IN SILICO STRUCTURE PREDICTION OF *GLOSSINA MORSITANS MORSITANS* ODORANT BINDING PROTEINS

E. D. Ndinya, S. G. Nyanjom and J. K. Kinyua
Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Abstract

Glossina morsitans morsitans is the primary vector for trypanosome parasite which is causative agent for sleeping sickness in human and nagana in animals, conditions that are prevalence in sub Saharan Africa. In this study functional annotation and structure prediction of *G. M. morsitans* odorant binding proteins was done by BLAST searches against NCBI and PDB databases respectively. Multiple sequence alignment was performed by clustalW while genetic relatedness of *G. m. morsitans* with mosquitoes (*Anopheles gambiae*, *Aedes aegypti* and *Culex quinquefasciatus*) and fruit fly (*Drosophila melanogaster*) was determined based on PHYML. The 3D-structure prediction was determined and viewed using Swiss model and Swiss-Pdb Viewer program respectively. The GmmOBP9 had the highest percentage identity of 77.9% and e-value of 1.42e-10 with DroOBP83b as its ortholog. GmmOBP10 recorded the highest percentage identity among the mosquitoes studied having 60.23%, 57.65% and 62.35% for *Anopheles*, *Aedes* and *Culex* respectively. The lowest percentage identity of 27.4% and e-value of 0.04 was recorded for DroOBP56i by GmmOBP17 while GmmOBP3 recorded the lowest percentage identity among the mosquitoes (*Anopheles* and *Aedes*) with 19.50% and 24.06% respectively. The Gmm-OBPs had amino acid sequence length ranging from 88 to 240 with GmmOBP7 being the longest (240 aa with MW of 28.3kda and pI 5.78). GmmOBP20 had the lowest MW of 10.3 kda. The GmmOBPs had a theoretical pI range of 4.41 to 9.30. Multiple sequence alignment revealed the six conserved cysteine while phylogenetic studies indicate that GmmOBPs are closely related to *Drosophila* OBPs. Structural prediction of the GmmOBPs showed presence of between four to six helices with multiple alignments with respective templates confirming the location of the conserved six cysteines. This study predicts the three dimensional structure of *G. m. morsitans* OBPs and open avenues for functional studies as they form potential targets for control of tsetse vectors.

Key words: *Glossina*, olfaction, odorant binding protein

1.0 Introduction

Tsetse flies (*Glossina*) are primary vectors of trypanosome parasites which cause Human African Trypanosomiasis (HAT), also known as sleeping sickness (Brun *et al*., 2010) and Animal African Trypanosomiasis (AAT), also known as nagana (Batista *et al.*, 2009). The genus *Glossina*, contains about 30 living taxa, 22 species and 8 subspecies which are grouped according to the habitat they occupy namely *Morsitans*, *Palpalis* and *Fusca*. *Morsitans* are the Savannah flies and examples include *Glossina morsitans morsitans* (vectors of HAT), *G. pallidipes and G. austeni* (vectors of AAT) (Welburn *et al*., 2001). *Palpalis* are the riverine tsetse flies inhabiting land masses near water bodies and examples include *G. fuscipes fuscipes, G. palpalis gambiensis* and *G. tachinoides* (vectors of HAT) (Bouyer *et al*., 2005). The forest flies are the *Fusca* group and some examples include *G. fusca fusca* (vector of nagana) (Leak *et al*., 1991), *G. tabaniformis* and *G. longipennis* which are not considered as important vectors of Trypanosomiases (Makumi *et al*., 2000).

The trypanosome parasite is transmitted through bite from infected *Glossina* to the animal host. It multiplies at the site of the bite, followed by entry to the lymphatic system and the blood stream, through which they reach other tissues and organs including central nervous system. In the animal the parasite is covered by a dense monolayer of identical glycoproteins which protect it from lysis by the host immune cells (Borst and Fairlamb, 1998). Thus, the parasite is able to evade the mammalian host humoral immune response and proliferate until new surface antigen coat is recognized by a new generation antibodies of immunoglobulin (Vanhamme *et al*., 2001). The blood stream forms of the parasite transform into procyclic trypomastigotes in the fly’s midgut and multiply by binary fission. Finally, the trypomastigotes leave the midgut and develop into epimastigotes before migrating to the fly's salivary glands accompanied by continued reproduction by binary fission (Tyler *et al*., 2001). The cycle then repeats itself when the infected tsetse fly bites another host. Examples of trypanosomes transmitted by *Glossina* species includes *Trypanosome brucei* transmitted by *G. morsitans, Trypanosoma rhodesience* transmitted by *G. pallidepes and G. fuscipes quanzensis* (Welburn *et al*., 2001).

Trypanosomiasis (HAT and AAT) are prevalent in Sub-Saharan Africa affecting both human beings and livestock. The consequences are that AAT has a negative impact on human nutrition and livelihood in terms of losses in milk production, hides, meat, blood and treatment costs (FAO, 1993). In human beings, HAT account for losses that can be manifested inform of bad health, death and loss of man hours. World Health Organization (WHO) estimate that about 66 million people are exposed to the risk of contracting HAT while about 45 million cattle are at risk of being infected by AAT. Economic losses are estimated to the tune of US$ 4.75 billion and US$ 1.2 billion per year in both agricultural and cattle productions respectively (WHO, 2000). This makes control of HAT and AAT vectors to be a research priority, a task taken by African Heads of State and Governments in attempt to eradicate tsetse flies from the African Continent leading to establishment of Pan African Tsetse and Trypanosomiasis Eradication Campaign (Kabayo, 2002).

Olfaction plays a critical role in tsetse flies in identification of hosts, breeding mates and larviposition. This role is mediated by olfactory proteins which include odorant binding proteins (OBPs), pheromone binding proteins (PBPs), Chemosensory proteins (CSPs) and Odorant receptors (Ors) and are localized in the tsetse fly antennae (Pelosi, 1996). It is postulated that odors from external environment get in through antennal pores and are picked by soluble olfactory proteins (OBPs, PBPs, CSPs), then transported through the sensillium lymph to the Ors located within olfactory receptor neurons (ORNs). Transduction of the odor message through axons to antennal lobe and higher brain centers lead to processing and decoding of the odor information (Rutzler and Zwiebel, 2005). With the sequencing of *Glossina morsitans morsitans* genome and identification of OBPs (Liu *et al*., 2011), detailed study on molecular structure and functional expression of OBPs is important as they play a key and initial role in the peri-receptor events thus are potential targets in the genetic manipulation of the tsetse fly. This study propose to examine the genetic relatedness of *G. m. morsitans* OBPs with other identified insects OBPs.

2.0 Materials and Methods

2.1 Bioinformatics and Sequence Retrieval

The 33 sequences of *G. m morsitans* OBPs were retrieved from NCBI database (www.ncbi.nlm.nih.gov/entrez/viewer.fcgi) by general search query using Entrez and opened in FASTA format. The FASTA sequences were saved with both gene name and accession numbers**.** Functional annotation was done by blast search against *D. melanogaster* genome (version FB2012_06) and mosquitos' genomes (*An. gambiae*, *Ae. aegypti* and *Cu. quinquefasciatus*) (Vb-2012-12) using Blastp (Altschul *et al*., 1990). Results that produced the highest percentage similarity with low E score (<0.0) were taken to be orthologs. The 33 Gmm-OBPs were analyzed for molecular weight (MW) and isoelectric point (pI) using protpharm tool at Expasy (http: www.expasy.ch/tools/ protparam) with default parameters. Multiple sequence alignment of *G. m morsitans* OBPs with selected orthologs from blast results was carried out using clustalW, the sequences were loaded, complete alignment carried out. (Larkin *et al*., 2007).

2.2 Phylogenetic Analysis

Genetic relatedness of *G. m. morsitans* with mosquitoes (*An. gambiae*, *Ae. aegypti* and *Cu. quinquefasciatus*) and fruit fly (Drosophila *melanogaster*) was determined based on PHYML. Evolutionary track was inferred using the neighbor- joining method (Saitou and Ni, 1987).

2.3 Prediction of 3D Structure

The 3D structure of the 33 *G. m morsitans* OBPs were predicted using Swiss-Model at Expasy hosted by Swiss PDB (http://www.swissmodel.expasy.org). The OBP sequences were used to interrogate Protein Data Bank (PDB) to ensure that the 3-dimensional structure of the OBP sequences were not available in PDB (http://www.rcsb.org/orgpdb/home/home.do). The protein sequences of *G. m morsitans* OBPs obtained from NCBI were edited and uploaded into Swiss model workspace. The program first identifies homologous proteins with known 3D structures using PSI-BLAST. The identified homologous sequences form the template that is aligned with the target sequence (*G. m morsitans* OBPs). The alignment extract geometrical restraints (dihedral angles and distances) for corresponding atoms between query and template sequences, performs 3D construction of the protein by using a distance geometry approach and finally predict the structure (Combet *et al*., 2001) which was opened and viewed with SPDB Viewer (http://www.spdbv.vital-it.ch). The Swiss-Model followed the flowchart in figure 1. The generated models and amino acid sequence alignments of target and template proteins were displayed using Swiss-Pdb Viewer programme "Deep-View" (Guex *et al*., 1997).

3.0 Results

3.1 Functional Analysis of GMM-OBPS

A total of 33 OBPs from *G. m. morsitans* were retrieved from GenBank database. The GmmOBP9 had the highest percentage identity of 77.9% and e-value of 1.42e-10 with DroOBP83b. Likewise GmmOBP10 recorded the highest percentage identity among the mosquitoes studied having 60.23%, 57.65% and 62.35% for *Anopheles*, *Aedes* and *Culex* respectively (Table 1). The lowest percentage identity of 27.4% and e-value of 0.04 was recorded in DroOBP56i by GmmOBP17; on the other hand GmmOBP3 recorded the lowest percentage identity among the mosquitoes (*Anopheles* and *Aedes*) with 19.50% and 24.06% respectively. GmmOBP3 never recorded any hit with *Culex* while GmmOBP4 recorded the lowest percentage identity of 20.17% and e-value of 5e-07 with CulexOBP28 (Table 1).

The Gmm-OBPs had amino acid sequence length ranging from 88 to 240 with GmmOBP7 being the longest having 240 aa with MW of 28316.4 and pI 5.78. The shortest Gmm-OBP was GmmOBP16 while GmmOBP20 had the lowest MW of 10.3 kda. Both GmmOBP 18 and GmmOBP19 had the same length of aa and pI while GmmOBP8 and GmmOBP9 had similar MW and pI but same length (Table 2).

3.2 Multiple Sequence Alignment

Multiple sequence alignment of GmmOBPs with the *Drosophila* and mosquito ortholgs revealed that OBP sequence are quite diverse and only conserved at certain residues (Figure 2). The OBPs had six conserved cysteine residues.

3.3 Phylogenetic Analysis of *G.M.MORSITANS* OBPS

Phylogenetic relationship revealed that GmmOBPs (GmmOBP3, GmmOBP22 and GmmOBP11) were closely related to *Drosophila* orthologs than the mosquitoes with percentage similarity of 94% (Figure 3). The exceptions were GmmOBP9 and GmmOBP10 that clustered with AngamOBP2 with percentage relatedness of 91%, DmelOBP99c was noted to have clustered with CqiOBP99a and AeayOBP99c with percentage relatedness of 72%. Another recording reveals that GmmOBP15 clustered with AeayOBP56a with percentage similarity of 63.

