Journal of Chemical Ecology, Vol. 28, No. 7, July 2002 (© 2002)

POLYMORPHISM IN MAJOR URINARY PROTEINS: MOLECULAR HETEROGENEITY IN A WILD MOUSE POPULATION

ROBERT J. BEYNON,[1,*] CHRISTINA VEGGERBY,[1] CAROLINE E. PAYNE,[2] DUNCAN H. L. ROBERTSON,[1] SIMON J. GASKELL,[3] RICK E. HUMPHRIES,[2] and JANE L. HURST,[2]

[1]*Protein Function Group*

[2]*Animal Behaviour Group*
Faculty of Veterinary Science
University of Liverpool
PO Box 147
Liverpool, L69 3BX, United Kingdom

[3]*Michael Barber Centre for Mass Spectrometry*
UMIST
PO Box 88
Manchester M60 1QD, United Kingdom

(Received November 13, 2001; accepted March 9, 2002)

Abstract—Major urinary proteins (MUPs) are present in high levels in the urine of mice, and the specific profile of MUPs varies considerably among wild-caught individuals. We have conducted a detailed study of the polymorphic variation within a geographically constrained island population, analyzing the MUP heterogeneity by isoelectric focusing and analytical ion exchange chromatography. Several MUPs were purified in sufficient quantities for analysis by electrospray ionization mass spectrometry and MALDI-TOF mass spectrometry of endopeptidase Lys-C peptide maps. The results of such analyses permitted the identification of three new MUP allelic variants. In each of these proteins, the sites of variation were located to a restricted segment of the polypeptide chain, projecting to a patch on the surface of the protein, and connected to the central lipocalin calyx through the polypeptide backbone. The restriction of the polymorphic variation to one segment of the polypeptide may be of functional significance, either in the modulation of ligand release or in communication of individuality signals within urinary scent marks.

Key Words—Major urinary proteins, mouse, MALDI-TOF mass spectrometry, protein purification, isoelectric focusing.

* To whom correspondence should be addressed. E-mail: r.beynon@liv.ac.uk

INTRODUCTION

The urine of house mice (*Mus domesticus*) contains a high concentration of protein, virtually all of which consists of members of an 18- to 19-kDa protein family termed "major urinary proteins," (MUPs). These MUPs are lipocalins (Flower et al., 1993), characterized by an archetypical eight-stranded β-barrel (Bocskei et al., 1992; Lucke et al., 1999; Zidek et al., 1999), enclosing a hydrophobic cavity. A generic role of the lipocalin family is the binding of small hydrophobic ligands (Flower et al., 1993, 2000), and in this respect, the MUPs are typical. Thus far, five endogenous MUP ligands have been reported; 2-*sec*-butyl-4,5-dihydrothiazole, 3,4-dehydro-*exo*-brevicomin, α- and β-farnesene, and 6-hydroxy-6-methyl-3-heptanone (Bocskei et al., 1992; Bacchini et al., 1992; Robertson et al., 1993; Novotny et al., 1999; Timm et al., 2001). All of these ligands are semiochemicals and are known to elicit primer pheromonal effects (Jemiolo et al., 1985, 1986).

Until recently, the necessity of protein binding for the activity of these ligands has not been understood, but we have provided direct behavioral and biochemical evidence that MUPs slow the release of semiochemicals from deposited urine samples (Hurst et al., 1998; Beynon et al., 1999). However, a simple role of ligand binding and release does not explain the observation that MUPs are encoded by a large, multigene complex and that urine from individual mice contains several MUPs, each a discrete gene product. Although there is some variation in ligand binding capability between different MUPs (Marie et al., 2001; Robertson et al., 2001), differential ligand release kinetics seems inadequate as an explanation for MUP diversity, since many MUPs show similar ligand binding properties. We have shown that protein fractions, depleted of natural volatile ligands by aging or by competitive displacement, elicit a strong countermarking response (Humphries et al., 1999) and, thus, MUPs or nonvolatile MUP–ligand complexes communicate chemical information in their own right. This is consistent with the hypothesis that proteins act as a stable signal of an animal's presence and are the medium through which scent marks evince the presence of a different individual. The growing evidence for other roles of MUPs in chemical signaling is also relevant. Brennan et al. (1999) suggested that these proteins provide one component of the strain recognition signal, and Mucignat-Caretta et al. (1995) showed that MUPs interact with receptors in the vomeronasal organ to exert an effect on puberty acceleration in females, even without the ligands. We have suggested that the MUPs, possibly in combination with ligands, have the ability to be used to communicate individuality within a mouse population (Beynon et al., 1999), and we recently provided direct evidence to support this hypothesis (Hurst et al., 2001).

Most of the information relating to MUPs has derived from studies of the proteins from relatively few inbred mouse strains (Robertson et al., 1996 and references therein). Since each strain is genetically homogeneous, it must be considered as a single snapshot of the wild mouse genome and, as such, yields little

information on the extent of genetic diversity in the wild house mouse. Accordingly, we have embarked on a study of the MUPs from wild mice. Our initial screen of a small number of wild mice from several sites in the UK yielded clear evidence of a hitherto unsuspected degree of polymorphism (Robertson et al., 1997; Pes et al., 1999; Payne et al., 2001). In this study, we have focused on MUP diversity from a single island population of mice to establish further the degree of polymorphism, particularly in a geographically and, therefore, genetically constrained population. The Isle of May (54 ha) in the Firth of Forth, approx. 10 km from the Scottish mainland, is uninhabited (except for wardens and a small number of researchers during the summer months) and contains a population of between a few hundred and a few thousand mice that vary cyclically over the year (Triggs, 1991). Indeed, until the introduction of a small number of mice from another island in 1982, the May mice were homozygous for over 70 loci, including the major histocompatibility complex (Berry and Peters, 1977; Berry et al., 1991). Subsequently, the new alleles penetrated rapidly through the population and reached equilibrium within three years (Figueroa et al., 1986; Jones et al., 1995). This rapid penetration may have been aided by the lack of territorial division in this population, leading to considerable genetic mixing. In contrast to the strongly territorial behavior normally seen among house mice living commensally in the built environment, competitive behavior appears to be virtually absent among feral males from the Isle of May (Gray and Hurst, 1998), and individual home ranges overlap extensively (Hurst, unpublished data). However, even within such an isolated and freely mixing population, where overall genetic variation is low, we should expect relatively high variation among individuals in cues that are used for individual recognition and perhaps also to recognize close relatives. Further identification of the sites of variation between the MUPs expressed by different individuals may indicate whether variation lies largely at the ligand-binding site (suggesting that individuality may be signaled through differential binding and release of volatile ligands) or lies on the surface of the protein for direct receptor-mediated detection. This paper reports detailed characterization of the MUPs expressed by different individuals within the Isle of May population and provides evidence for several new allelic variants not previously encountered.

