



**UNIVERSIDADE ESTADUAL DE FEIRA DE
SANTANA
PROGRAMA DE PÓS-GRADUAÇÃO EM
BIOTECNOLOGIA**



ISIS BUGIA SANTANA

**MODELAGEM COMPARATIVA E TRIAGEM VIRTUAL
HIERÁRQUICA PARA IDENTIFICAÇÃO DE MODULADORES
DAS OBPS DE *LUTZOMYIA LONGIPALPIS***

ISIS BUGIA SANTANA

**MODELAGEM COMPARATIVA E TRIAGEM VIRTUAL
HIERÁRQUICA PARA IDENTIFICAÇÃO DE MODULADORES
DAS OBPS DE *LUTZOMYIA LONGIPALPIS***

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia, da Universidade Estadual de Feira de Santana como requisito parcial para obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Manoelito Coelho dos Santos Junior.

Feira de Santana, BA
2016

Santana, Isis Bugia

S223m Modelagem comparativa e triagem virtual hierárquica para identificação de moduladores das OBPS de *LUTZOMYIA LONGIPALPIS* / Isis Bugia Santana. –Feira de Santana, 2016.
90 f.: il.

Orientador: Manoelito Coelho dos Santos Junior

Dissertação (Mestrado) – Universidade Estadual de Feira de Santana, Programa de Pós-Graduação em Biotecnologia, 2016.

1. Leishmaniose. 2. *Lutzomyia longipalpis* – Mosquito-palha. 3. Proteínas Ligadoras de Odor - OBPS I. Santos Junior, Manoelito Coelho dos, orient. II. Universidade Estadual de Feira de Santana. III. Título.

CDU: 616.993.161

Dedico este trabalho a Deus e à minha família, os únicos que sabem o quanto foi duro, difícil e longo o caminho para chegar até aqui.

AGRADECIMENTOS

Agradeço incondicionalmente e louvo ao Deus que me dá o fôlego da vida, que se faz presente e me encoraja a prosseguir. Agradeço à minha mãe, que tanto abdicou para zelar tão bem de sua família, e que junto aos meus irmãos faz do nosso lar um lugar de aconchego, pessoa cuja força e fé eu espero herdar. Aos meus irmãos, agradeço a oportunidade de abrir o caminho, de incentivar e de carregar o peso de dar bons exemplos, como irmã mais velha que sou. Tenham a certeza de que minha jornada seria, no mínimo, impalatável sem vocês. Sou grata a todos os que me motivaram, à família, aos que amo, aos amigos e conhecidos, inclusive a aqueles que utilizaram de uma abordagem reversa, pois a cada barreira que eu imaginava ser intransponível sentia que não tinha o direito de desistir. Aos meus gatos, em especial agradeço a Rubi, pelas tantas baratas recebidas nas madrugadas em que tentava estudar, agradeço também pela presença, apoio moral e carinho de Minho e seus queridos filhos. Ao grande amigo Reis (*in memoriam*) pela preocupação, por me ensinar um bocado sobre a vida, e por ter sido capaz de me contar por tantas vezes a mesma história com a mesma graça. Agradeço a toda turma do LMM (o laboratório mais zen e amistoso de que já tive notícia), em especial agradeço a Catiane pelos ensinamentos providenciais e pelo bom humor característico, a Franco pelo apoio, cordialidade e pelos tantos questionamentos semeados. Agradeço novamente a Deus por ter colocado na minha cabeça naquela tarde sem graça, a ideia de procurar o professor Manoelito, e com a cara e a coragem abraçar este prazeroso e novo desafio. Obrigada professor pelo ensino diário, pelo exemplo diário, pela confiança e enorme paciência.

“The beauty of a living thing is not the atoms that go into it, but the way those atoms are put together.”

— [Carl Sagan](#), [Cosmos](#)

RESUMO

A Leishmaniose Visceral (LV) é a segunda doença vetorial mais importante do mundo, transmitida nas Américas por *Lutzomyia longipalpis*, o controle do vetor é indispensável à prevenção da doença. Mas como não é possível identificar onde ocorre a oviposição, o combate é direcionado aos insetos adultos, utilizando armadilhas impregnadas com atrativos químicos. Considerando que as Proteínas Ligadoras de Odor (OBPs) atuam no primeiro nível de seleção dos odores, este trabalho utilizou uma metodologia *in silico* para identificar potenciais moduladores químicos olfativos do vetor baseando-se na estrutura das OBPs e de ligantes conhecidos. Para isso, foram preditas as estruturas tridimensionais (3D) de OBPs de *L. longipalpis* por três métodos de modelagem comparativa. O melhor modelo, predito pelo I-Tasser, foi refinado por Dinâmica Molecular no Gromacs. Então, numa abordagem hierárquica da triagem virtual, os compostos naturais do ZINC¹² mais próximos dos típicos ligantes de OBPs no espaço químico global, fornecido pelo ChemGPS-NP, foram avaliados e escalonados quanto à afinidade com o sítio ortostérico da OBP, pelo acoplamento molecular no DOCK6. Os compostos foram pontuados pelo Gridscore, em seguida, os cem melhores classificados foram submetidos à pontuação pelo Amberscore, que levou em conta a flexibilidade tanto da OBP como dos ligantes acoplados. As conformações de menor energia interagiram com um bolsão hidrofóbico através dos resíduos Met6, Ala9, Gly10, Glu11, Arg14, Met53, Leu74, Phe118, Phe119, Pro120; grupamentos amino formaram pontes salinas com a carboxila do Glu11. Além disso, os resíduos Phe119, Asn29 e Gln69 formaram ligações hidrogênio, sendo que, este último resíduo formou ligações-H aceptoras e doadoras.

Palavras-chave: Leishmaniose. *Lutzomyia longipalpis*. Ecologia química. *Odorant-binding protein*. Modelagem Comparativa. Triagem virtual.

ABSTRACT

The Visceral Leishmaniasis (VL) is the second most important vector-borne disease in the world, transmitted in the Americas by *Lutzomyia longipalpis*, vector control is essential for the prevention of the disease. But since it is not possible to identify the oviposition sites, the fight is directed to adult insects, using traps impregnated with chemical attractants. Whereas the Odorant Binding Proteins (OBPs) act in the first level of odor selection, this work used in silico methodology to identify putative vector olfactory chemical modulators based on the structure of OBPs and known ligands. For this, tridimensional (3D) structure of *L. longipalpis* OBPs were predicted by three comparative modeling methods. The best model, predicted by I-Tasser, was refined by Molecular Dynamics on Gromacs. Then, in a hierarchical virtual screening approach, natural compounds of ZINC12 closer to the typical OBP ligands in global chemical space, provided by ChemGPS-NP, were evaluated and staggered concerning affinity with the orthosteric site from the OBP, by molecular docking on DOCK6. The compounds were scored by GRIDSCORE, then the 100 best classified were submitted to AMBERScore, which took into account the flexibility from both OBP and the docked ligands. The lowest energy conformations interacted with a hydrophobic pocket through residues Met6, Gly10, Glu11, Ala9 Arg14, Leu74, Met53, Phe118, Phe119, Pro120, amino groups and formed ionic interaction with carboxyl of Glu11, Furthermore, Phe119, Asn29 and Gln69 formed hydrogen bonds, this last formed donor and acceptor H-bonds.

Keywords: Leishmaniasis. *Lutzomyia longipalpis*. Chemical Ecology. *Odorant-binding protein*. Comparative Modeling. Virtual screening.

LISTA DE FIGURAS

Figura 1 -	Condições de endemicidade da leishmaniose visceral no mundo, em 2012.	19
Figura 2 -	Ciclo de vida dos parasitas do gênero <i>Leishmania</i> , agentes etiológicos da Leishmaniose Visceral.	20
Figura 3 -	Filogenia de dípteros flebotomíneos, culicídeos e muscídeos por máxima verossimilhança, através da análise de genes codificantes ortólogos.	21
Figura 4 -	Feromônios de <i>Lutzomyia longipalpis</i> .	21
Figura 5 -	COVs de cães infectados por <i>Leishmania</i> .	22
Figura 6 -	Evolução do aparato olfativo nos insetos.	23
Figura 7 -	a - Sensilas de <i>Lutzomyia longipalpis</i> ; b – Representação esquemática de uma sensila olfativa.	24
Figura 8 -	Modelos de mecanismos transduccionais para os receptores olfativos (ORs).	25
Figura 9 -	Tipos de receptores olfativos de membrana nos insetos.	26
Figura 10 -	Esquema geral da percepção olfativa em insetos.	29
Figura 11 -	Crescimento do banco de sequências e do banco de estruturas proteicas.	31
Figura 12 -	Representação esquemática das etapas da Modelagem Comparativa.	32
Figura 13 -	Precisão média dos modelos comparativos em função da identidade sequencial entre alvo e molde.	32
Figura 14 -	Precisão e aplicações dos modelos 3D de proteína obtidos a partir de diferentes técnicas.	33
Figura 15 -	Ângulos torcionais diedros da cadeia principal (Φ , Ψ) e planaridade das ligações peptídicas.	35
Figura 16 -	Equação matemática típica de um campo de força.	36
Figura 17 -	Esquema geral da modelagem comparativa e da triagem virtual hierárquica.	42
Figura 18 -	Equações para o cálculo do <i>Score</i> Total.	46
Figura 19 -	Exemplo de predição de peptídeo sinal (seq. BAM004B04).	51
Figura 20 -	Alinhamento múltiplo das sequências de aminoácidos das OBPs de <i>Lutzomyia longipalpis</i> (sem o peptídeo sinal).	52
Figura 21 -	Similaridade sequencial entre as OBPs de <i>Lutzomyia longipalpis</i> .	54
Figura 22 -	Análise filogenética molecular pelo método de máxima verossimilhança.	55
Figura 23 -	Alinhamento entre as sequências alvo (OBPs 1 – 4) e moldes (3OGN e 3V2L).	57
Figura 24 -	Modelo OBP2 (azul) sobreposto ao Molde PDB ID 3V2L (vermelho).	60

Figura 25 -	Representação gráfica do RMSD do modelo OBP2 (cadeia principal) durante a simulação por DM.	61
Figura 26 -	Representação gráfica da variação do RMSF de cada aminoácido do modelo OBP2 durante a simulação de DM.	62
Figura 27 -	Variação do raio de giro (resultante dos eixos x, y e z) durante a simulação dos 100ns da DM.	62
Figura 28 -	Superfície de energia livre a 300K para o modelo OBP2 em termos de RMS e do raio de giro.	63
Figura 29 -	Modelo OBP2 otimizado e seu respectivo gráfico de Ramachandran.	64
Figura 30 -	Qualidade da cadeia principal do modelo OBP2: a – Qualidade do Ramachandran; b – Planaridade das lig. peptídicas – Ângulo ω ; c - Medida das interações ruins não-ligadas; d - Distorção do carbono- α tetraédrico; e - Energia das ligações de hidrogênio.	65
Figura 31 -	Moduladores conhecidos de OBPs.	67
Figura 32 -	a – Distribuição dos compostos distribuídos no gráfico 3D através da análise de agrupamento hierárquico (HCA); b – Dendrograma mostrando as distâncias euclidianas entre os agrupamentos dos compostos, definidos pelo método de ligação do centroide.	68
Figura 33 -	a – Representação esquemática do sítio 1 predito pelo MetaPocket; b – Região correspondente ao sítio ortostérico do molde PDB ID 3V2L.	70
Figura 34 -	RMSD entre a orientação do ligante cristalográfico (verde) e o ligante acoplado (lilás).	71
Figura 35 -	Mapa de interação 2D do Lig1 com a OBP2.	72
Figura 36 -	Mapa de interação 2D do Lig2 com a OBP2.	73
Figura 37 -	Mapa de interação 2D do Lig3 com a OBP2.	74
Figura 38 -	Mapa de interação 2D do Lig4 com a OBP2.	74
Figura 39 -	Mapa de interação 2D do Lig5 com a OBP2.	75
Figura 40 -	Mapa de interação 2D do Lig6 com a OBP2.	76
Figura 41 -	Mapa de interação 2D do Lig7 com a OBP2.	76
Figura 42 -	Mapa de interação 2D do Lig8 com a OBP2.	77

LISTA DE TABELAS

Tabela 1 -	Tipos de validação utilizadas para avaliar a qualidade do modelo 3D e seus pesos na pontuação relativa.	45
Tabela 2 -	Sequências de OBPs de <i>Lutzomyia longipalpis</i> organizadas quanto ao padrão de espaçamento das cisteínas.	50
Tabela 3 -	Sequências alvo com os moldes estruturais selecionados.	56
Tabela 4 -	Avaliação da qualidade estérica, energética e global dos modelos 3D construídos pelo SWISS-MODEL, I-TASSER e MODELLER.	59
Tabela 5 -	As 05 estruturas mais frequentes durante toda a simulação de DM.	63
Tabela 6 -	Coordenadas no espaço químico dos moduladores de OBPs.	67
Tabela 7 -	Matriz das distâncias euclidianas entre as moléculas extraídas de OBPs.	68
Tabela 8 -	Resultado da predição multiconsensual pelo MetaPocket.	69
Tabela 9 -	Ligantes de menor energia pontuados pelo AMBERSCORE.	71

LISTA DE SIGLAS E ABREVIATURAS

2D	Bidimensional
3D	Tridimensional
7TM	Receptores com sete domínios transmembranares
aa	Aminoácidos
ANOlea	<i>Atomic Non-Local Environment Assessment</i>
BAF	Biblioteca das Antenas de Fêmeas
BAM	Biblioteca das Antenas de Machos
cDNA	DNA complementar
COV	Compostos Orgânicos Voláteis
CSP	<i>Chemosensory Proteins</i>
DE	Distância Euclidiana
FES	<i>Free Energy Surface</i>
gDNA	DNA genômico
GOBP	<i>General Odorant Binding Protein</i>
GR	<i>Gustatory Receptors</i>
IR	Ionotropic Receptors
LV	Leishmaniose Visceral
MOP	<i>Mosquito Oviposition Pheromone</i>
NCBI	<i>National Center for Biotechnology Information</i>
NSO	Neurônio Sensorial Olfativo
OBP	<i>Odorant Binding Protein</i>
ODE	<i>Odorant Degrading Enzyme</i>
OR	<i>Olfactory Receptor</i>
ORCo	<i>Olfactory Co-receptor</i>
PBP	<i>Pheromone Binding Protein</i>
PDB	<i>Protein Data Bank</i>
PDB ID	<i>Protein Data Bank Identifier</i>
RMN	Ressonância Magnética Nuclear
RMSD	<i>Root Mean Square Deviation</i>

RMSF	<i>Root Mean Square Flutuaction</i>
Seq	Sequência
SNMP	<i>Sensory Neuron Membrane Protein</i>
TVBL	Triagem Virtual Baseada no Ligante
TVBR	Triagem Virtual Baseada no Receptor
TVH	Triagem Virtual Hierárquica

SUMÁRIO

1	INTRODUÇÃO	15
2	REVISÃO DA LITERATURA	18
2.1	LEISHMANIOSE VISCERAL	18
2.2	<i>LUTZOMYIA LONGIPALPIS</i> : ECOLOGIA E COMUNICAÇÃO OLFATIVA	20
2.3	PERCEPÇÃO OLFATIVA NOS INSETOS	23
2.3.1	Receptores de Odor (ORs)	25
2.3.2	Outros Receptores de Membrana – Ionotrópicos (IRs), Gustativos (GRs) e Sensoriais da Membrana Neuronal (SNMPs)	26
2.3.3	Receptores Quimiossensoriais Extracelulares – Enzimas Degradantes de Odor (ODEs), Proteínas Quimiossensoriais (CSPs) e Proteínas Ligadoras de Odor (OBPs)	27
2.4	MODELAGEM COMPARATIVA	30
2.5	SIMULAÇÕES DE DINÂMICA MOLECULAR	36
2.6	IDENTIFICAÇÃO/DEFINIÇÃO DO SÍTIO ORTOSTÉRICO	38
2.7	TRIAGEM VIRTUAL HIERÁRQUICA	39
2.7.1	Triagem Virtual Baseada no Ligante	40
2.7.2	Triagem Virtual Baseada no Receptor	40
3	MATERIAIS E MÉTODOS	42
3.1	MODELAGEM COMPARATIVA	43
3.1.1	Identificação das sequências de OBPs de <i>Lutzomyia longipalpis</i>	43
3.1.2	Seleção dos moldes estruturais	44
3.1.3	Predição da estrutura 3D das sequências alvo	44
3.1.4	Validação dos modelos 3D construídos	45
3.1.5	Seleção do melhor modelo	45
3.2	OTIMIZAÇÃO POR MECÂNICA MOLECULAR E REVALIDAÇÃO DO MODELO	46
3.3	IDENTIFICAÇÃO DO SÍTIO ORTOSTÉRICO DA OBP	47
3.4	TRIAGEM VIRTUAL HIERÁRQUICA	48
3.4.1	Triagem virtual baseada no ligante	48
3.4.2	Triagem virtual baseada no receptor (alvo 3D)	48

3.5	CONSTRUÇÃO DOS MAPAS DE INTERAÇÃO	49
4	RESULTADOS E DISCUSSÃO	50
4.1	MODELAGEM COMPARATIVA	50
4.1.1	Identificação e caracterização das sequências de OBP (Alvo)	50
4.1.2	Seleção das sequências alvo e dos moldes estruturais	56
4.1.3	Construção dos modelos comparativos	58
4.1.4	Validação e escolha do melhor modelo	58
4.1.5	Refinamento do modelo	61
4.1.6	Validação do modelo otimizado	64
4.2	TRIAGEM VIRTUAL BASEADA NO LIGANTE	66
4.2.1	Seleção do banco de compostos para as triagens	66
4.2.2	Triagem do banco por similaridade com ligantes conhecidos de OBPs	66
4.3	TRIAGEM VIRTUAL COM BASE NO RECEPTOR	69
4.3.1	Definição do sítio ortostérico da OBP	69
4.3.2	Triagem por acoplamento molecular	70
5	CONCLUSÃO	78
6	REFERÊNCIAS	80

1 INTRODUÇÃO

As doenças transmitidas por vetores estão entre os principais problemas de saúde pública negligenciados, afetam mais da metade da população e causam mais de um milhão de mortes por ano em todo mundo (WHO, 2014). Sua diversidade é abrangente, pois envolve uma pluralidade de agentes etiológicos como protozoários, nematoides, vírus, bactérias e até fungos que se adaptam à microbiota do vetor e dos hospedeiros, modificando neles diversos aspectos comportamentais e fisiológicos que favorecem à finalização do ciclo (READY, 2012; ROGERS, 2007). Em sua maioria, são transmitidas por dípteros hematófagos (mosquitos, flebotomos e moscas), como no caso da malária, tripanossomíase africana, leishmanioses, filariose, febre-amarela, dengue, zika, chikungunya e outras arboviroses (LEMON et al., 2008; WHO, 2014). Dentre os vetores, os mosquitos, em especial, matam (indiretamente) mais que qualquer outro animal no planeta. (LLC, 2014).

A leishmaniose visceral (LV) é uma infecção sistêmica extremamente letal quando não tratada, já que ataca os órgãos vitais do hospedeiro (BRASIL, 2003). Anualmente, são mais de 500.000 novos infectados e cerca de 60.000 óbitos (READY, 2014; WHO, 2010a). A LV é causada por protozoários do gênero *Leishmania*, e transmitida nas Américas por fêmeas flebotomíneas de *Lutzomyia longipalpis*, que localizam seu hospedeiro de maneira análoga aos mosquitos hematófagos, guiados pela pluma de odor dos compostos orgânicos voláteis que emanam dos hospedeiros (READY, 2013). Considerada uma zoonose, pois infecta tanto o homem como outros vertebrados, o cão é o principal reservatório de infecção da doença nos centros urbanos (BRASIL, 2003), no entanto, a tentativa de controlar a doença através da eutanásia dos cães infectados (implementada pelo Ministério da Saúde - Brasil) não conseguiu controlar sua expansão (DI et al., 2008). Além disso, ainda não há vacinas para humanos contra a LV (JAIN; JAIN, 2015), e os medicamentos disponíveis apresentam problemas relacionados à toxicidade, eficácia e resistência (FREITAS-JUNIOR et al., 2012), logo, o controle do vetor é estrategicamente indicado.

O desenvolvimento de novas estratégias para o controle dos vetores é essencial a qualquer programa de prevenção de doenças vetoriais (LEMON et al., 2008; TOWNSON, 2005; WHO, 2014). Nas últimas décadas, a utilização de inseticidas como o diclorodifeniltricloroetano (DDT) foi um importante auxílio no controle do quadro epidemiológico dessas doenças (WHO, 2011), todavia, danos ambientais e à saúde humana, causados pelo acúmulo do composto nos tecidos dos organismos a ele expostos, restringiram sua aplicação (SNEDEKER, 2001; STUETZ, 2006; VAN DEN BERG, 2010). Outro fator importante é que em função da pressão seletiva, tem sido relatado

amplamente o surgimento de vetores resistentes aos inseticidas (HEMINGWAY; RANSON, 2000; OWUSU et al., 2015), salientando a urgência na busca por alternativas no controle dos vetores.

Tendo em vista que os insetos utilizam a comunicação quimiossensorial na percepção do mundo que os cerca, para reconhecer parceiros sexuais, fontes de alimento, locais de oviposição, sinais de alerta (LEAL, 2005), e também na procura por hospedeiros (TAKKEN; VERHULST, 2013), desenvolver um mecanismo que consiga interferir nesta relação vetor-hospedeiro é um passo chave para modular o comportamento do vetor, e assim evitar a transmissão da doença (PICKETT et al., 2010). A percepção do estímulo químico à distância é função do sistema olfativo dos insetos, e, embora sua complexidade não tenha sido totalmente elucidada (GALIZIA; RÖSSLER, 2010), quanto à recepção dos sinais olfativos já se sabe que proteínas massivamente presentes nos órgãos olfativos, as Proteínas Ligadoras de Odor (OBPs - do inglês *odorant binding proteins*) – são o primeiro filtro de sensibilidade e seleção dos odores (LEAL, 2013; VOGT; RIDDIFORD, 1981; XU et al., 2005).

Com a função principal de transportar moléculas hidrofóbicas em fluídos aquosos, as OBPs tanto atuam na captação dos compostos voláteis nos órgãos sensoriais, como em sua liberação a partir das glândulas de feromônios/e ou sexuais do inseto (PELOSI et al., 2014a). Isto trouxe uma nova perspectiva na busca de moduladores olfativos, já que a utilização da estrutura da OBP poderia guiar a identificação de possíveis compostos com afinidade ao sítio do transportador (JAYANTHI et al., 2014; LEAL et al., 2008). Neste contexto, a aplicação das ferramentas da quimioinformática traz a possibilidade de analisar *in silico*, rapidamente, gigantescas bibliotecas de estruturas químicas, e assim, aumentar as chances da identificação de moléculas bioativas, um processo conhecido como triagem virtual (TV) (RODRIGUES et al., 2012).

As triagens virtuais podem ser divididas segundo as seguintes abordagens: i) Triagem Virtual Baseada no Ligante (TVBL) – que seleciona os compostos por meio da similaridade química entre moduladores já conhecidos; e ii) Triagem Virtual Baseada no Receptor - alvo molecular (TVBR) – que visa selecionar compostos capazes de interagir com o sítio ortostérico da estrutura tridimensional do alvo macromolecular (RODRIGUES et al., 2012). Nesta última, quando a estrutura 3D do alvo não foi elucidada, a modelagem comparativa possibilitaria utilizar o conhecimento de estruturas já conhecidas e próximas evolutivamente para predizer a estrutura do alvo, a julgar que durante a evolução, o arcabouço estrutural necessário para a função geralmente é conservado dentro de uma família proteica (WORTH; GONG; BLUNDELL, 2009).

Atualmente, cerca de 22 milhões de compostos estão disponíveis para triagem virtual (IRWIN et al., 2012), mas estudos recentes têm proposto que a busca por moléculas bioativas não exige uma amostragem tão ampla, visto que o espaço químico biologicamente relevante está confinado a uma série de padrões de esqueletos químicos capazes de interagir com os alvos

biológico, e por isso, da mesma forma que os alvos moleculares assumem um conjunto limitado de padrões estruturais, seus moduladores naturais também seriam conservados (HARVEY; EDRADA-EBEL; QUINN, 2015). Logo, os compostos naturais são os mais indicados para realização das triagens, visto que, inerentemente, ocupam o espaço químico biologicamente relevante (KELLENBERGER; HOFMANN; QUINN, 2011).

Deste modo, este trabalho teve como objetivo principal buscar potenciais moduladores químicos do comportamento de *Lutzomyia longipalpis*, através das seguintes metas: Identificação das sequências (alvo) - Proteínas Ligadoras de Odor (OBPs) de *L. longipalpis* e de seus moldes estruturais; Construção através da Modelagem Comparativa das estruturas tridimensionais (3D) das OBPs alvo por diferentes métodos; Avaliação da qualidade geral (Validação) dos modelos 3D construídos; Otimização do melhor modelo 3D validado; e Utilização da Triagem Virtual Hierárquica para selecionar compostos naturais com afinidade ao provável sítio ortostérico do Modelo Comparativo 3D da OBP de *Lutzomyia longipalpis* obtido nas etapas anteriores.

2 REVISÃO DA LITERATURA

2.1 LEISHMANIOSE VISCERAL

As leishmanioses são um complexo de doenças vetoriais importantes do ponto de vista clínico e epidemiológico (DESJEUX, 2004), associadas a fatores sociais como pobreza, migração, ocupação urbana não planejada, destruição ambiental, abandono de animais nas cidades, condições precárias de saneamento, habitação e desnutrição (WHO, 2010b). Causadas por protozoários do gênero *Leishmania* (Kinetoplastida, Trypanosomatidae), parasitas intracelulares obrigatórios das células do sistema fagocítico mononuclear, as leishmanioses afetam tanto o homem como outros vertebrados, e podem ser transmitidas de maneira cruzada, através da picada de vetores flebotomíneos, caracterizando assim seu aspecto zoonótico (BRASIL, 2003). A transmissão é realizada por fêmeas flebótomas, que têm comportamento hematófago no período que antecede à oviposição, já que a alimentação rica em nutrientes é indispensável para maturação de seus ovos (BRASIL, 2003; WHO, 2010a).

O tipo mais grave das leishmanioses é o visceral (LV), também conhecido como “Calazar” ou “Barriga d’água”, uma doença crônica que ataca os órgãos vitais do hospedeiro, com sintomatologia típica que envolve: febre, anemia, hepatoesplenomegalia, emagrecimento progressivo, desnutrição, e em alguns casos hemorragia (OMS, 2013), e que pode levar à morte (DRIEMEIER et al., 2015; PIMENTEL et al., 2014). Estima-se que anualmente haja mais de 60.000 óbitos em função da doença, e que cerca de 500.000 novas pessoas em todo mundo sejam infectadas (READY, 2014).

Os dados mais recentes da Organização Mundial da Saúde (WHO, 2010a) indicam que a maioria desses casos se concentra em apenas seis países: Bangladesh, Brasil, Etiópia, Índia, Nepal e Sudão e Sudão do Sul, (figura 1) (ALVAR et al., 2012). O Brasil é considerado um País endêmico para doença, e concentra 90% das ocorrências da América Latina (BRASIL, 2006). Além do mais, estudos apontam a expansão geográfica da infecção no país, com atenção à notável adaptação do vetor aos centros urbanos (WERNECK, 2010), e o registro dos mais altos índices de endemicidade e mortalidade para as regiões norte e nordeste (MARTINS-MELO, F. et al., 2014).

Figura 1: Condições de endemicidade da leishmaniose visceral no mundo, em 2012.



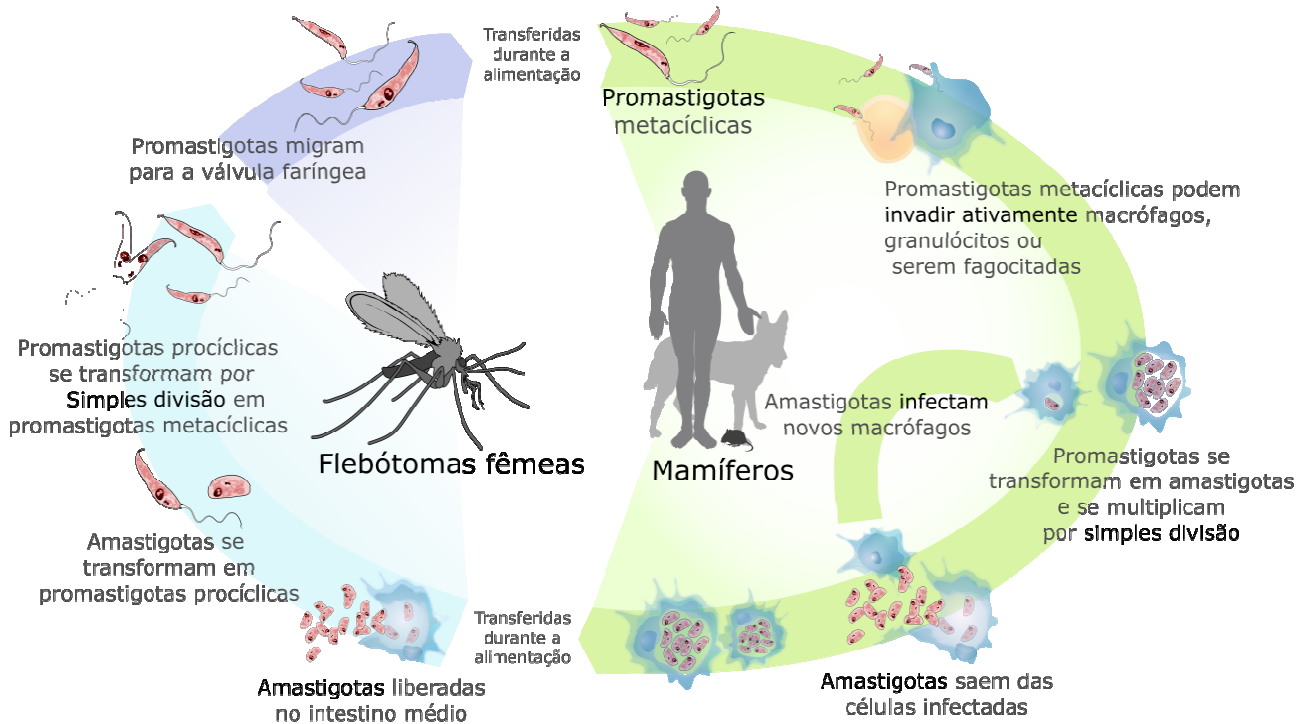
Nas Américas, o parasito *Leishmania infantum* é o agente etiológico da leishmaniose visceral (DANTAS-TORRES, 2006), e o flebotomíneo *Lutzomyia longipalpis* (Diptera, Psychodidae) o vetor mais importante (READY P. et al., 2014). A doença vem se alastrando rapidamente no meio urbano (WERNECK, 2010), onde os cães são os principais reservatórios (hospedeiros intermediários) da LV para o homem (hospedeiro definitivo) (BRASIL, 2006). Mas a infecção pelo parasita já foi registrada em gatos, ratos, aves domésticas, animais silvestres (QUINNELL; COURTENAY, 2009), e há evidências de que carrapatos transmitam mecanicamente a doença, mas neles, o ciclo do parasito não é continuado (CAMPOS; COSTA, 2014).

O ciclo de vida heteroxênico dos parasitas *Leishmania spp.* inicia quando fêmeas flebotomíneas infectadas regurgitam pela probóscide o parasita em sua forma flagelada infectante (promastigota), através da picada, em sua alimentação hematófaga. Já no hospedeiro, os parasitas são fagocitados por macrófagos e outras células de defesa, onde se diferenciam na forma aflagelada (amastigota) e se multiplicam por divisão binária. Ao serem novamente liberados na corrente sanguínea, continuam sendo fagocitados e multiplicam-se mais ainda.

O ciclo é continuado quando os hospedeiros infectados são expostos à picada de vetores flebotomíneos, e há ingestão acidental de células infectadas pela forma amastigota do parasito. A partir daí, uma nova diferenciação ocorre no intestino do vetor, das formas amastigotas para as promastigotas, e estas após se multiplicarem, emergem para o esôfago e faringe do vetor, sendo

passíveis de transmissão através da probóscide, em uma nova picada (figura 2) (DPDX, 2015; WALTERS et al., 1993; WHO, 2010a).

Figura 2: Ciclo de vida dos parasitas do gênero *Leishmania*, agentes etiológicos da Leishmaniose Visceral.



Fonte: (LADYOFHATS, 2008), adaptado

2.2 LUTZOMYIA LONGIPALPIS: ECOLOGIA E COMUNICAÇÃO OLFATIVA

Lutzomyia longipalpis (vetor da LV) é conhecido popularmente como “mosquito-palha”, possui coloração marrom clara, tem o corpo recoberto de pelos e mede de 1 a 3 mm de comprimento. Ativo no período crepuscular e noturno, geralmente, é encontrado dentro ou próximo das casas e de criações de animais (BRASIL, 2003). Devido às semelhanças fenotípicas compartilhadas entre os dípteros e ao comportamento hematófago de suas fêmeas, este flebotomíneo é confundido, às vezes, com os mosquitos, representantes de seu grupo irmão, a família Culicidae (figura 3).

