

Article

# A Comparative Evaluation of the Structural and Dynamic Properties of Insect Odorant Binding Proteins

George Tzotzos 

Ferrogasse 27, 1180 Vienna, Austria; gtzotzos@me.com

**Abstract:** Insects devote a major part of their metabolic resources to the production of odorant binding proteins (OBPs). Although initially, these proteins were implicated in the solubilisation, binding and transport of semiochemicals to olfactory receptors, it is now recognised that they may play diverse, as yet uncharacterised, roles in insect physiology. The structures of these OBPs, the majority of which are known as “classical” OBPs, have shed some light on their potential functional roles. However, the dynamic properties of these proteins have received little attention despite their functional importance. Structural dynamics are encoded in the native protein fold and enable the adaptation of proteins to substrate binding. This paper provides a comparative review of the structural and dynamic properties of OBPs, making use of sequence/structure analysis, statistical and theoretical physics-based methods. It provides a new layer of information and additional methodological tools useful in unravelling the relationship between structure, dynamics and function of insect OBPs. The dynamic properties of OBPs, studied by means of elastic network models, reflect the similarities/dissimilarities observed in their respective structures and provides insights regarding protein motions that may have important implications for ligand recognition and binding. Furthermore, it was shown that the OBPs studied in this paper share conserved structural ‘core’ that may be of evolutionary and functional importance.

**Keywords:** insect OBPs; structural analysis; covariance similarity analysis; elastic network models



**Citation:** Tzotzos, G. A Comparative Evaluation of the Structural and Dynamic Properties of Insect Odorant Binding Proteins. *Biomolecules* **2022**, *12*, 282. <https://doi.org/10.3390/biom12020282>

Academic Editor: Dov Borovsky

Received: 12 January 2022

Accepted: 24 January 2022

Published: 9 February 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Insects are the dominant animal species on earth. Their remarkable evolutionary success is largely attributed to their highly intricate olfactory systems. Early olfactory processing takes place at the peripheral nervous system, and involves solubilisation, and transport of semiochemicals (pheromones, and odorant or sapid molecules) to olfactory receptors (ORs) [1,2]. ORs are located on the dendrites of sensory neurons and bathed in an aqueous medium known as the sensillar lymph. They mediate the detection of chemical cues. Further processing of olfactory signals occurs in the central nervous system resulting in odour perception and thence adapted behaviour. Other protein components of the peripheral olfactory system include odorant binding proteins (OBPs), odorant degrading enzymes (ODEs) involved in the inactivation of excess odorants following OR activation, sensory neuron membrane proteins (SNMPs) functioning in tandem with ORs [3], and ionotropic receptors (IRs), which are thought to be an evolutionarily different mechanism of chemoreception [4]. OBPs, which are the subject of this review, are implicated in the solubilisation, binding and transport of semiochemicals to ORs.

The first insect OBP was isolated from the giant moth *Antheraea polyphemus* in the early 1980s [5]. Since then, the number of insect proteins classified as OBPs has grown enormously. The PFAM database [6] lists over 3000 sequences from 104 species classified as pheromone-binding and general odorant-binding proteins (Pfam ID: PF01395). The number of OBPs per insect species varies widely. For example, 69 OBPs have been associated with *A. gambiae*, whereas 52 have been associated with *D. melanogaster*, 21 with *A. mellifera* and 13 with *B. mori* [7,8].

Insects dedicate an enormous part of their metabolic resources to synthesise OBPs in their antennae, where their concentration may be as high as 10 mM [9]. RNAseq analysis has shown that the levels of RNA for the most abundantly produced OBPs in the olfactory segment of the *Drosophila* antenna are three orders of magnitude higher than those of ORs [7]. For many years, the prevailing hypothesis regarding the role of OBPs in insect olfaction associated them with the solubilisation, binding and transport of semiochemicals to ORs. Strong support for this hypothesis came from studies involving pheromones (PBP) and general odorant binding proteins (GOBPs) in Lepidopteran species demonstrating high specificity of these proteins for volatile compounds [10,11]. Added support for this hypothesis came from studies suggesting a possible release mechanism of bound odorants brought about by conformational changes in OBPs induced by the negatively charged environment at the dendritic membranes.

However, the above hypothesis has been challenged by research showing that OBPs:

1. display broad specificity of binding, and ability to bind more than one ligand simultaneously [12,13].
2. are expressed not only in the insect olfactory organs but also in other body parts [14]. For example, it was shown in proteomic studies that of the 66 genes encoding OBPs in *Anopheles gambiae* only 18 are detectable in olfactory tissues [15]. These results raise questions regarding the actual number of OBPs that may have a purely olfactory function.

In the light of the above, it is now thought that OBPs may play diverse, as yet uncharacterized, roles in insect physiology. Thus, in addition to the odorant transport hypothesis, OBPs may be involved in the protection of odorants from degradative enzymes, odorant clearance from the sensillum lymph after OR activation, or buffering against sudden changes in odour levels after the termination of an odorant pulse [7]. For OBPs expressed in non-olfactory tissues, it has been proposed that they may play a role in the modulation of mating behaviour, haematopoiesis, humidity detection and attraction or aversion of gustatory cues [8].

The rapidly increasing numbers of OBP structures, solved by X-ray crystallography and NMR, serve as a solid foundation to draw functional inferences. The number of structures deposited at the Protein Data Bank (PDB) has grown from 4 in 2004 [16] to 25 derived from 17 species belonging to 7 insect orders. These OBPs, known as “classical” OBPs, are globular proteins consisting of alpha-helices of different topologies bearing a characteristic signature of three conserved disulfide bonds. Four of the cysteines are equispaced with three amino acids between the second and the third cysteines (C<sub>2</sub>, C<sub>3</sub>), and eight amino acids between the fifth and the sixth cysteines (C<sub>5</sub>, C<sub>6</sub>). The disulfide bonds form the pattern C<sub>1</sub>–C<sub>3</sub>, C<sub>2</sub>–C<sub>5</sub>, C<sub>4</sub>–C<sub>6</sub>. “Classical” OBPs vary in the size and topology of their respective helices, and belong in the same PFAM family (PF01395) and have same fold in SCOPe (Fold a.39: EF Hand-like) [17]. In addition to the “classical” OBPs, other subgroups differ in the number of disulfide bonds: the so-called C-plus OBPs with >3 disulphide bonds, and the so-called C-minus OBPs with <3 disulfide bonds.

OBP structures have been described in detail by numerous papers as well as in several reviews [2,18–20]. Yet, X-ray structural models fail to capture structural fluctuations due to natural thermal motions, which are induced by equilibrium fluctuations and non-equilibrium effects [21]. Equilibrium fluctuations determine protein action, and as such, the dynamics of protein motion are critical in facilitating protein interactions, or even driving them. The 3D architecture of a protein, or fold, characterise its conformational space under physiological conditions, and this, in turn, circumscribes the spectrum of motions that are intrinsically accessible to the interacting partners such as odorants or proteins. By implication, evolutionary pressure may have selected specific motions that enable proteins to perform their biological functions. This process is encapsulated in the evolutionary paradigm “structure-encodes-dynamics-encodes-function” [22].

The intrinsic motions of biomolecules are usually described in terms of the spectrum of the available vibrational frequencies or normal modes (see Appendix D). The low-frequency

motions are highly collective, meaning that they involve cooperative movements of large portions of the structure, such as domain motions or structural rearrangements. These motions, frequently referred to as global or intrinsic motions, provide a useful measure of the relative rigidity in proteins. A small set of the lowest frequency vibrational modes suffices to quantify the large-scale protein motions [23]. The higher frequency modes are less collective and correspond to local changes in conformation and dynamics restricted to a few atoms. They are localised in the interior or surface of the protein and are strongly implicated in signal transmission or other internal processes [24]. Normal mode analysis (NMA) provides a powerful tool for the study of protein structure and dynamics. However, NMA is very expensive computationally and for this reason simplified coarse-grained models, such as the elastic network models (ENMs) have been developed for efficient normal mode computations [25] (see SI NMA). ENMs have been used extensively to study protein dynamics [23,26].

The aim of this review is to provide a comparative analysis of the similarities and dissimilarities of OBPs in terms of their structural and dynamic properties. The focus is not on individual OBPs but rather on this group of proteins as a whole.

## 2. Materials and Methods

The OBPs used in this study comprise exclusively the “classical” OBPs (Table 1). In order to ensure the quality of structural comparisons, classical OBPs with truncated structure (PDB IDs: 3l4a) or of low resolution (PDB IDs: 3l4a, 6vq5) or solved by NMR (PDB IDs: 2jpo, 1tuj, 6um9) were excluded. For the same reason, C-plus and C-minus OBPs were likewise excluded. The 20 OBPs in the dataset (Table 1) were compared in terms of sequence and structure similarity. They were then subjected to principal component analysis (PC) to identify patterns of structural variance, and finally to normal mode (NM) analysis using elastic network models (ENM) to determine their dynamic properties [27].

**Table 1.** The OBP dataset used in this study.

UniProt	Organism	Abbreviation	PDB ID	Resolution (Å)	No. Residues
Q8T6S0	<i>Anopheles gambiae</i>	AgamOBP1	3n7h	1.60	125
Q8T6R7	<i>Anopheles gambiae</i>	AgamOBP4	3q8i	2.00	123
Q7Q9J3	<i>Anopheles gambiae</i>	AgamOBP20	3v2l	1.80	120
Q6Y2R8	<i>Aedes aegypti</i>	AaegOBP1	3k1e	1.85	124
Q1HRL7	<i>Aedes aegypti</i>	AaegOBP22	6oii	1.85	120
Q8T6I2	<i>Culex quinquefasciatus</i>	CquiOBP1	3ogn	1.30	124
O02372	<i>Drosophila melanogaster</i>	DmelOBP76α	2gte	1.40	124
P54195	<i>Drosophila melanogaster</i>	DmelOBP28a	6qq4	2.00	121
W8W3V2	<i>Ceratitis capitata</i>	CcapOBP22	6hhe	1.52	116
P34174	<i>Bombyx mori</i>	BmorPBP1	1dqe	1.80	137
P34170	<i>Bombyx mori</i>	BmorGOBP2	2wc5	1.90	141
D0E9M1	<i>Amyelois transitella</i>	AtraPBP1	4inw	1.14	140
Q8WRW5	<i>Apis mellifera</i>	AmelASP1	3bjh	1.60	117
Q8WRW2	<i>Apis mellifera</i>	AmelASP5	3r72	1.15	122
Q3HM32	<i>Locusta migratoria</i>	LmigOBP1	4pt1	1.65	129
Q8MTC1	<i>Leucophaea maderae</i>	LmadPBP1	1org	1.70	118
A0A0S2E5N6	<i>Megoura viciae</i>	MvicOBP3	4z39	1.30	121
A0A0M4AUH6	<i>Nasonovia ribisnigri</i>	NribOBP3	4z45	2.02	118
L8B8J6	<i>Phormia regina</i>	PregOBP56a	5dic	1.18	115
A0A0R8PDN4	<i>Chrysopa pallens</i>	CpalOBP4	6jpm	2.10	119

The Bio3d v2.4-2 suite of programs was used for sequence, structural, principal component and normal mode analyses [28,29]. The nma function of Bio3d was used to for normal mode calculations employing an elastic network model (ENM) using the ‘calpha’ force field. This force field employs a spring force constant differentiating nearest-neighbour pairs of atoms along the backbone from pairs in spatial proximity. The resulting mode vectors were scaled by the thermal fluctuation amplitudes.

