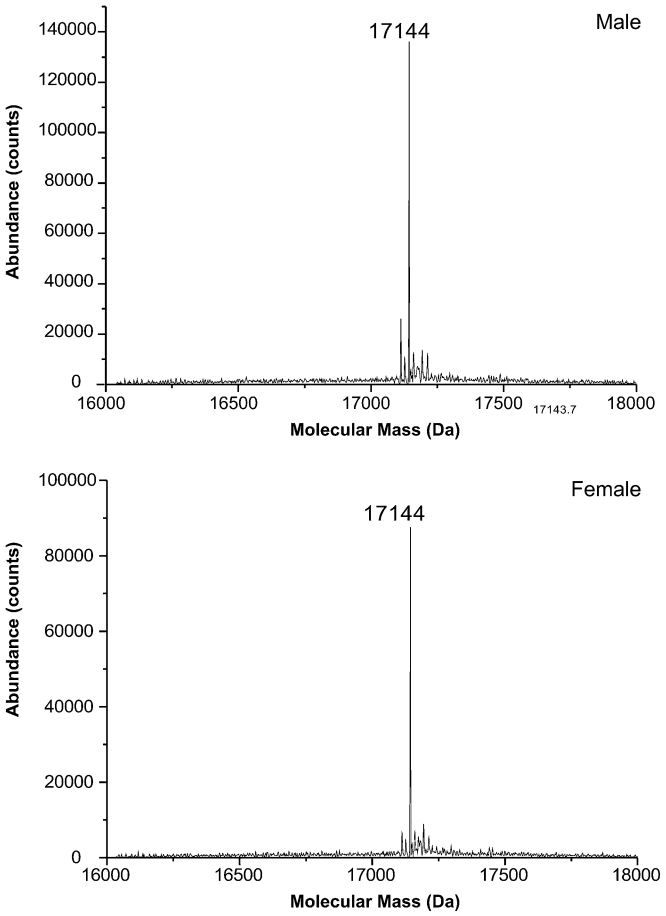

Figure 3 ESI-MS of intact Roborovski hamster urinary proteins. Urinary proteins from male and female Robrovski hamsters were obtained from cage washes and further concentrated and purified by anion exchange chromatography. These preparations were subsequently desalted and analyzed by ESI-MS. An initial low-resolution MaxEnt survey spectrum processed between 10 and 50 kDa (not shown) indicated that the urinary proteins had masses of approximately 17 100 Da. More accurate masses for each sample were obtained from a second higher resolution MaxEnt spectrum. Spectra obtained from male and female urine samples are in the upper and lower panels, respectively.

of another overlapping 839 m/z $[M+2H]^{2+}$ Glu-C peptide (G6) was interpreted to show a sequence linking L3 to another 620 m/z $[M+3H]^{3+}$ Lys-C peptide (L4), which extended the peptide chain by a further 12 residues (... NVDRAGQETNJJ ...). The sequence at the C-terminal region of this peptide was not clear from the MS/MS spectrum and is unassigned. The identity of the C-terminal residue, however, was inferred as a lysine residue, on the basis of the cleavage specificity of Lys-C. Further MS/MS analysis of a 445 m/z $[M+3H]^{3+}$ Lys-C peptide (L5) revealed the sequence ... SEHJTPEKHEK ... which, in the absence of an overlapping peptide, was placed adjacent to the sequence from L4 on the basis of alignment with aphrodisin. The MS/MS analysis of these peptides has established identities for 120 N amino acids at the N-terminus of this protein.

