# Crystal structure of a novel type of odorant binding protein from Anopheles gambiae, belonging to the C+ class

Amandine Lagarde, Silvia Spinelli, Huili Qiao, Mariella Tegoni, Paolo Pelosi, Christian Cambillau

# Crystal Structure of a Novel Type of Odorant Binding Protein from *Anopheles Gambiae,* Belonging to the C$^+$ Class

**Amandine Lagarde[1], Silvia Spinelli[1], Huili Qiao[2], Mariella Tegoni[1], Paolo Pelosi[2] and Christian Cambillau[1*]**

[1]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités of Marseille, 163 Av. de Luminy Case 932, 13288 Marseille Cedex 09, France;

[2] Department of Biology of Agricultural Plants, University of Pisa, via S. Michele, 4, 56124 Pisa, Italy.

Address correspondence to: Christian Cambillau. Fax:+33.491.266.720; E-mail: cambillau@afmb.univ-mrs.fr

Running title: Structure of a C+ Class Odorant Binding Protein

Keywords: olfactory proteins, Anopheles gambiae, odorant binding proteins, protein structure, binding proteins, odorant transport.

**Abstract**

***Anopheles gambiae (Agam)* relies on its olfactory system to target human prey, leading eventually to injection of *Plasmodium falciparum*, the malaria vector. Odorant-binding proteins (OBPs) are the first line of proteins involved in odorant recognition. They interact with olfactory receptors and thus constitute an interesting target for insect control. We undertook a large-scale study of proteins belonging to the olfactory system of Agam with the aim of preventing insect bites by designing strong olfactory repellents. We determined the 3D structures of several Agam OBPs alone or in complex with model compounds. Here, we report the first 3D structure of a member of the OBP $C^+$ class, AgamOBP47, which has a longer sequence than classical OBPs and contains 6 disulphide bridges. AgamOBP47 possesses a core of six α-helices and three disulphide bridges, similar to the classical OBP fold. Two extra loops and the N- and C-terminal extra segments contain two additional α-helices and are maintained together by three disulphide bridges. They are embrace the classical OBP core domain. The binding site of OBP47 is located between the core and the additional domains. Two crevices are observed on opposite sides of OBP47, which are joined together by a shallow channel of sufficient size to accommodate a model of the best tested ligand. The binding sites of $C^+$ class OBPs exhibit therefore different characteristics of their binding site, as compared to classical OBPs, which should leave to markedly diverse functional implications.**

*The abbreviations used are*: PBP: pheromone binding protein; OBP: odorant binding protein; GOBP: General odorant binding protein; ESRF: European Synchrotron Radiation Facility; OR: Olfactory receptor; SNMP: Sensory Neuron Membrane Protein; rmsd: root mean square deviation.

**Introduction**

Olfaction is a finely tuned sense and is able to detect and discriminate between thousands of volatile molecules at low concentrations, some differing by only one or a few atoms. Olfaction in insects relies on dedicated organs, the antennae, which gather the olfactory neurons located at the base of a large number of sensilla [1]. Each sensillum can be considered as an elementary module possessing all the components necessary for translating a chemical signal into an electrical stimulus. The axons of olfactory neurons harbour olfactory receptors (ORs), which are stimulated by odorant molecules [2-4]. OR bathe into the sensillary lymph containing a large concentration of ~13-20 kDa proteins, the odorant binding proteins (OBPs) [5, 6]. [5, 6]. The OBPs can convey signals from pheromones (pheromone-binding proteins, PBPs), or from general odorants (general odorant-binding proteins, GOBPs) [7]. The OBPs have been recently shown to be directly involved in the olfactory process [8-10]. In *Drosophila melanogaster* (Dmel), a PBP (named Lush) changes conformation upon pheromone binding, a trigger for OR activation [9]. By contrast, *Antheraea polyphemus* (Apol, a lepidopteran) ORs can directly bind pheromones, but with less specificity and 1000-times less efficiency than when presented by a specific PBP [10].

Several structures of PBPs and GOBPs have been characterised [11-17]. They exhibit remarkable similarities, such as a six α-helix core, an internal cavity and three disulphide bridges, features that define them as "classical OBPs". However, an in-depth look at the structures shows significant differences at their C-termini, which have functional implications. Lepidopteran PBPs, such as those from *Bombyx mori* (Bmor) or Apol, are the longest classical OBPs. They possess a C-terminal segment of ~15 residues whose structure can adopt an α-helical fold and is buried inside the PBP cavity at acidic pH. At neutral pH, the same segment is elongated, located at the protein surface, and leaves the internal cavity free [11, 12, 18].The OBPs Lush [13] (Dmel), ASP1 [14] from *Apis mellifera* (Amel) or OBP1 [15] from *Anopheles gambiae* (Agam) are somewhat shorter and possess an elongated C-terminal segment buried inside the protein core, forming a wall with the internal cavity. The last residue is tightly bound by two strong hydrogen bonds via its carboxylic acid moiety to the protein core. In Amel ASP1 PBP, a mechanism of domain swapping, depending upon the ligand and pH, has been described [19, 20]. A third sub-class, including such as the PBP of the cockroach *Leucophera maderae* [16] or ASP2 OBP from Amel [17], is even shorter, with their C-termini ending at the cavity opening. We have named these three kinds of classical OBPs long, medium and short C-terminus sub-classes, respectively [21].

The genomes of several insects are now available , eg those of Bmor [22], Dmel [23], Amel [24], Agam [25, 26] and Aedes [27], in which numerous olfactory proteins, ORs, OBPs and chemosensory proteins (CSPs) have been identified. In addition to classical OBPs, the large repertoire of OBPs reported is composed of several distinct other classes: OBPs with only 4 cysteines, the $C^-$ class [24], OBPs with a longer chain and 12 or more cysteines, the $C^+$ class [28], and OBPs with a double domain of classical OBPs, that we term here "double-OBPs" [23]. Many outliers, such as Bmor GOBP2 [29], do not fit in these large classes.