Figura 3: Filogenia de dípteros flebotomíneos, culicídeos e muscídeos por máxima verossimilhança, através da análise de genes codificantes ortólogos.

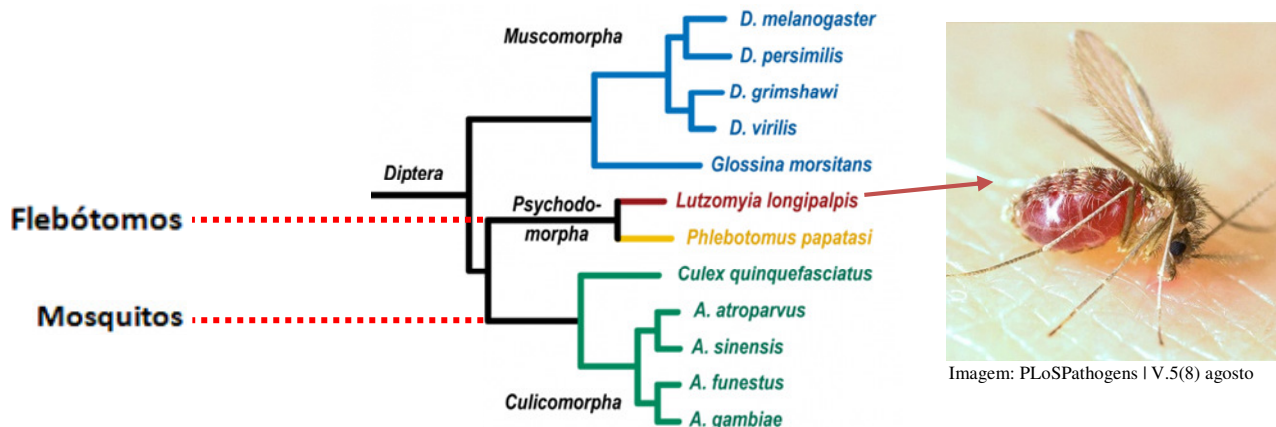


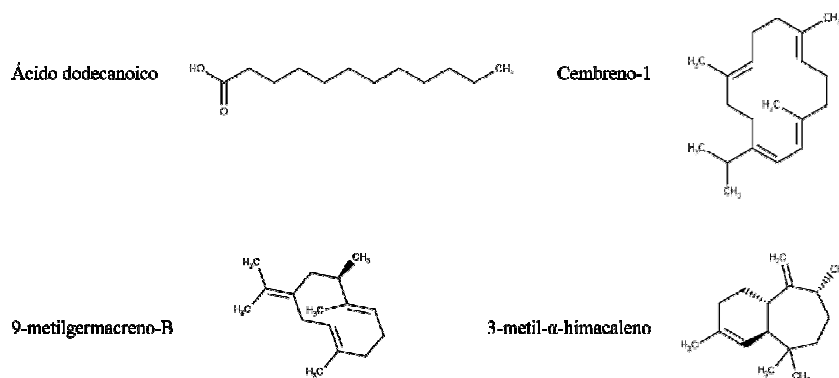
Imagem: PLoSPathogens | V.5(8) agosto

Fonte: (WIEGMANN et al., 2011), adaptado

As fêmeas do flebótomo ovipositam em criadouros de difícil identificação, em ambientes terrestres, úmidos, sombreados e com matéria orgânica (OLIVEIRA, 2013), de modo que, o controle do vetor somente é viável na fase adulta. A oviposição, em *Lutzomyia longipalpis*, é acompanhada da liberação do feromônio ácido dodecanoico (figura 4), que parece induzir as fêmeas a depositarem seus ovos de maneira agregada (quadro 1) (ELNAIEM; WARD, 1992).

Embora apenas as fêmeas sejam hematófagas, acredita-se que o característico comportamento de agregação dos machos em torno dos hospedeiros também seja influenciado por feromônios, alguns já identificados como o 3-metil- α -himacaleno, 9-metilgermacreno-B e Cembreno-1, facilitando assim o encontro de machos e fêmeas para o acasalamento (figura 4) (CASANOVA et al., 2006; JARVISI, 1992).

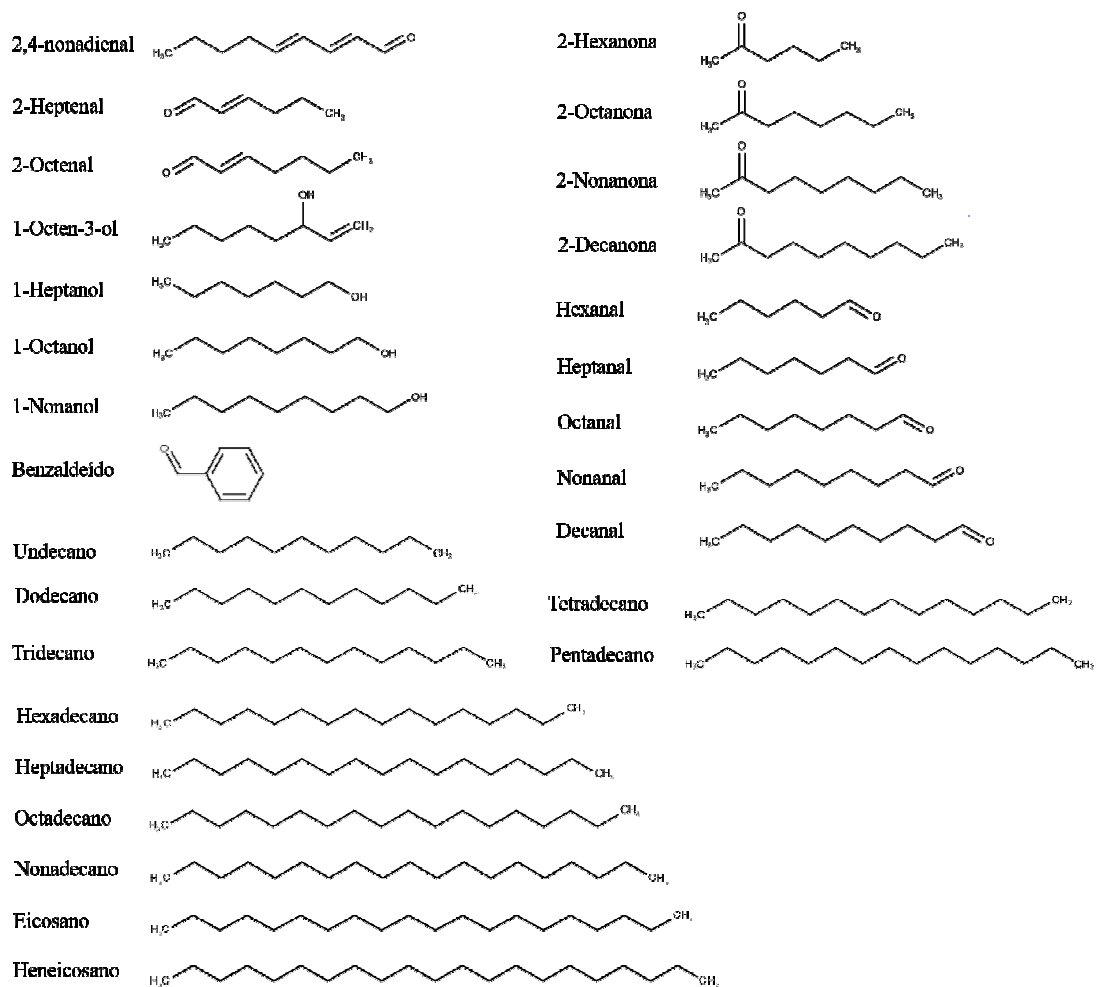
Figura 4: Feromônios de *Lutzomyia longipalpis*.



Fonte: (CASANOVA et al., 2006; JARVISI, 1992)

No estudo da LV, um importante aspecto da transmissão está relacionado à relação entre hospedeiro-vetor no que diz respeito à atração exercida pelo hospedeiro ao vetor (REBOLLAR-TELLEZ, et al., 1999), já que, em insetos vetores, a procura e a seleção de hospedeiros são mediadas por compostos orgânicos voláteis (COVs) que emanam dos hospedeiros (TAKKEN; KNOLS, 2010). Existem indícios, inclusive, de que o parasito interfira nessa relação (KNOLS; MEIJERINK, 1997), pois já foi demonstrado que cães infectados com *Leishmania spp.* produzem um perfil diferenciado de compostos orgânicos voláteis (COVs) que possivelmente exercem maior atração a *Lutzomyia longipalpis*, quando comparados aos cães saudáveis (ANDRADE et al., 2014; DE OLIVEIRA et al., 2008; MAGALHÃES-JUNIOR et al., 2014) (figura 5).

Figura 5: COVs de cães infectados por *Leishmania*.

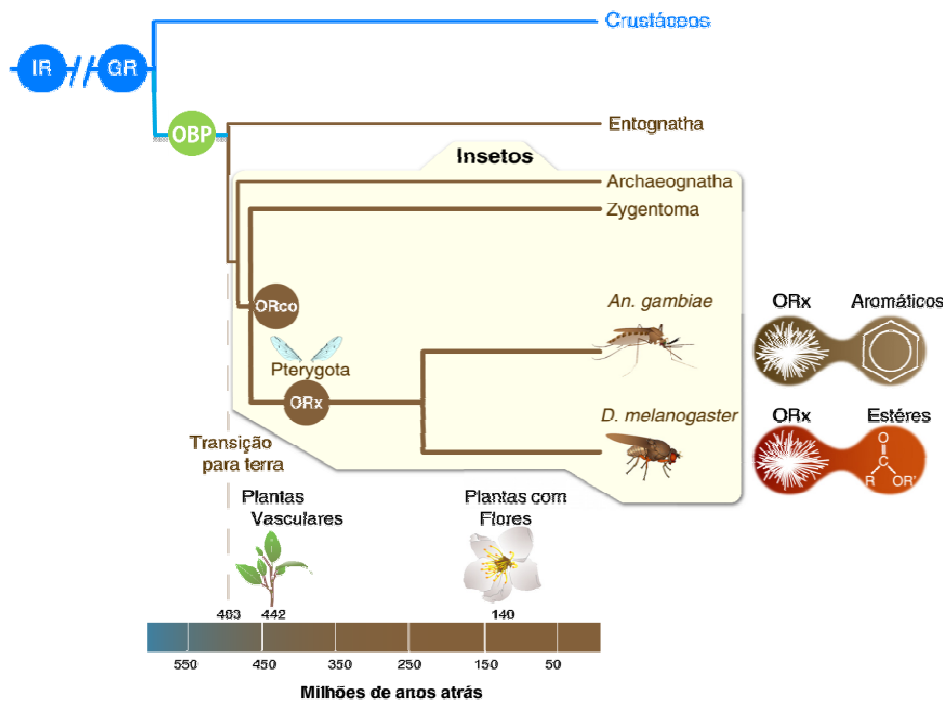


Fonte: (ANDRADE et al., 2014; DE OLIVEIRA et al., 2008; MAGALHÃES-JUNIOR et al., 2014)

2.3 PERCEPÇÃO OLFATIVA NOS INSETOS

Ao longo de sua adaptação evolutiva à vida terrestre, os insetos adquiriram a capacidade de responder a compostos voláteis hidrofóbicos que fornecem uma detalhada descrição do ambiente. Estudos recentes apontam para importantes modificações no sistema quimiossensorial desse grupo que coincidem com o surgimento dos primeiros insetos voadores “Pterigotos” (MISSBACH et al., 2014), uma evidência da estreita relação do voo com a habilidade de identificar e distinguir rapidamente odores fisiologicamente bioativos, em um universo de sinais químicos (figura 6).

Figura 6: Evolução do aparato olfativo nos insetos.

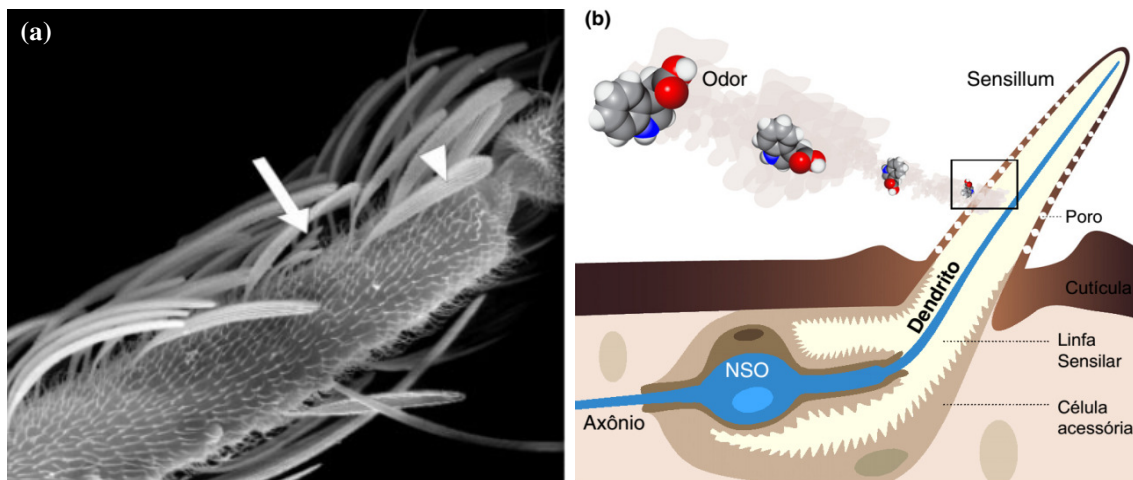


Fonte: (MISSBACH et al., 2014), adaptado

As estruturas morfológicas responsáveis pela olfação nos insetos consistem em regiões altamente especializadas da cutícula das antenas e palpos maxilares, conhecidas como sensilas. A depender de sua estrutura interna e externa, estes órgãos sensoriais também podem captar estímulos gustativos, proprioceptivos, mecânicos, higr- e termossensoriais (STEINBRECHT, 1996). Em *Lutzomyia longipalpis*, seis tipos diferentes de sensilas foram identificadas: tricoides, celocônicas, campaniformes, caéticas, basicônicas, e esquamiformes (FERNANDES et al., 2008).

Dentro das sensilas, e imersas em um fluido aquoso denominado “linfa sensilar”, estão abrigadas as terminações dendríticas dos neurônios sensoriais, cuja principal função é a transdução do estímulo ambiental em impulso elétrico neural. As sensilas especializadas na olfação (basicônicas, tricoides e celocônicas) são caracterizadas pela presença de múltiplos poros ao longo da parede cuticular, através dos quais, os compostos voláteis podem ter acesso à membrana dendrítica dos NSOs (figura 7). No entanto, como explicar a travessia de odores hidrofóbicos através da linfa sensilar?

Figura 7: a - Sensilas de *Lutzomyia longipalpis*; b - Representação esquemática de uma sensila olfativa.



Fonte: (SPIEGEL et al., 2005; SUH; BOHBOT; ZWIEBEL, 2014), adaptado

A descoberta de pequenas proteínas globulares e solúveis, secretadas em grande quantidade no lúmen das sensilas (VOGT; RIDDIFORD, 1981) indicou sua possível participação no transporte dos odores, no mecanismo olfativo pré-neural. Hoje, após uma série de estudos de inibição de expressão (SENGUL; TU, 2010; XU et al., 2005), análises de expressão (SCHYMURA, 2010), determinação estrutural (JANSEN et al., 2007), e das modificações estruturais promovidas pela interação com odores (YIN et al., 2015; ZIEMBA et al., 2013), o papel das Proteínas Ligadoras de Odor (OBPs) e das Proteínas Quimiossensoriais (CSPs) na solubilização dos odores é amplamente reconhecido, e em referência ao seu modo de ação, estão sendo chamadas de “encapsulinas” (LEAL, 2013; LI et al., 2015c; PELOSI; CALVELLO; BAN, 2005).

Apontadas como o primeiro filtro de seleção olfativa para os odores, as encapsulinas (OBPs e CSPs) transportam compostos de baixa solubilidade até a membrana dendrítica do NSOs, onde estão inseridos os receptores olfativos (ORs), ionotrópicos (IRs) e as proteínas sensoriais do sistema neuronal (SNMPs) (FAN et al., 2011), todos eles descritos detalhadamente nas próximas seções.

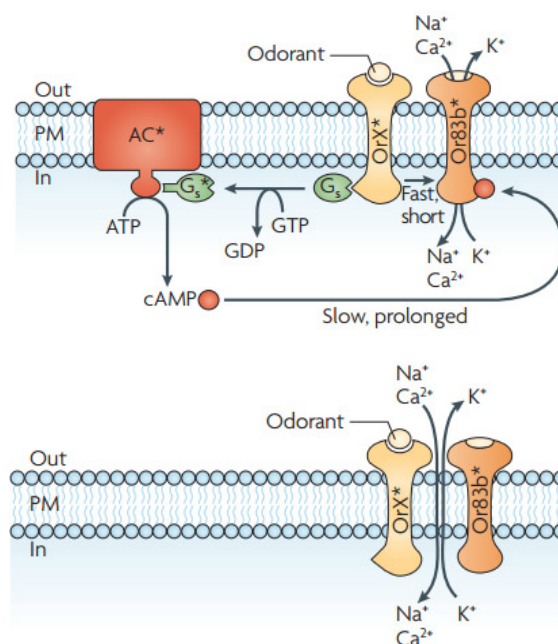
2.3.1 Receptores de Odor (ORs)

Os receptores de odor (ORs) dos insetos são proteínas compostas por sete domínios transmembranares (7TM) com a região N-terminal voltada para o citoplasma (LUNDIN et al., 2007), topologia reversa a do sistema olfativo dos vertebrados e nematoides (BARGMANN, 2006; BUCK; AXEL, 1991). Dois modelos de mecanismos transduccionais são propostos para os ORs:

1) Metabotrópico, no qual o receptor estaria acoplado a uma proteína G, e por meio de sua ativação, sinalizaria a produção de AMP cíclico que, por sua vez, atuaria como segundo mensageiro para abertura de canais iônicos, responsáveis pela despolarização da membrana dendrítica, e, conseqüentemente, pela geração do impulso nervoso (GETAHUN et al., 2013; WICHER et al., 2008);

2) Ionotrópico, no qual os ORs compõem a estrutura do próprio canal iônico (SATO et al., 2008) (figura 8). Este último mecanismo, de ação direta, estaria em conformidade, por exemplo, com a exigência de respostas extremamente rápidas aos estímulos odoríferos em insetos que voam através de uma pluma de odor (HA; SMITH, 2009). O canal iônico dependente de odor dos insetos é um complexo heteromérico formado por um tipo conservado de co-receptor (ORco) da família OR83b (KRIEGER et al., 2003) e por outro OR diverso que, possivelmente, conferiria a especificidade da percepção do estímulo (BENTON et al., 2006).

Figura 8: Modelos de mecanismos transduccionais para os receptores olfativos (ORs)



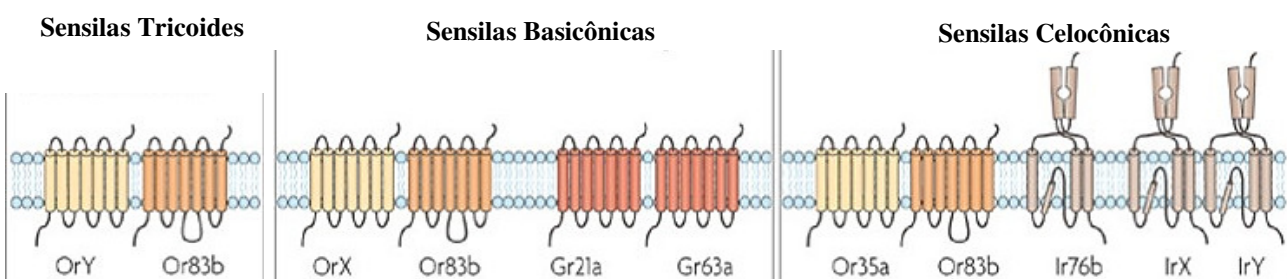
Fonte: (KAUPP, 2010)

2.3.2 Outros Receptores de Membrana - Ionotrópicos (IRs), Gustativos (GRs) e Sensoriais da Membrana Neuronal (SNMPs)

Há um tipo de neurônio olfativo nos insetos, no entanto, onde os co-receptores do tipo ORco não são expressos, e tem sido sugerido que o estímulo olfativo seja percebido por outra família proteica de receptores de odor, os receptores ionotrópicos (IRs). Eles são homólogos aos clássicos receptores ionotrópicos de glutamato, mas que se especializaram na percepção do odor (KAUPP, 2010), especialmente da detecção de ácidos orgânicos, aminas e álcoois (AI et al., 2010; BENTON et al., 2009). Assume-se que os IRs também possam atuar como canais iônicos, com seus próprios co-receptores (BENTON et al., 2009). E, como são expressos no sistema quimiossensorial dos crustáceos (COREY et al., 2013), provavelmente, sejam o tipo mais antigo de receptor olfativo (MISSBACH et al., 2014) (figura 6).

Outra família importante de receptores na detecção olfativa é a dos receptores gustativos (GRs), que, mesmo sendo expressos predominantemente em tecidos gustativos, são os responsáveis pela percepção de CO₂ nas antenas dos insetos (figura 9) (JONES et al., 2007). Os GRs não são homólogos dos ORs, e são reconhecidos por uma característica assinatura terminal carboxílica, além disso, também não se associam aos receptores Orco (SCOTT et al., 2001), ainda que sejam encontrados como dímeros nas membranas dos NSOs.

Figura 9: Tipos de receptores olfativos de membrana nos insetos



Fonte: (KAUPP, 2010)

As proteínas sensoriais da membrana neuronal (SNMPs), embora pouco conhecidas, têm sido apontadas como importantes na detecção odorífera, e associadas inclusive ao reconhecimento de feromônios (JIN; HA; SMITH, 2008). As SNMPs não se assemelham aos típicos receptores de odor 7TM, em vez disso, apresentam dois domínios transmembranares e são homólogas à família proteica CD36 (BENTON; VANNICE; VOSSHALL, 2007; JIN; HA; SMITH, 2008).

2.3.3 Receptores Quimiossensoriais Extracelulares - Enzimas Degradantes de Odor (ODEs), Proteínas Quimiossensoriais (CSPs) e Proteínas Ligadoras de Odor (OBPs)

A capacidade de orientação dos insetos, especialmente os pterigotos, através de uma pluma de odor está diretamente ligada a rápida inativação dos sinais químicos na região olfativa periférica. Esta função é desempenhada pelas Enzimas Degradantes de Odor (ODEs), que permitem que pequenas alterações da composição química dos voláteis sejam rapidamente detectadas (LEAL, 2013). Embora ainda haja relativamente poucos estudos sobre esta classe de proteínas, já foram relatadas como ODEs algumas esterases, citocromo P450, glutathione S-transferase (GST), UDP-glicosiltransferases (UGTs), aldeído oxidases e álcool desidrogenases, encontradas também em outros tecidos não correlacionados à função olfatória (YOUNUS et al., 2014).

Já, as duas classes de proteínas globulares, solúveis e de baixo peso molecular (~15 kDa) formam o grupo das encapsulinas, e, embora não possuam homologia, são capazes de transportar compostos hidrofílicos dentro de uma cavidade hidrofóbica em sua estrutura terciária. As Proteínas Ligadoras de Odor – OBPs (do inglês *Odorant Binding Proteins*) e as Proteínas Quimiossensoriais – CSPs (do inglês *Chemosensory Proteins*) já foram identificadas em diversos representantes dos insetos (hexápodes), todavia, existe uma atenção preponderantemente voltada para as OBPs devido a sua grande diversidade, e por serem mais estritamente ligadas ao transporte de odores do que as CSPs (PELOSI; CALVELLO; BAN, 2005).

As OBPs vêm sendo extensivamente estudadas desde que seu papel na solubilização de odores hidrofóbicos através da linfa sensilar foi apontado por Vogt e colaboradores, em estudos envolvendo o bicho da seda (VOGT; RIDDIFORD, 1981). A presença das OBPs na periferia dos NSOs é tida como uma adaptação evolutiva dos insetos à percepção sensorial da vida terrestre, e sua descoberta solucionou a incógnita de como um sistema sensorial que detectava compostos dissolvidos na água, se tornou capaz de receptor compostos voláteis hidrofóbicos mesmo mantendo um ambiente aquoso dentro da sensila (LEAL, 2013; ZHOU, 2010).

As OBPs são produzidas por células acessórias neuronais e secretadas na linfa sensilar (de acordo com a especificidade do NSO), onde desempenham a função de transportar compostos lipofílicos dos poros sensilares até os receptores da membrana dendrítica (STEINBRECHT, 1996) (figura x). A correspondência estrutural necessária para formação do complexo “OBP-odor” é tida como um fator seletivo primário às moléculas odoríferas (PELOSI; CALVELLO; BAN, 2005).

Por serem proteínas secretadas, suas sequências contém na porção N-terminal uma região de peptídeo sinal importante para o correto endereçamento celular (PUGSLEY, 1989). Após a clivagem do peptídeo, a estrutura composta por seis alfa-hélices adota uma conformação

extremamente estável (ISHIDA Y. LEAL, 2005; PELOSI et al., 2014b), mantida principalmente por ligações dissulfeto entre os resíduos conservados de cisteína que conferem a assinatura do grupo (LEAL, 2013). A família das OBPs (*Insect pheromone/ odorant-binding proteins*) é bastante diversa entre espécies diferentes, e até mesmo as sequências intraespecíficas possuem baixa similaridade, o que torna a identificação de novas proteínas da classe um desafio (XU et al., 2009). Por isso, o padrão de espaçamento entre as cisteínas é utilizado tanto na identificação como na classificação entre os subtipos: Clássico; “Menos C”; “Mais C” e Atípico (FAN et al., 2011).

Embora seu papel principal seja solubilizar odores nas sensilas olfativas, já foi demonstrado que elas podem transportar odores em glândulas que produzem feromônios, e nos locais onde eles são liberados, como por exemplo, nos órgãos reprodutivos (GONZÁLEZ-CABALLERO et al., 2013; LI et al., 2015b), e não se descarta sua participação em outros mecanismos além da olfação (LORENZO, M. G. MELO, 2012; PELOSI et al., 2014a).

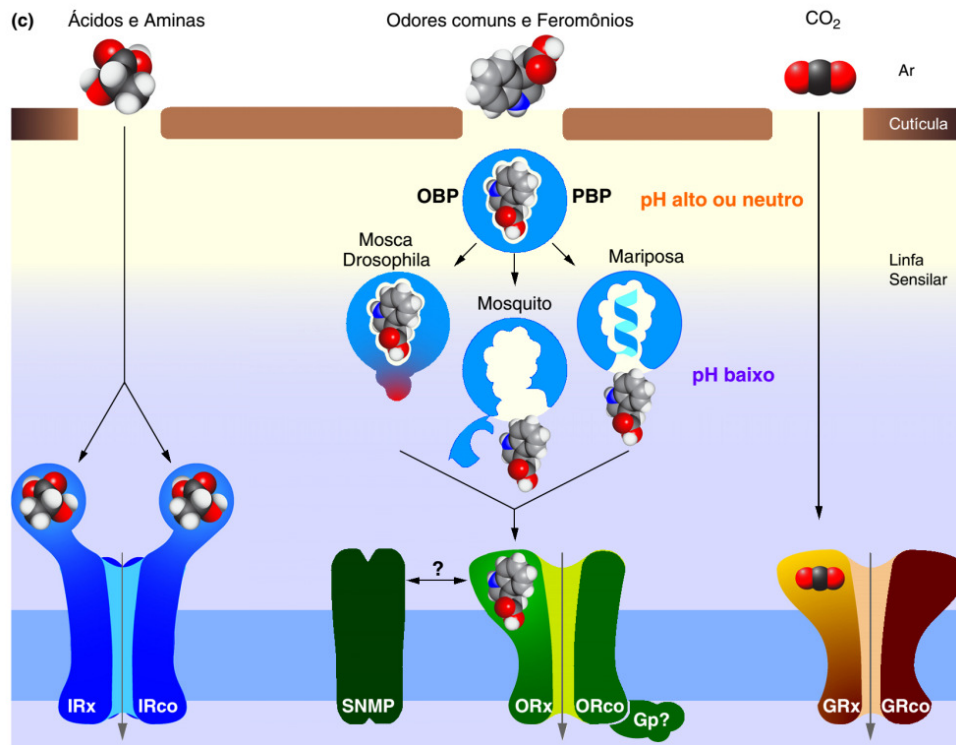
A importância das OBPs na receptação e seleção dos odores hidrofóbicos já foi muito discutida nos ensaios com a OBP do tipo *Lush*, em moscas da espécie *Drosophila melanogaster* (STOWERS; LOGAN, 2008). Inicialmente, o papel das OBPs foi proposto como essencial para a percepção do feromônio 11-*cis*-vaccenil acetato (cVA), já que moscas que não expressavam o gene *Lush* não responderam aos estímulos com o feromônio (KIM, 2001; XU et al., 2005), no entanto, um novo estudo mostrou que a percepção não havia sido totalmente bloqueada, mas que a ausência das OBPs diminuía consideravelmente a sensibilidade do inseto ao feromônio (LAUGHLIN, J. D., HA, T. S., JONES, D. N. M., AND SMITH, 2008).

Outro intenso debate surgiu em torno de qual mecanismo as OBPs utilizavam para interceptar os odores, e de que forma ocorria a subsequente ativação dos ORs, se os odores eram liberados próximo à membrana ou se a estrutura do complexo “OBP-odor” era responsável pela ativação. A OBP *Lush*, é capaz de se ligar a uma série de compostos, mas foi comprovado que apenas o feromônio induz na proteína uma alteração conformacional (ativa), responsável por ativar os receptores de odor. Adicionalmente, ensaios com uma OBP mutante que teve sua estrutura alterada para simular o efeito do cVA mostraram que os ORs foram ativados da mesma maneira (LAUGHLIN, J. D., HA, T. S., JONES, D. N. M., AND SMITH, 2008).

Por outro lado, na determinação estrutural da OBP do bicho-da-seda foram descobertas duas conformações diferentes, cada uma associada a um diferente valor de pH do meio (WOJTASEK; LEAL, 1999). Uma com as típicas seis α -hélices estava presente em valores de pH básicos, similares aos encontrados na linfa sensilar; e outra em pH ácido, na qual o último segmento de alça C-terminal, pela protonação dos resíduos ionizáveis, se transforma em uma sétima alfa-hélice e que ocupa o sítio ortostérico onde antes interagiu o ligante. Esta conformação poderia ocorrer em regiões próximas à membrana dendrítica, que é carregada negativamente, e justificaria um

mecanismo de liberação e interceptação do odor impulsionado pela variação do pH (figura 10) (DAMBERGER et al., 2013).

Figura 10: Esquema geral da percepção olfativa em insetos.



Fonte: (SUH; BOHBOT; ZWIEBEL, 2014), adaptado

As OBPs, no entanto, não são apenas ativadas por feromônios, e este tipo de interação é bem mais difícil de ser identificado. Alguns autores separaram as OBPs em dois grupos: As PBPs (do inglês *Pheromone Binding Protein*), associadas à percepção específica de feromônios, e as GOBPs (do inglês *General Odorant Binding Protein*), que transportam e são ativadas por mais de um ligante (não feromônio) (XU et al., 2009). Mas esta separação é subjetiva, não há nenhum padrão estrutural que a justifique, e certas OBPs referidas como generalistas (GOBPs) podem apenas não terem sido identificadas numa interação mais específica com determinado ligante, a exemplo da OBP *Lush* que só é ativada pelo feromônio cVA, mas consegue se ligar a vários compostos (STOWERS; LOGAN, 2008).

O progresso recente na determinação das especificidades químicas e propriedades funcionais de cada um dos componentes do sistema olfativo tem proporcionado conhecimento sobre os mecanismos subjacentes da codificação de odor em insetos (HALLEM; DAHANUKAR; CARLSON, 2006), no entanto, pela dificuldade em se estudar estruturalmente os receptores de

membrana, as pesquisas sobre a ecologia química dos insetos são substancialmente baseadas no isolamento e identificação de semioquímicos, e em buscar novas estruturas bioativas com base em sua similaridade química (DE OLIVEIRA et al., 2008; LOGAN; BIRKETT, 2007; VERHULST et al., 2010). Por outro lado, relacionar a interação entre os ligantes e seus alvos é fundamental para descoberta de novos potenciais moduladores do odor (BOHBOT; PITTS, 2015). Com este pressuposto, e, tendo em vista que: i) as OBPs interagem de maneira reversível às moléculas odoríferas com afinidade micromolar e atuam na percepção e seleção olfativas; ii) São proteínas extracelulares, de complexidade estrutural muito menor comparadas aos receptores transmembranares, as OBPs são alvos estratégicos para seleção de novos atrativos ou repelentes químicos, e para isto, vêm sendo utilizadas com sucesso numa abordagem conhecida como Ecologia Química Reversa, incluindo também o uso de ferramentas computacionais (JAYANTHI et al., 2014; LEAL et al., 2008).