METHODS AND MATERIALS

Chemicals. Acetonitrile, triflouroacetic acid, and methanol, all HiPerSolv grade, were obtained from BDH (Poole, England). Formic acid, NaCl, β-mercaptoethanol, EDTA, Tris, and trichloroacetic acid, all AnalaR grade, were obtained from BDH. Peptides (des-Arg bradykinin, neurotensin, ACTH, and insulin B-chain), horse heart myoglobin, Coomassie blue, Bistris, and α-cyano-4-hydroxycinnamic acid were obtained from Sigma (Poole, England). The Coomassie

Protein Plus assay including BSA standard was obtained from Pierce & Warriner, Chester, UK. The endoprotease Lys-C (sequencing grade) was obtained from Roche (Lewes, England).

Mice and Urine Collection. Male wild mice were captured from the population living on the Isle of May. The mice were housed individually in cages (30 × 13 × 12 cm) on peat substrate with shredded paper nest material and ad libitum food (Teklad Mouse/Rat diet 9609; Banting and Kingman, Hull, UK, and wheat grain) and water. Urine, voided in response to the mouse being held by the scruff of the neck, was collected into 1.5-ml Eppendorf tubes and frozen immediately. Many samples were no greater than 10–20 μl, which precluded exhaustive analysis.

Isoelectric Focusing (IEF). Isoelectric focusing was performed using Immobiline dryplates pH 4.2–4.9 (AP Biotech) on the Multiphor flatbed system cooled to 10°C with a Multitemp III thermostatic circulator (AP Biotech). Samples were diluted 1:10 in water, and 5 μl was loaded onto the gel at 200 V, 5 mA, and 15 W for 200 vh. Sample applicators were removed and the gel run at 3500 V, 5 mA, and 15 W for 14.8 kVh. The gel was stained with Coomassie Brilliant Blue in a Hoefer automated gel stainer.

Ion Exchange Chromatography (IEC). Urinary proteins were purified by high-resolution anion exchange chromatography on the SMART chromatography platform (AP Biotech). Prior to chromatography, 10 μl urine was desalted on micro Sephadex G25 spin columns and filtered through a 0.22-μm filter. Protein concentration of the desalted urine was measured (Coomassie Protein Plus assay). Fifteen micrograms of urinary protein was applied to a MiniQ PC 3.2/3 column (240 μl bed volume, Pharmacia) equilibrated with 20 mM Bistris buffer, pH 5.5, at a flow rate of 120 μl/min. Following sample application, the column was washed with 5 column volumes of 20 mM Bistris buffer, pH 5.5. Bound protein was eluted with a linear gradient of NaCl from 0 to 150 mM in the same buffer in 20 column volumes. Protein peaks, identified by absorbance at 280 nm, were collected. Peak fraction sizes were typically 60 μl, containing 1–3 μg of protein.

Electrospray Ionization Mass Spectrometry (ESI-MS). Peak fractions from IEC were desalted into 0.2% (v/v) formic acid on Microcon-10 concentrators (Millipore, Watford, England). Desalted samples were diluted 1:10 in 50% (v/v) acetonitrile–0.1% (v/v) formic acid. All analyses were performed on a VG Quattro mass spectrometer, upgraded to Quattro-II specifications (Micromass, Manchester, UK), fitted with an electrospray ionization source. Samples were introduced into the mass spectrometer as a continuous infusion at 5 μl/min. Raw data, gathered between *m/z* 700 and 1400, were subsequently refined and transformed to a true mass scale using Maximum Entropy software. Raw data were processed between 18,400 and 19,000 Da. The instrument was calibrated using a 2 pmol/μl solution of horse heart myoglobin in 50% (v/v) acetonitrile–0.1% (v/v) formic acid run after each sample. The calibration spectrum also served to determine the peak width parameter (0.75 Da) during Maximum Entropy processing.

Lys-C Digestion. The proteins in peak fractions from ion exchange chromatography were denatured in 4 M guanidinium thiocyanate for 1 hr and desalted into Lys-C digestion buffer (25 mM Tris, 1 mM EDTA, 2 mM β-mercaptoethanol, pH 8.5) on Microcon-10 concentrators. Sequencing grade Lys-C was added in a 1:100 (w/w) ratio. Digestion took place overnight at 37°C, and the reaction was stopped by the addition of 1 μl formic acid. A reference peptide digest of uMUP-I from BALB/C was prepared by digesting purified uMUP-I (IEC on the SMART system) as described above.

MALDI-TOF MS. Peptides were analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (PE Biosystems Voyager Elite, Warrington, UK), equipped with a 2-m linear and a 3-m reflectron flight mode and a delayed extraction function. A pulsed nitrogen laser emitting at 337 nm was used as a desorption/ionization source. In all experiments, mass spectrometry was performed in reflectron mode with positive ion detection. Instrument parameters were as follows: accelerating voltage 20 kV, grid voltage 73.5%, guide wire voltage 0.05%, and pulse delay time 175 nsec. Laser energy was ca. 1500. External mass calibration was performed with a mixture of des-Arg bradykinin, neurotensin, adrenocorticotrophic hormone, and insulin β-chain (50 nM each in 50% acetonitrile–0.1% triflouroacetic acid). To 1 μl of digested protein or peptide calibrants, 1 μl of α-cyano-4-hydroxycinnamic acid (saturated in 50% acetonitrile–0.1% triflouroacetic acid) was added, and 1 μl of the mixture was spotted onto the target. Signal-averaged mass spectra for 80–256 laser shots were collected, and data were analyzed using the PE Biosystems GRAMS/386 software.

RESULTS AND DISCUSSION

We have previously examined in detail the MUP allelomorphs from two inbred laboratory strains of mice: BALB/C and C57BL/6J (Robertson et al., 1996) in which we were able to reconcile urinary protein expression with known cDNA sequences and provide evidence, purely from protein analysis, for new gene products for which no cDNA sequences were known. These inbred strains, BALB/C and C57BL/6J, can be construed as two snapshots of the wild mouse genome, and our data provided evidence for at least 14 proteins potentially or actually expressed in these two strains. Further work on limited numbers of wild-caught specimens yielded clear evidence for remarkable diversity in MUP profiles (Pes et al., 1999; Payne et al., 2001), suggesting many as yet uncharacterized variants of these proteins. Accordingly, we embarked upon a systematic study of MUP expression in mice sampled from a closed, island population. In addition to yielding further information on MUP polymorphism, these samples were informative in studies on MUP pattern complexity within a closed but natural population, the results of which will be reported elsewhere. In this paper, we describe the isolation and characterization of the most abundant MUPs found in this population.