With the goal of determining the molecular features of a $C^+$ class OBP [28] (Fig. 1), we have expressed, characterised and solved the crystal structure of OBP47 from *Anopheles gambiae* (AgamOBP47). We show that AgamOBP47 has six disulphide bridges and exhibits a new fold consisting of the classical OBP core, which is decorated with large loops and N- and C-terminal extensions that form new open binding sites on either face of the OBP. We expect that these differences might have significant consequences in terms of ligand binding and function.

**MATERIALS AND METHODS**

**Protein cloning, expression and purification.**

The protein was produced using *Escherichia coli* BL21 T7 (DE3) transformed with a pET-5 b(+) expression vector containing the AgamOBP47 cDNA sequence [30]. Cultures of T7 were grown in Turbo Broth medium supplemented with ampicillin and chloramphenicol. Cultures were grown at 37°C to an $OD_{600}$ of 0.6, induced by the addition of 1 mM IPTG, and grown again for 4 h at 37°C. Cellular pellets were resuspended in lysis buffer (50 mM Tris, pH 8, 1 mM PMSF, 0.25 mg ml-1 lysozyme) and frozen at -80°C for at least 2 h. After rapid thawing, they were treated with DNase, sonicated and centrifuged for 30 minutes at 14,000 x g. Pellets were resuspended in washing buffer (50 mM Tris, pH 8, 1 M Urea, 0.1% Triton) and centrifuged again for 30 minutes at 14,000 x g. The washed pellet was then resuspended with solubilisation buffer (50 mM Tris, pH 8, 8 M Urea) and incubated at room temperature with stirring for 30 minutes. Cellular debris was removed by ultracentrifugation (1 h at 100,000 x g at 20°C). Refolding was performed after screening for the best conditions [31]. β-Mercaptoethanol (10 mM final) was added to denatured AgamOBP47 (diluted to 5 mg/mL), and the mixture was stirred for 1 h at room temperature. Refolding was initiated by pipetting the protein solution into 20 volumes of refolding solution (50 mM Tris, pH 8, 10 mM β-mercaptoethanol) with constant stirring at 4°C. The reaction was allowed to proceed for 24 h at 4°C.

AgamOBP47 was then purified by ion exchange chromatography and gel filtration in the absence of β-mercaptoethanol. Anion exchange (Source 30-Q resin, Pharmacia) was performed in 50 mM Tris, pH 8, with a 0–500 mM NaCl gradient. The protein was recovered in the flow-through and concentrated on a 10 kDa cut off Ultrafiltration Membrane (Regenerated Cellulose, Millipore). The sample was then applied onto a FPLC gel-filtration column (Superdex S75 Prep Grade Column, Pharmacia) equilibrated with a buffer containing 15 mM Tris pH 8, 25 mM NaCl. Elution was carried out in the same buffer. The AgamOBP47 fractions were pooled and the mass of the protein was confirmed through MALDI-TOF mass spectroscopy.

**Biophysical characterisation**

The integrity and monodispersity of purified AgamOBP47 were analysed by a combination of UV spectrophotometry, multi-angle static light scattering (MALS), and refractometry coupled on-line with an analytical size exclusion chromatography (SEC) column, as described by Veesler *et al.* [32]. A 15-ml KW-804 column (Shodex) was used on an Alliance HPLC 2695 system (Waters). The protein concentration was adjusted to 18 mg. ml-1 (~1mM) (Fig. S1). A buffer containing 50 mM Tris, pH 7.5, 50 mM NaCl and a flow rate of 0.5 ml/min were used for column equilibration and running. Circular dichroism spectra were recorded using 0.1 mg/ml AgamOBP47 in 10 mM NaHPO4, pH 7.5 on a Jasco J-810 spectropolarimeter. Spectra deconvolution was performed between 190 and 260 nm with CDNN software and confirmed AgamOBP47 as a α-helical protein.

**Crystallisation of OBP47, structure determination and refinement.**

Purified OBP47 was concentrated up to 12 mg/ml in 15 mM Tris, pH 8, 25 mM NaCl using an Amicon concentrator with a 10 kDa cut off. Crystallisation screening experiments were performed using several commercial kits with the sitting drop method and the nanodrop technology with Greiner plates [33]. The initial crystallisation droplet contained 200 nl of a 12 mg/ml protein solution in 15 mM Tris, pH 8.0, 25 mM NaCl mixed with 100 nl of well solution containing 36% PEG MME 5K and 0.1 M sodium acetate, pH 6.0. Optimisation experiments led to crystals growth within 4-7 days in3 0 to 35% PEG MME 5K and 0.1 M sodium acetate, pH 5.7 to 6.3.

In order to incorporate $Cs^+$ and $I^-$ in the crystals, we saoked several crystals in synthetic solutions containing 35% PEG MME 5K and 0,1M Sodium Acetate, pH6, with increasing

concentrations of NaI and CsI, up to the final concentration of 0.5 M for each salt. Crystals were frozen using a solution containing 35% PEG MME 5K and 0,1M Sodium Acetate, 0.5 M NaI and 0.5 M of CsI, pH 6.0.. Diffraction images for 0.5 M NaI-0.5 M CsI-soaked crystals were collected on the PROXIMA-1 beamline (SOLEIL) at wavelengths of 1.7712 Å and 0.9537 Å with an ADSC 315r detector. Data sets were processed using XDS and scaled with XSCALE[34], resulting in a data set at 2.3 Å resolution. The iodine positions were determined with SHELXC and SHELXD [35] and subsequently used for SAD phasing using PHASER [36]. These phases were improved by solvent flattening and histogram matching using the program PARROT [37] assuming 45% solvent content, resulting in interpretable maps from which the BUCCANEER program [37] was able to build an initial model comprising more than 90% of the expected amino acids. A native dataset was collected at ESRF's BM14 at 1.8 Å resolution. Structure refinement was performed with AutoBUSTER [38] and alternated with rebuilding using COOT [39] using TLS parameters. For the native data set, the NaI-CsI data set was used as a starting model, and refinement was performed in the same way as previously, resulting in a complete polypeptide chain (1-173) and 292 water molecules with $R/R_{free}$ of 18.7% and 22.3%, respectively, and all residues but one (Asn69) in Coot's Ramachandran preferred region. Figures were made with Pymol (*http://pymol.sourceforge.net/*), or Chimera [40] (Table 1). The coordinates and reflection files have been deposited in the Protein Data Bank at RCSB (http://www.rcsb.org/pdb/) with accession number 3PM2.

