



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA

Leonardo de Bortoli

**Identificação de interfaces de interação de dimerização de receptores
nucleares da ostra-do-pacífico por dinâmica molecular**

Florianópolis
2025

Leonardo de Bortoli

**Identificação de interfaces de interação de dimerização de receptores
nucleares da ostra-do-pacífico por dinâmica molecular**

Dissertação submetida ao Programa de Pós-
Graduação da Universidade Federal de Santa
Catarina como requisito parcial para a obtenção do
título de Mestre em Bioquímica
Orientador: Prof. Dr. Guilherme Razzera Maciel

Florianópolis

2025

Leonardo de Bortoli

**Identificação de interfaces de interação de dimerização de receptores
nucleares da ostra-do-pacífico por dinâmica molecular**

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em [dia] de [mês]
de [ano de defesa] pela banca examinadora composta pelos seguintes membros:

Prof. Rafael Diego da Rosa, Dr.
Universidade Federal de Santa Catarina

Prof. Hernán Francisco Terenzi, Dr.
Universidade Federal de Santa Catarina

Prof.^a Maria Risoleta Freire Marques, Dra.
Universidade Federal de Santa Catarina

Prof. Carlos Peres Silva, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi
julgado adequado para obtenção do título de Mestre em Bioquímica.



Documento assinado digitalmente

Ariane Zamoner Pacheco de Souza

Data: 05/11/2025 17:11:16-0300

CPF: ***.637.269-**

Verifique as assinaturas em <https://v.ufsc.br>

_ _ ordenação do Programa de Pós-Graduação

Prof. Guilherme Razzera Maciel, Dr.
Orientador



Documento assinado digitalmente

GUILHERME RAZZERA MACIEL

Data: 05/11/2025 05:56:51-0300

CPF: ***.635.709-**

Verifique as assinaturas em <https://v.ufsc.br>

Florianópolis, 2025.

Dedico este trabalho aos meus pais,
Sandra e Giovanni.

AGRADECIMENTOS

Gostaria de começar dizendo que, se houvesse caminhado só, este trabalho não haveria de ser. Por isso, quero dizer obrigado primeiramente para meus pais, Sandra e Giovani, pelo incentivo constante durante toda a minha vida. Para meu orientador Guilherme, por acreditar que eu seria capaz e me dar uma chance. Para a minha parceira Giulia, por todo o amor e paciência. E para meus amigos Theo, Karin, Tâmelá e Julia, pelas ideias, correções, ensinamentos e cafés.

Gostaria também de agradecer a Universidade Federal de Santa Catarina (UFSC) e o Programa de Pós-Graduação em Bioquímica (PPGBQA). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“Nada é orgânico, é tudo programado.”
(Pitty)

RESUMO

Os ecossistemas aquáticos são altamente impactados por atividades antrópicas, incluindo o despejo compostos xenobióticos como desreguladores endócrinos que alteram negativamente funções fisiológicas dos animais aquáticos. Assim, faz-se necessário identificar novos biomarcadores de poluição ambiental como os receptores nucleares (NR), uma superfamília de proteínas que regula vias metabólicas como o metabolismo de xenobióticos. Os NR são universais em metazoários e compartilham uma estrutura conservada, contendo um domínio de ligação ao DNA (DBD) e outro de ligação ao ligante (LBD), cruciais para suas atividades genômicas, bem como regiões menos conservadas. Quando ativados pelos ligantes, estes receptores iniciam o processo de dimerização que pode ser homo ou heterodimérico. A formação de dímeros é mediada pelo LBD, porém há pouca informação sobre quais interações são responsáveis para a manutenção deste processo, principalmente em espécies aquáticas. Assim, o objetivo deste trabalho foi identificar resíduos aminoacídicos e interfaces de interação ativas em dímeros de NR envolvidos em vias metabólicas essenciais para *Homo sapiens* e *Magallana gigas*. O receptor X retinoide (RXR) e seus parceiros de dimerização foram escolhidos como objeto de estudo, visto que este receptor é parceiro da maioria dos demais NR, formando um *hub* de dímeros diversos. Para isso, foram utilizados programas como AlphaFold 3 para modelagem tridimensional e análises de dinâmica molecular *coarse-grained* para identificar contatos entre receptores, além de atracamento e filogenia molecular para investigar e caracterizar o subgrupo *MgNR1P* identificado nas sequências. Foi possível simular e identificar interfaces de interação em cinco sistemas de dímeros entre RXR e parceiros de dimerização (PPAR, RAR e NR1P10), sendo dois controles positivos de *H. sapiens* contendo os domínios conservados, região da alça e moléculas de DNA, além de três sistemas de dímeros de *M. gigas* contendo os domínios conservados e região da alça. Os resultados apontam para a presença de interfaces conhecidas entre os LBD, além de interfaces inéditas entre os domínios conservados e maior relevância para interações entre DBD. A análise filogenética de NR de *M. gigas* posicionou o subgrupo NR1P como clado monofilético. Ainda, foi possível identificar uma nova disposição de α -hélices em seu LBD e a presença de um motivo conservado que possa ter relação com este achado. Ademais, identificamos possíveis ligantes de alta afinidade com os NR de *MgNR1P* como ácido retinoico e ácido araquidônico. Por fim, recomenda-se a inclusão dos sistemas de dímeros aqui estudados em ensaios de ecotoxicologia e metodologias de hibridização para averiguar o efeito de contaminantes no comportamento destes sistemas e para validar os dímeros e interfaces aqui reportados.

Palavras-chave: ecotoxicologia, biomarcadores, receptores nucleares, dinâmica molecular, interações proteína-proteína.

ABSTRACT

Aquatic ecosystems are highly impacted by anthropic activities, such as dumping of xenobiotic compounds including endocrine disruptors which negatively affect physiological functions of aquatic animals. Thus, it is essential to find new biomarker molecules, such as nuclear receptors (NR) which are a protein superfamily involved in the regulation of various metabolic pathways. NR are ubiquitous in metazoans and share a conserved structure, comprising a DNA binding domain (DBD) and a ligand binding domain (LBD), indispensable for their genomic functions, as well as less conserved regions. When activated by a ligand, these receptors begin the dimerization process, which can be homo or heterodimeric. Dimer formation is mainly mediated by the LBD, although limited information about which interactions are responsible for this process is available, especially in aquatic species. Therefore, the main objective of this study was to identify active aminoacidic residues and interaction interfaces of NR dimers involved in essential metabolic pathways for *Homo sapiens* and *Magallana gigas*. The retinoid X receptor (RXR) and its dimerization partners were chosen as the subject of this study due to RXR being the main partner of most NR, creating a hub of different dimers. To achieve this goal, programs such as AlphaFold 3 were used for tridimensional modeling and coarse-grained molecular dynamics analyses were employed to identify contacts between receptors. Additionally, molecular docking and phylogeny were conducted to assess and characterize the MgNR1P subgroup identified in the sequences. Five RXR and partner receptors (PPAR, RAR and NR1P10) dimer systems were simulated with identified interaction interfaces. Two systems were comprised of *H. sapiens* NR with both conserved domains, hinge region e DNA molecules, and three systems were comprised of *M. gigas* dimers containing the conserved domains and hinge region. The results point to the presence of known interaction interfaces among LBD, as well as novel interfaces between the conserved domains, especially the DBD. Phylogenetic analysis of *M. gigas* NR positioned the NR1P subgroup in a monophyletic clade. Additionally, a new alpha-helix disposition on the NR1P LBD was identified, together with a conserved motif which may be associated with this finding. Furthermore, possible ligands for the NR1P subgroup were identified, such as retinoic acid and arachidonic acid. Finally, we recommend the inclusion of the dimers systems presented here in ecotoxicological assays and hybridization experiments to assess the effect of contaminants in the behaviour of these systems and to validate the dimers and interfaces reported in this study.

Keywords: ecotoxicology, biomarkers, nuclear receptors, molecular dynamics, protein-protein interactions.

LISTA DE FIGURAS

Figura 1 – Estrutura tridimensional dos domínios de ligação ao DNA e de ligação ao ligante.....	16
Figura 2 – Alinhamentos múltiplos de sequência entre receptores nucleares das subfamílias NR1, NR2 e NR3 de <i>M. gigas</i>	33
Figura 3 – Valores de pLDDT para os modelos tridimensionais de dímeros de receptores nucleares de <i>H. sapiens</i> e <i>M. gigas</i>	35
Figura 4 – Alinhamentos estruturais entre estruturas resolvidas experimentalmente de receptores nucleares e modelos tridimensionais gerados pelo AlphaFold 3.....	36
Figura 5 – Modelo tridimensional do dímero RXR/ROR de <i>H. sapiens</i>	38
Figura 6 – Predição de interfaces de interação entre dímeros de RXR e PPAR, RAR, GR e NR1P10.....	40
Figura 7 – Frequência de contatos com menos de 5 Å entre <i>HsRXR</i> e <i>HsPPAR</i> entre todas as réplicas.....	44
Figura 8 – Valores médios de RMSD, RMSF e mapa de contatos do sistema <i>HsRXR/HsPPAR</i>	45
Figura 9 – Frequência de contatos com menos de 5 Å entre <i>MgRXR</i> e <i>MgPPAR</i> entre todas as réplicas.....	47
Figura 10 – Valores médios de RMSD, RMSF e mapa de contatos do sistema <i>MgRXR/MgPPAR</i>	49
Figura 11 – Frequência de contatos com menos de 5 Å entre <i>HsRXR</i> e <i>HsRAR</i> entre todas as réplicas.....	52
Figura 12 – Valores médios de RMSD, RMSF e mapa de contatos do sistema <i>HsRXR/HsRAR</i>	54
Figura 13 – Frequência de contatos com menos de 5 Å entre <i>MgRXR</i> e <i>MgRAR</i> entre todas as réplicas.....	55
Figura 14 – Valores médios de RMSD, RMSF e mapa de contatos do sistema <i>MgRXR/MgRAR</i>	56
Figura 15 – Frequência de contatos com menos de 5 Å entre <i>MgRXR</i> e <i>MgNR1P10</i> entre todas as réplicas.....	59
Figura 16 – Valores médios de RMSD, RMSF e mapa de contatos do sistema <i>MgRXR/MgNR1P10</i>	60
Figura 17 – Filogenia molecular dos receptores nucleares de <i>M. gigas</i>	62
Figura 18 – Modelos tridimensionais do LBD de receptores nucleares do subgrupo NR1P de <i>M. gigas</i>	64

Figura 19 – Alinhamento múltiplo de sequências de receptores nucleares da subfamília NR1 de <i>M. gigas</i>	65
Figura 20 – Predição da posição do sítio ativo de <i>MgNR1P10</i>	66

LISTA DE QUADROS

<u>Quadro 1 - Receptores nucleares de Homo sapiens e Magallana gigas e seus</u> <u>códigos de acesso.....</u>	<u>29</u>
--	-----------

LISTA DE TABELAS

<u>Tabela 1 – Comprimento das sequências de receptores nucleares do subgrupo NR1P de <i>M. gigas</i>.....</u>	<u>33</u>
<u>Tabela 2 – Volume predito do sítio ativo de NR do subgrupo NR1P.....</u>	<u>66</u>
<u>Tabela 3 – Valores das funções de energia do atracamento molecular entre receptores nucleares e ligantes.....</u>	<u>67</u>

LISTA DE ABREVIATURAS E SIGLAS

9REA – Ácido 9-cis-retinoico

ATRA – Ácido *all-trans* retinoico

BLAST – *Basic Local Alignment Search Tool*

CAR – Receptor constitutivo de androstano

CASP – *Critical Assessment of protein Structure Prediction*

CTE – Extensão C-terminal

DBD – Domínio de ligação ao ligante

DNA – Ácido desoxirribonucleico

DR – Repetição direta

EcR – Receptor de ecdisona

ER – Receptor de estrógeno

GR – Receptor glicocorticoide

HAP – Hidrocarboneto aromático policíclico

HNF4 – Fator hepático nuclear 4

Hs – Homo sapiens

ipTM – *interface predicted Template Modeling*

IR – Repetição invertida

LBD – Domínio de ligação ao ligante

LBP – Sítio de ligação ao ligante

LXR – Receptor hepático X

MD – Dinâmica molecular

Mg – Magallana gigas

MR – Receptor mineralocorticoide

MSA – Alinhamento múltiplo de sequências

NHR – Receptor hormonal nuclear

NR – Receptor nuclear

NR1P – Subgrupo P da subfamília NR1

PCB – Bifenila policlorada

PDB – *Protein Data Bank*

pLDDT – *predicted Local-Distance Difference Test*

PPAR – Receptor ativado por proliferador de perioxossomos

PPCP – Produtos de cuidado e uso pessoal

pTM – *predicted Template Modeling*

PXR – Receptor pregnano X

RAR – Receptor de ácido retinoico

RMSD – Desvio quadrático médio

RMSF – Flutuação quadrática média

RNA – Ácido ribonucleico

ROR - receptor órfão relacionado ao RAR

RXR – Receptor retinoide X

TBT – Tributilestanho

VDR – Receptor de vitamina D

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 OS AMBIENTES AQUÁTICOS E A ECOTOXICOLOGIA.....	13
1.2 A SUPERFAMÍLIA DOS RECEPTORES NUCLEARES.....	14
1.2.1 A dimerização dos receptores nucleares.....	16
1.3 A OSTRADO-PACÍFICO COMO MODELO DE ESTUDO.....	18
1.4 A BIOINFORMÁTICA ESTRUTURAL: MODELAGEM, ATRACAMENTO E DINÂMICA MOLECULAR.....	19
2. OBJETIVOS.....	27
2.1 OBJETIVO GERAL.....	27
2.2 OBJETIVOS ESPECÍFICOS.....	27
3. METODOLOGIA.....	23
3.1 OBTENÇÃO DE SEQUÊNCIAS.....	23
3.2 IDENTIFICAÇÃO DE DOMÍNIOS E RESÍDUOS CONSERVADOS.....	24
3.3 MODELAGEM TRIDIMENSIONAL.....	24
3.4 IDENTIFICAÇÃO DE SÍTIOS ATIVOS E ATRACAMENTO DE LIGANTES.....	25
3.5 DINÂMICA MOLECULAR.....	26
3.6 IDENTIFICAÇÃO DE INTERAÇÕES ENTRE PROTEÍNAS.....	27
3.7 ANÁLISES FILOGENÉTICAS.....	28
4. RESULTADOS E DISCUSSÃO.....	28
4.1 SEQUÊNCIAS DE NR IDENTIFICADAS NA OSTRADO-PACÍFICO.....	29
4.2 MODELAGEM TRIDIMENSIONAL.....	33
4.3 DINÂMICA MOLECULAR.....	41
4.3.1 O sistema RXR/PPAR.....	41
4.3.2 O sistema RXR/RAR.....	49
4.3.3 O sistema <i>MgRXR/MgNR1P10</i>.....	57
4.4 CARACTERIZAÇÃO DO SUBGRUPO NR1P.....	61
5. CONCLUSÃO.....	68
REFERÊNCIAS.....	70
APÊNDICE A - PARÂMETROS UTILIZADOS NAS ANÁLISES DE DINÂMICA MOLECULAR.....	71

APÊNDICE B - MÉTRICAS DE QUALIDADE DA MODELAGEM
TRIDIMENSIONAL DE RECEPTORES NUCLEARES DE *H. SAPIENS*.....72

APÊNDICE C - MÉTRICAS DE QUALIDADE DA MODELAGEM
TRIDIMENSIONAL DE RECEPTORES NUCLEARES DE *M. GIGAS*.....73

1. INTRODUÇÃO

1.1 OS AMBIENTES AQUÁTICOS E A ECOTOXICOLOGIA

Os ambientes aquáticos sofrem constantemente com o despejo de resíduos vindos de atividades antrópicas, tanto domiciliares quanto industriais, afetando negativamente a qualidade da água e a fisiologia dos seres vivos que habitam esses ecossistemas. Entre os poluentes comumente encontrados em rios e oceanos estão os chamados “contaminantes emergentes” que ainda não são devidamente gerenciados nas estações de tratamento e regulados, com efeitos ainda pouco compreendidos na fisiologia animal (Capitão, 2021; Gore, 2015; Bourguet, 2015). Entre estes compostos, pode-se citar os desreguladores endócrinos como os hidrocarbonetos aromáticos policíclicos (HAP), ftalatos, bifenilas policloradas (PCB), bisfenol, antibióticos, hormônios esteroides presentes em anticoncepcionais e outros produtos farmacêuticos e de cuidado pessoal (PPCP) (Carajaville, 2003; Bourguet, 2015; Vasilachi, 2021; Souza, 2016). Estes compostos podem desregular vias metabólicas humanas envolvidas na reprodução, desenvolvimento, morte celular, divisão celular, antioxidação e processos anticarcinogênicos (Choi, 2023).

Além disso, compostos sintéticos como a maioria dos desreguladores endócrinos, são muitas vezes projetados para possuir alta afinidade e interagir com receptores humanos, mimetizando a estrutura e efeito de compostos orgânicos encontrados não só em humanos. Assim, o risco potencial da presença desses contaminantes nos ambientes aquáticos é alto, visto que diversas espécies de mamíferos aquáticos têm esses locais como habitat e o efeito de sua presença ainda é pouco conhecido. Adicionalmente, estudos mostraram que esses compostos podem afetar espécies de invertebrados como insetos e moluscos, incluindo a ostra *Magallana gigas* (ostra-do-pacífico) (Miglioli, 2021; Reis, 2020; André, 2019; Soin, 2007;). Este cenário reforça a necessidade pela busca de biomarcadores que auxiliem no entendimento dos efeitos destes contaminantes, especialmente nos organismos aquáticos.

A Organização Mundial da Saúde define biomarcadores como qualquer medida que reflita a interação entre um sistema biológico e um agente ambiental, que pode ser químico, físico ou biológico (OMS, 1993; McCarty e Munkittrick, 1996). Ainda, biomarcadores podem pertencer a diferentes níveis organizacionais dentro de sistemas biológicos, como por exemplo enzimas,

estruturas ou processos fisiológicos, até populações e comunidades (OMS, 1993; McCarty e Munkittrick, 1996). Adicionalmente, os biomarcadores podem pertencer a três categorias: marcadores de exposição (a identificação de um xenobiótico em um tecido, por exemplo), efeito (alterações em processos bioquímicos, genéticos, fisiológicos ou comportamentais) ou susceptibilidade (indicadores da capacidade de um organismo responder ao contato com um xenobiótico) (Lionetto, 2012). Para avaliar os efeitos de poluentes ambientais, pode-se utilizar os biomarcadores de efeito, como alterações nos níveis de expressão de citocromos P450 ou de fatores de transcrição envolvidos no metabolismo de endo e xenobióticos, como os receptores nucleares (NR - *Nuclear Receptor*).

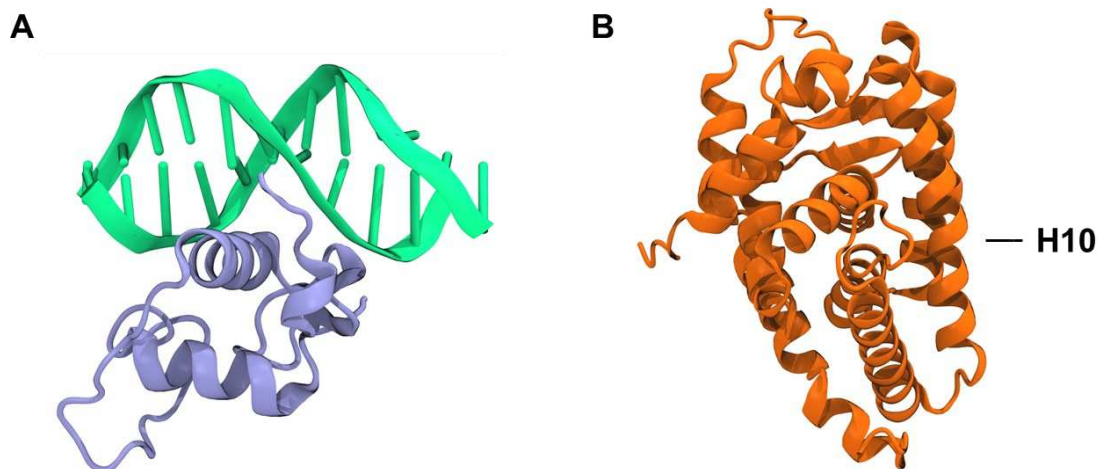
1.2 A SUPERFAMÍLIA DE RECEPTORES NUCLEARES

Os fatores de transcrição pertencentes à superfamília de receptores nucleares são exclusivos da linhagem animal, presentes tanto em *taxa* mais ancestrais como poríferos até grupos mais recentes como mamíferos e aves (Taubenheim, 2021; Bridgham, 2010), sendo divididos em sete subfamílias (NR0 a NR6). Em relação à sua função, os NR regulam a expressão de genes envolvidos em processos relacionados ao desenvolvimento embrionário, diferenciação celular, reprodução, apoptose, autofagia, metabolismo de lipídios, glicose, esteroides, colesteróis e xenobióticos (Taubenheim, 2021; Li, 2022). Esta regulação se dá, geralmente, pela interação com ligantes que causam o desacoplamento de correpressores e recrutamento de coativadores, ativando os NR. O contrário pode ocorrer com NR que possuem atividade inibitória, com o desacoplamento de coativadores e recrutamento de correpressores (Ortlund, 2018). Ambos os processos podem ocorrer tanto no citoplasma quanto no núcleo celular, porém, caso o NR encontre-se no citoplasma, este processo irá causar uma migração para o núcleo, onde o NR irá interagir com elementos de resposta específicos, modulando os genes-alvo.

A interação com elementos de resposta no DNA e com ligantes acontece por meio de duas regiões altamente conservadas presentes na maioria dos NR, respectivamente, a região C que contém o domínio de ligação ao DNA (DBD - *DNA Binding Domain*) e a região E, contendo o domínio de ligação ao ligante (LBD - *Ligand Binding Domain*). Existem ainda outras quatro regiões menos

conservadas presentes nos NR, são elas: as regiões A/B, contendo o domínio de ativação de função 1 (AF1) e tendo comprimento e estrutura variáveis; a região D, que atua como uma alça entre o DBD e LBD; e a região F, que nem sempre está presente e possui função desconhecida. O DBD é o domínio mais conservado dos NR tanto em termos de sequência, estrutura e função. Sua estrutura compreende duas alfa-hélices, duas folhas-beta, dois dedos de zinco tipo-IV que compõem um núcleo no centro do DBD e que interagem diretamente com o DNA, além das regiões P-box e D-box, que auxiliam na interação com elementos de resposta e na dimerização com outros NR, respectivamente (Ortlund, 2018) (**Figura 1**). Em alguns receptores, após o núcleo formado pelos dedos de zinco, há uma extensão C-terminal (CTE) que contém duas regiões denominadas A-box e T-box contendo resíduos positivos que favorecem adicionalmente a interação com elementos de resposta (Wilkum, 2018; Pawlak, 2012). De modo geral os elementos de resposta aos quais os NR se ligam possuem o motivo conservado “AGGTCA” (chamado também de meio-sítio ou *half-site*), que pode estar presente a) como um único sítio estendido reconhecido por monômeros de NR, ou dois sítios palíndromos em b) repetição direta (*direct repeat* - DR), c) em repetição invertida (*inverted repeat* - IR) ou d) em repetição evertida (*everted repeat* - ER) reconhecidos por dímeros (Ortlund, 2018; Pawlak, 2012). Além disso, estes sítios podem estar separados por espaçadores com comprimentos que variam de um a oito nucleotídeos, conferindo especificidade e estabilidade diferentes para cada dímero e aumentando o conjunto de interações entre NR e DNA (Pawlak, 2012).

Figura 1. Estrutura tridimensional dos domínios de ligação ao DNA e de ligação ao ligante.