3.4 Insilico Structure Prediction

The 31 GmmOBPs out of the 33 GmmOBPs were modeled. Modelling of GmmOBPs was based on varied protein templates with exception of some that were modeled based on the shared template. The templates were from protein data bank. GmmOBP29 and GmmOBP30, was never modeled. GmmOBP9 and GmmOBP12 was modeled based on 2erb which had the highest percentage (67.21%). The GmmOBP14, 10, 8A, 12, 26 and 28 models had percentage similarity between 45.61%-67.21% while the rest of GmmOBPs had percentage similarity as low as 20%. GmmOBP3, GmmOBP8A, GmmOBP21, GmmOBP25, GmmOBP19 and GmmOBP21 3D structure was modeled based on 3n7h template with GmmOBP8 having the highest percentage of 55.2% among the group, while 314a template was a platform for GmmOBP1, GmmOBP2, GmmOBP5, and GmmOBP22 3D structure (PDB: 314 chain A). GmmOBP13, GmmOBP14, GmmOBP16 was modeled based on 3v2l template (PDB: 3v2l chain A), 2WcJ [PDB: 2wcj Chain: A] template was used to modeled GmmOBP6 and GmmOBP10 (Zhou et al., 2009) while 30gn [PDB: 3ogn Chain: A] template was used as a template for GmmOBP10 (62.35%) (Mao *et al*., 2010). (Table 3)

4.0 Discussion and Conclusion

G.m.morsitans OBPs molecular structure and their functional expression is important as they play an important role in insect olfaction by mediating interactions between odorants and odorant receptors. They are crucial in feeding, mating and oviposition of the insects.

In the present study, Functional annotation of sequences from *G.m.morsitans* OBPs were analysed with reference to mosquitoes and fruitfly was done by blast search against *D. melanogaster* genome (version FB2012_06) and mosquitos' genomes (*An. gambiae*, *Ae. aegypti* and *Cu. quinquefasciatus*) (Vb-2012-12) using Blastp. Multiple sequence alignment through Clustal W of these insects sequences indicated 6 conserved cysteine a characteristic of odorant binding proteins. The alignment done between the GmmOBPs and the template used to predict the 3D-structure revealed these conserved cysteine. The conserved cysteine in position one in all templates (3v2La, 2erb, 1oohA, 3ogn, 3n7h and 3r1vA) except 2wcj were not completely aligned while the conserved cysteine in position

six in the alignment between the sequences and all the templates sequences exept the sequence of 3v2lA were completely aligned.The cysteines contain sulfide play arole in the folding of the protein in 3-Dimension structure via sulfide bond.
OBPs (OBP14, OBP13, OBP4 and OBP16) were modeled based on *Anopheles gambie* OBP20 (PDB: 3V2L chain A) this supported the fact modeling of the structure was based on the structure whose orthologue had highest percentage similarity with low E score (<0.0).

GmmOBP10 recorded the highest percentage identity among the mosquitoes studied having 60.23%, 57.65% and 62.35% for *Anopheles*, *Aedes* and *Culex* respectively. Mosquitoes being blood feeders it can be inferred that GmmOBP10 could have a role blood feeding whereas the GmmOBP9 had the highest percentage identity of 77.9% and e-value of 1.42e-10 with DroOBP83b, *Drosophila melanogaster* requires this OBP to sense the fruit in the environment, the GmmOBP9 could also play a role in sensing of the presence of food.

The protpharm analysis of GmmOBPs shows GmmOBP 4, OBP5, OBP14, OBP15 had a theoretical Isoelectric point between 8.20-9.30 this could translated to the fact that these OBPs work best in alkaline environment, GmmOBP1 was unique among the 22 GmmOBPs since it had a neutral pH of 7.50 indicating a working efficiency in a neutral environment while the remaining GmmOBPs recorded a pH of (4.41-6.27) indicating acidic working environment, hence it can be clearly deduced that majority of GmmOBPs work best in alkaline environment.

Phylogenetic analysis revealed that GmmOBPs were closely related to *Drosophila* orthologs than the mosquitoes these supported the facts on the ground where tsetsefly cluster together with fruitfly a situation that can be attributed to the facts that the 2 insects are closely related.

The conserved cysteines were present in the alignment of the sequences between the templates and the GmmOBPs that confirmed a characteristic of odorant binding, as they contain sulfide stabilizing disulfide bond in the 3-D structure. Hence like other proteins the odorant binding proteins have 3-D structure. The phylogenetic tree indicated the relationship between GmmOBPs and *D. melanogaster* genome (version FB2012_06) and mosquitoes genomes (*An. gambiae, Ae. aegypti and Cu. quinquefasciatus*) shows that the insects had a common ancestor.

Acknowledgement
To the Almighty God with thanks giving for this far He has brought me indeed His love, mercies endures forever and His grace is ever sufficient. The journey to this day has been long though with destiny helpers offering spiritual, intellectual material support and guidance its genesis looks like yesterday. My sincere gratitude goes to my supervisor Dr. Ger Nyanjom, a mentor, friend and my first supervisor for identification of the project and guiding in report writing, without forgetting his office and laboratory at ICEPE that housed much of the research practicals.

I am honour-bound to Dr. Kinyua Kang'ethe for report editing, project guidance, project materials. Much encouragement, advices and critics came from my lecturers these included Prof. Magiri, Dr. Maina, Mrs. Mutoro and my fellow students. My thanks also go to Sunday school children's who always encourage me to live my talk as a God minister for the Bible says we walk by faith and not by sight a true sign of sons of God. Glory to the most high God.

References

Batista, J. S., Oliveira, A. F., Rodrigues, C. M., Damasceno, C. A., Oliveiria, I. R., Alves, H. M., Paiva, E. S., Brito P. D., Medeiros , J. M., Teixeira ,M. M. and Rodrigues A. C. (2009). Infection by *Trypanosoma vivax* in goats and sheep in the Brazilian semiarid region. *Vetinery parasitology,* **165**, pp 131-135.

Borst, P. and Fairlamb, A. H. (1998). Surface receptors and transporters of *Trypanosoma brucei*. *Annual Review Microbiology* **52,** pp 745-778.

Brun, R., Blum, J., Chappuis, F. and Burri, C. (2010). Human African trypanosomiasis. *The Lancet,* **375**, pp 148 – 159.

Gao, Q. and Chess, A. (1999). Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* **60**, pp 31–39.

Hallem, E. A., Ho, M.G. and Carlson, J.R. (2004). The molecular basis of odor coding in the *Drosophila* antenna. *Cell,* **117**, pp 965–979.

Leak, S. G., Colardelle, C., D'Ieteren, G., Dumont, P., Feron, A., Jeannin, P., Minengu, M., Mulungu, M., Ngamuna, S. and Ordner, G. (1991). *Glossina fusca* group tsetse as vectors of cattle trypanosomiasis in Gabon and Zaire. *Medical and Veterinary Entomology* **5,** pp 111-120.

Logan, J.G., Cook, L. J., Mordoe, A. J. and Kline, L. D. (2010). Ecology and control of vector borne diseases. *olfaction in vector-vector host interaction,* **2,** pp 217-246.

Mowatt, M. R. and Clayton, C. E. (1987). Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. *Molecular cell biology,* **7**, pp 2838–2844.

Smart, R., Kiely, A., Beale, M., Vargas, E., Carraher, C., Kralicek, A. V., Christie, D. L., Chen, C. and Newcomb, R. D. (2008). *Drosophila* odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect Biochemistry and Molecular Biology* **38**, pp 770–780.

Shanbhag, S. R., Muller B. and Steinbrecht, R. A. (1999) Atlas of olfactory organs of *Drosophila melanogaster*—1.Types, external organization, innervations and distribution of olfactory sensilla. *Insect Morphology and Embryology A,* **28**, pp 377–397.

Vanhamme, L., Pays, E., McCulloch, R. and Barry**,** J. D. (2001). An update on antigenic variation in African trypanosomes. *Trends Parasitology,* **17**, pp 338-343.

Touhara, K. and Vosshall, L. B. (2009).Sensing odorants and pheromones with chemosensory receptors. *Annual Review Physiology,* **71**, pp 307–332.

Yao, C. A. and Carlson, J. R. (2010). Role of G-proteins in odor sensing and CO2-sensing neurons in *Drosophila*. *Annual review of Neuroscience,* **30**, pp 4562–4572.

Abbeele, V. J., Claes, Y., Bockstaele, D.V., Le Ray, D. and M. Coosemans, M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology,* **118**, pp 469-478.

Vickerman, K., Tetley, L., Hendry, A. and Turner, C.M. (1988). Biology of African trypanosomes in the tsetse fly. *Biology of Cell,* 64, pp 109-119.
FAO (1993). The tsetse fly and its effects on agriculture in sub-saharan Africa, FAO senior officers panel. Rome, Italy.

WHO (2002). Pan African tsetse and trypanosomiasis eradication campaign, WHO technical report. Geneva, France.

WHO (1993).Control and surveillance of African Trypanomiasis. WHO expert committee, Swiss franscisco. France.

WHO (1998). Control of African Trypanomiasis. *WHO technical report series* **881**, pp 1-27.

Bouyer, L., Guerrini, J., Cesar, S., Dela Rockque and Cuisane, D. (2005). Aphyto-Sociological analysis of riverine tsetse flies in Burkina faso. *Medical vertinary*, 19, pp 372-378.

Makumi, J. N., Stevenson, P. and Green, C. H. (2000). Control of *Glossina Longipennis* by insecticide- trated targets at Galana Ranch, Kenya, and confirmation of the role of Glossina *Longipennis* as a vector of cattle trypanomiasis. *Bulletin of entomological Research*, **90**, pp 397-406.

Vogt, R. G. (2002). Odorant binding protein homologues of the malaria *Anopheles gambiae*; possible orthlogues of the OS-E and OS-F OBPs of *Drosophila melanogaster*. *Journal Chemical Ecology,* **28**, pp 2371-2376.

Wang, C., Zhang, J., Tobe, S. S., and Bendena, W. G. (2012). Defining the contribution of select europeptides and their receptors in regulating sesquiterpenoid biosynthesis by *Drosophila melanogaster* ring gland/corpus allatum through RNAi analysis. *General comparative endocrinology*, **176** (3). Pp 347-353.

Hallem, E. A., Ho, M. G., Carlson, J. R. (2004). The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117**(7), pp 965-979.

Bos, J., Prince, D., Pitino, M., Maffei, M. E., and Win, J. (2010) A Functional Genomics Approach Identifies Candidate Effectors from the Aphid Species *Myzus persicae* (Green Peach Aphid). *PLoS Genet,* **6**(11): e1001216. doi:10.1371/journal.pgen.1001216.

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling**.** *Electrophoresis,* **18**, pp 2714-2723.

http://www.rcsb.org/orgpdb/home/home.do.

http://www.swissmodel.expasy.org.

Schillinger, T., Lisfi, M., Chi, J., Cullum, J. and Zingler, N. (2012). Analysis of a comprehensive dataset of diversity generating retroelements generated by the program DiGReF. *BMC Genomics,* **13**, pp 430.