## RESULTS

### Overall structure of apo-OBP47

The 3D structure of AgamOBP47 was solved using remote SAD techniques with $Cs^+$ and $I^-$ as anomalous scatterers and a data set collected at 1.77 Å wavelength with 2.3 Å resolution. Cesium and iodine from NaI/CsI crystal soaking provided a sufficient phasing power of ~6 anomalous electrons per $Cs^+$ or $I^-$ ion. A native data set was further obtained at 1.8 Å resolution, and the previous model was refined at this resolution (Table 1).

AgamOBP47 belongs to the $C^+$ OBP class, is 173 residues long and possess 13 cysteine residues. The whole chain was built in the electron density map. Similar to classical OBPs, the structure is mostly helical. However, eight α-helices are observed in OBP47 rather than six in classical OBPs such as BmoOBP1 or Lush OBP (Table 2). The OBP47 structure exhibits a long N-terminal stretch of residues (residues 3-25) devoid of defined secondary structures (helix or strand) and containing three Cys residues (Fig. 2A). They form three disulphide bridges with Cys residues in the C-terminal segment (149-173), namely disulphide bridges 5-170, 18-160 and 19-149. At residue 25, the long H1 helix (26-35) is followed by a kink and helix H2 (37-45). Helix H1 establishes a disulphide bridge with a Cys residue localised in a β-turn (Cys33-Cys59). Helices H3-H7 follow and are linked by two disulphide bridges, 55-130 and 103-140 (Table 2). Two extensions are observed: a hairpin turn between helices H3 and H4 (loop 64-72), a feature conserved in many classical OBPs, and a long connecting loop between H6 and H7 (loop 119-131). The C-terminal segment begins with a long stretch of residues (149-157) linking helix H7 to the C-terminal helix H8 (Table 2) and finishes at the end of helix 8. Overall, the structure of AgamOBP47 appears rather flat with two crevices on either side (Fig. 2B,C).

The OBPs may form functional homo- or hetero-dimers [30]. However, equivocal data have been reported that do not allow for forming firm conclusions for or against dimerisation of OBPs. Dimers have been observed for AgamOBP1, but the authors do not conclude that this observation may also apply to solutions. In the case of OBP47, we observed a monomer of 19 kDA using multi-angle static light scattering (MALS)/ dynamic light scattering (DLS) and refractometry (RI) coupled on-line with a SEC column (MALS/UV/RI/SEC) (Fig. S1). The concentrations of these samples were ~1 mM. We used sample concentrations of ~0.6 mM in

our crystallisation experiments. The PISA server [41] identified a putative dimer in the crystal. Its buried surface accounts for 1300 $Å^2$, 7.5 % of the total dimer surface. According to PISA, this makes this complex unstable, with an evaluated $\Delta G_{diss}$ of 1.1 Kcal, corresponding to a $K_{diss}$ of ~2 mM. Noteworthy, an OBP concentration of up to 10 mM was measured in the lymph.

**Comparison of OBP47 with classical OBPs**

Dmel Lush is the best-characterised OBP from Diptera and the OBP47 sequence remotely resembles the Lush sequence. However, the best alignments were obtained when structural data were included in the sequence alignment and when some Cys residues were forced to be aligned (Fig. 3A). When the 3D structure of OBP47 was superimposed to that of Lush, most of the secondary structural elements, helices, and disulphide bridges and some of the loops overlapped (Fig. 3, Movie S1). Five helices are decently superimposed between OBP47 and Lush: H1, H3, H4 and H5, as well as OBP47 H7 with Lush H6 (Table 2). Helix 2 from both OBPs occupies different places in space, and two extra helices (H6 and H8) are observed in OBP47 (Fig. 3B). When the conserved structures were compared to the non-conserved structures, the latter ones form a crescent-shaped structure embracing the classical OBP conserved domain (Fig. 3C,D,E). The two cavities identified in each face of the molecular surface are formed at the interface between the conserved and non-conserved domains (Fig. 3D,E).

**AgamOBP47 putative binding sites**

The molecular surface of AgamOBP47 does not have the usual features of the classical OBPs. The internal cavity of classical OBPs is not observed in OBP47, but instead two shallow crevices are observed on opposite faces and are linked by a channel of 8 Å diameter (Fig. 2B,C, 3D,E and S2, Movie S2).

Binding site 1 (Fig. S2) is ~ 12 Å x 10 Å and 14 Å deep. The lips of the crevice are mainly formed by the N-terminus segment 5-34,38-42 (Fig. S2A,B, blue area), which forms three of its faces, whereas the last cavity wall is formed by the C-terminus segment 165-173 (Fig. S2A,B, red area). Of note, the carboxylic acid group of the last residue, Leu 173, forms a strong hydrogen bond with His 34 atom ND1, a feature also found in classical medium-length OBPs [13, 15, 19, 20]. The walls of the cavity are formed by numerous hydrophobic residues, five aromatics (Phe, Tyr) and six aliphatics (Leu, Val, Met), Pro, Gly and one semi-polar residue Thr (Table 3).

