Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

addenda and errata

The X-ray structure of a recombinant major urinary protein at 1.75 Å resolution. A comparative study of X-ray and NMR-derived structures. Addendum

Paula R. Kuser,[a] Lorella Franzoni,[b] Elena Ferrari,[b] Alberto Spisni[b] and Igor Polikarpov[a]

[a]Laboratório Nacional de Luz Síncrotron, Caixa Postal 6192, CEP 13084-971 Campinas, SP, Brazil, and [b]Department of Experimental Medicine, Section of Chemistry and Structural Biochemistry, University of Parma, Via Volturno 39, 43100 Parma, Italy

In reference to the paper by Kuser *et al.* [*Acta Cryst.* (2001), D**57**, 1863–1869], the authors Igor Polikarpov and Alberto Spisni were jointly responsible for this paper. Professor Polikarpov can be contacted at ipolikarpov@if.sc.usp.br and Professor Spisni can be contacted at aspin@ipruniv.cce.unipr.it.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The X-ray structure of a recombinant major urinary protein at 1.75 Å resolution. A comparative study of X-ray and NMR-derived structures

Paula R. Kuser,[a] Lorella Franzoni,[b] Elena Ferrari,[b] Alberto Spisni[b]* and Igor Polikarpov[a]†

[a]Laboratório Nacional de Luz Síncrotron, Caixa Postal 6192, CEP 13084-971 Campinas, SP, Brazil, and [b]Department of Experimental Medicine, Section of Chemistry and Structural Biochemistry, University of Parma, Via Volturno 39, 43100 Parma, Italy

† Present address: USP-Instituto de Física de São Carlos, Departamento de Física e Informática, Av. Dr Carlos Botelho 1465, 13560-250 São Carlos, SP, Brazil.

Correspondence e-mail: aspin@unipr.it

Received 6 June 2001
Accepted 29 October 2001

PDB Reference: mouse major urinary protein, 1jv4.

Major urinary proteins belong to the lipocalin family and are present in the urine of rodents as an ensemble of isoforms with pheromonal activity. The crystal structure of a recombinant mouse MUP (rMUP) was solved by the molecular-replacement technique and refined to an R factor and $R_{\rm free}$ of 20 and 26.5%, respectively, at 1.75 Å resolution. The structure was compared with an NMR model and with a crystallographic structure of the wild-type form of the protein. The crystal structures determined in different space groups present significantly smaller conformational differences amongst themselves than in comparison with NMR models. Some, but not all, of the conformational differences between the crystal and solution structures can be explained by the influence of crystallographic contacts. Most of the differences between the NMR and X-ray structures were found in the N-terminus and loop regions. A number of side chains lining the hydrophobic pocket of the molecule are more tightly packed in the NMR structure than in the crystallographic model. Surprisingly, clear and continuous electron density for a ligand was observed inside the hydrophobic pocket of this recombinant protein. Conformation of the ligand modelled inside the density is coherent with the results of recent NMR experiments.

1. Introduction

It is now widely accepted that the pace of sexual maturation of female rodents is dependent on chemical signalling. Indeed, male rats and mice synthesize and excrete a large amount of highly homologous proteins namely α_2-globulin (α_{2u}) in rats and major urinary protein (MUP) in mice (Dinh *et al.*, 1965; Finlayson *et al.*, 1986). Both belong to a large group of structurally homologous proteins named calycins (Flower *et al.*, 1993) owing to their basket-shaped structure made up of antiparallel β-sheets.

The calycin protein superfamily subdivides itself into three families: the fatty acid-binding proteins (FABPs), the avidins and the lipocalins (Flower *et al.*, 1993). In spite of their poor sequence similarity, all these proteins are structurally characterized by an antiparallel β-barrel with +1 topology formed by ten β-strands in the case of FABPs and eight β-strands in the case of the lipocalins. They are all involved in the binding and transport of small hydrophobic molecules.

Lipocalins share the same N-terminal motif consisting of a 3_{10}-helix that leads into the first β-strand. They are specifically involved in the transport of retinoids, steroids and pheromones, in olfaction, in immune-system regulation and in the mediation of cell homeostasis.

MUPs are members of the lipocalin family. They form an ensemble of closely related proteins that, being expressed in the liver under androgenic control (Knopf *et al.*, 1983; Kuhn *et al.*, 1984), are secreted into the serum, filtered by the kidney and excreted in the urine. Not much attention was given to these proteins until a significant sequence similarity with the pyrazine-binding proteins from calf nasal mucosa was found and their possible pheromonal function was identified (Cavaggioni *et al.*, 1987, 1990; Mucignat-Carretta *et al.*, 1995).

Several volatile pheromones, in particular 2-*sec*-butyl-4,5-dihydrothiazole, were shown to bind to MUPs, being subsequently slowly released into the air as the urine dries out. These volatile components, trafficked by MUPs, are able to stimulate the sexual maturation of mice (Novotny *et al.*, 1999). However, whether MUPs are mere pheromone carriers or have some pheromonal activity of their own is still an open question.