The C-terminus of the protein was initially identified as an 877 m/z $[M+3H]^{3+}$ Lys-C peptide (L6) from which 23 amino acids were identified (AJHENNJPEENJHNVJATDTCPK). The C-terminal lysine residue in this peptide, the primary cleavage site of Lys-C, raised the possibility that this peptide might not represent the true C-terminus of the protein. This uncertainty was resolved by the presence of 2 Glu-C peptides (G9 and G10; 727 m/z $[M+3H]^{3+}$ and 806 m/z $[M+2H]^{2+}$, respectively), which shared 13 C-terminal amino acids with L6 including the C-terminal lysine, not recognized as a primary cleavage site for Glu-C. Additionally, there was no evidence of $[^{16}O:^{18}O, 1:1]$ incorporation into any of these peptides indicating that these peptides are derived from a region of the protein with a free carboxyl group, namely the C-terminus. The identities of a further 4 C-terminal amino acid residues were inferred from a $[M+H]^{+}$ 3097.61 m/z Glu-C peptide (G8), observed only following MALDI-ToF mass spectrometry. The amino acid sequence of this peptide was inferred to be KJVKAJHENNJPEENJHNVJATDTCPK on the basis of the peptide mass and the cleavage specificities of Lys-C and Glu-C. Specifically, this peptide was identified as belonging to the C-terminal portion of the peptide chain by virtue of the fact that it underwent an increase in mass of 57.14 Da following treatment with DTT and IAA (data not shown). This peptide was, however, 468.32 Da heavier than L6 indicating the presence of additional amino acid residues. One of these residues was assigned as a lysine on the basis of the Lys-C specificity in the cleavage of L6. Alignment of the aphrodisin sequence identified a glutamate at position 123, also observed in L4, which was considered the likely cleavage point for G8. This allowed the N-terminal residue of G8 to be identified as a lysine by comparison with L4. Following these 2 inferred identifications, the unassigned mass in G8, when compared with L6, was 212.14 Da. This difference in mass is consistent with 2 possible dipeptide combinations: PD (212.07 Da) or JV (212.15 Da). The latter combination (JV) was assigned on the basis of alignment to the aphrodisin sequence (LV) and closer agreement in mass.

The C-terminal sequence thus determined was linked to the 120 residue N-terminal sequence by alignment to aphrodisin. A comparison of the predicted molecular mass of the complete protein sequence (16777 Da) with that observed by ESI-MS (17144Da) revealed a mass deficit of 367 Da. This difference in mass is attributed to the residues in L4 which were not identified by MS/MS and is most likely a tripeptide or tetrapeptide. In consequence, the final sequence is likely to comprise 150 or 151 amino acid residues.

The positions of cysteine residues within the protein structure were confirmed by MALDI-ToF mass spectrometry, following treatment with either a reducing agent alone (20 mM DTT) or a reducing agent followed by IAA to alkylate free thiol groups. Under such conditions, a mass change of +57.016 Da (cysteine carbamidomethylation) is expected for each cysteine residue in the peptide. The Lys-C peptide L6, previously determined to be the C-terminal portion of the