O uso das ferramentas da quimioinformática amplia consideravelmente a capacidade de busca por moléculas de interesse, além de permitir que o pesquisador teste a eficácia, atividade, toxicidade e bioavaliabilidade, do composto antes de ser submetido aos ensaios *in vitro* e *in vivo*. Mas para isto, é preciso conhecer a estrutura dos alvos moleculares. Em relação ao flebótomo *Lutzomyia longipalpis*, estudos isolados já identificaram algumas OBPs, mas como as estruturas terciárias destas proteínas ainda não foram determinadas não há estudos que avaliem a interação dos semioquímicos e seus alvos moleculares, uma barreira que pode ser transposta com o auxílio das técnicas de modelagem molecular.

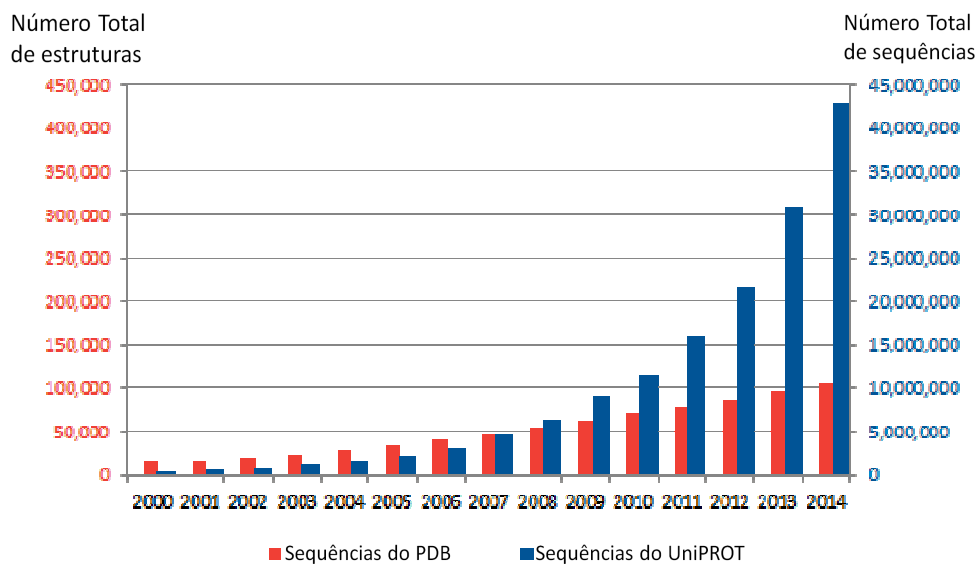
2.4 MODELAGEM COMPARATIVA

O conhecimento da estrutura tridimensional dos alvos moleculares é um requisito básico para o entendimento de sua função biológica e do modo de interação com outras proteínas e micromoléculas (WORTH; GONG; BLUNDELL, 2009). Os métodos experimentais mais utilizados para determinação estrutural das macromoléculas são a cristalografia de raios-x e a ressonância magnética nuclear (RMN), no entanto, eles são demasiadamente complexos e demorados, além de exigirem grandes aportes financeiros e recursos humanos especializados (SANTOS FILHO; ALENCASTRO, 2003).

Por outro lado, a identificação das estruturas primárias ocorre numa velocidade vertiginosa, principalmente na última década, após o surgimento dos sequenciamentos de próxima geração,

ultrapassando os 80 milhões de seqüências (UNIPROT, 2014). O que explica a enorme lacuna entre a quantidade de estruturas 3D conhecidas e a de seqüências proteicas disponíveis (figura 11).

Figura 11: Crescimento do banco de seqüências e do banco de estruturas proteicas.



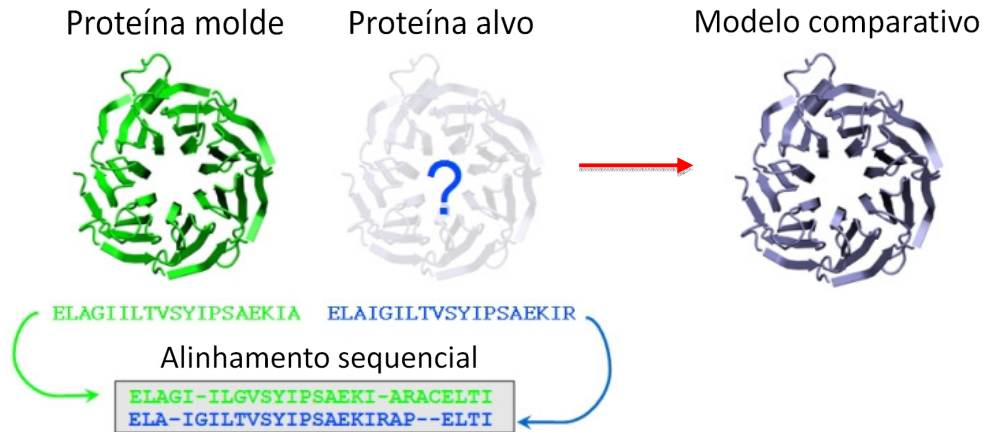
Fonte: (UNIPROT, 2014), adaptado

A estrutura terciária de uma proteína é determinada intrinsecamente pelas interações físico-químicas entre os aminoácidos que a compõem e pelo ambiente químico no qual está inserida, dessa forma, seqüências similares tendem a adotar um dobramento semelhante (CHOTHIA; LESK, 1986). No entanto, ao contrário do que se poderia supor, mesmo considerando a diversidade de funções desempenhadas e a possibilidade de combinações de aminoácidos numa seqüência, devido à pressão seletiva estrutural em favor da manutenção de determinada função biológica as proteínas adotam um restrito número de dobramentos e podem ser agrupadas em poucas famílias (GOVINDARAJAN; RECABARREN; GOLDSTEIN, 1999; WESCHLER; NAZAROFF, 2010). Portanto, proteínas evolutivamente relacionadas (homólogas) quase sempre guardam similaridade estrutural, ainda quando compartilham valores baixos de identidade seqüencial (20 – 30%) (KRISSINEL, 2007).

Tendo em vista a dificuldade na obtenção das estruturas pelos métodos experimentais, a similaridade entre as seqüências proteicas é amplamente utilizada na química computacional para prever *in silico* a estrutura de uma proteína (alvo) com base na estrutura de outra proteína já determinada experimentalmente (molde). Esta técnica conhecida é como modelagem comparativa (ou modelagem com base no molde/ modelagem por homologia), e permite prever muitas vezes modelos tridimensionais acurados que podem ser utilizados inclusive nas triagens virtuais,

fornecendo resultados comparáveis aos obtidos com os modelos experimentais (figura 12) (FAN et al., 2009; LIU; TANG; CAPRIOTTI, 2011; NAQSH E ZAHRA; KHATTAK; MIR, 2013).

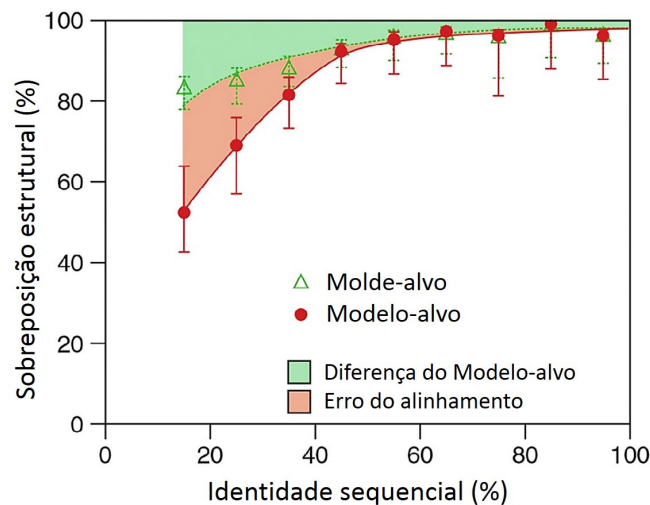
Figura 12: – Representação esquemática das etapas da Modelagem Comparativa.



Fonte: (UNIL, 2014), adaptado

Na modelagem comparativa, a busca por moldes estruturais tende a se concentrar na mesma família proteica, já que as melhores candidatas são aquelas que partilham maior identidade sequencial com o alvo biológico (ROST, 1999). Quando a conservação entre as seqüências de alvo e molde é alta o bastante para fornecer modelos precisos, a sobreposição estrutural é mais fidedigna e, por outro lado, a chance de ocorrerem erros no alinhamento é baixa, especialmente a partir dos 45% de identidade sequencial (figura 13).

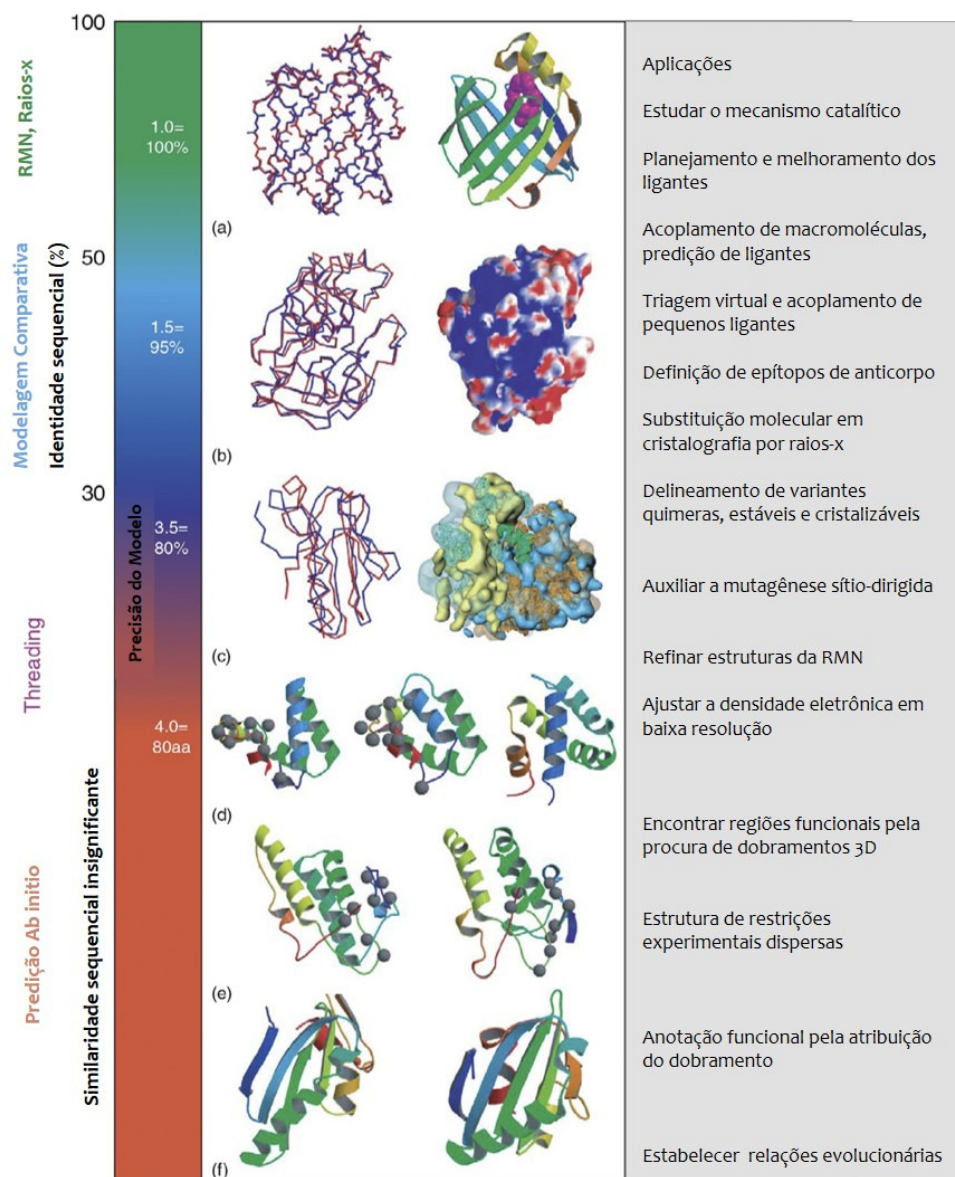
Figura 13: Precisão média dos modelos comparativos em função da identidade sequencial entre alvo e molde.



Fonte: (WEBB et al., 2014), adaptado

Quando se tem por objetivo utilizar um modelo estrutural para realizar triagens virtuais baseadas na estrutura do receptor, a modelagem comparativa é a única técnica *in silico* capaz de fornecer modelos de alta precisão (figura 14). Na escolha do molde, prefere-se optar por aqueles com alta resolução (< 2 Å), cuja quantidade de informações estruturais possibilita a construção de um modelo de maior qualidade. Além disso, é fundamental que os moldes tenham sido determinados em sua forma Holo (complexadas com ligante) (MCGOVERN; SHOICHET, 2003), pois isto assegura que o sítio ortostérico do modelo esteja numa conformação que permita o acoplamento de novos ligantes em potencial (FAN et al., 2009).

Figura 14: Precisão e aplicações dos modelos 3D de proteína obtidos a partir de diferentes técnicas.



Fonte: (WEBB et al., 2014), adaptado

As principais etapas na obtenção de um modelo comparativo são: 1) Identificação de moldes estruturais; 2) Alinhamento sequencial entre proteína-alvo e proteína-molde; 3) Construção do modelo tridimensional (primitivo) da proteína-alvo baseada na estrutura do molde; 4) Refinamento ou otimização do modelo primitivo; e 5) Validação do modelo final. Estes passos podem ser repetidos inúmeras vezes até que um resultado satisfatório seja alcançado, e cada um deles pode contribuir gradualmente para o acúmulo de falhas no modelo (SANTOS FILHO; ALENCASTRO, 2003).

A identificação dos moldes estruturais é facilmente realizada com o auxílio das ferramentas de busca online (PSIBLAST) que através de alinhamentos locais avaliam o grau de identidade dos resíduos correspondentes entre as duas sequências (alvo e molde), bem como a extensão do alinhamento, e o padrão “assinatura” da família proteica considerada para pontuar os melhores candidatos com estrutura 3D elucidada (ALTSCHUL, 1997).

Definidos os moldes, o alinhamento global responsável por especificar as regiões conservadas entre as duas moléculas pode ser realizado de maneira independente, como no próprio programa empregado para construir o modelo comparativo. O algoritmo mais utilizado para este fim é o CLUSTALW (THOMPSON; HIGGINS; GIBSON, 1994), disponível tanto servidores on-line ou pela interface de programas de bioinformática (EMBL-EBI, 2016; KUB, [s.d.]; PETERSEN et al., 2004; SIB, [s.d.]; TAMURA et al., 2013).

A construção dos modelos comparativos pode ser realizada através de vários programas e servidores. E dentre os mais utilizados para predição de modelos (cadeia principal e cadeias laterais) apropriados ao acoplamento molecular estão os programas SWISSMODEL (BIASINI et al., 2014), MODELLER (ESWAR et al., 2006) e o servidor I-TASSER (ZHANG, 2008). Existem três métodos de construção de modelos comparativos, que resumidamente podem ser divididos e conceituados da seguinte maneira:

i) **Corpos Rígidos** – (SWISSMODEL) que utiliza para construção da cadeia principal do modelo as mesmas coordenadas atômicas dos resíduos correspondentes no molde, isto é, das regiões conservadas no alinhamento. As posições espaciais das regiões não conservadas, geralmente, são calculadas empregando curtas minimizações através de um campo de força, por mecânica molecular. As cadeias laterais podem ser definidas a partir de bibliotecas dos confôrmeros encontrados com mais frequência em regiões similares as de outras proteínas, e os hidrogênios são designados nas posições que garantam o maior número de ligações de hidrogênio (BIASINI et al., 2014).

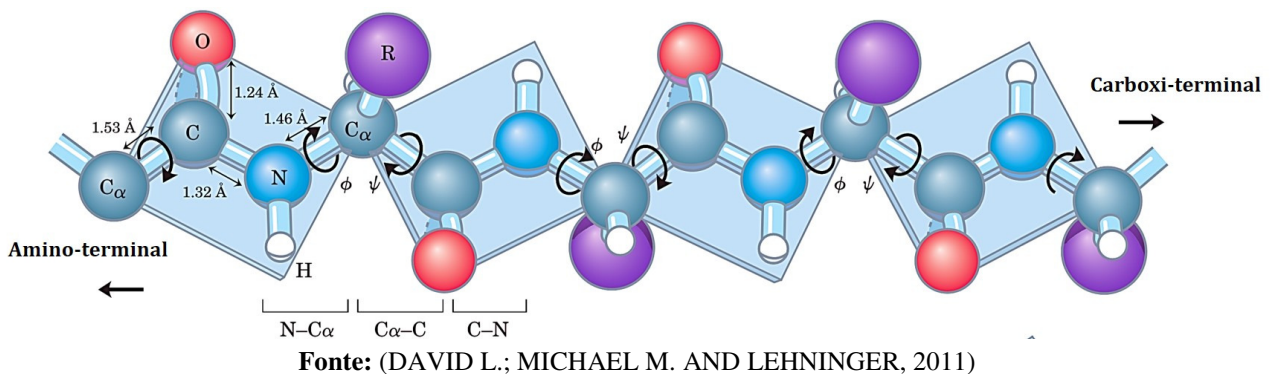
ii) **Satisfação de Restrições Espaciais** – (MODELLER) este método se baseia no alinhamento sequencial entre "molde-alvo" e em alinhamentos entre sequências homólogas (com estrutura conhecida) de diversas outras famílias proteicas, para definir estatisticamente uma zona de

restrições geométricas espaciais, que leva em conta várias características estruturais. Em seguida, o modelo comparativo é construído a partir da zona de restrições, e considerando as implicações energéticas estereoquímicas. Além disso, o modelo é submetido a uma curta otimização, e a estrutura que menos viola as restrições é escolhida. (ESWAR et al., 2006).

iii) Segmentação ou “*Threading*” – (I-TASSER) permite que outras proteínas, não necessariamente homólogas, sejam definidas para molde. Isto porque, regiões que foram bem pontuadas nos alinhamentos entre a proteína-alvo e outras proteínas com estrutura conhecida podem servir como molde para segmentos específicos do modelo, embora, também possa ser designado um molde homólogo já conhecido, como nos métodos anteriormente citados (ZHANG, 2008).

Depois de concluídas as etapas da determinação estrutural, é extremamente necessário que o modelo seja submetido ao refinamento, uma das etapas que mais tem ganhado atenção nos eventos de avaliação crítica da predição de estruturas de proteínas – (CASPs) (MOULT et al., 2011). Isto porque o processo permite corrigir falhas da construção do modelo primitivo, removendo choques estéricos entre os átomos não ligados e minimizando a energia potencial dos ângulos torcionais e de ligação da cadeia principal e das cadeias laterais com valores incorretos (SANTOS FILHO; ALENCASTRO, 2003) (figura 15), e conseqüentemente, aproximando-o do que seria a estrutura nativa da proteína (BRYNGELSON et al., 1995). Para isso, aplicam-se os algoritmos *Steepest Descent* (SD) (ARFKEN, 1985) e Gradiente Conjugado (GD) (HESTENES; STIEFEL, 1952).

Figura 15: Ângulos torcionais diedros da cadeia principal (ϕ , ψ) e planaridade das ligações peptídicas.



Nas etapas que seguem à construção do modelo, a validação teórica ocorre não apenas no final, mas em todo o processo, e consiste em avaliar a qualidade da estrutura predita. O programa PROCHECK (LASKOWSKI et al., 1993) é amplamente utilizado na avaliação estérica e gera diferentes gráficos que demonstram a qualidade geométrica, através das coordenadas da estrutura da proteína, estabelecendo uma pontuação para a qualidade estérica do modelo em estudo. Para validar

a qualidade energética, o servidor ANOLEA (MELO et al., 1997) avalia as interações não locais dos átomos da cadeia principal e laterais e fornece um gráfico, onde as regiões de alta energia potencial estão correlacionadas a erros de posicionamento atômico. Já a avaliação global do tipo Z-SCORE (ZHANG; SKOLNICK, 1998) consiste em calcular o desvio do valor de energia do modelo em sua conformação nativa em relação ao conjunto de possíveis estruturas não nativas, e utiliza-se da função global e local QMEAN6 (BENKERT; KÜNZLI; SCHWEDE, 2009) para comparar a qualidade do modelo frente à outras estruturas proteicas de tamanho similar de média a alta resolução, depositadas nas bases de dados.

Validado o modelo, recomenda-se que uma nova etapa de refinamento seja efetuada através de simulações por dinâmica molecular (DM), tendo em vista que a funcionalidade proteica está intimamente associada à sua plasticidade conformacional, e, por conseguinte, a utilização de um modelo que não considere esta dinâmica terá menos chances de representar realisticamente uma estrutura biologicamente importante (OROZCO, 2014).

2.5 SIMULAÇÕES DE DINÂMICA MOLECULAR

As proteínas possuem estruturas dinâmicas, constantemente em movimento de adaptação às interações físico-químicas internas e externas. Portanto, quando se determina *in silico* a estrutura de um modelo 3D, há de se ter em vista que a estrutura em questão pode não ser a biologicamente mais importante. A física teórica por meio das ferramentas computacionais possibilita a obtenção de uma valiosa quantidade de informações estruturais das proteínas, e é uma alternativa viável para a maioria dos estudos onde a sequência dos aminoácidos que compõem a macromolécula já é conhecida (OROZCO, 2014; XIANG, 2006).

O método de simulação empírica por dinâmica molecular utiliza aproximações quantum-mecânicas na descrição de cada átomo da molécula, cujos parâmetros moleculares como os comprimentos e ângulos das ligações e as distâncias das interações entre os átomos não ligados são descritos em um campo de força, através de uma expressão matemática (figura 16) que descreve a dependência da energia de um sistema de coordenadas sobre suas partículas (GONZÁLEZ, 2011).

Figura 16: Equação matemática típica de um campo de força.

$$U_{\text{pares}} = \sum U_{\text{ligações}} + \sum U_{\text{ângulos}} + \sum U_{\text{diédros}} + \sum U_{\text{interações não-ligantes}}$$

Fonte: (OROZCO, 2014)

Na equação exemplificada do campo de força (figura 16), os três primeiros termos (energia de estiramento das ligações, deformações angulares e rotação de ângulos diedros) referem-se às contribuições intramoleculares ou locais para a energia total, e o último termo descreve a energia das interações não ligadas, repulsão e atração de Van der Waals (potencial de Lennard-Jones 12-6) e interações de Coulomb (eletrostáticas) (GONZÁLEZ, 2011). Além disso, para cada átomo são definidas cargas parciais que contribuirão para o cálculo da energia total do sistema (GONZÁLEZ, 2011).

Por meio da resolução das equações clássicas (newtonianas) da mecânica molecular, a DM consegue calcular as forças que agem sobre cada átomo, bem como seu *momentum* angular correspondente a cada passo da simulação. Os valores obtidos em um passo são utilizados por um algoritmo integrador para definir as posições e energias atômicas do passo seguinte, que por sua vez, serão utilizados para o cálculo das forças e *momentum* subsequentes. Assim, pode-se traçar uma trajetória com as posições adotadas pelo conjunto de átomos da proteína durante toda a simulação, e ainda que a trajetória obtida consista apenas em uma representação da trajetória real, as informações energéticas e estruturais contidas nela podem ser utilizadas na definição de um modelo mais acurado (GONZÁLEZ, 2011) (NAMBA; SILVA; SILVA, 2008).

Isto pode ser realizado através de uma avaliação da variabilidade estrutural e energética do modelo ao longo da simulação. E, considerando que as estruturas nativas tendem a se localizar nas proximidades dos mínimos energéticos de energia potencial, convém que o arranjo 3D escolhido corresponda ao menor valor de energia da superfície de energia livre de Gibbs. Esta superfície pode ser gerada pelo método global dos estados visitados (*visited states method*), numa abordagem que relaciona inversamente a frequência das configurações amostradas aos valores de energia (CHRIST; MARK; VAN GUNSTEREN, 2010).

Um dos pacotes de programas de alto desempenho mais utilizados para realização das simulações de DM com biomoléculas é o GROMACS (*GRO*ningen *MA*chine for *C*hemical *S*imulation), que além de ter o código aberto, é referido entre os mais rápidos para realização dos cálculos de simulação (BERENDSEN; VAN DER SPOEL; VAN DRUNEN, 1995; PRONK et al., 2013). Os parâmetros contidos nos campos de força do GROMACS são disponíveis para todos os átomos, com exceção dos hidrogênios alifáticos, que neste caso são considerados como um só átomo, junto ao carbono a que estão ligados. O campo de força GROMOS53A6 do GROMACS é o mais adequado quando se deseja analisar um sistema solvatado, como o das proteínas globulares, pois seus parâmetros foram especialmente otimizados considerando o cálculo da energia livre da solvatação das cadeias laterais dos aminoácidos (OOSTENBRINK et al., 2004).

Tendo em vista que a estrutura das proteínas globulares é influenciada em grande parte pela interação dos diferentes aminoácidos com a água, as simulações por DM devem incluir o fundamental componente da solvatação, de maneira implícita ou explícita (ZHOU, 2003), sendo que na última, o modelo de água é realmente adicionado ao cálculo, e resulta em uma aproximação mais fiel dos sistemas biológicos (NOLTE et al., 2013). O modelo de água com três cargas parciais SPC-E (*extended simple point charge model*) (BERENDSEN; GRIGERA; STRAATSMA, 1987) implementado no GROMACS é o mais utilizado para solvatação explícita nas simulações de DM, pois representa de forma detalhada às características da água. Além disso, usualmente são adicionados íons ao sistema para que a superfície proteica seja neutralizada, isto é necessário para que as interações eletrostáticas de longo alcance sejam calculadas corretamente (IBRAGIMOVA; WADE, 1998).

A partir do refinamento pelos cálculos de DM, portanto, o modelo comparativo deixa de ser uma modesta extrapolação matemática, e se aproxima cada vez mais do real cenário conformacional responsável pelo desempenho de sua função (FAN; MARK, 2004). Com isto, espera-se que ele possua maior poder preditivo e menos erros associados, quando utilizado para o estudo *in silico* de suas propriedades e possíveis interações.

2.6 IDENTIFICAÇÃO/DEFINIÇÃO DO SÍTIO ORTOSTÉRICO

Nas estruturas proteicas determinadas experimentalmente em complexo com o ligante, o sítio ortostérico, responsável pela função biológica, pode ser facilmente visualizado através da análise das interações com a micromolécula (TURNBULL; EMSLEY, 2013). Mas, quando se trata de um modelo comparativo, não é possível fazer a mesma correlação. Com base no grau de conservação entre as sequências alvo e molde a primeira observação a ser feita, obviamente, é a que a correspondência sequencial e estrutural pode aumentar as chances de a região do sítio no modelo comparativo ser a mesma do molde (ALOY et al., 2003).

Existem, no entanto, outras formas de predição do sítio, implementadas por ferramentas computacionais capazes de avaliar criteriosamente a superfície proteica, e isto pode ser utilizado para aumentar a confiabilidade da identificação. Um dos servidores mais utilizados é o MetaPocket 2.0 (HUANG, 2009), que utiliza uma abordagem de predição consensual das cavidades da proteína, baseado em diferentes métodos de predição. De acordo com as predições de cada método o método pontua qual a cavidade possui mais indícios para ser definida como o sítio ortostérico do modelo. Caso a região predita seja correspondente a encontrada no molde, é possível selecionar os resíduos

que comporão a região do sítio pela proximidade espacial em relação ao ligante cristalográfico, na sobreposição estrutural entre molde e modelo.

2.7 TRIAGEM VIRTUAL HIERÁRQUICA

A procura por micromoléculas com aplicação biotecnológica é voltado, na maioria das vezes, ao desenvolvimento de fármacos (RODRIGUES et al., 2012). Mas em outras áreas o interesse por moléculas com características especiais para modulação de fenômenos biológicos tem crescido, como por exemplo o uso dos semioquímicos direcionado ao controle de pragas e de vetores de doenças (GOULART et al., 2015; LOGAN; BIRKETT, 2007; NAVARRO-SILVA; MARQUES; DUQUE L, 2009; PICKETT; BARASA; BIRKETT, 2014).

Nesse contexto, a utilização das técnicas computacionais possibilita uma rápida análise das bibliotecas virtuais de compostos químicos, aumentando as chances da identificação, além de viabilizar um maior planejamento e direcionamento para as potenciais moléculas bioativas, reduzindo bastante o custo e tempo empregado na pesquisa (VILAR et al., 2011; VYAS, 2008). Para isto, têm sido utilizadas as triagens virtuais (TV), sob duas abordagens básicas: a que se baseia no ligante (TVBL) e a que se baseia no receptor - alvo molecular (TVBR) (LAVECCHIA; DI GIOVANNI, 2013).

A primeira delas tem uma grande utilidade quando apenas são conhecidos os moduladores do fenômeno estudado, ao passo que, na segunda, o alvo biológico é conhecido, sua estrutura tridimensional foi elucidada e o sítio ortostérico definido (RODRIGUES et al., 2012). Entretanto, as duas abordagens usadas em paralelo, ou seja, de maneira isolada e independente, possuem alguns inconvenientes, como a baixa taxa de sucesso e o alto custo computacional (KUMAR; ZHANG, 2015; LI et al., 2015a; SAXENA; ROY, 2012).

Uma nova metodologia, referida como triagem virtual hierárquica ou híbrida, tem sido proposta para aproveitar sistematicamente, e de maneira progressiva, o conhecimento dos ligantes e da estrutura do alvo biológico nas triagens virtuais, com a aplicação gradual de filtros de seleção em diferentes níveis ao longo do processo (BALLESTER et al., 2012; KUMAR; ZHANG, 2015; SAXENA; ROY, 2012). Desse modo, a triagem virtual hierárquica (TVH) faz uso inicialmente da TVBL para filtrar as estruturas com base na similaridade química entre ligantes, que serão utilizadas pela TVBR na identificação compostos que possuam complementariedade com o sítio ortostérico do alvo. A principal vantagem deste método é otimizar o custo computacional utilizado

para o acoplamento molecular (TVBR), ainda impondo restrições aos estudos em larga escala (BALLESTER et al., 2012).

2.7.1 Triagem Virtual Baseada no Ligante

O tamanho e a diversidade da biblioteca virtual a ser filtrada são extremamente importantes na busca por novos compostos bioativos. O banco de compostos do ZINC12 (IRWIN et al., 2012) é um dos maiores acervos mundiais de estruturas químicas para estudos de pesquisa e desenvolvimento de novos compostos bioativos, e foi especialmente compilado e organizado com produtos de diversas coleções em todo o mundo, muitos deles disponíveis para compra. Destaca-se que neste catálogo há um subgrupo para produtos do semiárido baiano (<http://zinc.docking.org/catalogs/uefsnp>) que foi construído e depositado pela equipe do Laboratório de Modelagem Molecular da UEFS.

Na TVBL, a estrutura de ligantes conhecidos é utilizada como molde para identificação de características estruturais, eletrostáticas e hidrofóbicas que sejam importantes para a modulação do alvo biológico (MARSHALL, 2005). Nesta etapa, podem ser utilizados descritores físico-químicos para selecionar no banco de dados os compostos com características físico-químicas similares às das moléculas bioativas (XUE; BAJORATH, 2000), mas devido ao grande número de descritores existentes esta pode ser uma tarefa complexa. Condensar as informações contidas nos descritores físico-químicos em componentes principais (PCs) é a alternativa trazida pelo servidor ChemGPS-NP (LARSSON et al., 2007) para facilitar a análise do espaço químico multidimensional. Desse modo, os compostos de interesse são escolhidos com base na proximidade do espaço químico do ligante, através do cálculo das distâncias euclidianas (BUONFIGLIO et al., 2015).

2.7.2 Triagem Virtual Baseada no Receptor

Tendo em vista que a interação entre ligante e alvo molecular ocorre em regiões específicas da estrutura do alvo, o acoplamento molecular (uma das técnicas da TVBR) utiliza a informação estrutural do alvo para tentar prever quais esqueletos químicos possuem potencial para interagir com seu sítio ortostérico, bem como, as particularidades desta interação (EWING et al., 2001; RODRIGUES et al., 2012; VILAR et al., 2011; VYAS, 2008).

O processo ocorre em diferentes fases, e tem início com o uso de um algoritmo para orientar o ligante em uma série de conformações possíveis no sítio do receptor, a partir daí, os modos de ligação amostrados são pontuados por uma função, que tenta prever a força das interações não-ligadas entre a conformação acoplada dos ligantes e o sítio do alvo, para, finalmente, serem classificados de acordo com a energia de afinidade (EWING et al., 2001).