It was shown that MUPs, acting *via* the vomeronasal organ (VNO), can accelerate the onset of puberty in female mice, thus raising the hypothesis of the existence of MUP receptors in the VNO (Bocskei *et al.*, 1992). This conjecture was supported by the identification of genes encoding for putative pheromone receptors (Dulac & Axel, 1995; Ryba & Tirindelli, 1997). Furthermore, it was recently shown that proteins from rat urine were able to activate a specific type of G protein in the VNO (Krieger *et al.*, 1999). Three distinct families of pheromone receptors were identified in the VNO (Ryba & Tirindelli, 1997; Krieger *et al.*, 1999; Martini *et al.*, 2001), namely the V1R and V3R families specific for small hydrophobic molecules and the V2R family activated by pheromones of proteic nature.

Since the physiological cycle of MUPs is quite complex, understanding the structure–function relationship of these proteins requires a precise knowledge of their structure and dynamics. The crystal structure of the wild-type MUP (wtMUP) was solved and refined to 2.4 Å resolution in the tetragonal crystal form (Bocskei *et al.*, 1992). Recently, the solution structure of a recombinant MUP (rMUP) was determined (Lücke *et al.*, 1999). Although the overall fold of the protein in solution is similar to that reported in the crystal structure (Bocskei *et al.*, 1992), a number of local differences in the main-chain conformation were observed (Lücke *et al.*, 1999). This does not seem particularly surprising, as the X-ray structure was derived from the crystal obtained from a mixture of isoforms carrying a heterogeneous ligand population, while the NMR-derived structure is obtained from a single recombinant isoform.

In order to verify whether these differences are indeed representative of the crystal and solution state, the structure of rMUP crystallized in a monoclinic space group was solved and refined to 1.75 Å resolution; a comparison between the existing structures of this protein is presented.

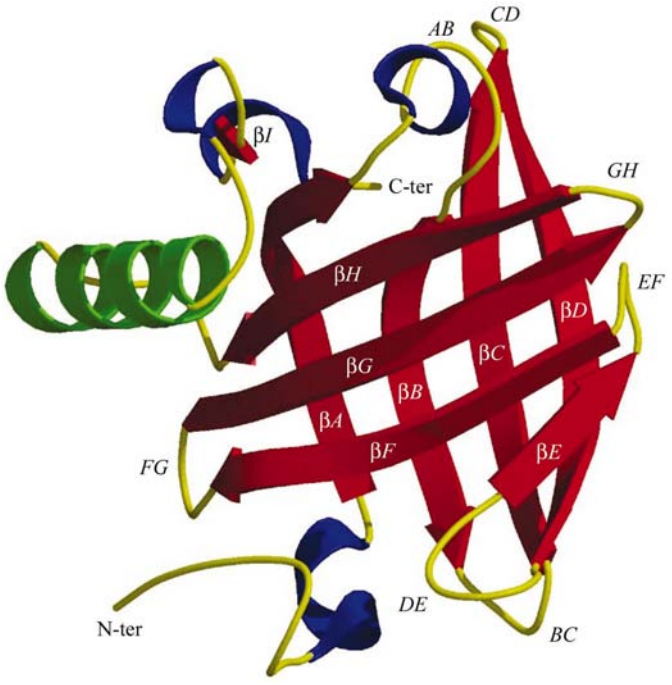

Figure 1
Overall fold of the rMUP polypeptide chain. The calyx-shaped β-barrel is formed by eight β-strands (red). The α-helix is coloured green and four short 3_{10}-helices are coloured blue. One further β-strand occurs close to the C-terminal. The hydrophobic pocket is located inside the barrel. The diagram was drawn using *MOLSCRIPT* (Kraulis, 1991).

2. Materials and methods

2.1. Protein expression

The recombinant mouse major urinary protein (rMUP) was expressed in the methylotropic yeast *Pichia pastoris* and purified as described previously (Ferrari *et al.*, 1997). The sequence of this recombinant MUP is deposited in the EMBLNEW databank with accession code MMU309921.

2.2. Crystallization

The crystals used for data collection were grown by hanging-drop vapour diffusion at 291 K in sodium phosphate buffer at pH near 6.5. 60 m*M* $CdCl_2$ was used as the precipitant. Drops consisted of equal volumes of protein at 10 mg ml^{-1} concentration and well solution (well volume = 1.0 ml). Crystals of dimensions 0.4 × 0.15 × 0.1 mm grew in four weeks. Diffraction data were collected on a MAR Research 345 mm imaging plate at the Protein Crystallographic (PCr) beamline (Polikarpov *et al.*, 1998) at the Laboratório Nacional de Luz Síncrotron (LNLS), Campinas, Brazil. The initial diffraction resolution limit was 1.4 Å, but gradually decreased during data collection.

2.3. Data processing

The programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) were used for data processing. A data set consisting of 80 oscillation photographs with an oscillation range of 1.0° was collected from a single unfrozen crystal to 1.75 Å resolution. The overall $R_{\rm merge}$ was 7.9% and $R_{\rm merge}$ for

the highest shell was 23.3%. The crystal belongs to the monoclinic space group $P2_1$, with unit-cell parameters $a = 37.14$, $b = 55.79$, $c = 37.67$ Å, $\beta = 93.24^\circ$ (Kuser *et al.*, 1999). The asymmetric unit cell contains one rMUP molecule and the solvent content is about 48%.