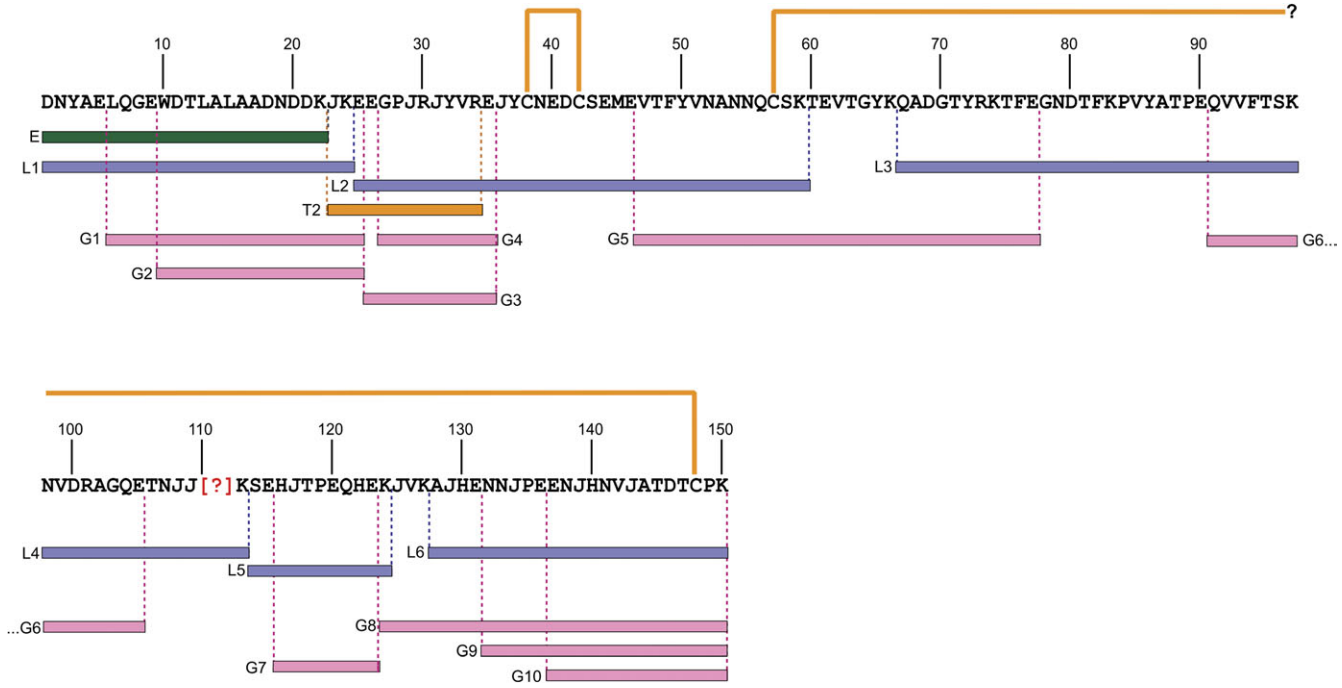

Figure 4 The amino acid sequence of the 17.1 kDa Roborovski urinary protein determined by differential proteolytic digestion, Edman degradation, and MS/MS. Aliquots of an anion exchange purified Roborovski urinary protein preparation were digested in solution with 3 proteases: Trypsin (peptide labeled T1), Endoproteinase Lys-C (peptides labeled L1 to L6), and Endoproteinase Glu-C (peptides labeled G1–G10). A further aliquot of the same sample was electroblotted onto a polyvinylidene fluoride membrane and subjected to Edman degradation (peptide labeled E). The position of each proteolytic peptide within the protein molecule was deduced from portions of amino acid sequence that overlapped those of other peptides or with the N-terminal sequence determined by Edman degradation. The position of a disulphide bridge between residues 38 and 42 was determined from MALDI-ToF mass spectrometry of proteolytic digests in the presence and absence of the reducing agent DTT. Isobaric leucine (L) and isoleucine (I) residues cannot be distinguished by MS/MS and are consequently represented by the letter "J" in the amino acid sequence.

protein, increased in mass from $[M+H]^+$ 2572.15 m/z under reducing conditions alone to $[M+H]^+$ 2629.19 m/z when reduced and treated with IAA. This mass increase (+57.04 Da) is consistent with the presence of a single alkylated cysteine (C*) residue within the peptide. The sequence of the alkylated peptide was interpreted as AJHENNJPEENJHNVJATDTC*PK from the MS/MS spectrum of an 877 m/z $[M+H]^{3+}$ precursor ion. This peptide has a high degree of sequence similarity to the C-terminus of aphrodisin and the cysteine residue aligns well with the highly conserved cysteine that forms the structural disulphide bond linking the C-terminus to the main beta barrel in most lipocalin structures (Flower et al. 1993).

A second trypsin peptide of $[M+H]^+$ 2931.0 was replaced by an ion of $[M+H]^+$ 3104.16 m/z following carbamidomethylation, a change of 173.16 Da. This mass difference was accounted for by the incorporation of 3 carbamidomethyl groups (+171.05 Da) and the reduction of an internal disulphide bridge (+2.02 Da). The formation of an internal disulphide bridge was confirmed by MALDI-ToF analysis of an $[M+H]^+$ 4144.02 m/z Lys-C peptide that was produced under weak reducing conditions (20 mM DTT). Carbamidomethylation yielded a peptide of $[M+H]^+$ 4317 m/z, a mass difference of 171.14 m/z, indicating that the disulphide bridge did indeed reform under mild reducing conditions. Under strong reducing conditions (200 mM DTT), the peptide ion was observed at $[M+H]^+$ 4146.0 m/z, a difference of +2.13 Da to that observed under weak reducing conditions. When this peptide was carbamidomethylated, a mass change of +172.92 Da was observed, yielding a peptide ion of $[M+H]^+$ 4316.94 m/z. The presence of a putative disulphide bond in a peptide treated with 20 mM DTT was unexpected and attributed to the formation of a tight disulphide bond between the 2 cysteine residues at the N-terminal end of the peptide that could reform. A similar disulphide bond arrangement has previously been observed in aphrodisin (Vincent et al. 2001). The full sequence of this region was determined as EEGPJRJYVREJYC*NEDC*SEMEVTFYVNANNKC*SK from MS/MS analysis of a 1080.5 m/z $[M+4H]^{4+}$ Lys-C peptide.

