

# SCIENTIFIC REPORTS



OPEN

## Crystal Structures and Binding Dynamics of Odorant-Binding Protein 3 from two aphid species *Megoura viciae* and *Nasonovia ribisnigri*

Received: 04 December 2015

Accepted: 05 April 2016

Published: 22 April 2016

Tom Northey<sup>1,2,\*</sup>, Herbert Venthur<sup>3,4,\*</sup>, Filomena De Biasio<sup>3,5,\*</sup>, Francois-Xavier Chauviac<sup>1,†</sup>, Ambrose Cole<sup>1</sup>, Karlos Antonio Lisboa Ribeiro Junior<sup>3</sup>, Gerarda Grossi<sup>5</sup>, Patrizia Falabella<sup>5</sup>, Linda M. Field<sup>3</sup>, Nicholas H. Keep<sup>1</sup> & Jing-Jiang Zhou<sup>3</sup>

Aphids use chemical cues to locate hosts and find mates. The vetch aphid *Megoura viciae* feeds exclusively on the Fabaceae, whereas the currant-lettuce aphid *Nasonovia ribisnigri* alternates hosts between the Grossulariaceae and Asteraceae. Both species use alarm pheromones to warn of dangers. For *N. ribisnigri* this pheromone is a single component (*E*)- $\beta$ -farnesene but *M. viciae* uses a mixture of (*E*)- $\beta$ -farnesene, (–)- $\alpha$ -pinene,  $\beta$ -pinene, and limonene. Odorant-binding proteins (OBP) are believed to capture and transport such semiochemicals to their receptors. Here, we report the first aphid OBP crystal structures and examine their molecular interactions with the alarm pheromone components. Our study reveals some unique structural features: 1) the lack of an internal ligand binding site; 2) a striking groove in the surface of the proteins as a putative binding site; 3) the N-terminus rather than the C-terminus occupies the site closing off the conventional OBP pocket. The results from fluorescent binding assays, molecular docking and dynamics demonstrate that OBP3 from *M. viciae* can bind to all four alarm pheromone components and the differential ligand binding between these very similar OBP3s from the two aphid species is determined mainly by the direct  $\pi$ - $\pi$  interactions between ligands and the aromatic residues of OBP3s in the binding pocket.

The accurate detection of semiochemicals (naturally-occurring behaviour/development-modifying chemical signals) in a surrounding complex environment is crucial for insects to survive, and a sensitive and specifically-tuned olfactory system is essential. In most insects, there is a highly selective and sensitive olfactory recognition system for individual volatile semiochemicals that are perceived at critical stages of the insect's life cycle. The antenna is the main olfactory organ, where two olfactory protein families, the odorant binding proteins (OBPs) and the odorant receptors (ORs), are responsible for receiving, transporting and triggering responses to semiochemicals. OBPs, including the pheromone-binding proteins (PBP), are small, globular polypeptides of 150–250 amino acids and have been described in a great number of insect species from most orders. The characteristic signature of most OBPs is the presence of six highly-conserved cysteine residues, with a conserved spacing, paired in three interlocked disulphide bridges, although atypical OBPs with more than six cysteine residues have been reported<sup>1</sup>.

<sup>1</sup>Crystallography, Institute for Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, London WC1E 7HX, UK. <sup>2</sup>Research Department of Structural and Molecular Biology, University College London, Gower St., London WC1E 6BT, UK. <sup>3</sup>Department of Biological Chemistry and Crop Protection, Rothamsted Research, Harpenden, AL5 2JQ, UK. <sup>4</sup>Laboratorio de Química Ecológica, Departamento de Ciencias Químicas y Recursos Naturales, Universidad de La Frontera, Temuco, Chile. <sup>5</sup>Dipartimento di Scienze, Università della Basilicata, via dell'Ateneo Lucano 10, 85100 Potenza, Italy. \*These authors contributed equally to this work. <sup>†</sup>Present address: Ministère des Affaires Sociales et de la Santé, 14 Avenue Duquesne, 75350 Paris SP 07, France. Correspondence and requests for materials should be addressed to J.-J.Z. (email: jing-jiang.zhou@rothamsted.ac.uk)

Aphids (Hemiptera: Aphididae) use semiochemicals, such as plant volatiles and pheromones, to locate their host plants, find mates and warn of danger. There is good evidence that aphids, like other insects, use OBPs to transport semiochemicals across the lymph into the sensilla on their antennae and deliver the signals to the ORs, where a behavioural response is initiated. This includes the detection of alarm pheromones. Research on the pea aphid, *Acyrtosiphon pisum*, has shown that the recognition of the alarm pheromone (*E*)- $\beta$ -farnesene (E $\beta$ f), and structurally related molecules, is mediated by ApisOBP3 and ApisOBP7<sup>2–4</sup>. Homologous genes to ApisOBP3 have also been identified in other aphid species<sup>5,6</sup> and in other insects such as the hoverfly, *Episyrphus balteatus* (EbalOBP3)<sup>3</sup> and the brown plant hopper, *Nilaparvata lugens* (NlugOBP3)<sup>7</sup> with these OBPs showing an apparent high affinity for E $\beta$ f. Indeed it was suggested that *E. balteatus* uses E $\beta$ f as a kairomone to locate its aphid prey<sup>8</sup>. The aphid OBPs have been immunochemically localised to the aphid antennal sensilla<sup>9,10</sup>, as well as in non-sensory organs<sup>10</sup>, and RNA interference of NlugOBP3 has shown that it is involved in both nymph olfaction and more general non-olfactory functions<sup>7</sup>.

The vetch aphid *Megoura viciae* (Buckton) feeds exclusively on members of the Leguminosae, whereas the currant-lettuce aphid *Nasonovia ribisnigri* alternates its hosts between members of the Grossulariaceae and Asteraceae. When attacked, both species release an alarm signal to warn neighbouring aphids of danger. For *N. ribisnigri*, like *A. pisum* and the peach-potato aphid *Myzus persicae*, this signal is a single compound E $\beta$ f<sup>11–16</sup>. However, for *M. viciae* the alarm signal is a mixture of several terpenes [(–)- $\alpha$ -pinene,  $\beta$ -pinene, E $\beta$ f and limonene], although behavioural assays have shown that (–)- $\alpha$ -pinene is the most active component<sup>13–15</sup>.

The aim of the present study was to identify and characterise OBP3 from *M. viciae* and *N. ribisnigri* by molecular, biochemical and structural approaches and to determine whether this OBP is involved in detection of aphid alarm pheromone components. For the first time we report the crystal structures of an aphid OBP and postulate the molecular interactions with E $\beta$ f using affinity measurements and *in silico* binding prediction by molecular docking and molecular dynamics analysis.

## Results and Discussion

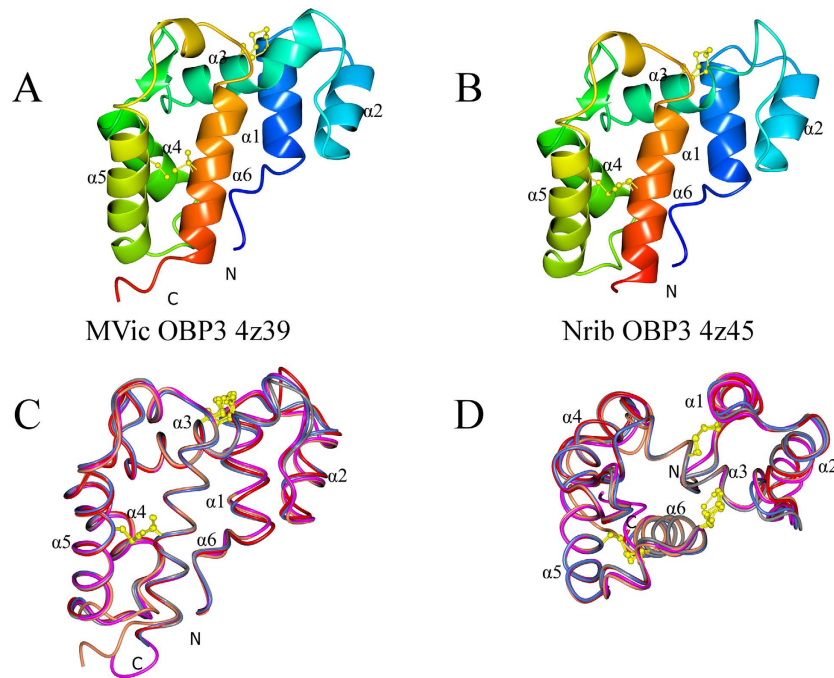
**Identification of OBP3 in *M. viciae* and *N. ribisnigri*.** To identify genes encoding OBP3 in *M. viciae* and *N. ribisnigri*, which do not have genome sequences available, we used primers designed to match the ApisOBP3 mature protein (GenBank accession: FM242529), a strategy likely to work because of the high similarity between OBPs of different aphid species<sup>6</sup>. This allowed the successful cloning of the genes encoding mature MvicOBP3 and NribOBP3 (GenBank accessions: KT750882 and KR139668), which like ApisOBP3, can be classified as “classic” OBPs, with the usual conserved spacing between the six conserved cysteines<sup>2</sup>. The mature proteins have 118 and 114 amino acids with theoretical pIs of 5.20 and 5.18, and molecular weights of 13.7 and 13.4 kDa, respectively. In contrast to the high diversity amongst OBPs from the same species<sup>6</sup>, the OBP3 sequences are very similar between aphid species with ApisOBP3 and MvicOBP3 having only three amino acid differences, and NribOBP3 having two amino acid differences plus four fewer residues at the C-terminus (Supplementary Figure 1). There is one amino acid difference between MvicOBP3 and NribOBP3 plus four fewer amino acids at the C-terminus. Since ApisOBP3<sup>2,17</sup> and the *Sitobion avenae* OBP3<sup>3</sup> have been shown to be involved in the binding of E $\beta$ f, we hypothesised that MvicOBP3 and NribOBP3 might also serve as a carrier for E $\beta$ f. However, since *M. viciae* also has other terpene components in its alarm pheromone blend; it might be that its OBP3 can also bind these compounds. We therefore decided to determine the interactions of MvicOBP3 and NribOBP3 with E $\beta$ f, (–)- $\alpha$ -pinene,  $\beta$ -pinene and (+)-limonene as well as the crystal structures of these proteins.