Estruturas tridimensionais dos domínios conservados de receptores nucleares. A) DBD em azul e meio sítio de elemento de resposta, em verde. B) LBD com hélice 10 destacada.

Em sequência, o LBD é o segundo domínio conservado dos NR, sendo o principal responsável pela interação com a ampla gama de ligantes conhecidos destes receptores. Este domínio é menos conservado que o DBD na sua estrutura primária, possivelmente pela afinidade por diferentes ligantes, porém apresenta uma conformação estrutural bastante conservada, apresentando 12 alfa-hélices e 2 a 4 folhas-beta que se organizam em três camadas anti paralelas, formando uma estrutura em “sanduíche”. No centro desta estrutura encontra-se o sítio ativo dos NR, chamado LBP (*Ligand Binding Pocket*) que possui volume variável entre 300 a 1500 Å³. Entre os ligantes que interagem com os NR, pode-se citar ácidos retinóicos, ácidos graxos, sais biliares, hormônios tireoidianos, hormônios esteroides e xenobióticos orgânicos e sintéticos, como os desreguladores endócrinos. A hélice 12 (conhecida como domínio de ativação de função 2 - AF2) atua como uma interface para a entrada de ligantes no LBP, encontrando-se fechada quando o NR está inativo e se abrindo à medida que seu motivo recrutador de coativadores LxxLL/LLxxL torna-se exposto, permitindo a entrada do ligante (Li, 2021).

1.2.1 A dimerização dos receptores nucleares

Sabe-se que o processo de dimerização entre NR é mediado em grande parte pela hélice 10 do LBD, além de outras hélices acessórias como H7, H9, H11 e loops entre as hélices 8-9 e 9-10 (Li, 2021). A dimerização é uma etapa

essencial para o funcionamento dos NR (com exceção daqueles monoméricos) que pode ocorrer tanto no citoplasma quanto no núcleo celular. A partir do recrutamento de um NR parceiro, o dímero liga-se aos elementos de resposta mencionados anteriormente e ocorre a modulação da expressão gênica. Os dímeros podem ser homo ou heterodímeros, dependendo do subgrupo a que pertencerem. Por exemplo, a subfamília NR3 é caracterizada por agrupar receptores exclusivamente homodiméricos quando desempenhando funções genômicas, enquanto a maioria dos membros da subfamília NR1 forma heterodímeros com o receptor retinóide X (RXR - NR2B). De fato, a heterodimerização aparenta ser mais comum, embora estudos sugiram que este modo de interação seja derivado evolutivamente, visto que o próprio RXR, receptor mais conservado, pode formar homodímeros, além de outros membros da subfamília NR2 (mais antiga evolutivamente) como o fator de transcrição hepático 4 (HNF4) (Beinsteiner, 2023).

Curiosamente, o RXR é o único receptor nuclear conhecido que possui a habilidade de realizar ambas as formas de dimerização (Lefebvre, 2010). Além disso, por dimerizar com receptores das subfamílias NR1, NR2 e NR4, o RXR reconhece também múltiplos elementos de resposta, tornando-o um candidato potencial para o entendimento das interações entre NR, DNA e ligantes (Lefebvre, 2010). Adicionalmente, estudos apontam que este receptor surgiu cedo na linhagem animal, próximo ao surgimento dos animais bilatérios e mantendo sua sequência, estrutura e função conservadas ao longo do processo evolutivo (Beinsteiner, 2021). Ainda, sequências ortólogas deste receptor e de seus parceiros estão depositadas em bancos de dados públicos, facilitando o estudo comparativo entre espécies distantes, como humanos e invertebrados marinhos, por exemplo. Entre os parceiros de RXR que apresentam relevância para estudos ecotoxicológicos, pode-se citar o receptor de ácido retinoico (RAR - NR1B), o receptor ativado por proliferação de peroxissomos (PPAR - NR1C), o receptor de ecdisona (EcR - NR1H1) e o receptor pregnano X (PXR - NR1I2), que quando dimerizados com RXR regulam processos como neurogênese, metabolismo de lipídios, formação da ecdise e metabolismo de xenobióticos, respectivamente (Evans & Mangelsdorf, 2014; Gutierrez-Mariegos, 2014; Maletta, 2014; Lefebvre, 2010; Wells, 2009).

No entanto, a maior parte da literatura se concentra em receptores humanos e nas interações com seus ligantes conhecidos, gerando um panorama limitado sobre os modos de ação de contaminantes emergentes e seus impactos em receptores com ligantes desconhecidos. Em relação aos invertebrados e seus NR, o cenário é ainda mais desconhecido, visto que receptores ortólogos em invertebrados podem apresentar comportamentos distintos em termos de dimerização, interação com ligantes e vias metabólicas envolvidas. Os primeiros esforços direcionados à caracterização e entendimento da função de NR em invertebrados focaram em artrópodes, especialmente insetos, expandindo-se para espécies aquáticas como crustáceos e moluscos (Welter, 1996; Castro, 2007). Atualmente, estudos abordando a interação de contaminantes emergentes com NR de invertebrados estão disponíveis, como por exemplo a interação entre o tributilestanho (TBT) e o dímero RXR/PPAR em moluscos bivalves e gastrópodes. Esses efeitos alteram a homeostase lipídica, podendo causar o efeito imposex, que se caracteriza pela reversão de sexo do animal, com diversas implicações fisiológicas e ecológicas. É conhecido que pesticidas contendo ecdisteróides modulam o EcR em insetos e crustáceos, e que xenoestrógenos como bisfenol A interagem com o ER no gastrópode *Nucella lapillus* (Laranjeiro, 2016; Capitão, 2021; Fahrbach, 2012; Chan, 2019; Castro, 2007). Além disso, um estudo de nosso grupo de pesquisa demonstrou a modulação de genes que codificam os receptores NR0B, NR5A e NR1P10 na ostra *Magallana gigas* mediante exposição à benzofenona (Saldaña, 2025).

Diante desses dados, fica evidente o papel central do RXR como sensor da regulação de diversas vias metabólicas e na interação com ligantes endógenos e exógenos, como os desreguladores endócrinos, em invertebrados. Além disso, os dados de Saldaña e colaboradores sugerem mais um possível subgrupo parceiro de RXR em moluscos bivalves com função desconhecida. O subgrupo NR1P foi identificado primeiramente em 2010 na ostra *M. gigas* (Vogeler, 2014) e estudos recentes indicam que este subgrupo possa ser exclusivo de moluscos bivalves. Porém, os dados disponíveis sobre o subgrupo NR1P limitam-se à sua relação filogenética com outros NR e seu perfil de expressão ao longo do desenvolvimento de *M. gigas* (Miglioli, 2024; Vogeler, 2014).

1.3 A OSTRADO-PACÍFICO COMO MODELO DE ESTUDO

A ostra-do-pacífico (*Magallana gigas*) é uma espécie nativa do leste asiático e atualmente possui uma distribuição geográfica global, sendo a terceira espécie de bivalve mais cultivada na aquicultura global (FAO, 2022). No Brasil, o cultivo desta espécie está estabelecido, com o estado de Santa Catarina liderando a sua produção, tendo a cidade de Florianópolis como principal produtor (EPAGRI, 2022). Em relação ao seu modo de vida, trata-se de um organismo sésil e filtrador (ou seja, possui a capacidade de bioacumular diversas moléculas ingeridas, inclusive xenobióticos), sendo considerado uma espécie sentinela para estudos de ecotoxicologia e identificação de biomarcadores. Além disso, *M. gigas* possui oito genomas completamente sequenciados disponíveis publicamente, sendo o mais recente de 2023 (Mrowicki, 2025), viabilizando metodologias *in silico* e estudos moleculares. Portanto, devido à sua importância ecológica, científica, econômica e por ser um organismo estudado extensamente pelo nosso grupo de pesquisa, escolhemos *M. gigas* como objeto de estudo desta pesquisa.

Assim, com o intuito de elucidar o processo de dimerização do RXR em moluscos bivalves e identificar possíveis parceiros de dimerização que possam ser investigados em estudos experimentais futuros, buscamos identificar resíduos e interfaces relevantes para o processo de dimerização do RXR de humanos e *M. gigas* com receptores parceiros conhecidos. Buscamos também caracterizar o subgrupo NR1P em termos de sequência e estrutura, averiguar se esse subgrupo possui potencial de dimerizar com o RXR e se possui afinidade por ligantes endógenos e xenobióticos.

1.4 A BIOINFORMÁTICA ESTRUTURAL: MODELAGEM, ATRACAMENTO E DINÂMICA MOLECULAR

Em cenários cuja disponibilidade de dados acerca da existência, expressão, modulação e interação de genes, proteínas ou ligantes é escassa, como no estudo de invertebrados marinhos, pode-se recorrer a ferramentas de bioinformática para ampliar o escopo conhecido dessas características. Para analisar a extensa gama de dados biológicos, diversas abordagens *in silico* podem ser empregadas como a bioinformática estrutural, por exemplo (Paiva, 2022). Esta área da bioinformática traz consigo técnicas como modelagem estrutural, atracamento e dinâmica molecular, interações proteína-proteína,

predições de sítios ativos e ligações, análises de conservação de resíduos e domínios, etc (Paiva, 2022). Além disso, ferramentas baseadas em sequências como alinhamentos múltiplos (MSA) e análises filogenéticas também podem ser utilizadas em conjunto com dados estruturais, auxiliando na caracterização de novos receptores nucleares. De modo geral, as ferramentas utilizadas se baseiam em algoritmos que executam modelos matemáticos, químicos e físicos para analisar grandes quantidades de dados ou para prever características e interações entre moléculas.

Inicialmente, os estudos em bioinformática estrutural se baseiam na modelagem tridimensional de proteínas, lipídios, ácidos nucleicos e/ou ligantes, que pode ser realizada por meio de métodos que utilizam estruturas resolvidas experimentalmente como referência ou com a abordagem *ab initio* que utiliza funções de energia como a hipótese termodinâmica de Anfinsen e amostragens conformacionais para gerar uma estrutura tridimensional (Paiva, 2022). Atualmente, os métodos baseados em estruturas experimentais são os mais populares, principalmente devido à publicação do modelo AlphaFold 2 em 2021 (Jumper, 2021) que utiliza algoritmos de inteligência artificial como redes neurais e aprendizado profundo (*deep learning*) para criar alinhamentos e prever a posição de resíduos aminoácidos com base em dados de coevolução, até compor uma estrutura tridimensional. Adicionalmente, uma nova versão deste modelo foi anunciada em 2024, AlphaFold 3, trazendo melhorias como a integração do atracamento molecular com a modelagem de complexos proteicos, DNA, RNA, lipídios e ligantes, além de mudanças pós-traducionais e a criação de um servidor gratuito (Abramson, 2024). Além disso, ambas as versões iniciais do AlphaFold obtiveram os melhores resultados na competição CASP (*Critical Assessment of protein Structure Prediction*) e as contribuições deste modelo resultaram em um Prêmio Nobel de Química em 2024 (Callaway, 2024). Assim, com a possibilidade de gerar estruturas tridimensionais com eficiência similar a métodos experimentais em poucos minutos, foi criado um banco de dados com cerca de 214 milhões de estruturas depositadas (disponível em: <https://alphafold.ebi.ac.uk/>). Este banco de dados está atualmente integrado em diversas outras plataformas que rapidamente acessam e utilizam essas estruturas para executar outros tipos de análises, como o atracamento molecular e predição de sítios ativos, por exemplo.

Novidades como modelos baseados em aprendizado automático facilitam o estudo de proteínas que ainda não estão completamente caracterizadas, como os receptores nucleares de invertebrados. A aplicação de métodos baseados em aprendizado automático se estende para a predição de sítios ativos, auxiliando as análises de atracamento molecular posteriores. O atracamento (ou *docking*) é uma técnica que visa identificar ligações entre uma proteína e um ligante, gerando poses do ligante que são ranqueadas de acordo com uma pontuação refletindo a quantidade de energia livre do sistema, específica para cada algoritmo. As poses dos ligantes podem ser identificadas de forma exaustiva, explorando todas as interfaces possíveis da proteína, ou de forma estocástica, gerando pontos de partida aleatórios para a busca. Métodos que utilizam ambas as abordagens simultaneamente existem atualmente, demonstrando uma performance superior aos métodos de abordagem única. Para ranquear as poses encontradas, são calculadas interações como ligações de hidrogênio, interações iônicas, contatos hidrofóbicos e π -*stacking*, as quais são consideradas na equação de soma das energias de Van der Waals, Coulomb, polares e não-polares (Röhrig, 2023; Bugnon, 2024). De forma similar, campos de força podem ser utilizados para calcular as energias das diferentes poses de um ligante, gerando um ranqueamento baseado nas diferentes energias do sistema analisado (Röhrig, 2023).

Atualmente, a aplicação de campos de força vem crescendo, estando presente não só nos algoritmos de atracamento, mas também no próprio AlphaFold. Porém, embora venha se popularizando, o uso de campos de força ainda é mais comum em estudos de dinâmica molecular (MD). O primeiro registro dessa técnica na bioinformática estrutural foi documentado em 1977 (McCammon & Karplus), quando o dobramento do inibidor da tripsina bovina foi simulado, iniciando o uso da dinâmica molecular nesta área. Essa técnica baseia-se na simulação de todos os átomos de uma molécula e suas interações com outros átomos ao longo do tempo, sujeitos a um campo de força, tornando possível analisar processos temporais como a entrada de ligantes em um sítio ativo, dobramento protéico, mudanças conformacionais, dimerização etc.

Um campo de força pode ser definido como a junção de funções matemáticas baseadas em dinâmica newtoniana que regem o comportamento dos átomos ao longo do tempo e espaço e parâmetros que definem

características geométricas e energéticas das ligações interatômicas (Guvench e MacKerell, 2008). De modo geral, todos os campos de força disponíveis atualmente seguem as seguintes equações para o cálculo das forças atuantes no sistema:

$$E_{covalente} = \sum_{\text{ligações}} K_b (b - b_0)^2 + \sum_{\text{ângulos}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{diedros}} K_\chi [1 + \cos(n\chi - \sigma)] \quad (1)$$

$$E_{sem-ligação} = \sum_{\text{pares não-ligados}} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 * \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right) \quad (2)$$

A Equação 1 refere-se à soma das energias das interações de ligação (*bonded interactions*), como a distância entre as ligações entre átomos ($\sum_{\text{ligações}}$), dobramentos de ângulos entre átomos ($\sum_{\text{ângulos}}$) e rotações dos ângulos dos diedros ($\sum_{\text{diedros}}$). Em sequência, a Equação 2 indica a soma das energias das interações não-covalentes por meio da junção das equações de Lennard-Jones (entre colchetes) e a Lei de Coulomb (termo final), que calculam as forças de atração/repulsão e as interações eletrostáticas entre átomos, respectivamente. Ainda, outras energias podem ser calculadas a depender das características específicas de cada campo de força, como o tratamento de diedros impróprios e de átomos com mesmo número atômico em diferentes contextos químicos, resultando na Equação 3.

$$E_{total} = E_{covalente} + E_{sem-ligação} + E_{outros} \quad (3)$$

Dessa forma, os cálculos acima são executados a cada passo da simulação de MD, gerando uma trajetória tridimensional com tempo definido para cada átomo do sistema. Além disso, antes da simulação iniciar de fato, etapas como a criação da própria caixa de simulação, solvatação e ionização do sistema, minimização de energia, e equilíbrio da temperatura e pressão são necessárias para relaxar a estrutura analisada e reduzir choques estéricos. Ainda, é possível customizar diversos parâmetros como o tipo de termostato e barostato utilizados, valores de compressibilidade, integradores de vetores, além de impor restrições de posição em átomos específicos. Porém, calcular todas estas interações repetidamente exige um poder computacional robusto, mesmo com o avanço constante de processadores computacionais e gráficos, fazendo com que as simulações de MD ocorram por poucos nanossegundos (Orozco, 2011; Gelpi, 2015).

Para contornar este problema, foram criados sistemas simplificados, chamados *coarse-grained* (granulação grossa). Nestes sistemas, a topologia das moléculas é alterada, substituindo grupos de átomos por esferas (*beads*) com características distintas, como tamanho e polaridade. Para simular estas estruturas modificadas, foram criados campos de força específicos, como o MARTINI (Marrink, 2007), que seguem as forças descritas nas Equações 1, 2 e 3, porém adaptado a modelos *coarse-grained*. Dessa forma, o custo computacional é drasticamente reduzido e a performance aumenta em até 1000 vezes, visto que o número de cálculos realizados diminui, tornando possível realizar simulações mais longas e observar processos ao longo de microssegundos, mantendo as propriedades físico-químicas das moléculas biológicas (Orozco, 2011). Essa abordagem é ideal para analisar processos em que a resolução atômica não tão é necessária e que ocorrem em escalas de tempo maiores, como interações entre um complexo protéico, a auto-montagem de uma membrana lipídica ou para realizar *screening* virtual de ligantes ou parceiros de dimerização. Visto que entre os objetivos deste trabalho estão o entendimento dos processos de dimerização de receptores nucleares e a identificação de parceiros de RXR, escolhemos utilizar simulações *coarse-grained* para fazer réplicas das simulações em escalas temporais maiores. Acreditamos que o entendimento de como se dão os processos de dimerização de receptores nucleares pode ajudar na identificação de novos alvos para regulação de xenobióticos em ambientes aquáticos, contribuindo para futuros estudos em organismos não modelo para a biologia estrutural.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Identificar interfaces de interação proteína-proteína e resíduos ativos no processo de dimerização de receptores nucleares de *Homo sapiens* e da ostra *Magallana gigas*.

2.2 OBJETIVOS ESPECÍFICOS

- Gerar modelos tridimensionais de dímeros de RXR e demais receptores nucleares de *H. sapiens* e *M. gigas*;

- Identificar interfaces de interação e resíduos de aminoácidos envolvidos em processos de dimerização entre dímeros de ambas as espécies via dinâmica molecular;
- Investigar possíveis dímeros entre o RXR e membros do grupo NR1P de *M. gigas*.

3. METODOLOGIA

3.1 OBTENÇÃO DE SEQUÊNCIAS

A sequência do RXR-gamma humano (Código de acesso Ensembl: ENSG00000143171) foi utilizada como entrada no banco de dados NCBI e BLAST (Camacho, 2009) para recuperar demais sequências de receptores nucleares humanos. Para obter os códigos das sequências de ostra, foi utilizada a listagem feita por Vogeler e colaboradores (2014). As sequências depositadas nesses códigos, pertencentes a um genoma mais antigo (Zhang, 2012), foram utilizadas como entrada na plataforma BLASTp contra o genoma mais recente de *M. gigas* (Mrowicki, 2025). O melhor resultado para cada receptor foi anotado, não havendo sequências redundantes.

3.2 IDENTIFICAÇÃO DE DOMÍNIOS E RESÍDUOS CONSERVADOS

A identificação das regiões de início e fim dos domínios LBD e DBD foi feita com o auxílio da plataforma ExPASy PROSITE (Sigrist, 2009). Para identificar resíduos conservados dentro dos domínios, foram feitos alinhamentos utilizando o conjunto de receptores de *H. sapiens* e *M. gigas* separadamente com o algoritmo MUSCLE. Por fim, para identificar a posição das alfa-hélices e folhas-beta foram utilizadas estruturas tridimensionais resolvidas experimentalmente depositadas no PDB como referência (código de acesso: 3DZY, 4NQA e 5UAN).

3.3 MODELAGEM TRIDIMENSIONAL

Inicialmente foram utilizados os programas I-TASSER e HADDOCK para modelagem de monômeros e atracamento das cadeias, respectivamente (Roy, 2010; Dominguez, 2003). A fim de comparação, utilizamos os programas AlphaFold2 e AlphaFold Multimer (Jumper, 2021; Evans, 2022) para modelagem de monômeros e dímeros de RXR, que se mostraram mais rápidos e mais precisos que os programas anteriores. Sendo assim, optamos por utilizar o

modelo AlphaFold 2 nas análises subsequentes. Este modelo trouxe novidades na forma de conduzir o processo de modelagem tridimensional ao incorporar redes neurais de aprendizado profundo com dados de co-evolução. O programa opera da seguinte maneira: i) uma sequência inicial é inserida pelo usuário; ii) um alinhamento múltiplo de sequências em larga escala é criado, junto com uma matriz deste alinhamento; iii) estruturas experimentais são escolhidas em bancos de dados como moldes com base em homologia de sequência; iv) uma representação do MSA e outra contendo informações de pares de resíduos como distância e orientação são criadas; v) estas representações são utilizadas como entrada no Evoformer, onde cada linha e coluna de cada matriz será analisada 48 vezes, havendo troca de informação entre as matrizes para inferir relações evolutivas e espaciais; vi) os resultados do Evoformer são inseridos como entrada no bloco estrutural que se repete 8 vezes, utilizando a representação em pares refinada para inferir uma estrutura tridimensional, iniciando pelos carbonos-alfa (*backbone*); vii) todo este processo é repetido mais duas vezes, terminando com a posição das cadeias laterais e um relaxamento da estrutura com o campo de força Amber; viii) métricas de qualidade como o teste de diferença de distância local prevista (*predicted local-distance difference test*, pLDDT), TM-Score predito (*predicted template modeling score*, pTM; Zhang, 2004) e TM-Score predito para interfaces de complexos protéicos (*interface predicted template modeling*, ipTM).

Posteriormente, com o lançamento do algoritmo atualizado do programa AlphaFold 3 (Abramson, 2024) com performance melhorada e modelagem de complexos protéicos integrada com ácidos nucleicos, os modelos preexistentes foram refeitos. Além disso, os blocos principais da rede neural sofreram alterações com esta nova versão, com a introdução do módulo Pairformer no lugar do Evoformer, reduzindo o peso do MSA para inferência da estrutura tridimensional e dando maior peso para a representação em pares. O bloco estrutural também sofreu alterações, sendo substituído por um módulo de difusão que opera com base em dados brutos. O número de iterações e reciclagens também foi reduzido, acelerando os cálculos. As métricas utilizadas para gerar pontuações e gráficos finais permaneceram as mesmas.

Ao todo, foram feitos modelos para os 48 receptores nucleares de *H. sapiens* em complexo com o HsRXR- α e para os 43 receptores de *M. gigas*

compreendendo os domínios conservados e a alça entre domínios em complexo com *MgRXR*. Adicionalmente, foram feitos modelos dos LBD de 11 receptores do subgrupo NR1P para as análises de identificação de sítios ativos e atracamento de ligantes. Por fim, foram feitos modelos para dímeros de RXR ligados ao DNA com estrutura experimental depositada no PDB (código PDB: 3DZY, 4NQA e 5UAN) para simulações de dinâmica molecular.

3.4 IDENTIFICAÇÃO DE SÍTIOS ATIVOS E ATRACAMENTO DE LIGANTES

A predição de possíveis sítios de ligação de ligantes foi feita com o auxílio do método P2Rank que utiliza dados de conservação de resíduos através de alinhamentos de sequências junto com dados estruturais e físico-químicos para identificar possíveis pontos de ligação na estrutura de uma proteína (Krivák, 2018). Os bancos de dados acessados para realizar os alinhamentos são o UniRef50 (UniProfReference Cluster/EMBL-EBI; Suzek, 2007) e o AlphaFold Database, ambos integrados na plataforma online PrankWeb. Assim, os sítios que obtiveram melhor resultado de acordo com as pontuações de probabilidade de ocorrência e conservação foram escolhidos para a etapa de atracamento de ligantes.