A molecular-replacement solution was found with the program *AMoRe* (Collaborative Computational Project, Number 4, 1994; Navaza, 1994) using the previously reported crystal structure of wtMUP determined at 2.4 Å resolution as a search model (PDB code 1mup; Bocskei *et al.*, 1992). The top solution for the rotation function has a correlation coefficient of 32.5%, whereas the next highest peak was below 17%. A translational position was found which yielded an *R* factor of 44% and a correlation coefficient of 68.7%. The packing arrangement of the molecules in the unit cell for this solution yielded no unfavourable intermolecular contacts.

2.4. Model building and refinement

Positional and temperature-factor refinement were carried out using *REFMAC* (Collaborative Computational Project, Number 4, 1994) against all measured reflections to a resolution of 1.75 Å. The refinement calculation was interleaved with several rounds of model building with the program *O* (Jones *et al.*, 1991). Water molecules were added using the program *ARP* (Collaborative Computational Project, Number 4, 1994). The final *R* factor was 20.6% and $R_{\rm free}$ was 26.5%. There was no density for residue Glu1 and the density for the first nine N-terminal residues was poorly defined. In addition, the side chain of Asn153 was disordered and atoms beyond C^β were not modelled. Ile32 was modelled in two conformations. Seven well defined Cd atoms were identified. Their presence seems to be important for protein crystallization. The refinement statistics are summarized in Table 1.

Table 1
Refinement statistics.

Values in parentheses refer to the last resolution shell.

Data collection	
Resolution limit of data (Å)	1.75
Completeness of data (%)	86.4 (89.6)
$R_{\rm symm}$† (%)	7.5 (29.9)
Refinement	
Resolution range used for refinement (Å)	14.0–1.75
Total No. of protein atoms	1262
Total No. of solvent molecules	235
R factor‡	20.6
$R_{\rm free}$§	26.5
Average *B* factors	
All protein atoms (Å^2)	26.5
Solvents (Å^2)	38.4
R.m.s.d.s from ideal geometry	
Bonds (Å)	0.028
Bond angles (°)	2.94

† $R_{\rm symm} = \sum |I - \langle I\rangle| / \sum I$. ‡ Crystallographic *R* factor $= \sum ||F_{\rm obs} - F_{\rm calc}|| / \sum |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. § $R_{\rm free}$ is the crystallographic *R* factor calculated for a subset of randomly selected reflections (5%) not used in the phasing process.

Table 2
Comparison of intermolecular contacts of rMUP and wtMUP with symmetry-related molecules.

Residue/atom	Residue/atom SYM	Distance (Å)
rMUP		
Glu2 $O^{\varepsilon1}$	Arg60 NH1	2.82
Glu2 $O^{\varepsilon2}$	Arg60 NH1	2.54
Glu18 N*	Glu132 $O^{\varepsilon2}$	3.10
Arg29 NH2	Glu66 $O^{\varepsilon2}$	2.43
wtMUP		
Glu22 N*	Glu136 $O^{\varepsilon2}$	3.03
Glu22 $O^{\varepsilon1}$	Gln140 $N^{\varepsilon2}$	3.20
His50 $N^{\delta1}$	Asp133 $O^{\delta1}$	3.06
Asn54 $N^{\delta2}$	Asp133 $O^{\delta2}$	3.23
Lys59 N^{ζ}	Asp76 $O^{\delta1}$	2.51
Lys59 N^{ζ}	Asp76 $O^{\delta2}$	2.84
Glu66 $O^{\varepsilon2}$	Gln119 $N^{\varepsilon2}$	3.13

3. Results

3.1. Overall three-dimensional structure

The crystal structure of rMUP is dominated by eight β-strands forming a calyx-shaped β-barrel that encloses an internal ligand-binding site (Fig. 1). The Ramachandran plot of rMUP shows that essentially all residues are in the allowed regions, with the exception of Tyr97 ($\varphi = 68.1$, $\psi = -47.2^\circ$). Tyr97 is present in a γ-turn that is conserved in all known lipocalin structures (Brownlow *et al.*, 1997). In addition, two residues of the γ-turn are part of the most highly conserved triplet of residues in the lipocalin family, the sequence motif Thr-Asp-Tyr. In rMUP this turn is maintained by a hydrogen bond between the main-chain O atom of Tyr97 and the side chain of Arg122, a well defined and structurally conserved residue.

3.2. Comparisons with other structures

The crystal structure of rMUP was compared with the NMR-derived solution model and with the wtMUP crystal structure. The structural differences were evaluated using the molecular-graphics package *O* (Jones *et al.*, 1991) and were quantified using the root-mean-square deviation (r.m.s.d.) for corresponding sets of atoms calculated with *LSQKAB* (Collaborative Computational Project, Number 4, 1994). The crystallographic *B* factors and intermolecular contacts in the crystal lattice were also examined. In order to facilitate this comparison, the protein sequence numbering of wtMUP used by Bocskei *et al.* (1992) was modified in such a way that the first residue became residue 1 instead of residue 5.