Existem basicamente três tipos de funções de pontuação: i) baseadas no campo de força: avaliam as interações de van der Waals, eletrostáticas e de torção para pontuar os ligantes; ii) empíricas: a pontuação é derivada da quantidade de interações entre o ligante e o sítio do alvo; e iii) baseadas no conhecimento: analisam estatisticamente as interações intermoleculares de alvos e ligantes depositados em bancos de dados, para derivar potenciais que avaliem a interação amostrada (SOTRIFFER, 1999). No pacote de programas do DOCK6, o uso da função *Grid* (baseada no campo de força), derivada dos parâmetros implementados pelo AMBER (EWING et al., 2001) resulta em análises rápidas e acuradas para avaliação de grandes conjuntos de dados. Um outro método disponível através da função *Amber*, mais elaborado para pontuar os ligantes, considera as modificações intrínsecas do complexo “proteína-ligante” (encaixe-induzido), realizando uma etapa de minimização, seguida de uma pequena simulação por DM e de outra minimização. Além disso o cálculo desta função considera a solvatação do complexo por meio do modelo implícito GB/SA (*Generalized Born/Surface Area*) (ALLEN et al., 2015).

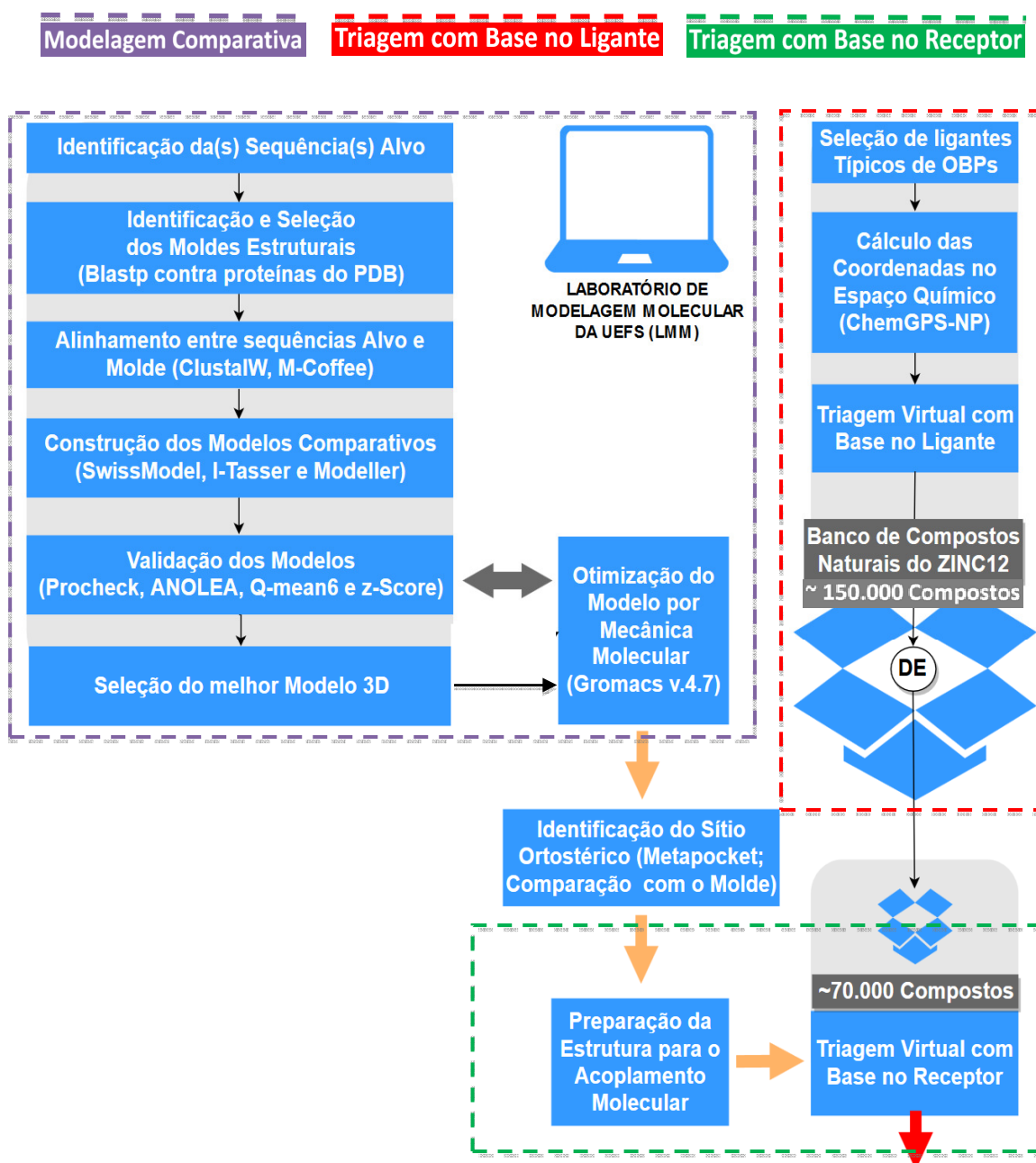
O modo de interação das conformações mais bem pontuadas pode ser cuidadosamente visualizado através de diagramas esquemáticos bidimensionais (2D) “mapas auto-organizáveis”. Nestes diagramas apenas a região de interesse é analisada, onde são apresentadas as interações entre os resíduos do sítio ativo da proteína e o ligante em escala atômica, bem como são especificados os tipos de interação (contatos hidrofóbicos, ligações de hidrogênio, pontes salinas, interações eletrostáticas e de van der Waals) e suas respectivas distâncias (WALLACE; LASKOWSKI; THORNTON, 1996).

Um dos poucos programas disponíveis para visualização das interações entre proteína e ligante em diagramas 2D é o PoseViewWeb (STIERAND; RAREY, 2007b, 2010). As interações são designadas de acordo com os tipos de átomos envolvidos e utilizando um simples critério geométrico. Os valores das distâncias e dos ângulos avaliados são retirados das coordenadas 3D dos arquivos PDB da proteína e mol2 do ligante. O PoseViewWeb consegue avaliar cinco tipos de interações: ligações de hidrogênio, interações metálicas, empilhamentos π - π , interações cátion- π , e contatos hidrofóbicos. A avaliação é feita de acordo com o tipo, valência, e hibridização do átomo do ligante e considerando os átomos do receptor que estão mais próximos ao ligante.

3 MATERIAIS E MÉTODOS

A metodologia geral deste trabalho pode ser dividida em três etapas: 1) Modelagem Comparativa de OBPs de *L. longipalpis*; 2) Triagem Virtual Baseada no Ligante 3) Triagem Virtual Baseada no Receptor, estas duas últimas numa abordagem Hierárquica (TVH) (figura 17).

Figura 17: Esquema geral da modelagem comparativa e da triagem virtual hierárquica.



A primeira parte consistiu em construir modelos comparativos 3D de sequências de OBPs do flebótomo. A segunda parte utilizou informações estruturais de ligantes cristalográficos de OBPs para refinar por similaridade química o espaço amostral dos compostos avaliados na etapa final. A terceira parte, então, usou informações estruturais obtidas na modelagem comparativa para tentar prever *in silico*, por ensaios de acoplamento molecular, quais compostos naturais, dentre os selecionados na segunda etapa, poderiam interagir com as OBPs, e, possivelmente, identificar potenciais moduladores olfativos do comportamento do vetor.

Todas estas análises foram realizadas no Laboratório de Modelagem Molecular (LMM) da Universidade Estadual de Feira de Santana. As etapas citadas foram descritas com maior detalhamento em seus respectivos tópicos a seguir.

3.1 MODELAGEM COMPARATIVA

3.1.1 Identificação das sequências de OBPs de *Lutzomyia longipalpis*

Embora o genoma de *Lutzomyia longipalpis* já tenha sido, em parte, sequenciado (<https://www.hgsc.bcm.edu/sand-fly-genome-project-0>), as famílias gênicas envolvidas em sua quimiorrecepção olfativa ainda não foram identificadas. Desse modo, as sequências das OBPs utilizadas neste estudo foram obtidas na literatura em conjunto com uma busca por prováveis OBPs do vetor nos bancos de dados biológicos do *Vectorbase* (<https://www.vectorbase.org/>) e NCBI (www.ncbi.nlm.nih.gov/).

Esta busca consistiu em utilizar sequências-problema das OBPs de mosquitos vetores e da mosca *Drosophila melanogaster* para uma série de alinhamentos locais (TBLASTX) contra os transcritos preditos no genoma do flebótomo, e com isso, identificar sequências de OBP homólogas. Estas sequências foram confirmadas através das comparações dos perfis dos Modelos Ocultos de Markov (HMMER – com *e-value* menor ou igual a 1, contra sequências proteicas não redundantes).

As sequências de OBPs encontradas foram então submetidas ao servidor SignalIP v. 4.1 (PETERSEN et al., 2011), que identificou as regiões de peptídeo sinal.

3.1.2 Seleção dos moldes estruturais

A seleção dos moldes estruturais para a modelagem comparativa foi realizada submetendo-se as sequências primárias de OBPs de *L. longipalpis* ao PSI-BLAST (ALTSCHUL, 1997) contra sequências com estruturas determinadas experimentalmente, do banco de dados *Protein Data Bank* (PDB) (BERMAN, 2000). Foram selecionados os moldes cristalográficos com no mínimo 45% de identidade sequencial com as sequências alvo, resolução de 2 Å, R-factor de até 20%, e que estivessem na forma Holo (FISER, 2010; GOLDSMITH-FISCHMAN; HONIG, 2003; XIANG, 2006). As sequências primárias dos alvos e moldes foram alinhadas no MEGA6.0 (TAMURA et al., 2013) através do algoritmo CLUSTALW (THOMPSON; HIGGINS; GIBSON, 1994), e então, observou-se o padrão de identidade/similaridade, e o espaçamento entre as cisteínas entre as sequências de OBPs utilizando o servidor WebLogo (CROOKS, 2004).

3.1.3 Predição da estrutura 3D das sequências alvo

Na determinação estrutural *in silico* das OBPs, cada sequência foi modelada por três métodos diferentes: 1) união de Corpos Rígidos – através do servidor SWISS-MODEL (<http://SWISS-MODEL.expasy.org/>) (BIASINI et al., 2014); 2) satisfação de restrições espaciais – no programa MODELLER (ESWAR et al., 2006), através da interface gráfica do CHIMERA (PETTERSEN et al., 2004) e 3) por segmentação (*Threading*) – utilizando a plataforma do I-TASSER (ZHANG, 2008) (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>).

Em cada método de predição, foram submetidos alinhamentos globais entre a sequência alvo e a sequência molde, bem como as informações estruturais do molde para construção dos modelos comparativos 3D. Os alinhamentos submetidos aos servidores SWISS-MODEL e I-TASSER foram realizados pelo algoritmo CLUSTALW, no MEGA6.0, os alinhamentos submetidos ao programa MODELLER foram gerados no CHIMERA 1.9.8, pelo algoritmo CLUSTAL ômega. Especialmente para as predições realizadas no I-TASSER e no MODELLER, onde foram gerados cinco modelos a cada submissão, o modelo melhor pontuado pelo próprio método foi considerado.

3.1.4 Validação dos modelos 3D construídos

Os modelos obtidos foram inspecionados visualmente nos programas UCSF Chimera 1.9.8 (PETTERSEN et al., 2004) e Discovery Studio Visualizer 4.1 (BIOVIA, 2015) com o intuito de verificar possíveis choques estéricos entre os átomos e a qualidade das cadeias laterais e principal. Em seguida, os modelos foram avaliados quanto à qualidade estérica utilizando o programa PROCHECK v.3.5.4 (LASKOWSKI et al., 1993) hospedado no servidor SWISSMODEL (BIASINI et al., 2014), e quanto à qualidade energética através do servidor ANOLEA (MELO et al., 1997) (<http://melolab.org/anolea/>). Além disso, a qualidade global foi obtida, comparando-os com outras estruturas de tamanho similar através das funções de pontuação QMEAN6 (BENKERT; KÜNZLI; SCHWEDE, 2009) e Z-SCORE (ZHANG; SKOLNICK, 1998).

3.1.5 Seleção do melhor modelo

Cada tipo de validação (estérica, energética e global) foi composto por uma ou mais métricas de avaliação, para as quais foi atribuído um peso específico. De modo que fossem privilegiados os modelos com mais resíduos nas regiões favoráveis e com menor valor de energia, e penalizados os que possuísem resíduos em regiões não favoráveis e com alta energia. As validações globais, por já serem relativas, receberam o peso um (1), sendo que, como a Z-SCORE apresentou alguns valores negativos, foi adicionada mais uma unidade (1) (tabela 1).

Tabela 1 – Tipos de validação utilizadas para avaliar a qualidade do modelo 3D e seus pesos na pontuação relativa

Validação	Métrica	Avaliação	Peso	Métrica	Avaliação	Peso
Estérica (Ramachandran)	R_1	% dos aminoácidos em Regiões favoráveis	14	R_3	% dos aminoácidos em Regiões gen. permitidas	-2
Estérica (Ramachandran)	R_2	% dos aminoácidos em Regiões permitidas	-1	R_4	% dos aminoácidos em Regiões gen. permitidas	-3
Energética	E_1	Energia total não local da proteína (E/kT)	15	E_2	% aa com alto valor energético	-5
Global	G_1	Q-Mean6	1	G_2	Z-Score	1

Para que cada validação relacionasse os modelos de uma mesma sequência entre si, os valores foram pontuados de acordo com as equações abaixo (figura 18), que resultaram em valores relativos, cuja somatória deu origem ao *ScoreTotal*, correspondente à qualidade do modelo. Assim, o modelo com maior qualidade geral foi escolhido para as etapas de refinamento e triagem virtual.

Figura 18: Equações para o cálculo do *ScoreTotal*. Onde: R = Ramachandran; E = Energia, e G = Global.

$$\left\{ \left[\left(\frac{\text{Peso}R_1}{R_{1\text{melhor modelo da seq. } x}} \right) \times R_{1\text{modelo da seq. } x} \right] + \left[\left(\frac{\text{Peso}R_2}{R_{2\text{melhor modelo da seq. } x}} \right) \times R_{2\text{modelo da seq. } x} \right] \right. \\ \left. + \left[\left(\frac{\text{Peso}R_3}{R_{3\text{melhor modelo da seq. } x}} \right) \times R_{3\text{modelo da seq. } x} \right] + \left[\left(\frac{\text{Peso}R_4}{R_{4\text{melhor modelo da seq. } x}} \right) \times R_{4\text{modelo da seq. } x} \right] \right\} = \mathbf{Rrelativo} \quad \sum_{\text{do modelo}} \mathbf{ScoreTotal} = \mathbf{Rrelativo} \\ + \mathbf{Erelativo} \\ + \mathbf{Grelativo}$$

$$\left\{ \left[\left(\frac{\text{Peso}E_1}{E_{1\text{melhor modelo da seq. } x}} \right) \times E_{1\text{modelo da seq. } x} \right] + \left[\left(\frac{\text{Peso}E_2}{E_{2\text{melhor modelo da seq. } x}} \right) \times E_{2\text{modelo da seq. } x} \right] \right\} = \mathbf{Erelativo}$$

$$\left\{ \left[\left(\frac{\text{Peso}G_1}{G_{1\text{melhor modelo da seq. } x}} \right) \times G_{1\text{modelo da seq. } x} \right] + \left[\left(\frac{\text{Peso}G_2}{G_{2\text{melhor modelo da seq. } x}} \right) \times G_{2\text{modelo da seq. } x} \right] \right\} = \mathbf{Grelativo}$$

3.2 OTIMIZAÇÃO POR MECÂNICA MOLECULAR E REVALIDAÇÃO DO MODELO

No módulo *pdb2gmx* do GROMACS os estados de protonação dos resíduos ionizáveis da OBP2 foram corrigidos de acordo com o valor aproximado do pH da linfa sensilar (7,5), com o auxílio do programa PROPKA (ROSTKOWSKI et al., 2011), onde foram calculados os valores de pKa dos resíduos de acordo com o seu microambiente químico .

Para reduzir eventuais contatos ruins dos modelos primitivos, procedeu-se à minimização de energia utilizando o GROMACS v4.5.7 (PRONK et al., 2013) empregando-se o método de mecânica molecular com o campo de força GROMOS96 53a6 (OOSTENBRINK et al., 2004). Os arquivos de coordenadas e topologia foram criados, e também adicionados os átomos de hidrogênio pelo módulo *pdb2gmx*. O modelo de solvatação utilizado foi o SPC-E, com caixa de simulação dodecaédrica com dimensões de 60 nm × 45 nm × 48 nm e os vértices espaçados a 13 Å do soluto, colocado no centro da caixa, além disso, avaliou-se a quantidade de íons necessários para neutralização do sistema.

A minimização de energia pelo algoritmo *Steepest Descent* (SD) ocorreu inicialmente com os movimentos da proteína restringidos por uma força de 200 kJ/mol, em seguida, o sistema foi minimizado sem restrições utilizando um critério de convergência energética, em que a energia potencial entre um passo e o outro subsequente da simulação se mantivesse em até 50 KJ/mol. Da mesma forma, a minimização pelo algoritmo Gradiente Conjugado (GC) ocorreu primeiro com restrições, e em seguida, sem restrições, também pelo critério de convergência energética, mas com o valor estipulado de 10 KJ/mol. Após as etapas de minimização de energia, utilizando o integrador Verlet (VERLET, 1967), o sistema foi equilibrado e aquecido à temperatura constante de 300K numa simulação de DM (1ns), o uso do algoritmo LINCS (HESS, B., BEKKER, H., BERENDSEN, H.J.C. & FRAAIJE, 1997) permitiu restringir a movimentação de todas as ligações entre os átomos da proteína, mantendo constantes o número de átomos, o volume da célula de simulação e a energia total do sistema (Conjunto Microcanônico - NVE).

A simulação ocorreu dentro de limites periódicos definidos para as coordenadas X, Y e Z no intuito de garantir que os átomos ficassem dentro da caixa de simulação. Para cada átomo foram consideradas as ligações de longo alcance que estavam próximas até 13 Å, e as interações eletrostáticas foram tratadas pelo método PME – *Particle Mesh Ewald* (DARDEN; YORK; PEDERSEN, 1993). Em seguida, nas mesmas condições, foi realizada uma simulação de DM de 100ns, sem restrições. Para avaliar o equilíbrio mecânico e energético da estrutura ao longo da trajetória, o desvio médio quadrático padrão (RMSD) dos átomos da cadeia principal foi calculado, bem como a flutuação desse desvio médio quadrático por resíduo (RMSF). A energia potencial foi utilizada para avaliar se a proteína se mantinha enovelada durante a simulação, e a análise do raio de giro teve por objetivo mensurar o empacotamento da estrutura.

A escolha da estrutura representativa da trajetória produtiva foi realizada através do método global de estados visitados, que de acordo com a frequência dos valores de RMSD e do raio de giro definiu uma superfície livre de energia de Gibbs (*Free Energy Surface*) (CHRIST; MARK; VAN GUNSTEREN, 2010). A estrutura com o menor valor energético (maior frequência encontrada) foi escolhida para a triagem virtual com base no receptor, o acoplamento molecular.

3.3 IDENTIFICAÇÃO DO SÍTIO ORTOSTÉRICO DA OBP

A determinação do sítio ativo foi feita tanto com base no programa de predição de sítio por análise consensual, MetaPocket 2.0 (HUANG, 2009; ZHANG et al., 2011), como pela análise da região ocupada pelo ligante cristalográfico do molde alinhado ao modelo de OBP, tendo em vista

que o grau de similaridade estrutural entre as duas proteínas, pode contribuir para que a região do sítio no modelo seja a mesma da proteína molde. Portanto, esta possível região do modelo foi avaliada utilizando o programa Discovery Studio 4.5 (BIOVIA, 2015). Definido o ponto central da do sítio ortostérico, foi delimitada uma proximidade de até 8Å para região do acoplamento molecular.

3.4 TRIAGEM VIRTUAL HIERÁRQUICA

3.4.1 Triagem virtual baseada no ligante

Foram selecionados os ligantes de todas as estruturas cristalográficas de OBPs de mosquitos complexadas (Holo), que serviriam de base para o filtro de similaridade química. Para compor o conjunto de compostos a ser filtrado, selecionou-se o banco de compostos naturais (biogênicos) do ZINC12 (IRWIN et al., 2012). Tanto os ligantes de OBPs como os compostos biogênicos foram submetidos ao servidor ChemGPS-NP (LARSSON et al., 2007), que devolveu as coordenadas químicas espaciais de cada um deles, das quais, apenas as coordenadas das três primeiras dimensões (PC1, PC2 e PC3) foram consideradas nas análises.

As posições químicas espaciais dos ligantes de OBPs foram utilizadas, no programa Minitab17 (AREND, 2010), para calcular a posição do “centroide” espacial. Com base na posição do centroide, foi estipulado o valor da distância euclidiana, para o qual, os compostos biogênicos seriam filtrados para construção de um novo banco de compostos para o acoplamento molecular.

3.4.2 Triagem virtual baseada no receptor (alvo 3D)

Esta etapa envolveu a utilização de um modelo comparativo otimizado e validado para os estudos de acoplamento molecular, utilizando o banco de compostos refinado pela Triagem Virtual com Base no Ligante. No programa UFSC CHIMERA 1.9.8 (PETTERSEN et al., 2004), a estrutura da OBP (monômero) foi preparada para o acoplamento com a remoção das moléculas de água e adição de cargas do tipo AMBER ff14SB.

A delimitação do espaço de acoplamento molecular foi realizada por meio dos programas acessórios do DOCK 6.7. Através da superfície molecular de acesso ao solvente do receptor gerada pelo programa DMS (FERRIN et al., 1988) (do inglês *Display Midas System*) na interface do CHIMERA (PETTERSEN et al., 2004), foi construída a imagem negativa da superfície molecular do sítio ortostérico (região de ligação do ligante cristalográfico). Em seguida, os programas SPHGEN e SPHERE_SELECTOR (KUNTZ et al., 1982) foram utilizados respectivamente para gerar e selecionar esferas dentro da região designada para o acoplamento.

As propriedades moleculares para a região do acoplamento foram calculadas pelo programa GRID em sua configuração padrão (MENG; SHOICHET; KUNTZ, 1992; SHOICHET; KUNTZ; BODIAN, 1992). Inicialmente foi utilizada a função de pontuação GRIDSCORE para triagem do banco, posteriormente, os 100 melhores compostos, ranqueados por número, foram submetidos a função AMBERScore com flexibilidade em todo o sistema (ligante e proteína), e aumentando-se o número de ciclos de simulação de DM padrão (3000) em uma ordem de grandeza.

Para analisar a capacidade de reprodução do acoplamento, o ligante cristalográfico retirado do molde que foi usado para construção do modelo da OBP foi minimizado, e, em seguida foi realizado o acoplamento dentro da caixa definida pelo GRID. As coordenadas dos átomos do ligante acoplado foram comparadas com as do ligante cristalizado através do cálculo do RMSD, e foi avaliada a capacidade do algoritmo em posicionar corretamente o ligante em sua pose inicial.

3.5 CONSTRUÇÃO DOS MAPAS DE INTERAÇÃO

O servidor PoseViewWeb (STIERAND; RAREY, 2007a, 2010) foi utilizado para a construção automática dos mapas de interações em 2D entre os resíduos do sítio ortostérico da OBP e os ligantes acoplados, que obtiveram os melhores valores de energia. Para confirmação dos resultados, adicionalmente os complexos “OBP-ligante” foram visualizados em 3D através dos programas CHIMERA (PETTERSEN et al., 2004) e DiscoveryStudio (BIOVIA, 2015).

4. RESULTADOS E DISCUSSÃO

4.1 MODELAGEM COMPARATIVA

4.1.1 Identificação e caracterização das sequências de OBP (Alvo)

Ao todo, foram analisadas 32 sequências de OBPs de *Lutzomyia longipalpis*, das quais 14 são sequências de peptídeos preditos a partir da biblioteca de cDNA extraído das antenas de machos (BAM-) e fêmeas (BAF-) (DIAS, 2008); 07 sequências peptídicas foram retiradas do transcriptoma de glândulas de feromônios sexuais (LLphg-) (GONZÁLEZ-CABALLERO et al., 2013), e 11 sequências foram identificadas a partir de gDNA traduzido (LLOJ-) (DILLON et al., 2006; MCDOWELL; RAMALHO-ORTIGAO; DILLON, 2006) (tabela 2).

Tabela 2: Sequências de OBPs de *Lutzomyia longipalpis* organizadas quanto ao padrão de espaçamento das cisteínas

Código ID ^a	Nº de Aa	Espaçamento Entre as Cisteínas	Código ID ^a	Nº de aa	Espaçamento Entre as Cisteínas
BAF008A01 ^b	128 aa	30/4/38/9/9	BAF004D01 ^f	122 aa	27/4/9/30/11/9
BAM001C09 ^b	128 aa	30/4/38/9/9	BAM009D05 ^f	122 aa	27/4/9/30/11/9
BAM004B04 ^c	120 aa	27/4/41/11/9	BAM012G09 ^f	122 aa	27/4/9/30/11/9
BAM006E08 ^c	120 aa	27/4/41/11/9	LLOJ004774	115 aa	28/4/40/9/9
LLOJ002269 ^c	120 aa	27/4/41/11/9	LLphg-contig_78	116 aa	28/4/40/9/9
BAF004A05 ^c	120 aa	27/4/41/11/9	LLphg-contig_80	116 aa	28/4/40/9/9
LLOJ002267	115 aa	27/4/41/11/9	LLphg-contig_81	116 aa	28/4/40/9/9
LLOJ002270 ^d	115 aa	27/4/41/3/8/9	LLphg-contig_1059	115 aa	28/4/40/9/9
LLOJ002268 ^d	115 aa	27/4/41/3/8/9	BAM005B12	119 aa	27/4/44/9/9
LLOJ002266	120 aa	27/4/41/3/8/9	LLphg-contig_954	122 aa	27/4/9/30/11/9
BAM009G03 ^c	117 aa	27/4/38/9/9	LLphg-contig_1236	122 aa	27/4/9/30/11/9
LLOJ008281 ^c	117 aa	27/4/38/9/9	BAM017F07 ^g	189 aa	18/81/4/43/13/9/9
BAM011G02	123 aa	27/4/38/9/9	BAM020D10 ^g	132 aa	4/65
LLOJ010013	115 aa	28/4/39/9/9	LLOJ000936 ^g	192 aa	19/37/27/4/41/11/9
LLOJ010014	115 aa	28/4/39/9/9	BAM003C01 ^g	139 aa	30/4/34/11/9
LLOJ003968	107 aa	28/4/39/12/9	LLphg-contig_79 ^g	116 aa	4/40/9/9

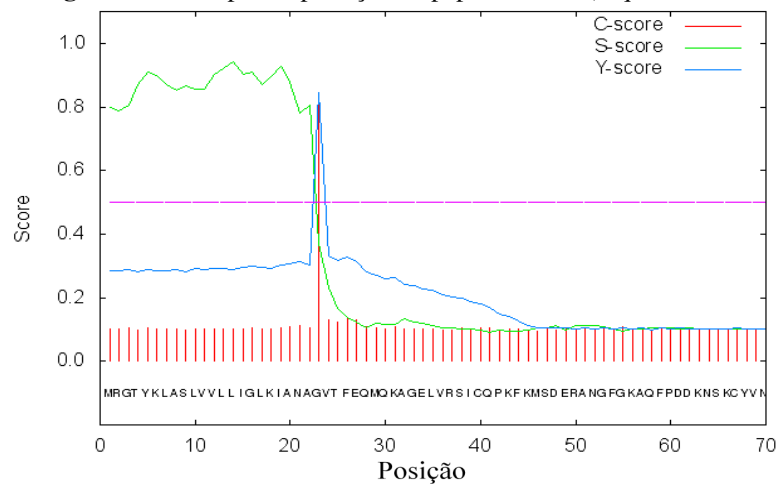
^a Códigos de identificação das OBPs iniciados em “BAF” e “BAM” - peptídeos de OBPs retirados da biblioteca de cDNA das antenas de *L. longipalpis*, iniciados em “LLOJ” - prováveis OBPs retiradas do genoma de *L. longipalpis*, identificadas com base na identidade/similaridade com outras OBPs de dípteros; Iniciados em “LLphg” foram obtidos do transcriptoma de glândulas de feromônios sexuais.

^b Peptídeos idênticos; ^c Peptídeos idênticos (com exceção de BAM004B04 que possui um aminoácido diferente);

^d Peptídeos idênticos; ^e Peptídeos idênticos; ^f Peptídeos idênticos (com exceção de BAM009D05 que possui um aminoácido diferente); ^g Sequências “atípicas” foram retiradas das análises.

Com exceção das sequências BAM017F07 e LLOJ000936, todas as outras apresentaram tamanho correspondente ao relatado para as OBPs de mosquitos (MANOHARAN et al., 2013a) e de outros insetos (PELOSI; CALVELLO; BAN, 2005). A sequência BAM020D10, provavelmente estava truncada ou não teria a estabilidade estrutural típica das OBPs, pois com três cisteínas, seria capaz de formar apenas uma ponte dissulfeto (tabela 2). Para facilitar a análise das sequências, foram realizados diversos alinhamentos múltiplos das OBPs de *Lutzomyia longipalpis* entre si, e com outras OBPs de mosquitos (MANOHARAN et al., 2013b), incluindo algumas que possuíam estrutura tridimensional determinada (MAO et al., 2010; TSITSANOU et al., 2012). Isto possibilitou a detecção de uma região inicial que não se alinhava a sequência das OBPs cristalográficas. Dessa forma, as sequências de OBPs de *L. longipalpis* foram submetidas a predição do peptídeo sinal, e foram identificadas regiões N-terminais preditas como peptídeo sinal para 27 sequências (figura 19). Estas regiões N-terminais (~22 aa) foram removidas das sequências de OBPs, já que não compõem a estrutura funcional das proteínas.

Figura 19: Exemplo de predição de peptídeo sinal (seq. BAM004B04).



Em seguida, um novo alinhamento múltiplo foi realizado (figura 20), (sem as sequências BAM017F07, BAM020D10, LLOJ000936 e LLphg-contig_79, devido à presença de muitos gaps) evidenciando a assinatura da superfamília PBP_GOBP, e ainda, permitindo classificá-las no subtipo das OBPs Clássicas.

Figura 20: Alinhamento múltiplo das seqüências de aminoácidos das OBPs de *Lutzomyia longipalpis* (sem o peptídeo sinal).

BAF008A01	PQTPRRDESYP	PRDILAEYRPIREQ	CLAETGATEKIIQEFS	SDGDHLS	SPVE-DEALKCYMNC	LFH	QI	QV	VDDT	71
BAM004B04	V-----	TFEQM	QKAGELVRSICQPKFKMS	DERANGFGKAQF---	PD-DKNSKCYVNC	I	L	E	N	M
BAM005B12	-----	SEENTKNMLKLAMK	C	KNKTGAKDE	DMHFMSKDL---	PD-TRTTKCLYAC	M	Q	E	L
BAM006E08	V-----	TFEQM	QKAGELVRSICQPKFKMS	DEQANGFGKAQF---	PD-DKNSKCYVNC	I	L	E	N	M
BAM009D05	W-----	RIQTAEDLAKHRTK	C	VEELKIEESAVTEYKKWNF---	TD-DEKTRCYIKC	I	F	N	Q	M
BAM009G03	V-----	EIPEELREGARLLHEHC	V	AQTNVDESLILASVNGYL---	PD-DRTLQCYIDC	L	F	R	T	T
BAM011G02	--ELRRDETYPPPELLRAIKIAHDV	C	VPQTGVLEEH	IKEFS	SDGE---AIE-DAALKCYMNC	L	F	H	E	F
BAM012G09	W-----	RIQTAEDLAKHRTK	C	VEELKIEESAVTEYKKWNF---	TD-DEKTRCYIKC	I	F	N	Q	M
LLphg-contig_954	W-----	RIQTAEDLAKHRTK	C	VEELKTEESAVTEYKKWNF---	TD-DEKTRCYIKC	I	F	N	Q	M
LLphg-contig_1236	W-----	RIQTAEDLAKHRTK	C	VEELKIEESAVTEYKKWNF---	TD-DEKTRCYIKC	I	F	N	Q	M
LLphg-contig_78	V-----	IPKEKEELGRQLLAL	C	SAEVGISDDKVM	TLRDGNF---DNVDVEVKCVIEC	G	F	K	T	G
LLphg-contig_80	V-----	IPKEKEELGRQLLAL	C	SAEVGISDDKVM	TLRDGNF---DNVDVEVKCVVEC	G	F	K	T	G
LLphg-contig_81	V-----	IPKEKEELGRQLLAL	C	SAEVGISDDKVM	TLRDGNF---DNVDVEVKCVVEC	G	F	K	T	G
LLphg-contig_1059	-----	VTDEQRKKAQEVLVS	C	SQENGVTL	DKVAQVRKGDF---ASADEKTQCVVD	C	F	F	K	V
LLOJ002266	V-----	SLEEMK	KAGMLVRSICQPKFNIPDDQADGYAKGIF---	PD-NKDSKCYVNC	V	F	E	N	M	Q
LLOJ002267	-----	MQKAGELVRSICQPKFKIS	DEQANGFGKAQF---	PD-DKNSKCYVNC	I	F	E	N	I	Q
LLOJ002270	-----	MKAGMLVRSICQPKFNIPDDQADGYAKGIF---	PD-DKNSKCYVNC	V	F	E	N	M	Q	S
LLOJ003968	-----	DIEDCAKETKLS	DEELARFQKDDL---	TSEDPTAHCFVK	C	L	L	A	K	N
LLOJ010014	-----	KGYVTKAIMELGLPCAQEVGLS	SSESLKSIQSGDI---	LSIQEGFGCFAD	C	I	A	K	L	G
LLOJ010013	-----	KGYVTRAIMELALPCAHEVGLS	RSLDFIESGNL---	PGLKGGFGCFAD	C	I	A	K	L	G
LLOJ004774	-----	VTDEQRKKAQEVLVS	C	SQENGVTL	DKIAQVRKGDF---ASADEKTQCVVD	C	F	F	K	V
			*	:	.	.	*	*	.	