O atracamento foi realizado utilizando o método *AttractingCavities 2.0* disponível na plataforma SwissDock (Röhrig, 2023). Este método é similar ao P2Rank no sentido de criar pontos representando possíveis locais de interação entre a proteína e ligantes e sobrepor esses pontos na estrutura tridimensional. Além disso, é utilizado o campo de força CHARMM em conjunto com o modelo FACTS (*Fast Analytical Continuum Treatment of Solvation*) para calcular tanto a energia livre do sistema proteína e ligante como entre esse sistema e o solvente, respectivamente. Os receptores nucleares escolhidos foram: hsRXR-alpha, hsRAR-alpha, hsPPAR-gamma, hsER-beta, cgRXR-alpha, cgRAR-alpha, cgPPAR-alpha, cgER e o subgrupo NR1P. Por fim, os ligantes escolhidos foram o ácido *all-trans* retinóico, ácido araquidônico, estrogênio e benzofenona.

3.5 DINÂMICA MOLECULAR

Foram feitas simulações de dinâmica molecular utilizando modelos *coarse-grained*, ou seja, modelos simplificados que substituem conjuntos de átomos por

esferas, diminuindo a resolução da técnica, mas aumentando consideravelmente o tempo de simulação e diminuindo o custo computacional. Para transformar as estruturas atomísticas de proteínas em *coarse-grained* foi utilizado o código “martinize 2” desenvolvido por X et al, enquanto que estruturas atomísticas de DNA foram convertidas em *coarse-grained* com auxílio do Laboratório de Modelação Multiescala (*Multiscaling Modeling Lab* - ITQB/NOVA) da Universidade de Lisboa. O campo de força Martini 3 (Souza, 2021) foi escolhido para as simulações por sua compatibilidade com modelos *coarse-grained* e por ser uma versão mais recente com parâmetros atualizados como esferas novas e interações de pontes de hidrogênio, garantindo maior flexibilidade para as estruturas submetidas ao campo de força e facilitando a interação entre proteínas. Adicionalmente, o pacote GROMACS v.2022 (Abraham, 2015) foi escolhido por proporcionar integração com o campo de força Martini 3 e possibilitar o uso de unidades de processamento computacional e gráficas simultaneamente.

Todas as simulações passaram por etapas de conversão da estrutura atomística para *coarse-grained*, solvatação, neutralização do sistema com adição de íons, minimização de energia do sistema, equilibração de temperatura e pressão, e produção da simulação. O modelo de restrição de estrutura secundária utilizado para construir os modelos *coarse-grained* de proteínas foi o “*Elastic networks*” (redes elásticas). Em sequência, o formato de caixa de simulação escolhido foi o dodecaedro, por possuir volume menor do que uma caixa cúbica, e conseqüentemente, requerer menos moléculas de solvente e diminuindo o tempo de computação. Os integradores escolhidos foram “*steepdescent*” (gradiente descendente) para a minimização e “*stochastic dynamics*” (dinâmica estocástica) para a equilibração e produção. Para estas duas últimas etapas foram utilizados o barostato “Parrinello-Rahman” e o termostato “*v-rescale*”, com um intervalo de integração (dt) de 0.02 picossegundos. Foram feitas duas etapas adicionais de equilibração com intervalos de integração menores (0,005 e 0,01) para o sistema RXR/RAR humano devido a constantes choques estéricos causados pelo resíduo L219 do RAR. Este resíduo teve sua estrutura secundária alterada de hélice para *loop*. Parâmetros adicionais das simulações estão detalhados no **Apêndice A**.

Cada simulação foi feita em triplicata, com 10 microssegundos de duração cada. Além disso, restrições de posição foram impostas em resíduos específicos de acordo com os mapas de densidade de elétrons disponíveis para estruturas experimentais depositadas no PDB. Em caso de indisponibilidade desses dados, as restrições foram impostas nas extremidades da alça entre os domínios DBD e LBD ou no DBD inteiro para simular o atracamento com o DNA. Tanto a imposição dessas restrições quanto as análises subsequentes das simulações foram feitas utilizando códigos e bibliotecas em Python 3 e *bash* (Michaud-Agrawal, 2011).

3.6 IDENTIFICAÇÃO DE INTERAÇÕES ENTRE PROTEÍNAS

O programa AlphaBridge (Álvarez-Salmoral, 2024) foi utilizado para detectar interações entre as cadeias de dímeros de RXR gerados pelo AlphaFold 3. Este programa utiliza as matrizes geradas pelo AlphaFold 3 para identificar contatos entre proteínas e gerar representações gráficas, além de integrar as matrizes PAE e pLDDT em uma nova matriz utilizada para identificar regiões ativas na estrutura dos modelos. Além disso, o programa possibilita a interação com a matriz PDE, permitindo ao usuário utilizar valores distintos de *cut-off* para o mascaramento da matriz, com valores menores tornando a busca mais ampla e menos criteriosa e vice-versa. Para este estudo, foi utilizado o valor padrão de *cut-off* de 0,7, dado que valores maiores omitiam contatos conhecidos em estruturas experimentais utilizadas como controle como o dímero de RXR α -PPAR γ humano.

Para detectar contatos entre cadeias protéicas nas análises de dinâmica molecular, foram feitos códigos em Python 3 utilizando bibliotecas como MDAnalysis, pandas, matplotlib e numpy (Michaud-Agrawal, 2011; McKinney, 2011; Oliphant, 2006; Hunter, 2007). Por analisar os contatos ao longo do tempo, esses códigos permitem observar flutuações nas distâncias entre as cadeias, o que não é possível ao detectar contatos entre modelos estáticos. Porém, apesar de ser mais sensível do que a análise feita com o AlphaBridge, a necessidade de haver uma análise de dinâmica molecular prévia torna esta abordagem mais custosa computacional e temporalmente. Assim, buscamos integrar ambas as abordagens para àqueles sistemas que foram submetidos à dinâmica molecular.

3.7 ANÁLISES FILOGENÉTICAS

Uma filogenia dos receptores de *M. gigas* foi feita utilizando sequências alinhadas contendo ambos os domínios DBD e LBD. O melhor modelo de substituição para o conjunto de dados foi identificado utilizando o programa IQTREE-2 (Minh, 2020), resultando no modelo LG (Le e Gascuel, 2008), com uma taxa de substituição definida em 6 e distribuição Gamma. Utilizando o programa MrBayes (Ronquist, 2012), a análise filogenética foi feita por meio de inferência bayesiana e os parâmetros foram definidos da seguinte forma: *prsetaamodelpr = lglsetnst = 6 rates = gamma mcmcncchains = 1 ngen = 5000000 samplefreq = 1000 diagnfreq = 1000 printfreq = 1000*, ou seja, uma corrente de Markov com 5.000.000 de gerações, sendo amostradas a cada 1.000. O *burn-in* foi definido no valor padrão de 25%. A árvore filogenética resultante foi visualizada e vetorizada pelo programa FigTree (Rambaut, 2007).

4. RESULTADOS E DISCUSSÃO

4.1 SEQUÊNCIAS DE NR IDENTIFICADAS NA OSTRADO-PACÍFICO

A busca por sequências identificou os 48 receptores nucleares humanos e 43 de ostra-do-pacífico (*Magallana gigas*), englobando todos os subgrupos conhecidos (**Quadro 1**). A nomenclatura utilizada para identificação das sequências encontradas de *M. gigas* foi àquela de Vogeler et al. (2014), dado que a anotação do genoma mais recente (Mrowicki, 2025) mostrou-se imprecisa, classificando subgrupos ainda pouco caracterizados como NR1J e NR1P como pertencendo a NR1F e NR1I1 (ROR e VDR, respectivamente). Após recuperar as sequências de NR de ambas as espécies, a identificação dos domínios conservados foi feita, indicando a presença do DBD e LBD em todas as sequências, exceto aquelas que possuem apenas um domínio como o subgrupo NR0, sugerindo boa qualidade das sequências recuperadas. Ainda, algumas sequências de *M. gigas* sem subgrupo definido também tiveram seus domínios identificados, como NHR40 (apenas DBD), NHR41 (apenas LBD), NHR42, NHR43 e 2DBDNr (contendo dois DBD). Um novo conjunto de dados contendo as sequências filtradas (contendo apenas os domínios DBD, LBD e alça entre domínios) foi criado para as análises de modelagem.

Quadro 1. Receptores nucleares de *Homo sapiens* e *Magallana gigas* e seus códigos de acesso.

Grupo	Subgrupo	NR	<i>Homo sapiens</i>	Código de acesso	<i>Magallana gigas</i>	Código de acesso
NR1	NR1A	A1	TR alpha	ADZ17370	MgNR1A	XP_011443300
		A2	TR beta	KAI4028673		
	NR1B	B1	RAR alpha	P10276	MgNR1B	XP_065933413
		B2	RAR beta	Q9W6B3		
		B3	RAR gamma	P13631		
	NR1C	C1	PPAR alpha	Q07869	MgNR1C	XP_011412920
		C2	PPAR beta	Q03181	MgNR1CDEFa	XP_034306943
		C3	PPAR gamma	P37231	MgNR1CDEFb	XP_011445721
	NR1D	D1	Rev-erb alpha	P20393	MgNR1D	XP_034303848
		D2	Rev-erb beta	Q14995		
	NR1E	E1	-	-	MgNR1E	XP_011419905
	NR1F	F1	ROR alpha	P35398	MgNR1F	XP_034321282
		F2	ROR beta	Q92753		
		F3	ROR gamma	P51449		
	NR1H	H1	-	-	MgNR1H	XP_011416718
		H2	LXR beta	P55055	-	-
		H3	LXR alpha	Q13133	-	-
		H4	FXR	Q96RI1	-	-
	NR1I	I1	VDR	P11473	-	-
		I2	PXR	O75469	-	-
		I3	CAR	Q14994	-	-
	NR1J	J1	-	-	MgNR1Ja	XP_011436006
		J2	-	-	MgNR1Jb	XP_034327479
		J3	-	-	MgNR1Jc	XP_011454352
	NR1P	P1	-	-	MgNR1P1	XP_065940816
		P2	-	-	MgNR1P2	XP_065943509
		P3	-	-	MgNR1P3	XP_011431575
		P4	-	-	MgNR1P4	XP_011439772
		P5	-	-	MgNR1P5	XP_011431105
		P6	-	-	MgNR1P6	XP_065937743
		P7	-	-	MgNR1P7	XP_011431570
		P8	-	-	MgNR1P8	XP_011431576

		P9	-	-	MgNR1P9	XP_065940820
		P10	-	-	MgNR1P10	XP_011431625
		P11	-	-	MgNR1P11	XP_011431632
NR2	NR2A	A1	HNF-4 Alpha	P41235	MgNR2A	XP_011451244
		A2	HNF-4 Beta	Q14541		
	NR2B	B1	RXR Alpha	P19793	MgNR2B	XP_011434498
		B2	RXR Beta	XP_054212168		
		B3	RXR Gamma	P48443		
	NR2C	C1	TR2	P13056	MgNR2CD	XP_065937975
		C2	TR4	P49116		
	NR2E	E1	TLX	Q9Y466	MgNR2E1	XP_011438581
		E2	-	-	MgNR2E2	XP_011438583
		E3	PNR	Q9Y5X4	MgNR2E3	XP_011416690
		E5	-	-	MgNR2E5	XP_011433241
	NR2F	F1	COUP- TF1	P10589	MgNR2F	XP_011456035
		F2	COUP- TF2	P24468		
		F6	EAR-2	P10588		
NR3	NR3A	A1	ER Alpha	P03372	MgNR3A	XP_065942990
		A2	ER Beta	Q92731		
	NR3B	B1	ERR Alpha	P11474	MgNR3B	XP_011414547
		B2	ERR Beta	O95718		
		B3	ERR Gamma	P62508		
	NR3C	C1	GR	P04150	-	-
		C2	MR	P08235	-	-
		C3	PR	P06401	-	-
		C4	AR	P10275	-	-
NR4	NR4A	A1	NUR77	P22736	MgNR4A	XP_065932343
		A2	NURR1	P43354		
		A3	NOR-1	Q92570		
NR5	NR5A	A1	STF-1	Q13285	MgNR5A	XP_065930513
		A2	LRH-1	O00482		
	NR5B	B1	-	-	MgNR5B	XP_019918672
NR6	NR6A	A1	GCNF	Q15406	-	-
NR0	NR0B	B1	DAX-1	P51843	MgNR0B	XP_011442427
		B2	SHP	Q15466		
Outros	2DBDNR		-	-	2DBD Gamma	XP_019919867
			-	-	2DBD Delta	XP_011428802

Apenas DBD	-	-	NHR40	AHV90298
Apenas LBD	-	-	NHR41	XP_065930999
NC*	-	-	NHR42	XP_011436094
	-	-	NHR43	XP_034336229

*NC = Não caracterizados.

Ambos os domínios, DBD e LBD, possuem sequência e estrutura altamente conservadas em comparação com os outros domínios de NR, podendo ser utilizados independentemente para análises de dinâmica molecular e filogenéticas, por exemplo (Brascher, 2024; Bortoli, 2022). O LBD é tido como o principal agente no processo de dimerização de NR, mais especificamente devido à sua 10^a alfa-hélice (H10), tida por alguns autores como interface “canônica” de dimerização entre NR (Beinsteiner 2021, Pawlak 2012, Brelivet, 2004). Outros trabalhos indicam a presença de uma região chamada “I-box” compreendendo a H10 e H11, com resíduos conservados ativos no processo de dimerização. Porém, neste trabalho optamos por utilizar os domínios conservados e a alça entre domínios a fim de tentar identificar interfaces de interação além do LBD e H10.

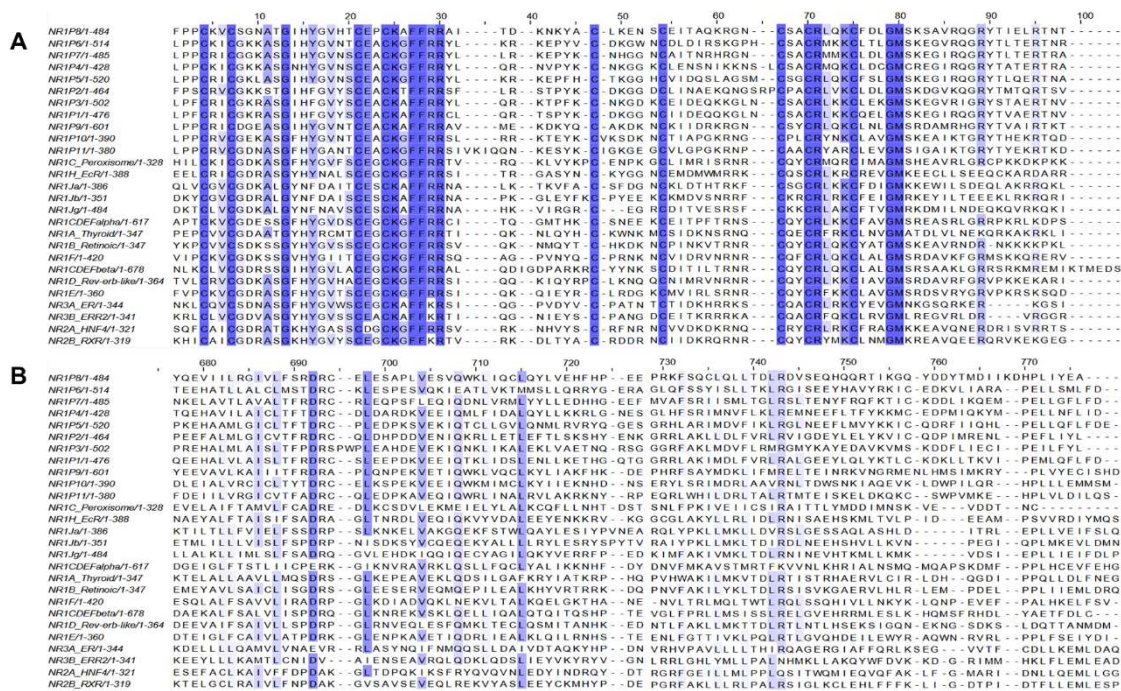
Em seguida, foram feitos alinhamentos múltiplos de sequência utilizando as sequências parálogas filtradas de cada espécie. Ainda, as posições dos elementos de estrutura secundária como alfa-hélices e folhas-beta foram determinadas utilizando a anotação de sequências do PDB como referência. Como esperado, o DBD apresentou alta similaridade de sequência entre diferentes NR da mesma espécie. Em relação ao LBD, foi observada uma maior diversidade de sequências, porém ainda foi possível identificar resíduos conservados, inclusive na região I-box, reforçando a importância desta região ao longo do processo evolutivo. A região da alça apresentou menor similaridade entre receptores parálogos, com sequência e comprimento altamente variáveis, visto que é uma região intrinsecamente desordenada (Choi, 2023). Assim, os resultados dos alinhamentos ilustram a conservação dos domínios envolvidos na dimerização e indicam regiões e resíduos conservados que possam estar envolvidos neste processo, inclusive no subgrupo NR1P de *M. gigas*, como o DBD e a I-box no LBD.

Ao todo foram recuperadas as sequências de 11 receptores nucleares do subgrupo NR1P, contendo todas as regiões e domínios conservados. As sequências variaram entre 488 e 721 resíduos de aminoácidos (**Tabela 1**), levando em consideração as sequências não filtradas. De modo geral, o DBD deste subgrupo é bastante conservado, apresentando bastante similaridade com os demais NR, seguido pelo LBD, que apesar de menos homogêneo, ainda possui uma identidade conservada entre o subgrupo e com os demais receptores membros da subfamília NR1 (**Figura 2**). É possível identificar resíduos conservados na região da I-box dos receptores do subgrupo NR1P como um resíduo de prolina na H12, arginina e leucina na H10, além de resíduos de aspartato e arginina no loop entre H8 e H9 que se situam próximos à H10, similar ao observado em receptores humanos, sugerindo conservação da propriedade de dimerização com RXR. Em contrapartida, a região da alça entre domínios apresentou sequência bastante heterogênea, tanto em relação ao seu tamanho quanto à identidade dos resíduos que as compõe, apresentando comprimento variável e estrutura tridimensional predita altamente desordenada para a maioria dos membros de NR1P, com exceção de NR1P10 e NR1P11.

Tabela 1. Comprimento das sequências de receptores nucleares do subgrupo NR1P de *M. gigas*.

	Sequência completa	Sequência filtrada
NR1P1	548	476
NR1P2	488	464
NR1P3	537	502
NR1P4	507	428
NR1P5	617	520
NR1P6	552	514
NR1P7	529	485
NR1P8	529	484
NR1P9	721	601
NR1P10	513	390
NR1P11	658	380

Figura 2. Alinhamentos múltiplos de sequência entre receptores nucleares das subfamílias NR1, NR2 e NR3 de *M. gigas*.



Alinhamentos de sequências filtradas de receptores nucleares de *M. gigas*. A) Sequências de DBD. B) Sequências de LBD. Resíduos aminoácidos estão destacados em azul de acordo com a taxa de conservação entre as demais sequências, tons mais fortes refletem resíduos mais conservados e vice-versa. Fonte: elaborado pelo autor (2025).

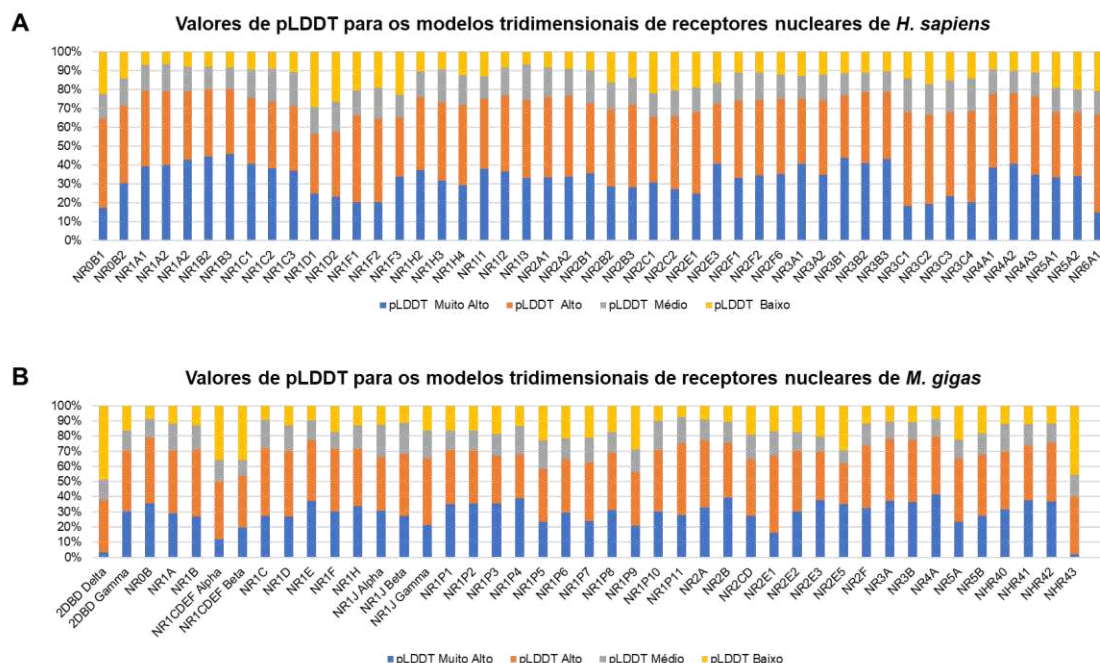
4.2 MODELOS TRIDIMENSIONAIS DE RECEPTORES NUCLEARES

Modelos tridimensionais foram gerados para todos os receptores nucleares de ambas as espécies utilizando as sequências filtradas como entrada no sistema AlphaFold 3. Todos os receptores foram modelados em complexo com o RXR, com o intuito de validar dímeros conhecidos, averiguar se o sistema AlphaFold 3 conseguiria modelar controles negativos com sucesso e identificar possíveis dímeros desconhecidos. Para averiguar a qualidade dos modelos, observamos tanto as pontuações resultantes das matrizes do AlphaFold 3 quanto a formação de dímeros simétricos.

De modo geral, a qualidade dos modelos foi satisfatória para ambas as espécies, com valores de pTM (*predicted template modeling*) e ipTM (*interface predicted template modeling*) variando entre 0,4 e 0,76 e de 0,15 e 0,8 para *H. sapiens* e de 0,32 e 0,74 e de 0,21 e 0,82 para *M. gigas*, respectivamente (**Figura 3; Apêndices B e C**). Estas métricas medem a qualidade geral das estruturas modeladas utilizando a pontuação *Template Modeling* (TM; Zhang e Skolnick,

2004; Xu e Zhang, 2010), com o pTM refletindo a qualidade de cadeias únicas, enquanto o ipTM tenta refletir a posição relativa das cadeias dentro de um complexo, sendo utilizado apenas na modelagem de dímeros ou oligômeros. Adicionalmente, de acordo com Jumper (2024), valores de pTM acima de 0,5 indicam uma estrutura similar à estrutura verdadeira, enquanto valores de ipTM acima de 0,8 indicam alta confiabilidade no modelo e valores entre 0,6 e 0,8 indicam uma “área cinza”, fazendo necessária a interpretação de um pesquisador. Os valores mais baixos estão presentes nos dados referentes a modelos que não formaram dímeros ou com regiões desordenadas. Como esperado, os modelos de receptores humanos obtiveram métricas mais altas devido à maior disponibilidade de dados disponíveis nos bancos de dados, inclusive àqueles acessados pelo AlphaFold 3 na sua busca por *templates*. Além disso, *Homo sapiens* é a única espécie com estruturas experimentais de receptores nucleares contendo ambos os domínios conservados depositadas no PDB.