3.2.1. Comparison of the crystal structures of rMUP and wtMUP. A previously determined structure of wtMUP purified from mice urine, was solved in the space group $P4_32_12$ and refined to 2.4 Å resolution (Bocskei *et al.*, 1992). The crystal structure of rMUP presented here was determined in the monoclinic space group $P2_1$. It was refined to 1.75 Å resolution. The overall fold of both structures is very similar, with an almost perfect overlap of the backbone atoms. Fig. 2(*a*) reports the r.m.s.d. for each residue obtained by superposition of the crystallographic structures of wtMUP and rMUP. The similarity between the two crystal structures is reflected by r.m.s.d. values of 0.44 and 0.54 Å between corresponding C^α and backbone atom positions, respectively. The differences

mainly arise from residues located at the N- and C-termini of the protein and in the regions containing residues 59–64 and 109–111. Some of the differences can be explained by crystal contacts (Table 2). The r.m.s.d. for all atoms of the two structures is 0.99 Å.

In Fig. 2(*a*) the residues involved in crystal contacts are highlighted with an arrow or a diamond depending whether the contacts are observed in the rMUP or wtMUP crystal structure. The side chain of Arg156 is the one that diverges the most between the two structures. This is probably caused by the interaction of its guanidinium group with the carbonyl group of Ser152 which is only present in the wtMUP model. The residues that make similar contacts in both structures, such as Glu18 and Glu132, have very low r.m.s.d.s both for the side-chain and the main-chain atoms.

3.2.2. Comparison of crystalline and solution structures of rMUP. The crystal structure of rMUP was compared with the ten conformers of the NMR-derived solution structure of rMUP deposited in the PDB with the accession code 1df3 (Lücke *et al.*, 1999).

The secondary-structure elements assigned to the structures of rMUP in the crystalline state and in solution are in good agreement. The protein scaffold formed by the β-barrel and by the helices is preserved. However, small differences were observed. While the two stretches 11–16 and 28–32 appear as 3_{10}-helices (helix $3_{10}a$ and $3_{10}b$) in the crystallographic models of rMUP and wtMUP, in the rMUP NMR-derived structures both are defined as α-helices. Furthermore, the 3_{10}-helix *d* and the β-strand I are not present in the solution structure.

The superposition between the crystal and the solution rMUP structures was performed for each of the ten NMR conformers individually. The r.m.s.d. values for the C^{α} atoms were calculated as an average of the r.m.s.d. values for each conformer, resulting in a final r.m.s.d. of 1.26 Å. The r.m.s.d.s for the main chain and all atoms are 1.35 and 2.10 Å, respectively.

No major differences are observed in the position of the residues forming the binding site, namely Leu40, Leu42, Phe90, Ala103 and Tyr120 (Zídek *et al.*, 1999), providing further evidence of the rigidity of this portion of the protein. Significant differences are observed in the side-chain orientations of residues Leu54, Phe56, Met69, Leu105 and Leu116 located inside the hydrophobic cavity but not involved in ligand binding. Relative positions of their side chains between the crystalline and the solution rMUP structures differ by 1.3 Å. In fact, many of these residues are more tightly packed and closer to the centre of the hydrophobic cavity in the NMR-derived structure compared with the corresponding residues in the crystal structure.

A number of significant differences in side-chain and main-chain conformation between the X-ray and solution structures are observed for residues located in the loop regions and the N- and C-termini (Fig. 2*b*). Other side chains with large r.m.s.d. values are located at the protein surface and contain polar or charged groups, namely Arg29, Arg39, Glu43, His46, Glu66, Ser68, Lys73, Lys94, Asn99, Glu132, Arg133, Cys138, Asn147 and Asp150, with the only exception being Ile32.

Large differences in main-chain and side-chain conformations are observed for residues 59–64 between β-strands *C* and *D*. The distance between the C^{α} atoms of the crystal and the NMR conformers is of approximately 3 Å. In the NMR structure the *CD* loop is in a slightly more open conformation. It is worthwhile to point out, however, that the quality of the NMR model in this region is very poor, with an r.m.s.d. over 2 Å between individual conformers, indicating high flexibility of this region.

Several significant differences in side-chain positions detected between the X-ray and NMR-derived structures are