*Identidade de todos os resíduos de aminoácidos alinhados na coluna indicada.

: Similaridade dos aminoácidos alinhados na coluna indicada

. Baixa similaridade dos aminoácidos alinhados na coluna indicada

- Não há similaridade entre os aminoácidos alinhados na coluna indicada

Seqüências idênticas não foram representadas para economia de espaço.

```

BAF008A01      GHVHFEKLRLLK----VP--DDLKDVGHNMIAQCENPVG---ANLCEKAYWLHTCFKRVDPVHYFL----V 128
BAM004B04      -KFQFEATKKQAEILLP--DDMKEATIKAMVACKGATDG- IKDPC EAAAYSLLKCLHSNHEDFFF-----P 120
BAM005B12      GRLKTQGIKNMAKLMFELDPYLRNVLMEIADECADVQD---DDRCELAYKTSSCAIEAAKRRNI-----D 119
BAM006E08      -KFQFEATKKQAEILLP--DDMKEATIKAMVACKGATDG- IKDPC EAAAYSLLKCLHSNHEDFFF-----P 120
BAM009D05      NGYNVEHLVEQLG--QS--GD-KDKVREQIVKCVDDNPN-KDDKCTWVFRGFNCFKASHLSLIKQSLKGD 122
BAM009G03      GHILFEE----IAHL--LPTEIKEIIVAVTSKCRTHG---KDECE TAYLTVKCYFEADPEHSML----I 117
BAM011G02      GHVHFEKLYNR----IP--ESMKERAKVTMDKCLDPVG---ADLCEKAYWLHKCWKTNDPSHYFL----A 123
BAM012G09      NGYNVEHLVEQLG--QS--GD-KEKVREQIVKCVDDNPN-KDDKCTWVFRGFNCFKANHLSLIKQSLKGD 122
LLphg-contig_954  NGYNVEHLVEQLG--QS--GD-KEKVREQIVKCVDDNPN-KDDKCTWVFRGFNCFKANHLSLIKQSLKGD 122
LLphg-contig_1236 NGYNVEHLVEQLG--QS--GD-KEKVREQIVKCVDDNPN-KDDKCTWVFRGFNCFKANHLSLIKQSLKGD 122
LLphg-contig_78  GVFQPDFTISAIS--MS--GVDSPKAPEAVNKCKDET---GSGNCEKAFNIFSCMFREFGPI-----F 116
LLphg-contig_80  GVFQPDFTISAIS--MS--GVDSPKAPEAVNKCKDET---GSGNCEKAFNIFSCMFREFGPI-----F 116
LLphg-contig_81  GVFQPDFTISAIS--MS--GVDSPKAPEAVNKCKDET---GSGNCEKAFNIFSCMFREFGPI-----F 116
LLphg-contig_1059 GVFQEKVTTDKLS--QG--EADPSKVSQMVQCKGESG---AGKCE TAFKIYKCTFEARAEI-----L 115
LLOJ002266      -KFQYEAMKKQAEILLP--PEIKDSNIKAMEVCKSCTDG- IKDHCEAGFALLLCMHKNTLFLVF----P 120
LLOJ002267      -KFQFEATKKQAEILLP--DDMREASIKAMVACKGATDG- IKDPC EAAAYSLLKCLHSNHEDFFF-----P 115
LLOJ002270      -KFQFEATKKQVELLLE--GELRDANIKAFEACKSCTDG- IKDHCEAGFSLLLCLHKNTKLFVF----P 115
LLOJ003968      GLPKKEVIAPYLR--GL--NPKD-VKEIMADCLKLI AEDGSGSCLTTYMRYKCSIQKGLVVL-----S 107
LLOJ010014      NSFHLGKYRERIG--KI--VLP-EVAAQLEETCKGSVG---TENCKISGSVLHCSLTKIVDIFY-----E 115
LLOJ010013      NFSFLEKYNERIG--KI--VQP-DIFGQITEACVGLAG---TENCKVAGSVLHCSITKIVDVFY-----E 115
LLOJ004774      GVFQEKVTTDKLS--QG--EADPSKVSQMVQCKGESG---AGKCE TAFKIYKCAFEARAEI-----L 115

```

*Identidade de todos os resíduos de aminoácidos alinhados na coluna indicada.

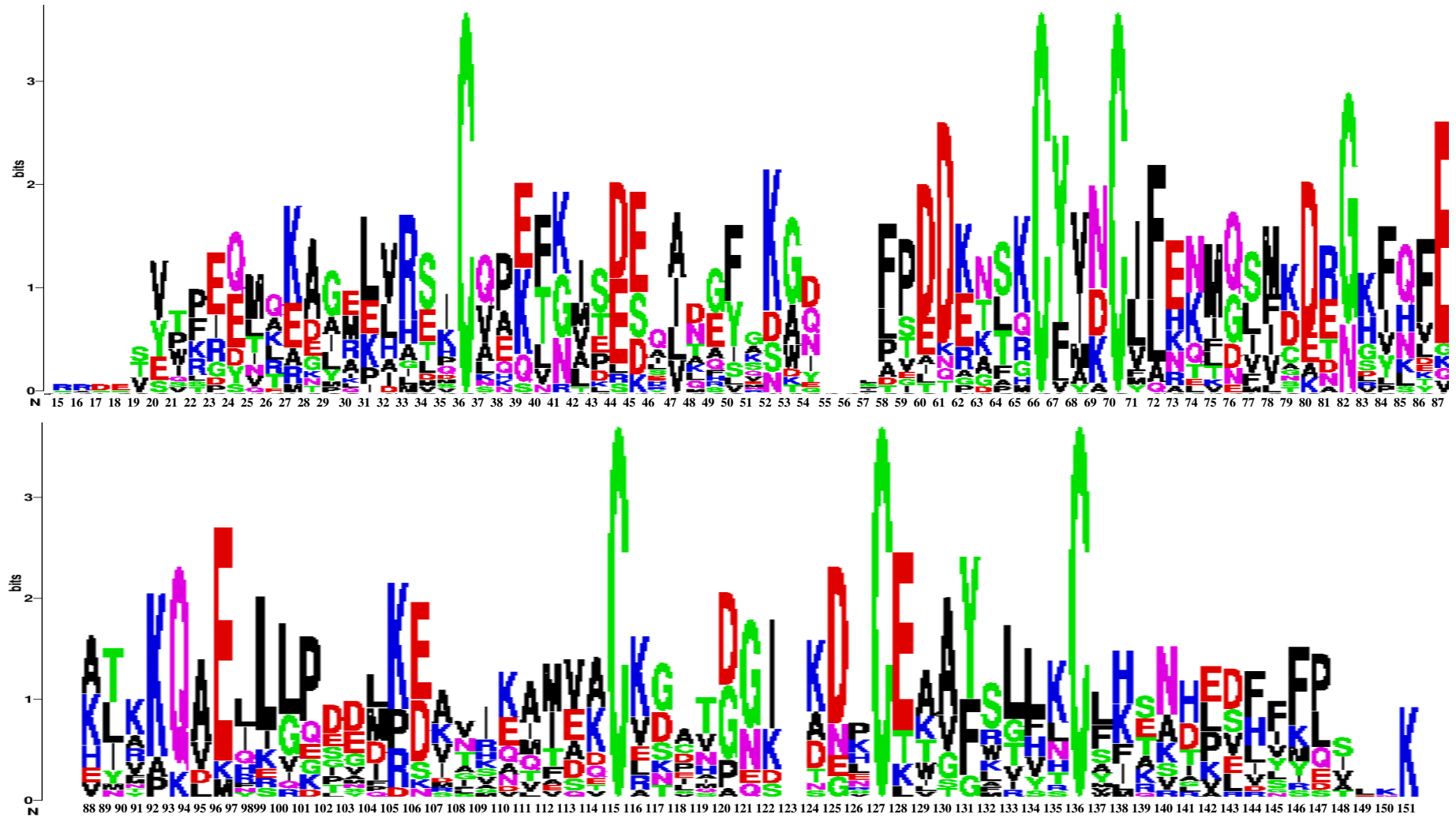
: Similaridade dos aminoácidos alinhados na coluna indicada

. Baixa similaridade dos aminoácidos alinhados na coluna indicada

- Não há similaridade entre os aminoácidos alinhados na coluna indicada

Sequências idênticas não foram representadas para economia de espaço.

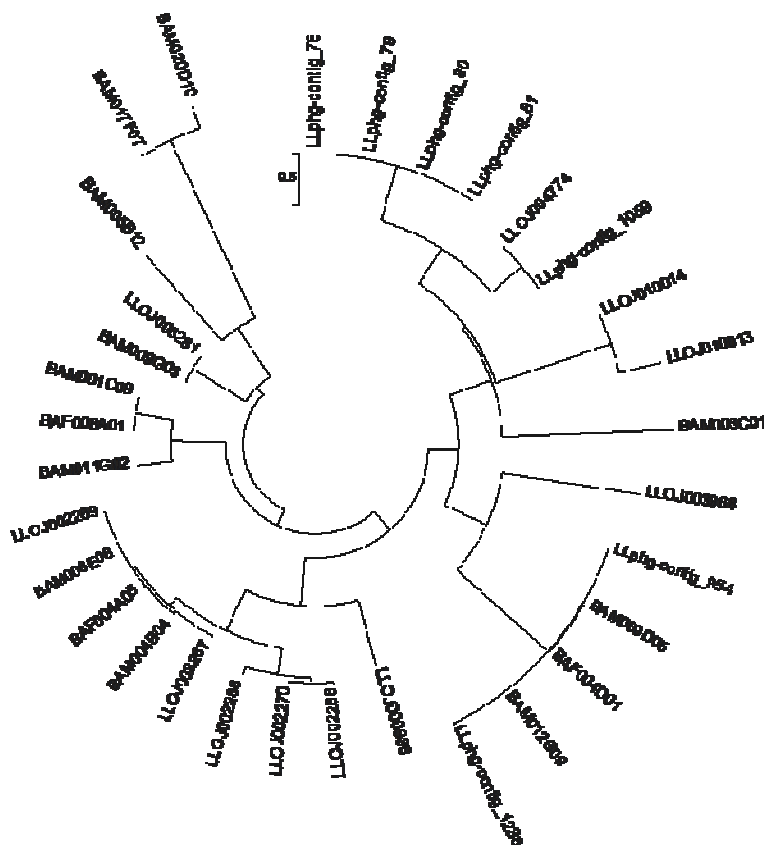
Figura 21: Similaridade sequencial entre as OBPs de *Lutzomyia longipalpis* (Os Gaps iniciais e finais foram removidos).



A baixa similaridade das OBPs é mais facilmente visualizada no gráfico gerado pelo WebLogo (figura 21), cuja frequência de cada aminoácido na posição do alinhamento é proporcional ao tamanho do caractere correspondente. Esta diversidade encontrada nas OBPs geralmente é associada a uma maior especialização no transporte de odor (PELOSI; CALVELLO; BAN, 2005).

As sequências alinhadas das OBPs de *L. longipalpis* foram utilizadas para construção de uma árvore filogenética, gerada a partir do método de máxima verossimilhança (figura 22). Observa-se, então, que algumas OBPs são correspondentes e até mesmo iguais (com diferença de até três aminoácidos). Este resultado foi importante pois tinha-se até então apenas as sequências peptídicas do cDNA das antenas e do transcriptoma das glândulas do flebótomo, portanto, as sequências de nucleotídeos correspondentes a algumas delas foram identificadas.

Figura 22: Análise filogenética molecular pelo método de máxima verossimilhança*.



*A história evolutiva foi inferida através do método de máxima verossimilhança baseado no modelo de matriz JTT (JONES; TAYLOR; THORNTON, 1992). A árvore com a maior probabilidade log (-3430.0831) é mostrada. Árvores iniciais para a busca heurística foram obtidas automaticamente aplicando os algoritmos de Neighbor-Join e BioNJ para uma matriz de distâncias pareadas estimada usando um modelo JTT, e em seguida, selecionando a topologia com valor de verossimilhança superior. A árvore é desenhada à escala, com comprimentos dos ramos medidos a partir do número de substituições por posição. A análise envolveu 32 sequências de aminoácidos. Todas as posições que contêm “gaps” e ausência de dados foram eliminadas. Houve um total de 94 posições no conjunto de dados final. As análises evolutivas foram realizadas no MEGA6 (TAMURA et al., 2013).

4.1.2 Seleção das sequências alvo e dos moldes estruturais

Das sequências de OBP analisadas, foram selecionadas quatro da biblioteca de cDNA das antenas de *Lutzomyia longipalpis* para construção dos modelos, tendo em vista que sua expressão em um órgão olfatório corrobora sua função no transporte de odores detectados pelo inseto (PELOSI et al., 2014a). As sequências de OBP escolhidas alinharam com mais de 45% de identidade com estruturas de OBPs depositadas no *RCSB Protein Data Bank*. E, de acordo com a melhor resolução, duas estruturas cristalográficas PDB-ID 3OGN e 3V2L foram escolhidas para modelar as quatro sequências alvo (tabela 3).

Tabela 3: Sequências alvo com os moldes estruturais selecionados

Sequência	Molde Selecionado	Identidade seq. “alvo-molde”	Cobertura do alinhamento	e-value ^a	ID do Alvo
BAM001C09	3OGN [1,3 Å]	53%	98%	6e-43	OBP1
BAM004B04	3V2L [1,8 Å]	45%	100%	2e-36	OBP2
BAM006E08	3V2L [1,8 Å]	45%	100%	2e-36	OBP3
BAM011G02	3OGN [1,3 Å]	61%	98%	3e-49	OBP4

^a Quanto mais baixo o e-valor ou mais próximo de zero, mais confiável é o alinhamento, $e = (10)^x$.

A estrutura PDB-ID 3OGN (MAO et al., 2010) corresponde à OBP1 do mosquito *Culex quinquefasciatus* (vetor da doença filariose ou elefantíase) em complexo com o feromônio de oviposição (5R,6S)-6-acetoxi-5-hexadecanolida, com 124 aminoácidos em sua sequência primária. A estrutura PDB-ID 3V2L (ZIEMBA et al., 2013) é a forma cristalizada da OBP20 do mosquito *Anopheles gambiae* (vetor da malária) complexada com polietilenoglicol, e composta por 120 aminoácidos. Os dois moldes selecionados possuem padrão de espaçamento das cisteínas igual ao das sequências alvo correspondentes.

Escolhidos os moldes estruturais, foi realizado o alinhamento entre cada sequência alvo e molde correspondente (figura 23). A análise da região compreendida pelo sítio ortostérico do molde, ou seja, os resíduos que estavam próximos em até 4 Å do ligante cristalográfico, possibilitou a comparação do grau de conservação entre os resíduos do sítio em cada alinhamento. Os sítios ortostéricos definidos para as sequências alvo BAM001C09 e BAM011G02 obtiveram, respectivamente, 36% e 57% de identidade com o sítio do molde PDB ID: 3OGN. Já para ambas as sequências BAM004B04 e BAM006E08, a identidade com o sítio do molde foi de 44%.

Figura 23: Alinhamentos entre as seqüências alvo (OBPs 1-4) e moldes (3OGN e 3V2L).

OBP1	BAM001C09	PQTPRRDESYPPRDILAEYRPIREQCLAETGATEKIIQEFSDG	43
	3OGN_PDBID	-VTPRRDAEYPPPELLEALKPLHDICAKKTGVTDEAIIIEFSDG	42
		***** .*** ::* :*::: * :*.*: : * *****	
OBP1	BAM001C09	DHLSPVDEALKCYMNCLFHFKIQVDDTGHVHFEEKLRLKVPDD	86
	3OGN_PDBID	K---IHEDEKLKCYMNCLFHEAKVDDNGDVHLEKLDHSLPNS	82
		. *** *****: :****.*.**:***: :*:..	
OBP1	BAM001C09	LKDVGHNMIACENPVGANLCEKAYWLHTCFKRVDPVHYFLV	128
	3OGN_PDBID	MHDIAMHMGKRCLYPEGENLCEKAFWLHKCWKQADPKHYFLV	124
		::*: . :* :* * * *****:***.*:*.** *****	
OBP2	BAM004B04	VTFEQMQAGELVRSICQPKFKMSDERANGFGKAQFPDDKNS	42
	3V2L_PDBID	MTVEQMMKSGEMIRSVCLGKTKVAEELVNGLRRESKFADVKEL	42
		:*.*** *:***:***:* * *:::* .** :*:.* *	
OBP2	BAM004B04	KCYVNCILENQSMKRGKFQFEATKKQAEILLPDDMKEATIK	84
	3V2L_PDBID	KCYVNCVMEMQTMKKGKLNVDASVKQIDTIMPDELAGPMRA	84
		*****:.* **:***:***:***: * * : :*:***: .	
OBP2	BAM004B04	AMVACKGATDGIKDPCEAAYSLLKCLHSNHEDFFFP	120
	3V2L_PDBID	ALDICRTVADGIKNNCDAAYVLLQCLSKNNPKFIIP	120
		*: * : .:****: *:*** **:*** .*: .:***	
OBP3	BAM006E08	VTFEQMQAGELVRSICQPKFKMSDEQANGFGKAQFPDDKNS	42
	3V2L_PDBID	MTVEQMMKSGEMIRSVCLGKTKVAEELVNGLRRESKFADVKEL	42
		:*.*** *:***:***:* * *:::* .** :*:.* *	
OBP3	BAM006E08	KCYVNCILENQSMKRGKFQFEATKKQAEILLPDDMKEATIK	84
	3V2L_PDBID	KCYVNCVMEMQTMKKGKLNVDASVKQIDTIMPDELAGPMRA	84
		*****:.* **:***:***:***: * * : :*:***: .	
OBP3	BAM006E08	AMVACKGATDGIKDPCEAAYSLLKCLHSNHEDFFFP	120
	3V2L_PDBID	ALDICRTVADGIKNNCDAAYVLLQCLSKNNPKFIIP	120
		*: * : .:****: *:*** **:*** .*: .:***	
OBP4	BAM011G02	E-LRRDETYPPPELLRAIKIAHDVCPQTGVLEEHIFEKSDGE	42
	3OGN_PDBID	VTPRRDAEYPPPELLEALKPLHDICAKKTGVTDEAIIIEFSDGK	43
		*** *****.*:* **:*. :*** :* * *****:	
OBP4	BAM011G02	AIEDAALKCYMNCLFHEFEVDDTGHVHFEEKLYNRIPESMKER	85
	3OGN_PDBID	IHEDEKLKCYMNCLFHEAKVDDNGDVHLEKLDHSLPNSMHD	86
		** *****:****.*.**:***: :*:***:	
OBP4	BAM011G02	AKVTMDKCLDPVGADLCEKAYWLHKCWKTNDPSHYFLA	123
	3OGN_PDBID	AMHMGKRCLYPEGENLCEKAFWLHKCWKQADPKHYFLV	124
		* .:*** * * :*****:***** **.***.	

*Identidade de todos os resíduos de aminoácidos alinhados na coluna indicada.

: Similaridade dos aminoácidos alinhados na coluna indicada

. Baixa similaridade dos aminoácidos alinhados na coluna indicada

- Não há similaridade entre os aminoácidos alinhados na coluna indicada

Marcações em azul são correspondentes ao sítio do molde; Marcações em amarelo correspondem aos resíduos mais facilmente ionizáveis da região C-terminal; Caracteres em vermelho correspondem às maiores regiões não conservadas do alinhamento.

Os graus de identidade e similaridade entre as sequências alvo e moldes foram satisfatórios para a realização da modelagem comparativa. Os alinhamentos apresentaram, no máximo, uma região não conservada compreendida em até seis resíduos adjacentes; no entanto, apenas os alinhamentos com o molde PDB ID 3OGN tiveram regiões curtas (até 3 aminoácidos) não correspondentes (“*gaps*”) em relação às sequências alvo. Interessantemente, todas as sequências alvo possuíam, na região C-terminal, os resíduos ionizáveis característicos das OBPs que captam e liberam seus ligantes de acordo com a variação do pH do meio (DAMBERGER et al., 2013), ainda que em um dos moldes esta característica não estivesse presente (PDB ID: 3V2L) (ZIEMBA et al., 2013) (figura 23). Cada alinhamento entre sequências alvo e molde (figura 19) foi submetido aos servidores SWISS-MODEL, I-TASSER e ao programa MODELLER para predição estrutural dos modelos OBPs 1, 2, 3 e 4.

4.1.3 Construção dos modelos comparativos

A construção de modelos 3D por três métodos diferentes para cada sequência alvo resultou em um total de doze modelos finais. No entanto, como o servidor I-TASSER e o programa MODELLER retornaram mais de um modelo 3D por alinhamento, o modelo mais acurado foi selecionado para as análises. A escolha baseou-se no relatório do I-TASSER, que inclui uma pré-avaliação e indica qual o melhor modelo. Do mesmo modo, o MODELLER através da função normalizada z-DOPE indica qual modelo provavelmente é o mais acurado. Para o SWISS-MODEL o resultado forneceu apenas um modelo por alinhamento.

4.1.4 Validação e escolha do melhor modelo

Ao avaliar as qualidades estérica, energética e global dos doze modelos 3D, o cálculo da pontuação relativa, por métrica, que considerou os modelos com a mesma sequência para compará-los entre si, permitiu que fosse gerada a pontuação final referente à qualidade total de cada modelo. O servidor I-TASSER conseguiu predizer todos os modelos de melhor qualidade, seguido pelo programa MODELLER e pelo servidor SWISS-MODEL (tabela 5). Numa comparação geral dos modelos, percebe-se que a presença de resíduos com altos valores de energia contribuiu para a baixa pontuação dos modelos das OBPs 1, 2 e 3 construídos pelo MODELLER e pelo SWISS-MODEL.

Já em relação ao modelo OBP 4, os resultados foram equilibrados entre os métodos de predição. Dentre os melhores modelos de cada OBP, o que mais contribuiu para redução da qualidade geral foram os resíduos localizados em regiões não permitidas do gráfico de Ramachandran. Com isto, a avaliação geral indicou como melhor o modelo OBP2 predito pelo I-TASSER (tabela 4).

Tabela 4 - Avaliação da qualidade estérica, energética e global dos modelos 3D construídos pelo SWISS-MODEL, I-TASSER e MODELLER

OBP	Ramachandran *				ANOLEA **		Qmean 6	Z-score	Score Total	
	F	P	GP	NP	ET (E/kT)	% aa c/ EA				
1	S	89,3	10,7	0	0	-532	11,81	0,65	-0,94	23,95
	I	88,4	10,7	0	0,9	-622	3,15	0,70	-0,38	25,52
	M	93,8	6,2	0	0	-240	26,77	0,70	-0,33	16,20
2	S	92,5	7,5	0	0	-207	24,37	0,73	-0,04	16,60
	I	88	9,3	2,8	0	-440	4,17	0,79	0,64	26,32
	M	93,5	6,5	0	0	-243	24,17	0,67	-0,65	17,68
3	S	93,5	6,5	0	0	-211	22,69	0,72	-0,18	16,66
	I	87	12	0	0,9	-463	7,50	0,72	-0,16	23,78
	M	95,4	4,6	0	0	-275	15,83	0,75	0,16	21,03
4	S	91,9	8,1	0	0	-557	6,45	0,68	-0,56	22,86
	I	87,4	12,6	0	0	-598	3,23	0,72	-0,20	25,68
	M	96,4	2,7	0	0,9	-457	1,61	0,77	0,40	23,00

*% aa nas regiões Favoráveis = F; Permitidas = P; Gen. Permitidas = GP e Não Permitidas = NP)

** (Energia Total = ET; Energia Alta = AE)

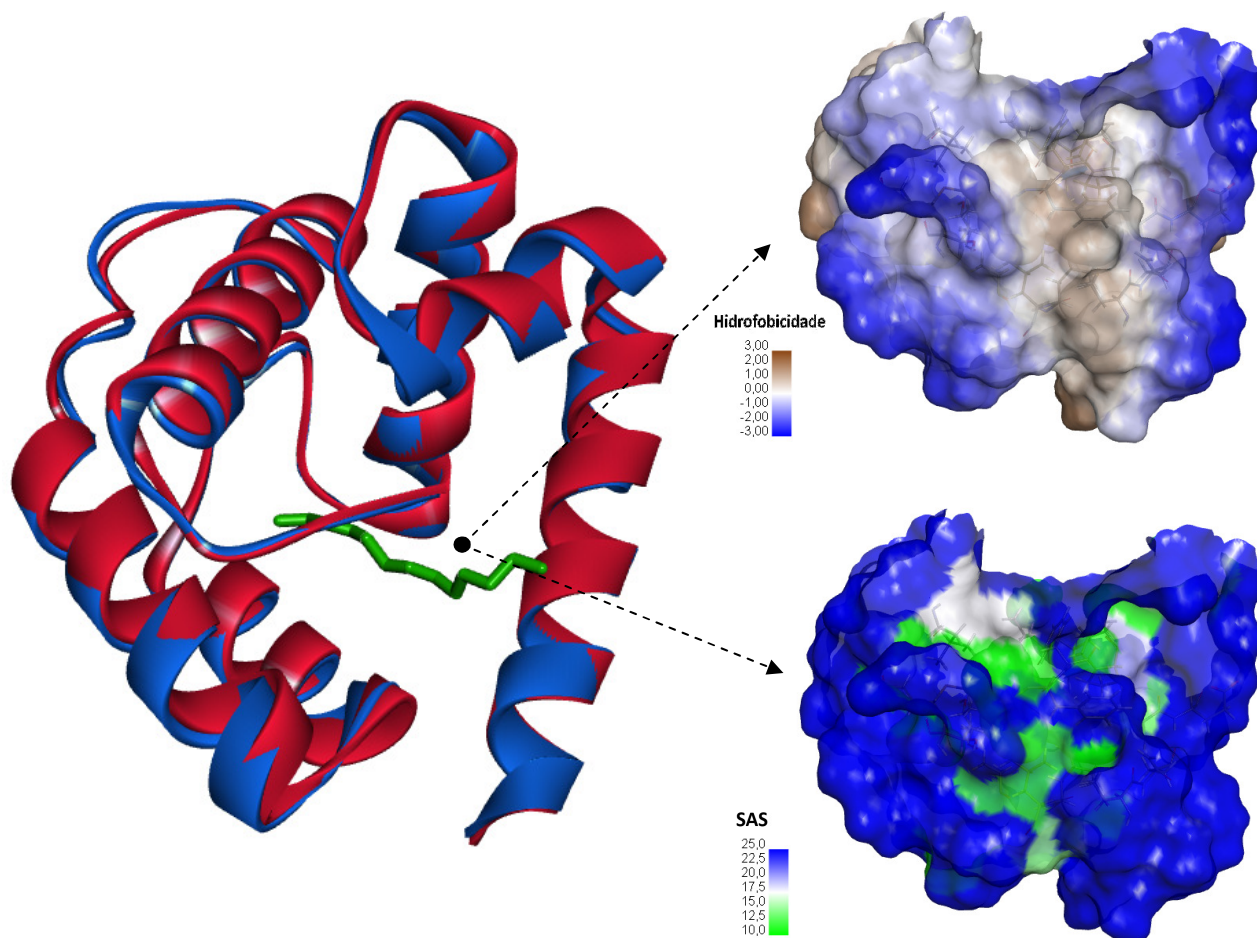
*** Servidores: S= SWISS-MODEL; I=I-TASSER; e programa: M=MODELLER

Os valores destacados em amarelo correspondem aos melhores modelos de cada alinhamento.

A escolha do modelo OBP2 (seq. BAM004B04) para as próximas análises também trouxe a possibilidade de confrontar as análises *in silico* com futuras análises *in vitro*, já que os únicos modelos construídos que possuem sequência de DNA conhecida são OBP2 e OBP3 (correspondentes à seq. LLOJ002269 identificada no genoma de *L. longipalpis*).

O RMSD é uma métrica frequentemente utilizada para avaliar o grau de variação entre estruturas (MAIOROV; CRIPPEN, 1994), e pode facilmente ser calculado entre duas estruturas sobrepostas no DiscoveryStudio4.0 (BIOVIA, 2015). Com este procedimento, obteve-se o valor de RMSD de 0,43 Å entre os átomos de carbono alfa do modelo primitivo (OBP2) e do molde (3V2L) (figura 20). De acordo com a literatura, os valores de RMSD entre os próprios modelos experimentais provenientes de uma mesma proteína alcançam até 0,5 Å, o que corrobora a alta qualidade do modelo (CHOTHIA; LESK, 1986).

Figura 24: Modelo OBP2 (azul) sobreposto ao Molde PDB ID 3V2L (vermelho). Ao lado direito, duas representações da região do Modelo OBP2 em torno do ligante cristalográfico com a superfície de hidrofobicidade e superfície de acesso ao solvente (SAS) geradas no DiscoveryStudio4.0.



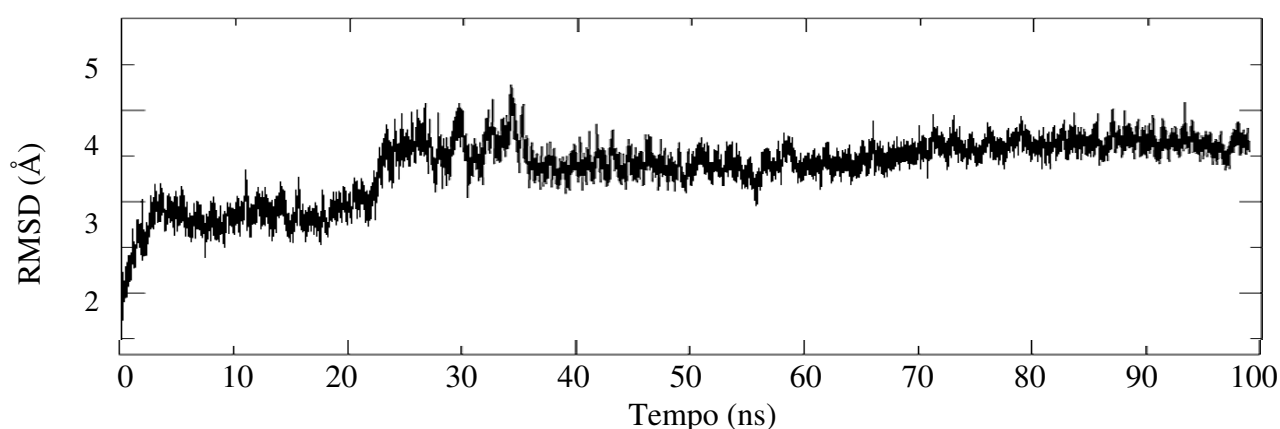
Analisando-se no modelo OBP2 a região correspondente ao sítio ortostérico do molde, percebe-se que o bolsão hidrofóbico foi conservado, assim como a restrita acessibilidade ao solvente (figura 24), características essenciais ao grupo das proteínas encapsulinas.

A predição da protonação dos resíduos de aminoácidos do OBP2 indicou que em um pH de 7,5 todos os resíduos de caráter ácido e histidinas estavam desprotonados, bem como os de caráter básico protonados. Observa-se que o mecanismo de liberação do odor por variação de pH é condizente com os valores de pKa dos resíduos His111, His114, Glu115 e Asp116, pois, considerando o pH ácido da membrana dendrítica (~ 4,5), três deles já estariam protonados, o que poderia ocasionar a formação de uma sétima alfa-hélice capaz de deslocar o ligante de sua cavidade hidrofóbica. Ademais, o pH ideal para estabilidade da proteína, calculado no PROPKA, foi de 7,8 - correspondente a um valor de energia livre de Gibbs de -33,0 kcal/mol à temperatura de 298K.

