Molecular heterogeneity in the Major Urinary Proteins of the house mouse *Mus musculus*

Duncan H. L. ROBERTSON*, Kathleen A. COX†, Simon J. GASKELL†, Richard P. EVERSHED‡ and Robert J. BEYNON*§

*Department of Biochemistry and Applied Molecular Biology, †Michael Barber Centre for Mass Spectrometry, UMIST, P.O. Box 88, Manchester M60 1QD, U.K., and ‡School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, U.K.

Major Urinary Proteins (MUPs) from different inbred strains of mouse have been analysed by high-resolution ion-exchange chromatography and mass spectrometry. MUPs from six strains were resolved chromatographically into four major protein peaks which characterized two distinct phenotypes, typified by the profiles obtained from the Balb/c and C57BL/6 inbred strains. A combination of ion-exchange chromatography and electrospray ionization mass spectrometry analysis of the MUPs from each strain identified five proteins, only one of which was common to both strains. The charge and mass data, together with N-terminal sequence analyses, were correlated with the masses of the proteins inferred from published cDNA sequences. Several members of the family of MUP sequences differ in only four positions, and in some circumstances the substitutions elicit a minimal change in protein mass (Lys/Gln; Lys/Glu). Peptide mapping with endopeptidase Lys-C, followed by matrix-assisted laser desorption ionization–time-of-flight mass spectrometry permitted identification of new MUPs that were correlated with partial cDNA sequence data. In the two strains there are at least 13 different MUPs, either observed or predicted, indicating the heterogeneity of expression of this group of proteins.

INTRODUCTION

As nocturnal burrowing animals, mice have evolved chemical methods of communication in place of the physical methods evolved by humans. Possibly the most important, and certainly the best documented, system of chemical communication in mice is olfaction. The main source of olfactory chemosignals in the mouse is urine, which has been shown to mediate many behavioural and physiological responses, including inter-male aggression [1], puberty onset [2] oestrous cycling [3] and mate selection [4]. Among all the animals that use olfaction, mice and a few closely related rodents are unique in that they secrete large amounts of protein into their urine. This protein fraction of the urine is reported to demonstrate pheromonal activity [5].

Further investigation of the protein content of mouse urine has shown it to consist predominantly of a group of closely related proteins termed the Major Urinary Proteins (MUPs). These are acidic proteins (pI values from 4.2 to 4.7) with molecular masses of approx. 19 kDa [6]. MUPs are the product of a multigene family of approx. 30 genes and pseudogenes located on chromosome 4 [7]. The expression of MUP differs between the sexes: males express considerably more protein than females do [8]. This sex-dependent expression pattern is probably regulated by circulating growth hormone levels. Females express male levels of MUPs when injected with growth hormone to mimic the male circulating pattern [9]. Further, the pattern of expressed MUPs varies between inbred strains [10], which demonstrate different MUP phenotypes when examined by isoelectric focusing [11]. It has been proposed that the differences in MUP phenotype arise as a result of allelic variation at at least four loci [12].

We and others have demonstrated that MUPs are associated with two pheromonally active ligands, 3,4-dehydro-exo-brevicomin and 2-sec-butyl-4,5-dihydrothiazole [13–15]. There is also some binding specificity of these two ligands to MUPs separated by ion-exchange chromatography. Thus MUPs play a role in binding small, hydrophobic ligands that are known to possess the capability of chemical signalling. On the basis of primary and tertiary structural homology, MUPs have been assigned to the lipocalin superfamily of proteins [16–18]. The lipocalins have a distinct tertiary structure consisting of eight, nine or ten β-sheets arranged in a β-barrel to form a central hydrophobic pocket or calyx into which the ligand is inserted. Current thinking emphasizes a role for the MUPs in either the timed release of odorants or in their protection from oxidation. To investigate differences in ligand binding between individual MUPs and between different strains, we have characterized MUPs from several inbred strains by ion-exchange chromatography and mass spectrometry. This information is an essential prerequisite for future studies on the tertiary structure of the molecule and its ligand binding characteristics. This protein characterization has also permitted correlation between MUPs that are proved to be expressed and the various published cDNA sequences.

MATERIALS AND METHODS

Animals and urine collection

Pooled urine was collected by bladder massage from male Balb/cJ, DBA/J, CBA/J, A/J, C57BL/6J and C57BL/10J mice housed in groups of 8–16 in standard conditions. Pooled urine was desalted on spun 5 ml Sephadex G-25 columns previously

Abbreviations used: ESI/MS, electrospray ionization mass spectrometry; MALDI-TOF/MS, matrix-assisted laser desorption ionization–time-of-flight mass spectrometry; MUP, Major Urinary Protein.

§ To whom correspondence should be addressed.

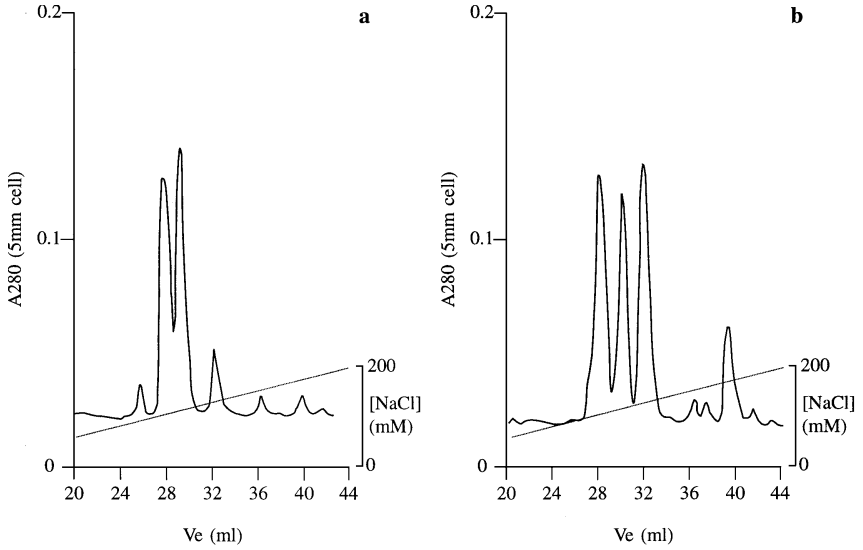

Figure 1 Anion-exchange chromatography of MUPs from different inbred strains of mouse

