A new non-classical fold of varroa odorant-binding proteins reveals a wide open internal cavity

Beatrice Amigues, Jiao Zhu, Anais Gaubert, Giovanni Renzone, Philippe Leone, Simona Arena, Isabella Maria Fischer, Harald Paulsen, Wolfgang Knoll, Andrea Scaloni, et al.

A new non-classical fold of varroa odorant-binding proteins reveals a wide open internal cavity

Beatrice Amigues[1#], Jiao Zhu[2,4#], Anais Gaubert[1], Giovanni Renzone[3], Philippe Leone[1], Simona Arena[3], Isabella Maria Fischer[2], Harald Paulsen[4], Wolfgang Knoll[2,5], Andrea Scaloni[3], Alain Roussel[1], Christian Cambillau[1*], Paolo Pelosi[2*]

[1]*Architecture et Fonction des Macromolécules Biologiques (AFMB, UMR 6098), Centre National de la Recherche Scientifique (CNRS) & Aix-Marseille Université (AMU), Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France*

[2]*Austrian Institute of Technology GmbH, Biosensor Technologies, Konrad-Lorenz Straße, 24, 3430 Tulln, Austria*

[3]*Proteomics & Mass Spectrometry Laboratory, ISPAAM, National Research Council, 80147 Napoli, Italy.*

[4]*Faculty of Biology, Institute of Molecular Physiology, Johannes Gutenberg-Universität, 55099 Mainz, Germany*

[5]*Department of Physics and Chemistry of Materials, Faculty of Medicine/Dental Medicine, Danube Private University, Krems, Austria*

[#] Contributed equally to the work

*Correspondence to CC (christian.cambillau@univ-amu.fr) or PP (paolo.pelosi@ait.ac.at; ppelosi.obp@gmail.com)

Abstract

Insect odorant-binding proteins (OBPs) are not present in other arthropods nor in other living organism. However, ticks, mites, spiders and millipedes contain genes encoding protein with sequence similarity to insect OBPs. In this work we have explored the structure and function of such non-insect OBPs in the mite *Varroa destructor*, a major pest for honey bees. Varroa OBPs present six cysteines, paired into three disulphide bridges, but their positions in the sequence and their connections are different from those of their insect counterparts. VdesOBP1 structure was determined in two crystal forms closely related and is likely in a monomeric state. Its structure assembles 5 α-helices linked by three disulphide bridges, one of them exhibiting a different connection as compared to their insect counterparts. Comparison with classical OBPs reveals that the second of the six α-helices is lacking in VdesOBP. Ligand-binding experiments revealed molecules able binding only to specific OBPs with a moderate affinity, suggesting that either unknown compounds have to be identified as optimal ligands, or post-translational modifications present in the native proteins may be essential for modulating protein binding activity, or else these OBPs represent a failed attempt in evolution and are not used by the mites.

Key words: *Varroa destructor*, Odorant-binding protein, Three-dimensional structure; Disulphide bridges; Ligand-binding assays; Evolution.

Introduction

Odorant-binding proteins (OBPs) are carriers of pheromones and odours in vertebrates and insects[1–4]. Insect OBPs, which represent a class of proteins structurally different[5,6] from those of vertebrates[7,8], have been detected so far only in Hexapoda[4,9]. They are polypeptides of 120-140 amino acids, folded in a compact structure generally made of six α-helical domains. A pattern of six conserved cysteines paired in three interlocked disulphide bridges is a key feature of these proteins[10–12]. A single hydrophobic pocket binds small organic volatile compounds, such as pheromones and general odorants, with dissociation constants of the micromolar order[13]. Besides OBPs, other soluble carriers such as chemosensory proteins (CSPs), Nieman-Pick C2 (NPC2) proteins and lipocalins, might be likely involved in chemical communication of arthropods[4,9,14–16]. In particular, NPC2 proteins, are expressed with several members in all arthropods and represented until a couple of years ago the only known class of putative semiochemical carriers in Chelicerata and Crustacea.

However, recent work unveiled the presence of proteins with some sequence similarity to insect OBPs in spiders, ticks and mites, as well as in Myriapoda, and have been tentatively named "OBP-like"[17–21]. Such polypeptides exhibit values of amino acid identity with insect OBPs around 10-15%. However, such low similarity can be significant when compared with comparably low values of identities between the OBPs of different insect Orders. Given their phylogenetic relationship with their insect counterparts, we shall simply refer to these proteins as OBPs. Members of this family have not been found so far in Crustacea nor in other species of living organisms.

In order to understand the role of OBPs in Chelicerata, we have focused our attention on the mite *Varroa destructor*, the well-known honey bee parasite. The genome of this species contains 5 genes encoding members of the OBP family[18,19,22]. The sequences of the five encoded proteins are aligned in Figure 1a. At the mature protein level, VdesOBP1 and VdesOBP2 are very similar to each other with 72% identity. Also VdesOBP3 and VdesOBP4 are closely related sharing 60% of their residues, but the identity levels drop to 15% or less between two groups and with VdesOBP5. Orthologues of these sequences are found also in other mites. Some representative examples are reported in the phylogenetic tree of Figure 1b, where the three clusters (OBP1-2, OBP3-4 and OBP5) are clearly separated. A number of sequences from other ticks and mites fall into different clusters. In particular, all four OBPs of the spider mite *Tetranychus urticae* belong to a group well distinct from the three highlighted for varroa. Besides, the four OBPs of *T. urticae* are much more similar to each

other with identities around 40%. This suggests that different mites may have adopted different sub-classes of OBPs for their chemoreception systems, although we cannot exclude that such differences might be partly due to incomplete annotation. In any case, OBPs are present in each species of Chelicerata with a small number of genes, poorly conserved between species and highly divergent within the same species. Unfortunately, due to only few genomes available and their poor annotation, the picture provided by this tree is to be considered as partial and preliminary. The amino acid sequences used to build the phylogenetic tree are reported in Supplementary Table S1 online.

Transcriptomic studies reported VdesOBP2 (XP_022645714) as the member with the highest expression in the forelegs[18], while the gene encoding VdesOBP3 (XP_022653293) was found to be upregulated in the forelegs, where chemosensory appendages are located, with respect to the second pair of legs [19]. At the proteomic level, four OBPs were detected: VdesOBP1 (XP_022672530), VdesOBP2, VdesOBP3 and VdesOBP4 (XP_022653281), the last two being more highly expressed in the forelegs than in hindlegs[22].

In this work, we have expressed and characterised the OBPs of this mite. We have also solved the crystal structure of VdesOBP1, which exhibits a folding markedly different from those of insect OBPs.

Results

Protein expression and purification

First we confirmed the five OBP sequences reported in the NCBI database by cloning the five genes from RNA extracted from whole individuals. Sequences were identical with the published ones at the amino acid level, while few single base substitutions were likely due to differences between the population of mites used in our study with respect to those used for the sequences stored in the NCBI database.

The proteins were expressed in bacteria, following standard protocols, as reported in the Methods section. We have successfully obtained VdesOBP1, VdesOBP2, VdesOBP4 and VdesOBP5 in medium to good yields. Instead, we were unable to express VdesOBP3 in amounts adequate for purification and consequently for its characterization. The proteins contained a starting methionine as the only addition to the natural sequence, and were purified by two or three chromatographic steps on anion-exchange columns, such as DE-52 and HiPrep-Q. Figure 2 reports a summary of expression and purification for the four OBPs.

Cysteine pairing assessment

VdesOBP1-5 contain six cysteines occurring at conserved positions (Figure 1), although in a pattern different from the one typical of insect OBPs. This fact suggested possible differences in the cysteine pairing with respect to insects[10–12]. To assess this issue, VdesOBP1, VdesOBP2 and VdesOBP5 were subjected to extensive alkylation with iodoacetamide under non-reducing denaturing conditions; after purification, they were then digested with trypsin followed by a second treatment with chymotrypsin. Resulting digests were analyzed by nanoLC-ESI-Q-Orbitrap-MS/MS, and S-S-crosslinked peptides were assigned using dedicated bioinformatic procedures (Supplementary Table S2 online). Examples of mass spectra of disulphide-bridged peptides from VdesOBP1 and VdesOBP5 digests are reported in Figure 3. These results and the absence of carboxyamidomethylated peptides in the protein digests demonstrated that all cysteines present in VdesOBPs are involved in disulphide bridges. Surprisingly, the ascertained, conserved disulphide pattern of VdesOBPs was unlike that of insect OBPs. In fact, a C1-C6, C2-C3 and C4-C5 arrangement was identified in VdesOBPs, which was completely different form the C1-C3, C2-C5, C4-C6 typical of insect OBPs[10–12]. This phenomenon brings two major consequences in the structure of VdesOBPs: i) the pairing of adjacent cysteines should make *V. destructor* proteins more flexible and less compact than insect OBPs; ii) linking C1-C6 brings together the protein N- and C-termini, while these two ends point generally in opposite directions in insect OBPs.