Survey of MUP Expression in Isle of May Mouse Urine Samples. Charge and exact mass are valuable parameters in the screening for novel MUPs (Robertson et al., 1997). Isle of May mouse urine samples separated on linear SDS–polyacrylamide gel electrophoresis showed a strong protein band at 18 kDa and little or no other protein constituents. When electroblotted to nitrocellulose, the 18 kDa band reacted strongly with an anti-MUP polyclonal antiserum (results not shown). Accordingly, we assumed that the detectable proteins in a mouse urine sample were predominantly MUPs, an assumption that has been repeatedly validated. For rapid initial screening of samples derived from Isle of May mice, urinary proteins from 17 wild caught animals were analyzed by IEF using the method described in Payne et al. (2001) (Figure 1). Whereas all individuals from any inbred mouse strain would have completely identical profiles in IEF (a BALB/C sample is included in the figure for comparison), the profiles of these 17 wild mice differed among individuals, although there was evidence of widespread expression of common protein subtypes. Between four and 10 bands were expressed in each of the Isle of May urine samples, but there was considerable overlap of protein expression in these individuals—a total of only 11 different bands could be discerned. By comparison with analyses of urine samples from mainland-caught mice, the patterns were much more similar among individuals, which may reflect the limited number of founder animals, or genetic bottlenecks due to poor overwinter survival

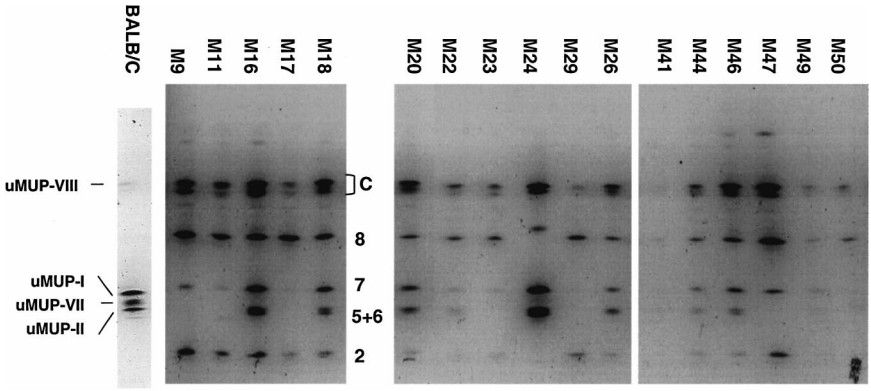

FIG. 1. Isoelectric focusing of MUPs from 17 wild mice from the Isle of May. Isoelectric focusing was performed using Immobiline dry plates, pH 4.2–4.9 (AP Biotech) on the Multiphor flatbed system (AP Biotech). Samples were diluted $1 + 9$ in water, 5 μl was applied and the gel was run at 3500 V, 5 mA, and 15 W for 14.8 kVh. A sample of urine from a BALB/C inbred mouse was included for comparison. Bands identified by analysis of proteins purified by ion exchange chromatography are labeled for cross-reference elsewhere in the paper. Known MUP variants from BALB/C urine are labelled with the standard nomenclature.

that is a feature of this geographically constrained population (Berry and Peters, 1977; Berry et al., 1991; Jones et al., 1995; Payne et al., 2001).

Chromatographic Separation of MUPs from Isle of May Mice. For further characterization of the proteins, MUPs from the wild-caught May mice and from inbred BALB/C and C57BL/6J mice were purified by high-resolution IEC (MiniQ column, SMART chromatography platform). A representative elution profile for one Isle of May mouse (M22), and the profiles from the well-characterized inbred strains BALB/C and C57BL/6J are given in Figure 2. A low loading (15 μg of protein) was necessary for reproducible and high-resolution separation of these mixtures of very similar proteins. The M22 mouse sample contained 11 discernible protein peaks, all eluting between 2.8 and 4.9 ml, thus displaying greater complexity than inbred mice (seven and five peaks for BALB/C and C57BL/6J, respectively). The first six and last three peaks were well separated, whereas the proteins at $V_e = 3.76$ ml and 3.86 ml overlapped and probably contained more than one protein each, as both were irregularly shaped with shoulders. The large, complex peak between $V_e = 3.76$ and 3.86 ml is a persistent feature of urine samples from Isle of May mice but is not present in either inbred strain. When comparing the peak elution volumes (V_e) to those of the inbred mice proteins, some overlap was found. Five peaks ($V_e = 2.97$, 3.38, 3.50, 3.76, and 4.82 ml) were matched by similar peaks in BALB/C ($V_e = 2.98$, 3.40, 3.49, 3.73, and 4.84 ml), and four May peaks ($V_e = 3.50$, 3.63, 3.76, and 4.82 ml) matched to equivalents in C57BL/6J ($V_e = 3.48$, 3.61, 3.77, and 4.83 ml), indicating the presence of proteins of similar charge in the three different samples.

Ion exchange chromatography identified more charge variants than IEF, although for the major proteins, the results of the two analyses correlated well. Chromatographically purified MUPs from M18 and from BALB/C were analyzed by IEF, and the peaks that could be identified are labeled in Figure 1. The BALB/C MUPs had been previously identified and were labeled according to the proposed nomenclature (Robertson et al., 1997). When the remainder of the Isle of May samples were analyzed by high-resolution IEC, similarly complex elution profiles were obtained, with 6–11 peaks expressed in each sample. These elution traces are not shown, but representative profiles are summarized in Figure 3. Because a constant amount of protein was applied to the column for each individual urine sample, comparison of the expression of different peaks was possible. The chromatographic traces revealed differences between almost every animal. Two proteins were found in a single individual: peak 1 was seen only in M15 and peak 9 was only expressed in M24. The V_e of these two proteins differed from the nearest groups by more than 2 SD and are, therefore, likely to be separate proteins. At the other extreme, several peaks (2, 3, 6, 7, 8, 11) were present in over 50% of the samples analyzed, the cluster proteins were found in all mice tested, and peaks 2, 7, and 8 were observed in all but one individual. Many individuals share the major proteins (2, 3, 5, 6, 7, 8, cluster, 11). Of the major proteins, peaks 5 and 6 demonstrated the