UCSF Chimera (Structure Measurements panel, Axes/Planes/Centroids section) was used for structure measurements described in Section 3.2 [30]. The distance and angles measurements were made based on sets of selected helical residues. The protein structure visualisation was performed with PyMOL v2.5.2 (Schrodinger pymol-open-source).

The EvoCouplings method was used for to find the residues in each OBP with the highest number of coupled interactions to their neighbours. The method is based on the direct information (DI) to compute a set of direct residue couplings that best explains all pair correlations observed in the multiple sequence alignment [31]. The EvoCouplings implementation of the methods is achieved through an open-source Python package for coevolutionary analysis [32,33].

As a prerequisite for the structural comparison of OBPs, we constructed a multiple structural alignment (MSA). Structural alignments have been found to perform better for the deduction of the structural and dynamic relatedness of the OBPs than the corresponding sequence-based ones, particularly in cases of low sequence identity [34]. Furthermore, the choice of the structural alignment method is critical because different alignment algorithms make different assumptions regarding the length of the consensus sequence, and the pairwise root-mean-square deviations (RMSD) within the set of aligned structures [35]. A local installation of Mustang v.3.2.3, interoperable with Bio3d, was used to construct the MSA [36]. The length of the sequences in the dataset ranges between 115 to 142 amino acids (average length 124 amino acids). The resulting MSA consists of 198 columns, 93 of which have 100% occupancy, i.e., columns containing no gaps (Appendix A Figure A1). An MSA in which columns with gaps have been removed is shown in Appendix A Figure A2.

### 3. Results

#### 3.1. Sequence Analysis

A percent identity value, in the form of a numeric score, was determined for each pair of aligned sequences. The identity score, which was normalised to values between 0 and 1, measures the number of identical residues in relation to the length of the alignment (Figure 1). The summary statistics of the calculated pairwise sequence identities are given in Table 2. Typically, the average sequence identity over all pairs in the dataset is around ~24%.

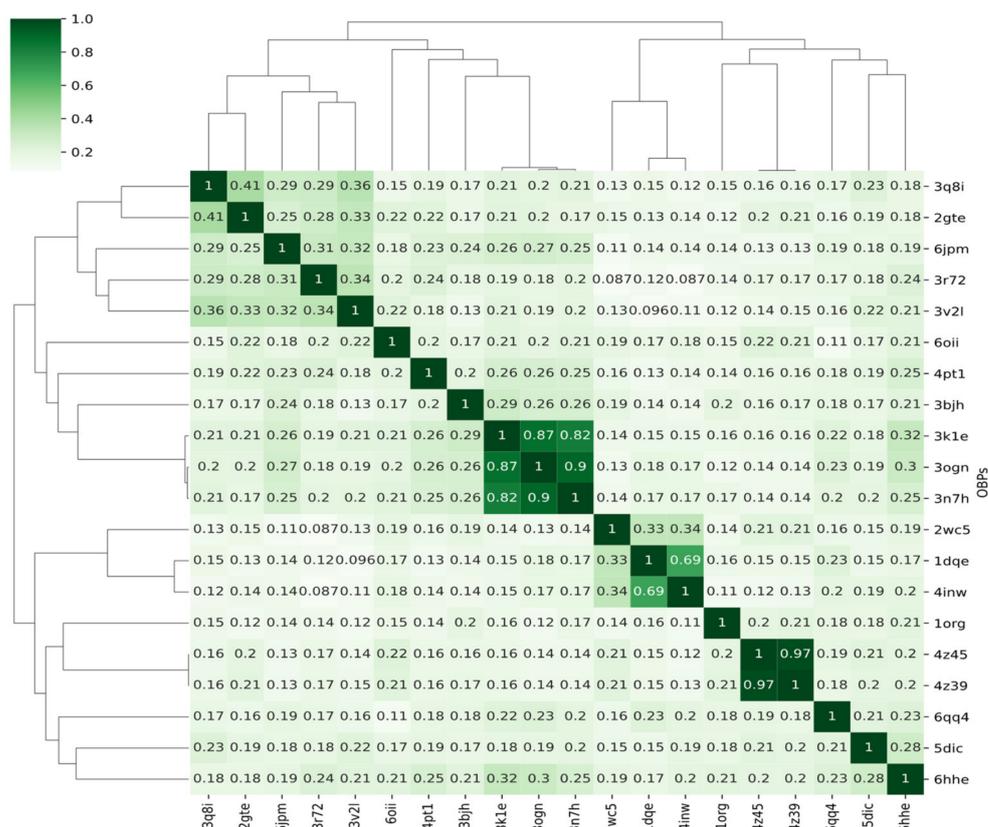
**Table 2.** Summary of the calculated OBP sequence identities.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.090	0.151	0.188	0.244	0.221	1.000

For most OBPs, the pairwise sequence identities are low, placing them in so-called “twilight-zone” where sequence identity falls between 10 and 30% (Figure 1). The “twilight-zone” is an operational term setting the boundaries of confidence for detecting evolutionary relatedness between proteins [37,38]. Despite the low sequence identities, the OBP structures have retained the same fold albeit with differences in the geometries of the packed secondary structure elements.

#### 3.2. Structure Analysis

To compare the similarity of the OBP structures, we used the structure coordinates of the C $\alpha$ -atoms and quantified it using the optimal RMSD over the sites that aligned in the MSA. RMSD is a similarity measure varying between 0 and  $\infty$  [39]. The calculation was made from the 93 equivalent positions in the MSA. The summary statistics of the calculated differences in RMSD are given in Table 3, and the heatmap and related dendrogram derived from the pair-wise similarity matrix of RMSD values (Figure 2) displays the structural relatedness of the OBPs.



**Figure 1.** Pairwise sequence similarity of OBPs. Comparisons of a protein to itself are shown along the diagonal. The dendrogram shows the hierarchical clustering obtained from the complete linkage algorithm as implemented in the Python package Seaborn.

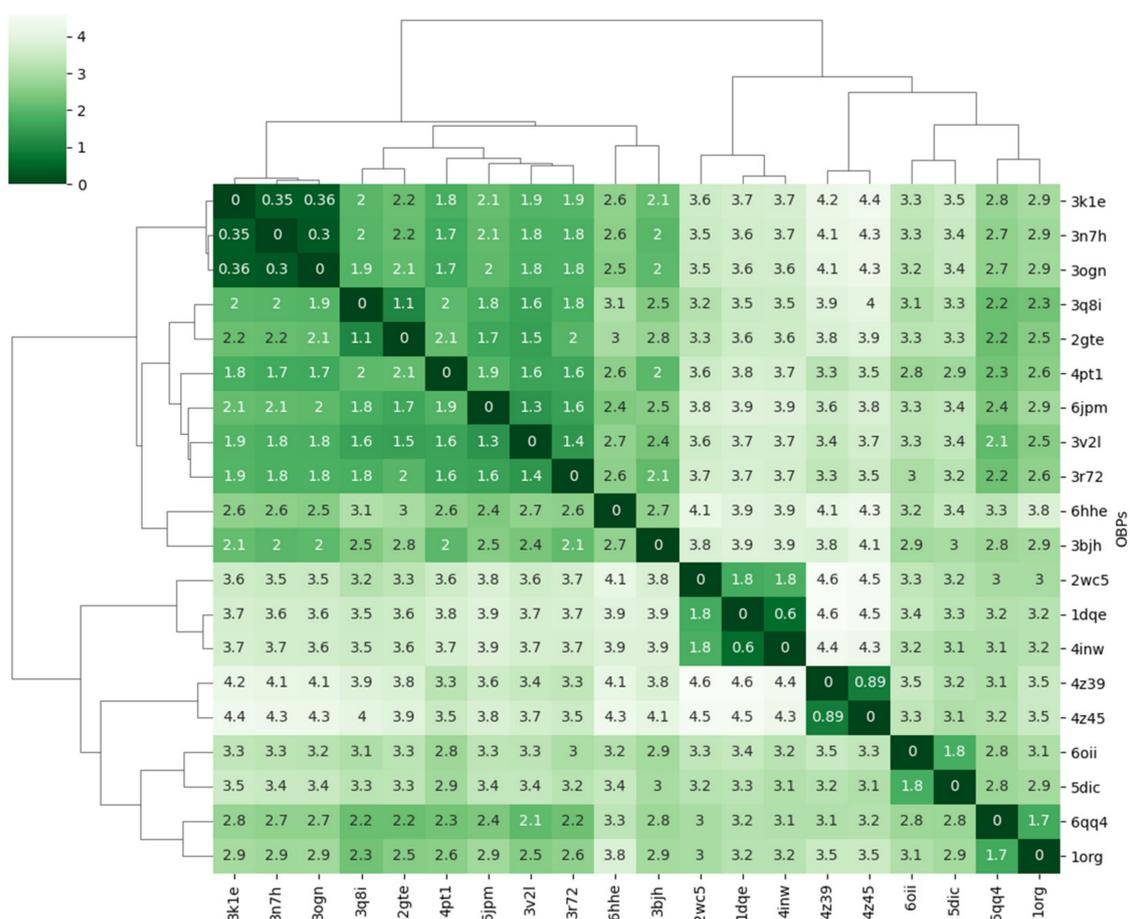
**Table 3.** Summary of the calculated OBP sequence identities (values in Å).

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.000	2.117	2.876	2.656	3.235	4.446

The dendrogram shows two main clades that split at 3 Å. One of these clades consists of 11 OBPs from 4 different insect orders, namely Diptera [AegOBP1(3k1e), AgamOBP1 (3n7h), CquiOPB1 (3ogn), AgamOBP4 (3q8i), AgamOBP20 (3v2l), DmelOBP76a (2gte), CcapOBP22 (6hhe), Hymenoptera [AmelASP1 (3bjh), AmelASP5 (3r72)], Orthoptera [LmigOBP1 (4pt1)], and Neuroptera [CpalOBP4 (6jpm)]. Three mosquito proteins (AgamOBP1, AegOBP1, CquiOBP1) are homologous (sequence identity > 80%), displaying the greatest structural similarity (RMSD 0.3–0.36 Å). The range of RMSD differences amongst the rest of the OBPs in this clade is between 1.1 Å and 3.0 Å.