Table 1: Functional annotation of Glossina morsitans morsitans OBPs

G. morsitans **OBP**	**Accession Number**	**Best match to *D. melanogaster* (Identities, E-value)**	**Best match to *An. Gambie* (Identities, E-value)**	**Best match to *Ae. aegypti* (Identities, E-value)**	**Best match to *C. quinquefasciatus* (Identities, E-value)**
OBP1	CBA11305.1	OBP44a (FBpp00087892) (62.1%, 9.39e-46)	OBP9 (AGAP000278) (48.91%,5e-33)	OBP99c (AAEL005772) (44.12%, 1e-31)	OBP99a (CPIJO17326) (48.53%, 1e-34)
OBP2A	CBA11306.1	OBP99b (FBpp0305400) (53.4%,5.89e-39)	OBP9 (AGAP000278) (44.35%, 4e-35)	OBP99c (AAEL005772) (38.16%, 4e-26)	OBP99c (CPIJ017326) (39.47%, 4e-26)
OBP3	CBA11307.1	OBP83ef (FBpp78233) 42.6%,1.59e-51	OBP32 (AGAP000638) (19.50%, 0.019)	OBP56e (AAEL002587 (24.06%, 0.015)	No hit.
OBP4	CBA11308.1	OBP56e (FBpp0085634) (34.8%,6.65e-12)	OBP18 (AGAP012319) (32.63%, 2e-08)	OBP56e (AAEL002591) (25.27%, 3e-06)	OBP28(CPIJ012716) (24.17%, 5e-07)
OBP5A	CBA11309.1	OBP19c (FBpp0077018) (32%,1.27e-18)	OBP10 (AGAP001189) (26.03%, 0.54)	OBP5 (AAEL002652) (40%, 0.60)	OBP56e (CPIJ018957) (25.27%, 0.13)
OBP6	CBA11310.1	OBP28a (FBpp0079083) (44.6%, 4.76e-28)	OBP14 (AGAP002905) (30.14%, 8e-08)	OBP56e (AAEL000035) (24.11%, 1e-05)	Novel protein (CPIJ016966) (29.23%, 9e-05)
OBP7	CBA11311.1	OBP83cd (FBpp0078233) (42.6%,5.14e-60)	OBP9 (AGAP000278) (25%, 0.42)	Hypothetical protein OBP99c (AAEL005772) (27.45%, 0.002)	OBP99a (CPIJ017326) (25%, 0.030)
OBP8A	CBA11312.1	OBP83a (FFpp0078305 (49%, 6.1e-37)	OBP17a (AGAP003309) (51.05%, 2e-36)	Novel protein (AAEL009449) (48.23%, 3e-35)	Conserved hypothetical protein (CPIJ007604) (48.59%,7e-35)
OBP9	CBA11313.1	OBP83b (FBpp0078304) (77.9%, 1.42e-60)	OBP17a (AGAP003309) (58.74%, 2e-47)	0BP9 (AAEL013018) (55.94%, 8e-45)	Conserved hypothetical protein(CPIJ007604) (60%, 6e-480)
OBP10	CBA11314.1	OBP83a (FBpp0078304) (72.6%, 3.5e-33)	OBP17a (AGAP003309) (60.23%, 9e-29)	(Novel protein) (AAEL13018) (57.65%, 3e-27)	Conserved hypothetical protein (CPIJ007604) (62.35%, 7e-29)
OBP11	CBA11315.1	OBP83g (FBpp78266) (34.2%, 1.42e-15)	OBP9 (AGAP000278) (33.33%, 2e-13)	OBP99c(AAEL005772) (32.46%, 9e-13)	Obp99a (CPIJ017326) (31.58%, 9e-13)
OBP12	CBA11316.1	OBP83a (FBpp0078305) (69.4%, 4.72e-33)	OBP17a (AGAP003309) (56.82%, 4e-26)	Novel protein (AAEL13018) (56.98%, 7e-26)	Conserved hypothetical protein (CPIJ007604) (58.14%, 9e-27)

OBP13	CBA11317.1	OBP56h (FBpp0292158) (43.5%, 2.38e-21)	OBP26 (AGAP012321) (36.11%, 5e-05)	OBP56e(AAEL002606 (37.84%, 5e-06)	OBP56d (CPIJ012719) (35.62%, 3e-06)
OBP14	CBA11318.1	OBP19a (FBpp0297995) (62.6%, 2.31e-42)	OBP20 (AGAP005208) (48.76%, 4e-30)	OBP56a(AAEL012377) (43.80%, 2e-27)	OBP56a (CPIJ006551) (44.63%, 7e-27)
OBP15	CBA11319.1	OBP56d (FBpp0085673) (33%, 4.08e-10)	OBP18 (AGAP0123121) (30.77%, 1e-05)	OBP56e(AAEL002596) (31.96%, 2e-07)	OBP56e (CPIJ018957) (26.04%, 2e-06)
OBP16	CBA11320.1	OBP57c(FBpp0112012 (35.4%,6.15e-09)	OBP15 (AGAP003307) (28.14%, 0.001)	OBP56a(AAEL000071) (31.25%, 0.003)	OBP13 (CPIJ016479) (24.73%, 0.022)
OBP17	CBA11321.1	OBP56i (FBpp00855639) (27.4%, 0.04)	OBP17b (AGAP005175) (32.39%, 0.065)	OBP56e(AAEL000073) (25%, 0.014)	Conserved hypothetical protein (CPIJ801707) (27%, 0.067)
OBP18	CBA11322.1	OBP69a (FBpp0075687) (44.4%, 6.42e-12)	OBP17 (AGAP003309) (29.11%, 4e-07)	OBP18(AAEL009449) (29.11%, 1e-06)	Conserved hypothetical protein (CPIJ007604) (27.85%, 8e-07)
OBP19	CBA11323.1	OBP69a (FBpp0075687) (44.4%, 6.42e-12)	OBP17a (AGAP003309) 29.11%, 4e-07	Novel protein (AAL009449) (29%, 1e-06)	OBP19 (CPIJ007604) (27.85%, 8e-07)
OBP20	CBA11324.1	OBP99a(FFpp0084816) (44%, 0.1)	OBP7 (AGAP001556) (47.62%, 0.24)	OBP50c(AAEL114830) (27.78%, 0.084)	D7protein (CPIJ018735) (29.41%, 0.11)
OBP21	CBA11325.1	OBP99c(FBpp0084829) (58.4%, 8.56e-44)	OBP9(AGAP000278) (31.93%, 3e-10)	OBP99c(AAELL005772) (30.59%, 3e-11)	OBP99a (CPIJ017326) (32.52%,2e-13)
OBP22	CBA11326.1	OBP8a(FBpp0071242) (35%, 1.86e-13)	OBP9(AGAP000278) (24%, 2e- 04)	OBP99c(AAEL005772) (27.4%, 5e-06)	Novel protein (CPIJ010782) (28.5%, 1e-05)
OBP23	CBA11327.1	OBP48b (FBpp1081115) (53%, 1.48e-32)	HP (AGAP002556) (33.6%, 2e-19)	OBP59 (AAEL015313) (39.6%, 1e-21)	OBP 14 (CPIJ009586) (37.8%, 4e-19)

OBP24	CBA11328.1	OBP19d (FBpp0076996) (45.5%, 6.37e-24)	OBP (AGAP002905) (37%, 2e-15)	OBP57 (AAEL0000035) (28.7%, 1e-11)	OBP28 (CPIJ0016965) (31%, 1e-16)
OBP25	CBA11329.1	OBP56e (FBpp0085634) (27.2%, 0.0002)	OBP26 (AGAP012321) (32%, 3e-08)	OBP13 (AAEL002591) (21.8%, 4e-04)	OBP17 (CPIJ012716) (23.2%, 5e-06)
OBP26	CBA11330.1	OBP Lush (FBpp0290704) (46.1%, 4.76e-24)	OBP (AGAP010489) (50.4%, 1e-36)	OBP1 (AAEL006454) (50.4%, 5e-34)	OBP6 (CPIJ008793) (47.8%, 1e-17)
OBP27	CBA11331.1	OBP59a (FBpp0071781) (38.2%, 2.58e-44)	HP (AGAP012867) (34.1%, 9e-36)	CP (AAEL011416) (29%, 8e-31)	CP (CPIJ010367) (31.3%, 9e-36)
OBP28	CBA11332.1	OBP19a (FBpp0297995) (59.4%, 1.7e-44)	OBP (AGAP05208) (49.6%, 5e-37)	OBP55 (AAEL012377) (43.3%, 3e-37)	OBP11 (CPIJ006551) (44.1%, 5e-38)
OBP29	CBA11333.1	OBP56i (FBpp00855639) (23.5%, 0.57)	OBP (AGAP012325) (27.4%, 0.14)	OBP59 (AAEL015313) (27.5%, 0.26)	OBP14 (CPIJ009586) (25.8%, 0.24)
OBP30	CBA11334.1	OBP73a (FBpp0112090) (55.4%, 2.3e-52)	HP(AGAP006368) (64.7%, 5e-44)	Modifier of mdg4 (AAEL010576) (28.9%, 0.67)	Conserved hypothetical protein (CPIJ017524) (63.8%, 2e-44)
OBP2B	CBA11306.1	OBP99d (FBpp0084821) (31 %, 4.1e-15)	OBP9 (AGAP000278) 44.4%, 1e-30	OBP22 (AAEL005772) (38.16%, 4e-26)	OBP43 (CPIJ017326) (39.47%, 4e-26)
OBP5B	CBA11309.1	OBP19c (FBpp0077018) (32%,1.27e-18)	OBP10 (AGAP001189) (26.03%, 0.54	OBP57 (AAEL002652) (40%, 0.60	OBP14 (CPIJ018957) (25.27%, 0.13)

Table 2: Prediction of Glossina morsitans morsitans OBPs Molecular weight (MW) and Isoelectric point (pI)

***Glossina morsitans morsitans* OBPs**	**Molecular weight (kda)**	**Theoritical Isoelectric point**	**Number of Amino acids**
1	16312.8	7.50	141
2	17900.3	6.15	153
3	27149.1	6.27	227
4	20535.1	9.30	178
5	20819.5	8.75	184
6	16013.5	4.91	145
7	28316.4	5.78	240
8	17474.1	5.42	150
9	17411.0	5.56	150
10	10836.2	5.03	94
11	14346.1	5.47	118
12	10851.2	4.81	95
13	12818.4	4.95	113
14	14588.8	8.80	138
15	12914.9	8.20	111
16	10317.6	4.41	88
17	11820.5	5.89	102
18	12782.5	5.21	109
19	12782.5	5.21	109
20	10376.1	8.50	88
21	15921.9	5.41	137
22	15335.3	5.71	126
23	12245.5	5.45	125
24	15811	4.41	144
25	15197	6.07	134
26	12943.1	7.58	114
27	1243.1	7.58	114
28	16584.4	8.59	149
29	16209.5	5.93	138
30	25901	4.9	157
2B	17900.3	6.15	153
5B	18178.5	4.48	157
8B	30064.5	5.57	261

Table 3: Template used for modeling Glossina morsitans morsitans OBPs

Gmm OBPs	**Template (Organism, Gene, PDB**	**% Similarity**	**E-Value**
1	*Culex quinquefasciatus;* CquiOBP1; 314a	25.46	1.70e-21
2A	*Culex quinquefasciatus;* CquiOBP1; 314a	22.81	3.80e-23
2B	*Culex quinquefasciatus;* CquiOBP1; 314a	22.81	4.1e-23
3	*Anopheles gambie ;* AgamOBP1; 3n7h	18.02	1.80e-5
4	*Anopheles gambie* ; AgamOBP20; 3v2lA	21.19	1.40e-20
5A	*Culex quinquefasciatus;* CquiOBP1; 314a	17.12	9.20e-7
5B	*Anopheles gambie* ; AgamOBP20; 3v2lA	19.01	3e-20
6	*Bombyx mori* ; GOBP2 ; 2wcj	24.22	1.90e-20
7	*Drosophila melanogaster;* OBPLush ; *1ooh*	18.87	5.40e-7
8A	*Anopheles gambie ;* AgamOBP1; 3n7h	55.20	2.90e-32
8B	*Anopheles gambie ;* AgamOBP1; 3n7h	57.5	2.1e-28
9	*Anopheles gambie* ; AgamOBP1 ; 2erbB	67.21	1.47e-45
10	*Culex quinquefasciatus;* CquiOBP ; 30gn	62.35	1.29e-26
11	*Bomb mori* ; GOBP2 ; 2wcj	20.91	1.80e-21
12	*Anopheles gambie ;* AgamOBP1; 3n7h	55.56	8.00e-26
13	*Anopheles gambie* ; AgamOBP20; 3v2lA	25	2.00e-25
14	*Anopheles gambie* ; AgamOBP20; 3v2lA	49.17	8.80e-34
15	*Anopheles gambie* ; AgamOBP7 ; 3r1vA	21.82	4.60e-34
16	*Anopheles gambie* ; AgamOBP20; 3v2lA	20.88	2.80e-14
17	*Anopheles gambie* ; AgamOBP7 ; 3r1vA	18.42	8.90e-5
18	*Anopheles gambie* ; AgamOBP7 ; 3r1vA	25.64	1.6e-28
19	*Anopheles gambie ;* AgamOBP1; 3n7h	18.75	7.1e-5
20	Crystal structure of AED7-Norepineprhine complex; 3dye	16.91	1.2e-7
21	*Anopheles gambie ;* AgamOBP1; 3n7h	19.17	9.6-26
22	*Culex quinquefasciatus;* CquiOBP1; 314a	19.09	2.2e-23
23	*Drosophila melanogaster;* OBPLush ; 1ooh	18.87	4e-18
24	*Bomb mori* ; Pherone Binding Protein ; 2fjy	22.13	1.1e-19
25	*Anopheles gambie ;* AgamOBP1;	18.92	1.1e-21