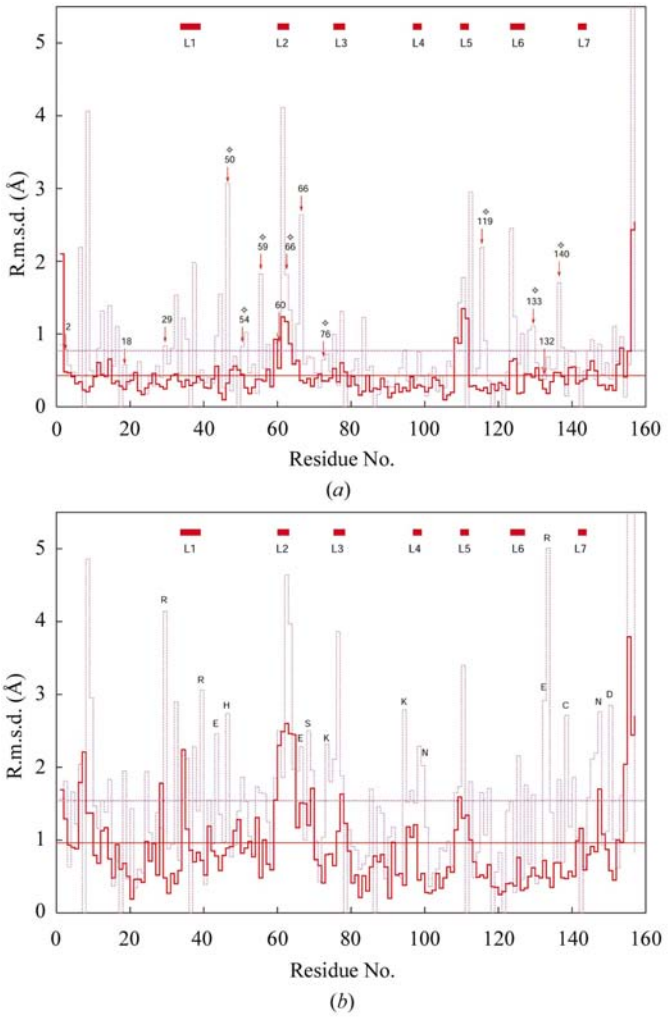

Figure 2
(*a*) Plot of the r.m.s.d.s between the main-chain and side-chain atoms of the wild-type MUP (PDB code 1mup) and rMUP (PDB code 1jv4) structures. The full and dashed lines represent main-chain and side-chain residues, respectively. The residues identified by arrows make crystal contacts in the rMUP structure. Residues marked with diamonds form crystal contacts in the wild-type MUP structure. The largest r.m.s. differences are for to the crystal contacts, the N- and C-terminal residues or residues belonging to the loop regions. (*b*) Plot of the r.m.s.d.s between the main-chain and side-chain atoms of rMUP in the crystal and in the solution averaged structure (PDB code 1df3; the plot represents the positions of an averaged structure based on the ten conformers of the NMR-derived structures). The top rectangles indicate the loop regions. Residues outside the loops regions and with large r.m.s.d.s are polar or charged residues located on the surface of the protein.

induced by crystallographic contacts. Residues Glu132 and Arg145, located at the surface of the molecule, provide an example of this. In the X-ray structure these residues form hydrogen bonds between their respective side chains (Arg145 NH1 and Glu132 OE1; Arg145 NH2 and Glu132 OE2). In the NMR structure these interactions do not exist. Residue Glu132 is pointing toward bulk solvent and Arg145 assumes different positions in each conformer.

In addition to the interactions mentioned, there are a number of hydrogen bonds and salt bridges that are present only in the X-ray structure. For example, His46, Glu43, Lys109, Glu112, Arg133 and Asp150 interact respectively with Gln44, His57, Asp110, Lys28, Glu2 and Lys131 in the crystal, whereas in the solution structure these interactions do not occur.

3.3. Conservative residues in lipocalins

An alignment of the rMUP and wtMUP amino-acid sequences with the primary structure of eight proteins from the lipocalin family is shown in Fig. 3. A number of conserved residues can be identified: Gly17, Trp19, Cys64, Thr95, Asp96, Tyr97, Arg122, Phe134 and Cys157; however, none of them belong to the ligand-binding pocket.

The conservation of these residues can be associated with specific structural features. Gly17 is responsible for changing the direction of the chain to go from 3_{10}-helix *a* to β-strand *A*. Trp19, which forms the bottom of the hydrophobic cavity and is involved in the definition of the binding site, has been reported to be important to stabilize the ligand inside the cavity and to protect it from the solvent (Katakura *et al.*, 1994). Cys64 makes a disulfide bridge with Cys157 that fastens the residues of the C-terminal to the rest of the molecule. Thr95, Asp96 and Tyr97 are the residues forming the conserved γ-turn.

The side chain of Arg122 makes hydrogen bonds with the carbonyl O atom of both Asn16 and Glu18, which seems to be important for the positioning of the eighth β-strand. Finally, Phe134 is surrounded by the hydrophobic residues Ile23, Met102, Met117, Leu119, Leu137 and Ile148 and is part of a hydrophobic patch located between the α-helix and the β-barrel.

As for the binding site, the fact that the amino acids in this region are poorly conserved can explain why these homologous proteins bind a variety of ligands with different affinity (Novotny *et al.*, 1985; Cavaggioni *et al.*, 1990; Bacchini *et al.*, 1992). However, they all possess a hydrophobic character.