4.1.5 Refinamento do modelo

Nas etapas de preparação do modelo que precedem o refinamento não houve necessidade da adição de cargas para neutralização do sistema. E após ser submetido à minimização de energia, a trajetória de 100ns da simulação por dinâmica molecular do modelo OBP2 solvatado pode ser visualizada, através da comparação entre os valores de RMSD dos átomos da cadeia principal ao longo do tempo (figura 25).

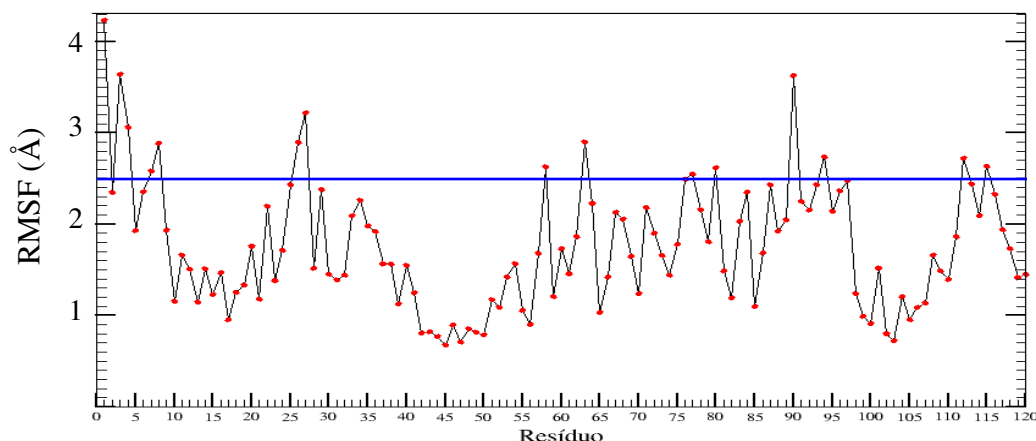
Figura 25: Representação gráfica do RMSD do modelo OBP2 (cadeia principal) durante a simulação por DM.



A análise da trajetória revelou que a OBP atingiu estabilidade após transcorridos 40 ns, em que os valores de RMSD mantiveram-se em torno de 4 Å, com pequenas flutuações. O desvio padrão médio dos valores de RMSD em toda trajetória foi de 3,42 Å. Os valores de RMSD encontrados foram compatíveis e até melhores que os obtidos em simulações de DM descritas na literatura utilizando modelos comparativos pequenos e médios (50 – 100aa) neste mesmo período de tempo (FAN; MARK, 2004), além disso a análise da energia potencial da estrutura indicou que os valores se mantiveram em torno de -3,75kJ/mol, corroborando a estabilidade do modelo.

Considerando que maiores valores de RMSD correspondem às regiões de alta instabilidade, a análise desta variação por resíduo RMSF (*Root Mean Square Fluctuation*) em toda a cadeia principal da proteína ao longo da simulação de DM mostrou que os seguintes resíduos apresentaram valores de RMSF proeminentes em relação aos demais (>2,5 Å): Val1 Phe3, Glu4, Gln7, Lys8, Glu26, Arg27, Arg58, Phe63, Asp77, Glu80, Lys90, Asp94, Ser112 e Glu115 (Figura 26).

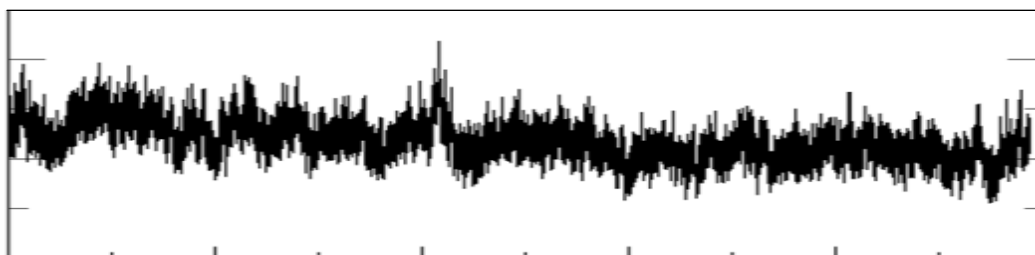
Figura 26: Representação gráfica da variação do *RMSF* de cada aminoácido do modelo OBP2 durante a simulação de DM (acima da linha azul estão os resíduos com mais de 2,5 Å de *RMSF*).



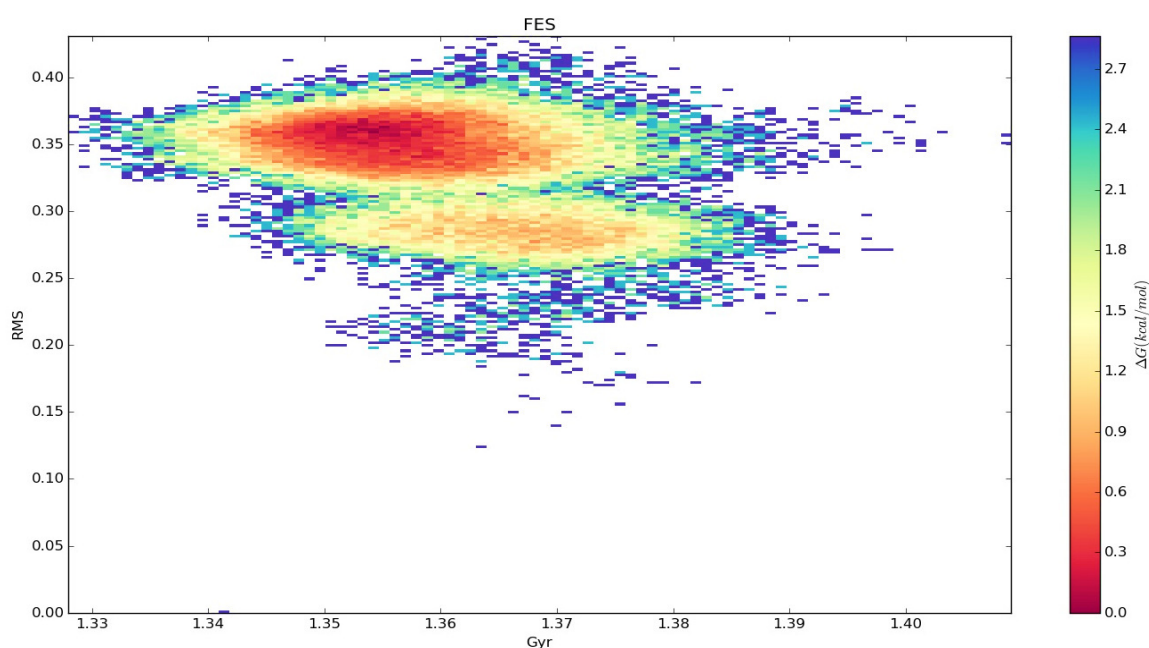
Os aminoácidos supracitados estão localizados em regiões de alça ou no início de alfa-hélices, locais que permitem maior flexibilidade, indicando que estes resíduos possuem um papel crítico para a movimentação interna da proteína, e conseqüentemente podem ser importantes para as transições estruturais entre os estados “ligado” e “não ligado”.

Assim como os valores de RMSD, o valor do raio de giro é uma importante medida da estabilidade, pois permite comparar o grau de empacotamento da proteína ao longo da simulação (LOBANOV; BOGATYREVA; GALZITSKAYA, 2008). A análise desta variável demonstrou que a proteína tornou-se mais compacta no decorrer da simulação, com uma pequena diminuição de aproximadamente 2,3 Å do raio de giro após 100ns de simulação. O valor de raio de giro final (~13,5 Å) foi extremamente parecido com o descrito na literatura em um estudo de DM com a OBP (ASP1), no mesmo pH (HAN et al., 2014) (figura 27).

Figura 27: Variação do raio de giro (resultante dos eixos x, y e z) durante a simulação dos 100ns da DM



A partir da amostragem conformacional produzida pela simulação de DM, selecionou-se a estrutura representativa do momento de maior estabilidade da trajetória, correspondente ao menor valor da Superfície de Energia Livre de Gibbs - ΔG (*Free Energy Surface* - FES) (figura 28). A análise da superfície indicou a presença de dois mínimos de energia, que puderam ser diferenciados de acordo com valores de RMS nos dois *platôs* da trajetória (0 - 22ns) (38ns - 100ns) (figura 25), sendo que no segundo (38ns - 100ns) houve um mínimo energético substancial.



As conformações que mais se repetiram durante a trajetória com base nos valores de ΔG estão elencadas na tabela a seguir, com seus respectivos valores de RMSD e raio de giro. A estrutura mais frequente da simulação (com o menor valor de ΔG) foi escolhida como representativa, e correspondeu ao passo dos 85030 ps da trajetória (~ 85 ns) (tabela 5).

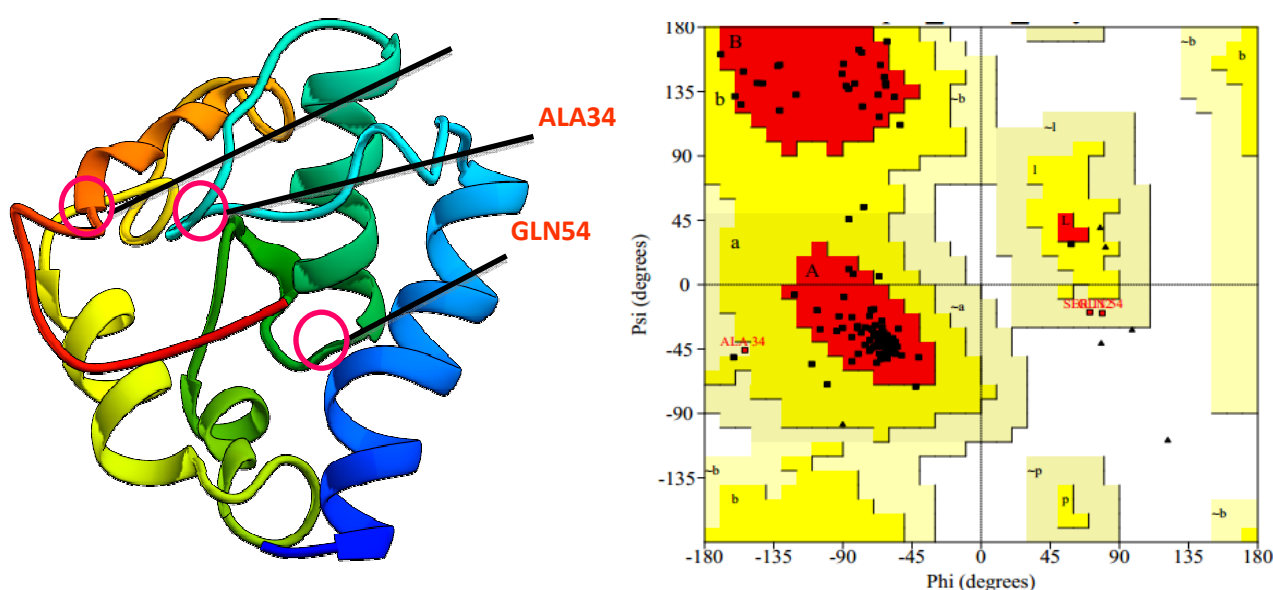
Tabela 5 – As 05 estruturas mais frequentes durante toda a simulação de DM

Dist.	Raio de giro	RMSD	Tempo (ps)
8,36	1,35425	0,3579349	85030,0
8,71	1,35425	0,3592428	58208,0
1,23	1,35427	0,359153	74898,0
1,27	1,35427	0,3575595	98620,0
2,81	1,35423	0,3587246	98144,0

A estrutura representativa da dinâmica foi submetida a uma nova etapa de minimização, desta vez complexada com o ligante cristalográfico, com o objetivo de adaptar as cadeias laterais da região correspondente ao sítio ortostérico no molde ao modelo OBP2, numa conformação que permitisse o acoplamento de outros ligantes. Em seguida, foi realizada uma nova avaliação da qualidade do modelo.

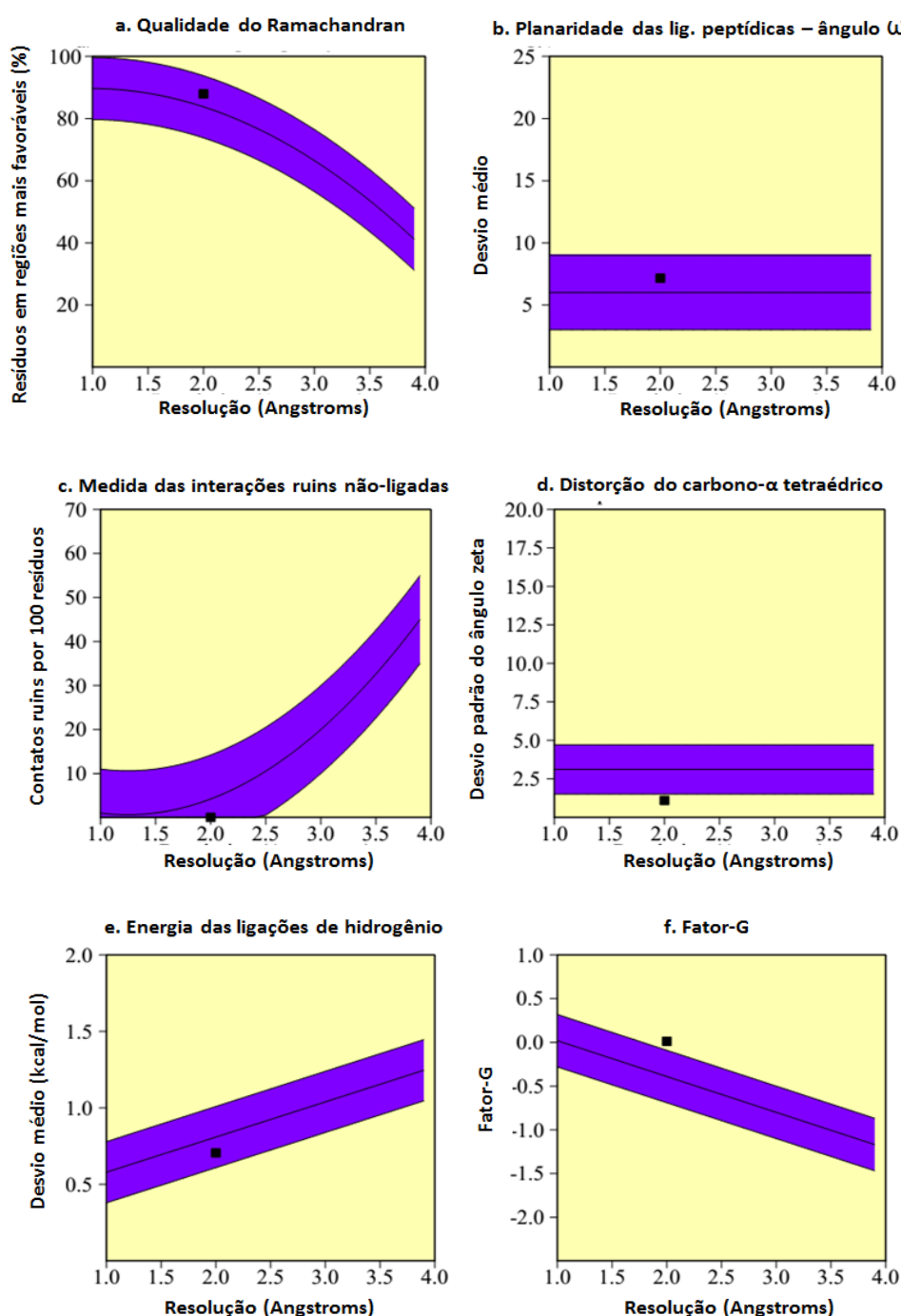
4.1.6 Validação do modelo otimizado

Segundo a validação energética do ANOLEA, o valor de energia total do modelo foi de -523 (unidades de E/kT), e não houve mais nenhum resíduo com valor alto de energia. Já a validação estérica do gráfico de Ramachandran indicou uma ótima qualidade estérica, com 97,3% dos resíduos estavam em regiões favoráveis e permitidas, apenas 2,8% (três resíduos) em regiões generosamente permitidas e nenhum resíduo em regiões não permitidas, isto porque os modelos com mais de 90% dos resíduos em regiões favoráveis, permitidas e generosamente permitidas são considerados de alta qualidade (LASKOWSKI et al., 1993; WHITEHEAD, 2006) (a análise não levou em conta os resíduos de glicina e prolina, pois estes adotam geralmente posições irregulares do gráfico de Ramachandran) (figura 29).



Quando se compara a qualidade da cadeia principal do modelo OBP2 com outras estruturas de boa resolução, em todos os parâmetros analisados o modelo tem melhor resultado que a média geral, com boa qualidade do Ramachandran e Planaridade das ligações peptídicas, nenhum contato ruim, pequena distorção do C- α , baixa energia das ligações de hidrogênio, 96,4% dos ângulos torcionais dentro dos limites esperados, e o fator-G (medida global dos fatores estereoquímicos) positivo (figura 30).

Figura 30: Qualidade da cadeia principal do modelo OBP2: a - Qualidade do Ramachandran; b - Planaridade das lig. peptídicas – Ângulo ω ; c - Medida das interações ruins não-ligadas; d - Distorção do carbono- α tetraédrico; e - Energia das ligações de hidrogênio.



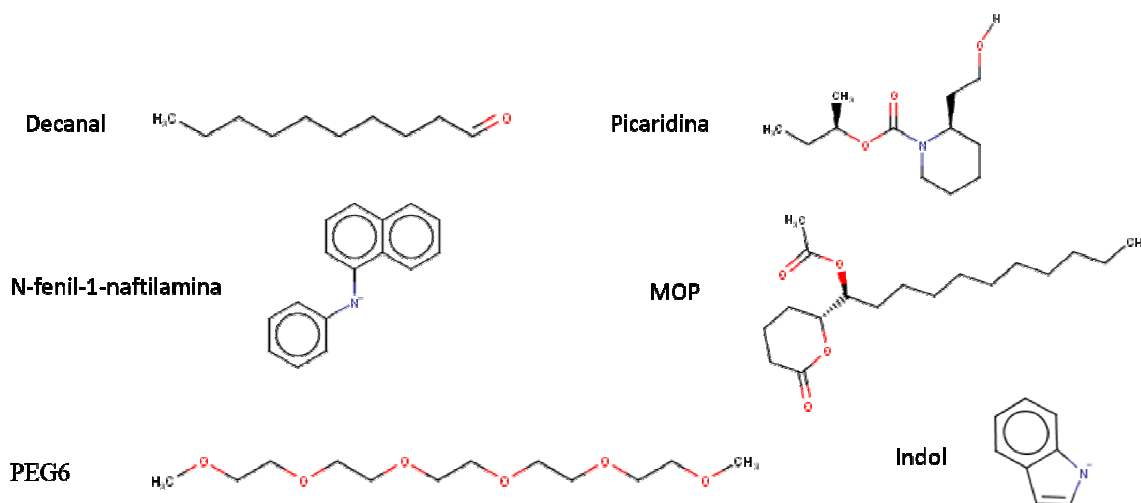
4.2 TRIAGEM VIRTUAL BASEADA NO LIGANTE

4.2.1 Seleção do banco de compostos para as triagens

A quantidade de compostos químicos disponíveis para triagens virtuais nos bancos de dados supera os 120 milhões (STERLING; IRWIN, 2015), no entanto a análise de todos eles na busca por compostos ativos inviabilizaria qualquer estudo, devido ao alto custo computacional e tempo dedicados à tarefa. Uma alternativa às triagens tão abrangentes seria utilizar os bancos de compostos naturais, que, no universo químico, possuem intrinsecamente maior similaridade aos compostos ativos. Logo, com o objetivo de identificar moduladores para as OBPs de *Lutzomyia longipalpis*, selecionou-se apenas o banco de produtos naturais (biogênicos) do ZINC12 (IRWIN et al., 2012), uma compilação com mais de 180 mil moléculas dos compostos naturais disponíveis através de diversos fornecedores.

4.2.2 Triagem do banco por similaridade com ligantes conhecidos de OBPs

Na tentativa de selecionar compostos com as mesmas características físico-químicas dos ligantes típicos de OBPs para os cálculos na triagem por acoplamento molecular, inicialmente, foi realizada uma triagem com base na similaridade química dos ligantes encontrados nas OBPs de mosquitos. Seis ligantes complexados com proteínas cristalográficas de OBPs, depositadas no *Protein Data Bank*, foram selecionados: Decanal (MAO et al., 2010), Indol (DAVRAZOU et al., 2011), N-fenil-1-naftilamina (LEAL; LEAL, 2015; TSITSANOOU et al., 2012), 1-(2-Metoxi-etoxi)-2-{2-[2-(2-metoxi-etoxi)-etoxi]-etano (PEG6) (ZIEMBA et al., 2013), Feromônio de oviposição de mosquito (MOP) (MAO et al., 2010) e Picaridina (YIN et al., 2015) (figura 31).

Figura 31: Moduladores conhecidos de OBPs.

Em seguida, a partir das estruturas dos moduladores conhecidos de OBPs, foi possível priorizar os compostos que compartilhavam as mesmas características físico-químicas, para serem submetidos ao acoplamento molecular. Tendo em vista a grande diversidade química dos compostos naturais, e conseqüentemente a dificuldade em encontrar descritores químicos capazes de compará-los acuradamente, o servidor ChemGPS-NP (LARSSON et al., 2007) foi utilizado para calcular as coordenadas multidimensionais do espaço químico biologicamente relevante para os potenciais ligantes de OBPs.

As coordenadas das dimensões mais relevantes foram aproveitadas, e corresponderam aos descritores químicos condensados nos componentes principais: PC1 - tamanho, formato e polarizabilidade, PC2 - propriedades aromáticas e conjugadas e PC3 - lipofilicidade, polaridade e capacidade de formar ligações de hidrogênio (tabela 6).

Tabela 6 – Coordenadas no espaço químico dos moduladores de OBPs

MOL-ID	PC1	PC2	PC3
Decanal	-4,16	-1,66	0,68
Indol	-4,18	3,25	-0,41
N-fenil-1-naftilamina	-2,29	4,47	0,91
PEG6	-2,73	-2,54	-0,21
MOP (5R,6S)-6-acetoxi-5-hexadecanolida	-1,94	-1,85	1,41
Picaridina (Bayrepel)	-3,12	-1,65	-0,21

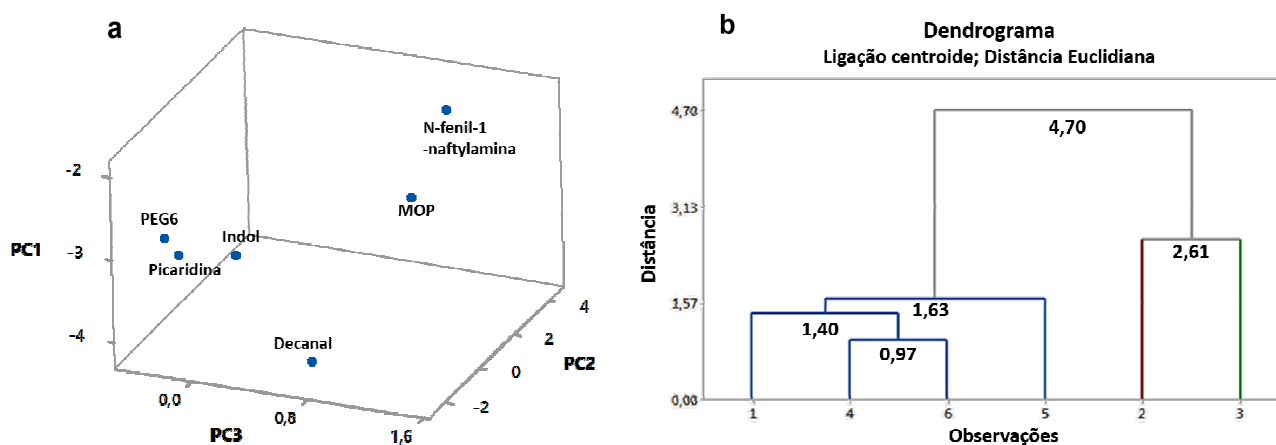
As coordenadas das PCs 1, 2 e 3 (tabela 6) foram utilizadas para calcular as distâncias euclidianas (baseando-se no teorema de Pitágoras) entre cada um dos compostos em relação a todos (tabela 7), e com isso comparar sua proximidade no espaço químico tridimensional.

Tabela 7 – Matriz das distâncias euclidianas entre as moléculas extraídas de OBPs

	Decanal	Indol	N-fenil-1	PG6	MOP	Picaridina
Decanal	0					
Indol	5,03	0				
N-fenil-1	6,41	1,92	0			
PEG6	1,90	5,95	7,11	0		
MOP	2,35	5,40	6,35	1,93	0	
Picaridina	1,37	5,00	6,28	0,96	2,02	0

De acordo com a matriz das distâncias euclidianas (tabela 7), os compostos Decanal, PEG, MOP e Picaridina estão mais próximos entre si que do que os compostos Indol e N-fenil-1-naftilamina, o que pode ser evidenciado através da visualização das estruturas dos compostos (figura 27), já que Indol e N-fenil-1-naftilamina possuem mais anéis aromáticos. No programa MINITAB17 foi gerado um gráfico 3D dos compostos (X = PC1, Y = PC2 e Z= PC3) (figura 32 a), e, a partir de uma análise estatística multivariada (observações por agrupamentos), definiu-se o ponto central no espaço químico (centroide) a partir do qual todos se distanciavam (figura 32 b).

Figura 32: **a** - Distribuição dos compostos distribuídos no gráfico 3D através da análise de agrupamento hierárquico (HCA); **b** – Dendrograma mostrando as distâncias euclidianas entre os agrupamentos dos compostos, definidos pelo método de ligação do centroide.



Isto auxiliou na definição da margem que deveria ser considerada, em unidades de Distância Euclidiana (DE), para busca por compostos com características físico-químicas semelhantes aos

ligantes de OBPs. Com esta avaliação, tomou-se como base o valor de 4,7 unidades de distância Euclidiana (DE), segundo a qual seria possível identificar os ligantes do banco de produtos naturais que estivessem dentro deste mesmo centroide. Desta forma, não seria necessário utilizar as coordenadas de todos os moduladores como filtro, logo, apenas os compostos Indol, PEG6, Decanal e MOP foram utilizados para seleção dos compostos naturais (biogênicos) do ZINC12.

Para isso, também foram geradas no ChemGPS-NP as coordenadas químicas espaciais de todos os compostos do banco, das quais apenas as que correspondiam às três primeiras dimensões foram consideradas. Depois de remover as duplicatas, 152,063 mil compostos biogênicos foram filtrados através da similaridade no espaço químico, e desses, cerca de 70 mil estavam localizados em no máximo 4,7 unidades de distância euclidiana (DE), e foram selecionados para as etapas de acoplamento molecular.

4.3 TRIAGEM VIRTUAL COM BASE NO RECEPTOR

4.3.1 Definição do sítio ortostérico da OBP

Após a construção do banco de ligantes para o acoplamento molecular (na triagem virtual baseada no ligante), a região da estrutura da OBP2 que seria utilizada para o acoplamento foi validada empiricamente através da predição multiconsensual do MetaPocket (HUANG, 2009). A predição pelo MetaPocket informou quatro regiões com características de sítio ortostérico de acordo com oito métodos de predição (tabela 8).

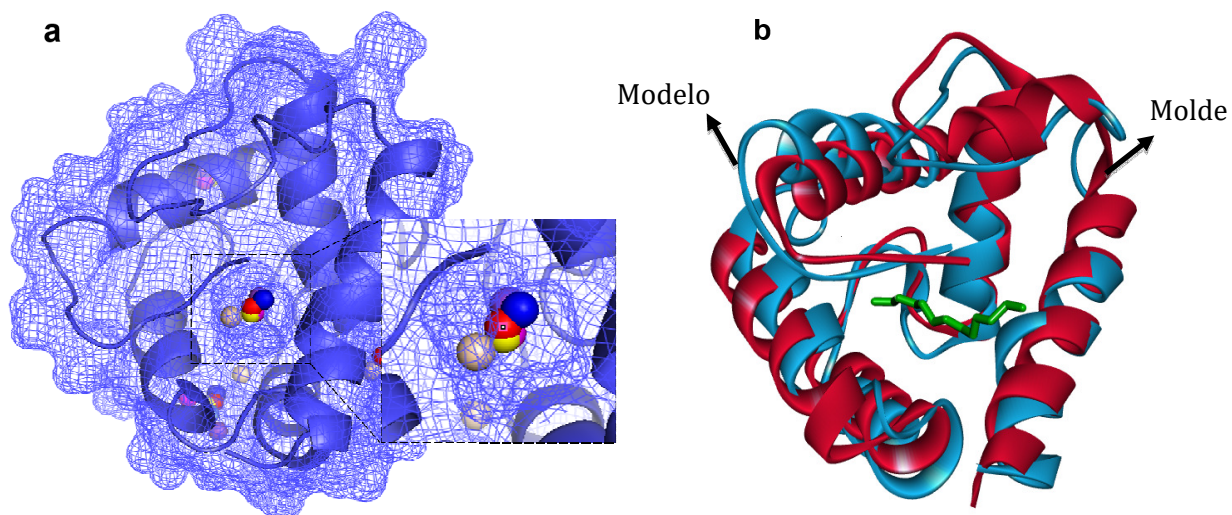
Tabela 8 – Resultado da predição multiconsensual pelo MetaPocket

Região	Pontuação	Diferentes métodos de predição por ordem de colocação					
Sítio 1	9.76	'LCS-1'	'FPK-1'	'SFN-2'	'GHE-2'	'CON-1'	
Sítio 2	7.70	'GHE-1'	'GHE-1'	'FPK-2'	'LCS-3'	'PAS-1'	'CON-2'
Sítio 3	3.40	'PAS-2'	'SFN-3'				
Sítio 4	2.43	'FPK-3'	'PAS-3'	'LCS-2'	'GHE-3'		

PASS11(PAS); LigsiteCS(LCS); GHECOM(GHE); Fpocket(FPK); SURFNET(SFN) e ConCavity(CON)

Apesar de a região mais bem colocada não ter sido a mais reconhecida por diferentes métodos, ela obteve as melhores classificações, e coincidiu com a mesma região do modelo OBP2 sobreposta ao sítio ortostérico da estrutura molde (PDB ID: 3V2L) (Figura 33).

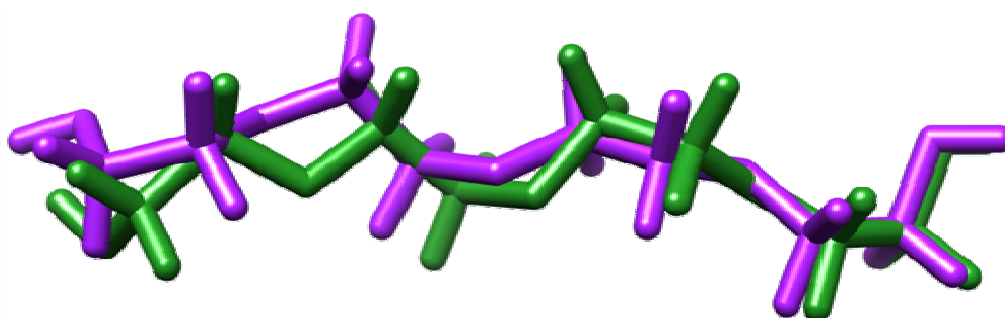
Figura 33 – a - Representação esquemática do sítio1 predito pelo MetaPocket; b - Região correspondente ao sítio ortostérico do molde PDB ID: 3V2L.



Portanto, mesmo considerando que o Sítio2 possa ser estudado em futuras triagens, deu-se preferência à utilização do Sítio1. Os resíduos da OBP2, que no acoplamento molecular seriam avaliados quanto à da interação com ligantes, foram escolhidos pela proximidade em relação ao centro do Sítio1 em até 5 Å de distância.

4.3.2 Triagem por acoplamento molecular

A capacidade de reprodução dos modos de ligação experimentais pelo acoplamento molecular foi considerada como bem sucedida, visto que a variação do RMSD entre todos os átomos do ligante cristalográfico e o ligante acoplado após a minimização foi de 0,81 Å (NIENHAUS, 2005) (figura 34).

Figura 34 – RMSD entre a orientação do ligante cristalográfico (verde) e o ligante acoplado (lilás).

O resultado do acoplamento molecular foi pontuado pela função de pontuação GRIDSCORE, e após a classificação das conformações dos ligantes de acordo com seus valores de energia selecionou-se os 100 compostos que apresentaram os menores valores de energia, para serem pontuados secundariamente pela função de pontuação AMBERSCORE. Dos compostos repontuados 13% apresentaram valores positivos de energia, isto indica que a avaliação da flexibilidade do complexo pode auxiliar na identificação dos resultados falso-positivos, permitindo uma maior confiabilidade da predição. Os menores valores de energia associados aos complexos de cada ligante acoplado estão descritos abaixo, dentre eles percebe-se que um se destaca pela enorme diferença energética quando comparada aos demais (tabela 9).