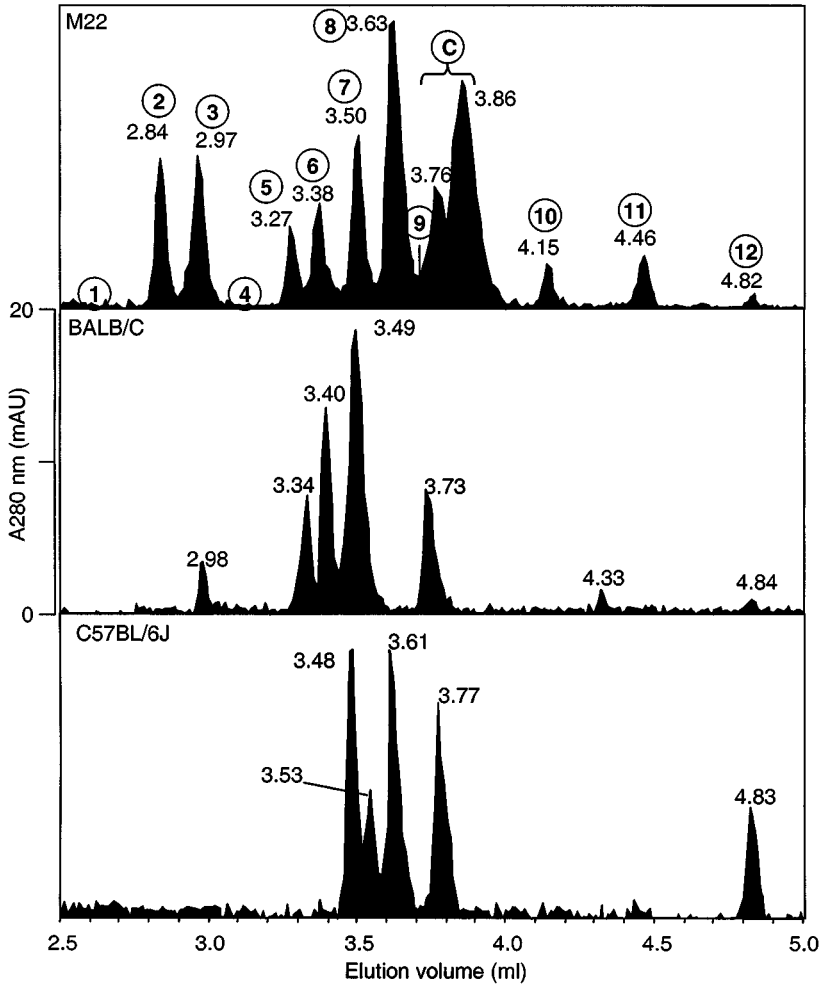

FIG. 2. Purification of MUPs from the M22 wild mouse. Urine samples from Isle of May mouse M22, BALB/C, and C57BL/6J were desalted into 20 mM Bistris buffer, pH 5.5, and 15 μg of protein was applied to a MiniQ column ($V_t = 240$ μl). Bound proteins were eluted with a linear salt gradient from 0 to 150 mM NaCl in 20 column volumes of the same buffer. The black areas define $A_{280\,nm}$. All three panels are displayed at the same scale, and each peak is annotated with the peak number and elution volume in milliliters.

greatest variability among individuals, but there was more pronounced variability in expression of the minor components. At present, we have no knowledge of the ability of mice to perceive differences in the relative concentrations of MUPs in a urine sample, and the significance of such variability is unknown.

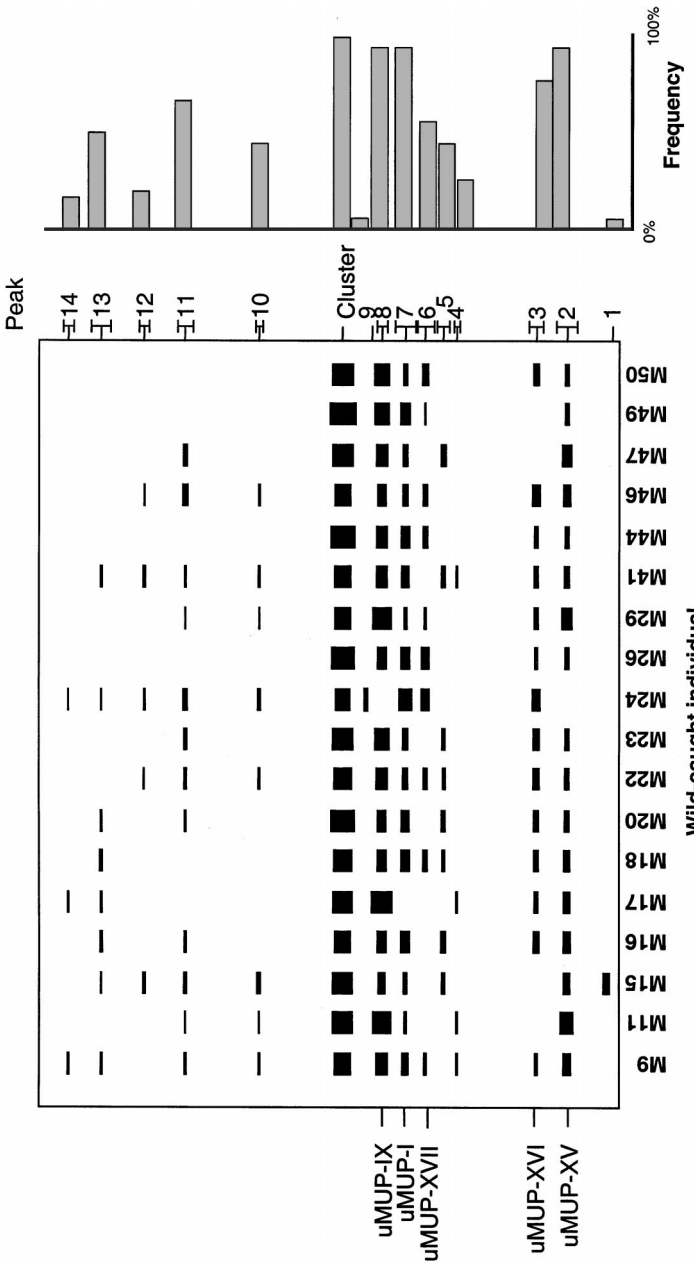

FIG. 3. Summary of IEC of MUPs from wild-caught mice from the Isle of May. MUPs from 18 Isle of May mice were purified by ion exchange chromatography on the SMART system, as described for M22 (see Figure 2 legend). Peaks from the different Isle of May mice were grouped according to V_e and are represented as bands at the average peak elution volume of the group. The width of the band represents the relative peak area (calculated as a percentage of the total peak area). The error bars to the right of the figure represent the mean $V_e \pm 2$ SD of each protein peak. The standard deviations of the elution volumes in the groups varied from 0.005 to 0.020 ml, which is comparable to that achieved from multiple runs of BALB/C (S.D. 0.01–0.02 ml for peaks 1–7, $N = 6$). The peaks between 3.69 and 3.88 ml are not completely resolved in any of the samples and are, therefore, presented as one cluster. Protein from groups annotated with an uMUP name are those that have been previously identified in inbred mouse strains or which were characterized as novel proteins by ESI-MS and peptide fingerprinting. The frequency of expression of each of the peaks is summarized to the right of the figure.

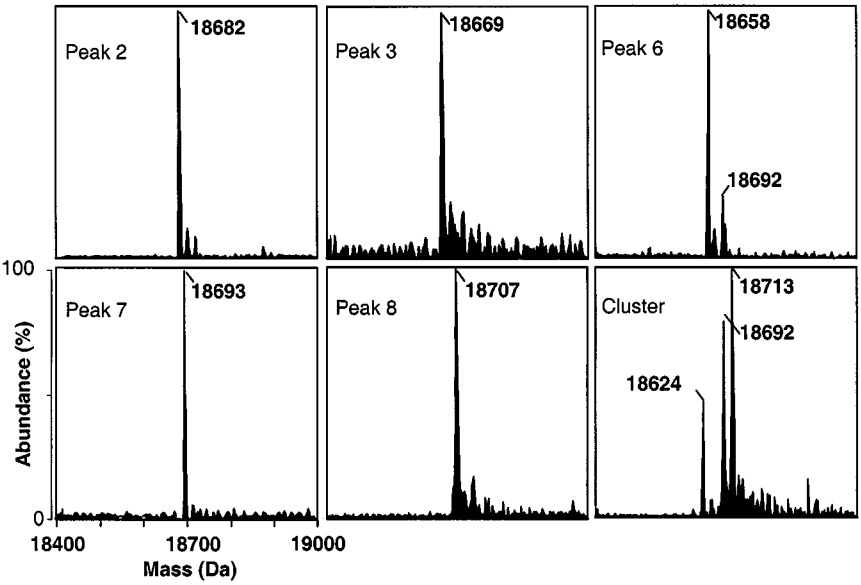

FIG. 4. Electrospray mass spectrometry analysis of purified M22 MUPs. IEC purified M22 urinary proteins were desalted into 0.2% formic acid and analyzed on a VG Quattro mass spectrometer, upgraded to Quattro-II specifications, fitted with an electrospray ionization source. Raw data, gathered between m/z 700 and 1400, were subsequently refined and transformed to a true mass scale using Maximum Entropy software. Raw data were processed between 18,400 and 19,000 Da. All six panels are displayed at the same scale.