Protein characterization in solution

The protein oligomeric state of VdesOBP1 was characterized in solution at dfferent pH values, namely 4.5, 7.0 and 9.7 using HPLC with multi-angle laser light scattering /UV/ refractometry (SEC-MALLS-RI) coupled detectors (Wyatt, Santa-Barbara, USA). The results point to a monomeric state, with measured masses of 15,708, 16,143 and 15,306 Da (16.496 kDa theoretical mass), respectively (Supplementary Fig. S1 online)

Overall structure of VdesOBP1

The three-dimensional structure of VdesOBP1 was solved by sulphur single-wavelength anomalous dispersion (SSAD) with a wavelength λ=1.512 Å on beamline proxima 2A at SOLEIL synchrotron (Saint-Aubin , France). Data were collected at 1.59 Å resolution on a $P3_221$ crystal form (Table 1). Another crystal forms was collected, with space group $P2_1$ at 1.2 Å resolution, and its structure was determined by molecular replacement using the $P3_221$ crystal form. VdesOBP1 crystal forms $P3_221$ and $P2_1$ 2 were crystallized at pH values of 7.5 and 9.5, respectively. They contain one and two molecules in the asymmetric unit,

respectively. After refinement, the final R/Rfree values were 18.9/23.0 % and 18.3 / 20.5 % for the two crystal forms, respectively (Table 1).

The cloned protein is 147 residues long and possesses 6 cysteine residues (Fig. 4). The chain was built in the electron density map of the $P2_1$ crystal form with the exception of residues 1–6 at the N-terminus while the $P3_221$ crystal form starts at residue 11 an lacks loop 50-53. We checked the integrity of the protein in the crystal and our analysis showed that the protein is intact, confirming that the N-terminal segment is present but not ordered in the crystal. The rmsd value between the structures of the two VdesOBP1 crystal forms is 1.55 Å.

Being the structure best defined, our description will apply to the $P2_1$ form. VdesOBP1 is mostly helical (Fig. 4). In VdesOBP1 five α-helices are observed in contrast with all classical insect OBP/PBPs such as BmorPBP[5], Lush[23] or LmadPBP[24] that assemble six α-helices. The N- and C-terminal stretches are elongated forming a pseudo β-sheet of two strands (Fig. 4). Three disulphide bridges (21/141;SS-N-5, 41/68;SS1-2 and 111/132;SS4-5, Fig. 4a) strengthen the 3D fold of VdesOBP1.

It has been reported that OBPs may form functional dimers[25,26]. However, it cannot to be concluded firmly for or against dimerization of OBPs in their biological context. In the crystal, dimers have been observed in the case of AgamOBP1, but this observation may not apply to solutions, due to an insufficient contact area[27]. In the case of VdesOBP1, we observe a monomer using multi-angle static light scattering (MALS)/dynamic light scattering and RI coupled on-line with a size-exclusion chromatography (SEC) column (MALS/UV/RI/SEC) at three pH values (Supplementary Fig. S1 online). In contrast, the $P2_1$ form contains two molecules in the asymmetric unit (Fig. 4b). A dimer can be reconstituted in the $P3_221$ form, using a symmetrical molecule in the crystal packing (Fig. 5b). However, the second monomer of one form does not display the same orientation relative to that of the second form when the first monomers are superimposed (Fig. 5b,c). Finally, the PISA server reported an ambiguous response for the $P2_1$ form (score=0.49) but discarded the possibility of a dimer in form $P3_221$ (score=0). All in all, it seems highly probable that the biologically relevant form is a monomer.

The binding crevice

The binding crevice, as detected by CavityPlus[28], is large and wide open in the monomer, with dimensions of the floor of ~18x18 Å (Fig. 6a). This large surface implies that numerous

side-chains form the walls of the binding site (Fig. 6b). Although most side chains are from hydrophobic/aromatic residues (16/24) two basic and four acidic residues complete the wall. The active site is empty, except in form $P2_1$ where a buffer molecule, 2-(N-cyclohexylamino)-ethane sulfonic acid (NCES), is docked against a wall of the binding site (Fig. 6a). Examination of the binding site in the dimer reveals that the N-terminus of each monomer protrudes deeply in the binding site of the other monomer, drastically reducing its size. This effect explains the serendipitous binding of NCES as this ligand interacts with the cell wall of one monomer and the C-terminal extension of the other one (Fig. 6c). The sulfonic acid moiety of NCES forms a ionic bond with Arg31A side chain (atom NH_2) and with the main chain N-H of Lys12B.

Comparison of VdesOBP1 with other OBPs/PBPs structures

A search on the PDB with the DALI server reported hits with good statistics, taking into account the Z-score and the r/m.s.d.[29]. The best hit is that of entry 5dic (Z=7; r.m.s.d. 2.9 Å on 96 / 115 residues), the fatty acid binding protein OBP56a from the oral disk of the blowfly *Phormia regina* (PDB ID: 5dic, by Ishida, Y., Leal, W.S., Wilson, D.K., unpublished) with a non-classical OBP fold. This FBP/OBP has been shown to bind the C16 palmitic acid (but not the acetate), as well as the C18 stearic, oleic and linoleic acids, but not decanoic acid[30]. The structure is reported in complex with two molecules of palmitic acid in its wide cavity. Superposition of VdesOBP1 and PregOBP56a shows that VdesOBP1 helices 1, 2, 3, 4 and 5 superimpose well with PregOBP56a helices 1, 3, 4, 5 and 6. Helix 2 of PregOBP56a closes its binding cavity, while a similar effect is not observed in VdesOBP1. This PregOBP56a helix 2 is replaced in VdesOBP1 by a loop between helices 1 and 2, projected out of the protein core (Fig. 7a). Another good fit (1ow4; Z=5.4; r.m.s.d 3.3 Å on 8-90/120 residues) is observed with the PBP of the cockroach *Leucopheae maderae* (LmadPBP) (3X). In this case again, LmadPBP exhibits an extra helix instead of a loop between helices 1 and 2 in VdedOBP1 (Fig. 7b). Two disulphide bonds are conserved between the two proteins, the first bridging helices 1 and 2, the second bridging helices 4 and 5 of VdesOBP1 (Fig. 4a and Fig. 7b), although as result of different (C1-C6, C2-C3, C4-C5 in VdesOBP1 and C1-C3, C2-C5, C4-C6 in LmadPBP) cysteine connectivities. The third disulphide bond that bridges helices 3 and 6 in LmadPBP is not observed in VdesOBP1. Instead, the third disulphide bridge is observed between helix 5 (corresponding to helix 6 in LmadPBP) and the N-terminus extended stretch in VdeOBP1 (Fig. 7b).

Ligand-binding assays

The purified VdesOBP1 did not show any affinity to neither of the most common probes used in ligand-binding assays with OBPs, N- phenyl-1-naphthylamine (1-NPN) and 1-aminoanthracene (1-AMA). Extraction with dichloromethane to remove possible endogenous ligands did not prove effective. We also subjected the protein to a denaturing treatment with 8 M urea and 1 mM DTT, followed by slow renaturation by extensive dialysis, but we still could not observe any binding with neither of the two probes. We also tried 8-phenylamino-naphthalenesulfonic acid (ANS) as well as larger fluorescent reporters[31], such as N-naphthyl-1-naphthylamine (1-NNN) and N-phenyl-1-aminoanthracene (PAA) with no effect.

It is still possible that the fluorescent probe binds the protein, but the event is not accompanied by a change in the emission spectrum. In fact, when adding 1-NPN to the protein, we observed a dose-dependent quenching of the intrinsic tryptophan fluorescence, suggestive of some affinity of this probe for VdesOBP1 (data not shown). In any case, competitive binding using a fluorescent reporter could not be used with VdesOBP1 nor with VdesOBP2. Therefore, we attempted to find ligands for these two proteins monitoring the quenching of intrinsic tryptophan fluorescence. This method, on the other hand, is limited to chemicals able to accept the energy transferred by the tryptophan (therefore structures with a region of conjugated π-electrons), and is not suitable to evaluate binding affinities. Instead, with VdesOBP4 and VdesOBP5 we obtained clear binding curves using 1-NPN as the fluorescent reporter (Figure 8a). Therefore, we performed competitive binding experiments only with these two OBPs. Figures 8c and 8d report some of the displacement curves obtained with VdesOBP4 and VdesOBP5, respectively. The structures of the ligands used in these experiments are shown in Figure 8b. We can notice that only coniferyl aldehyde is able to decrease the 1-NPN fluorescence to 50% at the maximum concentration used, qualifying for a modest-poor ligand. None of the other chemicals used in the experiments (listed in Supplementary Table S3 online) with the four proteins proved to be a better ligand. Representative curves relative to quenching of the intrinsic fluorescence of tryptophan in VdesOBP1 and VdesOBP2 are reported in Supplementary Fig. S2 online.

Summarising these data, we can first observe that some of the tested chemicals proved to be ligands for VdesOBP4 and VdesOBP5 in 1-NPN displacement assays, while they determined significant quenching of intrinsic tryptophan fluorescence in VdesOBP1 and VdesOBP2. The same few compounds (coniferyl aldehyde, safranal and vanillin) were the best ligands with all four proteins, although with poor dissociation constants between 10 and 20 μM. These ligands are associated with plants, in particular with leaf damage, and do not bear any relationship to the varroa habits. In particular none of the honey bee pheromones,

such as 9-ketodecanoic acid, homovanillyl alcohol or fatty acid esters were able to bind any of the tested proteins.

Discussion

This work reports the original structural and functional characterization of four out of the five OBPs encoded in the genome of the mite *Varroa destructor*, a major pest for honey-bee. Mass spectrometric analysis demonstrated that the six cysteines VdesOBPs, whose sequence position do not match the pattern in insect OBPs, are paired into three disulphide bridges, (C1-C6, C2-C3 and C4-C5) different from that (C1-C3, C2-C5, C4-C6) of their insect counterparts. The overall folding, determined for VdesOBP1, is also novel and markedly different from the conserved structure of insect OBPs, although some similar aspects can be recognised.