Urine from six strains of mouse, Balb/cJ, DBA/J, CBA/J, A/J, C57BL/6J and C57BL/10J, was desalted into 50 mM Mes buffer, pH 5.0, and a 100 μl aliquot of each was applied to a Mono-Q anion exchange column (V_t 1 ml). After a brief wash with buffer, bound protein was eluted with a linear gradient of 0–200 mM NaCl. Proteins analysed in this manner fell into one of two phenotypes, (**a**) and (**b**), the profiles of which are shown.

equilibrated in 50 mM Mes buffer, pH 5.0. Aliquots (250 μl) were eluted by centrifugation at 200 ***g*** for 1 min at room temperature.

Analytical anion-exchange chromatography of MUPs

Analytical anion-exchange chromatography was performed with a Pharmacia FPLC system fitted with a Mono-Q column (V_t 1 ml). The column was equilibrated with 10 ml of 50 mM Mes buffer, pH 5.0, before application of 100 μl of desalted urine in the same buffer. This contained approx. 0.2 mg of protein as estimated with a BCA assay kit (Pierce Chemicals). Bound protein was eluted from the column with a linear salt gradient of 0–200 mM NaCl in 22 ml.

Purification of MUPs by anion-exchange chromatography

Purification of individual MUP peaks was also achieved by anion-exchange chromatography on a Pharmacia FPLC system. In this instance the system was fitted with a Resource-Q column (V_t 6 ml). The column was equilibrated with 60 ml of 50 mM Mes buffer, pH 5.0, before application of either 2 ml of Balb/c or 3 ml of C57BL/6 desalted urine. Bound protein was washed with 44 ml of the above buffer before being eluted from the column with a linear salt gradient of 0–200 mM in 144 ml. Fractions (1 ml) were collected between 107 and 188 ml. Peaks resulting from this separation were pooled and diluted 1:4 with starting buffer and reapplied to the column with the same protocol. Fractions containing the largest single eluted peak were pooled and used in further analyses.

Before electrospray ionization mass spectrometry (ESI/MS), samples purified in the above manner were concentrated by HPLC with a single-pump LKB Bromma system fitted with a Hichrom RP8 (C8) column. The column was equilibrated in 10 ml of 0.1 % trifluoroacetic acid at 1 ml/min before injecting 2 ml of purified protein as 10 × 200 μl injections at 1 min intervals. The column was washed with a further 5 ml of 0.1 % trifluoroacetic acid before application of a linear acetonitrile gradient of 0–80 % in 30 ml. Protein eluted in this manner was monitored by absorbance at 280 nm (detector sensitivity 1.0 absorption unit full scale), collected as a pool and used for subsequent ESI/MS analyses.

ESI/MS

ESI/MS was performed either with a VG Quattro mass spectrometer for whole desalted urines or with a similar instrument upgraded to Quattro II specifications for all other analyses. The instrument was tuned and calibrated with a 20 pmol/μl solution of horse heart myoglobin made up in 50 % (v/v) aqueous acetonitrile, 1 % (v/v) formic acid. For the samples purified from anion-exchange profiles, the horse heart myoglobin was added to the sample as an internal standard (2 μl per 100 μl of sample). In these cases the instrument was calibrated with the internal standard spectrum. Each sample was introduced into the mass spectrometer as 2 × 20 μl injections at 2 min intervals (purified MUPs) and as 5 × 20 μl injections (desalted urines). In the latter, samples were prepared by diluting 10 μl of desalted MUPs with 50 μl of 2 % (v/v) formic acid, 40 μl of de-ionized water and 100 μl of acetonitrile (HPLC grade). The samples were introduced into the source at a flow rate of 10 μl/min in a carrier solvent of 50 % (v/v) aqueous acetonitrile.

Acquisition of data was controlled by VG Lab-Base software (desalted MUPs) or by VG MassLynx (purified MUPs). Raw spectra were subsequently deconvoluted using the 'MaxEnt' maximum-entropy software incorporated into the VG MassLynx package. All spectra were processed at 1 Da per channel over a mass range of 18300–18900 Da.

Peptide mapping of purified MUPs

Balb/cJ MUPs were purified by ion-exchange chromatography as described above. The two latest eluting peaks (V_t 122.9 and 132.9 ml) were concentrated and desalted to deionized water in a