	3n7h		
26	*Anopheles gambie* ; AgamOBP 3q8i	45.61	1.7e-31
27	*Drosophila melanogaster;* OBPLush ; 1ooh	17.28	3.10e-6

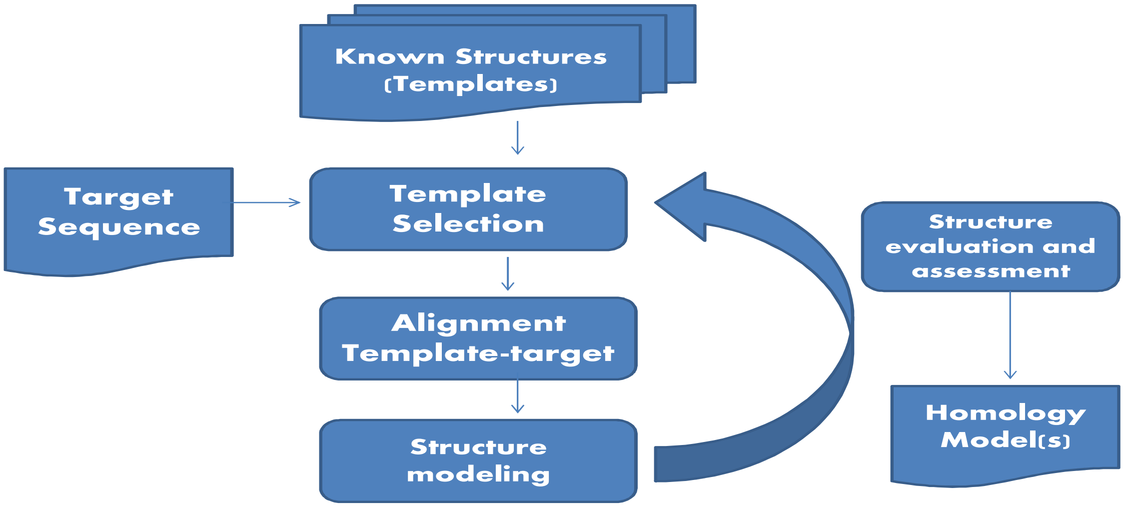


Figure 1: Flow chart of Swiss-Model (Lorensa et al., 1997)

```
DmelOBP56i      -----MHFFTCCA----LLLVVVTLPTCFVQAGPIKDQCMAAA-----------------
DmelOBP56d      --------------------MVVCVQRTQVQAGPIKDQCMAAA-----------------
DmelOBP56e      KVFFVFAALAALS----LASAVGLTDSQKAEAKQRAKACVKQE-----------------
GmmOBP15        ---------------------------TRETLQNYVKTCVIEE-----------------
GmmOBP4         RKVMFRVTLILLA----IVTSALFSENRYMEFLADFKHCKRER-----------------
AngamOBP18      NFKLYSSLFVFPS----PLQGARLEAEHVRRIHQNARECVKET-----------------
GmmOBP9         IVLLSAWTRAQQP----RRDDEYPPPAILKLAKPFHDICVEQT-----------------
DmelOBP83a      IALSLLSGALILPPAAAQRDENYPPPGILKMAKPFHDACVEKT-----------------
DmelOBPb83b     ILLLIGCAAAQEP----RRDGEWPPPAILKLGKHFHDICAPKT-----------------
AngamOBP17      LLCCSMTLGDTTP----RRDAEYPPPELLEALKPLHDICLGKT-----------------
CqiOBP18        LGIAVVVLADVTP----RRDAEYPPPELLEALKPLHDICAKKT-----------------
AeayON10        SCLVAVSIADVTP----RRDAEYPPPELLQALKPLRDICQKKT-----------------
AeayOBP14a      SALVSLSVGDVTP----RRDAEYPPPEFLEAMKPLREICIKKT-----------------
GmmOBP12        ------------------------------------------------------------
GmmOBP10        ------------------------------------------------------------
GmmOBP8         TLLMSFGLNNAQKP---RRDENYPPPDFLKSFKIIHDVCVEKT-----------------
AngamOBP15      SICLMATASANAP--------KSLSPELLQQMGQFRSECLRET-----------------
AngamOBP7       KMSNLVVVLVLLTMYIVLSAPFEIPDRYKKPAKMLHEICIAES-----------------
AngamOBP26      TFVAIAVVALIAG-------TFALTIDQKKKAEGYAAECVKTT-----------------
AeayPN13        FFVAIAVVALAAG-------AWALTIDQQKKAEAYAAECVKST-----------------
CqiOBP56d       TFVAIAVIALAAG-------AWSLTIEQQKKAEAYGAECVKST-----------------
AeayPN4         TFAAIVSFALIAG-------CMAVTEDQKEAARQLAGKCMQQT-----------------
CqiOBP28        SFVAIISLALVAS-------SMAVTEQEKEAARQLAGKCMQQT-----------------
Angamnovel      SIVAQTAEIPTLPWSGSELKAQYSGKKIKISSELFARGCVTTSDQGLVAAGKIGFPVMIK
AeayHNP5        SALAVWRTNPNPPSYRRGKRNPTVNPQQNNQSDQNSGTSKEQSPEQIGKFRHHG------
GmmOBP7         VLHQCLAPFGGYTLENDQRLQRFKQWSDTYEEFPCFTNCYLNNMFNIYN-----------
DmelOBP83cd     TINRCIQNYGGLTAENAERLERFKEWSDSYEEIPCFTRCYLSEMFDFYN-----------
GmmOBP3         WNKQRLVDELGANMYNYCRFELNRAFKNVCSFAFKGLKCLKQAEMN--------------
DmelOBP83ef     WRLKQLTEDLGADVYNYCRFELRRMGSDGCSFAYRGLRCLKQAEMH--------------
GmmOBP1         TAVILLALFALVS--------ADYKLRNQEDLNKARKECMEAK-----------------
DmelOBP44a      VAILLCALLGLAS-------ASDYKLRTAEDLQSARKECAASS-----------------
GmmOBP2         IVFLVTLATVWGHHHHEHHDDDDYVVKTREDLFKYRDECSNKLN----------------
DmelOBP99b      KVLIVLLLGLAFVLADHHHHHHDYVVKTHEDLTNYRTQCVEKVH----------------
AngamOBP9       FVVALLAFTAVVS--------AEFVVQTREDLLAYRAECVKSLG----------------
AeayOBP99c      VFIAVFALIAVAA--------AEFTVSTTEDLQRYRTECVSSLN----------------
CqiOBP99a       LFIAIFALIAVAT--------ADFTVKTTDDLQTYRSECVSSLS----------------
DmelOBP99a      VFVAICVLIGLAS--------ADYVVKNRHDMLAYRDECVKELA----------------
GmmOBP21        ------------------AEDEDWQPKTVADIKSIRNECLKEHP----------------
DmelOBP99c      YLIVALALCAVAH-------ADDWTPKTGEEIRKIRVDCLKENP----------------
GmmOBP11        ---------------------------TKDDALKAHEECHEEFQ----------------
DmelOBP83g      QSQSLLLIVAAVATFLVAQTTAKFLLKDHADAEKAFEECREDYY----------------
AngamOBP32      PAMRSFFHPDPDDCDYERRTYHCLNSQRLNHPSPHVDVCERAYESFRCYYEQYG------
AgamOBP10       VRVLIVFVALLTFAGQPFAVRGQQELSDLPEVKGYKLHCIESS-----------------
AeayPN3         LLISIVSFALVGAALSVPQQANLEDIGKIRNGETYALECLLAS-----------------
CqiOBP56e       AVLLVVAVACGVSAVPQQAPPNLEDISRIPNGEMYALECLLTS-----------------
AeayOBP50c      LTVIGLFAMACSQQPISQECFTRPNEGNPKDCCKAPNVIPPKDQ----------------
AeayPN15        LVVALLSVTIALN------QIKAFTLQQRQQGDIYAIECIAETG----------------
CqiOBP22        AVIVLALQAQLLP-------SSAWTPHSPEQFRRFEELCMDLA-----------------
CqiOBPD7        CSTAINVWTDSCKNCLHVIGSAVRRTAGHYSYQSRITACYSRCRGS--------------
DmelOBP69a      FFLALLILYDLIP----SNQGVEINPTIIKQVRKLRMRCLNQTG----------------
GmmOBP19        ------------------------------------------------------------
```

```
GmmOBP18       ------------------------------------------------------------
DmelOBP8a      LLSRLLLLLLVVELTPPAIPVPMRSSPQSLALLRARDQCGRELT----------------
GmmOBP22       -----------------------------DDFFQMSERCMRLEK----------------
GmmOBP20       LFTLFLIVFIFSSREASALNETSRFVLKEPNVRFAQMRCAEKYP----------------
GmmOBP17       ------------------------NIPGRFNLPNYS------------------------
GmmOBP16       ------------------------------------------------------------
GmmOBP14       ----------------------ATEEQMRSAANLMRDVCLPKFP----------------
DmelOBP19a     KFHLLLVCVAISLGPIPQSEAGVTEEQMWSAGKLMRDVCLPKYP----------------
AngamOBP20     MLFVFFTLLSCTKKKKIFPLRKSTVEQMMKSGEMIRSVCLGKTK----------------
AeayOBP14b     FGLFAVVTLFQTGLGGVGVEGKATVEQMTKTGEMIRNVCIGKLK----------------
CqiOBP56a      TRVELALLVWIAVWSTGKVEGKATVEQMMKTGEMIRSVCIGKAK----------------
AeayPN17       KVKLLFHVLLAVMLSLHTSESKSTMEQLAKASEMMRGVCVGKTK----------------
AeayOBP56a     MKTLSVIILGAWLVHLGGVMSSMTFEDMQETAKMMRGICQPKYG----------------
GmmOBP13       ---------------------------TKDDFEKILQSCREDMQ----------------
DmelOBP56h     FTLFCIALAAFLS-----------MGQCNPDFRQIMQQCMETN-----------------
GmmOBP6        LLLVTVLMLGILS-----VEAEIDVQEEIAKFILLANECREEVG----------------
DmelOBP28a     STPIILVAIVLLG---AALVRAFDEKEALAKLMESAESCMPEVG----------------
GmmOBP5        YAPAQYQLKPADNFASSPVNKRQMPTSDIPKNMQQFQDTLNEAKFK--------------
AngamOBP14     SAVLYFALLATAMVCRVQAGSAEELEQAKEMLRGLAAECKTKEG----------------
AeayPN6        SFCSLQLAWRFVTELQCANSDEEKKAQAKEMMRGMAEECKKKEG----------------
CqiOBP6        MVEIIINTG-------------PFKEILLAVRRGMLQDCKESLG----------------
CqiOBP13       -----MNLISAFG-------VFLAAAMVSADLSIQEEKCMKEEG----------------
DmelOBP19c     VVAVLLQTHCVRGQTQAFDLAKLLPKTGTEPIWAVIDRNLPQVQ----------------
DmelOBP57c     LWLICILTVSVVS----------IQSLSLLEETNYVSDCLASNN----------------


DmelOBP56i     SVKCFFR-------CFLENIGIIAD-NQIIPGAFDRVLGHIVTAEAVERME---------
DmelOBP56d     SVKCFFR-------CFLENIGIIAD-NQIIPGAFDRVLGHIVTAEAVERME---------
DmelOBP56e     KVKCFAN-------CFLEQTGLVAN-GQIKPDVVLAKLGPIAGEANVKEVQ---------
GmmOBP15       EGKCFFS-------CFHEKIGLTIN-GVLQKKIAFGHLKRIFDRETAEFVL---------
GmmOBP4        EAKCFLG-------CLYERTGILKN-GVLQNDVLKKNVGYIANRVLLDEVL---------
AngamOBP18     KAKCFVK-------CFLDKAGFIDDDGVIQQDVIREKLTVGIEAGKVNELI---------
GmmOBP9        ALKCYMN-------CLFHEFDVVDDNGDVHLEKLFSKIP-AALRDLLMEAS---------
DmelOBP83a     KLKCYMN-------CFFHEIEVVDDNGDVHLEKLFATVP-LSMRDKLMEMS---------
DmelOBPb83b    ALKCYMN-------CLFHEFEVVDDNGDVHMEKVLNAIPGEKLRNIMMEAS---------
AngamOBP17     KLKCYMN-------CLFHEAKVVDDNGDVHLEKLHDSLP-SSMHDIAMHMG---------
CqiOBP18       KLKCYMN-------CLFHEAKVVDDNGDVHLEKLHDSLP-NSMHDIAMHMG---------
AeayON10       KLKCYMN-------CLFHEAKVVDDTGHVHLEKLHDALP-DSMRDIAMHMG---------
AeayOBP14a     NLKCYMN-------CLFHEAKVVDDTGHVHLEKLHDALP-DSMHDIALHMG---------
GmmOBP12       ALKCYMN-------CFFHELGAVDDKGDVHLETLNLIMP-GSFVEAILKPA---------
GmmOBP10       ALKCYMN-------CFFHELGLVDDKGDVHLETLHQSMP-GSFVDLILKPA---------
GmmOBP8        ALKCYMN-------CLFHEVNVVDDAGELHFEKLVRMIP-EPFLEMVKHII---------
AngamOBP15     ELQCYMY-------CMFRLHNVTRPNGELDLIDVYHAIP-KQFNSIALKVL---------
AngamOBP7      AAKCYIH-------CLFDKIDVVDEATGRILLDRLLYIIPDDVKAAVDHLT---------
AngamOBP26     KTKCFAK-------CFLEKAGFMTDKGEIDEKTVIEKLSVDHDRAKVEGLV---------
AeayPN13       KTKCFSK-------CVLEKAGFMNEKGEIQEKTVIDKLSVDHDKAKVEATL---------
CqiOBP56d      KTKCFAK-------CVLEKAGFMNAAGDVQEKTVVEKLSIDHDKSKVEATL---------
AeayPN4        NTKCFVQ-------CFFQGAGVVDGEGNMQEAFVTEKLASEYGQAKAEEVV---------
CqiOBP28       NTRCFVQ-------CFFQGAGFVDADGNVQEEHVIEKMSAEFDRAKAEEVV---------
Angamnovel     LLADQYG-------NAISLFGRDCSIQRRHQKIIEEAPAVIADPAVFEEMERAAVRLAKM
AeayHNP5       SYRSDSD-------QHHLQNTVIEEEPEWVSAGPTSRLDTIELRGFDDDLSVS-------
```