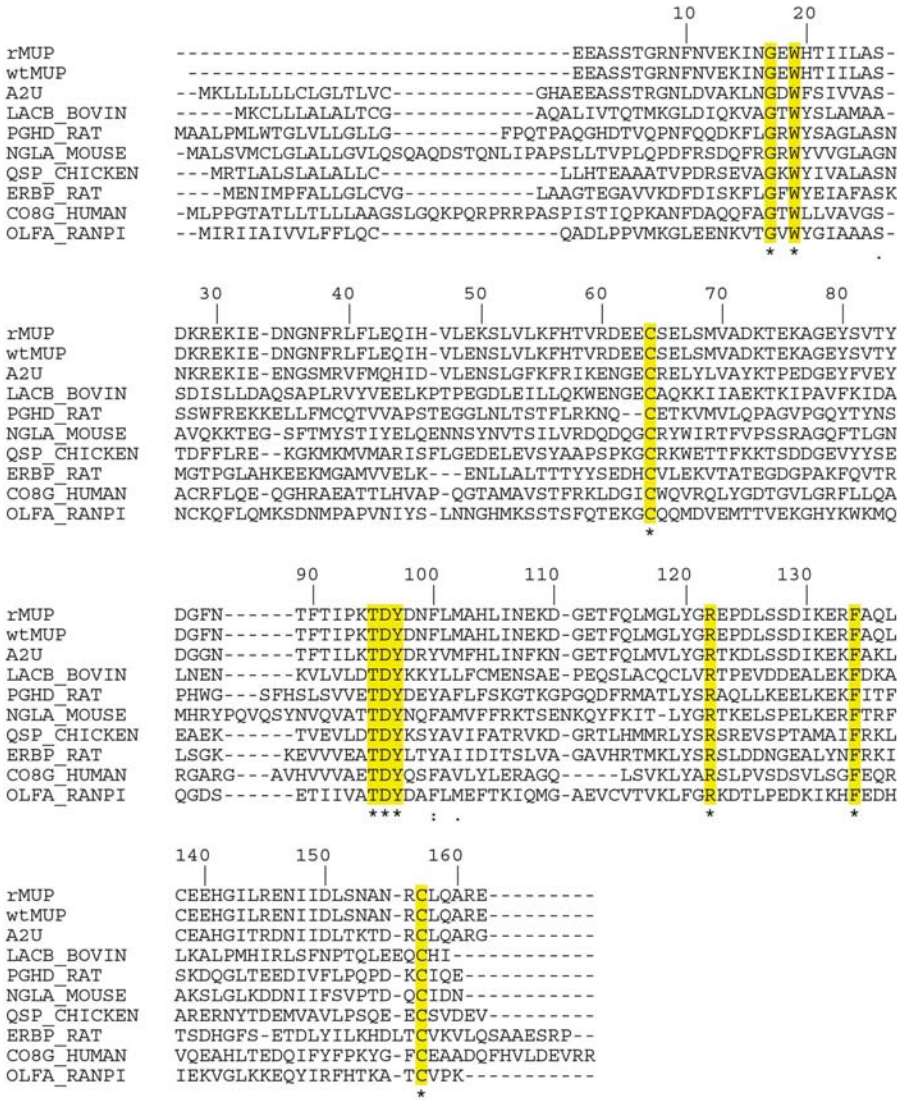

Figure 3
A sequence alignment of ten representative members of the lipocalin protein family. The few residues that are strictly conserved in all sequences are highlighted. This alignment was performed with the program *ClustalW*. The sequences listed are rMUP, wtMUP, α-2u-globulin (A2U), bovine β-lactoglobulin (LACB_BOVIN), rat prostaglandin-h2 D-isomerase precursor (PGHD_RAT), mouse neutrophil gelatinase-associated lipocalin (NGLA_MOUSE), chicken quiescence specific protein precursor (QSP_CHICKEN), rat epidymal retinoic acid binding protein (ERBP_RAT), human complement protein c8 γ (CO8G_HUMAN) and olfactory protein (OLFA_RANPI).

3.4. The hydrophobic pocket

A clear and continuous electron density is visible in the binding pocket of the crystal structure of rMUP. This is surprising, as no chemical substances of a size and shape which could fit the observed density was used in the expression and purification of rMUP. Therefore, such a compound must have been trapped and kept inside the calyx throughout the crystallization step. Unknown hydrophobic ligands trapped in the ligand-binding pockets have

previously been observed for several lipocalins (Tegoni *et al.*, 1996; Bianchet *et al.*, 1996; Folli *et al.*, 2001; Vincent *et al.*, 2001), including wtMUP (Bocskei *et al.*, 1992). In the wtMUP structure the hydrophobic ligand was modelled as 2-*sec*-butyl-4,5-dihydrothiazole (TZL), a natural pheromone isolated from male urine that enhances inter-male aggression (Novotny *et al.*, 1985) and induces oestrus synchronization in females (Jemiolo *et al.*, 1986).

In the rMUP molecule studied here, the shape of electron density observed in the $F_{obs} - F_{calc}$ map corresponds to a ring-shaped molecule (Fig. 4) reminiscent of TZL. Therefore, TZL was modelled in the hydrophobic pocket in order to compare its position and orientation to other structures. The exact chemical composition of the compound could not be judged from the electron density at 1.75 Å resolution. However, the $F_o - F_c$ and $2F_o - F_c$ electron-density maps for the ligand are well described in shape and size by a TZL molecule.

$C^{\varepsilon 2}$ and C^{γ} of Phe90 are the atoms closest to the electron density found in the binding pocket. The distance between them and the C9 atom of the modelled TZL is 3.5 Å, while residues Leu42, Leu105 and Tyr120 are at distances between 3.5 and 4.0 Å. The distances between the ligand atoms and the atoms of the protein are listed in Table 3.

Table 3
Distances between the modelled TZL and the protein residues.