Binding site 2 (Fig. S2) is larger than site and is ~ 14 Å x 11 Å and 14 Å deep. The lips of the crevice are formed by the N-terminal segment 1-9 (Fig. S2C,D, blue area) and by the C-terminal segment 165-173, which plunges from the surface into the protein (Fig. S2B, red area). The C-terminal segment forms a common wall located between the two crevices. The three last components of the crevice external ring are residues 45-52 (Fig. S2C,D, green area), residues 103-112 (Fig. S2C,D, pink area), and residues 135-143 (Fig. S2C,D, yellow area). The walls of the binding sites and channel are formed by some hydrophilic residues (Table 3), and a large number of water molecules cover them (Fig. S2C,D,F). The absence of an internal closed cavity, compared with classical OBPs such as Lush, results from the insertion of segments or side chains inside the protein core, at a place where the internal cavity is located in Lush. These insertions belong to the N- and C-terminal segments, residues 23-25 and 137, 141 and 145, respectively, and to three side chains of residues of helix H4 (Met 61 and Leu 77) and helix H5 (Tyr 81).

To investigate the binding capacity and geometry of AgamOBP47, we produced a model of a complex with 4-hydroxy-4'-isopropyl-azobenzene (AZO), the best tested ligand of OBP47 [30]. We docked the azo moiety in the thin channel, while letting the more bulky aromatic rings fit into the two lateral crevices. We then performed geometry optimisation using REFMAC. The model revealed that AZO fits well in the cavity and makes contacts with all of the residues from the cavities (Fig. S3, Table 3, Movie S2).

## DISCUSSION

The 3D structure of AgamOBP47 is the first reported structure from a member of the $C^+$ class OBPs. Contrary to all other PBPs/OBPs with known structures, AgamOBP47 features six disulphide bridges, eight α-helices, and no buried internal cavity. However, it clearly evolved from classical OBPs by the adjunction of long, disulphide-stabilised N- and C-termini to a central core, thus resembling many classical OBPs. Although some classical OBPs have N-terminal extensions, such as *Apis mellifera* ASP1 [19, 20] or *A. gambiae* OBP1 [15], or C-terminal extensions, such as *Bombyx mory* PBP [11] or GOBP2 [29], these extensions are much shorter. Moreover, an α-helix from classical OBPs is displaced in AgamOBP47 and an additional long loop is present. As a result of these important changes, AgamOBP47 exhibits two superficial crevices linked by an internal channel. This arrangement of surfaces and channels has not been observed in any insect OBP nor in any chemo sensory proteins (CSPs) [42]. However, Agam anti-inflammatory proteins, which form a tandem repeat of OBP domains, also exhibit surface-located binding cavities [43]. Although the implications of these changes are still unknown, they are likely related to the type of molecules that they are able to bind: less bulky with hydrophilic extremities and relatively straight.

Class $C^+$ OBPs have not been observed in Hymenoptera or in Lepidoptera. To date, they have been identified only in Diptera, *Anopheles*, *Culex* and *Drosophila*. They represent a unique and major step from an evolutionary point of view. The evolution from classical OBPs to the $C^-$ OBP class, as observed in Hymenoptera (e.g., *Apis*) and Diptera (e.g., *Drosophila*), involved only usual changes in sequences except for mutations of two cysteines. Moreover, an OBP presenting intermediate characters is observed in *Apis mellifera* (AmelOBP14). The other interesting class of "double-OBPs", i.e., proteins that possess a double classical OBP motif, may have emerged by gene duplication and fusion, followed by functional differentiation in OBPs and anti-inflammatory proteins expressed in the salivary gland [43]. However, it is more difficult to decipher the mechanism by which the $C^+$ OBPs class evolved. The $C^+$ OBP class is much more heterogeneous than the $C^-$ or classical OBPs classes. They exhibit large insertions/deletions of loops between disulphide bridges, different lengths of helices and variable N- and C-termini (Fig. 1). Therefore, homology modelling is challenging, except when using pairs of closely related members. For instance, AgamOBP47 presents 55% identity and 75% similarity with OBP48 (very abundant in antennae [30]) and only one gap. Both proteins share comparable surface properties and binding sites, as assessed by modelling (not shown). Considering that the termini form a large part of the binding cavities of $C^+$ class OBPs and that AgamOBP47 is among the shortest ones, we expect large functional variability in the $C^+$ Class OBP family through the formation of binding sites with different shapes and sizes.

**Accession codes.** The structure coordinates have been deposited to the Protein Data Bank with accession code 3PM2.

**Acknowledgements**

This work was supported in part by the PACA Region, by the Marseille-Nice-Genopole®, and by the FP7 EU project Enaromatic (FP7/2007-2013,Grant Agreement No. FP7-222927).

We are grateful to Dr. Pierre Legrand for help with data collection and structure determination and to the synchrotron Soleil (Saint-Aubin, France) and European Synchrotron Facility (Grenoble, France) for beamtime allocation.

**REFERENCES**

1-Krieger, J., Mameli, M. and Breer, H. (1997) Elements of the olfactory signaling pathways in insect antennae. Invert. Neurosci. **3**, 137-144

2-Vosshall, L. B. and Stocker, R. F. (2007) Molecular architecture of smell and taste in Drosophila. Annu. Rev. Neurosci. **30**, 505-533

3-Sato, K., Pellegrino, M., Nakagawa, T., Vosshall, L. B. and Touhara, K. (2008) Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature. **452**, 1002-1006

4-Wicher, D., Schafer, R., Bauernfeind, R., Stensmyr, M. C., Heller, R., Heinemann, S. H. and Hansson, B. S. (2008) Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature. **452**, 1007-1011

5-Vogt, R. G. and Riddiford, L. M. (1981) Pheromone binding and inactivation by moth antennae. Nature. **293**, 161-163

6-Krieger, J., Ganssle, H., Raming, K. and Breer, H. (1993) Odorant binding proteins of Heliothis virescens. Insect Biochem. Mol. Biol. **23**, 449-456

7-Stengl, M., Zufall, F., Hatt, H. and Hildebrand, J. G. (1992) Olfactory receptor neurons from antennae of developing male Manduca sexta respond to components of the species-specific sex pheromone in vitro. J. Neurosci. **12**, 2523-2531