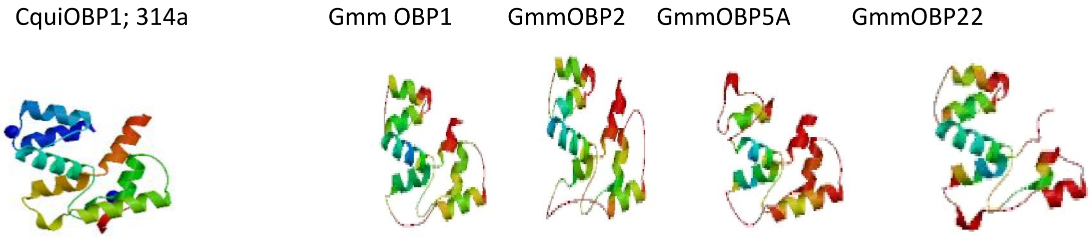
GmmOBP7	EKLIQGNNS-----CEIAYNGFHCLINREDDPFILIDNIEDISMEAKRAMK---------
DmelOBP83cd	KKLELPFESGESS-CKHAYEGFHCITNMESHPFTVIDNMPNISPSAKDAMK---------
GmmOBP3	LLEYYHFPQLEHIPCLFKCFADKSHLYTVNYEWNVLNWLKAFGPIRNENAD---------
DmelOBP83ef	LLQYSKLKSKEPIPCLFQCFADAMGFYDPDGNWRLENWKQAFGPSGNEDQSSG-------
GmmOBP1	ITRCYIE-------CIFDKFQLFDSQTGFKNDNLIAQLGQSKDNKDEVKAD---------
DmelOBP44a	ITRNYIQ-------CIFVKFDLFDEAKGFKVENLVAQLGQGKEDKAALKAD---------
GmmOBP2	VTKCYMK-------CMFEHFGFFNEKQGFDVHKIHKQLMGAHGTVDHSDET---------
DmelOBP99b	VTHCYLE-------CIFQKFGFYDTEHGFDVHKIHIQLAGPGVEVHESDEVHQK------
AngamOBP9	TTQCYIK-------CIFNKMQLFDDTNGPIVDNLVVQLAHGRDANEVREEI---------
AeayOBP99c	TTMCYIK-------CVFNKMQLFDDTEGPLVDNLVHQLAHGRDAEEVRTEV---------
CqiOBP99a	TTQCYIK-------CIFNKMELFDDNNGPIVDNLVLQLAHGRDADEVRAEI---------
DmelOBP99a	KTQCYIK-------CVFTKWGLFDVQSGFNVENIHQQLVGNHADHNEAFHAS--------
GmmOBP21	EVRQYLL-------CTALKMEVFCAHQGYHPNRIAKQFKMDMNEEEVLEIA---------
DmelOBP99c	DVRQYLT-------CSAIKLGIFCDQQGYHADRLAKQFKMDLSEEEALQIA---------
GmmOBP11	LTNCYVK-------CWVEKMGIFTENRGFNEKNIVAQYTYENFKNLESVRHG--------
DmelOBP83g	RTSCFVK-------CFLEKLELFSEKKGFDERAMIAQFTSKSSKDLSTVQHG--------
AngamOBP32	DVDCLAR-------CFLLRSGLYSEQHGPHLDRLYVQCNNYANETRFRETTG--------
AgamOBP10	PTKCFVQ-------CFFQKLRLMDEKGVVLKDKLEVFLTKLMDADKAKDYV---------
AeayPN3	RVKCLVK-------CFFEKTGFMDAEGNLNEEAIVTQLSQFMPKDQVETLV---------
CqiOBP56e	QVKCLVK-------CFFEKAGFMDSEGKLQQEVIVRQLGQMMGQDQVQKLV---------
AeayOBP50c	NHNCLAQ-------CMFEQQGIMADGAVSKDAAISKTVAVMGGSSEWEATTKN-------
AeayPN15	RSKCFIR-------CFFEKEGFMDSKGNLHTEKIADALAGDFNREKVETVL---------
CqiOBP22	RTHCFHR-------CLGIVSGLYSDREGADLGRVYAQFGGGRNETRFRDGA---------
CqiOBPD7	VLQVKYLKYFSLSTCLAAWAPLNPEETLYVYTSCFDEWAPKDGTQRKAVAETWFKSWDLK
DmelOBP69a	EIKCFLY-------CMFDMFGLIDSQNIMHLEALLEVLPEEIHKTINGLVS---------
GmmOBP19	SFKCYLQ-------CIFDSLGLVDSNNQVNLEKLINFAPTEIHEHILELHR---------
GmmOBP18	SFKCYLQ-------CIFDSLGLVDSNNQVNLEKLINFAPTEIHEHILELHR---------
DmelOBP8a	HVRHYLH-------CFWSRLQLWLDETGFQAQRIVQSFGGERRLNVEQALP---------
GmmOBP22	IVHKYIL-------CVNRELQIWDNNQGFDIEKIYQQYKGRANEEVVLPIIS--------
GmmOBP20	ANHCYVY-------CLFYKLGLIDLRSRDLDVVYLIKI---------------------
GmmOBP17	NAKCFLR-------CWYKKMGILKENLVTSAGPIPELRQHMRECNEVATEWAQ-------
GmmOBP16	KFKCYAH-------CLLSNLKYLNTFSGKFDIEDFKQQDGIEDEDVAVIAK---------
GmmOBP14	DAKCYIN-------CVMEMMRTMKKGKFLYEGALKQVDLLMPDSYKEEYRPG--------
DmelOBP19a	DTNCYIN-------CILEMMQAIKKGKFQLESTLKQMDIMLPDSYKDEYRKG--------
AngamOBP20	ELKCYVN-------CVMEMMQTMKKGKLNYDASVKQIDTIMPDELAGPMRAA--------
AeayOBP14b	ELKCYVN-------CIFEMMQVVKKGKLNYDAAMKQIDTIMPDELAEPMRIALN------
CqiOBP56a	EVKCYVN-------CALEMMQAMKKGKLNYDAMLKQIDTIMPDELAEPMRN---------
AeayPN17	DLKCYAN-------CVLEMMQAMRKGKVNADSAIKQVDLLIPPEIGEPTKK---------
AeayOBP56a	EFKCYAS-------CLMDLTHTAKRGKLNYEAAVKQITMLPDDFREPFRVG---------
GmmOBP13	GVKCYMK-------CVMEKQGHFKNGALLEEAVIKSLESSPADHNDQNQMS---------
DmelOBP56h	NLKCYTK-------CLMEKQGHLTNGQFNAQAMLDTLKNVPQIKDKMDEISSG-------
GmmOBP6	EGKCLRA-------CLMKKYEVLDANGKLVKSVALEHAKKFTNSDENKLKIAG-------
DmelOBP28a	AGKCLRA-------CVMKNIGILDANGKLDTEAGHEKAKQYTGNDPAKLKIALEIG----
GmmOBP5	REKCLMA-------CILKRMKLMDSDYKLSVPTISHIAGMISDENPLLISVAAAT-----
AngamOBP14	TQKCLAG-------CMQEQFGVSNGKAFQEDGFIEIAKMLMKGDETKIELAK--------
AeayPN6	VQKCFLS-------CFQHQFQISDGKRFNKDGFMQLSAMMFGEDQEKMATAEEIA-----
CqiOBP6	VQKCLAH-------CAMKQFGVLHGRKFNKQGFASVAKLVIFLDKRKSRYVD--------
CqiOBP13	VQKCYYG-------CLFQALGYLDAEGKRFNSEGFLKTTLPMAANNEKHTQGVHR-----
DmelOBP19c	KEKCLVE-------CVLKKIKLMDADNKLNVGQVEKLTSLVTQDNKMAIAVSSS------
DmelOBP57c	RYKCFIH-------CLAEKGNLLDTNGYLDVDKIDQIEPVSDELREILYDCKK-------

```
DmelOBP56i      --------------------------------------------ATCNMIK----------
DmelOBP56d      --------------------------------------------ATCNMIK----------
DmelOBP56e      --------------------------------------------AKC--------------
GmmOBP15        --------------------------------------------GEC--------------
GmmOBP4         --------------------------------------------PPC--------------
AngamOBP18      --------------------------------------------KKC--------------
GmmOBP9         --------------------------------------------KGC--------------
DmelOBP83a      --------------------------------------------KGC--------------
DmelOBPb83b     --------------------------------------------KGC--------------
AngamOBP17      --------------------------------------------KRC--------------
CqiOBP18        --------------------------------------------KRC--------------
AeayON10        --------------------------------------------KRC--------------
AeayOBP14a      --------------------------------------------KRC--------------
GmmOBP12        --------------------------------------------QHC--------------
GmmOBP10        --------------------------------------------QHC--------------
GmmOBP8         --------------------------------------------DACES------------
AngamOBP15      --------------------------------------------AKCN--K----------
AngamOBP7       --------------------------------------------RECS-------------
AngamOBP26      --------------------------------------------KKCN-------------
AeayPN13        --------------------------------------------KKCN-------------
CqiOBP56d       --------------------------------------------KKCN-------------
AeayPN4         --------------------------------------------QRCR-------------
CqiOBP28        --------------------------------------------SRCR-------------
Angamnovel      VGYVSAGTVEYLYDSEGKYFFLELNPRLQVEHPCTEMVADVNLPACQLQIGMGVPLYRIK
AeayHNP5        --------------------------------------------QASSDK-----------
GmmOBP7         ---------------------------------------------ECLHKFN---------
DmelOBP83cd     ---------------------------------------------DCLQDVHQD-------
GmmOBP3         -------------------------------------------ISICRVN-----------
DmelOBP83ef     -----------------------------------------ADYSGCRLSG----------
GmmOBP1         -------------------------------------------IEKCADKN----------
DmelOBP44a      -------------------------------------------IEKCAD------------
GmmOBP2         ----------------------------------------HEKIAKCAD------------
DmelOBP99b      -------------------------------------------IAHCAET-----------
AngamOBP9       --------------------------------------------VKCAG------------
AeayOBP99c      --------------------------------------------LKCVD------------
CqiOBP99a       --------------------------------------------LKCVD------------
DmelOBP99a      -------------------------------------------LAACVD------------
GmmOBP21        --------------------------------------------EKCHD------------
DmelOBP99c      --------------------------------------------QSCVD------------
GmmOBP11        -------------------------------------------LEKCID------------
DmelOBP83g      -------------------------------------------LEKCID------------
AngamOBP32      ---------------------------------------------TCYRR-----------
AgamOBP10       --------------------------------------------QQCD-------------
AeayPN3         --------------------------------------------KNCK-------------
CqiOBP56e       --------------------------------------------ENCN-------------
AeayOBP50c      ------------------------------------------VVEACFQKVSALG------
AeayPN15        --------------------------------------------ANCLT------------
CqiOBP22        --------------------------------------------ERCFR------------
CqiOBPD7        PDNPG-----------------------------------THCFAKCVLEG----------
DmelOBP69a      ---------------------------------------------SCG-------------
```

```
GmmOBP19      --------------------------------------------ACDTQR----------
GmmOBP18      --------------------------------------------ACDTQR----------
DmelOBP8a     -----------------------------------------AINGCNAKTS---------
GmmOBP22      --------------------------------------------QCN-------------
GmmOBP20      ------------------------------------------------------------
GmmOBP17      ------------------------------------------------------------
GmmOBP16      ---------------------------------------------CKK------------
GmmOBP14      ------------------------------------------LAKCKDS-----------
DmelOBP19a    ------------------------------------------INLCKDST----------
AngamOBP20    ------------------------------------------LDICRT------------
AeayOBP14b    --------------------------------------------ACRT------------
CqiOBP56a     -----------------------------------------AVNVCRN------------
AeayPN17      -----------------------------------------AFDMCRN------------
AeayOBP56a    ------------------------------------------LDSCRN------------
GmmOBP13      ----------------------------------------AIVKECKK------------
DmelOBP56h    ------------------------------------------VNACKD------------
GmmOBP6       ----------------------------------------TIIDMCS-------------
DmelOBP28a    -------------------------------------------DTCAA------------
GmmOBP5       ------------------------------------------ASNCNN------------
AngamOBP14    ----------------------------------------EIAADCK-------------
AeayPN6       -------------------------------------------EECSS------------
CqiOBP6       ----------------------------------------QVADECEK------------
CqiOBP13      -----------------------------------------LAKQCEG------------
DmelOBP19c    -----------------------------------------MAQACSR------------
DmelOBP57c    ------------------------------------------------------------


DmelOBP56i    --------SETSHDESCEFA-WQISECYEGVRLSDVKKG---------------------
DmelOBP56d    --------SETSHDESCEFA-WQISECYEGVRLSDVKKG---------------------
DmelOBP56e    --------DSTKGADKCDTS-YLLYKCYY-------------------------------
GmmOBP15      --------VNLVGKDKCETA-YQFEKCLFN------------------------------
GmmOBP4       --------YAVSGTNKCDIA-FELKKCFKNVGFDKVWIT---------------------
AngamOBP18    ---------SVEGTDACDTA-YQMYKCFFSNHKVPKELF---------------------
GmmOBP9       --------VHPEGDTLCHKA-WWFHQCWKK------------------------------
DmelOBP83a    --------VHPEGDTLCHKA-WWFHQCWKK------------------------------
DmelOBPb83b   --------IHPEGDTLCHKA-WWFHQCWKK------------------------------
AngamOBP17    --------LYPEGETLCDKA-FWLHKCWKQ------------------------------
CqiOBP18      --------LYPEGENLCEKA-FWLHKCWKQ------------------------------
AeayON10      --------LYPEGENLCEKA-FWLHKCWKE------------------------------
AeayOBP14a    --------LYPEGENLCEKA-FWLHKCWKE------------------------------
GmmOBP12      --------IHPEGDTLCHKA-WWFHQCWKK------------------------------
GmmOBP10      --------VHPEGDTLCHKA-WWFHQCWKK------------------------------
GmmOBP8       --------HIPKGETQCDRA-WSWHVCFKQ------------------------------
AngamOBP15    --------STGPIADACERA-YSHHRCWKETEP-----E---------------------
AngamOBP7     ---------HIVTPDKCETA-YETVKCYFN------------------------------
AngamOBP26    ----------HKEANPCETA-FKAYQCIYA------------------------------
AeayPN13      ----------QKGANACDTA-FKMTECFYN------------------------------
CqiOBP56d     ----------QKGANPCDTA-FKVYECFYN------------------------------
AeayPN4       ---------NNSGANACERS-FSLLQCYIAN-----------------------------
CqiOBP28      ---------NNAGPNACERS-FALLQCYIAN-----------------------------
Angamnovel    KNVWGYFSVAASGGLHEFAD-SQFGHCFSWGENRQQARENLVIALKELSIRGDFRTTVEY
```