Electrospray Ionization Mass Spectrometry of Purified Intact Proteins. To establish that the previously unseen species on IEF and IEC were new MUPs, the more abundant proteins were isolated by IEC of urine samples derived from individuals M9, M16, M18, M22, and M41. The major peaks (2, 3, 6, 7, 8, and the cluster mixture) were separately desalted into 0.2% formic acid and analyzed by electrospray ionization mass spectrometry (Figure 4). Each of the five main chromatographic peaks contained a predominant species, as assessed by accurate mass measurement (peak 6 contained a trace of a protein of mass 18,692 Da, possibly slight contamination from peak 7). The same chromatographic peaks from all five mice contained proteins of identical masses (mean $\pm$ SD): peak 2 was $18{,}682 \pm 0.3$ Da ($N = 5$), peak 3 was 18,668 Da (mean, $N = 2$), peak 6 was $18{,}658 \pm 0.9$ Da ($N = 3$), peak 7 was $18{,}693 \pm 0.5$ Da ($N = 5$), and peak 8 was $18{,}707 \pm 0.4$ Da ($N = 5$). From each mouse, the cluster contained protein species of three or four distinct masses, of which two (18,692 and 18,713 Da) had previously been observed in MUPs isolated from inbred animals (Robertson et al., 1997), and which could be reconciled to known cDNA sequences. It has not

proved possible to resolve the individual proteins in this cluster, and they have not been studied further.

Two of the proteins in the chromatographically pure peaks exhibited masses that were identical to the masses of proteins expressed in the urine of inbred mice (peak 7, 18,693 Da and peak 8, 18,707 Da) and that eluted in the same positions on IEC. The 18,693-Da peaks from BALB/C (uMUP-I) and Isle of May mice (peak 7) eluted at 3.49 ml and 3.50 ± 0.02 ml (mean $\pm$ SD), respectively, and the 18,707 Da peaks from C57BL/6J (uMUP-IX) and Isle of May mice (peak 8) eluted at 3.61 ml and 3.62 ± 0.01 ml, respectively. These proteins were, thus, indistinguishable on the basis of mass and charge, and we conjectured that these are the products of the same genes (subsequently proven by peptide mass fingerprinting, below).

Knowledge of the overall charge (isoelectric focusing and ion exchange chromatography) and intact protein mass are valuable parameters for the identification of new polymorphic variants of the proteins. However, these global parameters do not provide information on the site of protein sequence variation nor, indeed, do they exclude the possibility that these variants could be derived through posttranslational modifications. To characterize the new proteins further, we used MALDI-TOF peptide mass fingerprinting. The purified proteins were denatured and digested to limit peptides with the endopeptidase Lys-C, which cleaves specifically at the carboxyl side of lysyl residues. The resultant limit peptide mix was analyzed by MALDI-TOF mass spectrometry (Figure 5). Endopeptidase Lys-C is particularly valuable for the analysis of MUPs, since the limit digestion products (labeled L1 to L11 for the uMUP-I endopeptidase Lys-C limit peptides) were mostly large enough to be displaced from the chemical (mainly matrix) noise in the MALDI-TOF spectrum below 800 Da. All of the limit peptides, with the exception the two tripeptides L3 and L6, were observed and 96% (156/162) of the sequence was, thus, accounted for. As the protein sequence for uMUP-I was known (Genbank accession M16355), the peptide masses could be matched to that of the theoretical digest pattern for this protein, used as a reference for peptide mapping (Figure 6).

When comparing the peptide masses found for the five Isle of May MUPs to those of the BALB/C uMUP-I peptides, most of the Lys-C peptides were present in all samples. Those peptides corresponding to amino acid sequences from 1 to 28 (L1–L2) and 76 to 131 (L7–L9) were identical in all five MUPs analyzed. A further two peptides (L10 and L11) were variably present, due to a known amino acid polymorphism (Lys_{140} to Glu_{140}) that modifies a Lys-C site (Robertson et al., 1997) and, thus, in some proteins, the peptide (L10 + L11 − 18Da) was apparent. All other sequence variability in these proteins was located in peptides L4 and L5. For the Lys-C digest of the 18,682-Da protein (peak 2, expressed by all but one of the May mice), the expected 2095.9-Da L5 peptide was not present in the spectrum. Instead, a smaller 2047.4-Da peptide was present in the spectrum and is likely to correspond to the L5′ peptide arising from a $Phe_{56} \rightarrow Val_{56}$ mutation, as seen in uMUP-VII from BALB/C (Genbank accession X00907) (Kuhn et al., 1984;