Ligand atom	Protein atom	Distance (Å)
S	Leu105 $C^{\delta 1}$	4
C4	Leu105 $C^{\delta 1}$	3.9
C5	Leu105 $C^{\delta 1}$	3.9
C7	Tyr120 OH	3.7
C8	Leu42 $C^{\delta 2}$	3.9
C9	Phe90 C^{γ}	3.5
C9	Phe90 $C^{\varepsilon 2}$	3.5
C9	Phe90 C^{ζ}	3.8
C9	Phe90 $C^{\varepsilon 1}$	4

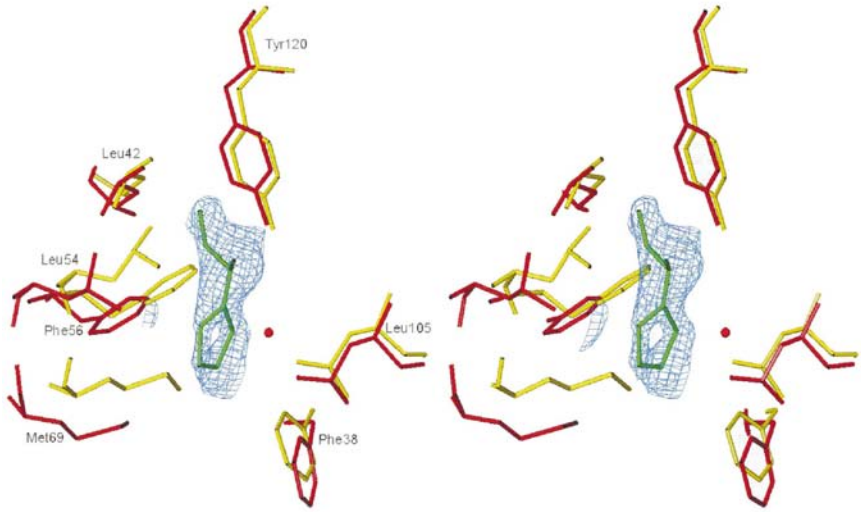

Figure 4
Superposition of the residues close to the binding site of the rMUP crystallographic structure (red) and averaged NMR structure (yellow) and the electron density (2σ in the $F_{obs} - F_{calc}$ map) observed in the binding cavity of the crystal rMUP. This density was modelled as 2-*sec*-butyl-4,5-dihydrothiazole.

Interestingly, the position of the modelled ligand in the structure presented here is inverted with respect to that described by Bocskei *et al.* (1992). In fact, this orientation is consistent with the ligand orientation reported in the NMR studies of the recombinant MUP-I (Zídek *et al.*, 1999).

Two water molecules (Wat172 and Wat235) were modelled in strong electron density in the interior of the hydrophobic pocket of rMUP. Wat172 is at a hydrogen-bonding distance from Tyr120 OH and from the main-chain carbonyl O atom of Leu40. Wat235 forms a hydrogen bond with the carbonyl group of Phe38. The temperature factors for Wat172 and Wat235 are 18 and 22 $Å^2$, respectively. These two water molecules are also present inside the hydrophobic pocket of the wild-type structure and seem to be structurally conserved.

Similar to other lipocalins such as OBP, the ligand-binding hydrophobic cavity of rMUP is rather large. Its internal volume calculated with *GRASP* (Nicholls *et al.*, 1991) is about 491 $Å^3$, while the volume occupied by the TZL is only about 120 $Å^3$. Such a difference may be related to the fact that the volatile pheromones must be released into the air and therefore there is a need for the protein to be able to modulate the conformation of the cavity.

4. Conclusions

Comparison of the present high-resolution X-ray structure of rMUP with its NMR-derived structure and with the crystallographic structure of wtMUP demonstrates that the overall fold of the protein is essentially the same. The similarity often extends to the local precision; that is, less well defined fragments in one structure are often also disordered in the other structure. However, there are a number of distinct differences in conformations of both the main chain and side chains of some residues. Positional differences are significantly smaller between the two crystallographic models, in spite of being determined in distinct crystal forms, than between the X-ray and NMR structures or within the ensemble of NMR structures themselves.

Some of these conformational differences were found to originate from the different environment found in the crystal and in solution. In the single-crystal form of the protein, differences may result from the crystallographic intermolecular interactions. It is not clear, however, whether the significantly higher r.m.s.d. values observed between the X-ray and NMR structures with respect to the values obtained by comparison of the X-ray models are a consequence of the different protein physical state or whether they reflect the precision of the various structure refinement protocols used. If we accept that in a crystal the interactions

of symmetry-related molecules may 'freeze' the protein in one of the possible conformational states existing in solution, different crystallographic contacts should lead to many conformational substructures 'randomly' chosen from an ensemble of possible conformations represented by NMR data. In this case, however, the r.m.s.d. between X-ray structures in different crystal forms should be comparable with the r.m.s.d. within the ensemble of the structures determined by NMR. In the current study this was not observed. On the contrary, the crystal structures differ less between themselves than with respect to the NMR model. This seems to be a general trend (MacArthur *et al.*, 1994), although the limited number of crystallographic models compared in the present study does not permit us to generalize this conclusion.

A clear electron density attributable to a possible ligand was observed in the hydrophobic pocket of rMUP. The electron density, corresponding to a small ring-shaped molecule, was modelled as 2-*sec*-butyl-4,5-dihydrothiazole, a natural ligand of wtMUP. The position but not the orientation of the molecule is similar to that observed in previous crystallographic study (Bocskei *et al.*, 1992).