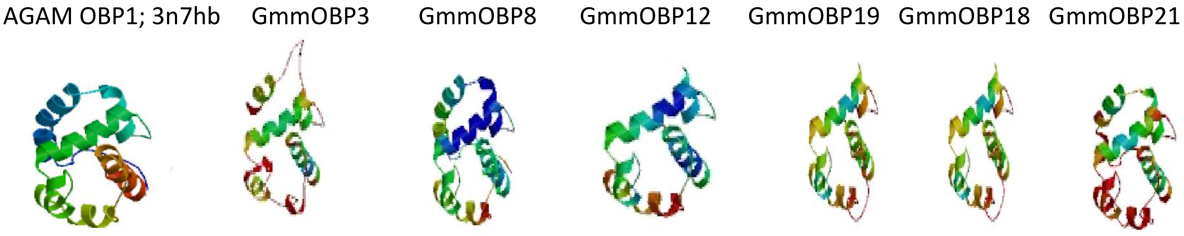
```
AeayHNP5      -------TSNGDGKNSSKNS-GKHISFYDELHHYEHVHAK---------------------
GmmOBP7       -----TDEWQYLSDYVRFPV-QEPIPCYTR-------------------------------
DmelOBP83cd   -------EWKSFDAFAYYPV-NEPIPCFTR-------------------------------
GmmOBP3       -------ANEREKMDICAIM-YEEYNCWERLNYN---------------------------
DmelOBP83ef   -------TQREVALSKCSWM-YHEYKCWER-------------------------------
GmmOBP1       ----------TEKSDSCTWA-FRGFKCFIS-------------------------------
DmelOBP44a    --------KNEQKSPANEWA-FRGFKCFLG-------------------------------
GmmOBP2       --------KKPEDTDPCAWA-YRGGVCFIN-------------------------------
DmelOBP99b    ---------HSKEGDSCSKA-YHAGMCFMN-------------------------------
AngamOBP9     ---------SNTDGNVCHWA-FRGFQCFQKN------------------------------
AeayOBP99c    ---------KNTDNNACHWA-FRGFKCFQKN------------------------------
CqiOBP99a     ---------KNTDDNSCHWA-FRGFKCFQTN------------------------------
DmelOBP99a    --------KNEQGSNACEWA-YRGATCLLKEN-----------------------------
GmmOBP21      --------SNPDNSSVDVWA-FRGHKCMMSS------------------------------
DmelOBP99c    --------DNAQKNPTDVWA-FRGHQCMMAS------------------------------
GmmOBP11      --------HNEWETDVCTWA-NRVFSCWLKVN-----------------------------
DmelOBP83g    --------HNEAESDVCTWA-NRVFSCWLPIN-----------------------------
AngamOBP32    --------LKSECQDECVLAGRFLRECFYEGGISIVN------------------------
AgamOBP10     ----------LRRTNPCDTA-YAVYDCYLG-------------------------------
AeayPN3       ----------IEGTDACDTA-YQATECYFKN------------------------------
CqiOBP56e     ----------VQGVDACDTA-YKATECYFKN------------------------------
AeayOBP50c    ---------AQKDSQGCSVMAGSFMDCMPSMMFT---------------------------
AeayPN15      -----------KEKTACETA-FRMYECFYN-------------------------------
CqiOBP22      --------WMLATEMGEGGTGIRLGKCERPYR-----------------------------
CqiOBPD7      --------VGLYDEKTNTMAVSRVVEQQKAFEKFN--------------------------
DmelOBP69a    ---------TQKGKDGCDTA-YETVKCYIAVN-----------------------------
GmmOBP19      ----KLVDVIPAGKDSCDIV-YTTSQCYYELKP----------------------------
GmmOBP18      ----KLVDVIPAGKDSCDIV-YTTSQCYYELKP----------------------------
DmelOBP8a     --------SRGSGAQTVVDWCFRAFVCVLATPVG---------------------------
GmmOBP22      ---------QDAKQRNYELWCYKAFLCILDTQVG---------------------------
GmmOBP20      --------------------------------------------------------------
GmmOBP17      ---------NQSNGDECEFA-WSFYTCMHES------------------------------
GmmOBP16      --------LYDNINDPCEYG-FNILQCILMF------------------------------
GmmOBP14      ---------ANGIKNNCDAA-YAVLSCLRG-------------------------------
DmelOBP19a    --------VGLKNAPNCDPA-HALLSCLKN-------------------------------
AngamOBP20    --------VADGIKNNCDAA-YVLLQCLSKN------------------------------
AeayOBP14b    --------ASDGIKNNCDAS-YAIAQCVAKN------------------------------
CqiOBP56a     --------SADGIKNNCEAS-YAVAKCISKN------------------------------
AeayPN17      --------SADGIKNNCEAA-WALVKCLHQK------------------------------
AeayOBP56a    --------AADGIDDYCEVA-YTLLKCFFKAS-----------------------------
GmmOBP13      ----------EIGSNECETA-FKVSMCLREH------------------------------
DmelOBP56h    ----------IKGTNDCDTA-FKVTMCLKEH------------------------------
GmmOBP6       --------AMDTVGDTCEAA-EQYSECFKKQ------------------------------
DmelOBP28a    ---------ITVPDDHCEAA-EAYGTCFRG-------------------------------
GmmOBP5       ---------AINAREPCEAA-NQINKCIAN-------------------------------
AngamOBP14    ---------AVANDDRCELA-VDIMNCLKES------------------------------
AeayPN6       ----------VENADRCQLS-VDIKECVEKAMD----------------------------
CqiOBP6       ----------IDNEDLCELG-AELYMCAVTG------------------------------
CqiOBP13      ----------VANEDRCELG-EDLMACLKDRG-----------------------------
DmelOBP19c    ---------GISSKNPCEVA-HLFNQCISRQLE----------------------------
DmelOBP57c    --------IYDEEEDHCEYA-FKMVTCLTESFEQ---------------------------
```