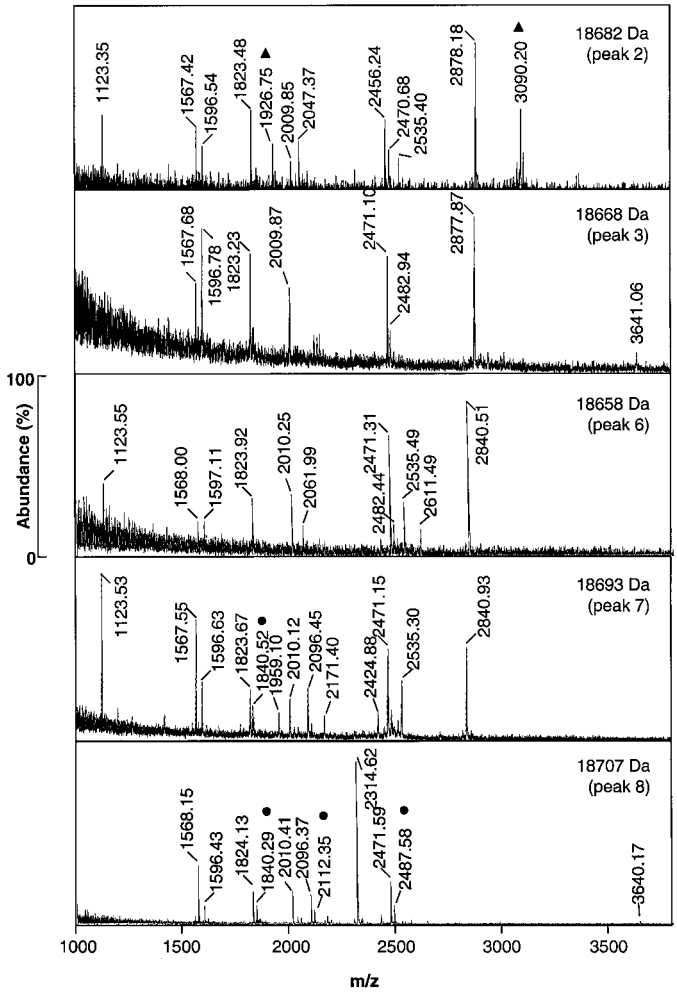

FIG. 5. MALDI-TOF mass spectrometry analysis of Lys-C digested purified M22 urinary proteins. Ion exchange chromatography-purified proteins were denatured in 4 M guanidine thiocyanate and desalted into Lys-C digestion buffer (25 mM Tris, 1 mM EDTA, 2 mM β-mercaptoethanol, pH 8.5) and digested with Lys-C overnight at 37°C. The peptides were analyzed on a PerSeptive Voyager Elite MALDI-TOF MS in reflectron mode. (**A**) Peak 2 (18682 Da), (**B**) Peak 3 (18669 Da), (**C**) Peak 6 (18658 Da), (**D**) Peak 7 (18693 Da), and (**E**) Peak 8 (18707 Da). Peaks marked with a closed circle are +18.003 Da peaks representing the oxidized methionine variant of the parent peak at lower mass. This serves as a useful confirmation of the presence of methionine in these peptides and their assignment to the peptide map. Peaks marked with a triangle are derived from autolysis of endopeptidase Lys-C (L1: 3090.49 Da and L2: 1926.83 Da).

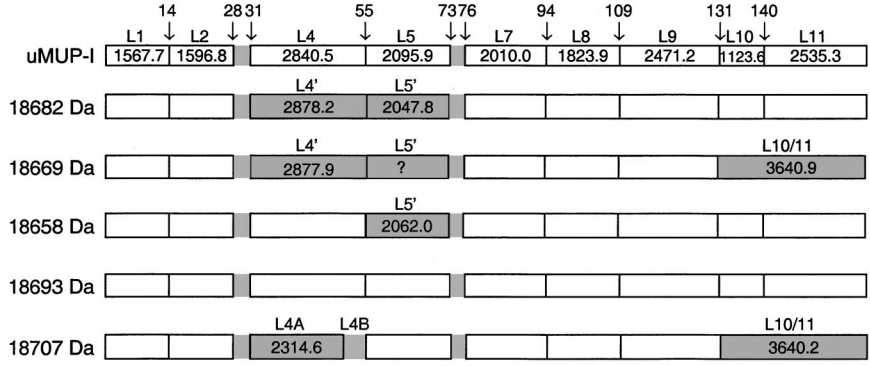

FIG. 6. Comparison of peptides obtained from Lys-C digestion of May mouse urinary proteins with that of a theoretical uMUP-I digest. A theoretical Lys-C digest of the protein sequence for uMUP-I was calculated using MassLynx software (Micromass, Manchester, UK) and compared with the MALDI-TOF results for the five Isle of May mouse peaks peaks. Lys-C peptides are numbered L1–L11. Some peptides are too small to be seen on the MALDI-TOF (L3, L6, and L4B). Blank boxes represent M22 peptides identical to the uMUP-I peptide. Peptide masses corresponding to known mutations found in inbred mice were also seen: $Asn_{50} \rightarrow Lys_{50} = 2314.2$ Da L4A and 559.4 Da L4B peptides, $Phe_{56} \rightarrow Val_{56} = 2047.9$ Da L5′ peptide, and $Lys_{140} \rightarrow Glu_{140} = 3640.8$ Da L10/11 peptide.

Robertson et al., 1997; Marie et al., 2001). Assuming that this was the L5′ peptide, an additional mass change had to account for the difference in the protein mass (10.7 Da smaller that uMUP-I). The 2840.5-Da L4 peptide was not present in the spectrum, but a prominent ion at 2878.2 Da was seen. This is 37.7 Da larger than the uMUP-I L4 peptide, and could, therefore, be a mutated L4′ peptide. The only single mass change of 38Da would be consequential to a Val to His mutation. Peptide L4′ contains two valyl residues, and it is not possible to be specific about which has been mutated. However, a Val to His mutation would also make the protein more cationic and less able to bind to the cation exchange column, which is exactly what is observed (Figure 2). The evidence clearly points to a new allelic variant, and this new protein was termed uMUP-XV according to the suggested nomenclature.

For the Lys-C digests of the 18,668-Da protein (peak 3, expressed by 14/18 males sampled), the L4, L5, L10, and L11 peptides of uMUP-I were not observed. All Lys-C peptide spectra for this protein contained ions at 3641 Da, which corresponds to the L10/11 peptide arising from the deletion of a Lys-C site by the mutation of Lys_{140} to Glu_{140}, as previously seen in the BALB/C protein uMUP-VIII [20]. The 2877.9-Da fragment is likely to be the L4′ peptide that was present in the digest of peak 2 material. Taking all these mass differences into account, peptide L5 would be expected to be 2037 Da. However, no peptide of this mass was present in any spectrum examined. It may be that L5′ in this protein contains

an additional Lys-C cleavage site that renders it too small for analysis. However, the implication is clear, inasmuch as there must be a further amino acid change in L5, which is the only peptide not accounted for. No peptides of the correct mass (i.e., the mass that would account for the remaining protein mass difference) were seen, and no substitution to Lys, thus generating an additional Lys-C cleavage site, could account for the mass differences or any observed peaks in the spectra. This too must be a new protein and is designated uMUP-XVI.

Lys-C digests of the 18,658-Da protein (peak 6, expressed by 10/16 males sampled) had only the L5 peptide mass peak missing, and the difference in protein mass must reside in this peptide. A 2062-Da peak was seen in the MALDI-TOF spectra, 34 Da smaller than the uMUP-I L5 fragment. This protein was termed uMUP-XVII. The third new MUP (uMUP-XVII) differed from uMUP-I at only one peptide: L5, by a decrease in mass of 34 Da. A number of single amino acid substitutions could account for this mass difference, of which a Phe→Ile/Leu substitution was deemed the most likely change. First, this was the only mutation that required a single base change. Second, the position at which the mutation occurred was F56, at which we had previously observed another mutation to valine (Kuhn et al., 1984; Robertson et al., 1997). Finally, the predicted average mass of the protein containing this mutation was 18,559 Da, in close agreement with the observed mass of 18,558 Da. This residue is directed in the central calyx of the MUP structure, and we have demonstrated that the F56V variant demonstrates altered binding of fluorescent probes (Marie et al., 2001). Thus, uMUP-VII from BALB/C, uMUP-X from C57BL/6J, and now uMUP-XVII from Isle of May mice are all variant at this position. All the amino acids at this position are hydrophobic, consistent with their disposition towards the hydrophobic central calyx. UMUP-XVII eluted at the same position on ion exchange chromatography as uMUP-VII from BALB/C, further evidence for the absence of amino substitutions that would evince a charge shift (such as Arg to Glu).