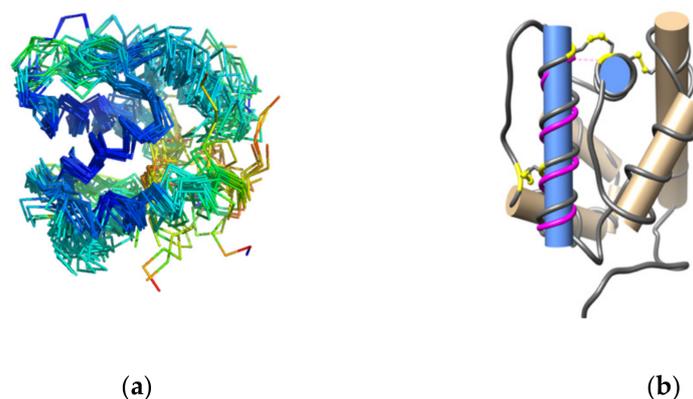
The second clade comprises OBPs that exhibit the greatest structural differences. They form the leaves of 4 distinct branches. One of the branches includes the aphid OBPs [MvicOBP3 (4z39), NribOBP3 (4z45)]. A second branch, the moth OBPs [BmorPBP1 (1dqe), BmorGOBP2 (2wc5), AtraPBP1 (4inw)]. The third branch, LmadPBP1 (1org) and DmelOBP28a (6qq4), and the fourth branch, AegOBP22 (6oii), PregOBP56a (5dic). All of these proteins are expressed in olfactory tissues with the exception of the last two. PregOBP56a is specifically expressed in a cluster of cells on the oral disk of the insect. PregOBP56a solubilises fatty acids from meat meal during feeding and delivers them to the midgut where it may help the process of reproduction [40]. AegOBP22 is expressed in multiple tissues including the antenna, and the male reproductive glands. Expressed in different antennal tissues, AegOBP22 may modulate chemosensory responses, but as a

secreted protein in the salivary gland is likely to be injected into the host during a blood meal [41].



**Figure 2.** Pairwise similarity matrix of RMSD of equivalent C $\alpha$  atoms in the MSA. Comparisons of a protein to itself are shown along the diagonal. The values are in Å. The dendrogram shows the hierarchical clustering obtained from the complete linkage algorithm as implemented in the Python package Seaborn.

The structural alignment of OBPs in the dataset involves a procedure in which (i) a score was calculated from pairwise residue-residue correspondences, followed by pairwise structural alignments; (ii) recalculation of scores in the context of multiple structures, and (iii) merging the series of pairwise alignments along a guide tree [36]. Visualisation of the superimposed structures shows helical regions where the RMSD difference of the backbone carbon atoms is the least (Figure 3, Panel a). To identify residues with the least RMSD difference, the aligned structures were subjected to an iterated superposition procedure, each round of which identified and eliminated residues displaying the largest positional differences. The subset of retained residues implies an evolutionarily conserved structural “core” consisting of 17 residues. In this case, 14 of these including OBP conserved cysteines C5 and C6, are found in the carboxy-terminal helix; the remaining 3 are found in the helix which contains C<sub>2</sub>. The pair-wise RMSD difference of the “core” residues is <1 Å. The angle between the two helices is in the range of ~78–88° and the distance between them is in the range of 6.8–8.1 Å (Figure 3, Panel b). The aphid OBPs (4z39, 4z45), and PregOBP56a (5dic) show the greatest deviations from this geometry (Appendix B, Table A1).

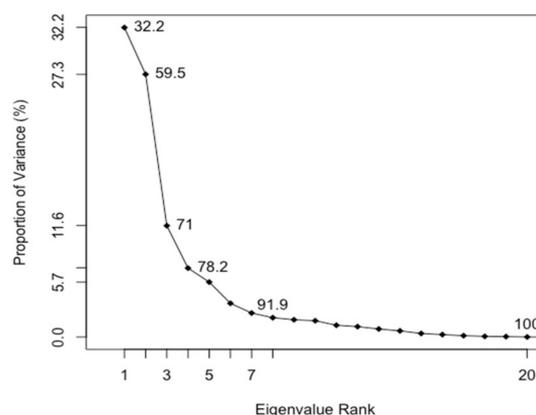


**Figure 3.** Visualisation of the superposed OBP structures. The helical regions coloured in blue show the least deviation in the positions of backbone Ca atoms (Panel a). Geometry of the residues comprising the conserved structural core (residues of the conserved core are depicted in the ribbon coloured in magenta). The two helices of the “conserved” geometry are shown in blue. AgamOBP1 (3n7h) was used in the example (Panel b).

### 3.3. Principal Component Analysis (PCA)

PCA was used to gain insights that best explain the underlying patterns of variance in the ensemble of aligned structures. When performed on an ensemble of interpolated X-ray structures, PCA captures concerted atomic displacements highlighting major structural differences between the structures. The basic premise of PCA is that it is possible to eliminate highly correlated variables in the data set without losing essential information (Appendix E).

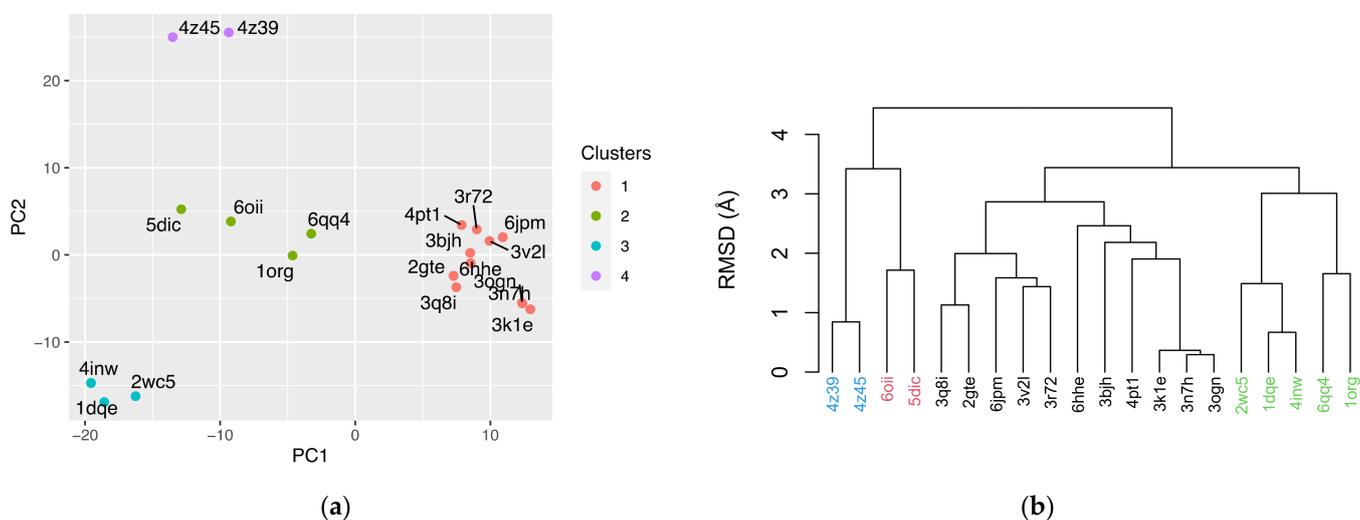
The PCA calculations were based on equivalent residues in the MSA. Approximately 82% of the total mean-square displacement (variance) of C $\alpha$  atom positional fluctuations is captured by six principal components, namely the axes of maximal variance of the distribution of structures (Figure 4). The first three principal components account for 71% of the structural variance. The eigenvalues capture the percentage of the total variance (i.e., the total mean square displacement) of atom positional fluctuations.



**Figure 4.** Scree plot. The cumulative variance accounted for by Principal Components. The eigenvalue rank represents the PC number.

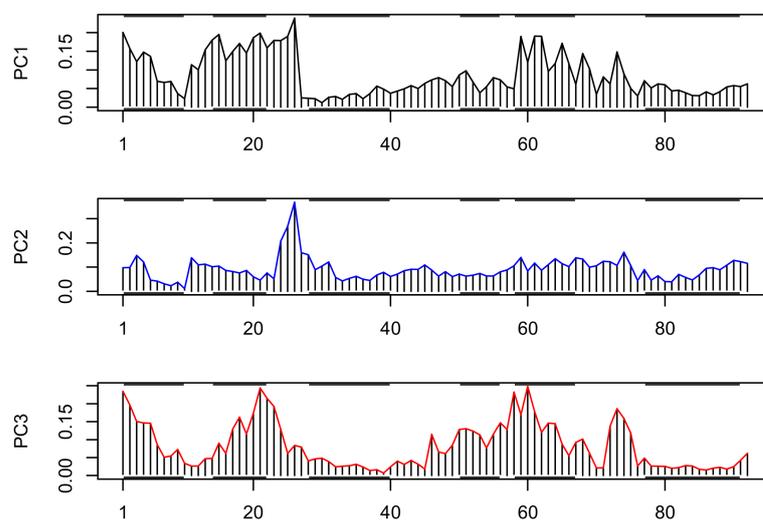
The influence of the individual OBPs on the structural variation and correlations within the ensemble of structures can be captured by means of score plots (see Appendix E). The score plot of the first two PCs, accounting to ~60% of the structural variance, maps the interrelation of the OBP structures (Figure 5a). In any given cluster depicted in the plot, the deviation of the equivalent C $\alpha$  atom positions is the least. Comparing the PCA clustering to the corresponding one obtained from the RMSD similarity matrix (Figure 5b), it can be

safely concluded that the score plot reflects the relatedness of OBPs as observed from the pairwise structural comparisons.



**Figure 5.** The PC scores in the 20 structures in the dataset, calculated from the deviation in  $\alpha$  positions. PC1 and PC2 capture 32.2% and 27.3% of the cumulative structural variance, respectively (Panel a). Clustering of OBPs based on RMSD similarity (Panel b).

The contribution of the individual residues, also known as loadings, to the structural variance covered by the first three PCs is presented in Figure 6. The figure shows that the least of the structural variance occurs in the 3rd and 6th helical regions where the residues of the structural “core” are located.



**Figure 6.** The loadings of the first three PCs. The X-axis represents the index of the residues in the multiple sequence alignment in which gaps were removed. The horizontal bars in black represent helical regions.