**Expression and purification of MvicOBP3 and NribOBP3.** We expressed MvicOBP3 and NribOBP3 in BL21(DE3) pLysS *E. coli* cells transformed with pET17b vectors containing the coding sequences for the mature proteins. The recombinant proteins, as is the case for most insect OBPs, were obtained as inclusion bodies and were therefore denatured with urea/DTT, refolded using the cysteine-cystine redox pair to enable disulphide bond formation and dialysed to obtain a soluble form. This process has been reported to give OBPs with native folding and with the correct cysteine bridges<sup>18–21</sup>. However, since it is also known that the expressed OBPs may contain endogenous ligands from the bacterial cells, they were delipidated prior to use<sup>22,23</sup>. The refolded OBPs were purified by anion-exchange chromatography followed by gel filtration, giving a single band by gel electrophoresis (Supplementary Figure 2).

**X-ray structure of MvicOBP3 and NribOBP3.** Experimentally-determined 3D structures of eighteen different classic and two plus-C insect OBPs are available in the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) and many of these have different structures dependent on their crystal form, pH or ligand. There are a further 8 salivary, haemolymph or venom protein structures that share the Interpro domain motif (IPR006170) with OBPs, but differ in length and other structural features. In total there are 104 PDB files of IPR006170 proteins containing 171 chains (August 2015 search).

The OBP structures are from Diptera (mosquitos and fruitflies), Hymenoptera (honey bee), Lepidoptera (moths), Orthoptera (locust) and Dictyoptera (cockroach), but to date there are no such structures for aphid OBPs. Indeed there are no reports of an OBP 3D structure for any Hemipteran, which comprise around 50,000–80,000 species of aphids, planthoppers, leafhoppers, shield bugs, and other crop pests despite there being more than 300 Hemipteran OBPs reported in NCBI GenBank. Thus, we present the first 3D structures of OBPs from aphids, *M. viciae* (MvicOBP3) (PDB: 4Z39) and *N. ribisnigri* (NribOBP3) (PDB: 4Z45), by X-ray diffraction (Supplementary Table 1).

The OBPs from both aphid species have a similar structure to other classical OBPs i.e. six  $\alpha$ -helices connected by extended loops, with three disulphide bridges formed by the six conserved cysteine residues between Cys17–Cys44; Cys40–Cys101 and Cys86–Cys110 (Fig. 1). The 5 chains in the two structures are reasonably similar



**Figure 1.** Cartoons of MvicOBP3 (top left) and NribOBP3 (top right) coloured from blue to red by residue number. The superposition of the 5 chains in these structure [MvicOBP3 (A) coral and (B) magenta, NribOBP3 (A) blue, (B) grey and (D) red] are shown in the same orientation (bottom left) and viewed from the top of the left view (bottom right) showing the variation is in helix 1 and 2 and the loop between helix 2 and 3.

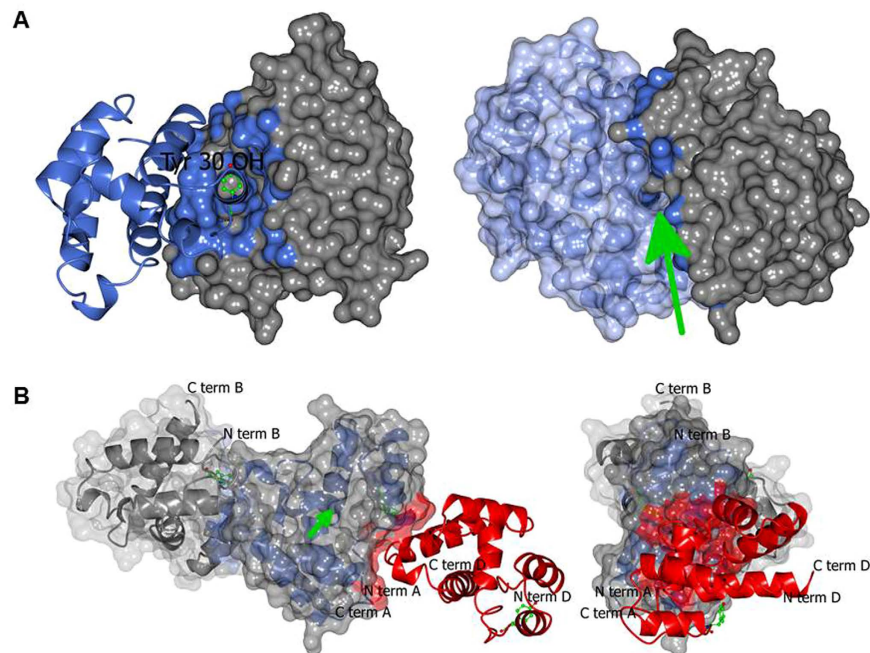
GESAMT	NribOBP3			MvicOBP3	
	A	B	D	A	B
NribOBP3 A	0.00/114	0.32/111	0.53/112	0.98/110	0.82/109
NribOBP3 B		0.00/112	0.52/110	0.94/108	0.8/108
NribOBP3 D			0.00/114	0.94/111	0.77/109
MvicOBP3 A				0.00/115	0.59/109
MvicOBP3 B					0.00/114

**Table 1.** Comparison between MvicOBP3 and NribOBP3. Calpha RMSD (Å) between chains of the aphid OBPs using GESAMT in CCP4MG<sup>54</sup>. The second number refers to the number of matching residues.

(Calpha RMSD less than 1.0 Å) (Table 1, Fig. 1) with the variation being mainly in the loops and the first two helices.

The two OBPs also have some unique structural features which are different from other classical OBPs. There is a significant groove in the surface of the aphid OBPs (indicated by an arrow in Fig. 2) formed primarily from the N-terminal region of the chains and parts of helices  $\alpha 1$  and  $\alpha 2$ . In the two molecules of the asymmetric unit of MvicOBP3, each chain partially occupies the other's groove, with Tyr30 inserted deep into the groove, forming an inter-chain hydrogen bond with the N of Phe5 in the other chain. The N-terminal residues of the other chain cover the second face of the groove and these symmetrical groove interactions form the centre of the monomer-monomer interface. The two molecules are related by very close to a twofold rotation of 179.4 degrees. A search of Protein Interfaces, Surfaces and Assemblies (PISA) ([http://www.ebi.ac.uk/pdbe/prot\\_int/pistart.html](http://www.ebi.ac.uk/pdbe/prot_int/pistart.html))<sup>24</sup> was used to analyse this interface, giving a calculated interface surface area of 1110 Å<sup>2</sup> per chain. PISA predicts that an interface of this size is biologically relevant rather than being a purely crystal packing interface. The formation of the monomer-monomer interface has the observed solvation free energy gain ( $\Delta iG$ ) of  $-6.5$  kcal/mol with the  $\Delta G_{diss}$  of 13.5 kcal/mol and a p value of 0.463. The interface generates a pocket, which assessed by CASTp<sup>25</sup> has a volume of 390 Å<sup>3</sup>. The volume of this pocket is bigger than the internal cavity of the *Antheraea polyphemus* PBP ApolPBP without ligand which has a volume of 282 Å<sup>3</sup><sup>26</sup> and that of *Bombyx mori* PBP BmorPBP1 which has a volume of 167 Å<sup>3</sup><sup>27</sup>. However, it is smaller than the 492 Å<sup>3</sup> of *A. mellifera* OBPAmeASP1<sup>28</sup> and the 435 Å<sup>3</sup> of *Leucophaea maderae* OBPLmadPBP<sup>29</sup>.