VdesOBP1 structure was determined in two crystal forms closely related. Its structure assembles 5 α-helices linked by three disulphide bridges. Comparison with classical OBPs reveals that the second of the six α-helices found in them is lacking in VdesOBP. Although dimer formation of VdesOBP1 is observed in the crystal, it most probably does not represent the functional form *in vivo*. This dimerization is probably driven by the presence of a wide and fully open cavity, able to accommodate the extended N- and C-terminal stretches of the opposite monomer. The large cavity opening can be ascribed to the absence of helix 2 in OBPs classical fold, that restrains and closes the internal binding site. In VdesOBP1, this helix is absent and replaced by a loop projected outside the protein core. Two disulphide bridges are conserved as compared to classical OBPs, although associated with different sequential cysteine connectivities. These observations suggest that VdesOBP1 possess an original fold remotely related to classical OBPs. Comparison with PregOBP56a suggests that two to three fatty acids could bind inside the VdesOBP1 cavity. Such an arrangement of three ligands was observed in *Mamestra brassicae* CSP (CSPMbraA6)[30].

The unique structure of VdesOBP1 and the lack of strong ligands pose questions that, at this stage, can only be addressed by formulating hypotheses and models to be verified by further experimental work. A first idea, based on the limitedbinding activity of these proteins and on the small number of genes (4-5) expressed in *V. destructor*, as well as in other mites and ticks, might suggest that these genes represent by-products of evolution, which are not used any more. In fact, at least two more families of carrier proteins could be active in Chelicerata:

NPC2 proteins and lipocalins, apart from a third family (CCPs) that so far seems to be specific of spiders[21]. NPCs and lipocalins are represented by with 6 and 9 members, respectively in varroa, while in other Chelicerata often 10-20 members of NPC2s have been identified, with exceptions of nearly 50 genes, as in the spider mite *Tetranychus urticae*. Although making sense for the reason reported above, the idea that OBPs in Chelicerata might represent a dead end in evolution and a failed attempt to make efficient carrier proteins contrasts with proteomic findings showing significant expression of such proteins [22]. In fact, four (OBP1-OBP4) of the five VdesOBPs were identified in different parts of *V. destructor*, with one of them (OBP4) particularly abundant in sensory organs, namely the first pair of legs and the mouth parts.

If then the expression of OBPs at the protein level indicate that they are functional, then, along with a second hypothesis, their poor affinity might be related to the absence of post-translational modifications. It has been recently shown that the pig OBP1 can bind fatty acids or steroids depending on the phosphorylation and glycosylation status of the protein[32]. Our varroa OBPs, being expressed in bacteria, are devoid of any post-translational modification. On the other hand, we do not have any information on possible phosphorylation and/or glycosylation of these proteins in their native forms. This is certainly an interesting aspect, although not easy to investigate, not only for the tiny size of sensory organs, but also because phosphorylation in particular may be present only in certain time windows related to the physiology and the behaviour of the mite. Indeed, protein phosphorylation is highly related to signal transduction events regulating essential physiological process. Thus, we searched for putative phosphorylation sites in the five OBP sequences, and found a very large number of serine and threonine residues that could be potentially phosphorylated (Supplementary Fig. S3 online). In addition we found few tyrosine residues as potential phosphorylation sites (2 in OBP1, 1 in OBP2 and OBP3, 4 in OBP5).

In the three-dimensional structure of VdesOBP1 no serine or threonine are found inside the binding pocket. However, phosphorylation of serine and threonine residues outside the binding pocket (as well as glycosylation of the protein) can affect the folding of the protein and influence its binding affinities. Further speculation would be inappropriate based on the available experimental data. Certainly, a structural study of the native proteins, regarding their post-translational modifications would be essential before proposing alternative modes of action.

Materials and Methods

Biological material

V. destructor mites were collected from the cells of an infected honeybee hive and used immediately for RNA extraction.

Reagents

Chemical compounds for buffers and solutions, as well as ligands for binding assays were of analytical grade from Merck, Austria, and. Ligands were dissolved in spectroscopic grade (Uvasol) methanol. Oligonucleotides and synthetic genes were custom synthesised at Eurofins Genomics, Germany. Enzymes and kits for DNA extraction and purification were from New England Biolabs, USA.

RNA extraction and gene cloning

Total RNA was extracted using Tri-reagent (Sigma-Aldrich) along with the protocol provided. cDNA was prepared from 2 μL of total RNA using the kit qScript cDNA SuperMix (Quanta Bio) and the manufacturer's procedure. For sequence verification, PCR was performed on cDNA using specific primers at the 5' end and at the 3' end. The primers, which were also used for cloning into the expression vector, also contained the restriction sites NdeI and EcoRI at the 5' and the 3' ends, respectively. After digestion, the PCR products were inserted into pET30 vector previously linearized with the same enzymes. This procedure was adopted for all the genes of *V. destructor* except for *V. destructor* OBP1 (used for structural studies) that was produced from a synthetic gene (Eurofins Genomics), whose nucleotide sequence was optimized for *E. coli* expression.

Gene cloning

Synthetic genes encoding VdesOBP1 and VdesOBP2 were custom synthesised at Eurofins Genomics, Germany, and optimised for *E. coli* expression. The sequences encoding VdesOBP4 and VdesOBP5, instead, were amplified from cDNA, digested and cloned into the pET-30 expression vector (Novagen, Darmstadt, Germany). The constructs encoded only the mature protein sequences preceded by an additional methionine, which is not considered here for amino acid numbering. Expression was performed in *E. coli* BL-21 (DE3) competent cells. The sequences of the primers used are reported in Supplementary Table S4 online.

Protein expression and purification

Proteins synthesis was induced by 0.4 mM IPTG of the *E.coli* culture, grown for two additional hours at 37 °C, then harvested by centrifugation, sonicated and centrifuged at high speed. The Vdes OBPs were obtained in good yield in the supernatant, except for

VdesOBP4, which was produced in very poor yield and was present in the pellet, from which it was solubilized with 8 M urea and 1 mM DTT. The protein was refolded by dialysis for three against 50 mM Tris-HCl, pH 7.4. All OBPs were purified 2-3 chromatographic steps on anion-exchange columns DE-52 and HiPrep-Q 16/10, 20 mL (Bio-Rad).

Protein alkylation and digestion, and mass spectrometry analysis

Samples of *V. destructor* OBP1, OBP2 and OBP5 (20-50 μg) dissolved in 0.1 M tetraethylammonium bicarbonate (TEAB), pH 6.5, containing 4 M guanidinium chloride, were treated with iodoacetamide (0.5 M final concentration) for 30 min, in the dark; then, proteins were insolubilized by addition of 6 vol of cold acetone, leaving the samples at -20 °C, overnight. After centrifugation at 12,000 rpm at 4° C, for 20 min, supernatants were removed, and recovered pellets were dried with a SpeedVac device (ThermoFisher Scientific, USA). Recovered proteins were dissolved in 0.05 M TEAB, pH 6.5 (2 μg/μL final concentration), treated with trypsin (1:10 w/w enzyme/substrate) for 16 h, at 37 °C, then with chymotrypsin (1:8 w/w enzyme/substrate) for 16 h, at 37 °C. Protein digests were desalted with ZipTip® C18 (Millipore, USA) and directly analyzed with a UltiMate 3000 HPLC RSLC nano-chromatographer (ThermoFisher Scientific) linked on-line through a nano-Spray ion source (Thermo Fisher Scientific) to a Q-ExactivePlus mass spectrometer (Thermo Fisher Scientific) [33]. Peptides were separated on an Acclaim PepMap RSLC C18 column (150 mm × 75 μm ID; 2 μm particle size; 100 Å pore size) (Thermo Fisher Scientific), at a flow rate of 300 nL/min, using a gradient of solvent B (19.92/80/0.08 v/v/v water/acetonitrile/formic acid) in solvent A (99.9/0.1 v/v water/formic acid). Solvent B started at 3%, increased linearly to 40% in 45 min, then raised to 80% in 5 min, where it remained for additional 4 min, before rapidly returning to 3%. Mass spectrometer worked in positive polarity using a data-dependent mode, performing a full MS1 scan in the range *m/z* 345-1350, at a nominal resolution of 70,000, followed by MS/MS scans of the 10 most abundant ions in high energy collisional dissociation mode[34]. MS/MS spectra were acquired in a dynamic *m/z* range, with a nominal resolution of 17,500, a normalized collision energy of 28%, an automatic gain control target of 50,000, a maximum ion injection time of 110 ms, an isolation window of 1.2 *m/z* and a dynamic exclusion setting of 20 s.