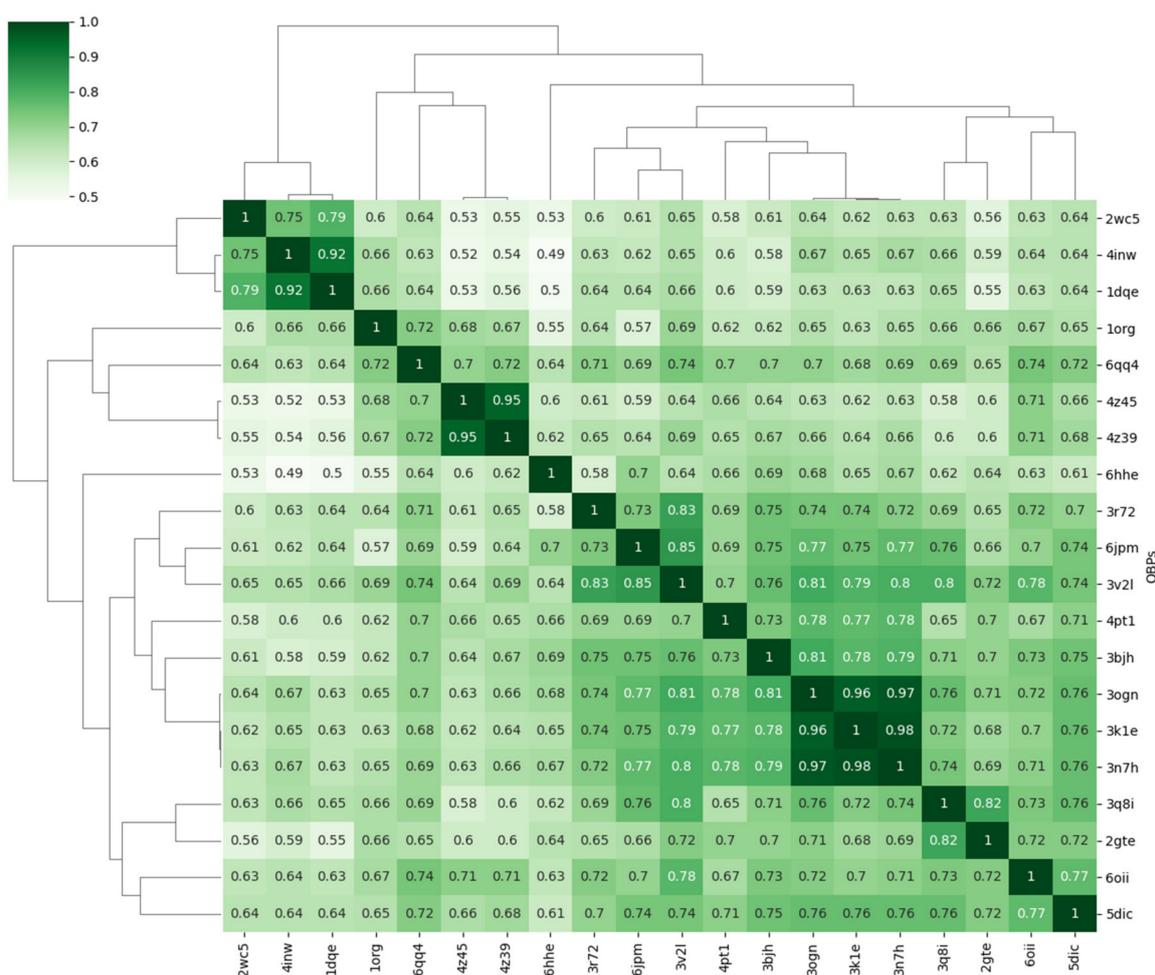
### 3.4. Normal Mode Analysis (NMA)

The intrinsic dynamics and flexibility of the OBPs were determined by the use of normal mode analysis based on the Elastic Network Models described in Section 2. The structural flexibility of proteins can be decomposed to high-frequency localised fluctuations of the individual amino acids, and to low-frequency motions of large rigid domains.

### 3.4.1. Covariance Similarity Analysis

To compare the similarity of the low-frequency motions of the OBPs, the 10 lowest modes for all pairs of structures were calculated. Two points are worth noting. First, there is no single well-accepted measure to describe dynamics similarity, and consequently linking the dynamics space to the structure space remains a challenge [39]. Second, dynamical similarity scores are extremely sensitive to MSA, particularly in cases where the sequence identities are very low. The root-mean square inner product (RMSIP) was used as a similarity metric [39]. RMSIP is sensitive to the number of modes used. Finer separations are discernible with smaller number of modes. When the number of modes included in the calculation approaches the full number, the dynamic distance order approaches the RMSD order. The RMSIP distance matrix was used to determine the relatedness of OBPs in terms of low-frequency modes.

The dynamics-based hierarchically-clustered heatmap or clustermap shows all pairwise comparisons for the aligned proteins (Figure 7). Comparison of a protein to itself along the diagonal have an RMSIP value equal to 1. Lower values indicate overlapping low-energy fluctuations of the aligned backbone carbon atoms. Fluctuations covering completely non-overlapping space have an RMSIP value equal to 0. The similarity clustering shown plot is in agreement with the structural classification of the OBPs as obtained from the pairwise RMSD measurements of the OBPs (*cf.* Figure 2).

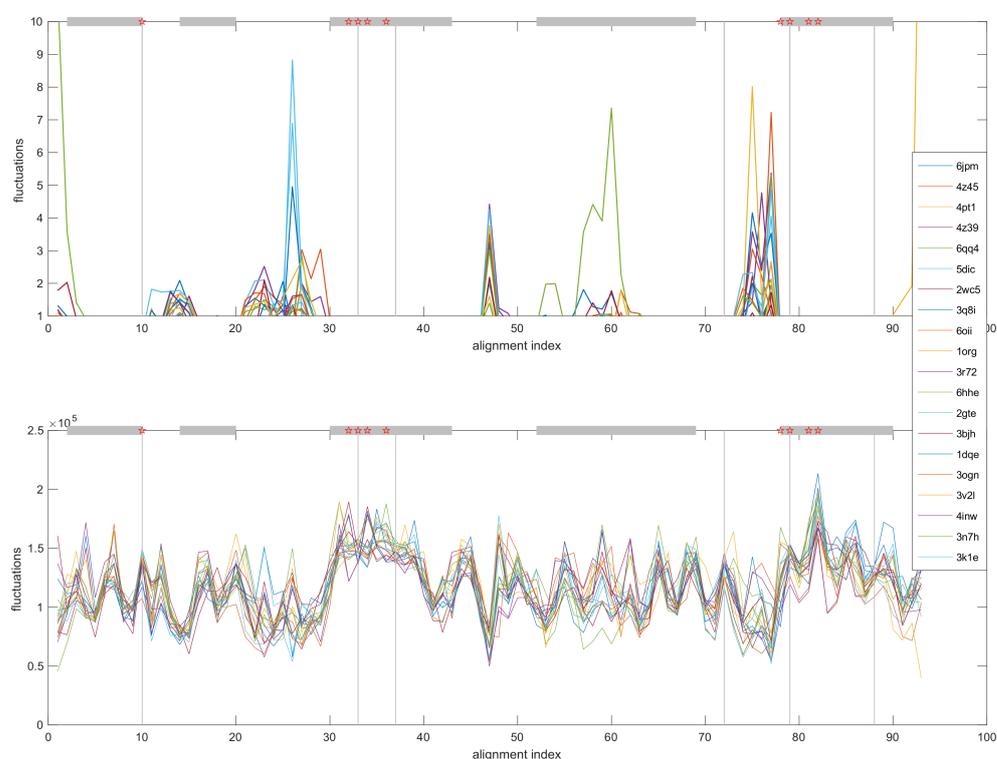


**Figure 7.** Dynamics-based hierarchically-clustered heatmap. The heatmap is represented such that each structure pairs with all others, including itself. This gives a diagonal with the maximum score of 1.0. The dendrogram was obtained using hierarchical clustering obtained from a complete linkage clustering with Euclidian distances.

### 3.4.2. Comparative Fluctuation and Deformation Analysis

The amplitude and directionality of residue fluctuations provide a measure of the local flexibility in any given protein. Fluctuations are defined as the sum of each atom's displacement along each mode, weighted by the reciprocal of the eigenvalues [42]. However, residue fluctuations are less informative regarding the amount of local flexibility of a given protein structure. Local deformation involves energy exchange arising from short-range atomic interactions. Atomic motions relative to neighbouring atoms can be calculated from the atomic energy contributions to deformation as a function of position. Thus, lower deformation energies may correspond to relatively rigid regions [24]. It is noted that deformation energies have no quantitative physical meaning, and therefore no units [43].

The fluctuation and deformation energy profiles of the conserved residues in the MSA of OBPs, averaged over all modes and normalised, are given in Figure 8.



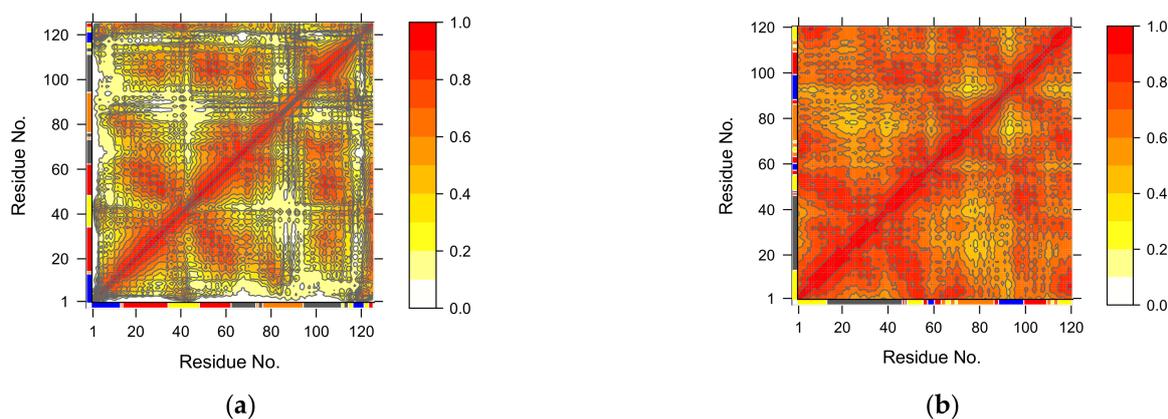
**Figure 8.** Calculated OBP residue square fluctuations and residue deformation energy profiles. The alignment index refers to columns with 100% occupancy. The horizontal bars in grey show helical regions. \* Alignment index positions showing high enrichment in evolutionary coupling interactions.