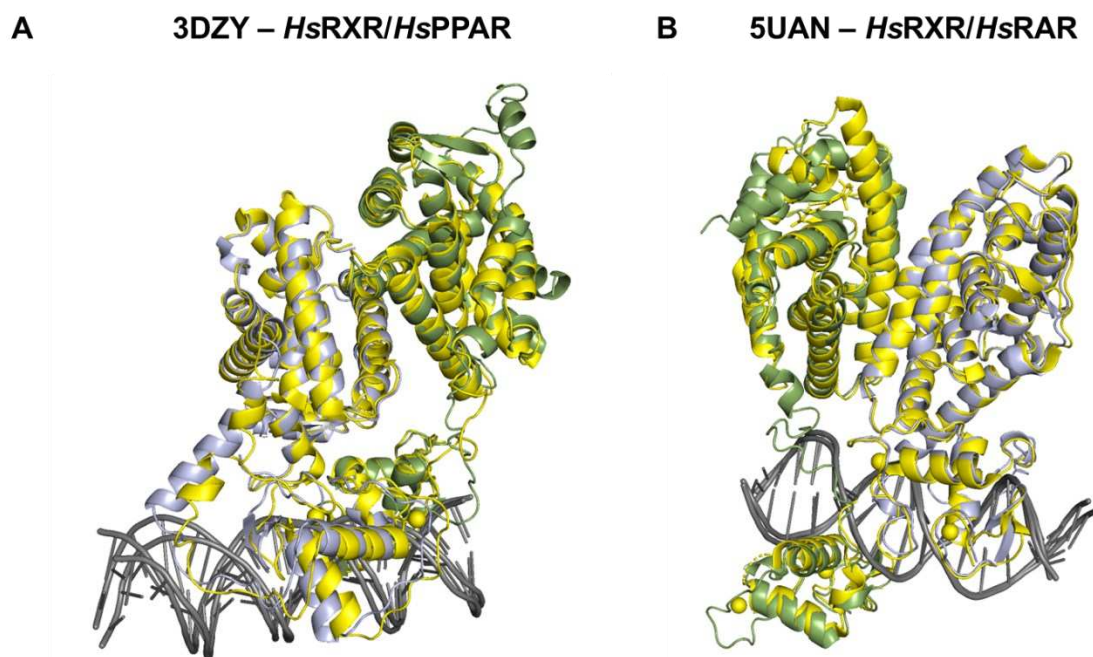
Figura 3. Valores de pLDDT para os modelos tridimensionais de dímeros de receptores nucleares de *H. sapiens* e *M. gigas*.



Qualidade da modelagem tridimensional dos modelos tridimensionais de dímeros de receptores nucleares. A) Receptores nucleares de *H. sapiens* em complexo com *HsRXR*. B) Receptores nucleares de *M. gigas* em complexo com *MgRXR*. Fonte: elaborado pelo autor (2025).

Entre estas estruturas, temos os sistemas RXR/PPAR e RXR/RAR, ambas com o seu respectivo elemento de resposta, ligantes e fragmento do coativador. Assim, esses foram os primeiros dímeros a serem modelados a fim de verificar a precisão em gerar modelos tridimensionais confiáveis de proteínas e de DNA. Os resultados indicam alta similaridade entre os modelos gerados para estes sistemas e suas estruturas experimentais, com valores de RMSD abaixo de 1 angstrom (0,1 nm) para RXR/PPAR e RXR/RAR (**Figura 4**), reforçando a precisão do AlphaFold 3. Desta forma, os sistemas RXR/PPAR e RXR/RAR foram escolhidos como controles positivos para a modelagem dos seus respectivos receptores ortólogos de *M. gigas* e posterior etapa de dinâmica molecular. Em relação à modelagem de controles negativos, foram escolhidos os modelos dos dímeros de RXR com GR e MR (Receptor Glicocorticóide e Receptor Mineralocorticóide), ambos da subfamília NR3C. A subfamília NR3 contém receptores de hormônios esteróides e acredita-se que estes realizam apenas homodimerização, não havendo registros até o momento de interações com RXR ou qualquer outro receptor.

Figura 4. Alinhamentos estruturais entre estruturas resolvidas experimentalmente de receptores nucleares e modelos tridimensionais gerados pelo AlphaFold 3.



Alinhamentos estruturais entre receptores nucleares com estruturas experimentais resolvidas e modelos dos mesmos receptores nucleares gerados pelo AlphaFold 3. A) Sistema *HsRXR/HsPPAR*. B) Sistema *HsRXR/HsRAR*. As estruturas experimentais estão coloridas em amarelo, *HsRXR* está colorido em verde, *HsPPAR* e *HsRAR* estão coloridos em azul claro e as fitas de DNA estão coloridas em cinza. Fonte: elaborado pelo autor (2025).

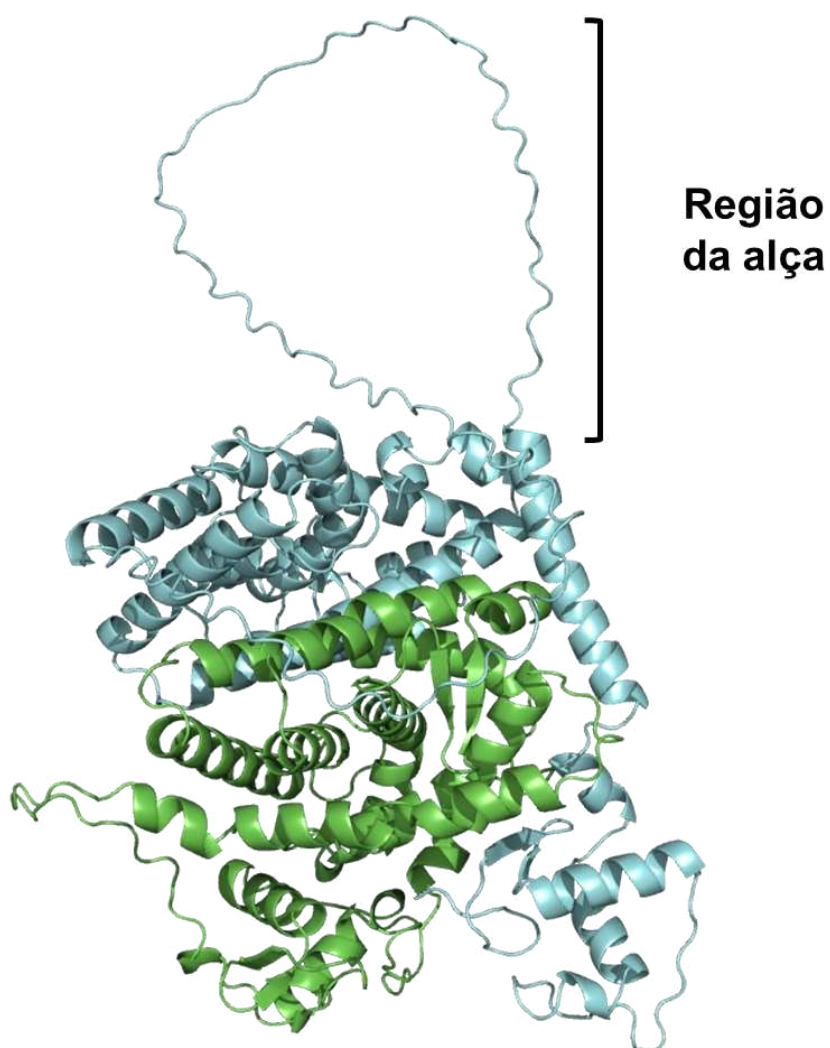
No entanto, estudos apontam que o receptor de estrógeno (ER) possa interagir de forma heterodimérica quando fora do núcleo, o que os autores classificaram como “funções não-genômicas” (Taylor & Heyland, 2022; Nadal, 2001), sendo identificado em compartimentos celulares como a membrana celular e em células especializadas como plaquetas. O modelo do dímero RXR/ER gerado pelo AlphaFold 3 mostra uma interface simétrica entre o LBD de ambos os receptores, em ambas as espécies. Em um primeiro momento, este modelo foi tido como um falso positivo visto que a subfamília NR3 não dimeriza com RXR, porém, dado que o ER forma heterodímeros fora do núcleo com receptores desconhecidos, acreditamos que esse possa ser um dímero válido.

Interessantemente, estudos demonstraram que o ER de *M. gigas* não possui um sítio ativo que comporte o estradiol (Bridgham, 2014), levantando dúvidas sobre sua função e se seria um receptor órfão. Ainda, Taylor e Heyland (2022) propõem que o caráter não-genômico poderia ser ancestral nos NR, havendo primeiramente apenas o LBD e posteriormente, um ganho de função

com o surgimento do DBD, a função genômica como fator de transcrição. Outro achado que dá suporte a essa visão é a identificação de receptores das subfamílias NR1 e NR2 fora do núcleo, como PPAR, RAR e o próprio RXR, que se acredita ser um dos NR mais antigos na história evolutiva dos NR.

Embora tenha sido possível criar modelos de controles positivos, negativos e identificar possíveis dímeros desconhecidos como RXR/ER, é importante ressaltar que a qualidade geral dos modelos foi alta devido aos domínios conservados. A região da alça entre domínios, como mencionado, é altamente desordenada e em alguns casos, é possível observar que nem mesmo algoritmos no estado da arte conseguem resolver estas estruturas atualmente. Isto fica mais evidente nos modelos gerados para receptores órfãos, aqueles que não possuem ligantes conhecidos, como o receptor órfão relacionado ao RAR (ROR) de *H. sapiens* e membros do subgrupo NR1P de *M. gigas* (**Figura 5**).

Figura 5. Modelo tridimensional do dímero RXR/ROR de *H. sapiens*.



Estrutura tridimensional do dímero *HsRXR/HsROR* onde é possível observar a região da alça de *HsROR* altamente desordenada. *HsRXR* está representado em verde, enquanto o *HsROR* está representado em azul claro. Fonte: elaborado pelo autor (2025).

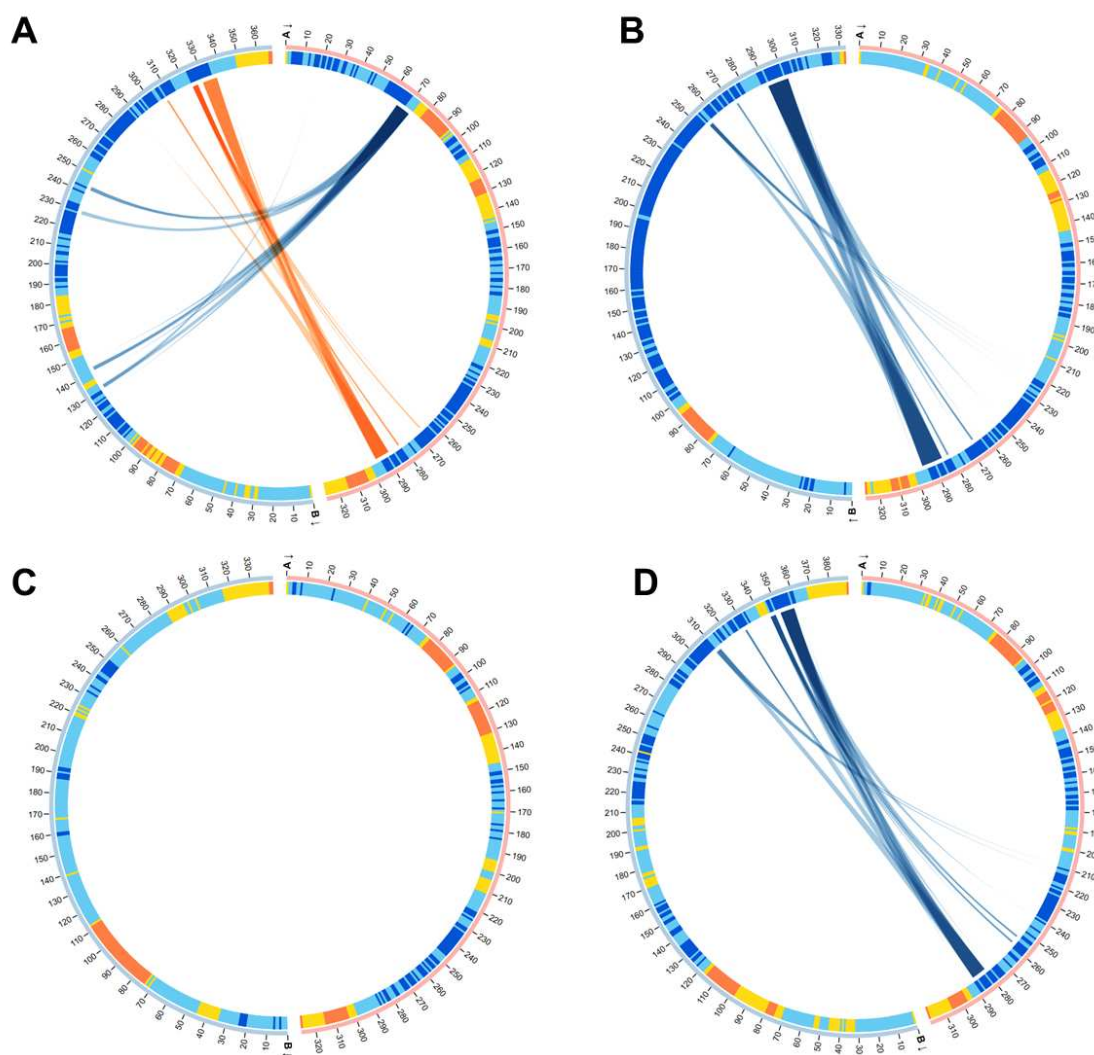
Essas regiões desordenadas não são necessariamente um problema na etapa de modelagem quando os domínios conservados são bem modelados, pois ainda é possível identificar uma interface simétrica entre os monômeros. Porém, para a etapa posterior de dinâmica molecular, seria inviável simular estruturas tão instáveis e com baixa confiabilidade de representarem uma estrutura verdadeira. Assim, esse foi um dos critérios de exclusão na escolha de um receptor membro do subgrupo NR1P para simulação de MD. Adicionalmente, em alguns casos é possível observar estruturas desordenadas em forma de alça em porções do LBD de alguns receptores, como no Rev-erb (NR1D) humano. O motivo para essas anomalias na modelagem tridimensional permanece desconhecido, mas sugerimos que isso possa se dar pela própria natureza

desordenada das regiões de *loops*, que mesmo em técnicas experimentais precisas como a cristalografia de raios-X não é possível determinar a posição da nuvem eletrônica dos átomos destas regiões. Outro argumento que possa ajudar a explicar este fenômeno pode ser a baixa disponibilidade ou inexistência de estruturas experimentais desses receptores para serem utilizadas como referência no processo de modelagem.

No entanto, apesar das limitações dos sistemas de modelagem atuais, é possível observar nas figuras acima, que boa parte dos sistemas modelados de ambas as espécies não apresentou alças altamente desordenadas, inclusive aqueles selecionados para as simulações. Percebe-se também que a integração dos resultados gerados pelo AlphaFold 3 por programas como o AlphaBridge facilita a busca por interfaces entre dímeros.

Como mencionado anteriormente, foram feitos modelos de dímeros para os 48 NR de *H. sapiens* e 43 de *M. gigas*. Todos estes modelos foram também submetidos à plataforma AlphaBridge para predição de interfaces de contato entre cadeias com base nas matrizes geradas pelo sistema AlphaFold 3. O resultado desta análise é um diagrama onde estão dispostas as cadeias modeladas, seus respectivos resíduos aminoacídicos e a rede de contatos de cada interface no centro. Além disso, a coloração da numeração dos resíduos segue àquela do índice pLDDT que reflete a qualidade de cada resíduo. Na **Figura 6** é possível observar exemplos destes diagramas para dímeros distintos, como os controles positivos, negativos e subgrupo NR1P, por exemplo. Nota-se que há uma concentração de contatos entre as regiões distais das cadeias, exatamente na região da I-box, onde se localizam a H10 e demais hélices e loops acessórios. Além disso, nenhuma interface ou contatos são identificados no modelo de controle negativo, reforçando a qualidade da modelagem e diminuindo a probabilidade de falsos positivos.

Figura 6. Predição de interfaces de interação entre dímeros de RXR e PPAR, RAR, GR e NR1P10.



Interfaces de interação entre dímeros de RXR e outros receptores nucleares preditas pelo programa AlphaBridge. A) *HsRXR/HsPPAR*. B) *HsRXR/HsRAR*. C) *HsRXR/HsGR*. D) *MgRXR/MgNR1P10*. Os números representam as posições dos resíduos aminoácídicos em suas respectivas cadeias. As faixas indicam as interfaces encontradas, coloridas em azul e laranja, quando presentes. As cores nas barras circulares representam os valores de pLDDT de cada resíduo. Fonte: AlphaBridge (em <https://alpha-bridge.eu>, 2025).

Esta análise, embora similar à proposta original do presente trabalho, não substitui a precisão e detalhes obtidos pelas análises de dinâmica molecular e trazem informações úteis para aqueles sistemas que não puderam ser submetidos às simulações. Além disso, é necessário lembrar que estes resultados são baseados em modelagens muitas vezes enviesadas para dados de *H. sapiens* e demais mamíferos, super-representados nos bancos de dados. Assim, estes resultados podem ser interpretados como indicadores de dímeros potenciais para análises mais profundas e extensas. Com os resultados promissores para os controles positivos e para o receptor NR1P10, foi dado início

às análises de dinâmica molecular, a fim de averiguar a estabilidade das interfaces propostas e o comportamento dos dímeros temporal e espacialmente.

4.3 DINÂMICA MOLECULAR DE DÍMEROS ENTRE RXR E PPAR, RAR E NR1P10

Foram feitas análises de dinâmica molecular por meio de simulações criadas a partir dos modelos gerados pelo AlphaFold 3, inclusive para aquelas que possuem moléculas de DNA associadas, como os sistemas utilizados como controle positivo. A versão mais recente do campo de força Martini foi escolhida devido a melhorias em relação à força das ligações entre esferas. Na versão anterior, Martini 2, havia uma tendência do modelo em manter as ligações unidas, não permitindo a separação e religação entre esferas. As simulações compreendem 10 μ s e foram feitas em triplicata, resultando em 30 μ s para cada sistema. De modo geral, foi possível simular todos os sistemas após realizar ajustes específicos para cada, detalhados abaixo.

4.3.1 O sistema RXR/PPAR

Entre os parceiros clássicos do RXR está o receptor ativado por proliferador de peroxissomos, ou PPAR. Este receptor está envolvido em vias metabólicas relacionadas principalmente ao metabolismo de lipídios e carboidratos (Wnuk & Lason 2018) e seus ligantes naturais incluem ácidos graxos e seus derivados (Lefebvre, 2010). Além dos ligantes naturais, compostos xenobióticos também podem interagir com estes receptores (Omiecinski & Peters 2010), como por exemplo o tributilestanho e trifenilestanho (Capitão, 2021), ftalatos (Issemann & Green, 1990), bisfenol A (Bourguet, 2015) e fármacos como rosiglitazona (Chandra, 2008), não só em humanos, mas como em invertebrados aquáticos (Miglioli, 2021). A ativação do receptor pelo ligante propicia a dimerização com o RXR e a interação com elementos de resposta em repetição direta separados por um ou dois nucleotídeos (DR1 e DR2; Chandra, 2008).

Ao todo, existem três isoformas conhecidas de PPAR em humanos – α , β/δ e γ –, sendo que todas realizam heterodimerização com o RXR (Wells, 2009). De modo geral, o PPAR forma dímeros com RXR pela mesma interface entre as décimas alfa-hélices do LBD identificada em outros dímeros de RN. No entanto, além desta interface mais comum, autores sugerem a presença de uma interface

entre os DBD de cada receptor, além de uma interface entre o LBD do PPAR-gama e o DBD do RXR-alfa (Chandra, 2008). Em relação às demais isoformas de PPAR, poucos estudos parecem abordá-las e as estruturas experimentais disponíveis não contemplam dímeros com RXR. Assim, como o PPAR-gama humano é a única isoforma representada em complexo com RXR, optamos por utilizá-la como controle positivo nas análises de dinâmica molecular.

Em relação aos invertebrados aquáticos, especialmente moluscos, foi identificada apenas uma isoforma de PPAR em espécies como *Magallana gigas* (Vogeler, ano), *Nucella lapillus* (Pascoal, Carvalho, Fang, 2012), *Mytilus galloprovincialis* (Miglioli, 2024) e duas isoformas em *Lottia gigantea* (Lockyer, Kaur, 2015). Evidências apontam que o surgimento das isoformas em moluscos, como os de *L. gigantea* tenha acontecido antes da separação com a linhagem dos vertebrados (Lockyer, Kaur, 2015) havendo homologia entre os PPAR de moluscos e humanos. Isso é reforçado pelo fato de que o PPAR se liga a ácidos graxos endógenos e pela similaridade estrutural entre os ortólogos destas espécies. Além disso, há evidência de que o dímero RXR/PPAR de *M. gigas* interaja com elementos de resposta DR1, reconhecendo a sequência 5'-AGGTCA-3' tal qual os ortólogos humanos fazem e que a região P-box do DBD de ambas as espécies são idênticas (Brascher, 2024). Assim, há indícios de que *MgRXR* e *MgPPAR* interajam entre si na forma de dímeros, porém, não há informações em relação a quais interfaces estão mediando esta interação. Para tentar elucidar esta questão, foram feitas simulações de dinâmica molecular do dímero *MgRXR/MgPPAR* utilizando os mesmos parâmetros e condições do dímero *HsRXR/HsPPAR*.

Em um primeiro momento, o sistema *HsRXR/HsPPAR* foi simulado sem restrições de posição e sem fitas de DNA associadas, compreendendo apenas os domínios conservados e a região da alça. Porém, foi observado que o sistema apresentava instabilidade, com os monômeros se separando rápida e definitivamente. Com a possibilidade de incluir fitas de DNA no modelo *coarse-grained* devido a parâmetros desenvolvidos por nosso grupo de pesquisa (dados não publicados), optamos por adicioná-las ao sistema. A estrutura experimental referente a este sistema (PDB: 3DZY) foi utilizada como referência para alinhar as fitas de DNA aos DBD do dímero. Assim, foi feita uma nova tentativa de simulação do sistema, que se mostrou mais estável devido à ancoragem do DBD

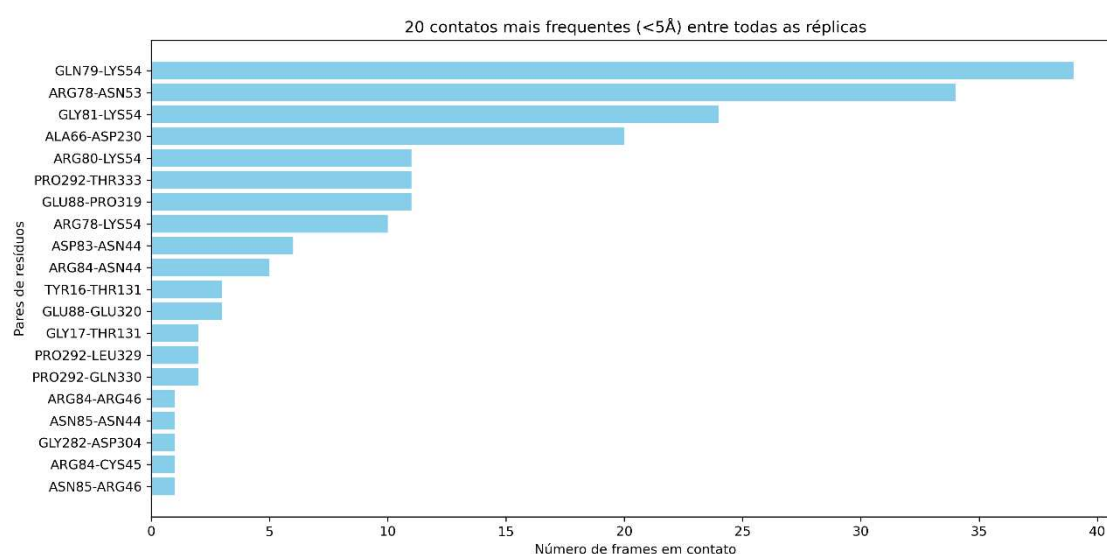
ao DNA, porém ainda bastante instável na região das alças, o que causou novamente uma separação definitiva dos monômeros. Portanto, foi decidido impor restrições de posição nas extremidades das alças de ambos os monômeros a fim de manter a estrutura do sistema mais coesa, porém sem remover a liberdade de movimento de uma região intrinsecamente desordenada.