Protein-protein interactions of OBPs has been reported for some mosquito OBPs, such as CquiOBP1, AgamOBP1 and AaegOBP1, where the OBPs form a hydrophobic tunnel, which is tuned to bind long molecules, such as (5R,6S)-6-acetoxy-5-hexadecanolide, the mosquito oviposition pheromone (MOP), which binds to



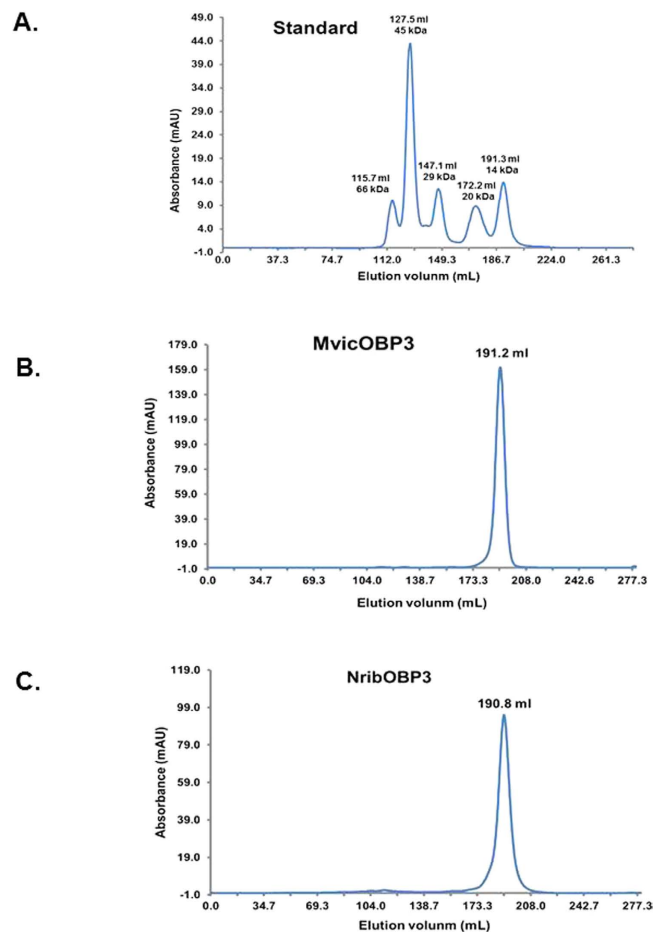
**Figure 2. The protein-protein interface in the crystal structure of MvicOBP3 (A) and NribOBP3 (B).** The green arrows indicate the binding pockets. Tyr30 is shown in green. (A) Left: the interface of chain A (grey surface with interface residues in blue) and chain B (blue ribbon) with Tyr30 (green). Right: the view of horizontal 90 degree rotation around the pocket entrance. Chain B is in a light blue surface. (B) Left: chains B (grey), A (blue) and D (red) in the filament. The surfaces of A and B are grey with the interface coloured by the interacting chain. Right: the end view of three fold rotation along the filament. The three green Tyr30s indicate the three fold.

CquiOBP1. Each CquiOBP1 binds a single molecule of MOP, but the aliphatic chains of the two molecules come close at the dimer interface, and make some contacts with the second chain. There is a fenestration to the binding pocket near the head group of MOP in each molecule so the dimer face need not open up to allow binding<sup>30</sup>. In the very similar (90% identity) AgamOBP1 dimer, DEET molecules make a similar level of contacts as MOP across the dimer interface, with polyethylene glycol (PEG) molecules occupying the side entrance to the pocket<sup>31</sup>. A single larger, but non-physiological PEG binds to both subunits in AegOBP1 and AgamOBP1 in the absence of other ligands<sup>32,33</sup>. The *A. mellifera* OBP AmelASP1 has a dimer formation that leads to domain swapping<sup>28</sup>. The interfaces of these OBPs bury 1200–1400 Å<sup>2</sup> of the proteins so they are not much larger than that of MvicOBP3. Also the plus-C OBP, OBPA48 from *A. gambiae*<sup>34</sup>, forms a predicted dimer burying an area of 2200 Å<sup>2</sup> from each chain generating a large pocket with a volume of 4330 Å<sup>3</sup> containing a PEG molecule.

However, the interface seen in MvicOBP3 is not recapitulated with NribOBP3 (Fig. 2B). Instead the three chains in NribOBP3 are in a filament generated by an approximate three fold screw axis perpendicular to the crystallographic three-fold screw (Angles 117.9, 122.4 and 119.8 degrees and translations of 28.9–29.3 Å summing to the a and b axis length). This contact involves some of the same residues as the symmetric putative protein interface contact in MvicOBP3 (Arg4, Phe5, Thr7 and Tyr30), but buries a much smaller area (average 608 Å<sup>2</sup>) and would not form a closed dimer assembly. This combined with gel filtration evidence that both MvicOBP3 and NribOBP3 are monomers in solution (Fig. 3), indicates that the proteins are likely to function as monomers. We did not observe any time dependent dimerisation of MvicOBP3 as reported for some other OBPs<sup>34,35</sup>.

**Comparison of MvicOBP3 and NribOBP3 with other OBPs.** To compare the aphid OBPs with other OBPs in the PDB, searches on PDBfold and DALI were run with chain A of the two OBPs. Using PDBfold MvicOBP3 gave 87 hits. Only the lowest hit by Z-score (PDB: 1M63) does not match IPR006170, this is also the 72<sup>nd</sup> by Z-score and only non-match to IPR006170 of the 75 matches to NribOBP3. RMSD Calpha for these hits varies from 2.1–3.7 Å, number of aligned residues from 71 to 104 (out of 118 or 121) and sequence identity from 8–20% (Supplementary Table 2). A DALI server search with MvicOBP3 and NribOBP3 gave 167 and 160 hits respectively, before the first non-IPR006170 protein, with all 160 common to both hit lists. The RMSD of these fits ranges from 2.7 to 5.1 Å, with the number of aligned residues ranging from 80 to 111 and the sequence identity from 7–20%.

These analyses confirm that the sequence identity with other OBPs is quite low, and the RMSD with OBPs structures is quite high. Not only is sequence diversity high in the OBPs, but there is also considerable conformational flexibility as agreements between different structures of the same OBPs are often above 1.0 Å RMSD. The sequence alignment of the PDB representatives of the OBPs family shows almost no conservation outside the conserved cysteines. This means it is not feasible to analyse the evolutionary conservation across the OBPs family.



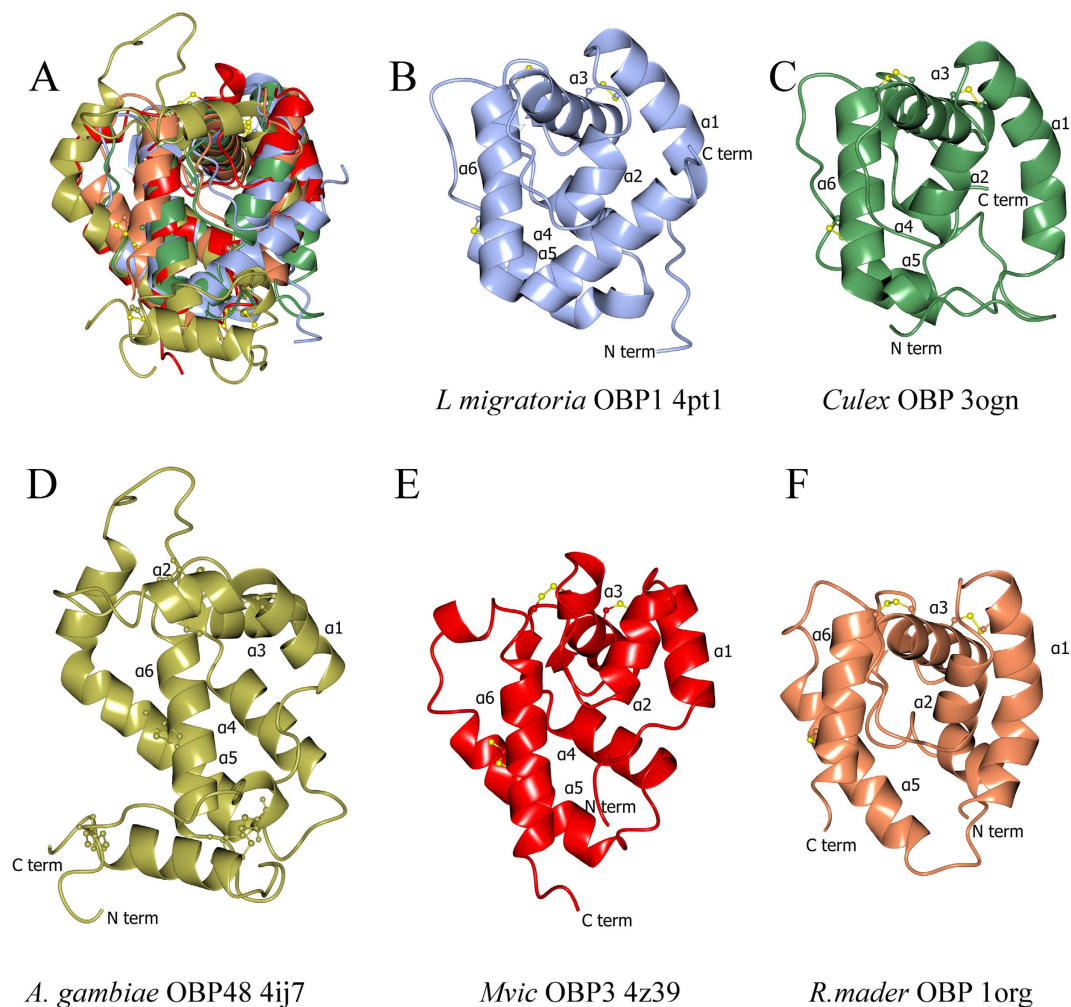
**Figure 3. Gel filtration trace of protein standards, MvicOBP3 and NribOBP3.** The protein elution volume is labelled above each peak. The protein standards are lactalbumin (14 kDa), trypsin inhibitor (20 kDa), carbonic anhydrase (29 kDa), ovalbumin (45 kDa) and BSA (66 kDa).