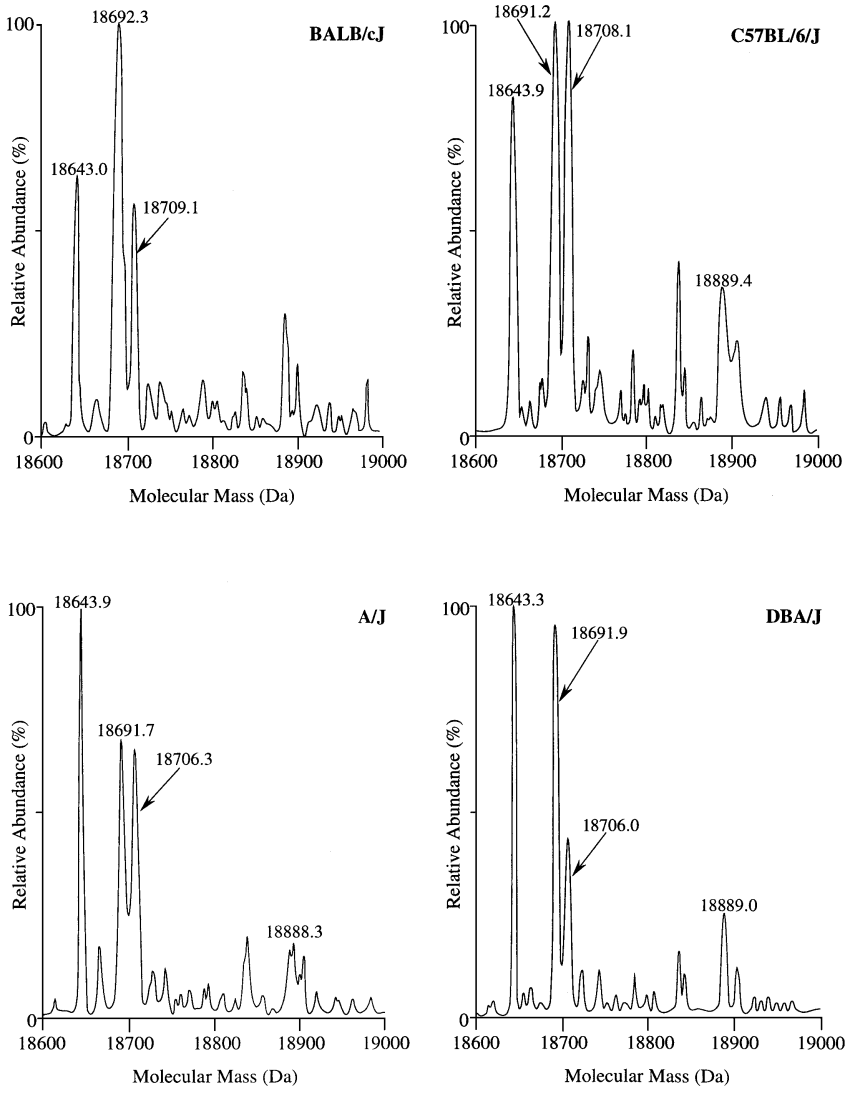

Figure 2 ESI/MS of MUPs from four different inbred strains of mouse

Desalted urine (10 μl) was diluted with 40 μl of deionized water, 50 μl of 2% (v/v) formic acid and 100 μl of acetonitrile. This solution was then injected into a VG Quattro triple quadrupole mass spectrometer fitted with an electrospray ionization source. The acquired mass-to-charge spectra were processed with maximum-entropy software included in the VG MassLynx package.

Centricon concentrator (Amicon) with a 10 kDa cut-off membrane. A 200 μl aliquot of this preparation was first denatured by addition of an equal volume of 8 M guanidinium thiocyanate (BDH) and then reduced with 20 μl of 2-mercaptoethanol followed by incubation at room temperature for 2 h. The solution was then treated with 30 μl of 10 M hydrochloric acid to precipitate the protein. The precipitate was pelleted by centrifugation at 11000 ***g*** for 2 min at room temperature in a microcentrifuge and washed twice with deionized water before being resuspended in 200 μl of digestion buffer (50 mM Tris/HCl, 2 mM EDTA, 2 mM 2-mercaptoethanol, pH 8.5).

For digestion with endopeptidase Lys-C, a 20 μl aliquot of the above suspension was added to an equal volume of 0.1 mg/ml sequencing-grade endopeptidase Lys-C (Boehringer-Mannheim) in 50 mM Tricine buffer, pH 8.0, containing 10 mM EDTA, as supplied by the manufacturers. This was incubated at 37 °C overnight and the reaction was then stopped by addition of 8 μl of formic acid.

Matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (MALDI-TOF/MS)

MALDI-TOF/MS was performed with a VG TofSpec-E instrument operated in the reflectron mode. Each Lys-C digest was diluted with four volumes of water containing 0.1% (v/v) trifluoroacetic acid. A portion was mixed with an equal volume of a saturated solution of 2,5-dihydroxybenzoic acid in 50% aqueous acetonitrile containing 0.1% trifluoroacetic acid and 50 mM fucose. A 2 μl aliquot of this mixture was air-dried on the MALDI sample target.

RESULTS

Anion-exchange chromatography (Mono-Q, FPLC) was performed on MUPs from six inbred strains of mouse: Balb/cJ, A/J, CBA/J, DBA1/J, C57BL/6J and C57BL/10J. The MUPs from these strains consistently and reproducibly demonstrated two distinct phenotypes (Figure 1). The first phenotype (Figure

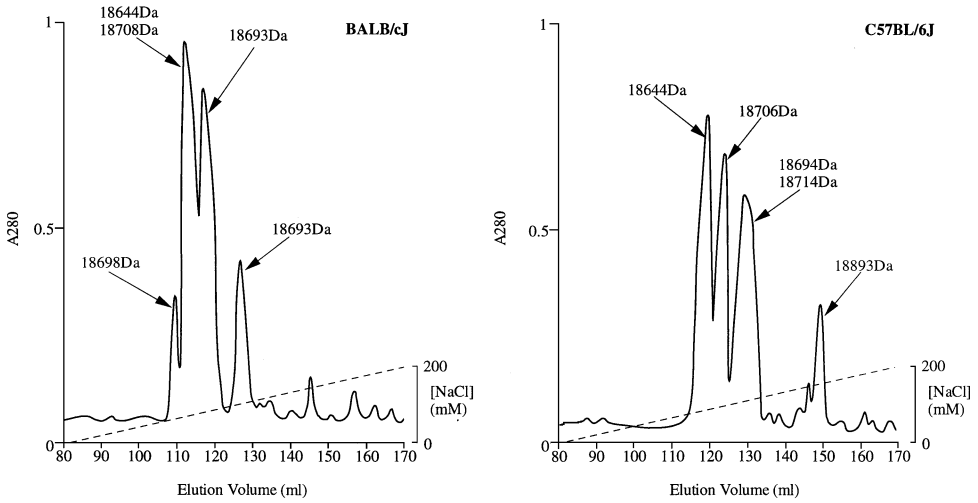

Figure 3 Analysis of Balb/cJ and C57BL/6J MUPs by anion-exchange chromatography and ESI/MS