Bioinformatics for peptide identification

Firstly, raw mass data files were analyzed with Proteome Discoverer v. 2.4 package (Thermo Fisher Scientific), running by Mascot v. 2.6.1 (Matrix Science, UK) and ByonicTM v. 2.6 (Protein Metrics, USA) software. Database searching was performed against a database containing the sequences of *V. destructor* OBP1, OBP2 and OBP5, trypsin, chymotrypsin and common protein contaminants. Parameters for database searching were variable

oxidation at Met, deamidation at Asn/Gln, pyroglutamate formation at Gln and carbamidomethylation at Cys. Mass tolerance setting included ± 10 ppm for precursor ions and ± 0.05 Da for fragmentation ions[35]. Proteolytic enzyme and maximum number of missed cleavages were set to trypsin, chymotrypsin and 5, respectively. Proteome Discoverer peptide candidates were considered confidently identified only when the following criteria were satisfied: i) protein and peptide false discovery rate (FDR) confidence: high; ii) peptide Mascot score: > 30; iii) peptide spectrum matches (PSMs): unambiguous; iv) peptide rank (rank of the peptide match): 1; v) Delta CN (normalized score difference between the selected PSM and the highest-scoring PSM for that spectrum): 0. Byonic peptide candidates were considered confidently identified only when the following criteria were satisfied: i) PEP 2D and PEP 1D: $<10 \times 10^{-5}$; ii) FDR: 0; iii) q-value 2D and q-value 1D: $<10 \times 10^{-5}$. Finally, disulphide bridge assignment was obtained using dedicated BioPharma Finder v. 4.0 (Thermo Fisher Scientific) and pLink v. 2.3.9 software[36]. Both programs were used for database searching, enabling the specific function of disulphide-linked peptides attribution, and applying the additional settings reported above for Proteome Discoverer and Byonic examinations. Disulphide-bridged peptide identification was considered reliable when BioPharma Finder and pLink results showed a "confidence score" > 95 and/or an "E-value" < 1.0^{-10}, respectively. Manual interpretation and verification of the candidate spectra were always performed.

Crystallization, data collection and structure determination of the $P3_221$ form

The protein was concentrated at 20 mg/ml. Screening experiments were performed using several commercial kits. The sitting-drop method in Swissci plates was used. The reservoirs of the Swcissi Plates were filled using a Genesis II (TECAN) pipetting robot. The nanodrops were dispensed by a MOSQUITO Crystal (TTP Labtech). Respectively 300nl, 200 nl and 100nl of protein were dispensed onto the central of three-sitting-drop shelves. 100nl of the reservoir solution were added to the protein drops. The plates were sealed with a transparent film and stored in a R 1000 crystallization hotel (Formulatrix). Small crystals were obtained in 0.2M Cadmium Chloride, 2.2 ammonium sulfate. Those crystals were used for micro-seeding. The sitting-drop method in Greiner plates was used and new crystals were obtained by doing a re-screen in the commercial screen PEGII with the initial condition in which the first crystals were obtained. The droplet with the thicker crystals contained 2/3 of the initial condition (0.2M Cadmium Chloride, 2.2 ammonium sulfate) + 1/3 of PEG II screen (0.2M calcium chloride, HEPES 0.1M pH 7,5; PEG 4000 30% w/v).
The Crystal was cryo-cooled at 100K within this crystallization liquor with the addition of 10% glycerol as cryoprotectant. Structure determination used SAD-phasing method at remote wavelengths with sulphur atom anomalous dispersion on native crystals. A strong theoretical

Bijvoet ratio of 1.12% was expected from the 10 sulphur atoms of the protein that contribute to the anomalous signal at wavelength of 1.77 Å. The sulphur-SAD data set was collected on the PROXIMA-2 beamline (SOLEIL) equipped with a high performance micro-diffractometer with a mini-kappa (MD2), an X-ray fluorescence detector (KETEK), and a fast, low-noise, photon-counting, area detector EIGER X 9M (238 fps in 9M mode, 750 fps in 4M mode). The wavelength was optimized to 1.77 Å and a redundant data set was collected at 1.81 Å resolution. including 3600 images of 0.1° rotation. The data set was processed with XDS[37]. The data sets were indexed and integrated with Pointless and scaled with CCP4 SCALA[38]. VdesOBP1 crystallized in space group $P3_221$ with cell dimensions a= 53.84 Å, b = 53.84 Å, and c = 91.74 Å and and γ= 120with one molecule in the asymmetric unit, corresponding to a solvent content of 50.87%. Six sulfur atom sites were found by using PHENIX Autosol[39]. The results from PHENIX include a PDB containing the anomalous scatterers, an MTZ file containing the experimentally determined phases, and the optimized phases from density modification. Structure refinement was performed with REFMAC5[40] and AutoBUSTER[41] alternated with rebuilding using COOT[42].

Crystallization data collection and structure determination of the $P2_1$ form

The protein at 20 mg/ml was was subjected to screening experiments were performed using several commercial kits. VdesOBP1 was set in 2/3 of the initial crystallization conditions (CHES pH9.5, 0.1M, 30% PEG3000 w/v) and 1/3 of ammonium sulfate screen (0.2 M sodium nitrate, 2.2M ammonium sulfate). Crystals of VdesOBP1 were briefly soaked in crystallization solution supplemented with 10 (v/v) polyethylenglycol. Diffraction data were collected at 1.2 Å resolution on beamline Proxima-1 at SOLEIL, Paris, France. The data sets were integrated with XDS[37] and were scaled with CCP4 SCALA[38] . VdesOBP1 crystallized in space group $P2_1$ with cell dimensions a= 31.6Å, b = 112.96 Å, and c = 41.82 Å and and β= 98.95 with two molecules in the asymmetric unit, corresponding to a solvent content of 44.48%. The structure was solved by molecular replacement with MOLREP[43] using the structure of VdesOBP1 form $P3_221$ as starting model. Refinement was performed with REFMAC5[40] and AutoBUSTER[41] and the structures were corrected with COOT[42].

Multi-Angle Light Scattering

100 μL of purified VdesOBP1 were analyzed by SEC-MALS using a HPLC system (Thermo Scientific) equipped with a Superdex 75 increase coupled with a multi-angle light scattering device (Dawn 8, Wyatt Technology) and an Optilab TrEX differential refractive index detector (Optilab TrEX, Wyatt Technology). At first, the column was first equilibrated with a buffer containing 0.02 % w/v NaN_3 for several hours, until the base line of the Optilab TrEX signal was stable. Then, light scattering and concentration data were analysed in ASTRA to

determine molecular weight and size distribution. Inter-detector band broadening and alignment were applied using the algorithms in ASTRA 7 (Wyatt Technology, USA) using BSA protein standard according to manufacturer's recommended methods. Three different buffers were used to understand the behaviour of the protein at different pH values, namely 50 mM CHES pH 10, 50 mM Hepes pH 7 and 50 mM sodium acetate pH.5). The temperature controlled-autosampler was kept at 10 °C.

Ligand-binding assays

Direct binding of fluorescent probes was monitored by adding aliquots of 1 mM methanol solution of the probe to a 2 μM solution of the protein in 50 mM Tris-HCl, pH 7.4 to final concentrations of 2 to 16 μM. The excitation wavelength was 337 nm and intensities were measured on the maximum of the peak, usually between around 410 and 420 nm. For VdesOBP1 and VdesOBP2 binding of ligands was evaluated by monitoring the intrinsic tryptophan quenching of the protein. Excitation wavelength was 295 nm and intensities were recorded around 335-340 nm. For VdesOBP4 and VdesOBP5 competitive binding experiments were titrating a solution of protein and 1-NPN both at 2 μM in Tris-HCl buffer, pH 7.4 with 1 mM methanolic solutions of ligands to final concentrations of 2-16 μM. All measurements were perfomed with a FL 6500 spectrofluorometer (PerkinElmer); slits were set at 5 nm for both excitation and emission, and 1 cm path quartz cuvettes were used. Dissociation constants for 1-NPN were evaluated using Prism software. Affinities of other ligands were calculated from the corresponding $[IC]_{50}$ values (the concentration of each ligand halving the initial value of fluorescence), from the equation:

$$K_d = [IC]_{50}/1 + [1\text{-NPN}]/K_{NPN}$$

where [1-NPN] is the concentration of free 1-NPN and K_{NPN} is the dissociation constant of the complex OBP/1-NPN.

Author contributions: J.Z., P.P. and C.C. designed research; J.Z., B.A., S.A, P.L., G.R. and A.G. performed research; P.P., C.C., A.R., J.Z., B.A., S.A, G.R., A.S. H.P. and W.K. analysed data and discussed results; P.P., A.S., A.R. and W.K. contributed all reagents, analytical tools and laboratory equipment; C.C., P.P. and A.S. wrote the paper. All authors read, revised and approved the final manuscript.

The authors declare no conflict of interest.