The correlation between the DALI and PDBefold lists is only moderate. The Pearson Rank correlation coefficient when ordered on Z-score is 0.46 between the PDBefold and the DALI order for MvicOBP3 and only 0.23 for NribOBP3. Changing the factor used to rank the hits also shows only a moderate correlation, indicating that the two aphid OBP3s are related to other OBPs but there is not a subset of existing structures that matches well to the two aphid OBP3s by this automated comparison.

Manual inspection shows that when compared to the vast majority of OBP structures the N terminus of the aphid OBP3 is in an unusual place close to the centre of the molecule, where the C terminus is normally found. Indeed only in the structures of the cockroach OBP (PDB 1ORG, 1OW4 and 1P28)<sup>29</sup> and the longer plus-C atypical OBPs (PDB 4IJ7, 4KYN, 3PM2)<sup>34,36</sup> does the N terminus head in a similar direction (Fig. 4).

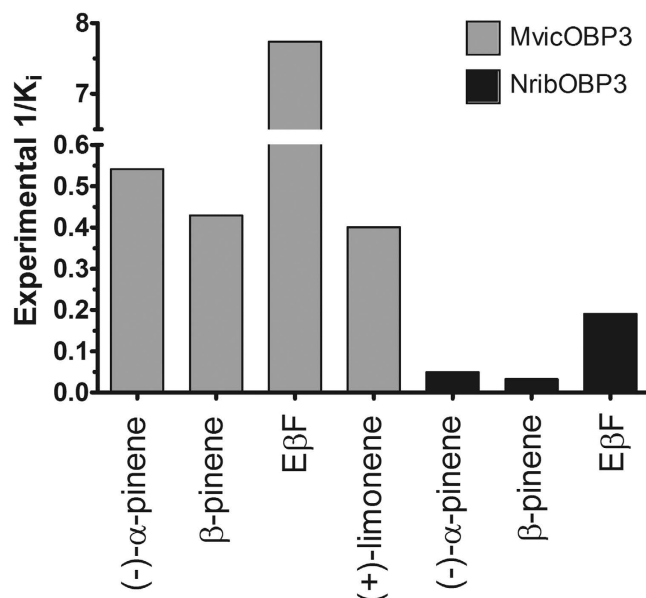
No significant internal binding pockets were predicted by CASTp<sup>37</sup> or SURFNET<sup>38</sup> for both MvicOBP3 and NribOBP3. This absence of an internal binding pocket is in contrast to other OBPs that are known to bind a ligand in an internal cavity<sup>26–28</sup>. Molecular dynamic (MD) analysis over 10 nsec simulation showed that the structural conformation changes are minor (Supplementary Figures 3 and 4), indicating that there is probably not some cryptic central pocket that was somehow hidden by crystallisation.

**Ligand binding studies with MvicOBP3 and NribOBP3.** The unique structural properties (i.e. no internal binding pocket, positioning of N-terminus) prompted us to carry out ligand binding studies. Recombinant MvicOBP3 and NribOBP3 bound the fluorescent probe, N-phenyl-1-naphthylamine (NPN), with high affinity ( $K_d$  1.9  $\mu$ M and 9.0  $\mu$ M, respectively) (Supplementary Figure 5), which allows other ligands to be tested in a competitive binding assay. Figure 5 reports the comparative inhibitor binding constants ( $K_i$ ) of four potential ligands, E $\beta$ f,  $\beta$ -pinene, (–)- $\alpha$ -pinene and (+)-limonene as determined by the displacement of NPN from MvicOBP3/NPN and NribOBP3/NPN complexes by increasing amounts of these chemicals. This shows that MvicOBP3 has a much higher affinity for E $\beta$ f ( $K_i$  0.1  $\mu$ M) than for the other alarm pheromone components  $\beta$ -pinene ( $K_i$  2.3  $\mu$ M), (–)- $\alpha$ -pinene ( $K_i$  1.8  $\mu$ M) and (+)-limonene ( $K_i$  2.5  $\mu$ M). NribOBP3 has a  $K_i$  value of 5.2  $\mu$ M for E $\beta$ f, 20.2  $\mu$ M for (–)- $\alpha$ -pinene and 30.7  $\mu$ M for  $\beta$ -pinene. Therefore, MvicOBP3 and NribOBP3 could be involved in detecting and recognising E $\beta$ f, which is consistent with previous reports of other aphid OBP3s<sup>2–4</sup>. MvicOBP3 seems to have a stronger binding affinity to these semiochemicals than NribOBP3 in the fluorescent competitive binding assay relative to the binding of NPN (Supplementary Figure 5).



**Figure 4.** (A) Superposition of five OBPs matching panels B–F in the same orientation. (B) *Locusta migratoria* OBP1 (PDB 4pt1)<sup>55</sup>. This is a highly ranked structural homologue but has the conventional C-terminus, in this case a helix, coming across the bottom of the molecule to the right hand side as drawn; (C) *Culex quinquefasciatus* OBP (PDB 3ogn)<sup>30</sup> has the C-terminus going across the bottom in a similar position to *L. migratoria*; although its N-terminus does come across to the left hand side; (D) *Anopheles gambiae* OBP48 structure (PDB 4ij7)<sup>34</sup> has the C-terminus not coming across the bottom of the molecule and staying on the left as drawn, with the N-terminal coming across the bottom of the molecule as does (E) *Mvic*OBP3 (PDB 4z39) and (F) *Rhyarobia mader* OBP (PDB 1org), although in this case the N-terminus does not block off the pocket fully.

**Molecular docking.** A blind molecular docking approach was used with E3f,  $\beta$ -pinene, (–)- $\alpha$ -pinene and (+)-limonene using Autogrid in Autodock Tools software with a grid box big enough to cover the whole chain A protein area of *Mvic*OBP3 and *Nrib*OBP3. This indicated that the four ligands were mainly docked in the region of Tyr30 and in the groove. To further investigate the ligand binding, directed-molecular docking was performed on the chain A of *Mvic*OBP3 using Autodock Vina, which is more precise and faster than its predecessor Autodock 4.0<sup>39</sup>. The best conformation for the lowest free binding energy for E3f,  $\beta$ -pinene, (–)- $\alpha$ -pinene and (+)-limonene are given in Table 2 and shown in Fig. 6. The docking results indicate that there is no major difference in the binding energy of *Mvic*OBP3 to (–)- $\alpha$ -pinene (–5.1 kcal/mol),  $\beta$ -pinene (–5.2 kcal/mol), (+)-limonene (–5.4 kcal/mol) and E3f (–5.5 kcal/mol). The competitive binding assay results give the order of strength of binding of E3f  $\gg$  (–)- $\alpha$ -pinene >  $\beta$ -pinene = (+)-limonene (Fig. 5). The stronger binding of E3f is not due to the formation of hydrogen bonds because there is a lack of functional groups such as alcohols, aldehydes or carbonyl groups in the ligands tested. The participation of residues Phe5, Val27, Tyr30 and Tyr105 in the docked conformations suggests that hydrophobic interactions are involved in the binding because of the hydrophobic nature of both the ligands and the protein residues. Some  $\pi$ - $\pi$  interactions could also be present between E3f and *Mvic*OBP3 involving Phe5 and Tyr105, due to the presence of unsaturated bonds in E3f, which are not present in the pinenes and limonene. There are differences in the binding affinities obtained between the fluorescent binding assay and the molecular docking, probably because molecular docking does not consider electronic properties, where the main structural differences of the two pinenes lies.



**Figure 5.** Competitive inhibitor binding constants ( $K_i$ ) of MvicOBP3 and NribOBP3 to the aphid alarm pheromone components measured experimentally by the fluorescence displacement binding assay. For a more clear presentation,  $1/K_i$  values have been plotted.