Note added in proof: while the current paper was being reviewed, the crystal structure of recombinant MUP-I complexed with synthetic pheromones was determined (Timm *et al.*, 2001). The position and orientation of the pheromones bound to rMUP-I is in close agreement with the position and orientation of the ligand found in the present structure.

PK and IP are grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for financial support *via* grants 99/03387-4 and 00/02093-6 and to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This work was partly supported by grant CNR No. 99.02608.CT04 and MURST Cofin 1999 (AS).

References

Bacchini, A., Gaetani, E. & Cavaggioni, A. (1992). *Experientia*, **48**, 419–421.
Bianchet, M. A., Bains, C., Pelosi, P., Pevsner, J., Snyder, S. H., Monaco, H. L. & Amzel, L. M. (1996). *Nature Struct. Biol.* **3**, 934–939.
Bocskei, Z. S., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggioni, A., Findlay, J. B. C. & North, A. C. T. (1992). *Nature (London)*, **360**, 186–189.
Brownlow, S., Morais-Cabral, J. H., Cooper, R., Flower, D. R., Yewdall, S. J., Polikarpov, I., North, A. C. T. & Sawyer, L. (1997). *Structure*, **5**, 481–495.
Cavaggioni, A., Findlay, J. B. C. & Tirindelli, R. (1990). *Comp. Biochem. Physiol. B*, **96**, 513–520.
Cavaggioni, A., Sorbi, R. T., Keen, J. N., Pappin D. J. C. & Findlay J. B. C. (1987). *FEBS Lett.* **212**, 225–228.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dinh, B., Tremblay, A. & Dufour, D. (1965). *J. Immunol.* **95**, 574–582.
Dulac, C. & Axel, R. (1995). *Cell*, **83**, 195–206.
Ferrari, E., Lodi, T., Sorbi, R. T., Tirindelli, R., Cavaggioni, A. & Spisni, A. (1997). *FEBS Lett.* **401**, 73–77.
Finlayson, J., Mushinski, J., Stonard, M. D. & Potter, M. (1986). *Biochem. Genet.* **2**, 127–140.
Flower, D. R., North, A. C. T. & Attwood, T. K. (1993). *Protein Sci.* **2**, 753–761.
Folli, C., Calderone, V., Ottonello, S., Bolchi, A., Zanotti, G., Stoppini, M. & Berni, R. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 3710–3715.
Jemiolo, B., Harvey, S. & Novotny, M. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 4576–4579.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Katakura, Y., Totsuka, A. A. & Kaminogawa, S. (1994). *Biochim. Biophys. Acta*, **1207**, 58–67.
Knopf, J. L., Gallagher, J. F. & Held, W. A. (1983). *Mol. Cell Biol.* **3**, 2232–2240.
Kraulis, J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Krieger, J., Schmitt, A., Lobell, L., Gudermann, T., Schultz, G., Breer, H. & Boekhoff, I. (1999). *J. Biol. Chem.* **274**, 4655–4662.
Kuhn, N. J., Woodworth-Gutai, M., Gross, K. W. & Held, W. A. (1984). *Nucleic Acid Res.* **12**, 6073–6090.
Kuser, P., Krauchenco, S., Fangel, A. & Polikarpov, I. (1999). *Acta Cryst.* D**55**, 1340–1341.
Lücke, C., Franzoni, L., Abbate, F., Löhr, F., Ferrari, E., Sorbi, T. R. & Spisni, A. (1999). *Eur. J. Biochem.* **266**, 1210–1218.
MacArthur, M. W., Laskowski, R. A. & Thornton, J. M. (1994). *Curr. Opin. Struct. Biol.* **4**, 731–737.
Martini, S., Silvotti, L., Shivazi, A., Ryba, N. J. & Tirindelli, R. (2001). *J. Neurosci.* **21**, 843–848.
Mucignat-Carretta, C., Carretta, A. & Cavaggioni, A. (1995). *J. Physiol.* **486**, 517–522.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Novotny, M., Harvey, S., Jemiolo, B. & Alberts, J. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 2059–2061.
Novotny, M. V., Ma, W., Wiesler, D. & Zídek, L. (1999). *Proc. R. Soc. London B*, **266**, 2017–2022.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). *J. Synchrotron Rad.* **5**, 72–76.
Ryba, N. J. P. & Tirindelli, R. (1997). *Neuron*, **19**, 371–379.
Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. & Cambillau, C. (1996). *Nature Struct. Biol.* **3**, 863–867.
Timm, D. E., Baker, L. J., Mueller, H., Zidek, L. & Novotny, M. V. (2001). *Protein Sci.* **10**, 997–1004.
Vincent, F., Lobel, D., Brown, K., Spinelli, S., Grote, P., Breer, H., Cambillau, C. & Tegoni, M. (2001). *J. Mol. Biol.* **305**, 459–469.
Zídek, L., Stone, M. J., Lato, S. M., Pagel, M. D., Miao, Z., Ellington, A. D. & Novotny, M. V. (1999). *Biochemistry*, **38**, 9850–9861.