Data deposition: The three-dimensional structure of VdesOBP1 reported in this paper has been deposited in the database with PDB ID 7NYJ and 7NZA

References

1. Pelosi, P., Baldaccini, N. E. & Pisanelli, A. M. Identification of a specific olfactory receptor for 2-isobutyl-3-methoxypyrazine. *Biochemical Journal* **201**, 245–248 (1982).
2. Vogt, R. G. & Riddiford, L. M. Pheromone binding and inactivation by moth antennae. *Nature* **293**, 161–163 (1981).
3. Leal, W. S. Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annual Review of Entomology* **58**, 373–391 (2013).
4. Pelosi, P., Iovinella, I., Zhu, J., Wang, G. & Dani, F. R. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects: Soluble olfactory proteins in insects. *Biol Rev* **93**, 184–200 (2018).
5. Sandler, B. H., Nikonova, L., Leal, W. S. & Clardy, J. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein–bombykol complex. *Chemistry & Biology* **7**, 143–151 (2000).
6. Tegoni, M., Campanacci, V. & Cambillau, C. Structural aspects of sexual attraction and chemical communication in insects. *Trends in Biochemical Sciences* **29**, 257–264 (2004).
7. Bianchet, M. A. *et al.* The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition. *Nature structural biology* **3**, 934 (1996).
8. Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. & Cambillau, C. Domain swapping creates a third putative combining site in bovine odorant binding protein dimer. *Nature Structural Biology* **3**, 863–867 (1996).
9. Pelosi, P., Iovinella, I., Felicioli, A. & Dani, F. R. Soluble proteins of chemical communication: an overview across arthropods. *Front. Physiol.* **5**, (2014).
10. Scaloni, A. Analysis of post-translational modifications in soluble proteins involved in chemical communication from mammals and insects. in *Methods in Enzymology* vol. 642 103–124 (2020).
11. Scaloni, A., Monti, M., Angeli, S. & Pelosi, P. Structural Analysis and Disulfide-Bridge Pairing of Two Odorant-Binding Proteins from Bombyx mori. *Biochemical and Biophysical Research Communications* **266**, 386–391 (1999).
12. Leal, W. S., Nikonova, L. & Peng, G. Disulfide structure of the pheromone binding protein from the silkworm moth, Bombyx mori. *FEBS Letters* **464**, 85–90 (1999).
13. Pelosi, P., Zhu, J. & Knoll, W. From Gas Sensors to Biomimetic Artificial Noses. *Chemosensors* **6**, 32 (2018).

14. Angeli, S. *et al.* Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from Schistocerca gregaria. *Eur J Biochem* **262**, 745–754 (1999).
15. Zhu, J. *et al.* Niemann-Pick C2 Proteins: A New Function for an Old Family. *Front. Physiol.* **9**, 52 (2018).
16. Ishida, Y. *et al.* Niemann–Pick type C2 protein mediating chemical communication in the worker ant. *PNAS* **111**, 3847–3852 (2014).
17. Renthal, R. *et al.* The chemosensory appendage proteome of Amblyomma americanum (Acari: Ixodidae) reveals putative odorant-binding and other chemoreception-related proteins. *Insect Science* **24**, 730–742 (2017).
18. Eliash, N. *et al.* Chemosensing of honeybee parasite, Varroa destructor : Transcriptomic analysis. *Scientific Reports* **7**, 1–11 (2017).
19. Eliash, N. *et al.* Varroa chemosensory proteins: some are conserved across Arthropoda but others are arachnid specific. *Insect Molecular Biology* **28**, 321–341 (2019).
20. Vizueta, J. *et al.* Evolution of Chemosensory Gene Families in Arthropods: Insight from the First Inclusive Comparative Transcriptome Analysis across Spider Appendages. *Genome Biol Evol* **9**, 178–196 (2017).
21. Vizueta, J., Rozas, J. & Sánchez-Gracia, A. Comparative Genomics Reveals Thousands of Novel Chemosensory Genes and Massive Changes in Chemoreceptor Repertories across Chelicerates. *Genome Biol Evol* **10**, 1221–1236 (2018).
22. Iovinella, I. *et al.* Proteomic analysis of chemosensory organs in the honey bee parasite Varroa destructor: A comprehensive examination of the potential carriers for semiochemicals. *Journal of Proteomics* **181**, 131–141 (2018).
23. Kruse, S. W., Zhao, R., Smith, D. P. & Jones, D. N. M. Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster. *Nature Structural & Molecular Biology* **10**, 694–700 (2003).
24. Lartigue, A. *et al.* The Crystal Structure of a Cockroach Pheromone-binding Protein Suggests a New Ligand Binding and Release Mechanism. *J. Biol. Chem.* **278**, 30213–30218 (2003).
25. Qiao, H. *et al.* Cooperative interactions between odorant-binding proteins of Anopheles gambiae. *Cell. Mol. Life Sci.* **68**, 1799–1813 (2011).
26. Biessmann, H. *et al.* The Anopheles gambiae Odorant Binding Protein 1 (AgamOBP1) Mediates Indole Recognition in the Antennae of Female Mosquitoes. *PLOS ONE* **5**, e9471 (2010).
27. Lartigue, A. *et al.* Sulfur Single-wavelength Anomalous Diffraction Crystal Structure of a Pheromone-Binding Protein from the Honeybee Apis mellifera L*. *Journal of Biological Chemistry* **279**, 4459–4464 (2004).

606 28. Xu, Y. *et al.* CavityPlus: a web server for protein cavity detection with pharmacophore
607 modelling, allosteric site identification and covalent ligand binding ability prediction.
608 *Nucleic Acids Research* **46**, W374–W379 (2018).
609 29. DALI and the persistence of protein shape - Holm - 2020 - Protein Science - Wiley
610 Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/pro.3749.
611 30. Campanacci, V. *et al.* Moth chemosensory protein exhibits drastic conformational
612 changes and cooperativity on ligand binding. *PNAS* **100**, 5069–5074 (2003).
613 31. Mastrogiacomo, R., Iovinella, I. & Napolitano, E. New fluorescent probes for ligand-
614 binding assays of odorant-binding proteins. *Biochemical and Biophysical Research*
615 *Communications* **446**, 137–142 (2014).
616 32. Nagnan-Le Meillour, P. *et al.* Binding Specificity of Native Odorant-Binding Protein
617 Isoforms Is Driven by Phosphorylation and O-N-Acetylglucosaminylation in the Pig Sus
618 scrofa. *Front. Endocrinol.* **9**, (2019).
619 33. Zhu, J. *et al.* Reverse chemical ecology: Olfactory proteins from the giant panda and
620 their interactions with putative pheromones and bamboo volatiles. *Proc Natl Acad Sci*
621 *USA* **114**, E9802–E9810 (2017).
622 34. Arena, S., Renzone, G. & Scaloni, A. A multi-approach peptidomic analysis of hen egg
623 white reveals novel putative bioactive molecules. *Journal of Proteomics* **215**, 103646
624 (2020).
625 35. Arena, S. & Scaloni, A. An Extensive Description of the Peptidomic Repertoire of the
626 Hen Egg Yolk Plasma. *J. Agric. Food Chem.* **66**, 3239–3255 (2018).
627 36. Chen, Z.-L. *et al.* A high-speed search engine pLink 2 with systematic evaluation for
628 proteome-scale identification of cross-linked peptides. *Nature Communications* **10**, 3404
629 (2019).
630 37. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, (2010).
631 38. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta*
632 *Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).
633 39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for
634 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
635 (2010).
636 40. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal
637 structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
638 41. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum
639 likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* **60**, 2210-2221 (2004).
640 42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of
641 Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

642 43. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D*
643 *Biol Crystallogr* **66**, 22-25 (2010).
644
645
646
647
648

Table 1: Data collection and refinement statistics.

Data collection		
Beamline	SOLEIL Proxima2	SOLEIL Proxima1
PDB ID	7NYJ	7NZA
Space group	$P3_221$	$P2_1$
Cell parameters, a, b, c (Å),	53.8, 53.8, 91.7 γ =120°	31.6, 113.0, 41.8 β=98.95°
Wavelength (Å)	1.770	0.9786
Resolution (Å)	46.6-1.81 (1.95-1.81)	40.09-1.20 (1.27-1.20)
Rmerge[a] (%)	7.9 (107)	6.5 (88.5)
Mean((I)/sd(I))[a]	13.26 (1.87)	13.15 (1.19)
cc(1/2)	99.9 (74.2)	99.8 (78.8)
No. of observations[a]	284352 (57758)	634776 (86086)
No. of unique reflections[a]	27234 (5620)	87891 (13358)
Completeness[a] (%)	100.0 (100.0)	97.2 (91.4)
Multiplicity[a]	10.4 (10.3)	7.0 (6.4)
Refinement		
Resolution (Å)	46.59-1.81 (1.857-1.810)	38.8-1.20 (1.23-1.20)
No. of reflections[a]	13827 (1009)	83470 (5555)
atoms:protein/ligands/solvent	2289/36/393	2325 / 36 / 340
No. of test set reflections	728 (53)	4402 (308)
Rwork / Rfree[a] (%)	18.6 (28.6) / 22.6 (31.0)	18.0 (33.0) / 20.1(34.1)
rmsd bonds (Å) / angles (°)	0.015 / 1.83	0.018/ 2.20
B-factors ($Å^2$)	40.5	19.0
Ramachandran: favored / allowed (%)	98.5 / 1.5	96.0/ 4.0

[a] Refers to the highest-resolution bin

Figures legends

Figure 1. (a) Alignment of the five OBPs of *V. destructor*. The six conserved cysteines are highlighted. OBP1 and OBP2 are very similar (72% identity), as OBP3 and OBP4 (60%), but proteins of the two groups share only about 15% of their residues between them as well as with OBP5. **(b) Phylogenetic tree of OBPs on ticks and mites.** Vdes: *Varroa destructor*; Turt: *Tetranychus urticae*; Hqin: *Haemaphysalis qinghaiensis*; Hfla: *Haemaphysalis flava*; Hlon: *Haemaphysalis longicornis*; Isca: *Ixodes scapularis*; Gocc: *Galendromus occidentalis*; Tmer: *Tropilaelaps mercedesae*; Rann: *Rhipicephalus annulatus*; Rsan: *Rhipicephalus sanguineus*; Ldel: *Leptotrombidium delicense*; Dtin: *Dinothrombium tinctorium*; Dpte: *Dermatophagoides pteronyssinus*; Ssca: *Sarcoptes scabiei*; Rmay: *Euroglyphus maynei*. All sequences used to build the tree are reported in Supplementary Table S1 online.