8-Benton, R., Vannice, K. S. and Vosshall, L. B. (2007) An essential role for a CD36-related receptor in pheromone detection in Drosophila. Nature. **450**, 289-293

9-Laughlin, J. D., Ha, T. S., Jones, D. N. and Smith, D. P. (2008) Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein. Cell. **133**, 1255-1265

10-Forstner, M., Breer, H. and Krieger, J. (2009) A receptor and binding protein interplay in the detection of a distinct pheromone component in the silkmoth Antheraea polyphemus. Int. J. Biol. Sci. **5**, 745-757

11-Sandler, B. H., Nikonova, L., Leal, W. S. and Clardy, J. (2000) Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. Chem. Biol. **7**, 143-151

12-Horst, R., Damberger, F., Luginbuhl, P., Guntert, P., Peng, G., Nikonova, L., Leal, W. S. and Wuthrich, K. (2001) NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. Proc. Natl. Acad. Sci. U.S.A.. **98**, 14374-14379

13-Kruse, S. W., Zhao, R., Smith, D. P. and Jones, D. N. (2003) Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster. Nat. Struct. Biol. **10**, 694-700

14-Lartigue, A., Gruez, A., Briand, L., Blon, F., Bezirard, V., Walsh, M., Pernollet, J. C., Tegoni, M. and Cambillau, C. (2004) Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee Apis mellifera L. J. Biol. Chem.. **279**, 4459-4464

15-Wogulis, M., Morgan, T., Ishida, Y., Leal, W. S. and Wilson, D. K. (2006) The crystal structure of an odorant binding protein from Anopheles gambiae: evidence for a common ligand release mechanism. Biochem. Biophys. Res. Commun. **339**, 157-164

16-Lartigue, A., Gruez, A., Spinelli, S., Riviere, S., Brossut, R., Tegoni, M. and Cambillau, C. (2003) The crystal structure of a cockroach pheromone-binding protein suggests a new ligand binding and release mechanism. J. Biol. Chem.. **278**, 30213-30218

17-Lescop, E., Briand, L., Pernollet, J. C. and Guittet, E. (2009) Structural basis of the broad specificity of a general odorant-binding protein from honeybee. Biochemistry. **48**, 2431-2441

18-Damberger, F. F., Ishida, Y., Leal, W. S. and Wuthrich, K. (2007) Structural Basis of Ligand Binding and Release in Insect Pheromone-binding Proteins: NMR Structure of Antheraea polyphemus PBP1 at pH 4.5. J. Mol. Biol.. **373**, 811-819

19-Pesenti, M. E., Spinelli, S., Bezirard, V., Briand, L., Pernollet, J. C., Campanacci, V., Tegoni, M. and Cambillau, C. (2009) Queen bee pheromone binding protein pH-induced domain swapping favors pheromone release. J. Mol. Biol.. **390**, 981-990
20-Pesenti, M. E., Spinelli, S., Bezirard, V., Briand, L., Pernollet, J. C., Tegoni, M. and Cambillau, C. (2008) Structural basis of the honey bee PBP pheromone and pH-induced conformational change. J. Mol. Biol.. **380**, 158-169
21-Tegoni, M., Campanacci, V. and Cambillau, C. (2004) Structural aspects of sexual attraction and chemical communication in insects. Trends Biochem. Sci. **29**, 257-264
22-Gong, D. P., Zhang, H. J., Zhao, P., Xia, Q. Y. and Xiang, Z. H. (2009) The odorant binding protein gene family from the genome of silkworm, Bombyx mori. BMC Genomics. **10**, 332
23-Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J. and Tanouye, M. A. (2002) Genome-wide analysis of the odorant-binding protein gene family in Drosophila melanogaster. Genome Res.. **12**, 1357-1369
24-Foret, S. and Maleszka, R. (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). Genome Res.. **16**, 1404-1413
25-Zdobnov, E. M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R. R., Christophides, G. K., Thomasova, D., Holt, R. A., Subramanian, G. M., Mueller, H. M., Dimopoulos, G., Law, J. H., Wells, M. A., Birney, E., Charlab, R., Halpern, A. L., Kokoza, E., Kraft, C. L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G. M., Salzberg, S. L., Sutton, G. G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F. H., Ribeiro, J., Gelbart, W. M., Kafatos, F. C. and Bork, P. (2002) Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science. **298**, 149-159
26-Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Venter, J. C., Weissenbach, J., Kafatos, F. C., Collins, F. H. and Hoffman, S. L. (2002) The genome sequence of the malaria mosquito Anopheles gambiae. Science. **298**, 129-149
27-Zhou, J. J., He, X. L., Pickett, J. A. and Field, L. M. (2008) Identification of odorant-binding proteins of the yellow fever mosquito Aedes aegypti: genome annotation and comparative analyses. Insect Mol. Biol.. **17**, 147-163
28-Li, Z., Shen, Z., Zhou, J. and Field, L. (2003) Bioinformatics-based identification of chemosensory proteins in African Malaria Mosquito, Anopheles gambiae. Genomics Proteomics Bioinformatics. **1**, 288-298