A 2 ml aliquot of urine from Balb/c or C57BL/6J mice was desalted into 50 mM Mes buffer, pH 5.0, and applied to a Resource-Q anion-exchange column (V_t 6 ml) pre-equilibrated in 60 ml of the same buffer. Bound protein was eluted from the column using a linear NaCl gradient of 0–200 mM in 144 ml. Fractions (1 ml) containing the four main peaks were collected and the individual peaks were pooled. Each pool was diluted 1:4 with the above buffer before being reapplied to the column and chromatographed under the same conditions. Fractions containing the largest peak were then pooled and concentrated by reverse-phase HPLC. The last two steps served to emphasize the homogeneity of the applied material and to effect a buffer change: no further separation was achieved. Protein prepared in this manner was then analysed by ESI/MS, and the molecular masses present in each peak are indicated on the trace.

1a) consists of four major peaks between elution volume (V_e) 24 and 34 ml, whereas the second phenotype (Figure 1b) has only three major peaks between V_e 24 and 34 ml with an additional peak at 39 ml. Each of these peaks consisted of a single band on SDS/PAGE, migrating at approx. 18 kDa, which reacted strongly with a rabbit polyclonal antiserum raised to MUPs (results not shown). Only the fourth major peak in phenotype (a) has a retention time identical with the third major peak in phenotype (b). The relative peak areas are also significantly different in both profiles. Balb/cJ, A/J, CBA/J and DBA1/J strains all exhibit the (a) phenotype, whereas C57BL/6J and C57BL/10J strains exhibit the (b) phenotype. From the genealogical tree of inbred mouse strains [19], all the strains of the (a) phenotype share a common ancestor distinct from the strains showing the (b) phenotype, suggesting that the different phenotypes had evolved before establishment of the inbred lineages.

MUPs from four different inbred strains were analysed by ESI/MS (Figure 2). Three of the strains, Balb/cJ, A/J and DBA1/J, are phenotype (a) whereas the remaining strain, C57BL/6J, is phenotype (b). All four strains express MUPs of three major masses, 18643(4), 18691(2) and 18706(7) Da, whereas the C57BL/6J MUPs show a higher abundance of protein at a mass of 18893 Da. The variations in the last significant figures are within the expected precision of mass assignment. The relative abundance (integrated peak area) of the three major masses also varies between the two different phenotypes. In phenotype (a) the relative abundances of the three major masses are 18692 Da > 18643(4) Da > 18706(9) Da, whereas in phenotype (b) they are 18708 Da > 18691 Da > 18643 Da.

The anion-exchange profiles and the ESI/MS analysis are, superficially, paradoxical. In phenotype (a) the three major observed masses have to account for four large anion-exchange peaks. In phenotype (b) MUPs, the four observed masses are accounted for by four anion-exchange peaks, but only one of these peaks has the same retention time, and therefore the same inferred charge, as a peak in the phenotype (a) profile. To clarify these anomalies we have used ESI/MS to screen the anion exchange profiles of Balb/cJ and C57BL/6J MUPs, strains representative of the two phenotypes. Direct screening of anion-exchange fractions with ESI/MS proved unsuccessful (results not shown). Therefore anion-exchange chromatography and rechromatography were used to purify individual peaks, which were then concentrated and desalted by reverse-phase HPLC before ESI/MS (HPLC traces not shown) The results of these analyses are shown in Figure 3 for Balb/cJ and C57BL/6J MUPs. In most instances, a single mass peak was associated with a protein peak on Mono-Q chromatography, but one peak in each strain (Balb/c, $V_e \approx 11$ ml; C57BL/6J, $V_e \approx 132$ ml) incorporated two proteins, co-eluting on two cycles of ion-exchange and reverse-phase chromatography. Comparison of the combined ESI/MS data in Figure 3 with the direct ESI/MS analysis of unfractionated MUPs in Figure 2 shows the presence of two MUPs (18698 and 18614 Da) in the former that are absent in the latter. This apparent anomaly is thought to be due to two factors: the relatively low levels of these proteins and the proximity of their molecular masses to that of other far more abundant MUPs, which has prevented their resolution by ESI/MS and maximum-entropy processing.

A summary analysis, relating elution volume on Mono-Q anion-exchange chromatography to mass and relative abundance for the two phenotypes, is particularly informative (Figure 4). It is evident that the same masses are present in both strains but with largely different elution volumes (and hence net charges). Both strains express proteins of different masses that are co-eluted. There are at least 10 different MUPs, only one of which is apparently common (in charge and mass) to both strains, attesting to the complexity of this group of proteins. One particularly interesting feature is the presence of two MUPs with the same mass, yet different charges, in the same strain (uMUP-I and uMUP-III, mass 18693 Da in Balb/cJ).