A tentativa de simulação com restrições de posição gerou resultados promissores, com um sistema estável ao longo de toda a trajetória das três réplicas. Ainda, foi possível identificar a interface de dimerização entre as H10, além da interface entre o DBD do RXR e LBD do PPAR, encontrada por Chandra e colaboradores (2008) em todas as simulações. Adicionalmente, foram identificados diversos contatos distintos e frequentes entre resíduos dos DBD de ambos os receptores, ressaltando a importância deste domínio não só para a interação com o DNA e elementos de resposta, mas talvez para o próprio processo de dimerização. É importante ressaltar que o sistema parte de uma posição inicial favorável à dimerização, com os monômeros pareados. Assim, não é possível inferir se o DBD participa do processo de dimerização antes da interação com o DNA ou apenas auxilia na manutenção da estabilidade do sistema. Todavia, este é o primeiro resultado reportando a simulação deste dímero com todos os domínios conservados e, ainda, com DNA associado.

Entre os contatos mais frequentes identificados nas três interfaces (**Figura 7**), pode-se dar destaque ao resíduo de prolina 292 do RXR que é identificado em todas as réplicas com frequência e se situa na H10. Esta prolina interage com outros três resíduos da H10 oposta, reforçando o papel desta interface no processo de dimerização e estabilidade da estrutura. Entre os resíduos que realizam contatos mais frequentes com a P292, está a treonina 333 do PPAR (por aproximadamente um quarto da simulação), seguida pela leucina 329 e glutamina 330 deste mesmo receptor. Estes resultados indicam um papel central deste resíduo em mediar contatos com diversos resíduos parceiros quase que simultaneamente na principal interface de contato do sistema *HsRXR/HsPPAR*. Em sequência, é possível observar uma predominância de contatos entre os DBD, ocupando as três primeiras posições por quase todo o período das simulações. Até então, o foco principal da literatura está na interface “canônica” entre H10 no LBD (Li, 2022; Pawlak, 2012), sendo tida como principal responsável pela dimerização de receptores nucleares. De fato, o LBD parece

ser o principal ponto para a dimerização e formação de uma estrutura simétrica, porém, as análises anteriores de dinâmica molecular foram realizadas utilizando apenas este domínio, sem levar em consideração a interação entre domínios e o papel central do DBD na estabilização do dímero. Com estes resultados, esperamos realçar a necessidade de analisar estruturas mais completas e de investigar interfaces além daquelas tidas como canônicas.

Figura 7. Contatos totais com menos de 5 Å entre *HsRXR* e *HsPPAR*.



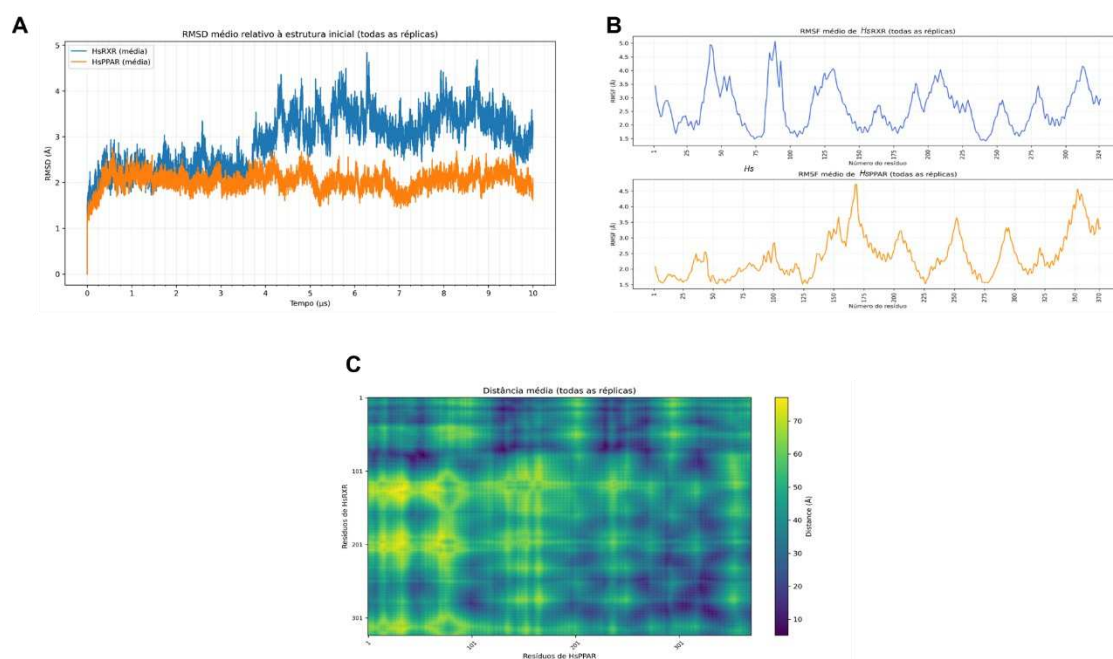
Soma de pares de contatos identificados entre *HsRXR* e *HsPPAR* em todas as réplicas. Resíduos estão nomeados de acordo com a nomenclatura de três letras, seguindo a disposição *HsRXR* – *HsPPAR*. Fonte: elaborado pelo autor (2025).

Como mencionado anteriormente, a interface entre DBD e LBD é representada por dois pares de contatos bastante frequentes – RXR-A66/PPAR-D230 e RXR-E88/PPAR-P319. Esta é uma interface até então exclusiva do sistema *HsRXR/HsPPAR* e conseguir identificá-la utilizando apenas métodos *in silico* reflete a precisão das técnicas atuais em replicar dados experimentais. Além disso, identificar os contatos responsáveis por esta interface e seus tempos de duração são dados inéditos e reforçam a necessidade das análises de dinâmica molecular para a descoberta e caracterização de interações proteína-proteína.

Adicionalmente, outros dados foram coletados a fim de investigar o comportamento dos monômeros ao longo das simulações e as distâncias entre eles, bem como mapas de contato para ilustrar regiões mais ativas de cada monômero. Os dados de variação de RMSD e RMSF, distâncias mínimas e

mapas de contato foram coletados para cada réplica e suas médias foram calculadas e representadas na **Figura 8A, B e C**. É possível observar que os valores de RMSD de ambos os monômeros foram bastantes estáveis, raramente apresentando variações acima de 4 Å, indicando que a posição do sistema se manteve similar àquela ao início da simulação. Os dados de RMSF refletem a flutuação média de cada resíduo ao longo da simulação, com valores menores indicando resíduos ou regiões que se mantiveram estáveis ao longo da trajetória, provavelmente devido a contatos realizados com a cadeia oposta. Por fim, o mapa de contato de fato mostra as regiões com maior interação com a cadeia oposta, sendo possível observar uma maior atividade entre os domínios conservados.

Figura 8. Valores médios de RMSD, RMSF e mapa de contatos do sistema *HsRXR/HsPPAR*.

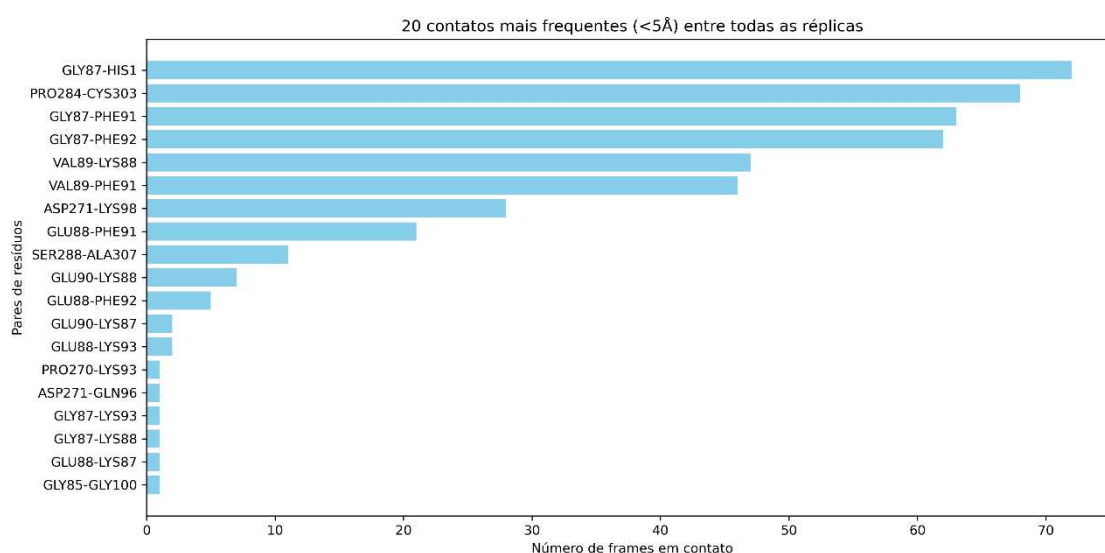


A) Valores médios de RMSD e B) RMSF de *HsRXR* e *HsPPAR* em azul e laranja, respectivamente. C) Mapa de contatos ilustrando a distância média entre resíduos de *HsRXR* e *HsPPAR*. Fonte: elaborado pelo autor (2025).

Em relação ao sistema ortólogo de *M. gigas*, *MgRXR/MgPPAR*, também foi possível obter resultados promissores em relação a interfaces mediando o processo de dimerização e quais resíduos as compõem (**Figura 9**). Duas interfaces de interação foram identificadas neste sistema, uma entre os DBD e outra entre os LBD, não sendo possível identificar a terceira interface encontrada no sistema de *H. sapiens*. Duas hipóteses surgem para explicara a ausência

desta última interface, a) a ausência da fita de DNA no sistema altera a disposição dos DBD e pode influenciar na proximidade entre eles ou na exposição de regiões específicas que não estão interagindo ou b) esta é uma novidade do dímero encontrado em humanos (ou vertebrados, mamíferos ou primatas). Para validar estas hipóteses seria necessário encontrar e modelar os elementos de resposta do sistema *MgRXR/MgPPAR* e inclui-lo em simulações futuras e, para o segundo cenário, replicar este sistema com sequências ortólogas de outras espécies de vertebrados. De toda maneira, estes são dados inéditos para receptores nucleares da ostra-do-pacífico e invertebrados no geral.

Figura 9. Contatos totais com menos de 5 Å entre *MgRXR* e *MgPPAR*.



Soma de pares de contatos identificados entre *MgRXR* e *MgPPAR* em todas as réplicas. Resíduos estão nomeados de acordo com a nomenclatura de três letras, seguindo a disposição *MgRXR* – *MgPPAR*. Fonte: elaborado pelo autor (2025).

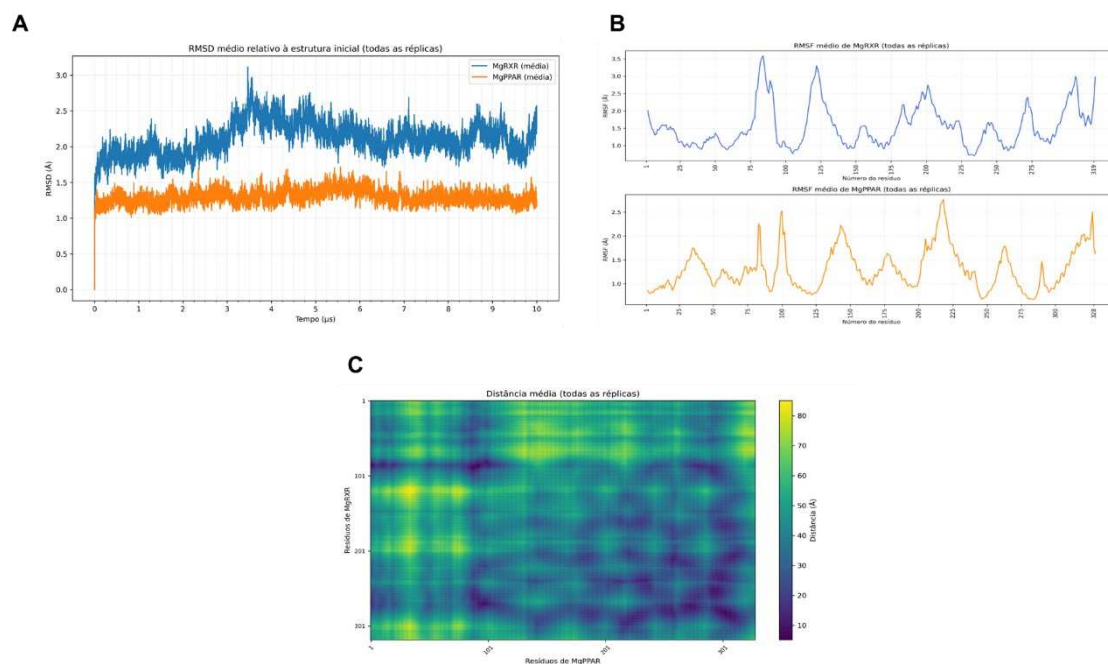
Entre os resíduos identificados como chave para a interação entre os monômeros estão um resíduo de glicina 87 do *MgRXR* que realiza contatos frequentes com outros seis resíduos do DBD oposto. Além desta glicina, outros resíduos do DBD de *MgRXR* participam ativamente na dimerização, como E88, V89 e E90. Ambos os glutamatos formam pares de contato com resíduos de lisina, carregados positivamente em pH neutro, sugerindo a presença de interações iônicas na estabilização do dímero. Em relação ao LBD, novamente é possível observar um resíduo de prolina situado na H10 realizando contatos com a cadeia oposta. De fato, este parece ser um resíduo conservado entre RXR de espécies diferentes e indica que este resíduo específico possa ter maior

relevância para o processo de dimerização. Outros estudos demonstraram a importância de resíduos de prolina para interações proteína-proteína, especialmente em proteínas que contêm domínios ricos em prolina (Ball, 2005).

Entre as características únicas deste aminoácido está a sua cadeia lateral aromática contendo o átomo de nitrogênio da cadeia principal, formando uma pirrolidina e deixando seu grupo amida sem hidrogênios, não sendo possível estabilizá-la através de ligações de hidrogênio (Pujals, 2007). Assim, este resíduo possui poucas conformações possíveis e maior rigidez quando comparado a outros resíduos aminoacídicos, resultando em maior estabilidade conformacional e menor custo energético para formar ligações predominantemente hidrofóbicas (Kay, 2000). Estas características, junto com o grande volume ocupado pela sua cadeia lateral, fazem com que a prolina tenha grande importância estrutural, geralmente causando quebras ou torções na estrutura tridimensional das proteínas e formando ligações de hidrogênio com outros resíduos com facilidade.

Adicionalmente, uma outra interface inédita pode estar presente no sistema RXR/PPAR de *M. gigas*, inversa àquela encontrada no sistema humano entre o DBD de RXR e LBD de PPAR. Foi possível identificar contatos próximos e frequentes entre três pares distintos de contatos entre resíduos do LBD de MgRXR e o DBD de MgPPAR. Destes, dois envolvem um resíduo de aspartato (D271) interagindo com resíduos de lisina e glutamina, novamente sugerindo a presença de interações iônicas entre as cadeias. O outro par identificado envolve um outro resíduo de prolina (P270) situado no loop entre H9 e H10, onde há uma torção na estrutura. Ademais, as métricas de desvio e flutuação quadrática média e mapas de contato (**Figura 10A, B e C**) apontam para um sistema estável durante as simulações, com valores de RMSD abaixo em torno de 1,5 a 2 Å e valores de RMSF abaixo de 1,5 a 1 Å para resíduos situados nos DBD e LBD principalmente. O mapa de contato é bastante similar àquele do sistema controle, indicando maior atividade entre os LBD, e em menor grau, entre os DBD.

Figura 10. Valores médios de RMSD, RMSF e mapa de contatos do sistema *MgRXR/MgPPAR*.



A) Valores médios de RMSD e B) RMSF de *MgRXR* e *MgPPAR* em azul e laranja, respectivamente. C) Mapa de contatos ilustrando a distância média entre resíduos de *MgRXR* e *MgPPAR*. Fonte: elaborado pelo autor (2025).

Estes resultados apontam para a existência de um dímero funcional e estável entre *MgRXR* e *MgPPAR*, com interfaces similares àquelas encontradas em receptores ortólogos humanos. Em relação à atividade deste sistema, estudos apontam a falta de informação sobre seus ligantes conhecidos em moluscos e dúvidas sobre o papel do PPAR no metabolismo de lipídios em invertebrados (Capitão, 2021). Porém, há registros da interação de organotinas como TBT interagindo tanto com o RXR quanto o PPAR de moluscos, indicando atividade no metabolismo de xenobióticos (Capitão, 2021; Vogeler, 2017). Assim, sugere-se a inclusão destes receptores em estudos de ecotoxicologia envolvendo não só *M. gigas*, mas como outros moluscos bivalves e invertebrados. Além disso, faz-se necessária a identificação dos elementos de resposta deste sistema a fim de investigar quais genes estão sendo modulados após sua ativação por ligantes e xenobióticos.

4.3.2 O sistema RXR/RAR

Outro parceiro de dimerização extensamente documentado em vertebrados é o receptor de ácido retinoico (RAR), também pertencente à

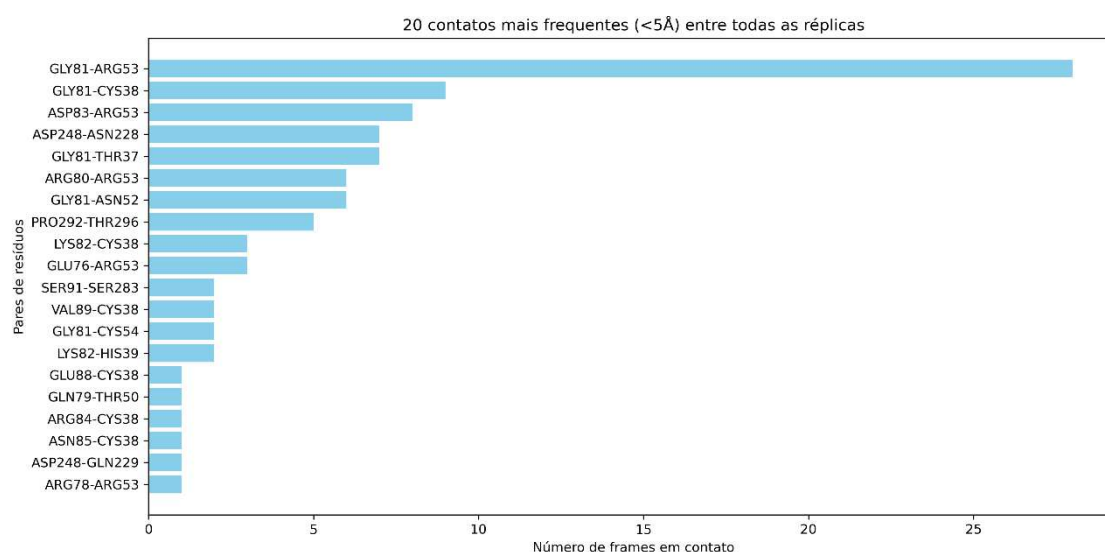
subfamília NR1. Entre os ligantes conhecidos deste receptor estão diferentes isoformas do ácido retinoico, como o ácido *all-trans*-retinoico (ATRA) e ácido 9-*cis*-retinoico (9REA), além de compostos exógenos como xenobióticos como pesticidas (Inoue, 2010). Em humanos e outros vertebrados, existem três isoformas do RAR (α , β e γ), enquanto em invertebrados como moluscos apenas uma isoforma foi identificada (Vogeler, 2014; Miglioli, 2024). Outros taxa de invertebrados como insetos não possuem RAR conhecidos até o momento, com o entendimento de que uma exclusão gênica tenha ocorrido neste grupo (Fonseca, 2019). Este dado é suportado pela identificação de um gene de RAR em uma espécie de Priapulida (Fonseca, 2019), um grupo derivado de Ecdysozoa e irmão de Panarthropoda, sugerindo a presença deste receptor no ancestral dos ecdisozoários. Outro ponto importante sobre a evolução deste receptor é a evidência de que a afinidade por ácidos retinoicos seja uma característica derivada e adquirida apenas em vertebrados. Estudos apontam que animais deuterostômios basais possuem uma maquinaria responsável pelo desenvolvimento do plano corporal e tecido nervoso controlada por ácidos retinoicos e RAR e que este sistema refinado seria uma novidade responsável em parte pela radiação deste grupo (Cañestro, 2006).

Ainda, outros dados apontam a baixa afinidade do RAR de invertebrados com ácidos retinoicos, principalmente em moluscos bivalves e gastrópodes (Vogeler, 2017; Gutierrez-Mariegos, 2014), embora haja alta similaridade entre as sequências de *MgRAR* e *HsRAR*, por exemplo. Assim, embora haja indícios para o surgimento de RAR em animais urbilatérios (Campo-Paysaa, 2008), o ganho de função para interação com ácidos retinoicos aparenta ser uma aquisição tardia. No entanto, há indícios para a interação do RAR com ácidos retinoicos quando pareado com o RXR, sugerindo um efeito sinérgico entre os receptores (André, 2019), visto que o RXR de moluscos conservou a habilidade de interagir com ácidos retinoicos (Bouton, 2005). Dessa forma, fica evidente a importância do dímero RXR/RAR em moluscos para a manutenção da atividade e função de ácidos retinoicos na diferenciação celular e maturação sexual deste grupo (Castro, 2007; Nishikawa, 2004) e em humanos e vertebrados para o desenvolvimento do plano corporal, por exemplo.

As simulações de dinâmica molecular do sistema *HsRXR/HsRAR*, utilizado como controle positivo para o sistema ortólogo de *M. gigas* foram realizadas com

sucesso, sem a necessidade de impor restrições de posição nas extremidades das alças. A inclusão das fitas de DNA contendo o elemento de resposta DR1 foi realizada alinhando a estrutura das fitas de DNA à estrutura experimental deposita no PDB (código de acesso: 5UAN). Todas as réplicas foram simuladas por completo, provendo resultados similares entre si que indicam estabilidade do sistema ao longo do tempo e contatos entre os monômeros por duas interfaces principais (**Figura 11**). A primeira interface sugerida é entre os DBD de ambos os monômeros, envolvendo principalmente o resíduo de glicina 81 de *HsRXR* e a arginina 53 de *HsRAR*, que formam cinco pares de contatos distintos ao longo das trajetórias, inclusive entre si. A G81 foi identificada no sistema *HsRXR/HsPPAR* realizando contatos próximos, embora em menor grau. Em relação à R53, nota-se que dois de seus pares formados são com resíduos carregados negativamente, sugerindo de novo a presença de interações iônicas entre as cadeias. Estas ligações, geralmente fracas, são frequentes em interações proteína-proteína e proteína-ácido nucleico, estando presentes em distâncias de até 5 Å, como observado nos resultados aqui apresentados (Tam, 2022; Yu, 2020). Ainda, quando interações iônicas ocorrem próximas e simultaneamente com ligações de hidrogênio, há a ocorrência de pontes salinas, que podem estar presentes no processo de dimerização destes receptores nucleares (Tam, 2022). Visto que os demais pares de contato da R53 não são resíduos carregados negativamente e envolvem resíduos próximos (R78, R80, G81 de *HsRXR*), há a possibilidade deste outro tipo de interação estar agindo na manutenção da dimerização.

Figura 11. Contatos totais com menos de 5 Å entre *HsRXR* e *HsRAR*.