29-Zhou, J. J., Robertson, G., He, X., Dufour, S., Hooper, A. M., Pickett, J. A., Keep, N. H. and Field, L. M. (2009) Characterisation of Bombyx mori Odorant-binding proteins reveals that a general odorant-binding protein discriminates between sex pheromone components. J. Mol. Biol.. **389**, 529-545
30-Qiao, H., He, X., Schymura, D., Ban, L., Field, L., Dani, F. R., Michelucci, E., Caputo, B., Torre, A. D., Iatrou, K., Zhou, J. J., Krieger, J. and Pelosi, P. (2011) Cooperative interactions between odorant-binding proteins of Anopheles gambiae. Cell. Mol. Life Sci.. **68**, 1799-813.
31-Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. and Bignon, C. (2004) High-throughput automated refolding screening of inclusion bodies. Protein Sci. **13**, 2782-2792
32-Veesler, D., Blangy, S., Siponen, M., Vincentelli, R., Cambillau, C. and Sciara, G. (2009) Production and biophysical characterization of the CorA transporter from Methanosarcina mazei. Anal. Biochem.. **388**, 115-121
33-Lartigue, A., Gruez, A., Briand, L., Pernollet, J. C., Spinelli, S., Tegoni, M. and Cambillau, C. (2003) Optimization of crystals from nanodrops: crystallization and preliminary crystallographic study of a pheromone-binding protein from the honeybee Apis mellifera L. Acta Crystallogr. D Biol. Crystallogr. **59**, 919-921
34-Kabsch, W. (2010) Xds. Acta Crystallogr. D Biol. Crystallogr. **66**, 125-132
35-Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. **58**, 1772-1779
36-McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J Appl. Crystallogr. **40**, 658-674
37-Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. **62**, 1002-1011
38-Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr. D Biol. Crystallogr. **60**, 2210-2221
39-Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**, 2126-2132
40-Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput. Chem. **25**, 1605-1612
41-Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol.. **372**, 774-797
42-Campanacci, V., Lartigue, A., Hallberg, B. M., Jones, T. A., Giudici-Orticoni, M. T., Tegoni, M. and Cambillau, C. (2003) Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. Proc. Natl. Acad. Sci. U.S.A.. **100**, 5069-5074
43-Calvo, E., Mans, B.J., Ribeiro,J.M.C., Andersen,J.F. (2009) Multifunctionality and mechanism of ligand binding in a mosquito antiinflamatory protein. Proc. Natl. Acad. Sci. U.S.A.. **106**, 3728-3733
44-Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. **2**, 953-971
45-Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res.. **16**, 10881-10890
46-Gouet, P., Courcelle, E., Stuart, D. I. and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. **15**, 305-308
47-Lin, C. Y., Lin, F. K., Lin, C. H., Lai, L. W., Hsu, H. J., Chen, S. H. and Hsiung, C. A. (2005) POWER: PhylOgenetic WEb Repeater--an integrated and user-optimized framework for biomolecular phylogenetic analysis. Nucleic Acids Res.. **33**, W553-556

## FIGURE LEGENDS

**Figure 1**: Evolutionary tree and sequence alignment of $C^+$ OBPs from *Anopheles gambiae*. The secondary structures (top line) were assigned according to the AgamOBP47 crystal structure. The Cys and other conserved residues are depicted by white letters on grey background. Other semi-conserved residues are depicted in grey and are boxed in grey. The disulfide bridges are numbered 1 to 6. The signal peptides, identified with the SignalP software at Expasy, have been removed [44]. Alignement has been performed with Multalin (http://multalin.toulouse.inra.fr/multalin/ [45]) and ESPript [46]. The tree has been generated with PhylOgenetic WEb Repeater (POWER at http://power.nhri.org.tw/power/home.htm) [47].

**Figure 2**: Structure of apo *Anopheles gambiae* OBP47. **A**) Ribbon representation with rainbow colouring mode from blue (N-terminus) to red (C-terminus). We identified eight α-helices (H1-H8) as well as the Cys residues forming the six disulphide bridges (black numbers). The single Cys residue is identified in italics. **B**) Molecular surface of AgamOBP47. The shallow cavity is identified by a blue dot. **C**) This molecular surface has been clipped (blue line) and rotated by 90°. The cavity mentioned in (B) is denoted with a blue arrow. On the opposite side, a second cavity (green arrow) is related to the former one by a thin channel.

**Figure 3**: Sequence and structure comparisons of *Anopheles gambiae* OBP47 with *Drosophila melanogaster* Lush. **A**) The secondary structures have been assigned according to AgamOBP47 (PDB:3PM2) and Lush (PDB: 1OOH)[13] crystal structures. Conserved residues have been boxed in red. The 6 extra Cys residues of OBP47 have been boxed in green. The six disulfide bridges of OBP47 are numbered 1 to 6, while the three disulfide bridges of Lush are numbered 4 to 6. Alignement has been performed as in Figure 1. **B**) stereo ribbon view of the superimposed structures of AgamOBP47 (rainbow mode, blue to red) and DmelLush (grey). The α-helices have been identified by white numbers(1-8; OBP47) and black numbers (1-6; Lush). **C**) Ribbon representation of AgamOBP47, in the same orientation as in (B). The structures which are conserved with Lush are depicted in violet (the « core »). The non-conserved domains are the N-terminus (3-24 ; blue), the C-terminus (149-173 ; red), the displaced domain 39-53 (light orange) and the additional loop 111-129 (light yellow). The disulfide bridges are depicted by green sticks and numbered. **D**) Surface representation of OBP47, with the same orientation and coloring mode as in (C). Note that the shallow cavity (white arrow) is located at the interface between the « core » and the non-conserved domains. **D**) Surface representation of OBP47, rotated 180° with respect to (D). The second cavity is identified by a white arrow.

Biochemical Journal Immediate Publication. Published on 12 May 2011 as manuscript BJ20110522

**Table 1**. Data collection and refinement statistics.

| | Native OBP47 | NaI-CsI OBP47 |
|---|---|---|
| **DATA COLLECTION** | | |
| Beamline | BM14 (ESRF ) | PROXIMA 1 (SOLEIL) |
| Space group /cell | $P4_12_12$ / a=b=36.3 Å, c=263.5 Å | $P4_12_12$ / a=b=35.6 Å, c=263.2 Å |
| Wavelength (Å) | 0.97625 | 1.7712 |
| Resolution limits[a] (Å) | 36.3-1.8 (1.9-1.8) | 45-2.3 (2.36-2.3) |
| Rmerge[a] (%) | 16.7 (55.1) | 12.4 (46.4) |
| Nr. of observations[a] | 201596 (21799) | 135651 (5194) |
| Nr. unique reflections[a] | 17578 (2484) | 14657 (1027) |
| Mean((I)/sd(I)) [a] | 19.3 (4.0) | 12.8 (3.41) |
| Completeness[a] (%) | 99.8 (99.7) | 99.2 (93.4) |
| Multiplicity[a] | 11.5 (8.8) | 9.25 (5.1) |
| **REFINEMENT** | | |
| Resolution[a] (Å) | 35.0-1.8 (1.92-1.8) | |
| Nr of reflections[a] | 17530 (2756) | |
| Nr of protein/water atoms | 1307/292 | |
| Nr test set reflections | 887 | |
| Rwork/Rfree[a] (%) | 18.7/22.3 (21/26.2) | |
| r.m.s.d.bonds(Å)/angles (°) | 0.010/1.01 | |
| B-wilson / B-average | 23.4 / 26.5 | |
| Coot's ramachandran Preferred / allowed % | 99.42 / 0.58 | |

[a] parenthesis refer to the highest resolution bin.