Figure 2: Multiple sequence alignment of GmmOBPs with the Drosophila and mosquito ortholgs



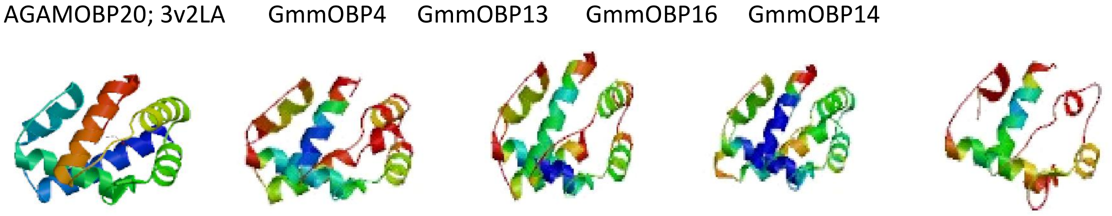
```
GmmOBP2       MKTIIVIVFLVTLATVWGHHHHEHHDDDDYVVKTREDLFKYRDECSNKLNVPADLLEKYK 60
GmmOBP1       ------------------------------------------------------------
314A       ------------------------------------------------------------
GmmOBP22      ------------------------------------------------------------
GmmOBP5A      ------------------------------------------------------------


GmmOBP2       KWQYPDDEVTKCYMKCMFEHFGFFNEKQGFDVHKIHKQLMGAHGTVDHSDETHEKIAKCA 120
GmmOBP1       ------------------------------------------------------------
314A       ------------------------------------------------------------
GmmOBP22      ------------------------------------------------------------
GmmOBP5A      --------------------------------------------------------MRFH 4


GmmOBP2       DKKPEDTDPCAWAYRGGVCFINSNLQLVKSSVNMKTTAVILLALFALVSADYKLRNQEDL 180
GmmOBP1       ---------------------------------MKTTAVILLALFALVSADYKLRNQEDL 27
314A       -----------------------------------MNSLLLIGGVLVVLNVQFVTAADNN 25
GmmOBP22      ---------------------------------------------------------DDF 3
GmmOBP5A      IILKLMSWMCLMCAIESKNVIDLLEGKMYAPAQYQLKPADNFASSPVNKRQMPTSDIPKN 64
                                                 .

GmmOBP2       NKARKECMEAKK--VTPELVEKYKKFDFPDDE---ITRCYIECIFDKFQLFDSQTGFKND 235
GmmOBP1       NKARKECMEAKK--VTPELVEKYKKFDFPDDE---ITRCYIECIFDKFQLFDSQTGFKND 82
314A       ESVIESCSNAVQGAANDELKVHYRANEFPDDP---VTHCFVRCIGLELNLYDDKYGVDLQ 82
GmmOBP22      FQMSERCMRLEK--VPDRYKAQFTEFQFPNDP---IVHKYILCVNRELQIWDNNQGFDIE 58
GmmOBP5A      MQQFQDTLNEAKFKCARAMRLDSNKLLMYEDQPSLREKCLMACILKRMKLMDSDYKLSVP 124
          . :   . :        .    : :*     :  : *:  .::: *..  ..

GmmOBP2       NLIAQLGQSKDNKDEVKADIEKCADKN---TEKSDSCTWAFRGFKCFISKNLPLVMESLK 292
GmmOBP1       NLIAQLGQSKDNKDEVKADIEKCADKN---TEKSDSCTWAFRGFKCFISKNLPLVMESLK 139
314A       ANWENLGNSDDADEEFVAKHRACLEAKN-LETIEDLCERAYSAFQCLREDYEMYQNNNNA 141
GmmOBP22      KIYQQYKG-RANEEVVLPIISQCNQDA----KQRNYELWCYKAFLCILDTQVGEWFKEDV 113
GmmOBP5A      TISHIAGMISDENPLLISVAAATASNCNNAINAREPCEAANQINKCIANELKAHKLNLIY 184
                 . . .     .       :    .     *: .      :

GmmOBP2       KN----------- 294
GmmOBP1       KN----------- 141
314A       TSE---------- 144
GmmOBP22      RRQQTRTLTNGHQ 126
GmmOBP5A      -------------
```

Figure 3: Amino acid sequence of Glossina morsitans morsitans OBPs (OBP22, OBP2, OBP1 and OBP5) aligned with that of Culex quinquefasciatus CquiOBP1 (PDB: 314 chain A), used as template for the predicted three-dimensional structure



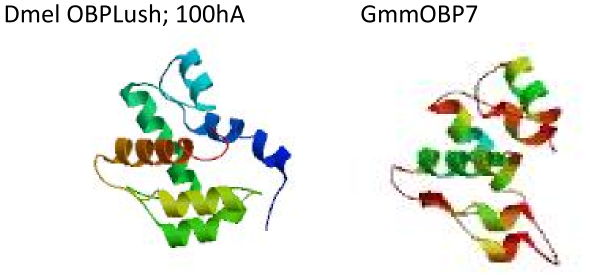
```
3n7hb        ------------------------------MK---LVTFVFAALLCCSMTLGDT-TPRRD 26
GmmOBP8A       ------------------------------MKKYHIYIVTFAITLLMSFGLNNAQKPRRD 30
GmmOBP12       -------------------------------------------------------------
GmmOBP18       -------------------------------------------------------------
GmmOBP19       -------------------------------------------------------------
GmmOBP3      NSEILRKCFEEIDQSQYNSSEVLLEKFKNYAYWSHEEIPCFARCIASEKGWFDIDLSRWN 60
GmmOBP21       ------------------------------------------------------AEDEDWQ 7



3n7hb        AEYPPPEL----------------------LEALKPLHDICLGKTGVTEEAIKKFSDEE 63
GmmOBP8        ENYPPPDF----------------------LKSFKIIHDVCVEKTGATEEAIKEFSDGE 67
GmmOBP12       ------------------------------------------------EAIREFSDGE 10
GmmOBP18       ------------------------------------------HMRYAEYLLAALYNNT 16
GmmOBP19       ------------------------------------------HMRYAEYLLAALYNNT 16
GmmOBP3      KQRLVDELGANMYNYCRFELNRAFKNVCSFAFKGLKCLKQAEMNVIITHNNLLECVKEKS 120
GmmOBP21       PKT---------------------------VADIKSIRNECLKEHPLSNEQITKMKNFE 39
                                       :  . :


3n7hb        IHEDEKLKCYMN-------CLFHEAKVVDDNGDVHLEKLHDSLPSSMH-DIAMHMGKRCL 115
GmmOBP8        IHEDPALKCYMN-------CLFHEVNVVDDAGELHFEKLVRMIPEPFL-EMVKHIIDACE 119
GmmOBP12       VHEDEALKCYMN-------CFFHELGAVDDKGDVHLETLNLIMPGSFV-EAILKPAQHCI 62
GmmOBP18       EHTES-FKCYLQ-------CIFDSLGLVDSNNQVNLEKLINFAPTEIH-EHILELHRACD 67
GmmOBP19       EHTES-FKCYLQ-------CIFDSLGLVDSNNQVNLEKLINFAPTEIH-EHILELHRACD 67
GmmOBP3      ISMDQLLEYYHFPQLEHIPCLFKCFADKSHLYTVNYEWNVLNWLKAFG-PIRNENADISI 179
GmmOBP21       FPDEEEVRQYLL-------CTALKMEVFCAHQGYHPNRIAKQFKMDMNEEEVLEIAEKCH 92
            :  .. *      *          : :      :     .    .


3n7hb        --------YP--ETLCDKAFWLHKCWKQSDPKHYFLV------------- 142
GmmOBP8        SH------IPKGETQCDRAWSWHVCFKQTDPVLYFLP------------- 150
GmmOBP12       --------HPEGDTLCHKAWWFHQCWKKADPEVSNLAQESL---------- 95
GmmOBP18       TQRKLVDVIPAGKDSCDIVYTTSQCYYELKPASREYIEYMMH--------- 109
GmmOBP19       TQRKLVDVIPAGKDSCDIVYTTSQCYYELKPASREYIEYMMH--------- 109
GmmOBP3      CRVN--ANEREKMDICAIMYEEYNCWERLNYNTDGISVTYKKALKKIFNF- 227
GmmOBP21       DS------NPDNSSVDVWAFRGHKCMMSSAIGDKVKAYIKKRQEENAAKNA 137
                      :    *
```

Figure 4: Amino acid sequence of Glossina morsitans morsitans OBPs (OBP8, OBP12, OBP18, OBP19, OBP3 and OBP 21) aligned with that of Anopheles gambie OBP1 (PDB: 3n7h chain B), used as template for the predicted three-dimensional structure



```
3V2LA       ----MLFVFFTLLSCTKKKKIFPLRKSTVEQMMK---SGEMIRSVCLGKT-KVAEELVNG 52
GmmOBP14    --------------------------ATEEQMRS---AANLMRDVCLPKFPKVSKETADG 31
GmmOBP13    ---------------------------TKDDFEK---ILQSCR-----EDMQINENDLRT 25
GmmOBP4     MKSKIIFDTRVRRKVMFRVTLILLAIVTSALFSENRYMEFLADFKHCKRERGVGRFELDR 60
GmmOBP16    -------------------------------------------------------ENFNA 5


3V2LA       LRESKFAD-VKELKCYVNCVMEMMQTMKKGKLNYDASVKQIDTIM--------------P 97
GmmOBP14    IRNGNLSD-NKDAKCYINCVMEMMRTMKKGKFLYEGALKQVDLLM--------------P 76
GmmOBP13    LSASPNDV-SEGVKCYMKCVMEKQGHFKNGALLEEAVIKSLESSP--------------A 70
GmmOBP4     LRVGNLAYPSYEAKCFLGCLYERTGILKNGVLQNDVLKKNVGYIANRVLLDEVLPPCYAV 120
GmmOBP16    FQSIDMEP-DRKFKCYAHCLLSNLKYLN--TFSGKFDIEDFKQ----------------Q 46
      :     **:  *:.   ::  :  .  :..

3V2LA       DELAG-PMRAALDICRTVADG--------IKNNCDAAYVLLQCLSKNNPKFIFP---- 142
GmmOBP14    DSYKE-EYRPGLAKCKDSANG--------IKNNCDAAYAVLSCLRGEITQFVFP---- 121
GmmOBP13    DHNDQNQMSAIVKECKKEIG----------SNECETAFKVSMCLREHKVDFEI----- 113
GmmOBP4     SGTNKCDIAFELKKCFKNVGFDKVWITVPWEDNTDPQYIAAMKLIDDLANVKYRVAFA 178
GmmOBP16    DGIED-EDVAVIAKCKKLYDN--------INDPCEYGFNILQCILMFEPTE------- 88
```