Sequence divergence between some of the MUPs is slight [20]. It was therefore likely that the two proteins of same mass and different charges were derived from an isobaric substitution that

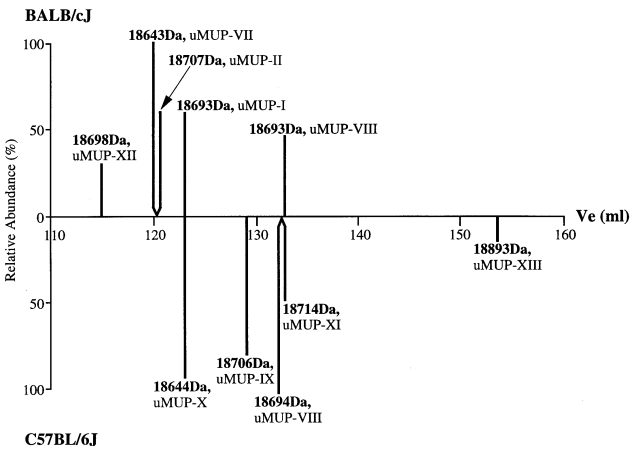

Figure 4 Summary comparison of Balb/cJ and C57BL/6 MUPs analysed by anion-exchange chromatography and ESI/MS

The central axis represents the anion-exchange elution volume (V_e) and the heights of the peaks represent their relative abundances. The annotated figure on each peak is the molecular mass as determined by ESI/MS. Finally, the peaks are labelled with the names suggested in Table 1 for ease of cross-reference.

caused a shift in overall charge. The only changes capable of this effect are Lys/Glx substitutions. Comparison of the aligned sequences indicated that uMUP-IX (our nomenclature, see Table 1) and uMUP-II differed by a single amino acid at position 136: in uMUP-II this is a Lys residue; in uMUP-IX it is Gln. The cDNA sequence of uMUP-II was derived from a Balb/c mouse (accession number M16356 [21]); that of uMUP-IX was derived from a C57BL/6J mouse (accession number X00908 [22]). This sequence difference would predict that the MUP containing Lys^{136} from Balb/c should be eluted earlier than the MUP containing Gln^{136} from C57BL/6J; this is exactly what is observed (Figure 4).

A similar phenomenon could be invoked to explain the charge difference between the two proteins of molecular mass 18693 Da in the Balb/c strain. In this instance, there were no sequence data to support our reasoning, and we analysed these two proteins further, The two 18693 Da MUPs from Balb/c mice were purified by chromatography/rechromatography and mapped by MALDI-TOF/MS. Because we expected an isobaric substitution involving a Lys residue, we reasoned that a Lys-C peptide map would be most informative. Figure 5 shows the mass spectra obtained. For the spectrum of the component eluting earlier (Figure 5, upper spectrum), all the observed proteolytic peptides could be assigned to the known sequence of a Balb/c MUP (Table 1, uMUP-I, our nomenclature); two predicted tripeptide fragments (29–31 and 74–76) were not observed. The MUP eluting later gave a map on MALDI-TOF/MS that was very similar to the first, with one important exception. The MUP eluting earlier has fragments of mass 1123 Da (residues 132–140) and 2535 Da (residues 141–162) that are absent from the spectrum of the MUP eluting later. Instead, the late-eluting MUP possesses a fragment of 3640 Da that is absent from the early-eluting MUP spectrum. This corresponds to a combination of the two

Table 1 A proposed classification and nomenclature for uMUPs

Abbreviations: cds, coding sequence; Glyc, glycosylation.

Name	Predicted/obs. mass/charge	Genbank accession no.	Genbank locus	References/notes	Balb/c Pred./Obs.	C57BL/6 Pred./Obs.
uMUP-I	18694 Da $z = -11.3$	M16355	MUSMUPI	[21]	Yes/Yes	–/No
uMUP-II	18709 Da $z = -11.3$	M16356	MUSMUPII	[21]	Yes/Yes	–/No
		X04115(?)	MMMUP8R	Incomplete cds [24]		
uMUP-III		M16359(?)	MUSMUP3B	Incomplete cds[21]		
uMUP-IV	18817 Da $z = -7.2$	M16358	MUSMUPIV	[21]	Yes/–	–/No
uMUP-V	18997 Da $z = -13.3$	M16360	MUSMUPV	[21]	Yes/No	–/No
uMUP-VI	19007 Da + Glyc $z = -13.7 \pm ?$	X03525	MMMUP15R	[24]	Yes/No?	–/–
		M27608	MUSMUPG	[24]		
		X00909(?)	MUSMUPC	Incomplete cds [22]		
		M16357(?)	MUSMUP3A	Incomplete cds [21]		
uMUP-VII	18646 Da $z = -11.3$	X00907	MUSMUPA	[22]	Yes/Yes	
uMUP-VIII	18695 Da $z = -13.3$	This study		This publication	Yes/Yes	–/Yes
		X03524(?)	MMMUP11R	Incomplete cds [24]		
		M27609(?)	MUSMUPF	Incomplete cds [24]		
uMUP-IX	18709 Da $z = -12.3$	X00908	MUSMUPB	[22]	–/No	Yes/Yes
uMUP-X	18644 Da $z \approx -11$	None	None	Observed protein, but no known cDNA sequence	–/–	–/Yes
uMUP-XI	18714 Da $z = -13.3$	None	None	Observed protein, but no known cDNA sequence	–/–	–/Yes
uMUP-XII	18698 Da $z \approx -12.3$	None	None	Observed protein, but no known cDNA sequence	–/Yes	–/–
uMUP-XIII	18893 Da $z \approx -15$	None	None	Observed protein, but no known cDNA sequence	–/Yes	–/Yes
uMUP-XIV	18739 Da $z = -13.1$	X03208	MMMUPBS8	[25]	Yes/No	–/No