**Table 2**: helical segments definition (A) and disulfide bridges localization (B) in AgamOBP47 structure. The signs =, ≠ and ~ indicate that the hecices in the comparison are at identical, different or similar postions, respectively.

A/

| Helix OBP47 / Lush | Helix AgamOBP47 | Helix Dmel Lush |
|---|---|---|
| H1 ~ H1 | E 26 – S 35 | T 2 – A 18 |
| H2 ≠ H2 | I 37 – M 45 | K 24 – V 33 |
| H3 = H3 | C 54 – S 63 | S 41 – G 56 |
| H4 ~ H4 | D 73 – T 85 | N 65 – V 77 |
| H5 = H5 | E 91 – D 108 | M 81 – R 93 |
| H6 new | S 110 – A 118 | - |
| H7 ~ H6 | P 132 – E 148 | E 101 – N 116 |
| H8 new | D 158 – Y 166 | - |

B/

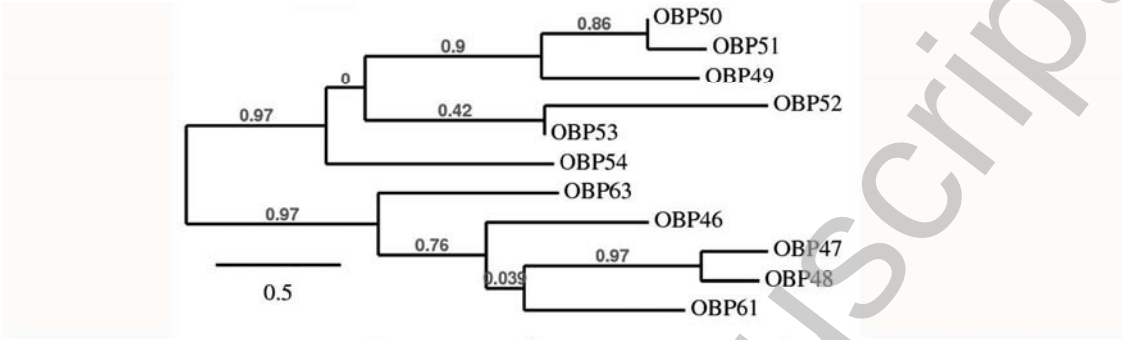
| Cys-Cys OBP47 | Cys-Cys Lush | Domains |
|---|---|---|
| 5 – 170 | - | N-term - Cterm |
| 18 - 160 | - | N-term – Cterm H8 |
| 19 - 149 | - | N-term - Cterm |
| 33 - 59 | 17 - 50 | core H1 – core turn |
| 55 - 130 | 46 - 103 | Core H3 – core H5 |
| 103 - 140 | 92 - 112 | Core H5 – core H7 (H6 lush) |
| Cys 54 | | single |

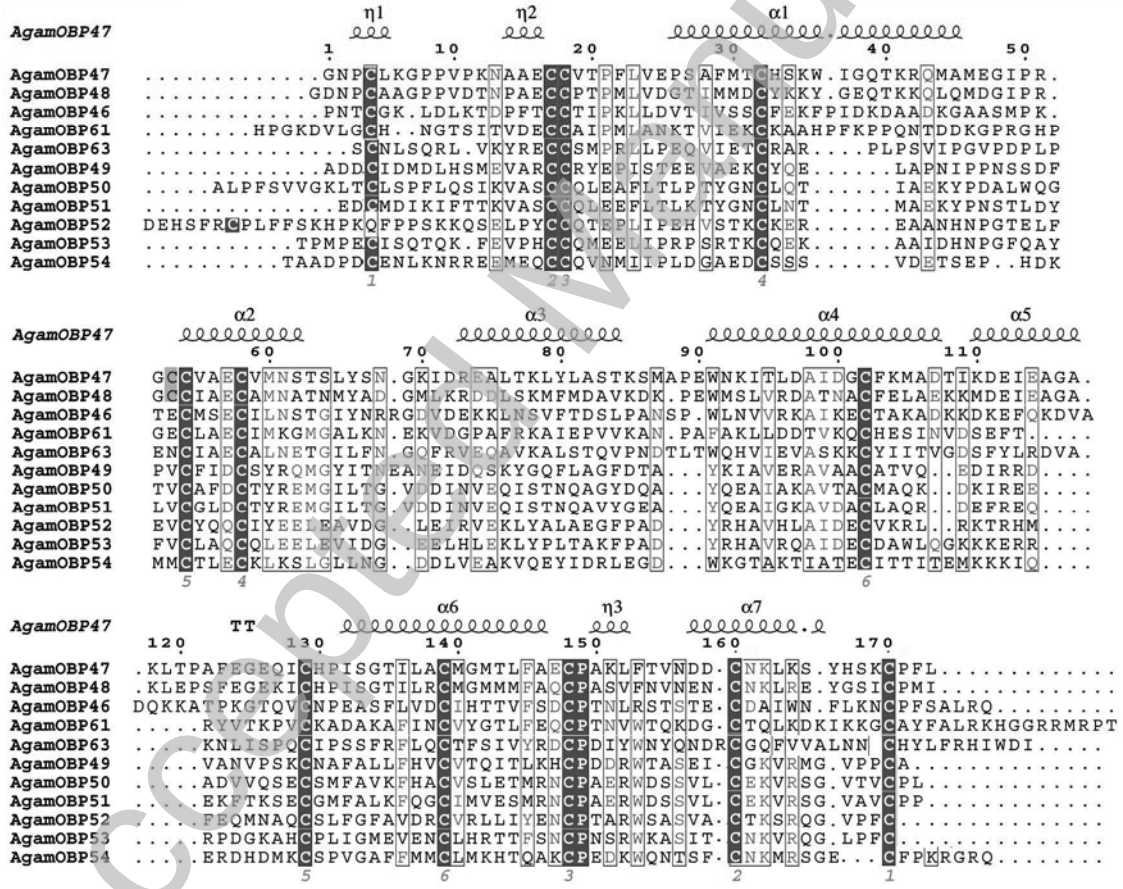
**Table 3**: AgamOBP47 residues forming the two binding crevices, linked by the internal tunnel, located on opposite faces of OBP46 (see Figures 3A and Figure S2A,B). Some of these residues also interact with 4-Hydroxy-4'-isopropylazobenzene in the modelled complex (labelled with *) (Figure S3).