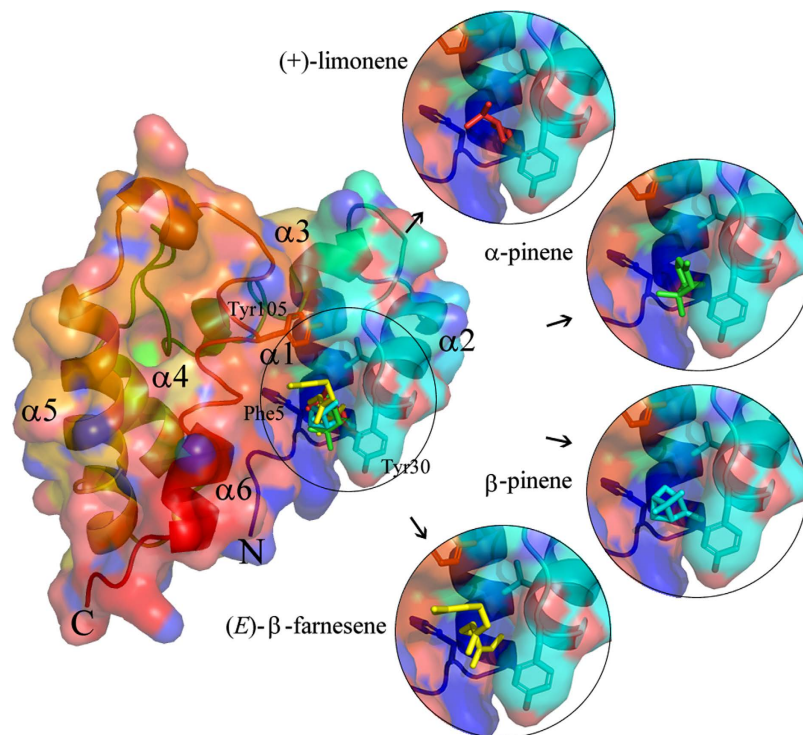
Molecule	MvicOBP3		NribOBP3	
	Binding energy (kcal mol <sup>-1</sup> )	$K_i$ (μM)	Binding energy (kcal mol <sup>-1</sup> )	$K_i$ (μM)
NPN	-7.3	4.4	-7.4	3.7
(-)-α-pinene	-5.1	181	-5.1	181
β-pinene	-5.2	153	-5.2	153
(+)-limonene	-5.4	109	-4.3	699
(E)-β-farnesene	-5.5	92	-5.1	181

**Table 2.** *In silico* binding energies and  $K_i$  for MvicOBP3 and NribOBP3.

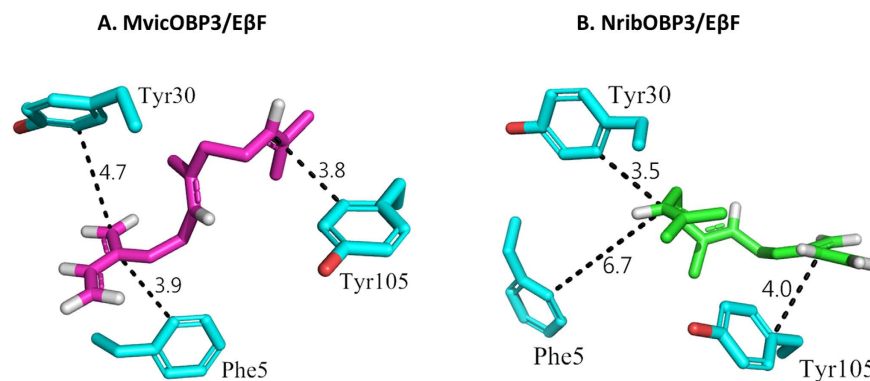
The molecular docking gives a marginally weaker NribOBP3/Eβf complex than MvicOBP3/Eβf complex (-5.5 kcal/mol for MvicOBP3/Eβf and -5.1 kcal/mol for NribOBP3/Eβf) and subsequent  $K_i$  (92 and 181 μM, respectively) (Table 2). This is also in agreement with the experimentally determined binding affinity, where MvicOBP3 has higher affinity ( $K_i$  of 0.1 μM) than NribOBP3 ( $K_i$  of 5.2 μM) (Fig. 5). In MvicOBP3/Eβf complex, the double bonds of Eβf have a more direct interaction with the aromatic rings of Phe5 and Tyr105 ( $\pi$ - $\pi$  interactions) in the binding pocket. Moreover, Eβf has more contacts with residues, such as Val27, Met108, Phe5, Lys28 and Tyr105. On the other hand, for NribOBP3/Eβf, the Eβf interacts more weakly with Tyr30, Arg4, Tyr105 and Ala109 with fewer contacts and has only one direct interaction with the aromatic ring of Tyr105. Tyr30 is inserted deep into the groove (Fig. 2 and Supplementary Figures 3 and 4) and quite close to the double bonds of Eβf (3.5 Å) (Fig. 7) but molecular docking suggests that its aromatic ring is not interacting directly with the double bonds as is in MvicOBP3/Eβf. This is supported by the C-like curved arrangement of Eβf in the NribOBP3/Eβf docked pose and the fact that the Phe5, an important residue in MvicOBP3/Eβf docked pose, is quite far away from the Eβf in NribOBP3/Eβf (6.7 Å) (Fig. 7). Docking studies of ligands to the MvicOBP3 dimer show that the best ligand/MvicOBP3 conformations are very similar to monomer MvicOBP3 complexes; all ligands are docked into the same surface binding site as in monomer MvicOBP3 although not as deep as with the monomer but with higher binding energy. This indicates that the binding pocket is the likely binding site.

Overall the docking studies indicate that the difference in the ligand binding between MvicOBP3 and NribOBP3 is probably not due to the last four amino acid residues at the C-terminus, (which is the major sequence difference), as they are on the opposite end of the protein to the ligand binding site as is the point mutation at Ile98 in MvicOBP3 and Val98 in NribOBP3 (Fig. 6 and Supplementary Figure 3). Another main structural difference between the OBPs is in the loop between helix 2 and 3 (Fig. 1), but the docking results indicate that in both OBPs all close interactions between ligands and protein are with amino acids from helices 6 and 2. This suggests that sequence difference is almost certainly not hiding a real difference in the ligand binding energy between these two very similar OBPs.

The MD results of MvicOBP3 showed that Tyr30 is displaced towards the protein (Supplementary Figures 3 and 4). In contrast, the MD analysis for NribOBP3 shows that Tyr30 is not significantly displaced. The conformational energy of MvicOBP3 monomer is higher than that of NribOBP3 (722 kcal/mol vs 677 kcal/mol, respectively)



**Figure 6.** Superimposed binding modes of (–)- $\alpha$ -pinene (green),  $\beta$ -pinene (light blue), (*E*)- $\beta$ -farnesene (yellow) and (+)-limonene (red) in the interface of MvicOBP3 predicted by blind docking.



**Figure 7.** Docked conformation of (*E*)- $\beta$ -farnesene (sticks) into the putative binding pocket on chain A of MvicOBP3 and NribOBP3. Hydrogens (white) in the ligand adjacent to unsaturated bonds are shown. Distance of the amino acid residues Tyr30, Phe5 and Tyr105 to (*E*)- $\beta$ -farnesene in the binding pocket of MvicOBP3 (A) and NribOBP3 (B) is shown in Å and as black dashed lines.

indicating that it is not in a conformation as stable as NribOBP3 conformation. The most likely explanation is that Tyr30 is involved in the crystal packing/dimerisation in MvicOBP3 and is relaxing towards its probable monomeric conformation more like that seen in NribOBP3. It is also possible that these differences between the OBP structures are because of the different crystallisation pHs (NribOBP3 has been crystallised at pH 5.0 whereas MvicOBP3 at pH 7.5). However as Tyr30 lies close to the predicted ligand binding site, its flexibility may well also be involved in ligand binding either directly or via alterations to the shape of the binding site. We cannot rule out that the optimum binding site is in some other conformation of Tyr30. The conformational energy for the MvicOBP3 conformation point mutated to the NribOBP3 sequence is 710 kcal/mol which is 12 kcal/mol less than MvicOBP3 (722 kcal/mol). That implies that a conformational energy of 33 kcal/mol of the total difference in energy of 44 kcal/mole between NribOBP3 (677 kcal/mol) and MvicOBP3 (722 kcal/mol) could be ascribed to the movement of Tyr30 between the two conformations.

## Conclusions

Two OBPs from the aphids *M. viciae* and *N. ribisnigri* were characterised by molecular, biochemical and structural approaches. For *M. viciae* OBP3 the fluorescence competition binding assays and molecular docking



demonstrate that it binds strongly to E $\beta$ f, as seen for the *A. pisum* OBP3<sup>2-5</sup>. However, it also binds to  $\beta$ -pinene, (-)- $\alpha$ -pinene and (+)-limonene so this one OBP could be responsible for transporting all four components of the alarm pheromone blend of this aphid. Behavioural experiments have shown that (-)- $\alpha$ -pinene is the most active compound in producing the alarm response of *M. viciae*<sup>40,41</sup> but in the present study both the fluorescence competitive binding assay and the molecular docking failed to show any real differential binding between (-)- $\alpha$ -pinene and  $\beta$ -pinene. In contrast NribOBP3 binds to E $\beta$ f with a lower affinity than does MvicOBP3 and has even lower affinity to the pinenes, which is consistent with the fact that *N. ribisnigri* does not use pinenes as its alarm signal.

The observation that most aphids studied so far just use E $\beta$ f as their alarm signal suggests that this is the 'ancestral' state and that *M. viciae*'s use of four components is relatively new. It would be interesting to see if *M. viciae* has other OBPs involved in discrimination of the alarm pheromone components. To this end there are 11 full length OBPs annotated in the *A. pisum* genome<sup>6</sup> and homologous genes have been identified in *M. viciae*<sup>42</sup>.