Tabela 9: Ligantes de menor energia pontuados pelo AMBERSCORE

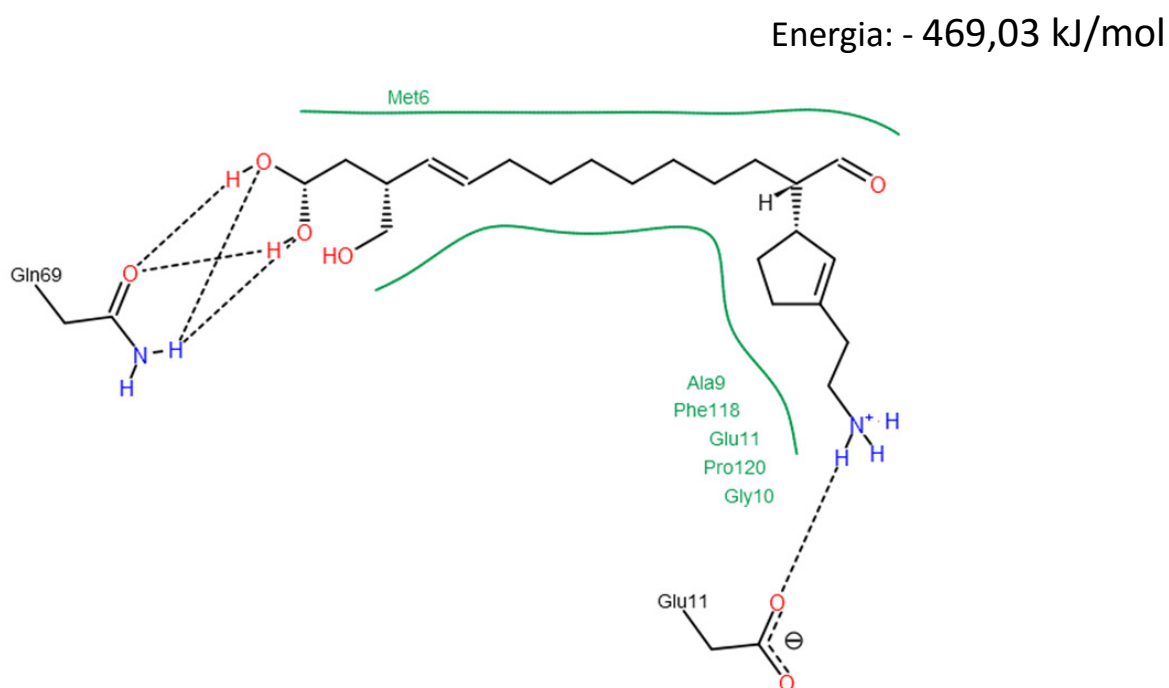
Código ZINC do ligante acoplado	Valor de energia (kJ/mol)
ZINC85489174	-469,03
ZINC85489219	-67,82
ZINC11867897	-62,59
ZINC04041684	-62,15
ZINC67913248	-61,23
ZINC85489217	-49,26
ZINC85510419	-46,99
ZINC85488305	-46,14
ZINC02110828	-45,68
ZINC72332904	-45,51
ZINC15215583	-44,62
ZINC15117857	-44,04
ZINC08964505	-43,97
ZINC03999406	-43,74
ZINC85490903	-43,19

Os diagramas 2D apresentados abaixo foram gerados a partir dos melhores resultados da pontuação pelo AMBERSCORE, e isso facilitou a análise detalhada do potencial de interação entre o ligante acoplado e o sítio da proteína.

O ligante Lig1 (figura 35) apresentou interações hidrofóbicas com os resíduos Met6, Ala9, Gly10, Glu11, Phe118 e Pro120, muitos dos quais também eram importantes nas interações vistas entre o ligante cristalográfico e a OBP molde. As cadeias apolares dos aminoácidos Ala, Met e Phe contribuíram para a formação desse tipo de interação com a cadeia alifática apolar do ligante. As hidroxilas terminais do ligante coordenam-se com o grupo amino da Gln69 formando duas interações de hidrogênio aceptoras. Neste mesmo local, também há formação de duas interações de hidrogênio doadoras com a carbonila da Gln69, e o grupo amino protonado do ligante interage com o Glu11 através de uma ligação de hidrogênio doadora.

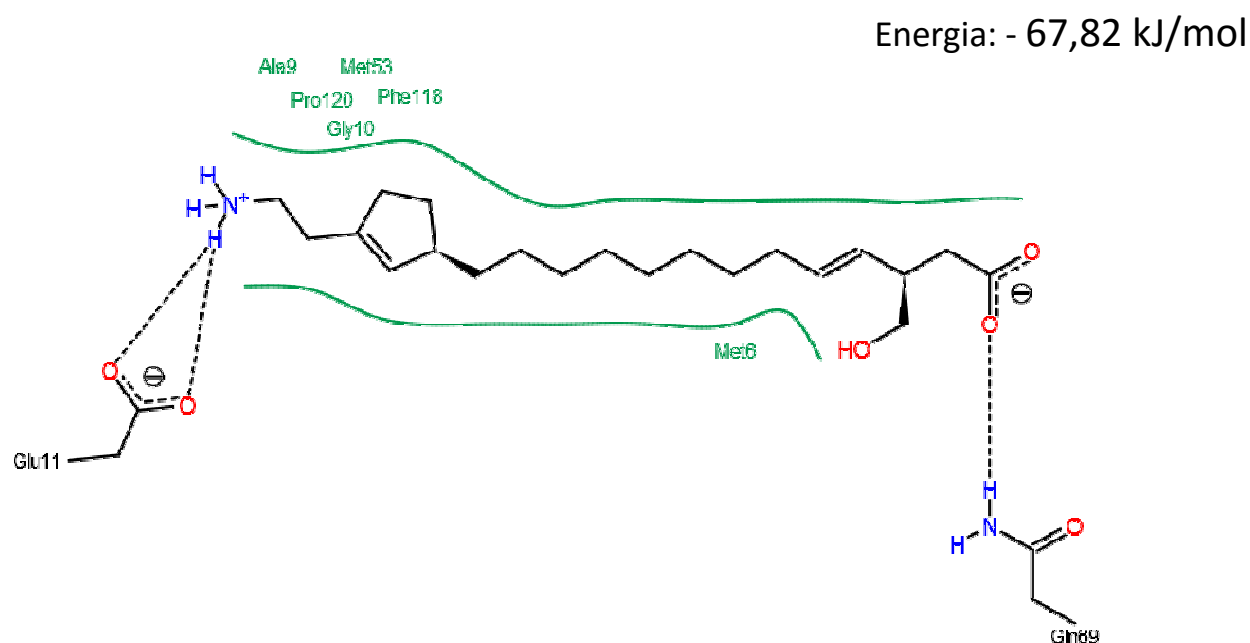
Quando são comparadas as interações aqui descritas com aquelas presentes no complexo do molde e ligante cristalizados (PDB ID: 3V2L), o resíduo Gln69 está voltado para fora da cavidade onde está inserido o ligante no molde. Já na OBP2 a dupla capacidade de participar de ligações de hidrogênio permite apontar o resíduo Gln69 como essencial para estabilização de ligantes com terminações doadoras ou aceptoras de hidrogênio.

Figura 35: Mapa de interação 2D do Lig1 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



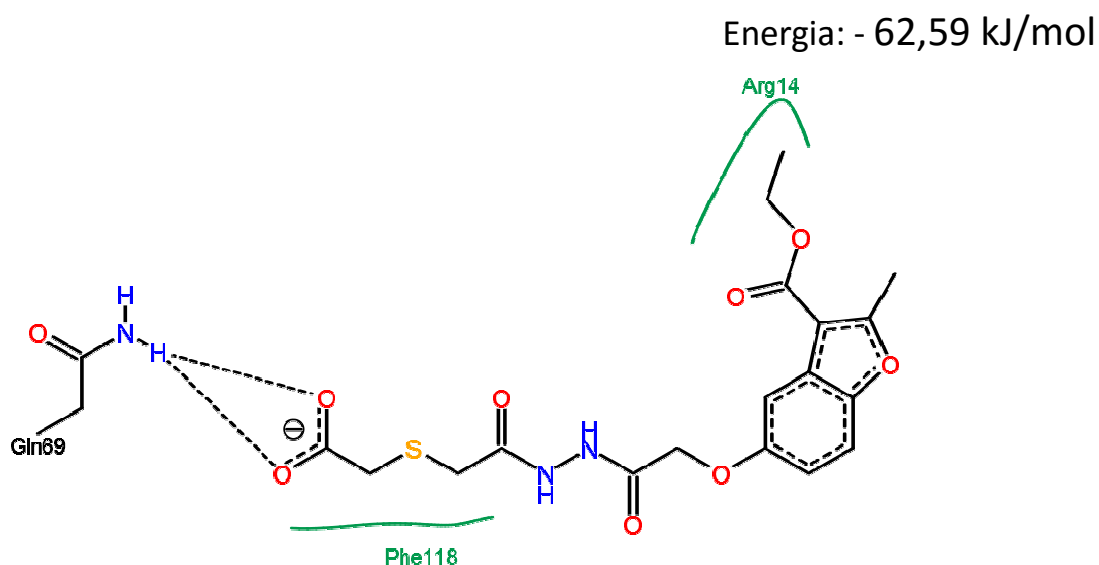
O ligante Lig2 (figura 36) apresentou interações hidrofóbicas com os mesmos resíduos, que interagem com o Lig1: Ala9, Gly10, Met6, Phe118, Pro120 e adicionalmente o resíduo Met53, que é conservado e também interage com o ligante no molde. A carbonila do ligante se coordenou com a Gln69 por meio de interação de hidrogênio aceitadora. Além disso, o grupo amino protonado do ligante interagiu com o resíduo Glu11 formando uma ligação de hidrogênio doadora, que não foi encontrada nas interações da estrutura molde cristalizada, provavelmente, porque os estudos que avaliam as interações nos complexos formados por OBPs quase sempre utilizam o mesmo grupo de ligantes, ou compostos pouco diversos (LEAL et al., 2008).

Figura 36: Mapa de interação 2D do Lig2 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



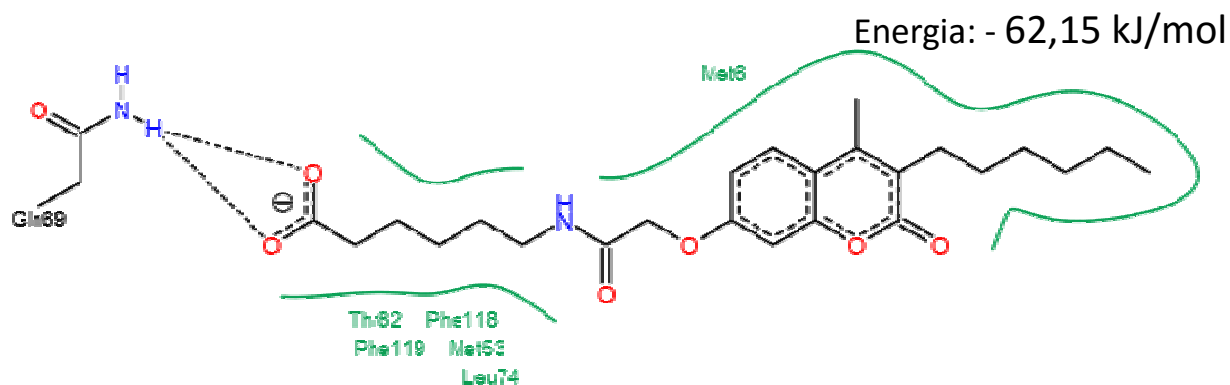
As interações hidrofóbicas do ligante Lig3 (figura 37) com o sítio da proteína ocorreram apenas entre a curta região apolar do ligante (intercalada por um átomo de enxofre) e o resíduo Phe118, e entre o grupo etil terminal do ligante e o resíduo Arg14, cuja cadeia lateral está desprotonada nesse pH. A redução nas interações hidrofóbicas, consequência dos grupamentos cetonas e aminos na cadeia alifática do ligante, foi balanceada pelas duas interações de hidrogênio entre a porção carboxi-terminal aceitadora do ligante e o grupo amino doador da Gln69.

Figura 37: Mapa de interação 2D do Lig3 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



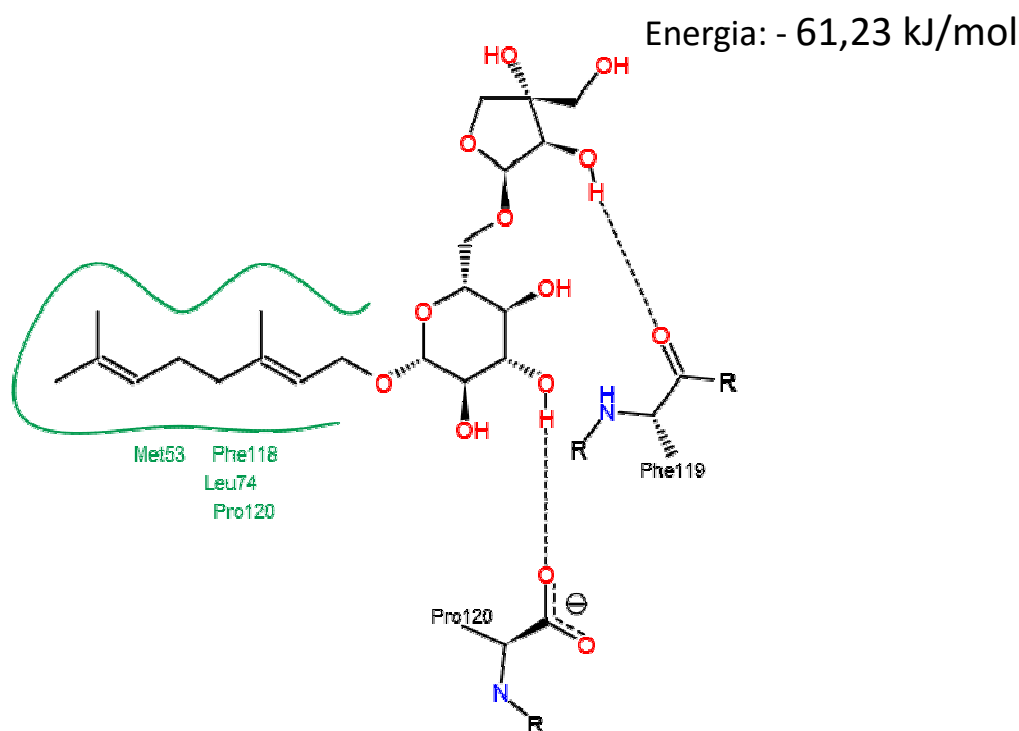
Duas regiões da molécula Lig4 (figura 38) apresentaram interações hidrofóbicas, uma compreendida entre uma curta cadeia alifática e a região apolar de dois anéis aromáticos com o resíduo Met6, e a outra, uma cadeia alifática que interagiu com os resíduos apolares Phe118, Phe119, Met53, Leu74 e em menor grau com a Thr82. E assim como nos outros ligantes, a carboxila do ligante também formou duas interaçõesceptoras de hidrogênio com a amina doadora da Gln69.

Figura 38: Mapa de interação 2D do Lig4 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



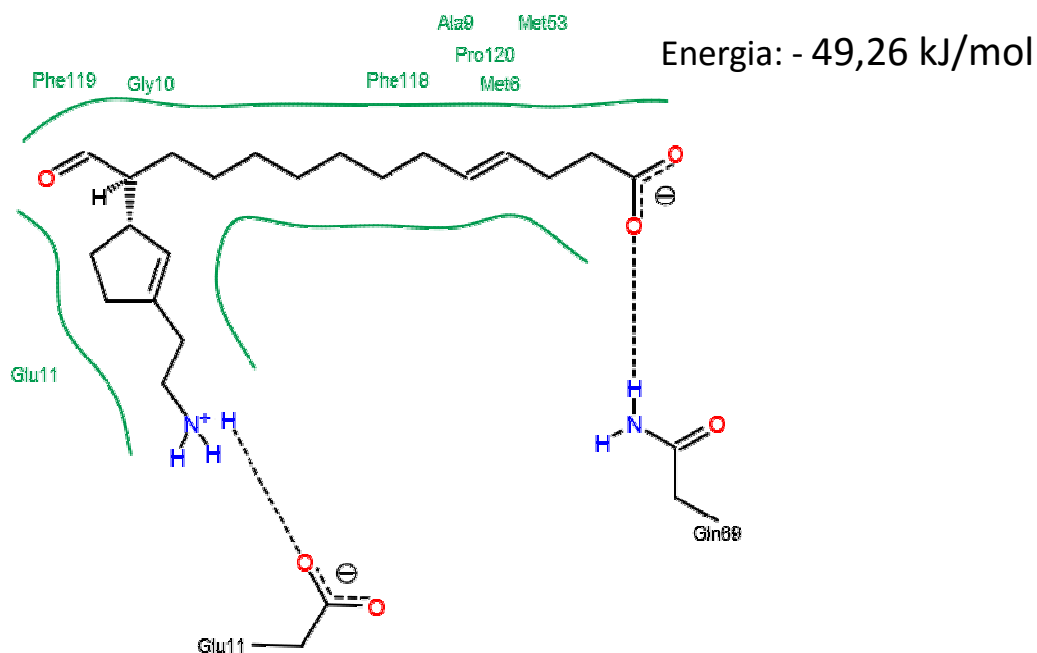
A cadeia alifática do Lig5 (figura 39) forma interações hidrofóbicas com os resíduos Met53, Phe118, Leu74 e Pro120. Desta vez, o ligante acoplado não interagiu com a Gln69, mas uma hidroxila ligada ao anel oxolano formou uma ligação de hidrogênio doadora com o grupo carbonila da Phe119.

Figura 39: Mapa de interação 2D do Lig5 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



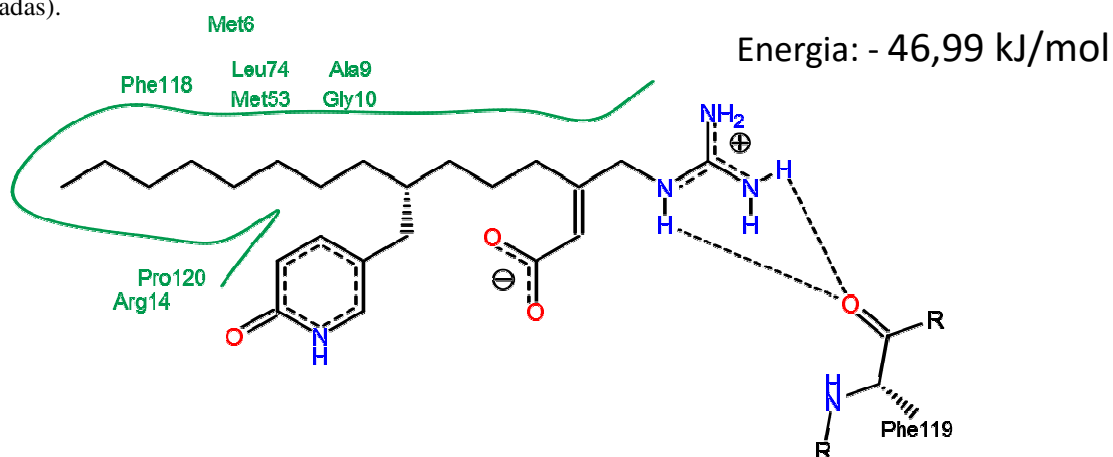
O Lig6 (figura 40) apresentou interações hidrofóbicas com os resíduos Glu11, Met6, Gly10, Phe118, Phe119, Ala9, Met53 e Pro120. A carbonila do ligante formou uma ligação de hidrogênio aceitadora com a amina da Gln69. Na outra terminação, um grupamento amino protonado do ligante formou uma ligação de hidrogênio doadora com a carboxila do Glu11.

Figura 40: Mapa de interação 2D do Lig6 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



O ligante Lig7 (figura 41) apresentou interações hidrofóbicas entre sua cadeia alifática apolar e os resíduos Ala9, Gly10, Met6, Phe118, Met53, Leu74, Pro120 e Arg14. O grupamento amino alifático do ligante formou duas ligações de hidrogênio, cada uma entre uma amina doadora e a carbonila aceptora da Phe119.

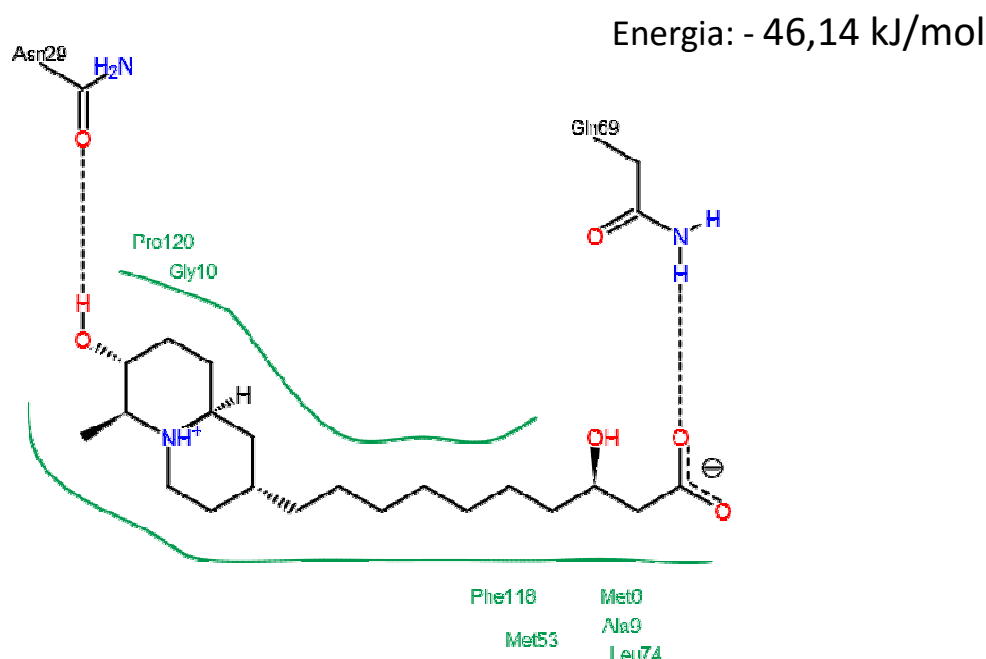
Figura 41: Mapa de interação 2D do Lig7 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



O ligante Lig8 (figura 42) se coordenou com a proteína por meio de interações hidrofóbicas com os resíduos Ala9, Gly10, Met6, Phe118, Pro120, Met53 e Leu74. A típica interação entre uma carbonila com a Gln69 através ligação de hidrogênio aceptora também foi encontrada.

Foi possível observar a formação de uma ligação de hidrogênio doadora entre a hidroxila ligada ao anel decahidroquinolizina do ligante e a carbonila da Asn29.

Figura 42: Mapa de interação 2D do Lig8 com a OBP2 (Os resíduos delineados em verde correspondem às interações hidrofóbicas; Linhas tracejadas correspondem às ligações de hidrogênio, e às interações iônicas quando na presença de regiões carregadas).



A observação das estruturas dos melhores ligantes pontuados evidencia que além dos contatos hidrofóbicos necessários para o transporte de moléculas não hidrossolúveis, a presença de grupos doadores e/ou aceptores de hidrogênio como hidroxilas, aminas e ácidos carboxílicos e de regiões carregadas nas terminações dos ligantes é um importante fator para o planejamento de moléculas que sejam capazes de formar complexos mais estáveis com o sítio da OBP analisada, em função, principalmente, da presença e orientação da Gln69, Glu11, Phe119, e Ans30. E, de fato, essas características estão presentes em uma série de compostos com poder atrativo ao flebótomo (MAGALHÃES-JUNIOR et al., 2014).

5. CONCLUSÃO

O entendimento dos mecanismos moleculares responsáveis pela modulação da olfação nos insetos, a longo prazo, poderá revolucionar a maneira como lidamos com polinizadores, pragas, e vetores de doenças em geral. Nos últimos anos, tem sido devastador o efeito da negligência e do atraso tecnológico a respeito do controle de doenças vetoriais. Em relação à leishmaniose, o controle do vetor é extremamente difícil de ser realizado, o que justifica, portanto, o interesse por estratégias focadas na modulação olfativa do seu comportamento com o objetivo de auxiliar na prevenção da doença.

A utilização das Proteínas Ligadoras de Odor (OBPs) como alvos moleculares importantes para o estudo da modulação olfativa dos vetores tem ganhado cada vez mais espaço na ecologia química, para *Lutzomyia longipalpis* algumas dessas proteínas já foram identificadas, ainda que suas sequências nucleotídicas não estejam depositadas em bancos de dados curados, como o NCBI (*National Center for Biotechnology Information*), além disso, há possibilidade de que mais proteínas da família sejam reconhecidas a partir do genoma do vetor, como foi realizado neste trabalho, com a identificação de 11 sequências de gDNA de OBPs, das quais uma já havia sido identificada no transcriptoma das glândulas de feromônio (LLOJ004774), e outras duas também possuíam correspondentes na Biblioteca de cDNA das antenas do flebótomo (LLOJ002269 e LLOJ008281). Estas sequências poderão ser analisadas quanto à expressão gênica de acordo com as diferentes fases de desenvolvimento dos machos e fêmeas, bem como em função do tipo de alimentação e luminosidade a que são expostos, dentre outros, e isto fornecerá um perfil importante na definição da importância de cada OBP para a ecologia química do inseto.

A estrutura tridimensional das OBPs do *Lutzomyia longipalpis*, no entanto, nunca havia sido investigada, e nem sua interação com ligantes. Portanto, este foi o primeiro estudo que possibilitou conhecer aspectos da biologia estrutural dessas moléculas de maneira detalhada através das técnicas de modelagem comparativa. A validação da qualidade dos modelos construídos permitiu inferir que o servidor I-TASSER possui grande utilidade na predição de modelos 3D de OBPs de alta qualidade, e também possibilitou priorizar o melhor modelo (OBP2) para o refinamento, incluindo o uso de simulações de dinâmica molecular. Com os resultados da DM foi possível escolher uma estrutura final estável, correspondente ao mínimo global da superfície de energia livre de Gibbs, para o qual se espera uma conformação corretamente enovelada e biologicamente ativa para os ensaios de acoplamento molecular.

A utilização da abordagem hierárquica da triagem virtual trouxe a possibilidade de filtrar sistematicamente e com alta eficiência mais de 150,000 compostos naturais para seleção daqueles que fossem semelhantes no espaço químico aos moduladores já conhecidos de OBPs, para isso o ChemGPS-NP simplificou a complexidade da descrição estrutural dos compostos naturais e facilitou a seleção dos ligantes relevantes para avaliação quanto a interação com o sítio ortostérico da proteína. A qualidade do modelo final otimizado foi considerada satisfatória de acordo com os valores energéticos e estéricos, nesse contexto, este estudo também se mostra pioneiro já que dentre os modelos comparativos de OBPs descritos na literatura há pouquíssimos relatos do emprego do refinamento para estas estruturas, mesmo daquelas que são utilizadas em ensaios de acoplamento.

Os ligantes acoplados foram pontuados também de maneira hierárquica, utilizando a função GRIDSCORE como primária e a AMBERSCORE como secundária. A análise das interações entre a OBP2 e cada ligante acoplado mostrou que os melhores resultados apresentaram contatos hidrofóbicos com os resíduos Met6, Ala9, Gly10, Glu11, Arg14, Met53, Leu74, Phe118, Phe119, Pro120, e formaram ligações hidrogênio doadoras com a carboxila da Glu12, Phe119 e Asn29, e doadoras e receptoras com o resíduo Gln69, indicando que estes resíduos são importantes para o transporte dos odores pela OBP avaliada.

Este estudo, portanto, trouxe uma nova perspectiva à busca de metodologias para a modulação do comportamento do vetor da LV, conhecimento este que poderia ser empregado para controle do inseto e/ou proteção dos indivíduos a ele expostos. De maneira que os compostos identificados como potenciais moduladores podem ser importantes protótipos no desenvolvimento de atrativos e repelentes ao *Lutzomyia longipalpis*. Para isto, pretende-se agora realizar simulações de dinâmica molecular com estes complexos “odor-OBP”, e, a partir dos resultados obtidos selecionar os ligantes que melhor interajam com a OBP para avaliação *in vivo* utilizando o detector eletroantenógrafo.

REFERÊNCIAS

- AI, M. et al. Acid sensing by the *Drosophila* olfactory system. **Nature**, v. 468, n. 7324, p. 691–5, 2 dez. 2010.
- ALLEN, W. J. et al. DOCK 6: Impact of new features and current docking performance. **Journal of computational chemistry**, v. 36, n. 15, p. 1132–56, 5 jun. 2015.
- ALOY, P. et al. The relationship between sequence and interaction divergence in proteins. **Journal of molecular biology**, v. 332, n. 5, p. 989–98, 3 out. 2003.
- ALTSCHUL, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, n. 17, p. 3389–3402, 1 set. 1997.
- ALVAR, J. et al. Leishmaniasis worldwide and global estimates of its incidence. **PloS one**, v. 7, n. 5, p. e35671, jan. 2012.
- ANDRADE, J. B. DE et al. Perspectives for Diagnosis and Control of Leishmaniasis Based on Volatile Organic Compounds. In: **Leishmaniasis - Trends in Epidemiology, Diagnosis and Treatment**. InTech, 2014.
- AREND, D. N. **Minitab 17 Statistical Software**. Minitab, Inc., , 2010.
- ARFKEN, G. The Method of Steepest Descents. In: **Mathematical Methods for Physicists**. 3rd. ed. Orlando: Academic Press, 1985. p. 428–436.
- BALLESTER, P. J. et al. Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. **Journal of the Royal Society, Interface / the Royal Society**, v. 9, n. 77, p. 3196–207, 7 dez. 2012.
- BARGMANN, C. I. Comparative chemosensation from receptors to ecology. **Nature**, v. 444, n. 7117, p. 295–301, 16 nov. 2006.
- BENKERT, P.; KÜNZLI, M.; SCHWEDE, T. QMEAN server for protein model quality estimation. **Nucleic acids research**, v. 37, n. Web Server issue, p. W510–4, jul. 2009.
- BENTON, R. et al. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. **PLoS biology**, v. 4, n. 2, p. e20, fev. 2006.
- BENTON, R. et al. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. **Cell**, v. 136, n. 1, p. 149–62, 9 jan. 2009.
- BENTON, R.; VANNICE, K. S.; VOSSHALL, L. B. An essential role for a CD36-related receptor in pheromone detection in *Drosophila*. **Nature**, v. 450, n. 7167, p. 289–93, 8 nov. 2007.
- BERENDSEN, H. J. C.; GRIGERA, J. R.; STRAATSMA, T. P. The missing term in effective pair potentials. **The Journal of Physical Chemistry**, v. 91, n. 24, p. 6269–6271, nov. 1987.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1-3, p. 43–56, set. 1995.
- BERMAN, H. M. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 1 jan. 2000.
- BIASINI, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic acids research**, v. 42, n. Web Server issue, p. W252–8, 29 jul. 2014.

- BIOVIA, D. S. **Discovery Studio Modeling Environment**. San Diego: Dassault Systèmes, , 2015.
- BOHBOT, J. D.; PITTS, R. J. The narrowing olfactory landscape of insect odorant receptors. **Frontiers in Ecology and Evolution**, v. 3, 10 abr. 2015.
- BRASIL, M. DA S. S. DE V. EM S. D. DE V. E. **Manual de vigilância e controle da leishmaniose visceral**. 1. ed. Brasília: Editora do Ministério da Saúde, 2003.
- BRYNGELSON, J. D. et al. Funnels, pathways, and the energy landscape of protein folding: a synthesis. **Proteins**, v. 21, n. 3, p. 167–95, mar. 1995.
- BUCK, L.; AXEL, R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. **Cell**, v. 65, n. 1, p. 175–187, abr. 1991.
- BUONFIGLIO, R. et al. Investigating Pharmacological Similarity by Charting Chemical Space. **Journal of chemical information and modeling**, v. 55, n. 11, p. 2375–90, 23 dez. 2015.
- CAMPOS, J. H. F.; COSTA, F. A. L. Participation of ticks in the infectious cycle of canine visceral leishmaniasis, in Teresina, Piauí, Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 56, n. 4, p. 297–300, jan. 2014.
- CASANOVA, C. et al. Identification of sex pheromones of *Lutzomyia longipalpis* (Lutz & Neiva, 1912) populations from the state of São Paulo, Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 101, n. 1, p. 113–115, fev. 2006.
- CHOTHIA, C.; LESK, A. M. The relation between the divergence of sequence and structure in proteins. **The EMBO journal**, v. 5, n. 4, p. 823–6, abr. 1986.
- CHRIST, C. D.; MARK, A. E.; VAN GUNSTEREN, W. F. Basic ingredients of free energy calculations: a review. **Journal of computational chemistry**, v. 31, n. 8, p. 1569–82, jun. 2010.
- COREY, E. A. et al. Ionotropic crustacean olfactory receptors. **PloS one**, v. 8, n. 4, p. e60551, jan. 2013.
- CROOKS, G. E. WebLogo: A Sequence Logo Generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, 12 maio 2004.
- DAMBERGER, F. F. et al. Pheromone discrimination by a pH-tuned polymorphism of the *Bombyx mori* pheromone-binding protein. **Proceedings of the National Academy of Sciences**, v. 110, n. 46, p. 18680–18685, 12 nov. 2013.
- DANTAS-TORRES, F. *Leishmania infantum* versus *Leishmania chagasi*: do not forget the law of priority. **Memórias do Instituto Oswaldo Cruz**, v. 101, n. 1, p. 117–118, fev. 2006.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089, 1993.
- DAVID L., A. L. N.; MICHAEL M. AND LEHNINGER, C. **Principles of Biochemistry**. New York: W. H. Freeman, 2011. v. 5
- DAVRAZOU, F. et al. New insights into the mechanism of odorant detection by the malaria-transmitting mosquito *Anopheles gambiae*. **The Journal of biological chemistry**, v. 286, n. 39, p. 34175–83, 30 set. 2011.
- DE OLIVEIRA, L. et al. Headspace solid phase microextraction/gas chromatography–mass spectrometry combined to chemometric analysis for volatile organic compounds determination in canine hair: A new tool to detect dog contamination by visceral leishmaniasis. **Journal of Chromatography B**, v. 875, n. 2, p. 392–398, 15 nov. 2008.
- DESJEUX, P. Leishmaniasis: current situation and new perspectives. **Comparative immunology, microbiology and infectious diseases**, v. 27, n. 5, p. 305–18, set. 2004.