MALDI-TOF mass spectrometry of the Lys-C digest of the 18,693-Da protein (peak 7, expressed in all 17 samples, Figure 3), matched exactly that of uMUP-I. This protein had the same overall charge (i.e., it eluted at the same volume in IEC), the same mass, and resulted in the same Lys-C peptides as uMUP-I, suggesting that it was the same protein. Similar data were obtained for the 18,707-Da protein (peak 8, expressed in all but one of 17 samples, Figure 3), which was identical to uMUP-IX with respect to overall charge, mass, and pattern of Lys-C peptides. These two proteins, initially characterized in inbred mouse strains were, thus, also present in this wild population.

From the five purified proteins mapped by MALDI-TOF mass spectrometry, three new MUPs were identified, each containing at least one mutation. Our previous investigations of MUPs in wild house mice have revealed a much higher degree of heterogeneity than in inbred mice, both as to the number of MUPs expressed per mouse and profile differences between mice even from the same population

(Robertson et al., 1997; Pes et al., 1999; Payne et al., 2001). Although the degree of MUP molecular heterogeneity in Isle of May mice is less than that from other wild mainland populations (Payne et al., 2001), and the extent of sharing of the major protein peaks is substantial, individual MUP profiles were generally more complex than inbred mice, and there was substantially more individual variability in the expression of minor MUP species. This is consistent with selection pressure to maintain a high level of heterozygosity at the MUP loci as predicted.

Although roles for MUPs in ligand release (Hurst et al., 1998) and in direct signaling (Humphries et al., 1999; Brennan et al., 1999) have been established, neither function offers an explanation for the considerable polymorphisms in the MUP family or the combinatorial diversity that ensues from coexpression of multiple proteins in a single urine scent mark. We have proposed that the patterns of MUPs in a scent mark might provide a nonvolatile, stable, and complex label of ownership of the scent mark and, thus, deliver the complex olfactory signal of health, sexual status, and relatedness in the context of an individual animal (Beynon et al., 2001; Malone et al., 2001). We have shown that MUPs contribute to individuality signaling in urine deposits (Hurst et al., 2001) although we do not yet know the relative contribution of the major and minor protein species to this process. This would imply a receptor, or family of receptors, for the MUPs, and the evidence for such a receptor is accumulating. Analysis of the primary sequence of uMUPs, whether from cDNA sequences, where there is direct proof of expression, or of new proteins that we have analyzed by mass spectrometry indicates that much of the heterogeneity resides in endopeptidase LysC limit peptides L4 and L5 (Figure 7).

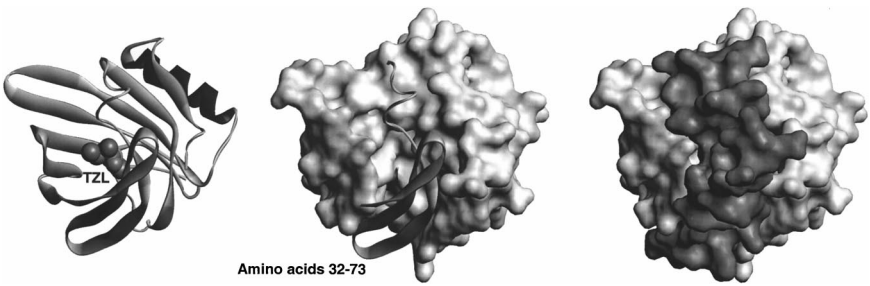

FIG. 7. Three-dimensional structure of MUP showing location of polymorphic variation. The structure of MUP [1MUP.PDB (Bocskei et al., 1992)] was used to generate these views of the protein. The left hand panel is a simple ribbon representation, showing the secondary structure and the approximate position of the ligand 2-*sec*-butyl-4,5-dihydrothiazole [although the disposition of this ligand has recently been revised (Zidek et al., 1999)]. Center panel: the segment of the polypeptide chain representing amino acids 32–73 (Lys C peptides L4 and L5, see text) is displayed in ribbon structure, superimposed on a space filling representation of the remainder of the protein. The right hand panel is a space filling representation in which the L4–L5 segment is shaded darker.

This area of the protein forms an extended patch on the surface of the protein, extended into a beta configuration. In such an extended conformation, the residues both protrude into the central cavity and towards solvent. If there were a selection for amino acid diversity directed to this patch of the surface, it might also lead to changes in cavity residues, which might, therefore, be consequential to the surface effects. Our initial analyses of the kinetics of ligand release have indicated that there are no major differences in the overall release–time curve when a comparison is made of urine or MUP samples derived from different strains and containing different proteins (Robertson et al., 2001). However, such studies do not reflect the environment proximal to a putative MUP receptor, and it is not unreasonable to consider coupling between externally facing residues that interact with a receptor and internally facing cavity-forming residues, such that the binding of the former alters cavity binding characteristics and ligand release.

Acknowledgments—This work was funded by grants from the BBSRC to R.J.B. and J.L.H. and a grant from NERC to J.L.H. The help of Samantha Gray in collecting wild mice from the Isle of May is gratefully acknowledged.

REFERENCES

BACCHINI, A., GAETANI, E., and CAVAGGIONI, A. 1992. Pheromone binding proteins of the mouse, *Mus musculus. Experientia* 48:419–421.

BERRY, R. J. and PETERS, J. 1977. Heterogeneous heterozygosities in *Mus musculus* populations. *Proc. R. Soc. London* B 197:485–503.

BERRY, R. J., TRIGGS, G. S., KING, B., NASH, H. R., and NOBLE, L. R. 1991. Hybridization and gene flow in house mice introduced into an existing population on an island. *J. Zool. London* 225:615–632.

BEYNON, R. J., ROBERTSON, D. H. L., HUBBARD, S. J., GASKELL, S. J., and HURST, J. L. 1999. The role of protein binding in chemical communication: Major urinary proteins in the house mouse, pp. 137–147, *in* R. E. Johnston, D. Muller-Schwarze, and P. Sorensen (eds.). Advances in Chemical Communication in Vertebrates. Plenum Press, New York.

BEYNON, R. J., HURST, J. L., GASKELL, S. J., HUBBARD, S. J., HUMPHRIES, R. E., MALONE, N., MARIE, A. D., MARTINESEN, L., NEVISON, C. M., PAYNE, C. E., ROBERTSON, D. H. L., and VEGGERBY, C. 2001. Mice, MUPs and myths: Structure-function relationships of the major urinary proteins, pp. 149–156, *in* A. Marchelewska-Koj, D. Muller-Schwarze, and J. Lepri (eds.). Chemical Signals in Vertebrates, Vol. 9. Plenum Press, New York.

BOCSKEI, Z., GROOM, C. R., FLOWER, D. R., WRIGHT, C. E., PHILLIPS, S. E. V., CAVAGGIONI, A., FINDLAY, J. B. C., and NORTH, A. C. T. 1992. Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* 360:186–188.