The differences in the inhibitor binding constants (K<sub>i</sub>) for the aphid OBPs between the experimental values (Fig. 5) and the theoretical values (by molecular docking) (Table 2) could be due to the nature of the methods used in this study. In the competitive fluorescence binding experiments the inhibitor binding constants (K<sub>i</sub>) are measured indirectly through the displacement of the fluorescent probe and expressed as ligand concentration. Whereas, the theoretical ligand binding affinities are predicted by molecular docking (i.e. force field calculations based on molecular mechanics) with an empty binding pocket and expressed as free binding energies. Inhibitor binding constants are calculated from the free binding energies and expressed as K<sub>i</sub> by the docking software. Moreover, the ligands in this study are considered as flexible, allowing torsional rotations<sup>43</sup>. Another factor that may contribute is the dynamic nature of the ligand binding, during which protein conformation may change, whereas the protein was rigid in the docking<sup>44</sup>.

MvicOBP3 and NribOBP3 are almost identical, and neither the sequence differences nor their C-terminus are contributing factors to their different 3D structure and ligand binding affinity. Molecular dynamic analysis demonstrates the residue Tyr30 (its interactions with ligand and orientation before and after ligand binding) may play an important role in the ligand binding. MvicOBP3 is in an open conformation (Tyr30 away from the pocket) because of the interaction with another monomer. The MD indicates that the MvicOBP3 would prefer to have Tyr30 in the NribOBP3 position (towards the binding pocket) in its monomeric state. The *in silico* binding studies may not have fully captured its role as calculated binding to the 'closed' Tyr30 in NribOBP3 and 'open' Tyr30 in MvicOBP3 structures are more or less the same.

Aphids are the main insect pests of agricultural crops in temperate regions, often causing major economic losses. Although broad-spectrum insecticides are available for control, alternative and more targeted methods are needed for sustainable agriculture due to insecticide resistance, environmental pollution and reduced availability of insecticides. One alternative method for aphid control is to exploit their pheromone-mediated communication<sup>45</sup>. We appreciate that the unique features of the two aphid OBPs (no internal binding pocket, surface binding site and involvement of N-terminus in forming the binding pocket) needs further evaluation of the effects of these features on their biological functions and role in the aphids' responses to alarm pheromones. In the longer term this could contribute to alternative methods for aphid control, by exploiting their pheromone-mediated communication for the interruption of their reproduction and for the attraction of natural enemy parasitoids and predators.

## Methods

**Insect rearing.** *M. viciae* was reared on broad bean plants (*Vicia faba* L.) in pots in a chamber at 20 ± 1 °C, 75 ± 5% relative humidity (RH), under a light:dark 18:6 h photoperiod. The cultures were started from a clone originally collected from alfalfa plants in Southern Italy (Eboli, SA). *N. ribisnigri* was reared on lettuce (*Lactuca sativa*) in pots in a chamber at 23 ± 1 °C during the day and 19 ± 1 °C during the night with 60% RH and light:dark 14:10 h photoperiod.

**RNA extraction and cDNA synthesis.** Apterous adults (50 mg) of *M. viciae* and *N. ribisnigri* were homogenised separately in 1 mL TRIzol RNA extraction reagent (Sigma-Aldrich, St. Louis, MO) and total RNAs isolated following the manufacturer's protocol. The integrity and purity of the total RNA were determined by agarose gel electrophoresis and the concentration measured using a spectrophotometer (NanoDrop ND-1000). Before cDNA synthesis, the RNA samples were treated with amplification grade DNase I (Life Technologies, Carlsbad, CA), according to the manufacturer's instructions. cDNA was prepared from 1 µg of total RNA by reverse transcription, using SuperScript<sup>®</sup> III First-Strand Synthesis System for RT-PCR (Life Technologies, Carlsbad, CA), following the manufacturer's protocol.

**Cloning and subcloning of OBP3 genes.** First strand cDNA (1 µL) was amplified in a G-STORM thermocycler in a total reaction solution (20 µL) containing 1 unit of *Taq* DNA polymerase (Life Technologies, Carlsbad, CA), 0.1 mM of each dNTP, 1.5 mM MgCl<sub>2</sub>, 0.2 µM of each PCR primer in the *Taq* Buffer supplied with the polymerase. The PCR primers were designed to match the *A. pisum* OBP3 sequence (GenBank accession: FM242529) using the online tool Primer3 (<http://www.genome.wi.mit.edu/cgi-bin/primer/>) and excluding the signal peptide. The primers (Forward: 5'-AACATATGGAAAACAATCAACAAAACACTCAA-3'; Reverse: 5'-CTGAATTCTTACATGCTATTGCGTCTGAATT-3') contained restriction sites (*Nhe*I and *Eco*RI) for cloning into the expression vector, the ATG codon (forward primer) and a stop codon (reverse primer) for the production of mature recombinant proteins with an additional methionine at position 1. After a first denaturation step at 94 °C for 3 min, 35 amplification cycles (30 sec at 94 °C, 30 sec at 55 °C, 45 sec at 72 °C) were done, followed by a final step of 10 min at 72 °C. The PCR product, with the expected size was purified from the agarose gel using QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

The purified PCR product was then ligated into the pGEM T-Easy (Promega, Madison, WI) vector for 3 h at room temperature and then the product was used to transform *E. coli* Top10 chemically competent cells. The plasmid DNA was extracted from positive colonies using QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and sequenced by Eurofins MWG (London, UK).

The pGEM/MvicOBP3 plasmid DNA containing the appropriate OBP sequence was digested with *NheI* and *BamHI* restriction enzymes for 3 h at 37 °C and the digestion products were separated on a 1% agarose gel. The OBP insert was purified from the gel using QIAQuick Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, and ligated into the expression vector pET17b (Novagen, Darmstadt, Germany), linearised with the same enzymes. The resulting pET17b/MvicOBP3 constructs were sequenced to confirm that they encoded the mature protein. Similarly, the bacterial expression construct pET17b/NribOBP3 for NribOBP3 was obtained by subcloning NribOBP3 genes in pGEM/NribOBP3 into pET17b vector but flanked with *NdeI* at 5'-end and *EcoRI* at 3'-end.

The computation of various physical and chemical parameters for the OBPs was done using the online tool ProtParam on the ExPasy SIB Bioinformatics Resource Portal (<http://expasy.org/>)<sup>46</sup>. Multiple sequence alignment of the OBPs of *N. ribisnigri* and *M. viciae* was performed using ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The signal peptides were predicted by SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>).

**Protein expression and purification of recombinant OBPs.** For expression of recombinant proteins, each pET17b vector containing the MvicOBP3 sequence was used for transformation into BL21(DE3) pLysS *E. coli* cells. The protein expression was induced at  $A_{600}$  of 0.5–0.8 by addition of IPTG to a final concentration of 1 mM when the culture had reached a value of  $OD_{600} = 0.8$ . The cells were grown for a further 3 h at 37 °C, then harvested by centrifugation and sonicated. After centrifugation, the OBP was present as inclusion bodies and solubilisation was performed by denaturation in urea/DTT, renaturation and extensive dialysis against 20 mM Tris buffer pH 7.4, using a method described previously<sup>19</sup>. The protein was purified by anion-exchange chromatography as described previously<sup>47</sup> on XK 26 column (Pharmacia, Uppsala, Sweden) filled with DE-52 resin (Whatman, Kent, UK), followed by gel filtration on Sephacryl S-200 HR (GE healthcare, Piscataway, NJ). The purified protein was stored at –20 °C in 20 mM Tris–HCl at pH 7.4. The gel filtration buffer is 20 mM Tris plus 150 mM NaCl at pH 7.4, and the storage buffer is 20 mM Tris pH 7.4. The protein was desalted prior crystallisation by dialysis with 20 mM Tris pH 7.4 buffer.