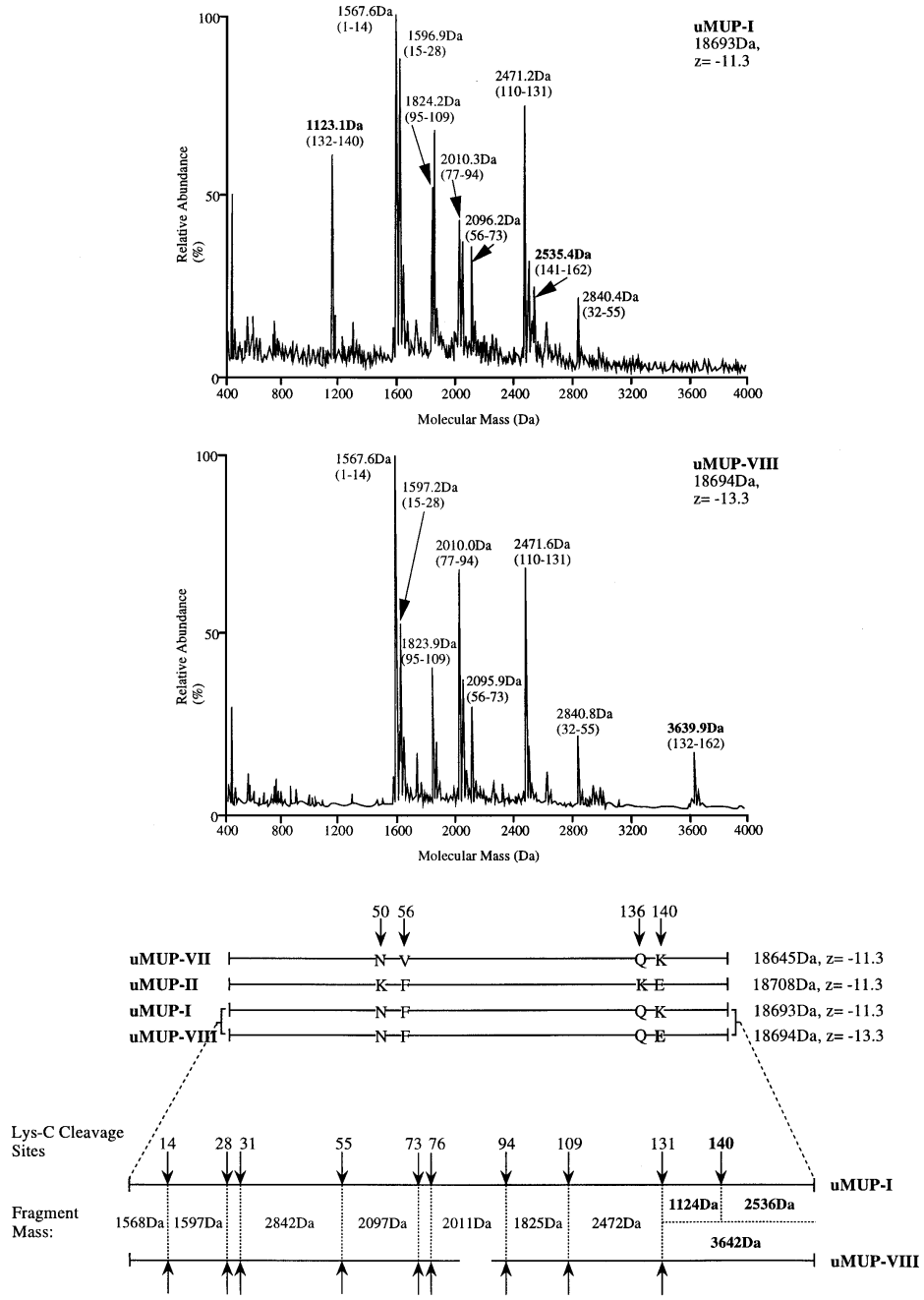

Figure 5 Peptide mapping of MUPs

uMUP-I and uMUP-VIII (both of molecular mass 18693 Da) were isolated by ion-exchange chromatography and were subjected to peptide mapping with endopeptidase Lys-C. The digests were analysed by MALDI-TOF/MS. The numbers in parentheses attached to each peak are the residues of the aligned sequence(s) that would give a fragment of that particular mass (bottom panel).

sequences, consistent with the elimination of a Lys-C site, and therefore comprises residues 132–162 (Figure 5, bottom panel). Loss of a Lys-C site with no detectable change in mass (at the precision of these ESI/MS and MALDI-TOF/MS analyses) can be achieved only by a Lys/Glx replacement. To confirm the substitution, the Lys-C peptide 132–162 was isolated by reverse-phase HPLC (results not shown) and sequenced by automated Edman degradation. The sequence so derived was E^{132}RFAQLX$EEHGI^{144}$, demonstrating clearly that this MUP is a new sequence that has Glu^{140} in place of Lys^{140}. The Lys → Glu change involves the gain of an acidic group as well as the loss of a basic group, and it is unsurprising that the new MUP is eluted considerably later in the salt gradient than the other protein.

DISCUSSION

MUPs have been used in studies of mouse genetics for some time [10,11], partly because of the accessibility of the protein product