- DI, C. et al. Visceral leishmaniasis in large Brazilian cities : challenges for control A leishmaniose visceral nos grandes centros urbanos : desafio para o controle. v. 24, n. 12, p. 2953–2958, 2008.
- DIAS, D. B. DOS S. **Análise de genes expressos nas antenas de Lutzomyia longipalpis (Lutz, Neiva, 1912) (Diptera: Psychodidae: Phlebotominae)**. Instituto Oswaldo Cruz, 2008.
- DILLON, R. J. et al. Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect-parasite relationship. **Genomics**, v. 88, n. 6, p. 831–40, dez. 2006.
- DNASTAR. <http://www.dnastar.com/blog/wp-content/uploads/2015/08/ProteinDBGrowthBar3.png>. Disponível em: <<https://www.dnastar.com/blog/structural-biology/why-structure-prediction-matters/>>. Acesso em: 16 fev. 2016.
- DPDX, L. I. OF P. D. OF P. H. C. **Parasites - Leishmaniasis**. Disponível em: <<http://www.cdc.gov/parasites/leishmaniasis/biology.html>>. Acesso em: 5 nov. 2015.
- DRIEMEIER, M. et al. Late diagnosis: a factor associated with death from visceral leishmaniasis in elderly patients. **Pathogens and Global Health**, v. 109, n. 6, p. 283–289, set. 2015.
- ELNAIEM, D. E.; WARD, R. D. Oviposition attractants and stimulants for the sandfly *Lutzomyia longipalpis* (Diptera: Psychodidae). **Journal of medical entomology**, v. 29, n. 1, p. 5–12, jan. 1992.
- EMBL-EBI. **EMBL-EBI**. Disponível em: <<http://www.ebi.ac.uk/Tools/msa/clustalo/>>. Acesso em: 16 fev. 2016.
- ESWAR, N. et al. Comparative protein structure modeling using Modeller. **Current protocols in bioinformatics**, v. Chapter 5, p. Unit 5.6, out. 2006.
- EWING, T. J. et al. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. **Journal of computer-aided molecular design**, v. 15, n. 5, p. 411–28, maio 2001.
- FAN, H. et al. Molecular docking screens using comparative models of proteins. **Journal of chemical information and modeling**, v. 49, n. 11, p. 2512–27, nov. 2009.
- FAN, H.; MARK, A. E. Refinement of homology-based protein structures by molecular dynamics simulation techniques. **Protein science : a publication of the Protein Society**, v. 13, n. 1, p. 211–20, jan. 2004.
- FAN, J. et al. An overview of odorant-binding protein functions in insect peripheral olfactory reception. **Genetics and molecular research : GMR**, v. 10, n. 4, p. 3056–69, jan. 2011.
- FERNANDES, F. D. F. et al. Fine structure and distribution pattern of antennal sensilla of *Lutzomyia longipalpis* (Diptera: Psychodidae) sand flies. **Journal of medical entomology**, v. 45, n. 6, p. 982–90, nov. 2008.
- FERRIN, T. E. et al. The MIDAS display system. **Journal of Molecular Graphics**, v. 6, n. 1, p. 13–27, mar. 1988.
- FISER, A. Template-based protein structure modeling. **Methods in molecular biology (Clifton, N.J.)**, v. 673, p. 73–94, jan. 2010.
- FREITAS-JUNIOR, L. H. et al. Visceral leishmaniasis treatment: What do we have, what do we need and how to deliver it? **International Journal for Parasitology: Drugs and Drug Resistance**, v. 2, p. 11–19, dez. 2012.
- GALIZIA, C. G.; RÖSSLER, W. Parallel Olfactory Systems in Insects: Anatomy and Function. **Annual Review of Entomology**, v. 55, n. 1, p. 399–420, jan. 2010.
- GETAHUN, M. N. et al. Insect odorant response sensitivity is tuned by metabotropically

autoregulated olfactory receptors. **PloS one**, v. 8, n. 3, p. e58889, jan. 2013.

GOLDSMITH-FISCHMAN, S.; HONIG, B. Structural genomics: computational methods for structure analysis. **Protein science : a publication of the Protein Society**, v. 12, n. 9, p. 1813–21, set. 2003.

GONZÁLEZ, M. A. A. Force fields and molecular dynamics simulations. **École thématique de la Société Française de la Neutronique**, v. 12, p. 169–200, 9 jun. 2011.

GONZÁLEZ-CABALLERO, N. et al. Transcriptome exploration of the sex pheromone gland of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). **Parasites & Vectors**, v. 6, n. 1, p. 56, jan. 2013.

GOULART, H. F. et al. Pheromones: A Green Alternative for the Integrated Pest Management. **Revista Virtual de Química**, v. 7, n. 4, p. 1205–1224, 2015.

GOVINDARAJAN, S.; RECARREN, R.; GOLDSTEIN, R. A. Estimating the total number of protein folds. **Proteins**, v. 35, n. 4, p. 408–14, 1 jun. 1999.

HA, T. S.; SMITH, D. P. Odorant and pheromone receptors in insects. **Frontiers in cellular neuroscience**, v. 3, n. September, p. 10, jan. 2009.

HALLEM, E. A.; DAHANUKAR, A.; CARLSON, J. R. Insect odor and taste receptors. **Annual review of entomology**, v. 51, n. 10, p. 113–35, jan. 2006.

HAN, L. et al. Operating mechanism and molecular dynamics of pheromone-binding protein ASP1 as influenced by pH. **PloS one**, v. 9, n. 10, p. e110565, jan. 2014.

HARVEY, A. L.; EDRADA-EBEL, R.; QUINN, R. J. The re-emergence of natural products for drug discovery in the genomics era. **Nature Reviews Drug Discovery**, v. 14, n. 2, p. 111–129, 23 jan. 2015.

HEMINGWAY, J.; RANSON, H. Insecticide Resistance in Insect Vectors of Human Disease. **Annual Review of Entomology**, v. 45, n. 1, p. 371–391, jan. 2000.

HESS, B., BEKKER, H., BERENDSEN, H.J.C. & FRAAIJE, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*. **Journal of Computational Chemistry**, n. 18, p. 1463–1472, 1997.

HESTENES, M. R.; STIEFEL, E. Methods of conjugate gradients for solving linear systems. **Journal of Research of the National Bureau of Standards**, v. 49, n. 6, p. 409, dez. 1952.

HUANG, B. MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. **OMICS: A Journal of Integrative Biology**, v. 13, n. 4, p. 325–330, ago. 2009.

IBRAGIMOVA, G. T.; WADE, R. C. Importance of explicit salt ions for protein stability in molecular dynamics simulation. **Biophysical journal**, v. 74, n. 6, p. 2906–11, jun. 1998.

IRWIN, J. J. et al. ZINC: A Free Tool to Discover Chemistry for Biology. **Journal of Chemical Information and Modeling**, v. 52, n. 7, p. 1757–1768, 23 jul. 2012.

ISHIDA Y. LEAL, W. S. Rapid inactivation of a moth pheromone. **Proc. Natl. Acad. Sci. USA**, 2005.

JAIN, K.; JAIN, N. K. Vaccines for visceral leishmaniasis: A review. **Journal of Immunological Methods**, v. 422, p. 1–12, jul. 2015.

JANSEN, S. et al. Structure of *Bombyx mori* chemosensory protein 1 in solution. **Archives of Insect Biochemistry and Physiology**, v. 66, n. 3, p. 135–145, nov. 2007.

JARVISI, E. K. . L. C. R. Laboratory Observations on Mating and Leklike Aggregations in *Lutzomyia longipalpis* (Diptera: Psychodidae). **Journal of medical entomology**, v. 29, n. 2, p.

171–177, 1992.

JAYANTHI, K. P. D. et al. Computational reverse chemical ecology: virtual screening and predicting behaviorally active semiochemicals for *Bactrocera dorsalis*. **BMC genomics**, v. 15, n. 1, p. 209, jan. 2014.

JIN, X.; HA, T. S.; SMITH, D. P. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 31, p. 10996–1001, 5 ago. 2008.

JONES, D. T.; TAYLOR, W. R.; THORNTON, J. M. The rapid generation of mutation data matrices from protein sequences. **Bioinformatics**, v. 8, n. 3, p. 275–282, 27 maio 1992.

JONES, W. D. et al. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. **Nature**, v. 445, n. 7123, p. 86–90, 4 jan. 2007.

KAUPP, U. B. Olfactory signalling in vertebrates and insects: differences and commonalities. **Nature Reviews Neuroscience**, 10 fev. 2010.

KELLENBERGER, E.; HOFMANN, A.; QUINN, R. J. Similar interactions of natural products with biosynthetic enzymes and therapeutic targets could explain why nature produces such a large proportion of existing drugs. **Natural Product Reports**, v. 28, n. 9, p. 1483, 2011.

KIM, M.-S. The Invertebrate Odorant-binding Protein LUSH is required for Normal Olfactory Behavior in *Drosophila*. **Chemical Senses**, v. 26, n. 2, p. 195–199, 1 fev. 2001.

KRIEGER, J. et al. A candidate olfactory receptor subtype highly conserved across different insect orders. **Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology**, v. 189, n. 7, p. 519–526, 1 jul. 2003.

KRISSINEL, E. On the relationship between sequence and structure similarities in proteomics. **Bioinformatics (Oxford, England)**, v. 23, n. 6, p. 717–23, 15 mar. 2007.

KUB. **No Title**. Disponível em: <<http://www.genome.jp/tools/clustalw/>>. Acesso em: 16 fev. 2016.

KUMAR, A.; ZHANG, K. Y. J. Hierarchical virtual screening approaches in small molecule drug discovery. **Methods (San Diego, Calif.)**, v. 71, p. 26–37, jan. 2015.

KUNTZ, I. D. et al. A geometric approach to macromolecule-ligand interactions. **Journal of molecular biology**, v. 161, n. 2, p. 269–88, 25 out. 1982.

LADYOFHATS. **Ciclo de vida dos parasitas do gênero Leishmania, a causa da doença Leishmaniose**. Disponível em: <https://commons.wikimedia.org/wiki/File:Leishmaniasis_life_cycle_diagram_en.svg?uselang=pt-br>. Acesso em: 7 abr. 2015.

LARSSON, J. et al. ChemGPS-NP: Tuned for navigation in biologically relevant chemical space. **Journal of Natural Products**, v. 70, n. 5, p. 789–794, 2007.

LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, 1 abr. 1993.

LAUGHLIN, J. D., HA, T. S., JONES, D. N. M., AND SMITH, D. P. Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein. **Cell**, v. 133, p. 1255–1265, 2008.

LAVECCHIA, A.; DI GIOVANNI, C. Virtual screening strategies in drug discovery: a critical review. **Current medicinal chemistry**, v. 20, n. 23, p. 2839–60, jan. 2013.

LEAL, G. M.; LEAL, W. S. Binding of a fluorescence reporter and a ligand to an odorant-binding protein of the yellow fever mosquito, *Aedes aegypti*. **F1000Research**, 9 jan. 2015.

- LEAL, W. S. Pheromone Reception. p. 1–36, 2005.
- LEAL, W. S. et al. Reverse and conventional chemical ecology approaches for the development of oviposition attractants for *Culex* mosquitoes. **PloS one**, v. 3, n. 8, p. e3045, jan. 2008.
- LEAL, W. S. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. **Annual review of entomology**, v. 58, n. September 2012, p. 373–91, jan. 2013.
- LEMON, S. M. et al. Vector-borne diseases: understanding the environmental, human health, and ecological connections. In: **Vector-Borne Diseases**. Washington (DC): National Academies Press (US), 2008. p. 1–69.
- LI, N. et al. Using Hierarchical Virtual Screening To Combat Drug Resistance of the HIV-1 Protease. **Journal of chemical information and modeling**, v. 55, n. 7, p. 1400–12, 27 jul. 2015a.
- LI, Z.-Q. et al. Transcriptome comparison of the sex pheromone glands from two sibling *Helicoverpa* species with opposite sex pheromone components. **Scientific reports**, v. 5, p. 9324, 20 jan. 2015b.
- LI, Z.-Q. et al. Odorant-binding proteins display high affinities for behavioral attractants and repellents in the natural predator *Chrysopa pallens*. **Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology**, v. 185, p. 51–57, jul. 2015c.
- LIU, T.; TANG, G. W.; CAPRIOTTI, E. Comparative modeling: the state of the art and protein drug target structure prediction. **Combinatorial chemistry & high throughput screening**, v. 14, n. 6, p. 532–47, jul. 2011.
- LLC, T. G. N. **Gates Notes**. Disponível em: <http://www.gatesnotes.com/Health/Most-Lethal-Animal-Mosquito-Week?WT.mc_id=MosquitoWeek2014_SharkWeek_tw&WT.tsrc=Twitter>. Acesso em: 25 dez. 2015.
- LOBANOV, M. Y.; BOGATYREVA, N. S.; GALZITSKAYA, O. V. Radius of gyration as an indicator of protein structure compactness. **Molecular Biology**, v. 42, n. 4, p. 623–628, 10 ago. 2008.
- LOGAN, J. G.; BIRKETT, M. A. Semiochemicals for biting fly control: their identification and exploitation. **Pest management science**, v. 63, n. 7, p. 647–57, jul. 2007.
- LORENZO, M. G. MELO, A. C. A. MARA. Olfacção e Comportamento. In: **Tópicos Avançados em Entomologia Molecular**. 1. ed. Rio de Janeiro: INCT-EM: Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular - INCTEM, 2012. p. 1–20.
- LUNDIN, C. et al. Membrane topology of the *Drosophila* OR83b odorant receptor. **FEBS letters**, v. 581, n. 29, p. 5601–4, 11 dez. 2007.
- MAGALHÃES-JUNIOR, J. T. et al. A laboratory evaluation of alcohols as attractants for the sandfly *Lutzomyia longipalpis* (Diptera:Psychodidae). **Parasites & Vectors**, v. 7, n. 1, p. 60, jan. 2014.
- MAIOROV, V. N.; CRIPPEN, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. **Journal of molecular biology**, v. 235, n. 2, p. 625–34, 14 jan. 1994.
- MANOHARAN, M. et al. Association of putative members to family of mosquito odorant binding proteins: scoring scheme using fuzzy functional templates and cys residue positions. **Bioinformatics and biology insights**, v. 7, p. 231–51, jan. 2013a.
- MANOHARAN, M. et al. Comparative genomics of odorant binding proteins in *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*. **Genome biology and evolution**, v. 5, n. 1, p. 163–80, 1 jan. 2013b.

- MAO, Y. et al. Crystal and solution structures of an odorant-binding protein from the southern house mosquito complexed with an oviposition pheromone. **Proceedings of the National Academy of Sciences**, v. 107, n. 44, p. 19102–19107, 2 nov. 2010.
- MARSHALL, G. R. Introduction to Chemoinformatics in Drug Discovery - A Personal View. In: Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2005. p. 1–22.
- MCDOWELL, M. A.; RAMALHO-ORTIGAO, M.; DILLON, R. **Proposal for Sequencing the Genome of the Sand Flies , *Lutzomyia longipalpis* and *Phlebotomus papatasi***. Disponível em: <<https://www.genome.gov/Pages/Research/Sequencing/SeqProposals/>>. Acesso em: 5 nov. 2015.
- MCGOVERN, S. L.; SHOICHET, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. **Journal of medicinal chemistry**, v. 46, n. 14, p. 2895–907, 3 jul. 2003.
- MELO, F. et al. ANOLEA: a www server to assess protein structures. **International Conference on Intelligent Systems for Molecular Biology**, v. 5, p. 187–90, jan. 1997.
- MENG, E. C.; SHOICHET, B. K.; KUNTZ, I. D. Automated docking with grid-based energy evaluation. **Journal of Computational Chemistry**, v. 13, n. 4, p. 505–524, maio 1992.
- MISSBACH, C. et al. Evolution of insect olfactory receptors. **eLife**, v. 3, 26 mar. 2014.
- MOULT, J. et al. Critical assessment of methods of protein structure prediction (CASP)-round IX. **Proteins: Structure, Function, and Bioinformatics**, v. 79, n. S10, p. 1–5, 2011.
- NAMBA, A. M.; SILVA, V. B. DA; SILVA, C. H. T. P. DA. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, v. 33, n. 4, p. 13–24, dez. 2008.
- NAQSH E ZAHRA, S.; KHATTAK, N. A.; MIR, A. Comparative modeling and docking studies of p16ink4/cyclin D1/Rb pathway genes in lung cancer revealed functionally interactive residue of RB1 and its functional partner E2F1. **Theoretical biology & medical modelling**, v. 10, n. 1, p. 1, 1 jan. 2013.
- NAVARRO-SILVA, M. A.; MARQUES, F. A.; DUQUE L, J. E. Review of semiochemicals that mediate the oviposition of mosquitoes: a possible sustainable tool for the control and monitoring of Culicidae. **Revista Brasileira de Entomologia**, v. 53, n. 1, p. 1–6, 2009.
- NIENHAUS, G. U. **Protein-Ligand Interactions**. Germany: Wiley- VCH, 2005. v. 305
- NOLTE, A. et al. In situ tip-recordings found no evidence for an Orco-based ionotropic mechanism of pheromone-transduction in *Manduca sexta*. **PloS one**, v. 8, n. 5, p. e62648, jan. 2013.
- OLIVEIRA, D. F. DE. **Identificação dos Criadouros Naturais de *Lutzomyia longipalpis* (DIPTERA: PSYCHODIDAE) em área endêmica para a Leishmaniose Visceral, do estado da Bahia, Brasil**. Salvador: Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa, 2013.
- OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. **Journal of computational chemistry**, v. 25, n. 13, p. 1656–76, out. 2004.
- OROZCO, M. A theoretical view of protein dynamics. **Chemical Society Reviews**, v. 43, n. 14, p. 5051, 21 jul. 2014.
- OWUSU, H. F. et al. Comparability between insecticide resistance bioassays for mosquito vectors: time to review current methodology? **Parasites & Vectors**, v. 8, n. 1, p. 357, 7 dez. 2015.
- PELOSI, P. et al. Soluble proteins of chemical communication: An overview across arthropods. **Frontiers in Physiology**, v. 5 AUG, 2014a.

- PELOSI, P. et al. Structure and biotechnological applications of odorant-binding proteins. **Applied Microbiology and Biotechnology**, v. 98, n. 1, p. 61–70, 22 jan. 2014b.
- PELOSI, P.; CALVELLO, M.; BAN, L. Diversity of odorant-binding proteins and chemosensory proteins in insects. **Chemical senses**, v. 30 Suppl 1, n. suppl 1, p. i291–2, jan. 2005.
- PETERSEN, T. N. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, v. 8, n. 10, p. 785–786, 29 set. 2011.
- PETTERSEN, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of computational chemistry**, v. 25, n. 13, p. 1605–12, out. 2004.
- PICKETT, J. A. et al. Chemical Ecology of Animal and Human Pathogen Vectors in a Changing Global Climate. **Journal of Chemical Ecology**, v. 36, n. 1, p. 113–121, 1 jan. 2010.
- PICKETT, J. A.; BARASA, S.; BIRKETT, M. A. Vertebrate pheromones and other semiochemicals: the potential for accommodating complexity in signalling by volatile compounds for vertebrate management. **Biochemical Society Transactions**, v. 42, n. 4, p. 846–850, 1 ago. 2014.
- PIMENTEL, M. I. F. et al. High visceral leishmaniasis mortality rate in Barra Mansa, a new area of visceral leishmaniasis transmission in the State of Rio de Janeiro, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 47, n. 4, p. 521–523, jul. 2014.
- PRONK, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics (Oxford, England)**, v. 29, n. 7, p. 845–54, 1 abr. 2013.
- PUGSLEY, A. P. **Protein targeting**. San Diego: Academic Press, 1989.
- QUINNELL, R. J.; COURTENAY, O. Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. **Parasitology**, v. 136, n. 14, p. 1915, 16 dez. 2009.
- READY, P. D. Biology of phlebotomine sand flies as vectors of disease agents. **Annual review of entomology**, v. 58, p. 227–50, jan. 2013.
- READY, P. D. Epidemiology of visceral leishmaniasis. **Clinical epidemiology**, v. 6, p. 147–54, jan. 2014.
- READY, P. D.; ROGERS, M. E. Behaviour of sandflies infected with *Leishmania*. In: CJM KOENRAADT, W. T. (Ed.). **Ecology of parasite-vector interactions**. Wageningen: The Neth.: Wageningen Acad. Publ., 2013. v. 3p. 167–178.
- RODRIGUES, R. P. et al. Estratégias de Triagem Virtual no Planejamento de Fármacos. **Revista Virtual de Química**, v. 4, n. 6, p. 739–776, 2012.
- ROGERS ME, B. P. *Leishmania* manipulation of sand fly feeding behaviour results in enhanced transmission. **PLoS Pathog**, v. 3, n. 91, 2007.
- ROST, B. Twilight zone of protein sequence alignments. **Protein Engineering Design and Selection**, v. 12, n. 2, p. 85–94, 1 fev. 1999.
- ROSTKOWSKI, M. et al. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. **BMC structural biology**, v. 11, n. 1, p. 6, 26 jan. 2011.
- SANTOS FILHO, O. A.; ALENCASTRO, R. B. DE. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253–259, mar. 2003.
- SATO, K. et al. Insect olfactory receptors are heteromeric ligand-gated ion channels. **Nature**, v. 452, n. 7190, p. 1002–6, 24 abr. 2008.
- SAXENA, A. K.; ROY, K. K. Hierarchical virtual screening: identification of potential high-affinity and selective $\beta(3)$ -adrenergic receptor agonists. **SAR and QSAR in environmental**

research, v. 23, n. 5-6, p. 389–407, 20 jul. 2012.

SCHYMURA, D. Antennal expression pattern of two olfactory receptors and an odorant binding protein implicated in host odor detection by the malaria vector *Anopheles gambiae*. **International Journal of Biological Sciences**, p. 614–626, 2010.

SCOTT, K. et al. A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. **Cell**, v. 104, n. 5, p. 661–73, 9 mar. 2001.

SENGUL, M. S.; TU, Z. Expression analysis and knockdown of two antennal odorant-binding protein genes in *Aedes aegypti*. **Journal of insect science (Online)**, v. 10, n. 171, p. 171, jan. 2010.

SHOICHET, B. K.; KUNTZ, I. D.; BODIAN, D. L. Molecular docking using shape descriptors. **Journal of Computational Chemistry**, v. 13, n. 3, p. 380–397, abr. 1992.

SIB. **ClustalW IN: Expasy Bioinformatics Resource Portal**. Disponível em: <<http://embnet.vital-it.ch/software/ClustalW.html>>. Acesso em: 16 fev. 2016.

SNEDEKER, S. M. Pesticides and breast cancer risk: A review of DDT, DDE and dieldrin. **Environmental Health Perspectives**, v. 109, n. SUPPL. 1, p. 35–47, 2001.

SOTRIFFER, C. **Virtual Screening**. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 1999.

SPIEGEL, C. N. et al. Structure and distribution of sensilla on maxillary palps and labella of *Lutzomyia longipalpis* (Diptera: Psychodidae) sand flies. **Microscopy Research and Technique**, v. 66, n. 6, p. 321–330, 15 abr. 2005.

STEINBRECHT, R. A. Structure and function of insect olfactory sensilla. **Ciba Foundation symposium**, v. 200, p. 158–74; discussion 174–7, jan. 1996.

STERLING, T.; IRWIN, J. J. ZINC 15 - Ligand Discovery for Everyone. **Journal of chemical information and modeling**, v. 55, n. 11, p. 2324–2337, 19 out. 2015.

STIERAND, K.; RAREY, M. PoseViewWeb: Two Dimensional Diagrams of Complexes with known Complex Structures. 2007a.

STIERAND, K.; RAREY, M. From Modeling to Medicinal Chemistry: Automatic Generation of Two-Dimensional Complex Diagrams. **ChemMedChem**, v. 2, n. 6, p. 853–860, 11 jun. 2007b.

STIERAND, K.; RAREY, M. Drawing the PDB: Protein–Ligand Complexes in Two Dimensions. **ACS Medicinal Chemistry Letters**, v. 1, n. 9, p. 540–545, 9 dez. 2010.

STOWERS, L.; LOGAN, D. W. LUSH Shapes Up for a Starring Role in Olfaction. **Cell**, v. 133, n. 7, p. 1137–1139, jun. 2008.

STUETZ, W. Global surveillance of DDT and DDE levels in human tissues. **International journal of occupational medicine and environmental health**, v. 19, n. 1, p. 83, jan. 2006.

SUH, E.; BOHBOT, J. D.; ZWIEBEL, L. J. Peripheral olfactory signaling in insects. **Current Opinion in Insect Science**, v. 6, p. 86–92, dez. 2014.

TAKKEN, W.; KNOLS, B. G. J. **Olfaction in Vector-host Interactions**. Wageningen. v.2. 2010.

TAKKEN, W.; VERHULST, N. O. Host preferences of blood-feeding mosquitoes. **Annual review of entomology**, v. 58, p. 433–53, jan. 2013.

TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725–2729, 1 dez. 2013.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

- penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673–4680, 11 nov. 1994.
- TOWNSON, H. ET AL. Exploiting the potential of vector control for disease prevention. **Bulletin of the World Health Organization**, v. 83, n. (12), 2005.
- TSITSANO, K. E. et al. Anopheles gambiae odorant binding protein crystal complex with the synthetic repellent DEET: implications for structure-based design of novel mosquito repellents. **Cellular and Molecular Life Sciences**, v. 69, n. 2, p. 283–297, 14 jan. 2012.
- TURNBULL, A. P.; EMSLEY, P. Studying protein-ligand interactions using X-ray crystallography. **Methods in molecular biology (Clifton, N.J.)**, v. 1008, p. 457–77, jan. 2013.
- UNIL. **Homology modeling.** Disponível em: <<http://www.unil.ch/pmf/en/home/menuinst/technologies/homology-modeling.html>>. Acesso em: 16 fev. 2016.
- UNIPROT. UniProt: a hub for protein information. **Nucleic Acids Research**, v. 43, n. D1, p. D204–212, 27 out. 2014.
- VAN DEN BERG, H. DDT and Malaria Prevention: van den Berg Responds. **Environmental Health Perspectives**, 1 jan. 2010.
- VERHULST, N. O. et al. Differential attraction of malaria mosquitoes to volatile blends produced by human skin bacteria. **PloS one**, v. 5, n. 12, p. e15829, jan. 2010.
- VERLET, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. **Physical Review**, v. 159, n. 1, p. 98–103, 5 jul. 1967.
- VILAR, S. et al. Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models. **Journal of molecular graphics & modelling**, v. 29, n. 5, p. 614–23, fev. 2011.
- VOGT, R. G.; RIDDIFORD, L. M. Pheromone binding and inactivation by moth antennae. **Nature**, v. 293, n. 5828, p. 161–163, 10 set. 1981.
- VYAS, V. Virtual Screening: A Fast Tool for Drug Design. **Scientia Pharmaceutica**, v. 76, n. 3, p. 333–360, 2008.
- WALTERS, L. L. et al. Life Cycle of Leishmania major (Kinetoplastida: Trypanosomatidae) in the Neotropical Sand Fly Lutzomyia longipalpis (Diptera: Psychodidae). **Journal of Medical Entomology**, v. 30, n. 4, p. 699–718, 1 jul. 1993.
- WEBB, B. et al. Comparative Modeling of Drug Target Proteins☆. In: **Reference Module in Chemistry, Molecular Sciences and Chemical Engineering**. San Francisco: Elsevier, 2014.
- WERNECK, G. L. Geographic spread of visceral leishmaniasis in Brazil. **Cad Saúde Pública**, v. 26, n. 644, p. 5, 2010.
- WESCHLER, C. J.; NAZAROFF, W. W. SVOC partitioning between the gas phase and settled dust indoors. **Atmospheric Environment**, v. 44, n. 30, p. 3609–3620, 2010.
- WHITEHEAD, T. L. Molecular Modeling: Basic Principles and Applications, 2nd Edition (Hans-Dieter Höltje, Wolfgang Sippl, Didier Rognan, and Gerd Folkers). **Journal of Chemical Education**, v. 83, n. 6, p. 116–118, jun. 2006.
- WHO. **Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of Leishmaniases**. WHO Techn ed. Geneva: WHO Press, 2010a.
- WHO. Control of the leishmaniasis. **World Health Organization technical report series**, n. 949, p. xii–xiii, 1–186, back cover, jan. 2010b.

- WHO. The use of DDT in malaria vector control. 2011.
- WHO. **A global brief on vector-borne diseases**. Geneva: World Health Organization, 2014.
- WHO. **World : Endemicity of visceral leishmaniasis**. Disponível em: <http://gamapsserver.who.int/mapLibrary/Files/Maps/Leishmaniasis_2013_VL.png>. Acesso em: 4 fev. 2016.
- WICHER, D. et al. Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. **Nature**, v. 452, n. 7190, p. 1007–1011, 24 abr. 2008.
- WIEGMANN, B. M. et al. Episodic radiations in the fly tree of life. **Proceedings of the National Academy of Sciences**, v. 108, n. 14, p. 5690–5695, 5 abr. 2011.
- WOJTASEK, H.; LEAL, W. S. Conformational change in the pheromone-binding protein from *Bombyx mori* induced by pH and by interaction with membranes. **The Journal of biological chemistry**, v. 274, n. 43, p. 30950–6, 22 out. 1999.
- WORTH, C. L.; GONG, S.; BLUNDELL, T. L. Structural and functional constraints in the evolution of protein families. **Nature Reviews Molecular Cell Biology**, 2 dez. 2009.
- XIANG, Z. Advances in homology protein structure modeling. **Current protein & peptide science**, v. 7, n. 3, p. 217–27, jun. 2006.
- XU, P. et al. Drosophila OBP LUSH is required for activity of pheromone-sensitive neurons. **Neuron**, v. 45, n. 2, p. 193–200, 20 jan. 2005.
- XU, Y.-L. et al. Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects. **BMC genomics**, v. 10, p. 632, jan. 2009.
- XUE, L.; BAJORATH, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. **Combinatorial chemistry & high throughput screening**, v. 3, n. 5, p. 363–72, out. 2000.
- YIN, J. et al. Selectivity of odorant-binding proteins from the southern house mosquito tested against physiologically relevant ligands. **Frontiers in physiology**, v. 6, p. 56, jan. 2015.
- YOUNUS, F. et al. Identification of candidate odorant degrading gene/enzyme systems in the antennal transcriptome of *Drosophila melanogaster*. **Insect Biochemistry and Molecular Biology**, v. 53, p. 30–43, out. 2014.
- ZHANG, L.; SKOLNICK, J. What should the Z-score of native protein structures be? **Protein science : a publication of the Protein Society**, v. 7, n. 5, p. 1201–7, maio 1998.
- ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC bioinformatics**, v. 9, n. 1, p. 40, 23 jan. 2008.
- ZHANG, Z. et al. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. **Bioinformatics**, v. 27, n. 15, p. 2083–2088, 1 ago. 2011.
- ZHOU, J. J. Odorant-binding proteins in insects. In: **Vitamins and Hormones**. San Diego: Elsevier Inc., 2010. v. 83p. 241–272.
- ZHOU, R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. **Proteins**, v. 53, n. 2, p. 148–61, 1 nov. 2003.
- ZIEMBA, B. P. et al. A novel mechanism of ligand binding and release in the odorant binding protein 20 from the malaria mosquito *Anopheles gambiae*. **Protein Science**, v. 22, n. 1, p. 11–21, jan. 2013.