Soma de pares de contatos identificados entre *HsRXR* e *HsRAR* em todas as réplicas. Resíduos estão nomeados de acordo com a nomenclatura de três letras, seguindo a disposição *HsRXR* – *HsRAR*. Fonte: elaborado pelo autor (2025).

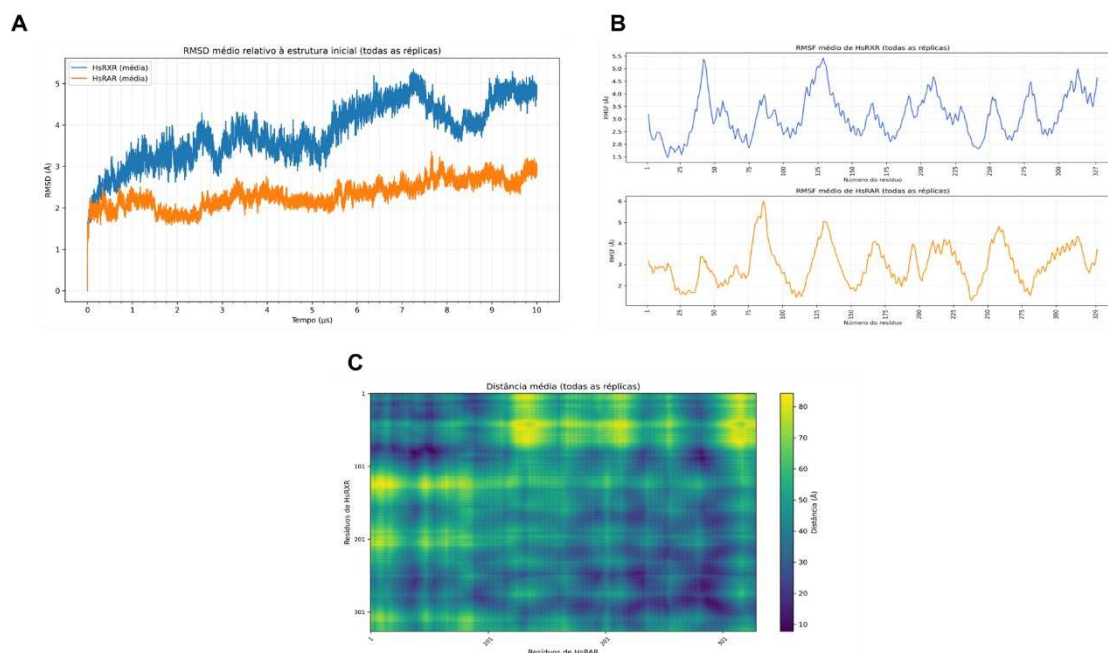
Em relação ao LBD, três pares contatos principais foram encontrados. Dois deles envolvem o resíduo de aspartato 248 de *HsRXR* com a glutamina 229 e asparagina 228 de *HsRAR*, ambos com potencial iônico positivo. Todos estes pares se encontram em hélices e loops próximos a interface principal, como o loop entre H8 e H9 de *HsRXR* e a H7 de *HsRAR*. Interessantemente, estudos apontaram estas regiões como acessórias para a manutenção da interface entre as H10 (Beinsteiner, 2021), principalmente entre *HsRXR/HsRAR* por assumir uma estrutura levemente assimétrica, havendo maior participação das H9 na sua dimerização. Ainda, os autores apontam a presença de um motivo único na H7 de RXR, definido como “ π -turn” (curva pi, em tradução livre), que interage com resíduos das H8 e H9 de RAR (Beinsteiner, 2021). Os resultados das simulações aqui apresentadas não apontam estas interações, indicando na realidade contatos dispostos de maneira exatamente inversa ao observado por Beinsteiner e colaboradores, sugerindo que talvez as hélices de ambos os receptores tenham a capacidade de realizá-las. Apesar disso, o motivo π -turn aparenta ser conservado ao longo da escala evolutiva, com a função primária de facilitação estrutural ao estabilizar a H10 e H11, permitindo primeiramente a homodimerização em receptores ancestrais e posteriormente garantindo a

habilidade do RXR de realizar ambas as formas de dimerização (Beinsteiner, 2021).

Em sequência, o terceiro contato observado entre LBD envolve novamente a P292 encontrada no sistema anterior, interagindo novamente com um resíduo de treonina, neste caso a T296, situada também na H10 do LBD. Observa-se no sistema RXR/RAR humano uma menor participação da H10 quando comparado ao sistema RXR/PPAR, com maior participação das regiões acessórias em mediar as interações entre as cadeias. Estes dados compelem a busca por interfaces além da H10, sugerindo uma expansão desta interface clássica no sistema RXR/RAR humano, e a relevância dos contatos entre DBD na dimerização.

Em relação às métricas de qualidade das simulações, foram calculadas as médias de valores de RMSD, RMSF e dos mapas de contato entre as réplicas a fim de minimizar o efeito de pontos fora da curva (**Figura 12A, B e C**). Os resultados de desvio quadrático médio apontam desvios baixos em relação à posição inicial do sistema, abaixo de 5 Å, o que indica estabilidade da estrutura geral ao longo das simulações. Em sequência, os dados das flutuações quadráticas médias para cada resíduo mostram quais regiões se mantiveram mais próximas à cadeia oposta ao longo das trajetórias, sendo possível observar a predominância de resíduos dos DBD e LBD abaixo de 2 Å. Por fim, o mapa de contato mostra um padrão similar àquele observado em ambos os sistemas RXR/PPAR, com as regiões dos DBD e LBD mais ativas entre si.

Figura 12. Valores médios de RMSD, RMSF e mapa de contatos do sistema HsRXR/HsRAR.



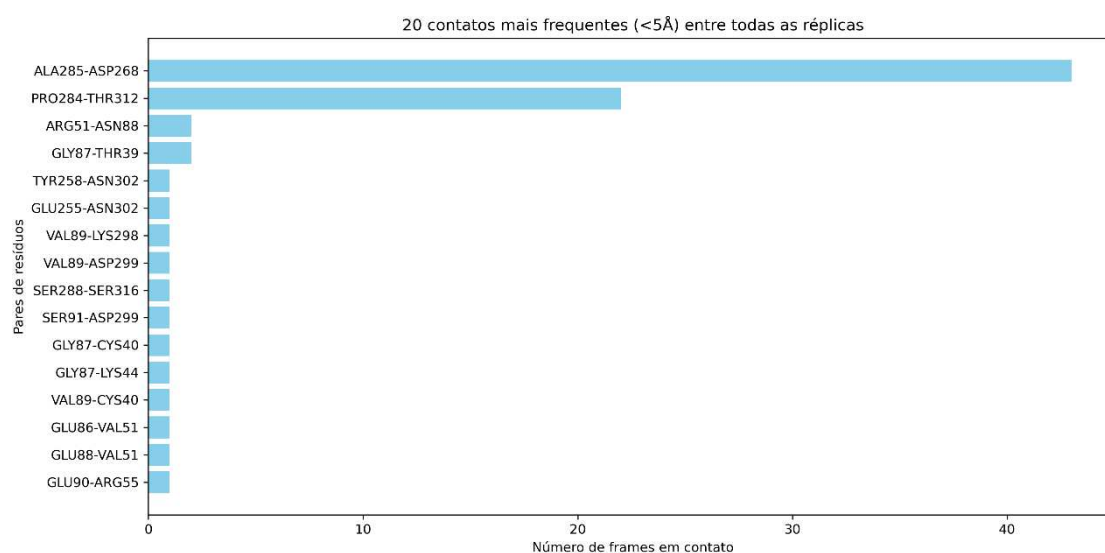
A) Valores médios de RMSD e B) RMSF de HsRXR e HsRAR em azul e laranja, respectivamente. C) Mapa de contatos ilustrando a distância média entre resíduos de HsRXR e HsRAR. Fonte: elaborado pelo autor (2025).

Seguindo para o sistema RXR/RAR de *M. gigas*, foi possível identificar interfaces de interação entre os monômeros, reforçando a capacidade destes receptores de se ligarem, embora ainda não se tenha muitos indícios da funcionalidade deste sistema ou sobre quais vias metabólicas são moduladas pela sua ativação. Até o momento, informações sobre este sistema em invertebrados focaram mais em espécies de cordados como o cefalocordado *Branchiostoma lanceolatum* e o urocordado *Botrylloides leachi*, com evidências para a formação de um dímero entre RXR e RAR destas espécies (Rinkevich, 2007; Escriva, 2006). Como mencionado anteriormente, a presença deste sistema em moluscos se restringiu principalmente a gastrópodes (Gutierrez-Mariegos, 2014; Urushitani, 2013), e até o presente momento, pouca informação está disponível em relação a espécies de bivalves como *M. gigas*.

Neste trabalho, foi possível identificar uma possível interface principal de interação entre MgRXR e MgRAR, situada entre os LBD (**Figura 13**). Dois pares de contatos foram identificados com alta frequência, um envolvendo novamente um resíduo de prolina situado na posição 284 (H10) do RXR, interagindo com a treonina 312 de RAR (H10). Este achado segue o mesmo padrão observado no

sistema controle, sendo observado durante boa parte das simulações numa distância menor que 5 Å. A presença de um mesmo resíduo na mesma hélice entre receptores de espécies tão distantes reforça a alta taxa de conservação destes receptores ao longo da evolução e pode sugerir conservação de função, por mais que o RAR de moluscos não possua afinidade por ácidos retinóicos. Em sequência, o outro contato observado, ainda mais frequente que o anterior, envolve a alanina 285 de *MgRXR* (vizinha à P284) e o aspartato 268 de *MgRAR*, situado no loop entre H8 e H9, sugerindo a participação de regiões acessórias na interface de interação, tal qual o sistema controle.

Figura 13. Contatos totais com menos de 5 Å entre *MgRXR* e *MgRAR*.



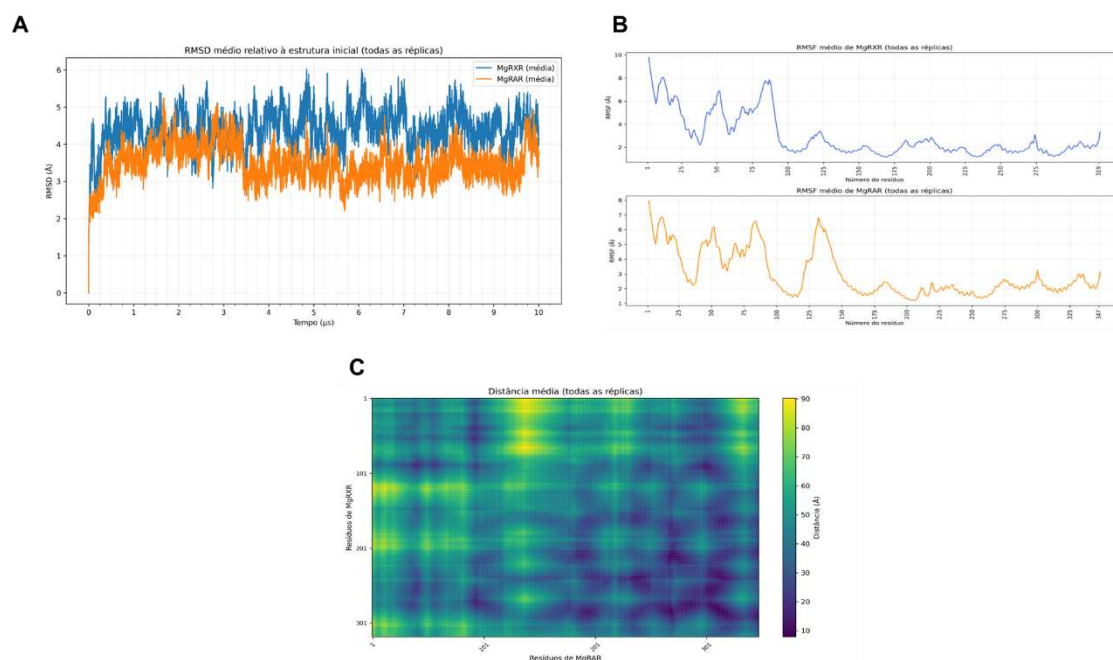
Soma de pares de contatos identificados entre *MgRXR* e *MgRAR* em todas as réplicas. Resíduos estão nomeados de acordo com a nomenclatura de três letras, seguindo a disposição *MgRXR* – *MgRAR*. Fonte: elaborado pelo autor (2025).

Alguns outros pares de contatos entre os LBD e DBD foram identificados, porém em frequências bastante baixas, não sendo possível inferir uma interface de interação estável entre eles. O motivo para a inexistência de uma interface entre os DBD pode ser a ausência das fitas de DNA contendo um elemento de resposta ancorando estes domínios. Embora contatos entre DBD e DNA não tenham sido mapeados neste trabalho, a presença das fitas de DNA garante maior estabilidade estrutural ao sistema. Assim, recomenda-se refazer estas análises quando os elementos de resposta do sistema *MgRXR/MgRAR* forem identificados e sua modelagem possível. Todavia, estes são dados inéditos que contribuem para o entendimento do comportamento e reconhecimento destes

receptores em bivalves. Visto que há registros tanto da interação do RXR quanto do RAR com contaminantes em moluscos (André, 2019), é imprescindível entender como e com quais outros receptores estas moléculas interagem.

As métricas de qualidade deste sistema reforçam a precisão das análises em identificar contatos e interfaces de interação, sendo possível observar valores de RMSD em torno de 4 Å para ambos receptores em relação às suas posições iniciais (**Figura 14A**). Os dados de RMSF indicam valores maiores para a região do DBD, indo ao encontro do padrão observado na listagem de contatos frequentes, indicando que os resíduos dos DBD apresentaram maior movimentação durante as simulações, enquanto o oposto é observado nos resíduos compondo os LBD, onde está a única interface de interação (**Figura 14B**). No mapa de contato entre as réplicas, é possível observar maior atividade na região do LBD, em especial nas porções mais distais englobando a H10 e hélices próximas, com menor atividade entre os DBD (**Figura 14C**).

Figura 14. Valores médios de RMSD, RMSF e mapa de contatos do sistema *MgRXR/MgRAR*.



A) Valores médios de RMSD e B) RMSF de *MgRXR* e *MgRAR* em azul e laranja, respectivamente. C) Mapa de contatos ilustrando a distância média entre resíduos de *MgRXR* e *MgRAR*. Fonte: elaborado pelo autor (2025).

De forma similar ao observado no sistema RXR/PPAR, o padrão observado nos controles parece ser conservado na ostra-do-pacífico. Embora haja

particularidades entre os sistemas de cada espécie, as semelhanças são suficientes para supor conservação no modo de interação entre estes receptores. Visto que há conservação em termos de sequência, estrutura, e agora interação, faz-se necessário investigar a função destes dímeros em *M. gigas* e outros invertebrados afetados por contaminantes, especialmente quando indícios de interação com xenobióticos crescem continuamente. Com os métodos atuais é possível investigar a interação destes receptores com ligantes em simulações *coarse-grained* para escalas de tempo suficientes para a entrada de um ligante no sítio ativo, por exemplo. Porém, o processo de parametrização das coordenadas da estrutura atômica de um ligante para uma estrutura *coarse-grained* é bastante complexo e custoso computacionalmente, não sendo possível atualmente realizá-lo em larga escala. Assim, as análises de dinâmica molecular *coarse-grained* com receptores nucleares se limitam principalmente à interação entre receptores, exceto para aqueles poucos com ligantes parametrizados. No entanto, estas análises *in silico* podem servir como triagem de potenciais candidatos para análises experimentais que possam validar estes resultados, como é o caso dos sistemas aqui apresentados.

4.3.3 O sistema *MgRXR/MgNR1P10*

O último sistema escolhido para análise envolve um receptor membro de um subgrupo até então exclusivo de moluscos bivalves, o subgrupo NR1P (Miglioli, 2024; Vogeler, 2014). Este subgrupo foi identificado pela primeira vez em *M. gigas* contendo 11 receptores distintos (Vogeler, 2014) e, posteriormente, outros 14 membros deste subgrupo foram encontrados em *M. galloprovincialis* (Miglioli, 2024). Até então, nenhum outro registro deste subgrupo foi feito em outras espécies e pouco se sabe sobre a sua função ou sobre a presença de ortólogos em outras espécies. No entanto, foi observado em ambos os estudos que o subgrupo NR1P se situa próximo aos subgrupos NR1C, NR1D, NR1E e NR1F em análises filogenéticas, sugerindo uma origem comum entre estes receptores (Vogeler, 2014; Miglioli, 2024). O subgrupo NR1C abriga o *MgPPAR*, presente no metabolismo de lipídios em bivalves, enquanto NR1D contém um homólogo do receptor Rev-Erb que, em vertebrados, liga-se ao grupo heme e participa de processos envolvendo proliferação celular, por exemplo (Sulli, 2018). Adicionalmente, o subgrupo NR1E aparenta estar presente apenas em

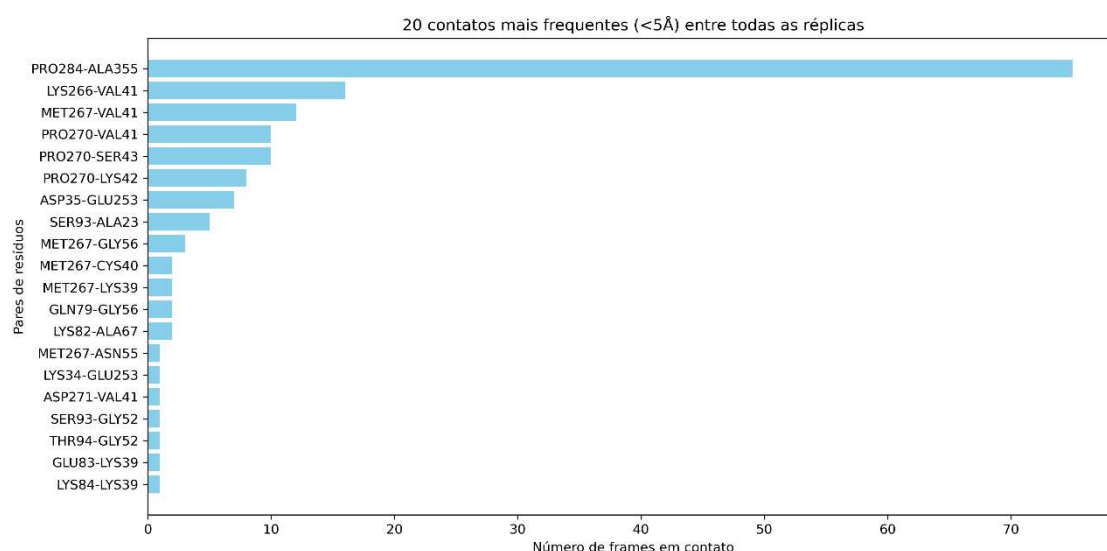
invertebrados como moluscos e insetos, sem ortólogos conhecidos em vertebrados (Vogeler, 2014). O subgrupo NR1F abriga o ortólogo do receptor órfão relacionado ao RAR (ROR) que, em humanos, liga-se a esteroides e participa em processos inflamatórios e metabolismo de lipídios e glicose (Jetten, 2013). Visto que o subgrupo NR1P ainda não foi caracterizado e seus ligantes e funções são desconhecidas, pode-se supor baseado em seus subgrupos irmãos que este subgrupo possa estar envolvido no metabolismo de alguma biomolécula como açúcares e lipídios, por exemplo. Vale ressaltar que apenas os receptores do subgrupo NR1C possuem registros de interação com RXR, enquanto NR1D e NR1F agem de forma monomérica ou homodimérica (Rastinejad, 2001; Harding, 1997).

Todavia, as análises de modelagem estrutural e predição de interfaces de interação resultaram em um dímero simétrico entre NR1P10 e NR1P11 com o RXR, indicando as H10 como principal região de contatos entre estes receptores. Com base nestes dados, decidimos investigar o comportamento de NR1P10 com RXR, visto que estudos de nosso grupo de pesquisa apontam para a co-expressão de ambos os receptores em ensaios ecotoxicológicos (Saldaña, 2025) e para a afinidade de NR1P10 com moléculas de estradiol (Madaloz, 2022). Embora não haja ortólogos deste receptor em vertebrados nem o registro da sua presença em qualquer outra espécie, não foi possível incluir um controle positivo para este sistema, porém, as etapas de preparação, minimização, equilibração e produção foram realizadas de maneira similar aos outros sistemas. Para este sistema, as fitas de DNA contendo elementos de resposta não foram incluídas e restrições de posição não foram impostas nas regiões das alças.

Foi possível concluir todas as simulações sem intercorrências e os resultados apontam para um sistema estável ao longo da trajetória. Estes resultados são inesperados visto que a modelagem foi realizada sem a utilização de muitas sequências homólogas, dado que elas não existem para a maioria das espécies, o que poderia sugerir um enviesamento do modelo AlphaFold 3 em relação a conformações características de outros subgrupos de NR utilizados como referência. Ao mesmo tempo, a qualidade do modelo de NR1P10 pode também reforçar a precisão do AlphaFold 3 em atingir estruturas tridimensionais coerentes mesmo com escassez de sequências homólogas próximas.

De toda maneira, foi possível encontrar pares de contatos em diferentes regiões destes receptores, indicando duas interfaces de contato principais (**Figura 15**). Como observado nos sistemas anteriores, um dos resíduos mais ativos de *MgRXR* é a P292, que neste caso forma um par com a alanina 355 de *MgNR1P10*, ambas situadas nas suas respectivas H10. Este achado sugere um padrão de interação entre RXR e seus parceiros de interação envolvendo este resíduo específico, presente em todos os sistemas analisados neste trabalho. De fato, ainda não é possível afirmar que *MgNR1P10* é um parceiro de dimerização de *MgRXR*, porém, com estes resultados há fortes indícios para a inclusão destes receptores em experimentos de hibridização por exemplo. Além disso, foi observado que não só *MgNR1P10* é co-expresso com *MgRXR*, mas outros membros do subgrupo NR1P também o fazem, indicando a possibilidade de mais parceiros de *MgRXR* neste subgrupo (Vogeler, 2014).

Figura 15. Contatos totais com menos de 5 Å entre *MgRXR* e *MgNR1P10*.

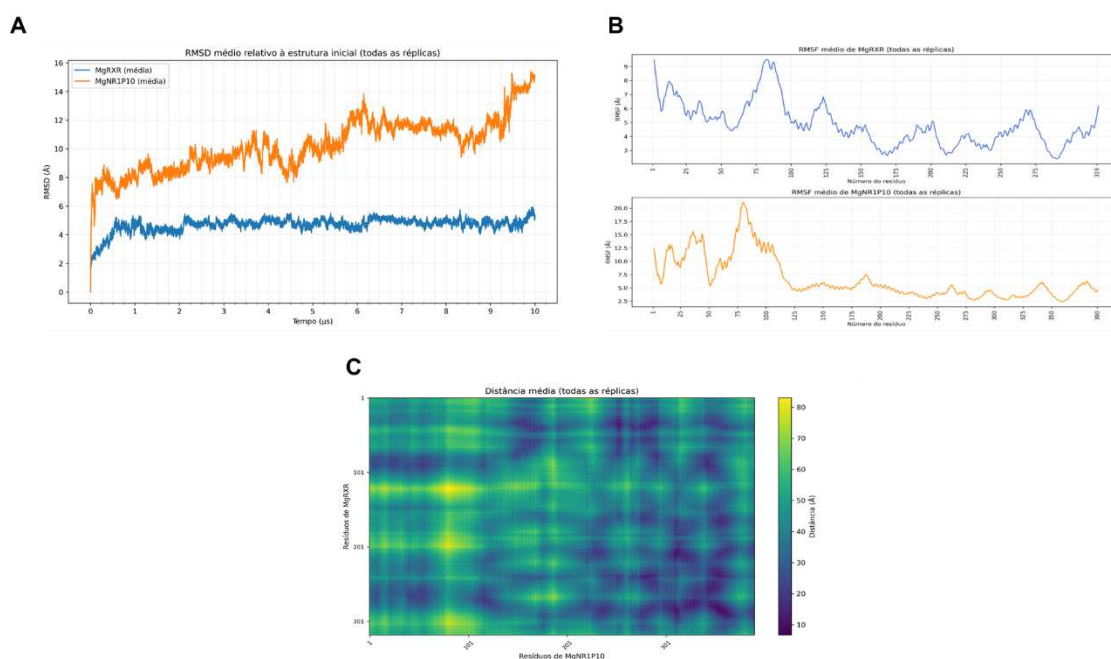


Soma de pares de contatos identificados entre *MgRXR* e *MgNR1P10* em todas as réplicas. Resíduos estão nomeados de acordo com a nomenclatura de três letras, seguindo a disposição *MgRXR* – *MgNR1P10*. Fonte: elaborado pelo autor (2025).