The plot in Figure 8 shows that the largest amplitude fluctuations occur almost exclusively in non-helical regions. Residues with the highest deformation energies lie in the helical regions that contain C<sub>2</sub>, C<sub>3</sub>, C<sub>5</sub> and C<sub>6</sub> and are coincident with the conserved structural core. The high deformation energy of these residues is indicative of interactions with spatially proximal residues [24,42]. Favourable interactions between spatially proximal residues are a major constraint in the evolutionary variation of proteins belonging to the same fold. Evolutionary coupling analysis of the individual OBPs showed C<sub>1</sub>, C<sub>2</sub>, C<sub>5</sub>, and residues adjacent to them in the sequence, form ‘dense’ coupling interactions with residues that are in close 3D proximity (data available in Supplementary Materials, evolutionary\_couplings.zip).

Moving beyond the structural and dynamic classification of OBPs, single structure analysis using GNM can provide additional inferences. Examples are given below.

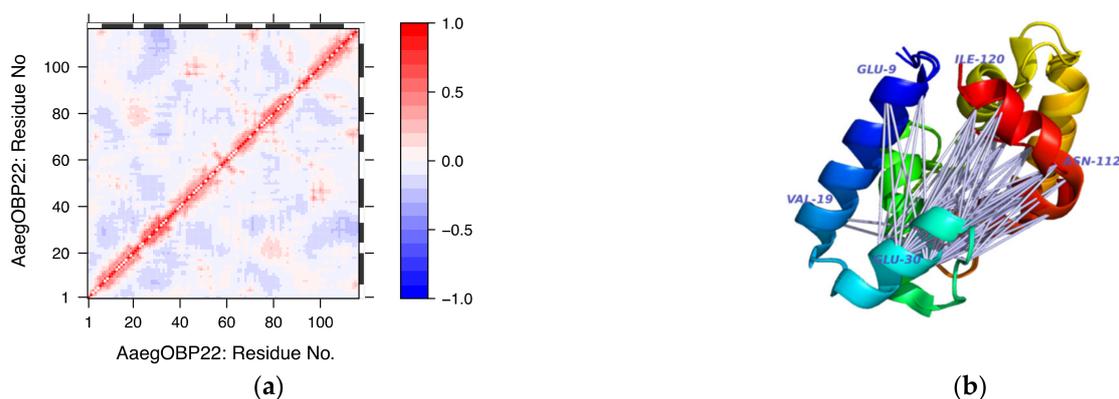
### 3.4.3. Single Analysis on Selected Examples

While the covariance matrix calculated from the low-frequency modes (see Figure 7) allows the classification of OBPs in terms of similarity of motions, it provides no information regarding their relative flexibilities. Protein motions entail the movement of relatively rigid structural domains, especially regions of secondary structure, with intervening regions of local flexibility [44]. During motion, these “dynamic” domains remain internally rigid in all conformations but move relative to each other. In OBPs such motions have been observed in studies of the *A. mellifera* antennal-specific protein, which at pH 7.0 is a domain-swapped dimer [45]. The relative flexibility of the OBPs in the dataset was determined by means of the Bio3D implementation of the Geometrically Stable Substructures (GeoStaS) algorithm [46]. Although the evaluation of residue cross-correlation coefficients is the standard method to detect atomic motions, GeoStaS has the advantage of detecting both rotations and translations, in contrast with residue cross-correlation analysis which detects translations only. The GeoStaS calculations based on the 5 lowest-frequency non-trivial modes showed considerable variation amongst the OBPs both in terms of the sizes of the “dynamic” domains, as well as the correlation of domain movements. This variation is exemplified in Figure 9 and additional material can be found in Supplementary Materials, GeoStaS and cross-correlation analysis.zip; Videos S1 and S2: global motions.



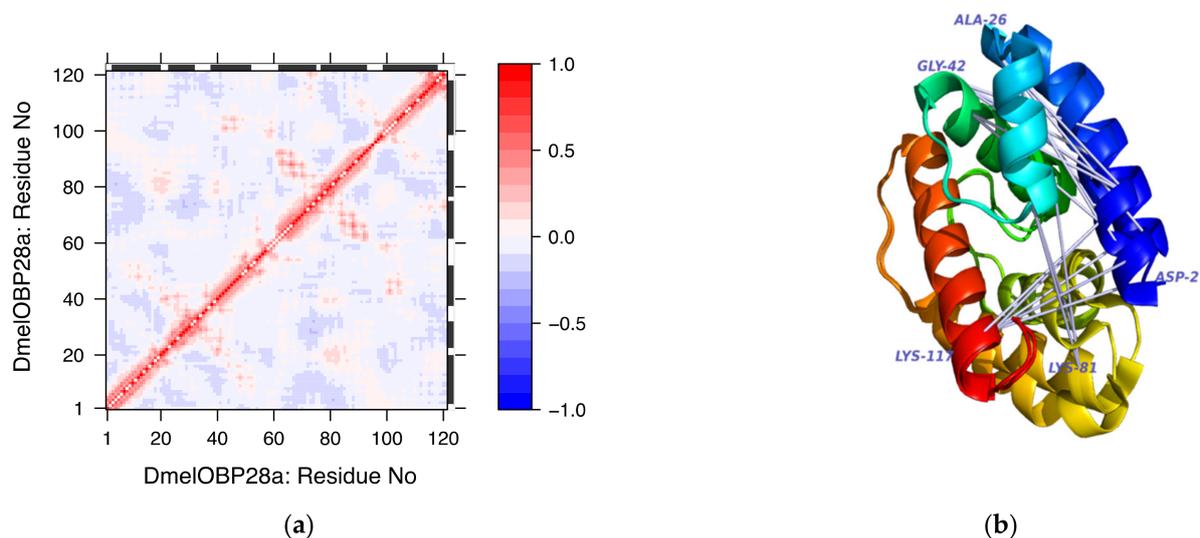
**Figure 9.** Visualization of the results obtained from the GeoStaS atomic movement similarity matrix (AMSM). AgamOBP1 domain assignment (panel a); AegOBP22 domain assignment (panel b). The colour bars at the x and y axes depict the residues assigned to the different domains that were found. The colour scale is the same in both matrices: red indicates strong correlations, and white strong anticorrelations.

Two additional examples are presented below. These involve AegOBP22 (6oii) and DmelOBP28a (6qq4). The choice of these proteins was based on the availability of crystal structures both in free and liganded forms. It is noted that, in addition to these structures, few other structures in the dataset have been crystallised in ligand-free form, namely, BmorPBP1, BmorGOBP2, AgamOBP20 and AmelASP1. AegOBP22 has been crystallised in open liganded form (PDB ID: 6oii) and closed ligand-free form (PDB ID: 6og0). Ligand binding has been associated with a conformational change at the carboxy-terminal end leading to the formation of an enlarged binding cavity [41]. The ligand-free structure is truncated missing 4 residues at the carboxy-terminus. The RMSD difference between the two structures is 1.35 Å. Subjecting the ligand-free structure to cross-correlation analysis shows significant anticorrelated motions involving residues of the 1st, 2nd and 6th helices (Figure 10, panel a). These can be visualised in the corresponding three-dimensional model of the protein (Figure 10, panel b). These motions may result in conformational changes that could accommodate ligand binding and the concomitant widening of the opening to the internal cavity of the protein.



**Figure 10.** Pairwise residue cross-correlations of AegOBP22. The cross-correlation matrices were calculated from the complete set of modes. The atomic fluctuations within a system are reflected in the magnitude of all pairwise cross-correlation coefficients for atoms  $i$  and  $j$ . Correlated fluctuations imply that atoms  $i$  and  $j$  fluctuate with the same period and phase, whereas anticorrelated fluctuations imply that atoms  $i$  and  $j$  fluctuate with the same period but opposite phase (Panel a). Visualisation of residues showing anticorrelated motions on the superposed structures of the AegOBP22 in ligand-free and ligand-bound conformations. Correlated motions are not shown (Panel b).

In the case of DmelOBP28a, the protein in liganded and free (apo) forms revealed a large conformational reorganization induced by ligand binding. This conformational change involved the 1st, 4th and 5th helices [47]. The residue-residue cross-correlation matrix shown in Figure 11, panel a is consistent with the structural observations as it identifies residues helices 1, 4 and 5 involved in anti-correlated movements



**Figure 11.** Pairwise residue cross-correlation matrix of DmelOBP28a (Panel a). Superposed structures of the DmelOBP28a in ligand-free and ligand-bound conformations. Visualisation of anticorrelated motions shown on the superimposed ligand-free and liganded conformers. Correlated motions are not shown (Panel b).

The normalised fluctuations and deformation energies of  $C\alpha$  atoms of AegOBP22 (PDB ID: 6oii) and DmelOBP28a (PDB ID: 6qq4, chain A), are included in Figure 7. Comparison of the fluctuation and deformation energy profiles against the corresponding conformers in ligand-free form, showed an increase in the magnitude of  $C\alpha$  atom mobilities in some of the interhelical regions (Appendix C, Figures 1 and 2). Analysis of a greater

number of models would be required to draw firm conclusions regarding the difference in fluctuations between ligand-free and ligand-bound forms.

#### 4. Discussion

From the preceding analysis it is obvious that, in most cases, the sequence identities between pairs of OBPs fall below the twilight zone and, therefore, it is not possible to draw conclusions regarding the possibility of divergent evolutionary relatedness of these proteins [37]. In terms of structure, the pairwise RMSD calculations and PC analysis result in a clustering of OBPs into subgroups that reflects, in part, phylogenetic classifications. The foregoing analysis identifies a conserved structural 'core' that represents 15–20% of the total sequence length of the OBPs. The 'core' residues, which include C<sub>2</sub>, C<sub>3</sub>, C<sub>5</sub> and C<sub>6</sub>, are located in the helical regions with the least structural variance (Figures 5 and 6). In particular 14 of the 17 'core' residues are found in the carboxy-terminal helix. The high deformation energies of the residues of these helical regions (Figure 8) are indicative of interactions with spatially proximal residues [24]. Further evidence of such interactions was derived from evolutionary coupling analysis. Evolutionary coupling interactions have a strong bearing on the overall architecture of the fold [31,48]. Generally, the similarity of 3D structure amongst the members of a particular fold is the result the constrained evolutionary boundaries ensuring conservation of function. This, in turn, implies the co-evolution of residues in spatial proximity in order to maintain energetically favourable interactions. Consequently, the high density of coupling interactions is indicative of 'hotspots' in the evolution of proteins and, as such, may be of functional importance [31]. From the preceding, it is not unreasonable to suggest that the residues adjacent to C<sub>2</sub>, C<sub>3</sub>, C<sub>5</sub> and C<sub>6</sub> may play an important role in ligand binding or recognition.