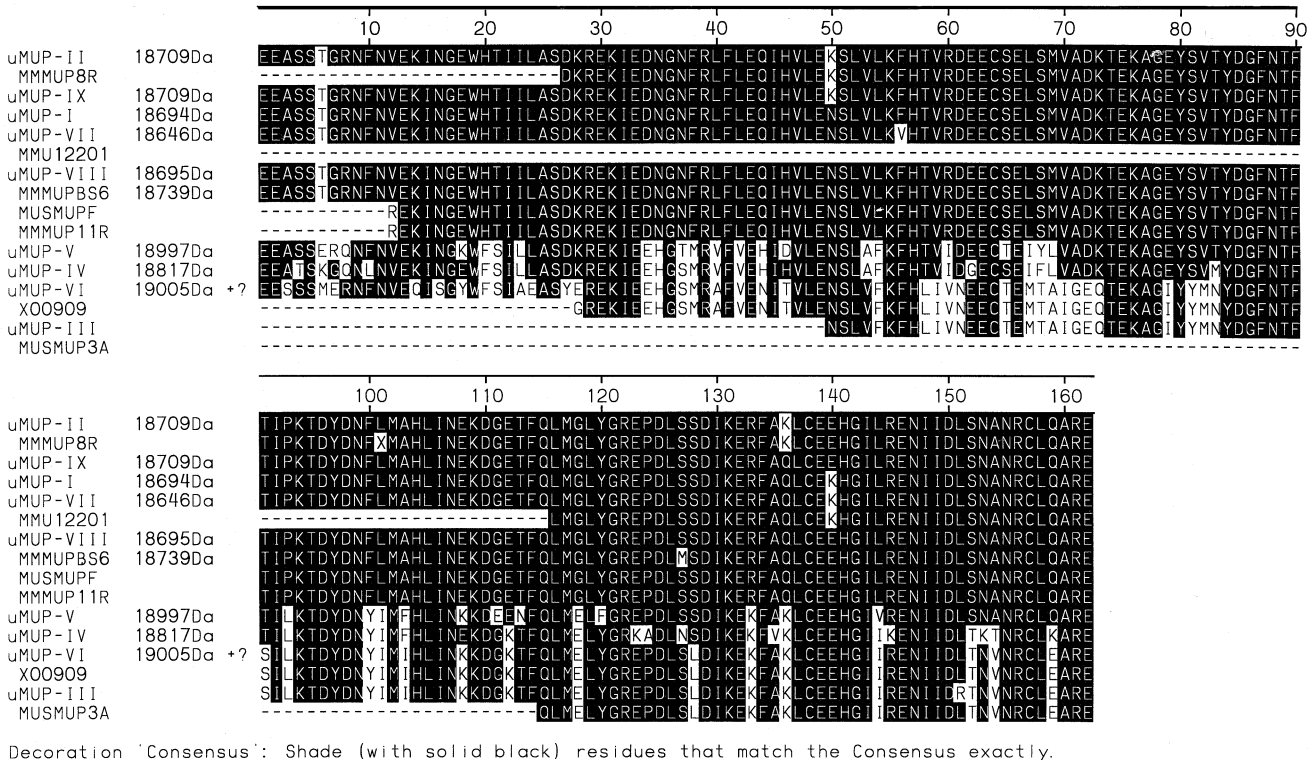

Figure 6 Sequence alignment of uMUPs

The inferred full or partial protein sequences of MUPs were aligned using the Clustal algorithm and a PAM250 similarity matrix. Residues that are in the majority in each position are highlighted on a black background. The sequences are named according to the data in Table 1. The complete sequence uMUP-II and the incomplete sequence MMMUP8R are identical except for an undetermined residue at position 101 in the latter. These are highly likely to be the same sequence. MMMUPBS6 differs from uMUP VIII at only one position (127), where the former possesses a methionine residue that is a serine in the latter. The predicted molecular mass of MMMUPBS6 is 18739 Da, whereas that of uMUP-VIII is 18695 Da. Only this latter mass is observed in Balb/c mice. The shortest aligned sequence is MMU12201 (residues 115–162), derived from C57BL/6 mice, which is identical with the corresponding residues in uMUP-I and uMUP-VII, both derived from Balb/c mice. The lack of proteins of the same mass and charge as uMUP-I and uMUP-VII in C57BL/6 mice leads to the conclusion that the full sequence of the gene from which MMU12201 is derived is not that of uMUP-I or uMUP-VII.

and partly because of the complexity of the phenotypes [12]. One outcome of this interest has been the derivation of complete or incomplete cDNA sequences for these proteins. In principle the accurate mass that can be derived from ESI/MS allows the correlation of the expressed proteins with the masses of proteins inferred from cDNA sequences. However, the cDNAs have been derived from two strains (Balb/c and C57BL/6), and several of the sequences are incomplete. We have therefore collated and aligned the inferred protein sequences (Figure 6), and propose a nomenclature for urinary MUPs only that encompasses the detailed data on expression and the inferred sequence data (Table 1). It is clear that multiple entries in the sequence databases refer to the same protein, as indicated by multiple sequence entries for a single MUP. Nearly all of the sequences have been derived from the Balb/c mouse, yet only four have been positively identified as urinary MUPs in this animal. A further two or three inferred protein sequences yield masses that have not been observed by ESI/MS. Thus high-level expression of these sequences remains questionable. All ESI/MS analyses were conducted on freshly expressed urine, and there is no possibility of rapid preferential degradation of specific proteins, despite the existence of proteolytic enzymes in mouse urine [23]. One of the MUPs (uMUP-VI, also known as MUP15) has been identified as a minor protein, and has been demonstrated to be glycosylated [25]. This sequence includes an NIT motif for N-linked glycosylation. The mass of the core sequence of uMUP-VI is 19005 Da, and the glycosylation would add further mass to this core.