The full deduced sequence (Figure 5), arbitrarily nominating all leucine–isoleucine residues as L, was used as a query sequence in a protein BLAST search against the nonredundant protein database. All matches were to lipocalins, confirming that the Roborovski urinary protein was indeed another member of this family and we have termed this protein roborovskin. The highest matches were to urinary odorant–binding proteins (OBPs) (in the order—forms 3, 2, and 1)

Figure 5 BLAST analysis of the Roborovski hamster urinary protein. The amino acid sequence of the Roborovski hamster urinary protein was used for a BLAST search to identify similar amino acid sequences. The highest match was to urinary olfactory–binding proteins aphrodisin from *Myodes glareolus*, with strong matches to aphrodisin from *Mesocricetus auratus*. The isobaric residues leucine (L) and isoleucine (I) were nominated as leucine for the purpose of the BLAST search.

from *Myodes glareolus* (Stopkova et al. 2010), with 60%, 57%, and 57% identity, followed by aphrodisin from *M. auratus* and *Cricetus cricetus* (Figure 5). The next highest match was to a mouse OBP at 44% identity. Interestingly, other urinary lipocalins, notably the MUPs from mouse and rat, were not matched strongly, preceded in rank by OBPs from the same species. The best alignments all required a single gap at the position of the short unknown region (Figure 4), corresponding to a 4 residues sequence segment in aphrodisin (VVAG). These different levels of homology between different categories of vertebrate semiochemical lipocalins suggest that they fall into 2 distinct groups: one associated with the Muridae, comprising MUPs from mice and rats and a second associated with the Cricetidae, comprising aphrodisin, submandibular protein, and roborovskin.

The deduced sequence was also used to predict a 3D structure using the Phyre server (Bennett-Lovsey et al. 2008; Kelley and Sternberg 2009), after insertion of the hypothetical tetrapeptide VVAG (derived from the comparable aphrodisin sequence) into the undetermined "gap." The solved structure that was most closely matched by Phyre to the roborovskin sequence was surprisingly not aphrodisin (1E5P.PDB) but another lipocalin family member—pig salivary lipocalins (1DKZ.PDB). All the other homologous structures were lipocalins (e values from 9.3×10^{-18} to 4.8×10^{-16}) and all generated a good quality structure. However, the match to aphrodisin (e value 4.2×10^{-17}) had the highest sequence similarity (58%, with the caveats that Leu/Ile are unknown and a theoretical tetrapeptide has been inserted), and this has been used to direct the construction of the theoretical model of roborovskin. The missing gap is located in one of the beta sheet strands, which, together with the knowledge of the mass deficit between the observed and the predicted sequence, suggests that this region of the beta strand extends further. Furthermore, in the modeled structure, the location of the cysteine residues matched those of aphrodisin, adding further evidence for the correct assignment of the 2 disulphide bonds, although the disposition of the short internal disulphide bond (C*NEDC*) suggests that the model may not be precise at this location. (Figure 6).

Although this protein is strongly expressed in urine, it does not appear to be an MUP. Given the observations of the extreme genomic complexity and protein polymorphism of MUPs (McFadyen and Locke 2000; Mudge et al. 2008; Cheetham et al. 2009), the apparently sex-independent expression of a single protein species in the Roborovski hamster is consistent with the protein having functions that are generic to the species rather than a sex-specific function or encoding an individual-specific signal. Of course, urinary roborovskin expression has thus far only been examined for adult breeding animals of both sexes, and further studies would be required to explore expression during development and in wild populations.