Em relação à segunda possível interface de interação encontrada neste sistema, foram identificados diversos pares de contato envolvendo resíduos situados no LBD de *MgRXR* com resíduos do DBD de *MgNR1P10*. Estes contatos envolvem resíduos de lisina, metionina e prolina (K266, M267 e P270) de *MgRXR* interagindo principalmente com a valina 41 de *MgNR1P10* e outros

resíduos próximos em menor grau. Devido à natureza destes resíduos, supõe-se a presença de ligações de hidrogênio entre os contatos desta possível interface entre LBD e DBD, similar aos resultados observados no sistema *MgRXR/MgPPAR*. Outro ponto similar entre este sistema e os outros são as métricas de qualidade geral das estruturas (**Figura 16A, B e C**), sendo possível observar valores e padrões similares de RMSD, RMSF e atividade entre cadeias. No geral os valores são de RMSD e RMSF são bastante baixos e o mapa de contato indica interações principalmente entre os LBD.

Figura 16. Valores médios de RMSD, RMSF e mapa de contatos do sistema *MgRXR/MgNR1P10*.



A) Valores médios de RMSD e B) RMSF de *MgRXR* e *MgNR1P10* em azul e laranja, respectivamente. C) Mapa de contatos ilustrando a distância média entre resíduos de *MgRXR* e *MgNR1P10*. Fonte: elaborado pelo autor (2025).

Novamente é necessário levar em consideração a ausência da molécula de DNA na composição do sistema e a importância de suas alterações na disposição espacial dos DBD. Porém, como mencionado, há uma proximidade evolutiva entre NR1P e NR1C em termos de sequência, que pode ser refletida também na conservação de padrões na estrutura terciária. Estes resultados, aliados aos anteriores, mostram a diversidade de regiões de contato existentes entre receptores nucleares durante a dimerização e trazem protagonismo para o DBD, até então tido como coadjuvante neste processo. No entanto, estes

achados também mostram como cada dímero é particular e possui características específicas, tanto em relação aos seus domínios conservados como na região da alça, por exemplo, requerendo ajustes finos em muitos casos, como a inserção de restrições de posição.

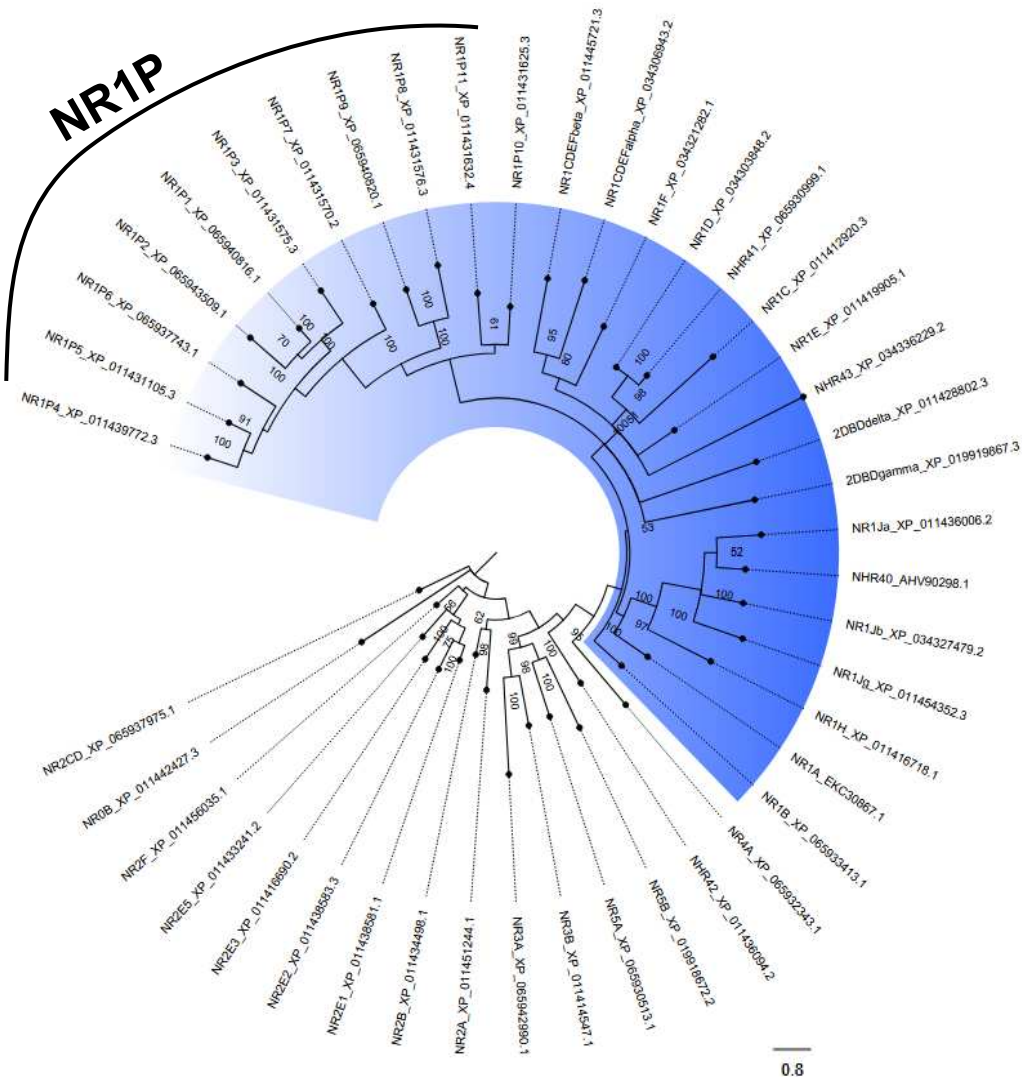
4.4 CARACTERIZAÇÃO DO SUBGRUPO NR1P

O subgrupo NR1P foi descoberto em 2014 e, desde então, apenas três estudos focaram na sua caracterização (Vogeler 2016; Vogeler, 2014; Miglioli, 2024). O estudo de Vogeler e colaboradores (2014) identificou 43 NR em *M. gigas*, dos quais 11 foram agrupados em um clado inédito e monofilético, situado dentro da subfamília NR1 que agrupa a maioria dos receptores parceiros de RXR. Posteriormente, foram descobertos 47 NR em *M. galloprovincialis*, incluindo 14 receptores classificados como NR1P por homologia àqueles de *M. gigas* (Miglioli, 2024). Interessantemente, neste último estudo também foram utilizadas sequências do genoma de *Mytilus coruscus* como referência para a busca e validação dos genes de *M. galloprovincialis* e para as análises filogenéticas, onde é possível observar a presença de 11 genes identificados como NR1P. Estes dados vão ao encontro da hipótese de que o subgrupo NR1P é exclusivo de moluscos bivalves e, visto que diversas espécies de bivalves possuem seus genomas sequenciados atualmente, talvez seja possível identificar mais ortólogos de NR1P. Adicionalmente, análises de perfil transcricional foram feitas para *M. gigas* e *M. galloprovincialis* onde é possível observar a co-expressão de receptores do subgrupo NR1P e RXR (Vogeler, 2016; Miglioli 2024).

Outros estudos fazem menção ao subgrupo NR1P no que diz respeito à sua afinidade pelo estradiol ou à modulação da sua expressão gênica mediante a exposição por contaminantes ambientais como benzofenona (Saldaña, 2025; Madaloz, 2023). No entanto, informações sobre a caracterização estrutural deste subgrupo ainda não foram reportadas, como a disposição de seus domínios e estruturas secundárias, bem como de seus sítios ativos e possíveis ligantes. Assim, as sequências primárias filtradas (contendo as regiões do DBD, alça e LBD) e os modelos tridimensionais dos LBD de receptores do subgrupo NR1P foram escolhidos para as análises subsequentes.

Embora uma análise filogenética dos NR de *M. gigas* tenha sido realizada previamente, inclusive com sequências de outras espécies de invertebrados e vertebrados (Vogeler, 2014), optamos por refazê-la utilizando sequências do genoma mais recente de *M. gigas* (**Figura 17**). Os resultados desta nova análise filogenética sustentam a posição do subgrupo NR1P como monofilético e sendo parte da subfamília NR1. Ainda, as relações internas do subgrupo são mantidas na mesma disposição reportada por Vogeler e colaboradores, enquanto a posição dos demais subgrupos difere, principalmente pela inclusão do subgrupo 2DBD e dos receptores NHR40 e NHR41 na subfamília NR1. Em relação ao restante das subfamílias, é possível observar a posição de NR2 nos ramos mais basais da árvore filogenética, de acordo com as evidências de que esta é a subfamília mais antiga de NR (Beinsteiner, 2023), com NR2A e NR2B (*MgHNF4* e *MgRXR*) formando um grupo único. Acredita-se que o receptor nuclear ancestral comum possuía a habilidade de realizar homodimerização devido a NR2A e NR2B poderem fazê-lo e por estes subgrupos serem os mais antigos e conservados na história evolutiva da superfamília de receptores nucleares (Beinsteiner, 2021; 2023). Estudos sugerem que a heterodimerização tenha sido facilitada pelo motivo estrutural π -turn citado anteriormente, que evoluiu paralelamente ao processo de dimerização como um todo em NR (Beinsteiner, 2021). Este tipo de hélice é termodinamicamente menos estável que α -hélices comuns e está geralmente associado a vantagens evolutivas como facilitação de interações com ligantes ou outras proteínas, por exemplo (Beinsteiner, 2021). De fato, este parece ser o principal ganho no surgimento deste motivo nos NR, por favorecer uma estrutura mais simétrica e estável entre os homodímeros ancestrais de RXR e HNF4 e, posteriormente, auxiliando na estabilização dos heterodímeros de RXR, mais assimétricos (Beinsteiner, 2021).

Figura 17. Filogenia molecular dos receptores nucleares de *M. gigas*.



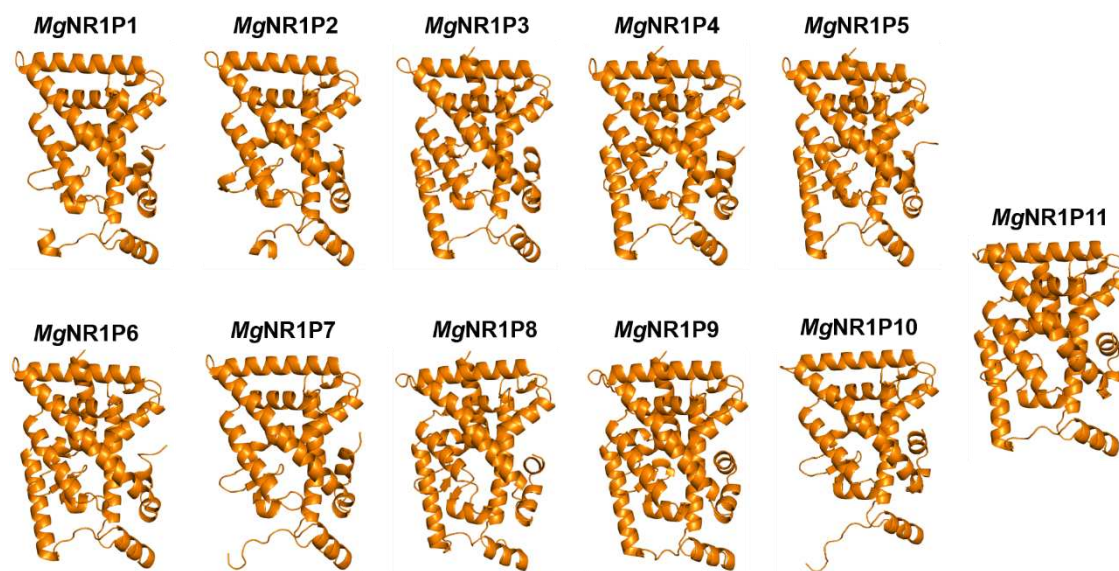
Análise filogenética baseada nas sequências filtradas dos NR de *M. gigas*. Números entre os nós indicam a probabilidade de cada grupo ser formado e pontos indicam ramos referentes a cada NR. A subfamília NR1 está destacada em azul. Fonte: elaborado pelo autor (2025).

Entre os principais parceiros de RXR estão os receptores da subfamília NR1, que apresentam uma história evolutiva bastante diversa entre vertebrados e invertebrados (Miglioli, 2024; Beinstainer, 2021; Cruzeiro, 2016; Castro, 2012). Alguns autores sugerem que o subgrupo NR1I/J, por exemplo, tenha um ancestral comum urbilatério submetido a diferentes processos evolutivos como deleção dos receptores NR1I e conservação de NR1J em animais protostômios e um processo inverso para a maioria dos animais deuterostômios, como humanos que possuem apenas NR1I (VDR, CAR, PXR; Cruzeiro, 2016). Um processo similar parece ter ocorrido com o subgrupo NR1P, havendo um possível episódio de expansão de linhagem entre o receptor ancestral comum

de NR1C, NR1D, NR1E e NR1F nos moluscos bivalves, gerando um novo subgrupo (Vogeler, 2014; Miglioli, 2024).

Para tentar investigar as características deste subgrupo, principalmente no que diz respeito ao seu LBD, sítios ativos e possíveis ligantes, foram utilizados modelos tridimensionais do LBD de cada um dos onze NR do subgrupo NR1P (**Figura 18**). Ao analisar cada domínio separadamente, observamos que a disposição de hélices deste subgrupo não segue exatamente àquela da maioria dos NR. A principal diferença está nas hélices 3 e 12 que compõem o AF-2, ou domínio de ativação de função, essencial para a ativação do receptor e entrada do ligante no sítio ativo (Spyrakis, 2012). No subgrupo NR1P, todos os receptores apresentam uma H3 com a porção proximal mais próxima à H12, limitando o espaço de movimento da H12 e, conseqüentemente, o caminho de entrada para o ligante. Ao alinhar a região proximal da H3 com outros NR, é possível observar a presença de resíduos conservados específicos de NR1P formando um motivo TxxDxDxR compreendendo o loop entre H2 e H3 e o início da H3 (**Figura 19**).

Figura 18. Modelos tridimensionais do LBD de receptores nucleares do subgrupo NR1P de *M. gigas*.



Mg = *Magallana gigas*. Fonte: elaborado pelo autor (2025).

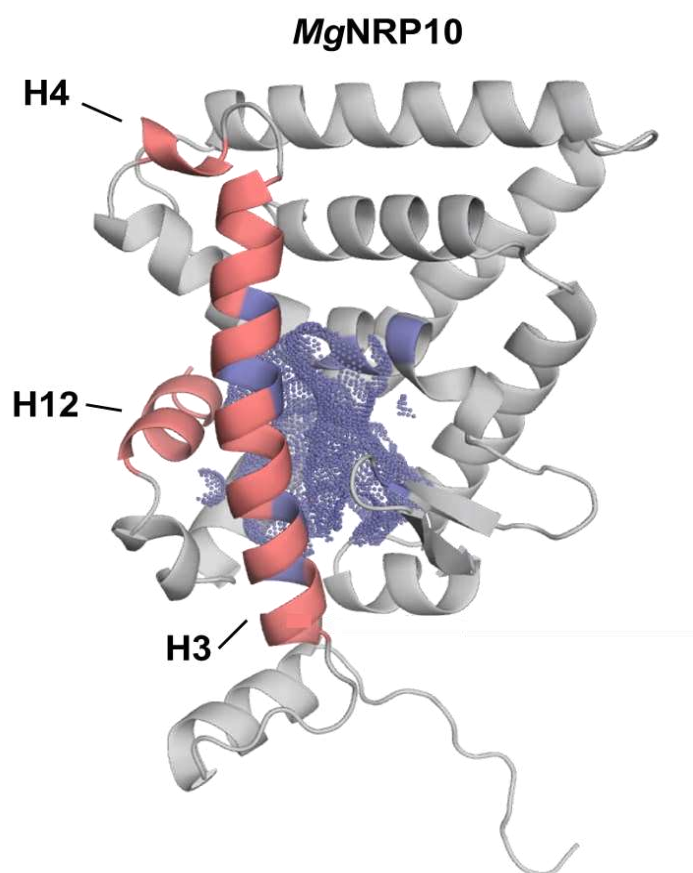
Figura 19. Alinhamento múltiplo de sequências de receptores nucleares da subfamília NR1 de *M. gigas*.

	470	480	490	500	510
NR1P8/1-484	-----	-----	-----	I K S L L D K E E F E K V Y R A T G I D V D G R L E F -	-----
NR1P6/1-514	-----	-----	-----	N L D P I P T K E Y N E I F E K T N L D V D S R T Q N -	-----
NR1P7/1-485	-----	-----	-----	P L N A L S P E E Y N A I Y E K T Q I D V D G R K Q H -	-----
NR1P4/1-428	-----	-----	-----	N M D P V P A G V Y S E I Y S N T Q M D I D G R M E I -	-----
NR1P5/1-520	-----	-----	-----	S M D P L P A N V Y N E I F N I T N M D I D G R M K I -	-----
NR1P2/1-464	-----	-----	-----	K L E Y L P T D Q Y N D I Y E K T Q L D V D G R K K L -	-----
NR1P3/1-502	-----	-----	-----	K M E Y L S I Q E Y K E I Y D K T K I D V D G R K E L -	-----
NR1P1/1-476	-----	-----	-----	N L E F L P A D E Y N E I Y E K T Q V D V D G R M K L -	-----
NR1P9/1-601	-----	-----	-----	L K K L L S T E E F N S F Y L A T G I D V D G R M E M -	-----
NR1P10/1-390	-----	-----	-----	S M N N L S K E E Y Q Q F Y S Q T G I D L D N R M M L -	-----
NR1P11/1-380	-----	-----	-----	T M S V I P K E V W D E F Y K E T G I D L D N R Q D L -	-----
NR1C_Peroxisome/1-328	-----	-----	-----	H L K M T C E N D -	-----
NR1Ja/1-386	-----	-----	-----	A V F D Q P Y N E T Q Y K K L T E S P R S -	-----
NR1Jb/1-351	-----	-----	-----	-----	-----
NR1Jg/1-484	-----	-----	-----	E L R I A H E Q S S F L A N T T T S I K A P P T N -	-----
NR1H_EcR/1-388	-----	-----	-----	K F E F P E E S K V N N A I D I D T A K E E S G E H V -	-----
NR1CDEFalpha/1-617	-----	-----	-----	A N D R F A E K M A S G D L P D F S K L V I N P N S I -	-----
NR2A_HNF4/1-321	-----	-----	-----	-----	-----
NR2B_RXR/1-319	-----	-----	-----	-----	-----
NR1A_Thyroid/1-347	-----	-----	-----	P V S Q E D L I S S D P E A -	-----
NR1B_Retinoic/1-347	-----	-----	-----	D E A S K E N S N Q K K E K S M L -	-----
NR1F/1-420	-----	-----	-----	I Q S S I T P Q E I I D A F K T M T R S Q L -	-----
NR1CDEFbeta/1-678	A V V A P I V K P N I Y I H R E E D T K L T I P L L T F K I H E S F N Q T F T F L K L Q L N H M E E I L				
NR1D_Rev-erb-like/1-364	-----	-----	-----	Q M A C P L N G N V T Q S S E D T Q G -	-----
NR1E/1-360	-----	-----	-----	Q I S I P S T P V Q F S D E E L E A Q K L S M -	-----

Motivo conservado TxxDxDxR de receptor nucleares do subgrupo NR1P. Resíduos conservados estão destacados em azul. Fonte: elaborado pelo autor (2025).

Após a identificação desta configuração de hélices distinta dos demais NR, os modelos dos LBD de *MgNR1P* foram submetidos à análise de predição de sítios ativos. Foi possível identificar sítios ativos em todos os modelos, situados no bolsão entre H3, H4 e H12, indicando que apesar da diferença na posição das hélices, a localização dos sítios ativos foi mantida (**Figura 20**). Além da posição do sítio ativo e de quais resíduos os compõem, seus volumes também foram calculados e estão listados na **Tabela 2**. É preciso ressaltar que apesar destes receptores possuírem sítios ativos prováveis, isto não necessariamente reflete a ligação com um ligante, como é o caso do ER de *M. gigas* que possui um sítio ativo identificado e caracterizado, mas que não comporta a molécula de estradiol e, portanto, não apresenta a mesma função que seus ortólogos vertebrados (Bridgham, 2014; Matsumoto, 2007). Apesar disso, foram feitas análises de atracamento molecular entre os onze receptores de *MgNR1P* tanto com ligantes endógenos conhecidos de outros receptores da subfamília NR1 e NR3 como ácido retinoico, ácido araquidônico e estradiol, quanto com ligantes xenobióticos como benzofenona.

Figura 20. Predição da posição do sítio ativo de *MgNR1P10*.



Posição do sítio ativo de *MgNR1P10*. O volume do sítio ativo está representado em azul, α -hélices que compõem o AF-2 estão representadas em vermelho e numeradas, enquanto o restante do LBD está em cinza. Fonte: elaborado pelo autor (2025).

Tabela 2. Volume predito do sítio ativo de NR do subgrupo NR1P.

Receptor nuclear	Volume ($\text{\AA}^3$)
MgNR1P1	3021,1
MgNR1P2	2168,9
MgNR1P3	2390,5
MgNR1P4	1739,7
MgNR1P5	2017
MgNR1P6	1788,4
MgNR1P7	2621,5
MgNR1P8	3096,3
MgNR1P9	1808,2
MgNR1P10	3257
MgNR1P11	2628,4

Os resultados listados na **Tabela 3** mostram os valores de ΔG que reflete a estabilidade do complexo ligante-receptor e os valores da pontuação do algoritmo Attracting Cavities 2.0 que mede a energia livre do sistema ligante-receptor-solvente utilizando o campo de força CHARMM (Röhrig, 2023). Os receptores clássicos dos ligantes endógenos foram utilizados como controles positivos para o subgrupo NR1P, exceto para benzofenona que não é um ligante endógeno. Nota-se que os valores mais baixos de energia livre encontrados para o subgrupo NR1P concentram-se nos atracamentos com o ácido araquidônico, ligante endógeno do PPAR, pertencente ao grupo irmão NR1C. Outro ponto curioso é o fato de receptores de NR1P apresentarem uma suposta afinidade maior pelo ácido araquidônico do que os próprios PPAR, fenômeno observado também nos atracamentos com o ácido retinóico em comparação com RXR e RAR. Este comportamento só não é observado nos atracamentos com estradiol, visto que o *MgER* realmente não interage com este ligante. Em relação à benzofenona, todos os receptores apresentaram pontuações mais baixas quando comparadas aos outros ligantes, o que pode sugerir que apesar deste xenobiótico ser teoricamente comportado no sítio ativo dos NR, ele não necessariamente se liga a eles e os ativa.