| Binding site 1 | Binding site 2 |
|---|---|
| Thr 21 | Asn 3 |
| Pro 22 | Cys 5 |
| Phe 23 * | Thr 41 |
| Leu 24 | Gln 44 |
| Val 25 | Met 45 |
| Phe 30 * | Pro 51 |
| Tyr 81 | Cys 54 |
| Leu 138 * | AlA 139 * |
| Met 141 | Pro 171 * |
| Gly 142 | Leu 173 |
| Met 143 | |
| Leu 145 | |
| Phe 146 | |
| Phe 172 * | |

Biochemical Journal Immediate Publication. Published on 12 May 2011 as manuscript BJ20110522

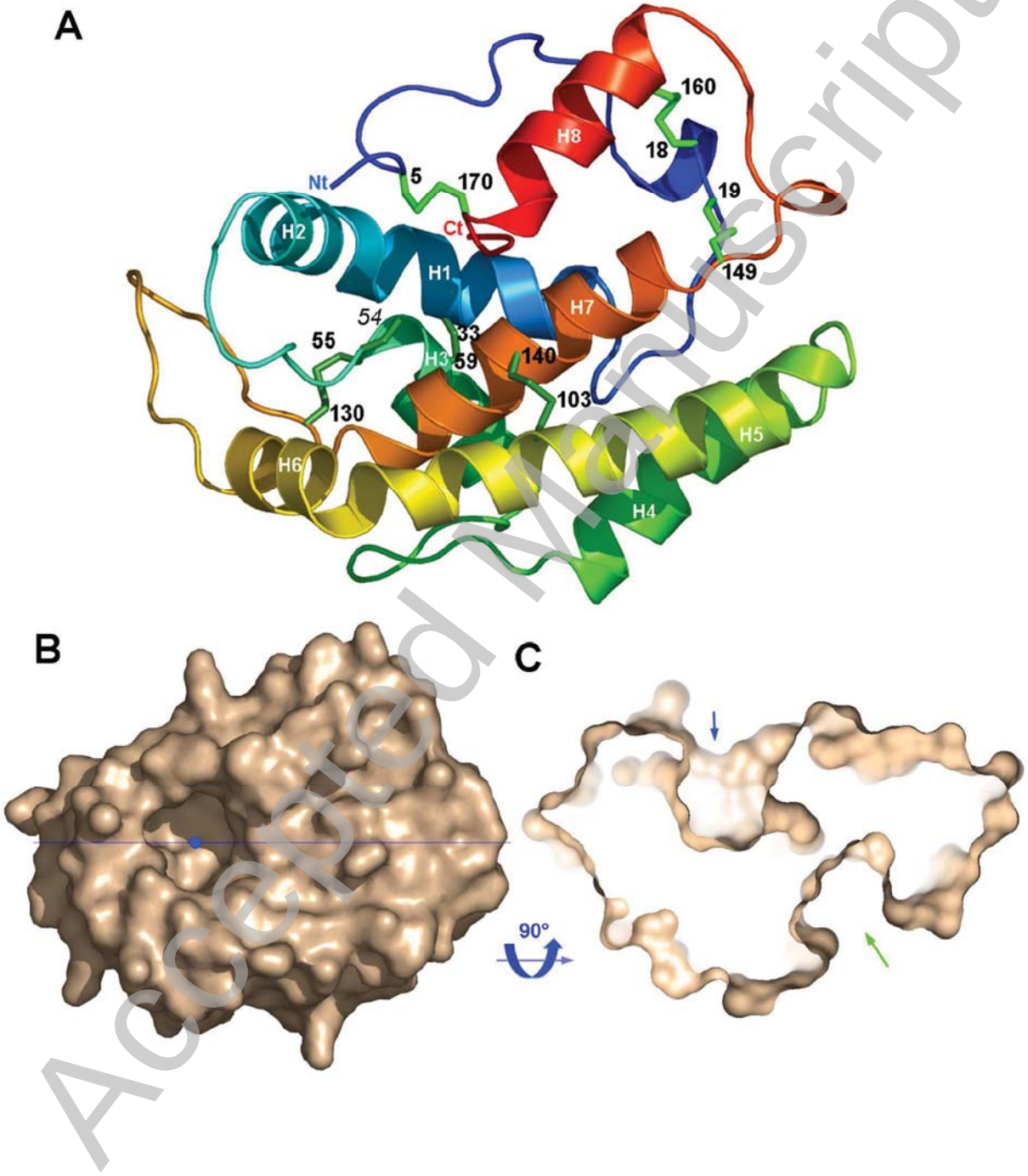




Biochemical Journal Immediate Publication. Published on 12 May 2011 as manuscript BJ20110522





Biochemical Journal Immediate Publication. Published on 12 May 2011 as manuscript BJ20110522