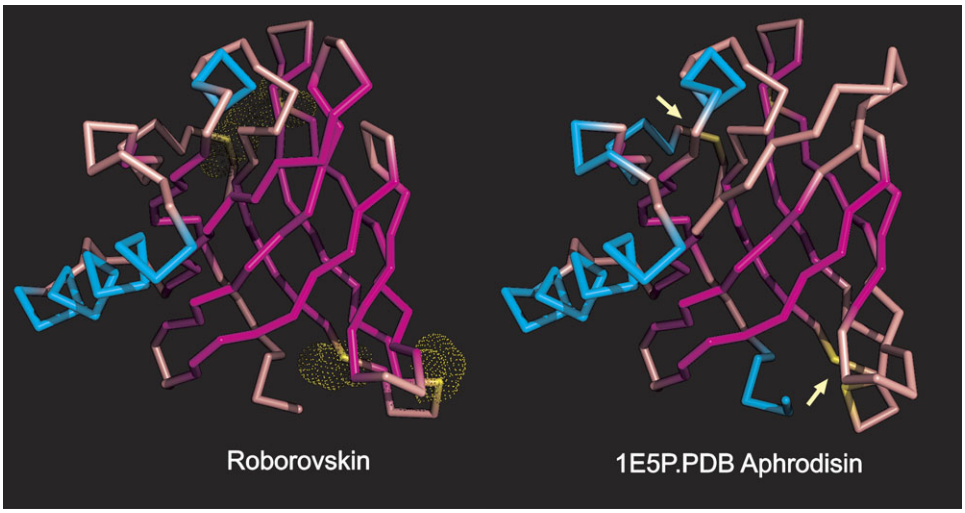

Figure 6 The 3D structure of roborovskin determined by Phyre ensemble fold recognition. The deduced sequence of roborovskin, with the isobaric leucine and isoleucine residues all represented as leucine, was used as an input to the Phyre server (www.sbg.bio.ic.ac.uk/phyre). The ensemble fold recognition process used by the Phyre server found high confidence matches to a series of lipocalins and the highest sequence similarity was to aphrodisin (1e5p.pdb), which was subsequently used to build a homology model of roborovskin. The known disulphide bonds in aphrodisin are highlighted with arrows, in roborovskin, the cysteine residues are highlighted as dot surfaces.

With the caveat above, the expression pattern of roborovskin was remarkably consistent between both individuals and between the sexes. There was no evidence of multiple roborovskin variants as observed with house mice, where differential expression of the MUP gene cluster family gives rise to a mixture of closely related proteins. In the light of this consistent urinary expression, the homology to aphrodisin is perhaps surprising. Aphrodisin is expressed in female hamster vaginal secretions during estrus and possesses aphrodisiac properties that facilitate the sexual behavior of the male (Singer 1991). Expression of aphrodisin is thus extremely sexually dimorphic, whereas in breeding adults, roborovskin expression shows no sexual dimorphism. Although roborovskin shares a greater degree of homology with aphrodisin, it is very unlikely to act in the same role in sexual reproduction.

Although the proteins themselves are unlikely to fulfill a sex-specific function, a characteristic of lipocalins is their ability to bind low molecular mass ligands. A sex-specific protein–ligand complex might have the potential to signal this type of information, although to date, no endogenous ligands of aphrodisin have been discovered and we have not investigated the ligand status of roborovskin. Analysis of the built model for a central cavity, using the CASTp algorithm (http://sts.bioengr.uic.edu/castp/calculation.php) (Dundas et al. 2006) revealed a central cavity of approximately 480 $Å^3$ that could accommodate a ligand of reasonable size. If the relatively unchanging expression pattern of roborovskin is manifest in wild individuals, it is possible that roborovskin could act as a species recognition signal between conspecifics, for marking resources such as nest sites and food sources. It could act as a navigation cue between such resources, of great relevance to a nocturnal burrowing animal. Finally, although the expression of the protein is unchanging, it does not preclude sex-specific seasonal or behaviorally driven variation in deposition patterns of a urinary chemosignal. Further exploration of these ideas will require studies on wild animals.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council (grant BB/C/503897); and the University of Liverpool.

Acknowledgement

We are grateful to Dr Rick Humphries and animal care staff for practical support with the animals.

References

Albone E. 1984. Mammalian semiochemistry. New York: John Wiley.

Armstrong SD, Robertson DHL, Cheetham SA, Hurst JL, Beynon RJ. 2005. Structural and functional differences in isoforms of mouse major urinary proteins: a male-specific protein that preferentially binds a male pheromone. Biochem J. 391:343–350.

Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA. 2008. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins. 70:611–625.

Beynon RJ, Hurst JL. 2004. Urinary proteins and the modulation of chemical scents in mice and rats. Peptides. 25:1553–1563.

Beynon RJ, Veggerby C, Payne CE, Robertson DH, Gaskell SJ, Humphries RE, Hurst JL. 2002. Polymorphism in major urinary proteins: molecular heterogeneity in a wild mouse population. J Chem Ecol. 28:1429–1446.

Cavaggioni A, Mucignat-Caretta C. 2000. Major urinary proteins, alpha(2U)-globulins and aphrodisin. Biochim Biophys Acta. 1482:218–228.

Chamero P, Marton TF, Logan DW, Flanagan K, Cruz JR, Saghatelian A, Cravatt BF, Stowers L. 2007. Identification of protein pheromones that promote aggressive behaviour. Nature. 450:899–902.

Cheetham SA, Smith AL, Armstrong SD, Beynon RJ, Hurst JL. 2009. Limited variation in the major urinary proteins of laboratory mice. Physiol Behav. 96:253–261.

Cheetham SA, Thom MD, Jury F, Ollier WE, Beynon RJ, Hurst JL. 2007. The genetic basis of individual-recognition signals in the mouse. Curr Biol. 17:1771–1777.