BRENNAN, P. A., SCHELLINCK, H. M., and KEVERNE, E. B. 1999. Patterns of expression of the immediate-early gene egr-1 in the accessory olfactory bulb of female mice exposed to pheromonal constituents of male urine. *Neuroscience* 90:1463–1470.

FIGUEROA, F., TICHY, H., BERRY, R. J., and KLEIN, J. 1986. MHC polymorphism in island populations of mice. *Curr. Top. Microbiol. Immunol.* 127:100–105.

FLOWER, D. R., NORTH, A. C. T., and ATTWOOD, T. K. 1993. Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci.* 2:753–761.

FLOWER, D. R., NORTH, A. C., and SANSOM, C. E. 2000. The lipocalin protein family: Structural and sequence overview. *Biochim. Biophys Acta* 1482:9–24.

GRAY, S. J. and HURST, J. L. 1998. Competitive behaviour in an island population of house mice, *Mus domesticus. Anim. Behav.* 56:1291–1299.

HUMPHRIES, R. E., ROBERTSON, D. H. L., BEYNON, R. J., and HURST, J. L. 1999. Unravelling the chemical basis of competitive scent marking in house mice. *Anim. Behav.* 58:1177–1190.

HURST, J. L., ROBERTSON, D. H. L., TOLLADAY, U., and BEYNON, R. J. 1998. Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. *Anim. Behav.* 55:1289–1297.

HURST, J. L., PAYNE, C. E., NEVISON, C. M., MARIE, A. D., HUMPHRIES, R. E., ROBERTSON, D. H. L., CAVAGGIONI, A., and BEYNON, R. J. 2001. Individual recognition in mice mediated by major urinary proteins. *Nature* 414:631–634.

JEMIOLO, D., ALBERTS, J., SOCHINSKI-WIGGINS, S., HARVEY, S., and NOVOTNY, M. 1985. Behavioural and endocrine responses of female mice to synthetic analogs of volatile compounds in male urine. *Anim. Behav.* 33:1114–1118.

JEMIOLO, B., HARVEY, S., and NOVOTNY, M. 1986. Promotion of the Whitten effect in female mice by synthetic analogs of male urinary constituents. *Proc. Natl. Acad. Sci. USA* 83:4576–4579.

JONES, C. S., NOBLE, L. R., JONES, J. S., TEGELSTROM, H., TRIGGS, G. S., and BERRY, R. J. 1995. Differential male genetic success determines gene flow in an experimentally manipulated mouse population. *Proc. R. Soc. London B* 260:251–256.

KUHN, N. J., WOODWORTH-GUTAI, M., GROSS, K. W., and HELD, W. A. 1984. Subfamilies of the mouse major urinary protein (MUP) multi-gene family: Sequence analysis of cDNA clones and differential regulation in the liver. *Nucleic Acids Res.* 12:6073–6090.

LUCKE, C., FRANZONI, L., ABBATE, F., LOHR, F., FERRARI, E., SORBI, R. T., RUTERJANS, H., and SPISNI, A. 1999. Solution structure of a recombinant mouse major urinary protein. *Eur. J. Biochem.* 266:1210–1218.

MALONE, N., PAYNE, C. E., BEYNON, R. J., and HURST, J. L. 2001. Social status, odour communication and mate choice in wild house mice, pp. 217–224, *in* A. Marchlewska-Koj, D. Muller-Schwarze, and J. Lepri (eds.). Chemical Signals in Vertebrates, Vol 9. Plenum Press, New York.

MARIE, A. D., VEGGERBY, C., ROBERTSON, D. H. L., GASKELL, S. J., HUBBARD, S. J., MARTINSEN, L., HURST, J. L., and BEYNON, R. J. 2001. Effect of polymorphisms on ligand binding by mouse major urinary proteins. *Protein Sci.* 10:411–417.

MUCIGNAT-CARETTA, C., CARETTA, A., and CAVAGGIONI, A. 1995. Acceleration of puberty onset in female mice by male urinary proteins. *J. Physiol.* (*London*) 486:517–522.

NOVOTNY, M. V., MA, W., WIESLER, D., and ZIDEK, L. 1999. Positive identification of the puberty-accelerating pheromone of the house mouse: The volatile ligands associating with the major urinary protein. *Proc. R. Soc. London B* 266:2017–2022.

PAYNE, C. E., MALONE, N., HUMPHRIES, R. E., BRADBROOK, C., VEGGERBY, C., BEYNON, R. J., and HURST, J. L. 2001. Heterogeneity of Major Urinary Proteins in house mice: Population and sex differences, pp. 233–240, *in* A. Marchelewska-Koj, D. Muller-Schwarze, and J. Lepri (eds.). Chemical Signals in Vertebrates, Vol. 9. Plenum Press, New York.

PES, D., ROBERTSON, D. H. L., HURST, J. L., GASKELL, S. J., and BEYNON, R. J. 1999. How many major urinary proteins are produced by the house mouse *Mus domesticus*? pp. 149–161, *in* R. E. Johnston, D. Muller-Schwarze, and P. Sorensen (eds.). Advances in Chemical Communication in Vertebrates, Plenum Press, New York.

ROBERTSON, D. H. L., BEYNON, R. J., and EVERSHED, R. P. 1993. Extraction, characterization and binding analysis of two pheromonally active ligands associated with major urinary protein of house mouse (*Mus musculus*). *J. Chem. Ecol.* 19:1405–1416.

ROBERTSON, D. H., COX, K. A., GASKELL, S. J., EVERSHED, R. P., and BEYNON, R. J. 1996. Molecular heterogeneity in the major urinary proteins of the house mouse *Mus musculus. Biochem. J.* 316:65–272.

ROBERTSON, D. H. L., HURST, J. L., BOLGAR, M. S., GASKELL, S. J., and BEYNON, R. J. 1997. Molecular heterogeneity of urinary proteins in wild house mouse populations. *Rapid Commun. Mass Spectrum.* 11:786–790.

ROBERTSON, D. H. L., MARIE, A. D., VEGGERBY, C., HURST, J. L., and BEYNON, R. J. 2001. Characteristics of ligand binding and release by major urinary proteins. pp. 169–176, *in* A. Marchlewska-Koj, D. Muller-Schwarze, and J. Lepri (eds.), Chemical Signals in Vertebrates, Plenum Press, New York.

TIMM, D. E., BAKER, L. J., MUELLER, H., ZIDEK, L., and NOVOTNY, M. V. 2001. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Sci.* 10:997–1004.

TRIGGS, G. S. 1991. The population ecology of house mice (*Mus domesticus*) on the Isle of May, Scotland. *J. Zool. London* 225:449–468.

ZIDEK, L., STONE, M. J., LATO, S. M., PAGEL, M. D., MIAO, Z., ELLINGTON, A. D., and NOVOTNY, M. V. 1999. NMR mapping of the recombinant mouse major urinary protein I binding site occupied by the pheromone 2-sec-butyl-4,5-dihydrothiazole. *Biochemistry* 38:9850–9861.