Figure 5: Amino acid sequence of Glossina morsitans morsitans OBPs (OBP14, OBP13, OBP4 and OBP16) aligned with that of Anopheles gambie OBP20 (PDB: 3V2L chain A), used as template for the predicted three-dimensional

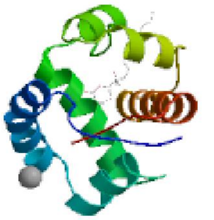


```
1oohA       --------------------------------------------------MKHW------ 4
GmmOBP7     MKLITVIVFSIDFLLFIDASPSGVQEGIVLHQCLAPFGGYTLENDQRLQRFKQWSDTYEE 60
                                                          :*:*

1oohA       ----------------------------KRRSSAVFAIVLQVLVLLLPDPAVAMTMEQFL 36
GmmOBP7     FPCFTNCYLNNMFNIYNETQGFNEENVIKRFGRSVYNACKEKLIQGNNSCEIAYNGFHCL 120
                                        ** .:*:   : *:   . :* .  : *

1oohA       TS--------LDMIRSGCAPKFKLKTEDLDRLRVGDFN-------FPPSQDLMCYTKCVS 81
GmmOBP7     INREDDPFILIDNIEDISMEAKRAMKECLHKFNTDEWQYLSDYVRFPVQEPIPCYTRCFV 180
```

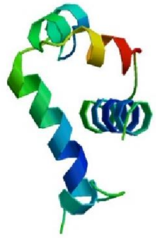
```
          .      :* *.. .    :  .* *.::...:::       ** .: : ***:*.

1oohA         LMAGTVNKKGEFNAPKALAQLPHLVPPEMMEMSRKSVEACRDTHKQFKESCERVYQTAKC 141
GmmOBP7         YKMQLYNHRLRSWNIAAMQRLLG-VPAEHANIEN----CLSLSKRRNNNMCAWIYKEMTC 235
               *:: .     *: :*    **.*  ::..    .   :::: :: *   :*:  .*

1oohA         FSENADGQFMWP 153
GmmOBP7         FSLSQ------- 240
           ** .
```

Figure 6: Amino acid sequence of Glossina morsitans morsitans OBP 9 aligned with that of Drosophila melanogaster OBP Lush(PDB: 1ooh chain A), used as template for the predicted three-dimensional structure

CquiOBP; 30gn



GmmOBP10

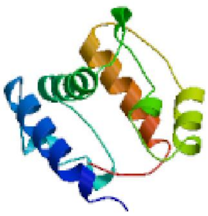


```
30gn          MAARCAKTLVLFSAVLGIAVVVLADVTPRRDAEYPPPELLEALKPLHDICAKKTGVTDEA 60
GmmOBP10        ----------------------------------------------------------EA 2
                                               **

30gn          IIEFSDGKIHEDEKLKCYMNCLFHEAKVVDDNGDVHLEKLHDSLPNSMHDIAMHMGKRC- 119
GmmOBP10        IKEFSEGNIHEDEALKCYMNCFFHELGLVDDKGDVHLETLHQSMPGSFVDLILKPAQHCV 62
           * ***:*:***** *******:***  :***:******.**:*:*.*: *: :: ..:*

30gn          YPEGENLCEKAFWLHKCWKQADPKHYFLV--- 148
GmmOBP10        HPEGDTLCHKAWWFHQCWKKADPVVSNLMEET 94
           :***:.**.**:*:*:***:***    *:
```

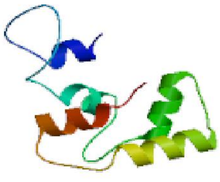
Figure 7: Amino acid sequence of Glossina morsitans morsitans OBP 10 aligned with that of Culex quinquefasciatus CquiOBP (PDB: 30gn), used as template for the predicted three-dimensional structure.

OBP 24

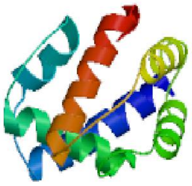
OBP 18



OBP2A



OBP 13



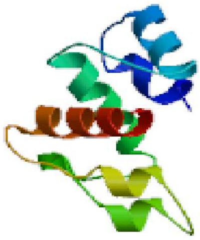
OBP 25

OBP 7

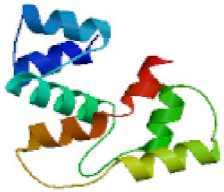
OBP8A

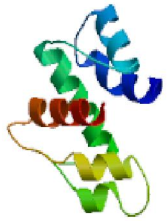
OBP1

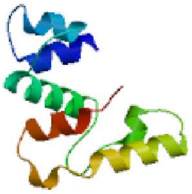

OBP 23

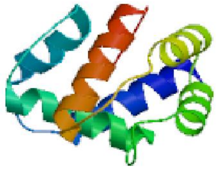

OBP 16

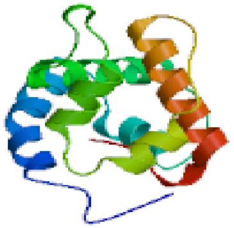

OBP5B

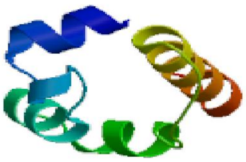

OBP12

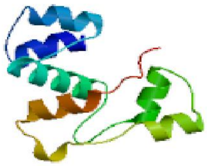

OBP 10

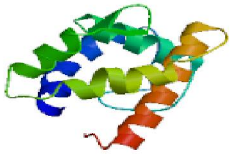

OBP 22

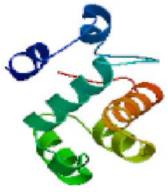

OBP6


OBP26

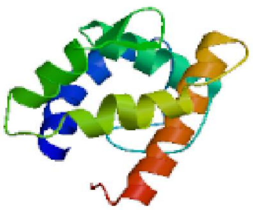

OBP   17


OBP 6

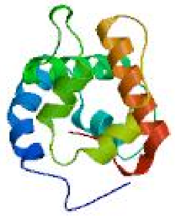

OBP9


OBP 20

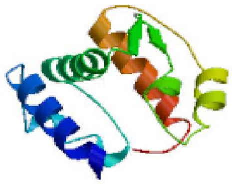

OBP 21

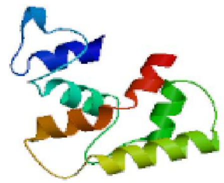

OBP   18

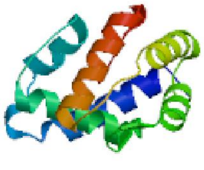

OBP5A


OBP 28


OBP 19


OBP 3


OBP 4


OBP 8B

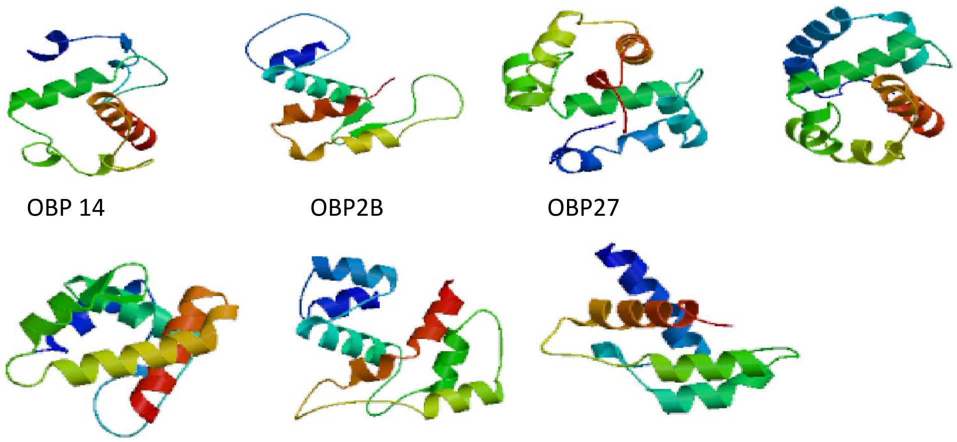


Figure 8: Predicted Glossina morsitans morsitans OBPs structure

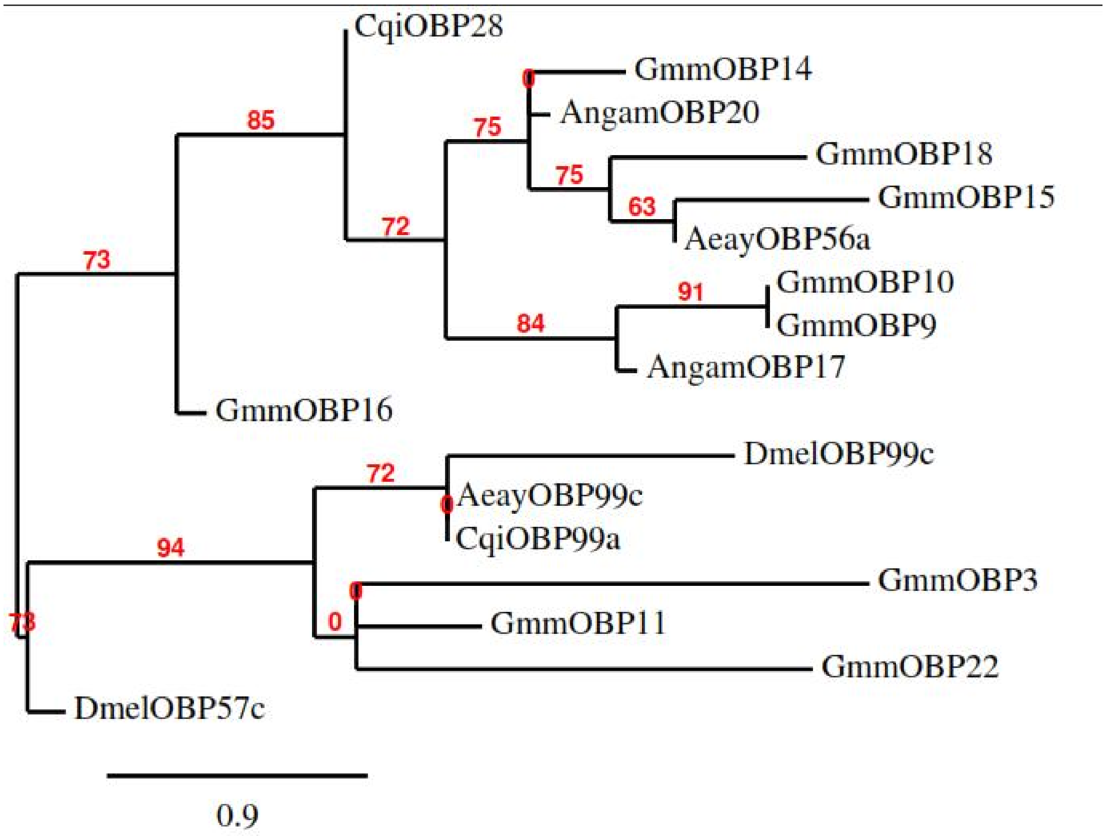


Figure 9: A dendrogram of retrieved GmmOBPs sequences against mosquitoes (An. gambiae, Ae. aegypti and Cu. quinquefasciatus) and fruitfly (D. melanogaster)