**Competitive fluorescence binding assays.** Proteins were dissolved in 20 mM Tris-HCl buffer pH 7.4 and emission fluorescence spectra were recorded on a LS50 Luminescence Spectrometer (PerkinElmer, Buckinghamshire, UK) at 22 °C in a right angle configuration, with a 1 cm light path quartz cuvette and 5–10 nm slits for excitation and emission. To measure the affinity of the fluorescent probe NPN to the protein, a 2  $\mu$ M solution of the protein in 20 mM Tris-HCl pH 7.4 was titrated with aliquots of 1 mM NPN in methanol to final concentrations from 1 to 14  $\mu$ M. The probe was excited at 337 nm and emission spectra were recorded between 380 and 450 nm. The affinity of other ligands was measured in competitive binding assays, using NPN as the fluorescent reporter at 2  $\mu$ M concentration titrated with each competitive ligand at between 0.1 and 15  $\mu$ M concentrations. For determining binding constants, the intensity values corresponding to the maximum fluorescence emission were plotted against free ligand concentrations. We use GraphPad ([http://www.graphpad.com/guides/prism/6/curve-fitting/index.htm?reg\\_one\\_site\\_competition\\_ki.htm](http://www.graphpad.com/guides/prism/6/curve-fitting/index.htm?reg_one_site_competition_ki.htm)) to determine the equilibrium dissociation constant  $K_i$  of ligand by measuring its competition for NPN and nonlinear regression curve fitting to the one site competitive binding model. This model fits the  $K_i$  of the ligand directly using the fixed concentration of NPN and the dissociation constant for NPN  $K_{NPN}$  obtained from a direct fluorescence titration and nonlinear regression curve fitting to one site binding model  $Y = B_{max} * X / (K_{NPN} + X) + NS * X$  ( $X$  is the NPN concentration,  $B_{max}$  is the maximum specific binding in the same units as  $Y$ ,  $NS$  is the slope of nonspecific binding in  $Y$  units divided by  $X$  units). The corrected  $K_i$  is fitted directly rather than fitting an  $IC_{50}$  and modifying it with the Cheng and Prusoff equation<sup>48</sup>. The analysis assumes that the protein has one binding site and 100% active, and that the binding is reversible and at equilibrium. Tested compounds were E $\beta$ F (Bedoukian Research, Danbury, CT, USA),  $\beta$ -pinene, (–)- $\alpha$ -pinene and (+)-limonene (Sigma, St. Louis, MO).

**Crystallisation and 3D structure solution.** Purified and delipidated protein samples were concentrated by centrifugation for 2 minutes at 4 °C using a Vivaspinn 20 with a 5,000 molecular weight cut off filter and an Eppendorf Centrifuge S415R set at 15600 rpm. An absorbance value at 280 nm was recorded using an ND-1000 Spectrophotometer and protein concentrations were calculated using the OBP primary sequence and Protein Calculator v3.3 (<http://protcalc.sourceforge.net/>). MvicOBP3 crystallised from a mother liquor of 0.3 M AmSO<sub>4</sub>, 30% (w/v) PEG 4000, 7.5% (v/v) Glycerol. NribOBP3 was crystallised from 20% PEG 6000, 0.1 M Sodium Citrate pH 5 10 mM ZnCl<sub>2</sub>. All crystals were harvested in 2  $\mu$ L of mother liquor with glycerol adjusted to 30% (v/v). Crystal data were collected at Diamond Synchrotron, Harwell beamlines I24 and I04-1. Data was processed automatically through xia2<sup>49</sup> to determine space groups and unit cell dimensions; in all cases the 3dii operation mode, which used XDS<sup>50</sup> and scala<sup>51</sup> and a wide scan of images for indexing, gave the best quality results. The structure of MvicOBP3 was solved with Phaser crystallographic software<sup>52</sup> using sequential searches for two copies of three separate helices taken from the structure of the *Drosophila melanogaster* PBP LUSH ( $\alpha$ -helices 2, 3 and 5). Phaser was run with MvicOBP3 resolution limited to 1.7 Å. This model was then run through SHELXE to improve phases by phase extension and density modification, including iterative model building and free lunch protocol<sup>53</sup>. The resulting map from SHELXE was autobuilt using Arp/wARP. The final Arp/wARP output was then run through 9 iterative stages of model refinement using COOT and RefMac5. Models were refined in COOT using difference map peak analysis, along with rotamer and density fit analysis. The final stage of refinement involved running Refmac5 with a resolution limit of 1.3 Å. NribOBP3 was solved using the MvicOBP3 structure as a search model

in Phaser. All 121 residues, including 3 vector residues are visible in both chains of MvicOBP3. NribOBP3 has the last residue not modelled in 2 copies and the last two residues in one copy, but the vector residues are ordered.

**Binding site prediction by CASTp and blind docking approach.** To identify putative binding pockets, the coordinate files for the monomers of MvicOBP3 and NribOBP3 were submitted to the Computed Atlas of Surface Topography of proteins (CASTp) server<sup>37</sup>. Residues, area and volume of the top-three pockets in the ranking were recorded. To corroborate the pocket prediction, a blind docking approach was done using Autodock Tools to set up the grid box for the OBP. Dimension and orientation for MvicOBP3 and NribOBP3 were  $114 \times 106 \times 112$  points (0.375 Å of spacing) and 7.992 (x-centre), -2.365 (y-centre) and 15.789 (z-centre). Dockings were run using Autodock Vina<sup>39</sup> including the PDBQT files created for the protein and each ligand. Similar-structure interfaces were searched for comparison using the PISA interface search. Search results are ranked by a Q score; a structural alignment score from 0 to 1 that take into account alignment length and C-alpha RMSD.

**Refinement of Molecular docking.** Binding affinities of MvicOBP3 and NribOBP3 were tested against (-)- $\alpha$ -pinene,  $\beta$ -pinene and E $\beta$ F. The protein and ligand structures were prepared using Autodock Tools as PDBQT files. A grid box with dimension  $38 \times 38 \times 38$  points (0.375 Å spacing) and positioned in the centre of pocket A -0.056 (x-centre), 6.556 (y-centre) and 6.556 (z-centre) for MvicOBP3 was prepared. Similarly, dimension  $38 \times 38 \times 38$  points and coordinates 20.706 (x-centre), -21.413 (y-centre) and 1.748 (z-centre) were prepared for NribOBP3. Ligands were energy minimised and considered flexible while the protein was rigid. Autodock Vina was run through the command line in Windows. Every docked conformation and binding energy was recorded in a log file. Thus, binding energies were used to choose the best conformation and to calculate the theoretical inhibition constant ( $K_i$ ) with formula (1).

$$K_i = e^{\frac{\Delta G}{RT}} \quad (1)$$

where  $\Delta G$  is the binding energy, R is the gas constant and T is standard temperature. Docked ligands were visualised using PyMOL (<http://www.pymol.org>).

**Molecular dynamics.** Simulations of conformational stability on both MvicOBP3 and NribOBP3 were performed using NAMD v2.9 and CHARMM36 force field. Proteins (chain A) were solvated with water (TIP3P model) in a cubic box with a minimum distance of 10 Å between the protein and the edge of the box. Likewise, the system was neutralised by adding Na<sup>+</sup> or Cl<sup>-</sup> randomly placed in the box. All protein preparation was carried out using Visual Molecular Dynamics software (VMD). Configuration files were prepared in order that the system was simulated under periodic boundary conditions with a cutoff radius of 12 Å for non-bonded interactions and a time step of 2 fs. Extensive energy minimisations (50000 steps) were performed followed by heating through short simulations of 1 ps at 50, 100, 150, 200, 250 and 300 K. Long simulations were kept at 300 K and 1 bar pressure in the NTP (referred to a constant number of particles, temperature and pressure) during 10 ns for each aphid OBP. Root-mean-square deviation (RMSD) trajectory tool in VMD was used to calculate the RMSD with reference to the starting structure. When the plotted RMSD did not showed any big changes, coordinates were analysed every 50 frames to obtain the best representative structure (lowest energy). Conformational changes were visually analysed by PyMOL software. Conformational energies (i.e. bond, angles, dihedral and improper features) for both MvicOBP3 and NribOBP3 were calculated using the DCD files resulted from MD simulations through NAMDenergy plugin provided by VMD software, and plotted for the same software. A point mutation was performed on NribOBP3 structure using PyMOL, where Val98 was replaced by Ile (i.e. same residue as MvicOBP3) with the aim to assess overall conformational energy and the role of this main difference in sequence. After in silico mutation, 10000 steps of energy minimization were carried out by NAMD through VMD software. Conformational energy was calculated by NAMDenergy plugin.

## References

- Zhou, J. J., Huang, W., Zhang, G. A., Pickett, J. A. & Field, L. M. "Plus-C" odorant-binding protein genes in two *Drosophila* species and the malaria mosquito *Anopheles gambiae*. *Gene* **327**, 117–129 (2004).
- Qiao, H. *et al.* Discrimination of alarm pheromone (E)-beta-farnesene by aphid odorant-binding proteins. *Insect Biochem Mol Biol.* **39**, 414–419 (2009).
- Vanderloten, S., Francis, F., Haubruge, E. & Leal, W. S. Conserved odorant-binding proteins from aphids and eavesdropping predators. *Plos One* **6**, e23608 (2011).
- Sun, Y. F. *et al.* Two odorant-binding proteins mediate the behavioural response of aphids to the alarm pheromone (E)-beta-farnesene and structural analogues. *Plos One* **7**, e32759 (2012).
- Zhong, T., Yin, J., Deng, S., Li, K. & Cao, Y. Fluorescence competition assay for the assessment of green leaf volatiles and trans-beta-farnesene bound to three odorant-binding proteins in the wheat aphid *Sitobion avenae* (Fabricius). *J Insect Physiol.* **58**, 771–781 (2012).
- Zhou, J. J. *et al.* Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrtosiphon pisum*. *Insect Mol Biol.* **19** Suppl 2, 113–122 (2010).
- He, P. *et al.* Distinct expression profiles and different functions of odorant binding proteins in *Nilaparvata lugens* Stal. *Plos One* **6**, e28921 (2011).
- Micha, S. G. & Wyss, U. Aphid alarm pheromone (E)-beta-farnesene, a host finding kairomone for the aphid primary parasitoid *Aphidius uzbekistanicus* (Hymenoptera, Aphidiinae). *Chemoecology* **7**, 8 (1966).
- Sun, Y. P., Zhao, L. J., Sun, L., Zhang, S. G. & Ban, L. P. Immunolocalization of odorant-binding proteins on antennal chemosensilla of the peach aphid *Myzus persicae* (Sulzer). *Chemical Senses.* **38**, 129–136 (2013).
- De Biasio, F. *et al.* Expression pattern analysis of odorant-binding proteins in the pea aphid *Acyrtosiphon pisum*. *Insect Sci.* **22**, 220–234 (2014).