Darwish Marie A, Veggerby C, Robertson DH, Gaskell SJ, Hubbard SJ, Martinsen L, Hurst JL, Beynon RJ. 2001. Effect of polymorphisms on ligand binding by mouse major urinary proteins. Protein Sci. 10: 411–417.

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 34:W116–W118.

Feoktistova NY, Naidenko SV. 2006. Hormonal response to conspecific chemical signals as an indicator of seasonal reproduction dynamics in the desert hamster, Phodopus roborovskii. Russ J Ecol. 37:426–430.

Flower DR, North AC, Attwood TK. 1993. Structure and sequence relationships in the lipocalins and related proteins. Protein Sci. 2: 753–761.

Flower DR. 2000. Experimentally determined lipocalin structures. Biochim Biophys Acta. 1482:46–56.

Flower DR, North AC, Sansom CE. 2000. The lipocalin protein family: structural and sequence overview. Biochim Biophys Acta. 1482:9–24.

Henzel WJ, Rodriguez H, Singer AG, Stults JT, Macrides F, Agosta WC, Niall H. 1988. The primary structure of aphrodisin. J Biol Chem. 263: 16682–16687.

Hurst JL, Beynon RJ. 2010. Making progress in genetic kin recognition among vertebrates. J Biol. 9:13.

Hurst JL, Payne CE, Nevison CM, Marie AD, Humphries RE, Robertson DH, Cavaggioni A, Beynon RJ. 2001. Individual recognition in mice mediated by major urinary proteins. Nature. 414:631–634.

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 4:363–371.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227:680–685.

McFadyen DA, Addison W, Locke J. 1999. Genomic organization of the rat alpha 2u-globulin gene cluster. Mamm Genome. 10:463–470.

McFadyen DA, Locke J. 2000. High-resolution FISH mapping of the rat alpha2u-globulin multigene family. Mamm Genome. 11: 292–299.

Mucignat-Caretta C, Caretta A, Cavaggioni A. 1995. Acceleration of puberty onset in female mice by male urinary proteins. J Physiol (Lond). 486:517–522.

Mudge JM, Armstrong SD, McLaren K, Beynon RJ, Hurst JL, Nicholson C, Robertson DH, Wilming LG, Harrow JL. 2008. Dynamic instability of the major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and 129 strain mice. Genome Biol. 9:R91.

Natochin Iu V, Meshcherskii IG, Goncharevskaia OA, Makarenko IG, Shakhmatova EI, Ugriumov MV, Feoktistova N, Alonso G. 1994. [A comparative study of the osmoregulating system in the hamsters *Phodopus roborovskii* and *Phodopus sungorus*]. Zh Evol Biokhim Fiziol. 30:344–357.

Roberts SA, Simpson DM, Armstrong SD, Davidson AJ, Robertson DH, McLean L, Beynon RJ, Hurst JL. 2010. Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. BMC Biol. 8:75.

Robertson DH, Cox KA, Gaskell SJ, Evershed RP, Beynon RJ. 1996. Molecular heterogeneity in the Major Urinary Proteins of the house mouse Mus musculus. Biochem J. 316(Pt 1):265–272.

Robertson DH, Hurst JL, Searle JB, Gunduz I, Beynon RJ. 2007. Characterization and comparison of major urinary proteins from the house mouse, *Mus musculus domesticus*, and the aboriginal mouse, *Mus macedonicus*. J Chem Ecol. 33:613–630.

Schwende FJ, Wiesler D, Novotny M. 1984. Volatile compounds associated with estrus in mouse urine: potential pheromones. Experientia. 40: 213–215.

Sherborne AL, Thom MD, Paterson S, Jury F, Ollier WE, Stockley P, Beynon RJ, Hurst JL. 2007. The genetic basis of inbreeding avoidance in house mice. Curr Biol. 17:2061–2066.

Singer AG. 1991. A chemistry of mammalian pheromones. J Steroid Biochem Mol Biol. 39:627–632.

Stopkova R, Zdrahal Z, Ryba S, Sedo O, Sandera M, Stopka P. 2010. Novel OBP genes similar to hamster Aphrodisin in the bank vole, *Myodes glareolus*. BMC Genomics. 11:45.

Thom MD, Stockley P, Jury F, Ollier WE, Beynon RJ, Hurst JL. 2008. The direct assessment of genetic heterozygosity through scent in the mouse. Curr Biol. 18:619–623.

Vincent F, Lobel D, Brown K, Spinelli S, Grote P, Breer H, Cambillau C, Tegoni M. 2001. Crystal structure of aphrodisin, a sex pheromone from female hamster. J Mol Biol. 305:459–469.