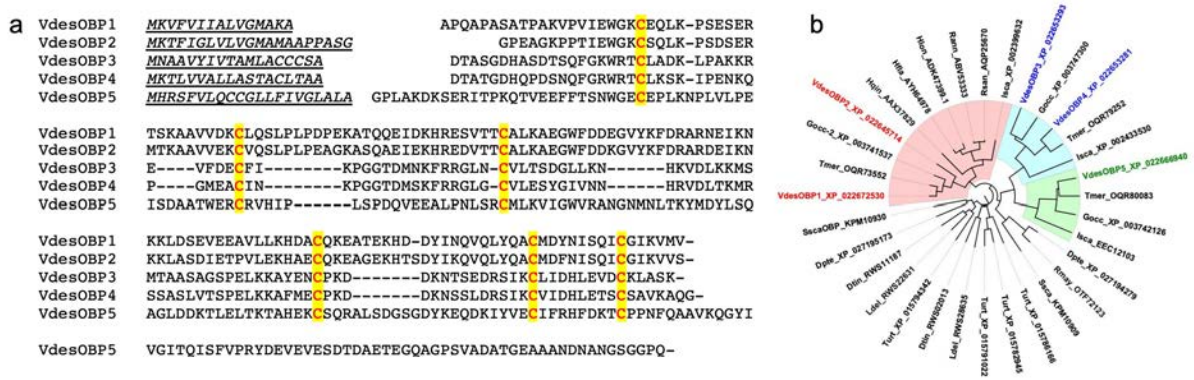


Figure 2. Expression and purification of four OBPs of V. *destructor*. While OBP1, OBP2 and OBP5 were easily produced in bacteria, OBP3 and OBP4, which are structurally related, proved difficult to express. In fact, OBP4 was obtained only in very low yield, while we could not observe expression of OBP3 in any of the different constructs adopted. M: Molecular weigth markers; bef: crude sample before culture induction; aft: crude sample after induction with IPTG; SN: supernatant; P: pellet; OBP: sample of purified protein.

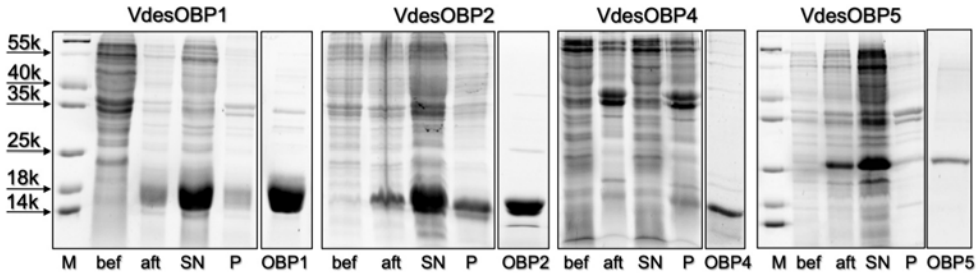


Figure 3. Exemplificative fragmentation spectra of disulfide-bridged peptides identified in the tryptic-chymotryptic digest of *V. destructor* OBP1 and OBP5 as revealed by nanoLC-ESI-Q-Orbitrap-MS/MS analysis. Panels A, C and E report peptides from OBP1 digest, while panels B, D and F report peptides from OBP5 digest. The fragments are highlighted in different colour depending on peptide present in S-S-linked species, and corresponding b and y ion series. Complete data on disulfide-bridged peptides are reported in Table S1.

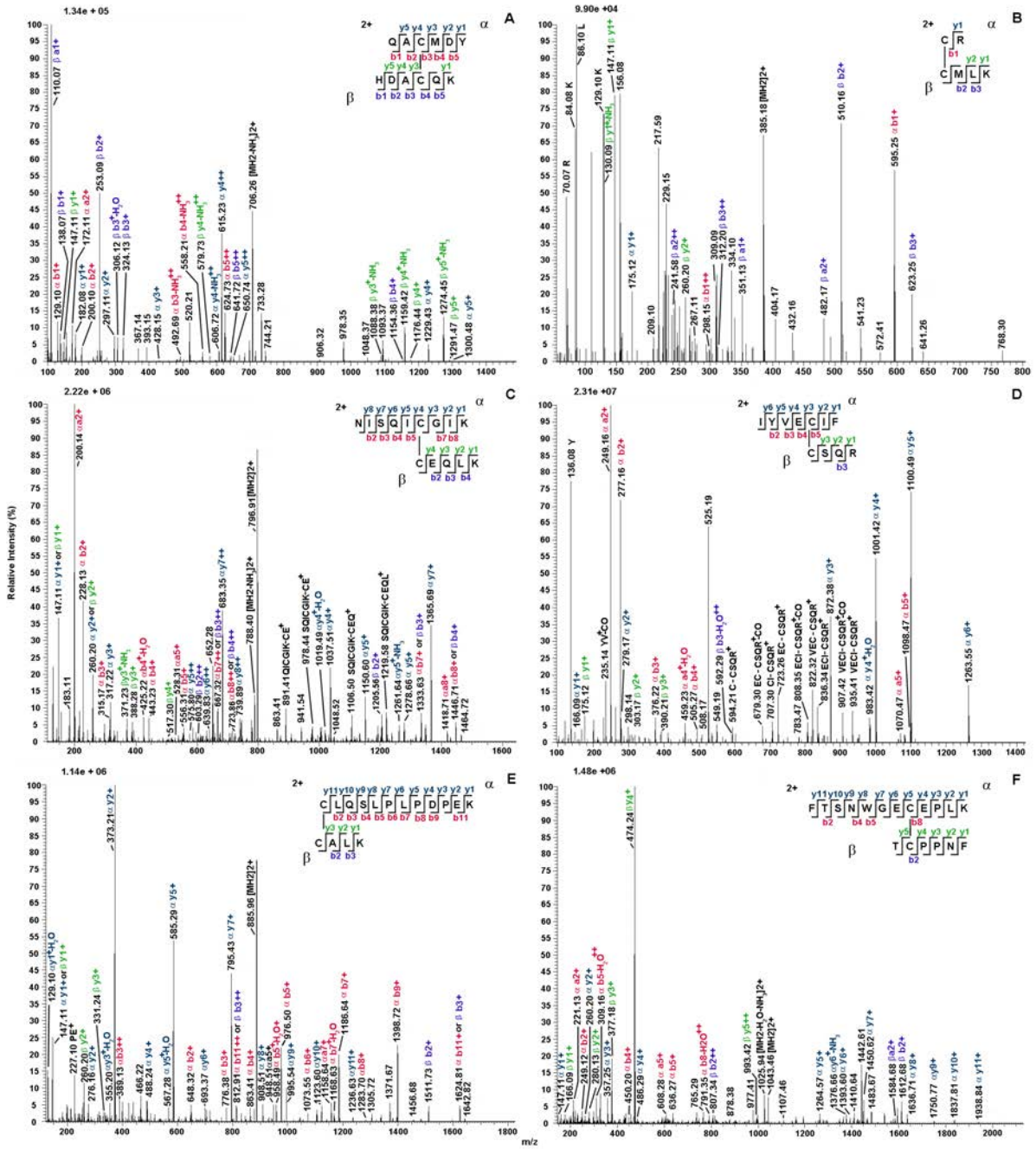


Figure 4. Crystal structure of VdesOBP1 form $P2_1$. (a) Ribbon view of a monomer displaying helices 1-5 and disulphide bridges. The amino-acid chain is rainbow coloured,

from blue (N-terminus) to red (C-terminus). The disulphide are identified by the secondary structures they belong to. **(b)** Ribbon view of the dimer. One monomer is rainbow coloured, the other one is coloured brown. The serendipitously bound buffer molecule 2-(N-cyclohexylamino)-ethane sulfonic acid (NCES) is displayed as atomic spheres (C: white, N: blue and O: red).