11. Bowers, W. S., Nault, L. R., Webb, R. E. & Dutky, S. R. Aphid alarm pheromone: isolation, identification, synthesis. *Science* **177**, 1121–1122 (1972).
12. Edwards, L. J., Siddall, J. B., Dunham, L. L., Uden, P. & Kislw, C. J. Trans-betafarnesene, alarm pheromone of green peach aphid, *Myzus persicae* (Sulzer). *Nature* **241**, 1 (1973).
13. Griffiths, D. C. & Pickett, J. A. A potential application of aphid alarm pheromones. *Entomol Exp Appl.* **27**, 3 (1983).
14. Pickett, J. A., Wadhams, L. J., Woodcock, C. M. & Hardie, J. The chemical ecology of aphids. *Annu Rev Entomol.* **37**, 4 (1992).
15. Francis, F., Martin, T., Lognay, G. & Haubruge, E. Role of (E)- $\beta$ -farnesene in systematic aphid prey location by *Episyrphus balteatus* larvae (Diptera, Syrphidae). *Eur J Entomol.* **102**, 5 (2005).
16. Dewhirst, S. Y., Pickett, J. A. & Hardie, J. Aphid pheromones. *Vitam Horm* **83**, 551–574 (2010).
17. Sun, Y. *et al.* New analogues of (E)-beta-farnesene with insecticidal activity and binding affinity to aphid odorant-binding proteins. *J Agric Food Chem.* **59**, 2456–2461 (2011).
18. Angeli, S. *et al.* Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *Eur J Biochem.* **262**, 745–754 (1999).
19. Prestwich, G. D. Bacterial expression and photoaffinity labeling of a pheromone binding protein. *Protein Sci.* **2**, 420–428 (1993).
20. Plettner, E., Lazar, J., Prestwich, E. G. & Prestwich, G. D. Discrimination of pheromone enantiomers by two pheromone binding proteins from the gypsy moth *Lymantria dispar*. *Biochemistry* **39**, 8953–8962 (2000).
21. Kruse, S. W., Zhao, R., Smith, D. P. & Jones, D. N. Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from *Drosophila melanogaster*. *Nat Struct Biol.* **10**, 694–700 (2003).
22. Lagarde, A. *et al.* The crystal structure of odorant binding protein 7 from *Anopheles gambiae* exhibits an outstanding adaptability of its binding site. *J Mol Biol.* **414**, 401–412 (2011).
23. Siciliano, P. *et al.* Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Biochem Mol Biol.* **48C**, 51–62 (2014).
24. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* **372**, 774–797 (2007).
25. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–118 (2006).
26. Mohanty, S., Zubkov, S. & Gronenborn, A. M. The solution NMR structure of *Antheraea polyphemus* PBP provides new insight into pheromone recognition by pheromone-binding proteins. *J Mol Biol.* **337**, 443–451 (2004).
27. Lautenschlager, C., Leal, W. S. & Clardy, J. Coil-to-helix transition and ligand release of *Bombyx mori* pheromone-binding protein. *Biochem Biophys Res Commun.* **335**, 1044–1050 (2005).
28. Pesenti, M. E. *et al.* Structural basis of the honey bee PBP pheromone and pH-induced conformational change. *J Mol Biol.* **380**, 158–169 (2008).
29. 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).
30. Mao, Y. *et al.* Crystal and solution structures of an odorant-binding protein from the southern house mosquito complexed with an oviposition pheromone. *Proc Natl Acad Sci. USA* **107**, 19102–19107 (2010).
31. Tsitsanou, K. E. *et al.* *Anopheles gambiae* odorant binding protein crystal complex with the synthetic repellent DEET: implications for structure-based design of novel mosquito repellents. *Cell Mol Life Sci.* **69**, 283–297 (2012).
32. Leite, N. R. *et al.* Structure of an odorant-binding protein from the mosquito *Aedes aegypti* suggests a binding pocket covered by a pH-sensitive “Lid”. *Plos One* **4**, e8006 (2009).
33. Wogulis, M., Morgan, T., Ishida, Y., Leal, W. S. & Wilson, D. K. 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 (2006).
34. Tsitsanou, K. E. *et al.* Crystal and solution studies of the “Plus-C” odorant-binding protein 48 from *Anopheles gambiae*: control of binding specificity through three-dimensional domain swapping. *J Biol Chem.* **288**, 33427–33438 (2013).
35. Leal, W. S. *et al.* Reverse and conventional chemical ecology approaches for the development of oviposition attractants for *Culex* mosquitoes. *Plos One* **3**, e3045 (2008).
36. Lagarde, A. *et al.* Crystal structure of a novel type of odorant-binding protein from *Anopheles gambiae*, belonging to the C-plus class. *Biochem J.* **437**, 423–430 (2011).
37. Dundas, J. *et al.* CASTp, computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl Acids Res.* **34**, W116–W118 (2002).
38. Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph.* **13**, 323–330, 307–328 (1995).
39. Trott, O. & Olson, A. J. AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* **31**, 7 (2010).
40. Pickett, J. A. & Griffiths, D. C. Composition of aphid alarm pheromones. *J Chem Ecol.* **6**, 11 (1980).
41. Francis, F., Vandermoten, S., Verheggen, F., Lognay, G. & Haubruge, E. Is the (E)- $\beta$ -farnesene only volatile terpenoid in aphids? *J Appl Entomol.* **129**, 6 (2005).
42. Zhou, J. J. Odorant-binding proteins in insects. *Vitam. Horm.* **83**, 241–272 (2010).
43. Cosconati, S. *et al.* Virtual Screening with AutoDock: Theory and Practice. *Exp Opin Drug Discov.* **5**, 597–607 (2010).
44. Venthur, H., Mutis, A., Zhou, J. J. & Quiroz, A. Ligand binding and homology modelling of insect odorant-binding proteins. *Physiol Entomol.* **39**, 16 (2014).
45. Birkett, M. A. & Pickett, J. A. Aphid sex pheromones: from discovery to commercial production. *Phytochemistry* **62**, 651–656 (2003).
46. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy Server. 571–607 (Humana Press Inc., 2005).
47. He, X. L. *et al.* Binding of the general odorant binding protein of *Bombyx mori* BmorGOBP2 to the moth sex pheromone components. *J Chem Ecol.* **36**, 1293–1305 (2010).
48. Campanacci, V. *et al.* Chemosensory protein from the moth *Mamestra brassicae*. Expression and secondary structure from  $^1\text{H}$  and  $^{15}\text{N}$  NMR. *Eur J Biochem.* **268**, 4731–4739 (2001).
49. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Cryst.* **43**, 4 (2010).
50. Kabsh, W. X. D. S. *Acta Crystallogr Sect D: Biol Crystallogr.* **66**, 7 (2010).
51. Evans, R. R. Scaling and assessment of data quality. *Acta Cryst.* **D62**, 10 (2006).
52. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Cryst.* **40**, 17 (2007).
53. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Cryst.* **D66**, 7 (2010).
54. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr Sect D: Biol Crystallogr.* **67**, 386–394 (2011).
55. Zheng, J. *et al.* Crystal structure of the *Locusta migratoria* odorant binding protein. *Biochem Biophys Res Commun.* **456**, 737–742 (2015).

## Acknowledgements

We thank Diamond Light Source for access to beamline I24 and I04-1 (BAG mx7197) that contributed to the results presented here. Rothamsted Research receives grant-aided support from the Biotechnology and Biological

Sciences Research Council of the UK. J.-J. Zhou acknowledges the financial support from the Royal Society, UK (JP100849) and BBSRC (BB/J020281/1 and BB/L001683/1).

### Author Contributions

J.-J.Z. and N.H.K. conceived, designed and coordinated the study. J.-J.Z., N.H.K., L.M.F. and P.F. wrote the paper. T.N., F.X.C., A.C. and N.H.K. crystallised MvicOBP3 and NribOBP3 proteins and determined two X-ray crystal structures. H.V., F.D.B., K.A.L.R.J. and G.G. performed gene cloning, expression and recombinant protein production. H.V. and J.-J.Z. carried out the molecular docking and dynamic analysis. All authors reviewed the results and approved the final version of the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Northey, T. *et al.* Crystal Structures and Binding Dynamics of Odorant-Binding Protein 3 from two aphid species *Megoura viciae* and *Nasonovia ribisnigri*. *Sci. Rep.* **6**, 24739; doi: 10.1038/srep24739 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>