Several cDNA sequences, first given Genbank accession numbers X00907, X00908, X00909 and J00607 [22], seem subsequently to have been cross-referenced into a full gene sequence, Genbank locus MMMUPBS6, accession number X03208. However, when the coding sequence is reconstructed from the exon data for this gene sequence, the open reading frame predicts a protein of the correct length (162 amino acids) and a mass (18739 Da) that is not observed in urine. Closer inspection of the sequences indicates that X00907 and X00909 were derived from a Balb/c mouse, and X00908 was derived from a C57BL/6J mouse. The MMMUPBS6 sequence, from Balb/c, reports the four variant amino acid positions as $N^{50} \ldots F^{56} \ldots Q^{136} \ldots E^{140}$, which in X00908 (C57BL/6) is $K^{50} \ldots F^{56} \ldots Q^{136} \ldots E^{140}$ and in X00907 (Balb/c) is $N^{50} \ldots V^{56} \ldots Q^{136} \ldots K^{140}$ (Table 1). Furthermore MMMUPBS6 is a unique sequence by virtue of the substitution of Ser^{127} for Met^{127}. We have positively identified both of the Balb/c protein products implied by these sequences from both mass and charge data. The lack of a protein of molecular mass 18739 Da, as predicted by the MMMUPBS6 sequence, is not entirely unexpected, as previous work has shown that MUP genes can be transcriptionally silent [26]. So far at least 14 uMUPs can be identified by a combination of DNA sequences and analysis of urinary secreted proteins. Several of

these MUPs are generated by variation in just four positions: $N/K^{50} \ldots V/F^{56} \ldots Q/K^{136} \ldots K/E^{140}$.

The current data pertaining to a role for MUPs suggest that they are involved in olfactory communication by virtue of the fact that they bind mouse pheromones [13,14,16]. On the basis of the results presented here we propose that the differences shown between the two inbred strains Balb/cJ and C57BL/6J represent a polymorphism present in wild mice brought about by the co-dominant expression of multiple alleles. The similarities seen between many of the strains reflect the fact that they have been bred from common ancestors. To a nocturnal, burrowing animal such as the mouse, the possession of a unique set of pheromone binding proteins by an individual, reflecting its genetic origin, could mediate an olfactory signature useful in a number of behavioural contexts including kin recognition, orientation and territorial demarcation. Such a hypothesis presupposes that these small substitutions of amino acids alter the molecules' ability to bind ligands differentially. We have already presented some preliminary evidence for specificity of ligand binding [13], but this area would benefit from a more extensive study.

We thank VG Organic (Manchester, U.K.) for access to the MALDI-TOF mass spectrometer and Dr. Brian Green of VG Organic for advice concerning electrospray mass spectrometry. This work was funded by the BBSRC.

REFERENCES

1 Novotny, M., Harvey, S., Jemilo, B. and Alberts, J. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 2059–2061
2 Drickamer, L. C. (1986) Dev. Psychobiol. **19**, 155–161
3 Barkley, M., DelgadoDeLeon, D. and Weste, R. (1993) J. Exp. Zool. **265**, 558–566
4 Williams, J. R. and Lenington, S. (1993) Behav. Genet. **23**, 51–58
5 Vandenbergh, J. G., Whitsett, J. M. and Lombardi, J. R. (1975) J. Reprod. Fert. **43**, 515–523
6 Duncan, R., Matthai, R., Huppi, K., Roderick, T. and Potter, M. (1988) Mol. Cell. Biol. **8**, 2705–2712
7 Bishop, J. O., Clark, A. J., Clissold, P. M., Hainey, S. and Francke, U. (1982) EMBO J. **1**, 615–620
8 Wicks, L. F. (1941) Proc. Soc. Exp. Biol. Med. **48**, 395–400
9 Al-Shawi, R., Wallace, H., Harrison, S., Jones, C., Johnson, D. and Bishop, J. O. (1992) Mol. Endocrinol. **6**, 181–190
10 Finlayson, J. S., Potter, M. and Runner, C. R. (1963) J. Natl. Cancer Inst. **31**, 91–107
11 Clissold, P. M., Hainey, S. and Bishop, J. O. (1983) Biochem. Genet. **22**, 379–387
12 Hainey, S. and Bishop, J. O. (1982) Genet. Res. **39**, 31–39
13 Robertson, D. H. L., Beynon, R. J. and Evershed, R. P. (1993) J. Chem. Ecol. **19**, 1405–1416
14 Bacchini, A., Gaetani, E. and Cavaggioni, A. (1992) Experientia **48**, 419–421
15 Bocskei, Z., Findlay, J. B. C., North, A. C. T., Phillips, S. E. V., Somers, W. S., Wright, C. E., Lionetti, C., Tirindelli, R. and Cavaggioni, A. (1991) J. Mol. Biol. **218**, 699–701
16 Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavagioni, A., Findlay, J. B. C. and North, A. C. T. (1992) Nature (London) **360**, 186–188
17 Adams, P. and Sawyer, L. (1990) Biochem. Soc. Trans. **18**, 936–937
18 Cavaggioni, A., Sorbi, R. T., Keen, J. N., Pappin, D. J. C. and Findlay, J. B. C. (1987) FEBS Lett. **212**, 225–228
19 Festing, M. F. W. (1987) International Index of Laboratory Animals, Whitaker, London
20 Evershed, R. P., Robertson, D. H. L., Beynon, R. J. and Green, B. N. (1993) Rapid Commun. Mass Spec. **7**, 882–886
21 Shahan, K., Gilmartin, M. and Derman, E. (1987) Mol. Cell. Biol. **7**, 1938–1946
22 Khun, N. J., Woodworth-Gutai, M., Gross, K. W. and Held, W. A. (1984) Nucleic Acids Res. **12**, 6073–6090
23 Beynon, R. J., Oliver, S. and Robertson, D. H. L. (1995) Biochem. J. **315**, 461–466
24 Clark, A. J., Chave-Cox, A., Ma, X. and Bishop, J. O. (1985) EMBO J. **4**, 3167–3171
25 Clark, A. J., Ghazal, P., Bingham, R. W., Barrett, D. and Bishop, J. O. (1985) EMBO J. **4**, 3159–3165
26 Shi, Y., Son, H. J., Shahan, K., Rodriguez, M., Constantini, F. and Derman, E. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 4584–4588

Received 15 November 1995/23 January 1996; accepted 29 January 1996