The covariance similarity analysis of the low-frequency motions of the OBPs (Figure 7) shows that their relatedness in terms of intrinsic dynamics mirrors that observed from the RMSD calculations and PC analysis (Figures 2 and 5). Further insights were obtained from the analysis of the motions of the individual proteins. The analysis showed that in some OBPs the motions are highly collective and thus may be significant in protein-ligand interactions (see examples in Section 3.4.3) [44,49]. Structural flexibility ensures the predisposition of proteins to attain alternative conformations. The pool of accessible conformations allows proteins to achieve their biological function in the presence of ligand binding, which, in turn, depends on changes in external conditions shifting the equilibrium between conformational states. Thus, although the examples described in Section 3.4.3 point to the importance of the intrinsic motions in determining the functional dynamics of the OBPs, it would be unwise to make broad generalisations in the absence of knowledge of the physiological milieu in which these proteins function.

#### 5. Conclusions

In this review, the "classical" insect OBPs were compared in terms of their structural features, and intrinsic dynamics. It was found that the structural differentiation of OBPs into subgroups is reflected in their low-frequency motions. The small dataset of available crystal structures was a limiting factor in the analysis. Without a sufficiently large dataset, it is not possible to make an in-depth evaluation of the relation between the conservation of structural dynamics to the evolution of sequence and structure in insect OBPs. Nevertheless, the study of OBPs, individually or collectively, by means of NMA using EN models can provide useful insights into the interplay between structure and dynamics at little computational cost. NMA using EN models has been applied to other heterogenous protein families of which the globin family is a case in point. The globin fold is an all- $\alpha$  fold made up of 6–7 helices that define the architecture of a well-defined haem pocket [50]. Although the divergent sequence evolution of the family induced changes in the relative disposition of helices, these had small effect on their packing. This is because the position of the haem group is essential for protein function [51]. In the case of OBPs, it is tempting to speculate that natural selection operated in the opposite direction by allowing promiscuous

changes in architecture of the fold in order to permit fast adaptation to the diverse chemical ecologies of insect.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/biom12020282/s1>. Archived files (1) evolutionary\_couplings.zip; (2) cross-correlation analysis.zip; Video S1.zip; Video S2.zip.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Jing-Jiang Zhou (State Key Laboratory of Green Pesticide and Agricultural Bio-engineering, Ministry of Education, Guizhou University, Huaxi District, Guiyang 550025, China) provided useful comments on Section 1, Section 2, Section 3.1, Section 3.2, Section 3.3 of the manuscript.

**Conflicts of Interest:** The author declares no conflict of interest.

### Appendix A

#### Multiple Structural Alignment of OBPs



Figure A1. Structural alignment of OBPs.

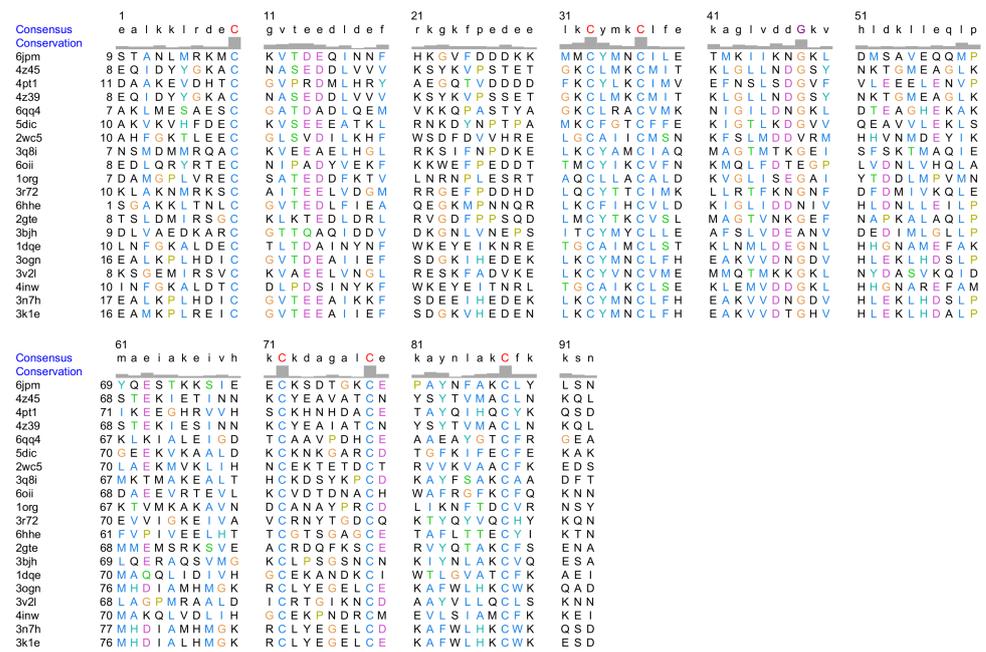


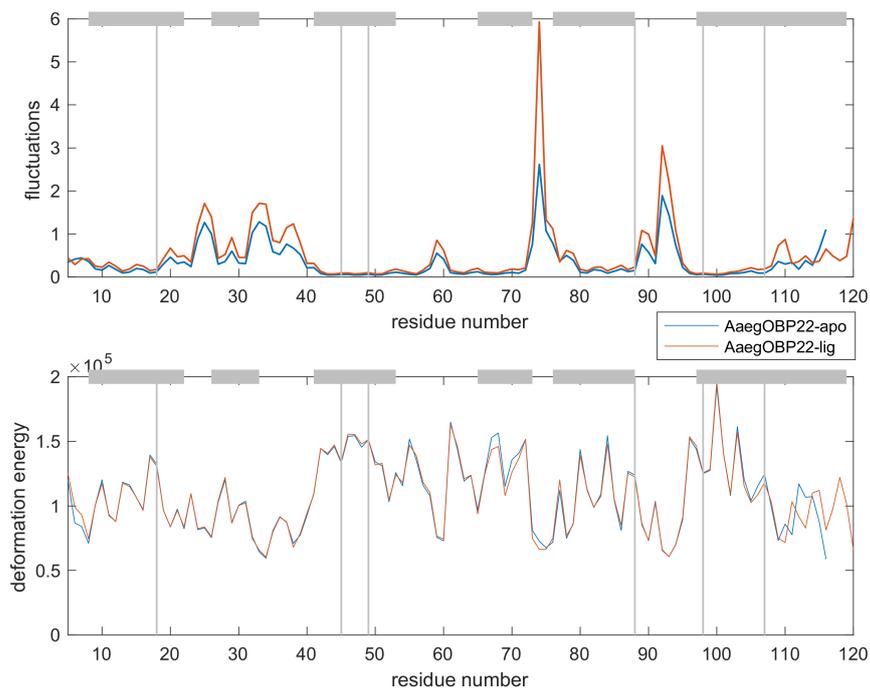
Figure A2. Same as Figure A1 after removing columns containing gaps.

### Appendix B

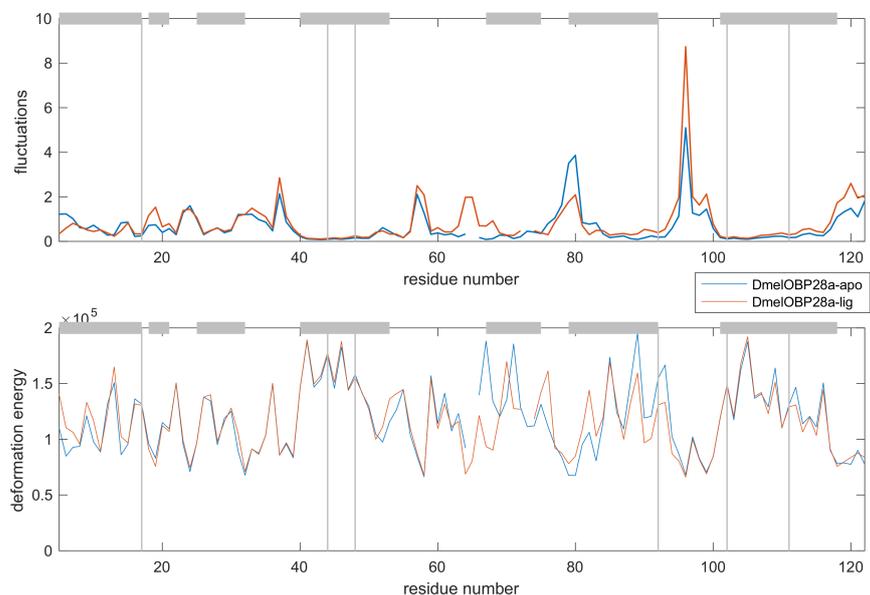
Table A1. Geometric features of the conserved ‘core’ helices.

Title 1	PDB ID	Angle (Degrees)	Distance (Å)
1	1dqe	83.4	7.5
2	1org	85.7	8.1
3	2gte	77.9	7.8
4	2wc5	83.0	6.8
5	3bjh	88.5	7.4
6	3k1e	87.3	7.6
7	3n7h	88.0	7.8
8	3ogn	88.1	7.8
9	3q8i	85.0	7.4
10	3r72	88.2	7.4
11	3v2l	81.0	7.7
12	4inw	81.5	7.3
13	4pt1	84.2	7.5
14	4z39	54.4	7.9
15	4z45	66.0	7.6
16	5dic	75.4	8.1
17	6hhe	83.8	7.8
18	6jpm	81.4	7.3
19	6oii	81.8	7.7
20	6qq4	78.9	7.7

## Appendix C



**Figure A3.** Fluctuation and deformation energy profiles of AgamOBP22 in ligand-free and ligand-bound conformations.



**Figure A4.** Fluctuation and deformation energy profiles of DmelOBP28a in ligand-free and ligand-bound conformations.

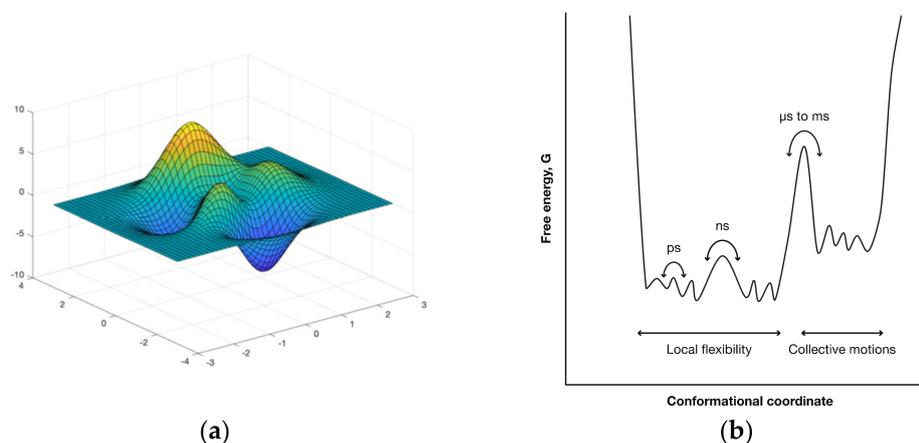
## Appendix D

### Normal Modes Analysis Using Elastic Network Models

The following is an explanatory note on some basic principles underpinning the use of Normal Mode Analysis.