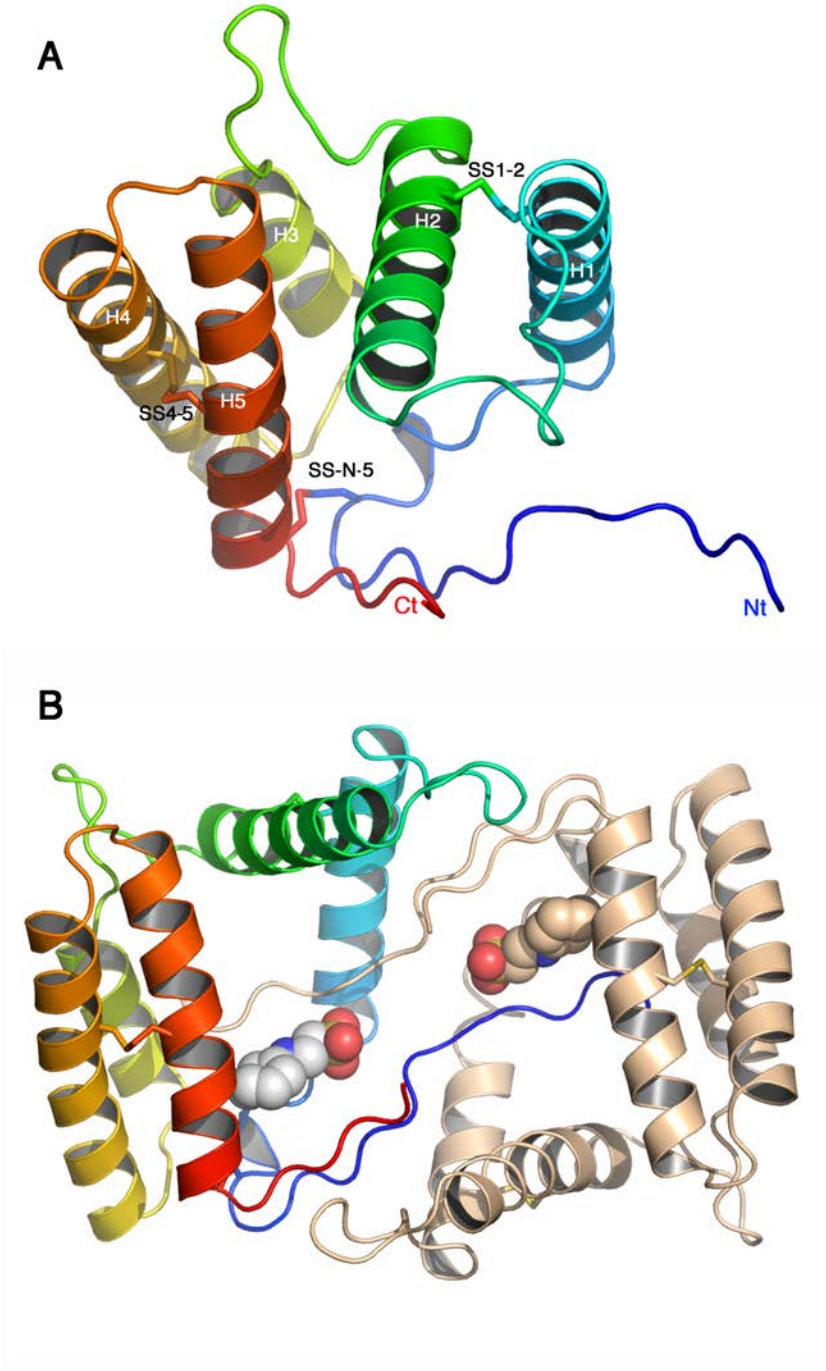


Figure 5. Comparison of crystal structures of VdesOBP1 forms $P2_1$ and $P3_221$. (a) Ribbon view of the superposition of monomers belonging to forms $P2_1$ (rainbow coloured) and $P3_221$ (grey). **(b)** Ribbon view of the dimers of forms $P2_1$ and $P3_221$. In form $P2_1$ one

monomer is rainbow coloured, the other one is coloured brown. The dimer of $P3_221$ form is coloured light grey. One monomer of each form has been superimposed (left). The second monomers do not superimpose. B/same view rotated 90° around a vertical axis. Note the rotation of ~80° of the second monomer of form $P3_221$ relative to the second monomer of form $P2_1$.

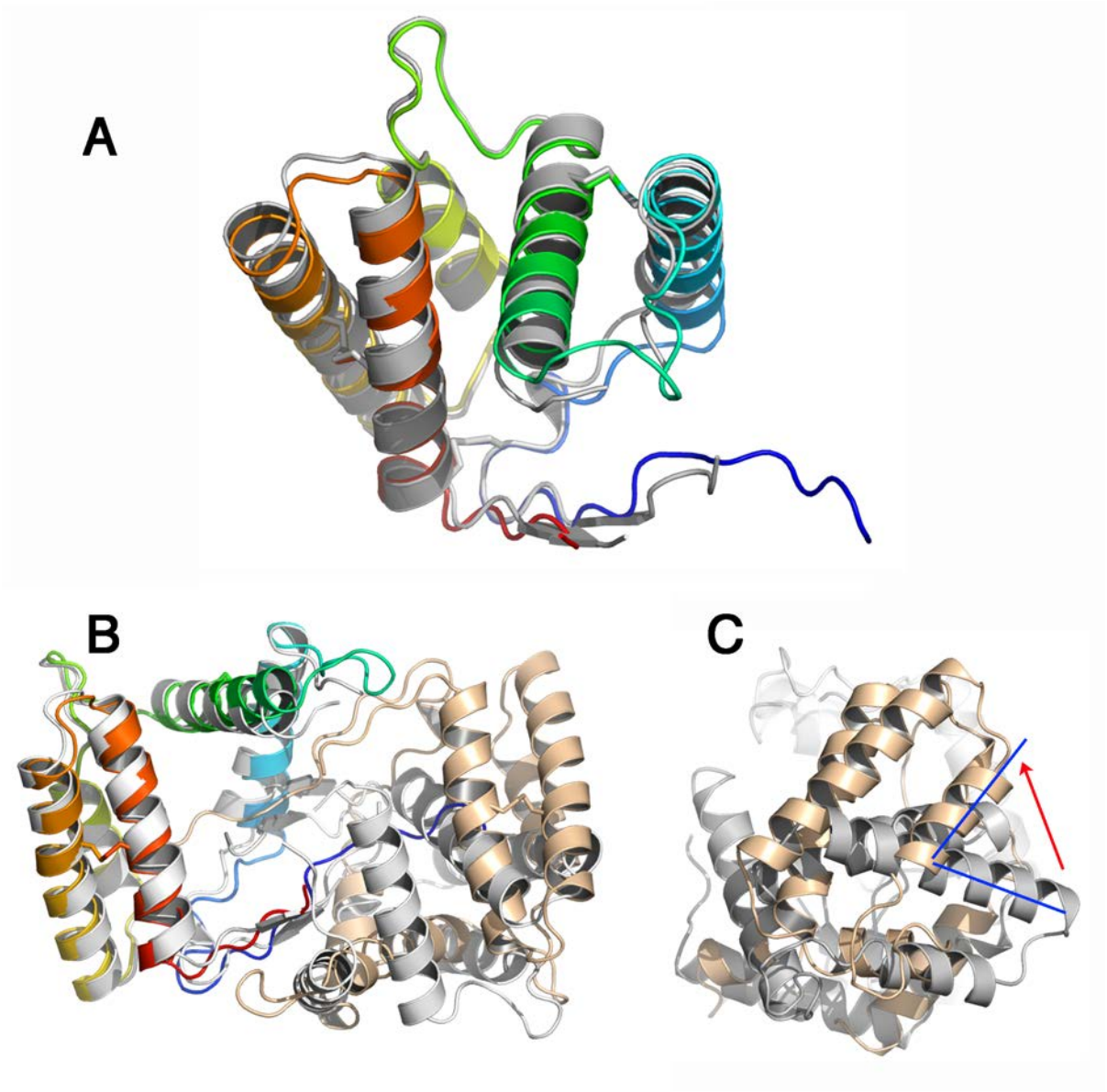


Figure 6. The internal cavity of VdesOBP1 form $P2_1$. (a) Ribbon view of a monomer rainbow coloured. The cavity is represented by a grey surface and is wide open. The serendipitously bound buffer molecule 2-(N-cyclohexylamino)-ethane sulfonic acid (NCES) is displayed as atomic spheres (C: white, N: blue and O: red), and is stacked against the cavity wall. **(b)** List of residues forming the cavity wall. Hydrophobic residues are coloured black, acidic residues are in red and basic are blue. hydrophilic non-charged residues are coloured green. **(c)** Ribbon view of a dimer rainbow coloured. The cavity is represented by a green

surface originating from monomer 1 and a blue surface from monomer 2. The serendipitously bound buffer molecule NCES is displayed as atomic spheres (C: white, N: blue and O: red), and is stacked against the cavity walls from both monomers.