Protein atoms are in thermal movement ( $E = kT$ ), and as a consequence protein structure changes as a function of time. Structural models (X-ray, NMR) are essentially snapshots of such movements. In physiological conditions, the protein shape changes

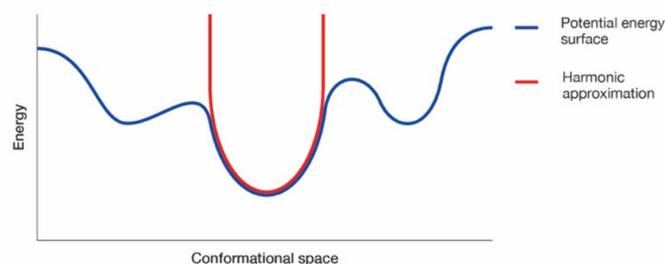
constantly. Consequently, the shape of binding site(s) may also change giving rise to more than one stable structures. The conformational space of a naturally occurring polypeptide is represented by a multidimensional potential energy surface. Briefly, the potential energy is a function of the positions  $R$ , where represents the coordinates of  $N$  atoms, each with coordinates  $(x,y,z)$ .



**Figure A5.** Representation of potential energy surface. A deep minimum in the potential energy surface indicates a stable structure (Panel a). One-dimensional cross-section through the multidimensional energy landscape of a protein. A conformational state is defined as a minimum in the energy surface. The difference in the Gibbs free energy between two states determines the rate of interconversion. Any change in the system, such as ligand binding, external conditions or mutation, will shift the equilibrium between states (Panel b).

The dynamics of proteins can be modelled means of molecular dynamics (MD). The application of MD assumes (i) a starting structure (existing X-ray or NMR model), (ii) assignment of a set of vectors for the coordinates of each atom in the structure, (iii) assignment of initial forces and velocities, and (iv) a timestep. Newton's equations are used to acquire a new position at each time interval, after which a new force is calculated for the determination of the next position, etc. The thermal energy allows the protein model to cross energy barriers thus sampling its entire conformational space. Given enough time, MD can sample the entire conformational space (Figure A5). MD provides detailed and realistic simulation of atomic movements but requires significant computational effort, particularly for nano-scale motions where demanding statistical analyses are required over a number of simulation trials.

In Normal Modes Analysis (NMA), protein motion is restricted to a single minimum by applying a harmonic approximation. At equilibrium, the thermal movement becomes harmonic and the structure is trapped in an energy minimum. When the harmonic approximation is applied at an energy minimum point, the potential surface is represented as a parabolic multidimensional surface (Figure A6).

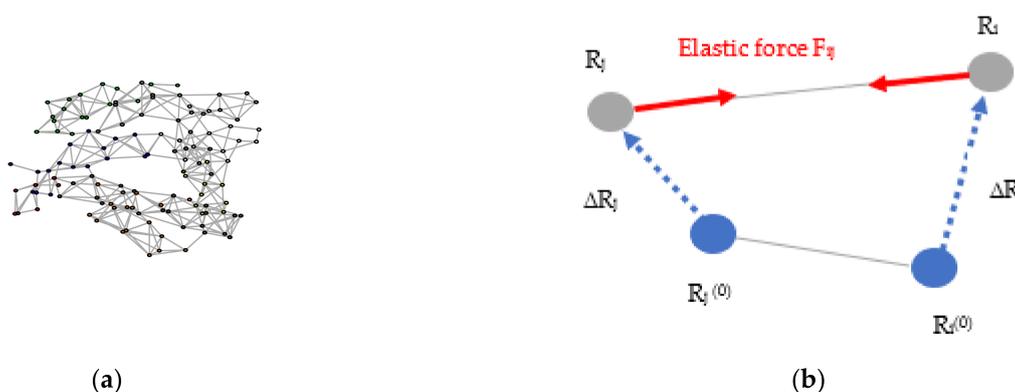


**Figure A6.** Energy landscape and harmonic approximation.

Based on this approximation, NMA is formulated and performed to study the dynamic properties of proteins. For a polyatomic system, such as protein, the Newton equations of motions can be expressed as in terms of the time evolution of a set of so-called mass-weighted Cartesian coordinates comprising of potential energy and kinetic energy terms. Assuming the harmonic approximation condition, normal mode analysis yields analytical solutions to the equations of motion. These solutions are given in terms of normal mode vectors (aka eigenvectors) that describe the direction and the displacement of each particle in the system with respect to the others, and eigenvalues that are the squares of the so-called normal mode vibrational frequencies. Assuming that the system undergoes some form of sinusoidal time evolution, and dividing the vibrational energy equally across all modes, the average amplitude of motion of any given mode scales with the inverse of the oscillation frequency. Thus, high frequency modes result in energetically more expensive localised displacements involving only a few directly bonded atoms. At low frequencies, the modes represent “collective” movements, that is they involve cooperative movements of large portions of the structure, such as domain motions or structural rearrangements.

NMA is computationally very expensive. For this reason, several coarse-grain methods have emerged in the form of elastic network models (ENMs), in which proteins are represented as assemblies of point-like particles joined by linear springs (Figure A7). In the simplest form, the backbone carbon atoms constitute the particles and the behaviour of the springs is linear-elastic, that is it obeys Hooke’s law. The connectivity of the particles is determined by calculating the distance between each pair of particles in the reference structure and using only particles falling within a cut-off distance, usually around 10 Å. NMA is then used to determine fluctuations or structural transitions in ENMs. The vibrational modes are orthogonal or normal and thus independent of each other. This means that the general motion of the system is described by the superposition of all modes.

The advantages and disadvantages of these models, as well as their application in structural biology have been reviewed extensively in references [34,52,53].



**Figure A7.** Representation of an Elastic Network Model of AgamOBP1 (a). Simplified diagram of the elastic network model (ENM). Each spring ( $i,j$ ) obeys Hooke’s law.  $R_i(0)$ ,  $R_j(0)$  are the cartesian coordinates of particles  $i$ ,  $j$  at equilibrium.  $R_i$ ,  $R_j$  are the particle coordinates following perturbation of the system. The elastic force  $F_i$  on particle  $i$  is the sum of forces  $F_{ij}$  exerted by all springs connected to the particle (not shown) and is proportional to its perturbation from the equilibrium length (b).

## Appendix E

### Principal Components Analysis

PCA is an unsupervised statistical technique. PCA forms the basis of multivariate data analysis [54]. The basic premise of the technique is that if some variable in the dataset are highly correlated, it is possible to replace them by a smaller number of variables referred to as Principal Components. PCA takes data points from an original variable space (the dataset) and maps these onto a PC space. This allows a drastic dimensionality reduction of the original variable space with little or no loss of information.

PCA performed on protein structure datasets takes as an input the set of coordinates in each structure, usually the coordinates of  $C_\alpha$  atoms, and computes the average position of each point in the structure as  $\langle r_i \rangle$ . The covariance for all pairs of points  $i, j$  is computed according to:

$$c_{i,j} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle$$

The brackets  $\langle \rangle$  denote averages over the entire dataset of structures. Often a correlation matrix is used instead of the covariance one. This ensures normalisation of the variable to prevent large atomic displacements from skewing the results. Assuming a data model of  $n$  members (e.g., OBPs) and each member consisting of  $k$  variables (e.g., coordinates of  $C_\alpha$  atoms), the data model can be represented by the correlation matrix (C) in the l.h.s of the equation below. Each of the observed variables  $x_{nk}$  can be decomposed to a complete set of orthogonal vectors (eigenvectors,  $t_n$ ), each with a corresponding eigenvalue  $p_n$  (variance) that characterises the displacement.

$$\begin{bmatrix} x_{11} & x_{1k} \\ x_{21} & x_{2k} \\ \dots & \dots \\ x_{n1} & x_{nk} \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_k \end{bmatrix} \times (p_1 \dots p_n)$$

Notice that for a cartesian coordinate system, the C matrix of the equation above is three dimensional ( $3n \times 3n$ ) symmetrical matrix. Matrix decomposition leads to a dimensionality reduction with  $3n$  eigenvectors. Component loading vectors are the result of multiplication of each eigenvector with the square root of the associated eigenvalue.

If each observation in the data model (row of the C matrix) is projected onto the direction of, say PC1, the score value of that observation is the distance from the origin up to the projection point. Considering a simplified example of four PCs, the number of resultant score plots would be six, one for each pairwise combination of PCs. Score plots are useful in that they allow separation of the data into clusters.

References for in-depth description of PCA: [55,56].

## References

1. Kaupp, U.B. Olfactory signalling in vertebrates and insects: Differences and commonalities. *Nat. Rev. Neurosci.* **2010**, *11*, 188–200. [CrossRef] [PubMed]
2. Leal, W.S. Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annu. Rev. Entomol.* **2013**, *58*, 373–391. [CrossRef] [PubMed]
3. Jin, X.; Ha, T.S.; Smith, D.P. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10996–11001. [CrossRef] [PubMed]
4. Bohbot, J.D.; Pitts, R.J. The narrowing olfactory landscape of insect odorant receptors. *Front. Ecol. Evol.* **2015**, *3*, 39. [CrossRef]
5. Vogt, R.G.; Riddiford, L.M. Pheromone binding and inactivation by moth antennae. *Nature* **1981**, *293*, 161–163. [CrossRef]
6. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
7. Sun, J.S.; Xiao, S.; Carlson, J.R. The diverse small proteins called odorant-binding proteins. *Open Biol.* **2018**, *8*, 180208. [CrossRef]
8. Rihani, K.; Ferveur, J.F.; Briand, L. The 40-Year Mystery of Insect Odorant-Binding Proteins. *Biomolecules* **2021**, *11*, 509. [CrossRef]
9. Pelosi, P.; Zhou, J.J.; Ban, L.P.; Calvello, M. Soluble proteins in insect chemical communication. *Cell Mol. Life Sci.* **2006**, *63*, 1658–1676. [CrossRef]
10. Sandler, B.H.; Nikonova, L.; Leal, W.S.; Clardy, J. Sexual attraction in the silkworm moth: Structure of the pheromone-binding-protein-bombykol complex. *Chem. Biol.* **2000**, *7*, 143–151. [CrossRef]
11. Maida, R.; Ziegelberger, G.; Kaissling, K.E. Ligand binding to six recombinant pheromone-binding proteins of *Antheraea polyphemus* and *Antheraea pernyi*. *J. Comp. Physiol. B* **2003**, *173*, 565–573. [CrossRef] [PubMed]
12. Lautenschlager, C.; Leal, W.S.; Clardy, J. *Bombyx mori* pheromone-binding protein binding nonpheromone ligands: Implications for pheromone recognition. *Structure* **2007**, *15*, 1148–1154. [CrossRef] [PubMed]
13. Drakou, C.E.; Tsitsanou, K.E.; Potamitis, C.; Fessas, D.; Zervou, M.; Zographos, S.E. The crystal structure of the AgamOBP1\*Icaridin complex reveals alternative binding modes and stereo-selective repellent recognition. *Cell. Mol. Life Sci.* **2017**, *74*, 319–338. [CrossRef] [PubMed]
14. Biessmann, H.; Nguyen, Q.K.; Le, D.; Walter, M.F. Microarray-based survey of a subset of putative olfactory genes in the mosquito *Anopheles gambiae*. *Insect Mol. Biol.* **2005**, *14*, 575–589. [CrossRef]

15. Mastrobuoni, G.; Qiao, H.; Iovinella, I.; Sagona, S.; Niccolini, A.; Boscaro, F.; Caputo, B.; Orejuela, M.R.; Torre, A.d.; Kempa, S.; et al. A Proteomic Investigation of Soluble Olfactory Proteins in *Anopheles gambiae*. *PLoS ONE* **2013**, *8*, e75162. [[CrossRef](#)]
16. Tegoni, M.; Campanacci, V.; Cambillau, C. Structural aspects of sexual attraction and chemical communication in insects. *Trends Biochem. Sci.* **2004**, *29*, 257–264. [[CrossRef](#)]
17. Fox, N.K.; Brenner, S.E.; Chandonia, J.M. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.* **2014**, *42*, D304–D309. [[CrossRef](#)]
18. Zhou, J.J. Odorant-binding proteins in insects. *Vitam. Horm.* **2010**, *83*, 241–272. [[CrossRef](#)]
19. Brito, N.F.; Moreira, M.F.; Melo, A.C.A. A look inside odorant-binding proteins in insect chemoreception. *J. Insect Physiol.* **2016**, *95*, 51–65. [[CrossRef](#)]
20. Venthur, H.; Zhou, J.J. Odorant Receptors and Odorant-Binding Proteins as Insect Pest Control Targets: A Comparative Analysis. *Front. Physiol.* **2018**, *9*, 1163. [[CrossRef](#)]
21. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–972. [[CrossRef](#)] [[PubMed](#)]
22. Bahar, I.; Lezon, T.R.; Bakan, A.; Shrivastava, I.H. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chem. Rev.* **2010**, *110*, 1463–1497. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, S.; Li, H.; Krieger, J.M.; Bahar, I. Shared Signature Dynamics Tempered by Local Fluctuations Enables Fold Adaptability and Specificity. *Mol. Biol. Evol.* **2019**, *36*, 2053–2068. [[CrossRef](#)] [[PubMed](#)]
24. Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **1998**, *33*, 417–429. [[CrossRef](#)]
25. Na, H.; Jernigan, R.L.; Song, G. Bridging between NMA and Elastic Network Models: Preserving All-Atom Accuracy in Coarse-Grained Models. *PLoS Comput. Biol.* **2015**, *11*, e1004542. [[CrossRef](#)]
26. Tiwari, S.P.; Reuter, N. Similarity in Shape Dictates Signature Intrinsic Dynamics Despite No Functional Conservation in TIM Barrel Enzymes. *PLoS Comput. Biol.* **2016**, *12*, e1004834. [[CrossRef](#)]
27. Skjaerven, L.; Yao, X.Q.; Scarabelli, G.; Grant, B.J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinform.* **2014**, *15*, 399. [[CrossRef](#)]
28. Grant, B.J.; Rodrigues, A.P.; ElSawy, K.M.; McCammon, J.A.; Caves, L.S. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695–2696. [[CrossRef](#)]
29. Grant, B.J.; Skjaerven, L.; Yao, X.Q. The Bio3D packages for structural bioinformatics. *Protein Sci.* **2021**, *30*, 20–30. [[CrossRef](#)]
30. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [[CrossRef](#)]
31. Marks, D.S.; Colwell, L.J.; Sheridan, R.; Hopf, T.A.; Pagnani, A.; Zecchina, R.; Sander, C. Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* **2011**, *6*, e28766. [[CrossRef](#)] [[PubMed](#)]
32. Sheridan, R.; Fieldhouse, R.J.; Hayat, S.; Sun, Y.; Antipin, Y.; Yang, L.; Hopf, T.; Marks, D.S.; Sander, C. EVfold.org: Evolutionary Couplings and Protein 3D Structure Prediction. *bioRxiv* **2015**, 021022. [[CrossRef](#)]
33. Hopf, T.A.; Green, A.G.; Schubert, B.; Mersmann, S.; Scharfe, C.P.I.; Ingraham, J.B.; Toth-Petroczy, A.; Brock, K.; Riesselman, A.J.; Palmedo, P.; et al. The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **2019**, *35*, 1582–1584. [[CrossRef](#)]
34. Fuglebakk, E.; Tiwari, S.P.; Reuter, N. Comparing the intrinsic dynamics of multiple protein structures using elastic network models. *Biochim. Biophys. Acta* **2015**, *1850*, 911–922. [[CrossRef](#)] [[PubMed](#)]
35. Ma, J.; Wang, S. Chapter Five—Algorithms, Applications, and Challenges of Protein Structure Alignment. In *Advances in Protein Chemistry and Structural Biology*; Donev, R., Ed.; Academic Press: Cambridge, MA, USA, 2014; Volume 94, pp. 121–175.
36. Konagurthu, A.S.; Whisstock, J.C.; Stuckey, P.J.; Lesk, A.M. MUSTANG: A multiple structural alignment algorithm. *Proteins* **2006**, *64*, 559–574. [[CrossRef](#)]
37. Chung, S.Y.; Subbiah, S. A structural explanation for the twilight zone of protein sequence homology. *Structure* **1996**, *4*, 1123–1127. [[CrossRef](#)]
38. Khor, B.Y.; Tye, G.J.; Lim, T.S.; Choong, Y.S. General overview on structure prediction of twilight-zone proteins. *Theor. Biol. Med. Model.* **2015**, *12*, 15. [[CrossRef](#)]
39. Fuglebakk, E.; Echave, J.; Reuter, N. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics* **2012**, *28*, 2431–2440. [[CrossRef](#)]
40. Ishida, Y.; Ishibashi, J.; Leal, W.S. Fatty acid solubilizer from the oral disk of the blowfly. *PLoS ONE* **2013**, *8*, e51779. [[CrossRef](#)]
41. Wang, J.; Murphy, E.J.; Nix, J.C.; Jones, D.N.M. *Aedes aegypti* Odorant Binding Protein 22 selectively binds fatty acids through a conformational change in its C-terminal tail. *Sci. Rep.* **2020**, *10*, 3300. [[CrossRef](#)]
42. Tiwari, S.P.; Fuglebakk, E.; Hollup, S.M.; Skjaerven, L.; Cragnolini, T.; Grindhaug, S.H.; Tekle, K.M.; Reuter, N. WEBnm@ v2.0: Web server and services for comparing protein flexibility. *BMC Bioinform.* **2014**, *15*, 427. [[CrossRef](#)] [[PubMed](#)]
43. Hollup, S.M.; Salensminde, G.; Reuter, N. WEBnm@: A web application for normal mode analyses of proteins. *BMC Bioinform.* **2005**, *6*, 52. [[CrossRef](#)] [[PubMed](#)]
44. Ichiye, T.; Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **1991**, *11*, 205–217. [[CrossRef](#)] [[PubMed](#)]
45. Pesenti, M.E.; Spinelli, S.; Bezirard, V.; Briand, L.; Pernollet, J.-C.; Campanacci, V.; Tegoni, M.; Cambillau, C. Queen Bee Pheromone Binding Protein pH-Induced Domain Swapping Favors Pheromone Release. *J. Mol. Biol.* **2009**, *390*, 981–990. [[CrossRef](#)]

46. Romanowska, J.; Nowinski, K.S.; Trylska, J. Determining Geometrically Stable Domains in Molecular Conformation Sets. *J. Chem. Theory Comput.* **2012**, *8*, 2588–2599. [[CrossRef](#)]
47. Gonzalez, D.; Rihani, K.; Neiers, F.; Poirier, N.; Fraichard, S.; Gotthard, G.; Cheretemps, T.; Maibeche, M.; Ferveur, J.F.; Briand, L. The Drosophila odorant-binding protein 28a is involved in the detection of the floral odour  $\alpha$ -ionone. *Cell Mol. Life Sci.* **2020**, *77*, 2565–2577. [[CrossRef](#)]
48. Bakan, A.; Dutta, A.; Mao, W.; Liu, Y.; Chennubhotla, C.; Lezon, T.R.; Bahar, I. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **2014**, *30*, 2681–2683. [[CrossRef](#)]
49. Bahar, I.; Lezon, T.R.; Yang, L.W.; Eyal, E. Global dynamics of proteins: Bridging between structure and function. *Annu. Rev. Biophys.* **2010**, *39*, 23–42. [[CrossRef](#)]
50. Aronson, H.E.; Royer, W.E., Jr.; Hendrickson, W.A. Quantification of tertiary structural conservation despite primary sequence drift in the globin fold. *Protein Sci.* **1994**, *3*, 1706–1711. [[CrossRef](#)]
51. Maguid, S.; Fernandez-Alberti, S.; Ferrelli, L.; Echave, J. Exploring the common dynamics of homologous proteins. Application to the globin family. *Biophys. J.* **2005**, *89*, 3–13. [[CrossRef](#)]
52. Togashi, Y.; Flechsig, H. Coarse-Grained Protein Dynamics Studies Using Elastic Network Models. *Int. J. Mol. Sci.* **2018**, *19*, 3899. [[CrossRef](#)] [[PubMed](#)]
53. Yamato, T.; Laprevote, O. Normal mode analysis and beyond. *Biophys. Physicobiol.* **2019**, *16*, 322–327. [[CrossRef](#)] [[PubMed](#)]
54. Erickson, L.; Byrne, T.; Johansson, E.; Tygg, J.; Vikstrom, C. *Multi- and Megavariate Data Analysis Basic Principles and Applications*; Umetrics Academy: Malmö, Sweden, 2013; Volume 1, p. 521.
55. Howe, P.W. Principal components analysis of protein structure ensembles calculated using NMR data. *J. Biomol. NMR* **2001**, *20*, 61–70. [[CrossRef](#)] [[PubMed](#)]
56. David, C.C.; Jacobs, D.J. Principal component analysis: A method for determining the essential dynamics of proteins. *Methods Mol. Biol.* **2014**, *1084*, 193–226. [[CrossRef](#)] [[PubMed](#)]