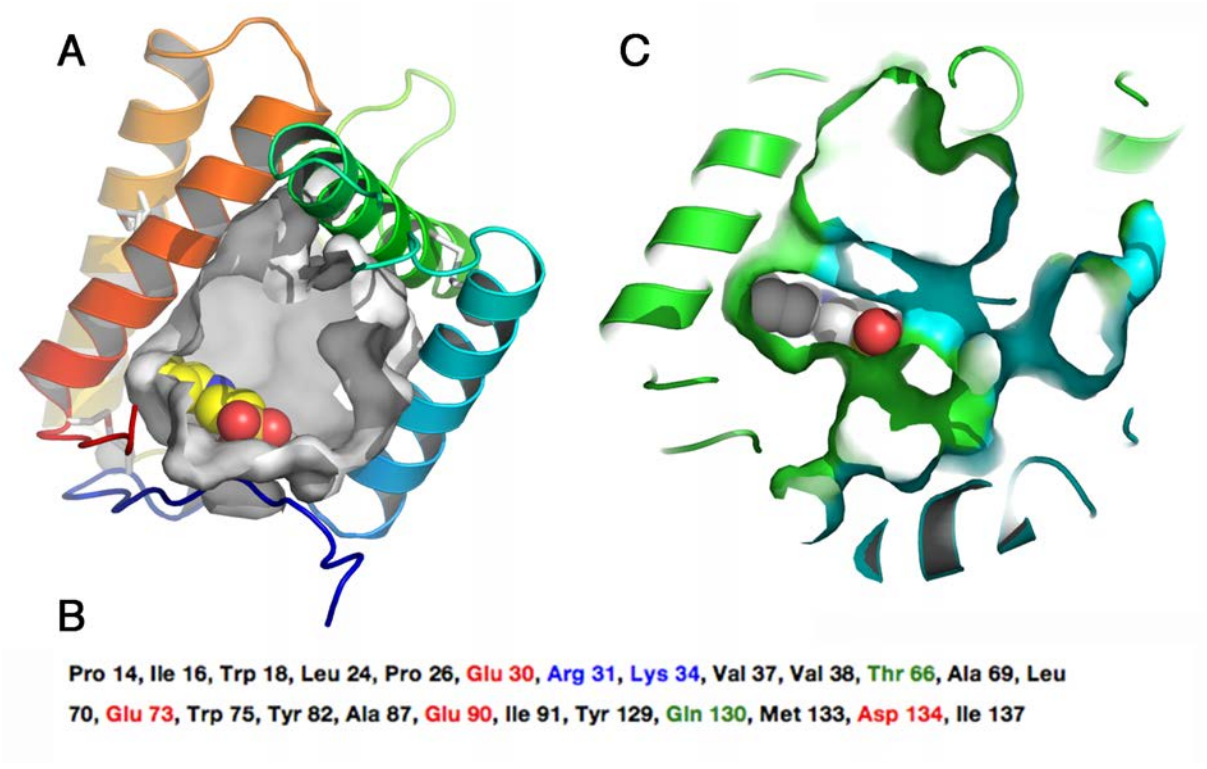


Figure 7. Comparison of crystal structures of VdesOBP1 with the closest related OBPs from *Phormia regina* (PregOBP56) and *Leucophera madereae* (LmadPBP). (a) Ribbon view of the superposition of monomers from VdesOBP1 (rainbow coloured) and PregOBP56 (light grey) with its two fatty acid ligands (grey sticks). Note the absence of the classical OBP fold helix 2 (red squared) in VdesOBP1**. (b)** Ribbon view of the superposition of monomers from VdesOBP1 (rainbow coloured) and Lmad (light grey) with its two fatty acid ligands (grey sticks). Note the absence of the classical OBP fold helix 2 (red squared) in VdesOBP1 as compared to LmadPBP (red squared) and the displacement of one disulphide bridge (red arrow). The 3D positions of the two other disulphide bridges are conserved (purple arrow), in the final fold, although as result of different C1-C6, C2-C3, C4-C5 (VdesOBP1) and C1-C3, C2-C5, C4-C6 (LmadPBP) connectivities.

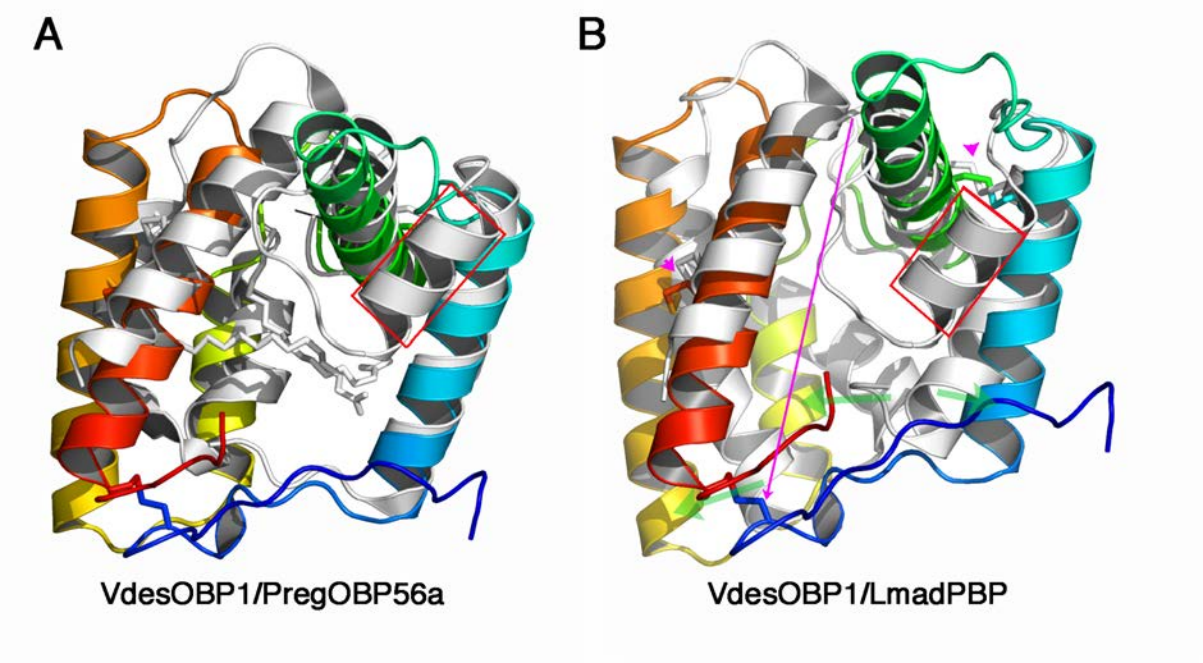
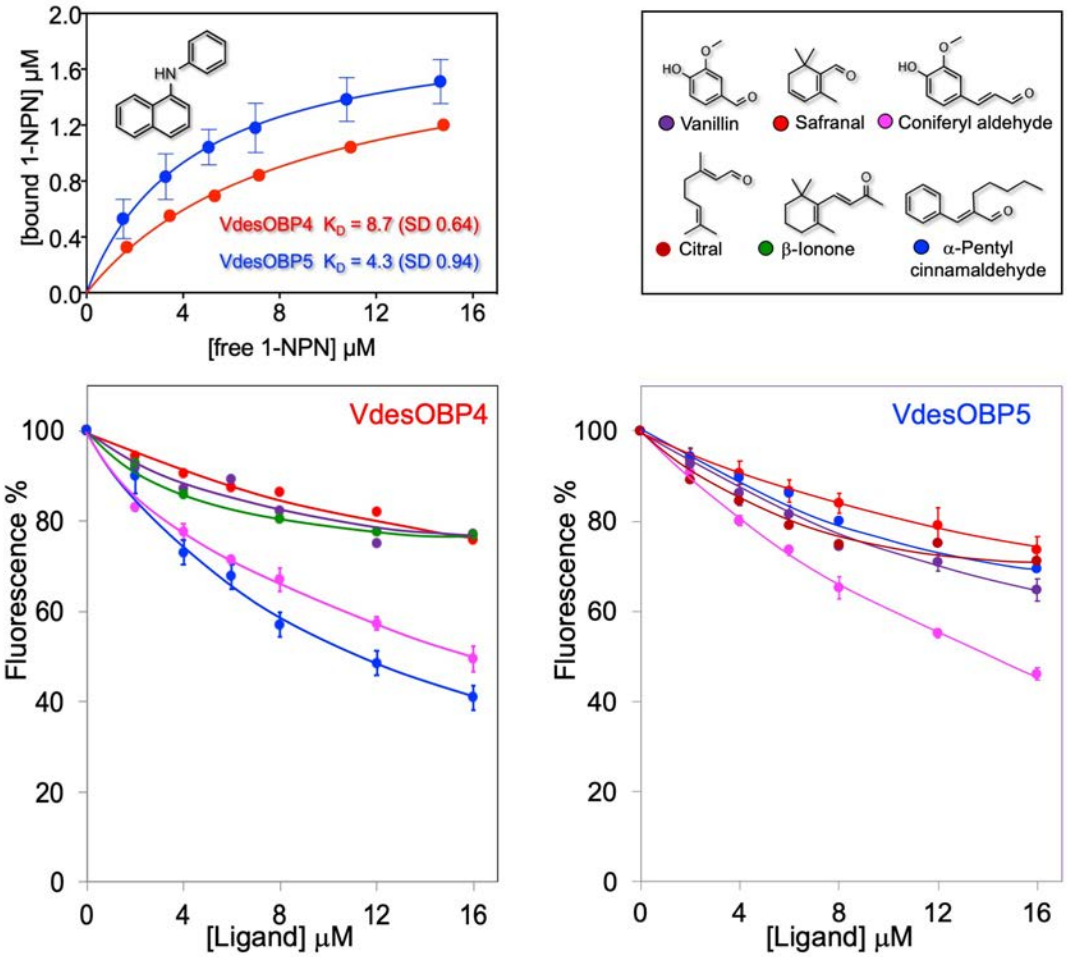


Figure 8. Ligand-binding assays. (**a**) **Binding of the fluorescent probe N-phenylnaphthylamine (1-NPN) to VdesOBP4 and VdesOBP5.** With the other two proteins expressed (VdesOBP1 and VdesOBP2) we could not observe the blue shift usually associated with binding of the fluorescent probe. (**c, d**) **Competitive binding assays**. Only few chemicals were able to displace 1-NPN from the complex with VdesOBP4 and VdesOBP5. The same chemicals were also the best ligands for the other VdesOBPs (Supplementary Fig. S2 online). (**b**) Structures of the ligand used in the competitive binding experiments.


[bound 1-NPN] µM
[free 1-NPN] µM
VdesOBP4 K_D = 8.7 (SD 0.64)
VdesOBP5 K_D = 4.3 (SD 0.94)
Vanillin
Safranal
Coniferyl aldehyde
Citral
β-Ionone
α-Pentyl cinnamaldehyde
VdesOBP4
VdesOBP5
Fluorescence %
[Ligand] µM

736