Tabela 3. Valores das funções de energia do atracamento molecular entre receptores nucleares e ligantes

Receptor nuclear	ΔG				AC-score			
	ACD	EST	REA	BZQ	ACD	EST	REA	BZQ
<i>HsRXR</i>	X	X	-6.99	-6,25	X	X	-42.99	29,73
<i>HsRAR</i>	X	X	-7.68	-6,99	X	X	-46.21	22,55
<i>HsPPAR</i>	-8.55	X	X	-6,84	-51.70	X	X	24,5
<i>HsER</i>	X	-8.27	X	-6,99	X	-18.84	X	20,68
<i>MgRXR</i>	X	X	-7.29	-6,84	X	X	-38.28	22,16
<i>MgRAR</i>	X	X	-8.53	-6,86	X	X	-58.51	22,47
<i>MgPPAR</i>	-7.39	X	X	X	-42.85	X	X	X
<i>MgER*</i>	X	-6.63	X	X	X	-0.32	X	X
<i>MgNR1P1</i>	-8.27	-7.42	-8.13	-6,67	-54.44	-5.78	-40.86	25,53
<i>MgNR1P2</i>	-8.26	-7.33	-8.12	-6,47	-54.07	-7.51	-36.77	26,24
<i>MgNR1P3</i>	-8.21	-7.47	-8.31	-6,57	-56.47	-10.14	-39.83	24,95
<i>MgNR1P4</i>	-8.62	-7.43	-8.53	-6,62	-57.22	-7.62	-40.38	25,01

<i>MgNR1P5</i>	-8.74	-7.42	-8.56	-6,81	-61.04	-7.83	-44.07	22,43
<i>MgNR1P6</i>	-7.72	-6.94	-7.95	-6,24	-49.38	-1.68	-29.87	29,41
<i>MgNR1P7</i>	-7.56	-6.95	-7.59	-6,27	-46.37	-2.66	-28.99	28,53
<i>MgNR1P8</i>	-7.94	-7.15	-7.85	-6,46	-45.29	-3.26	-29.63	27,24
<i>MgNR1P9</i>	-8.64	-7.17	-8.25	-6,68	-51.47	-2.91	-35.96	24,5
<i>MgNR1P10</i>	-9.02	-7.56	-8.30	-7	-58.27	-4.37	-30.61	21,79
<i>MgNR1P11</i>	-8.70	-7.13	-8.17	-6,72	-60.15	-4.09	-35.65	23,44

Novamente é importante ressaltar que estes são resultados obtidos *in silico*, baseados em modelos e que experimentos mais precisos são necessários para confirmá-los. Assim, com base nos resultados aqui apresentados, sugere-se a inclusão dos receptores nucleares do subgrupo NR1P em estudos de ecotoxicologia como fizeram Madaloz et al. (2023) e Saldaño et al. (2025) dado o cenário atual de contaminação aquática por compostos xenobióticos. Além disso, a continuação de estudos *in silico* utilizando dinâmica molecular com ligantes pode trazer informações relevantes em relação ao comportamento das hélices 3 e 12 e a entrada de ligantes nos sítios ativos que estas estruturas controlam, contribuindo para o melhor entendimento da interação não só dos receptores nucleares entre si e sua ativação, mas também o seu comportamento quando exposto a um contaminante ou ligante desconhecido.

5. CONCLUSÃO

O presente trabalho buscou investigar a interação entre receptores nucleares de *Homo sapiens* e *Magallana gigas* durante o processo de dimerização, característica fundamental do seu comportamento frente à exposição a ligantes que os ativam. Dado o atual cenário de contaminação aquática por compostos orgânicos como desreguladores endócrinos, faz-se necessário o entendimento de como estes receptores interagem entre si para manter sua atividade de modulação da expressão gênica. As espécies escolhidas foram baseadas tanto pela disponibilidade de dados e possibilidade de uso como controle no caso de *H. sapiens*, como pela relevância ecológica, científica e econômica no caso *M. gigas*. Adicionalmente, foi dado enfoque para o receptor retinoide X que possui a capacidade de realizar homo e

heterodimerização e por ser o principal parceiro de interação com a maioria dos NR.

Para realizar os objetivos deste trabalho, foram utilizadas apenas ferramentas *in silico* como modelagem estrutural e simulação de dinâmica molecular a partir de sequências de ambas as espécies disponíveis em bancos de dados públicos. Além disso, metodologias como atracamento e filogenia molecular foram empregadas a fim de avançar o processo de caracterização de um subgrupo de NR específico de moluscos bivalves, como *M. gigas*.

Como resultado do uso dessas técnicas, foi possível criar modelos de dímeros entre RXR e demais NR contendo ambos os domínios conservados, DBD e LBD, em conjunto com a região da alça que os conecta, para todos os receptores de ambas as espécies. Além disso, este é o primeiro registro de dinâmica molecular *coarse-grained* de sistemas completos de NR, utilizando inclusive moléculas de DNA contendo elementos de resposta específicos. Ao todo foram simulados cinco sistemas de RXR e receptores parceiros, sendo dois para *H. sapiens* e três para *M. gigas*, incluindo um sistema entre RXR e o subgrupo exclusivo de bivalves NR1P. Foi possível identificar interfaces de interação conhecidas em todos os sistemas, assim como interfaces inéditas em sistemas específicos. Ademais, foi realizada uma breve caracterização dos onze NR do subgrupo NR1P de *M. gigas*, sendo possível identificar uma nova configuração de α -hélices em seus LBD e um motivo conservado exclusivo de NR1P que pode estar envolvido neste achado. Por fim, análises de atracamento molecular com ligantes endógenos e xenobióticos conhecidos de NR foram feitas com receptores do subgrupo NR1P a fim de identificar possíveis ligantes que estes receptores possam ter afinidade.

Com base nos resultados apresentados, recomenda-se a inclusão dos NR estudados, principalmente o subgrupo NR1P, em estudos de ensaios ecotoxicológicos e de hibridização, a fim de averiguar o efeito de contaminantes na sua atividade e processo de dimerização, e confirmar a presença dos dímeros e interfaces identificadas *in silico*, respectivamente.

REFERÊNCIAS

- ABRAHAM, Mark James et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, v. 1-2, p. 19-25, 2015. DOI: 10.1016/j.softx.2015.06.001.
- ABRAMSON, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, v. 630, p. 493-501, 2024. <https://doi.org/10.1038/s41586-024-07487-w>.
- ÁLVAREZ-SALMORAL, Daniel et al. AlphaBridge: tools for the analysis of predicted macromolecular complexes. *bioRxiv*, 2024. DOI: 10.1101/2024.10.23.619601.
- ANDRÉ, A. et al. The retinoic acid receptor (RAR) in molluscs: Function, evolution and endocrine disruption insights. *Aquatic Toxicology*, v. 208, 2019. <https://doi.org/10.1016/j.aquatox.2019.01.002>.
- BALL, Linda J. et al. Proline-Rich Motifs by Protein-Protein Interaction Domains. *Angewandte Chemie International Edition*, v. 44, p. 2852-2869, 2005. DOI: 10.1002/anie.200400618.
- BEINSTEINER, B. et al. A structural signature motif enlightens the origin and diversification of nuclear receptors. *PLoS Genetics*, v. 17, n. 4, e1009492, 2021. <https://doi.org/10.1371/journal.pgen.1009492>.
- BEINSTEINER, B.; BILLAS, I. M. L.; MORAS, D. Structural insights into the HNF4 biology. *Frontiers in Endocrinology*, v. 14, 1197063, 2023. <https://doi.org/10.3389/fendo.2023.1197063>.
- BORTOLI, Leonardo de. Filogenia de genes da superfamília dos receptores nucleares em espécies aquáticas. 2022. Trabalho de Conclusão de Curso (Licenciatura em Ciências Biológicas) – Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Florianópolis, 2022.
- BOURGUET, William et al. Signature of the oligomeric behaviour of nuclear receptors at the sequence and structural level. *EMBO Reports*, v. 5, n. 4, p. 423-429, 2004. DOI: 10.1038/sj.embor.7400119.
- BOUTON, D. et al. A conserved retinoid X receptor (RXR) from the mollusk *Biomphalaria glabrata* transactivates transcription in the presence of retinoids. *Journal of Molecular Endocrinology*, v. 34, n. 3, p. 567-582, 2005. DOI: 10.1677/jme.1.01766.
- BRASCHER, Theo Cardozo. Caracterização dos domínios de ligação ao DNA dos receptores nucleares de *Crassostrea gigas* e a regulação da transcrição de genes de citocromos P450: um estudo in silico. 2024. Tese (Doutorado em Bioquímica) – Programa de Pós-Graduação em Bioquímica, Universidade Federal de Santa Catarina, Florianópolis, 2024.

BRIDGHAM, J. T. et al. Protein evolution by molecular tinkering: Diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS Biology*, v. 8, n. 10, e1000497, 2010. <https://doi.org/10.1371/journal.pbio.1000497>.

BRIDGHAM, Jamie T. et al. Vestigialization of an Allosteric Switch: Genetic and Structural Mechanisms for the Evolution of Constitutive Activity in a Steroid Hormone Receptor. *PLoS Genetics*, v. 10, n. 1, e1004058, 2014. DOI: 10.1371/journal.pgen.1004058.

BUGNON, M. et al. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Research*, v. 52, W1, p. W324-W332, 2024. <https://doi.org/10.1093/nar/gkae300>.

CAJARAVILLE, M. P. et al. Peroxisome proliferation as a biomarker in environmental pollution assessment. *Microscopy Research and Technique*, v. 61, n. 2, p. 191-202, 2003. <https://doi.org/10.1002/jemt.10329>.

CALLAWAY, E. Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures. *Nature*, 2024. <https://doi.org/10.1038/d41586-024-03214-7>

CAMACHO, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics*, v. 10, 421, 2009. <https://doi.org/10.1186/1471-2105-10-421>.

CAMPO-PAYSAA, Florent et al. Retinoic Acid Signaling in Development: Tissue-Specific Functions and Evolutionary Origins. *genesis*, v. 46, n. 11, p. 640-656, 2008. DOI: 10.1002/dvg.20444.

CANESTRO, Cristian et al. Is retinoic acid genetic machinery a chordate innovation? *Evolution & Development*, v. 8, n. 5, p. 394-406, 2006. DOI: 10.1111/j.1525-142X.2006.00112.x.

CAPITÃO, A. M. F. et al. An ancestral nuclear receptor couple, PPAR-RXR, is exploited by organotins. *Science of The Total Environment*, v. 787, 149044, 2021. Disponível em: <https://doi.org/10.1016/j.scitotenv.2021.149044>.

CASTRO, L. F. C. et al. Imposex induction is mediated through the Retinoid X Receptor signalling pathway in the neogastropod *Nucella lapillus*. *Aquatic Toxicology*, v. 85, n. 1, p. 57-66, 2007.

CHAN, L. S. A.; WELLS, R. A. Cross-talk between PPARs and the partners of RXR: A molecular perspective. *PPAR Research*, v. 2009, 925309, 2009. <https://doi.org/10.1155/2009/925309>.

CHAN, Y. H.; CHU, K. H.; CHAN, K. M. Ecdysteroid-mimicking compounds act as both agonists and antagonists to the crustacean ecdysone receptor. *Chemosphere*, v. 236, 124551, 2019. <https://doi.org/10.1016/j.chemosphere.2019.124551>.

CHANDRA, Vikas et al. Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA. *Nature*, v. 456, n. 7220, p. 350-356, 2008. DOI: 10.1038/nature07413.

CHOI, W. J.; HARATIPOUR, Z.; BLIND, R. D. Full-length nuclear receptor allosteric regulation. *Journal of Lipid Research*, v. 64, n. 8, 100406, 2023. <https://doi.org/10.1016/j.jlr.2023.100406>.

CRUZEIRO, C. et al. A mollusk VDR/PXR/CAR-like (NR1J) nuclear receptor provides insight into ancient detoxification mechanisms. *Aquatic Toxicology*, v. 174, p. 61–69, 2016. <https://doi.org/10.1016/j.aquatox.2016.02.007>.

DELFOSSÉ, V. et al. A structural perspective on nuclear receptors as targets of environmental compounds. *Acta Pharmacologica Sinica*, v. 36, n. 1, p. 88-101, jan. 2015. <https://doi.org/10.1038/aps.2014.133>.

DOMINGUEZ, C.; BOELEN, R.; BONVIN, A. M. J. J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *Journal of the American Chemical Society*, v. 125, n. 7, p. 1731-1737, 2003. <https://doi.org/10.1021/ja026939x>.

EPAGRI/CEPA. Síntese Anual da Agricultura de Santa Catarina 2021-2022. Florianópolis: Epagri/Cepa, 2023.

ESCRIVA, Hector et al. Neofunctionalization in Vertebrates: The Example of Retinoic Acid Receptors. *PLoS Genetics*, v. 2, n. 7, e102, 2006. DOI: 10.1371/journal.pgen.0020102.

EVANS, R. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021. <https://doi.org/10.1101/2021.10.04.463034>.

EVANS, R. M.; MANGELSDORF, D. J. Nuclear Receptors, RXR, and the Big Bang. *Cell*, v. 157, n. 1, p. 255-266, 2014. <https://doi.org/10.1016/j.cell.2014.03.012>.

FAHRBACH, S. E.; SMAGGHE, G.; VELARDE, R. A. Nuclear receptors in the insect endocrine system. *Annual Review of Entomology*, v. 57, p. 83-106, 2012. <https://doi.org/10.1146/annurev-ento-120710-100607>.

FAO. The State of World Fisheries and Aquaculture 2020. Sustainability in action. Rome: Food and Agriculture Organization of the United Nations, 2020. <https://doi.org/10.4060/ca9229en>.

FONSECA, Elza S. S. et al. An Orthologue of the Retinoic Acid Receptor (RAR) Is Present in the Ecdysozoa Phylum Priapulida. *Genes*, v. 10, n. 12, 985, 2019. DOI: 10.3390/genes10120985.

GORE, A. C. et al. Executive Summary to EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocrine*

Reviews, v. 36, n. 6, p. 593-602, dez. 2015. <https://doi.org/10.1210/er.2015-1093>.

GUTIERREZ-MAZARIEGOS, J. et al. A mollusk retinoic acid receptor (RAR) ortholog sheds light on the evolution of ligand binding. *Endocrinology*, v. 155, n. 11, p. 4275-4286, 2014. <https://doi.org/10.1210/en.2014-1181>.

GUVENCH, O.; MACKERELL JR., A. D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In: *Molecular Modeling of Proteins*. Humana Press, 2008. p. 63-88. (Methods in Molecular Biology, v. 443).

HARDING, H. P. et al. Transcriptional activation and repression by ROR α , an orphan nuclear receptor required for cerebellar development. *Molecular Endocrinology*, v. 11, n. 11, p. 1737–1746, 1997.

HOSPITAL, A. et al. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, v. 8, p. 37-47, 2015. <https://doi.org/10.2147/AABC.S70333>.

HUNTER, John D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, v. 9, n. 3, p. 90-95, 2007. DOI: 10.1109/MCSE.2007.55.

INOUE, Daisuke; SEI, Kazunari; IKE, Michihiko. Disruption of Retinoic Acid Receptor Signaling by Environmental Pollutants. *Journal of Health Science*, v. 56, n. 3, p. 221-230, 2010.

ISSEMAN, I.; GREEN, S. Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. *Nature*, v. 347, p. 645-650, 1990. DOI: 10.1038/347645a0.

JETTEN, A. M.; KANG, H. S.; TAKEDA, Y. Retinoic acid-related orphan receptors α and γ : key regulators of lipid/glucose metabolism, inflammation, and insulin sensitivity. *Frontiers in Endocrinology*, v. 4, art. 1, jan. 2013.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, v. 596, n. 7873, p. 583-589, 2021. <https://doi.org/10.1038/s41586-021-03819-2>.

KAUR, Satwant et al. The Nuclear Receptors of *Biomphalaria glabrata* and *Lottia gigantea*: Implications for Developing New Model Organisms. **PLoS ONE**, v. 10, n. 4, e0121259, 2015. DOI: 10.1371/journal.pone.0121259.

KAY, Brian K.; WILLIAMSON, Michael P.; SUDOL, Marius. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal*, v. 14, n. 2, p. 231-241, 2000. DOI: 10.1096/fasebj.14.2.231.

KRIVÁK, Radoslav; HOKSZA, David. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure.

Journal of Cheminformatics, v. 10, n. 1, p. 39, 2018. Disponível em:
<https://doi.org/10.1186/s13321-018-0285-8>.

LARANJEIRO, F. et al. Triphenyltin induces imposex in *Nucella lapillus* through an aphallic route. *Aquatic Toxicology*, 2016.
<https://doi.org/10.1016/j.aquatox.2016.03.005>.

LE, Si Quang; GASCUEL, Olivier. An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, v. 25, n. 7, p. 1307-1320, 2008. DOI: 10.1093/molbev/msn067.

LEFEBVRE, P.; BENOMAR, Y.; STAEELS, B. Retinoid X receptors: common heterodimerization partners with distinct functions. *Trends in Endocrinology and Metabolism*, v. 21, n. 11, p. 676-683, 2010.
<https://doi.org/10.1016/j.tem.2010.06.009>.

LI, F.; SONG, C.; ZHANG, Y.; WU, D. Structural overview and perspectives of the nuclear receptors, a major family as the direct targets for small-molecule drugs. *Acta Biochimica et Biophysica Sinica*, v. 54, n. 1, p. 12-24, 2022.
<https://doi.org/10.3724/abbs.2021001>.

LIONETTO, M. G.; CARICATO, R.; GIORDANO, M. E. Pollution biomarkers in environmental and human biomonitoring. *The Open Biomarkers Journal*, v. 9, p. 1-9, 2019. <https://doi.org/10.2174/1875318301909010001>.

MADALUZ, T. Z. et al. Nuclear receptor superfamily structural diversity in pacific oyster: In silico identification of estradiol binding candidates. *Chemosphere*, v. 339, p. 139877, 2023.

MALETTA, M. et al. The palindromic DNA-bound USP/EcR nuclear receptor adopts an asymmetric organization with allosteric domain positioning. *Nature Communications*, v. 5, 4139, 2014. <https://doi.org/10.1038/ncomms5139>.

MARRINK, S. J. et al. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, v. 111, n. 27, p. 7812-7824, 2007. <https://doi.org/10.1021/jp071097f>.

MATSUMOTO, T. et al. (2007). Oyster estrogen receptor: cDNA cloning and immunolocalization. *General and Comparative Endocrinology*, 150(2), 313–322.
<https://doi.org/10.1016/j.ygcen.2007.01.016>

McCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamics of folded proteins. *Nature*, v. 267, p. 585-590, 1977. <https://doi.org/10.1038/267585a0>.

McCARTY, L. S.; MUNKITTRICK, K. R. Environmental biomarkers in aquatic toxicology: Fiction, fantasy, or functional? *Human and Ecological Risk Assessment*, v. 2, n. 2, p. 268-274, 1996. <https://doi.org/10.1080/10807039609383607>.

MCKINNEY, Wes. pandas: a Foundational Python Library for Data Analysis and Statistics. In: *PYTHON IN SCIENCE CONFERENCE (SCIPY)*, 2011. *Proceedings...* [S.l.: s.n.], 2011. p. 56-61.

MICHAUD-AGRAWAL, Naveen et al. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, v. 32, n. 10, p. 2319-2327, 2011. DOI: 10.1002/jcc.21787.

MIGLIOLI, A. et al. First characterization of the nuclear receptor superfamily in the Mediterranean mussel *Mytilus galloprovincialis*: developmental expression dynamics and potential susceptibility to environmental chemicals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 379, 20220500, 2024. <https://doi.org/10.1098/rstb.2022.0500>.

MIGLIOLI, A. et al. Nuclear receptors and development of marine invertebrates. *Genes*, v. 12, n. 1, 83, 2021. <https://doi.org/10.3390/genes12010083>.

MINH, Bui Quang et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, v. 37, n. 5, p. 1530-1534, 2020. DOI: 10.1093/molbev/msaa015.

MROWICKI, R. et al. The genome sequence of the Pacific oyster, *Magallana gigas* (Thunberg, 1793). *Wellcome Open Research*, v. 9, 284, 2025. <https://doi.org/10.12688/wellcomeopenres.22255.2>.

NADAL, Angel; DÍAZ, Mario; VALVERDE, Miguel A. The Estrogen Trinity: Membrane, Cytosolic, and Nuclear Effects. *News in Physiological Sciences*, v. 16, p. 251-255, 2001.

NISHIKAWA, Jun-ichi et al. Involvement of the Retinoid X Receptor in the Development of Imposex Caused by Organotins in Gastropods. *Environmental Science & Technology*, v. 38, n. 23, p. 6271-6276, 2004. DOI: 10.1021/es049593u.

OLIPHANT, Travis E. *Guide to NumPy*. USA: Trelgol Publishing, 2006.

OMIECINSKI, Curtis J. et al. Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. *Toxicological Sciences*, v. 120, Suppl. 1, p. S49-S75, 2011. DOI: 10.1093/toxsci/kfq338.

OROZCO, M. et al. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. *Advances in Protein Chemistry and Structural Biology*, v. 85, p. 183-212, 2011. <https://doi.org/10.1016/B978-0-12-386485-7.00005-3>.

PAIVA, V. A. et al. Protein structural bioinformatics: An overview. *Computers in Biology and Medicine*, v. 147, 105695, 2022. <https://doi.org/10.1016/j.compbiomed.2022.105695>.

PASCOAL, Sonia et al. Transcriptomics and in vivo tests reveal novel mechanisms underlying endocrine disruption in an ecological sentinel, *Nucella lapillus*. *Molecular Ecology*, v. 22, n. 6, p. 1589-1608, 2013. DOI: 10.1111/mec.12137.

PAWLAK, M.; LEFEBVRE, P.; STAELS, B. General molecular biology and architecture of nuclear receptors. *Current Topics in Medicinal Chemistry*, v. 12, n. 6, p. 486-504, 2012. <https://doi.org/10.2174/156802612799436641>.

PUJALS, Sílvia; GIRALT, Ernest. Proline-rich, amphipathic cell-penetrating peptides. *Advanced Drug Delivery Reviews*, v. 60, p. 473-484, 2008. DOI: 10.1016/j.addr.2007.09.012.

RAMBAUT, A. FigTree, a graphical viewer of phylogenetic trees. [S.I.], 2007. Disponível em: <http://tree.bio.ed.ac.uk/software/figtree/>.

REIS, I. M. M. dos et al. Differential responses in the biotransformation systems of the oyster *Crassostrea gigas* (Thunberg, 1789) elicited by pyrene and fluorene: Molecular, biochemical and histological approach – Part II. *Aquatic Toxicology*, 105565, 2020. <https://doi.org/10.1016/j.aquatox.2020.105565>.

RINKEVICH, Yuval et al. Systemic bud induction and retinoic acid signaling underlie whole body regeneration in the urochordate *Botrylloides leachi*. *PLoS Biology*, v. 5, n. 4, e71, 2007. DOI: 10.1371/journal.pbio.0050071.

RÖHRIG, U. F.; GOULLIEUX, M.; BUGNON, M.; ZOETE, V. Attracting Cavities 2.0: Improving the Flexibility and Robustness for Small-Molecule Docking. *Journal of Chemical Information and Modeling*, v. 63, n. 12, p. 3925-3940, 2023. <https://doi.org/10.1021/acs.jcim.3c00054>.

RONQUIST, Fredrik et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, v. 61, n. 3, p. 539-542, 2012. DOI: 10.1093/sysbio/sys029.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, v. 5, n. 4, p. 725-738, 2010. <https://doi.org/10.1038/nprot.2010.5>.

SALDANA-SERRANO, M. et al. Gene transcription in the oyster *Crassostrea gigas* exposed to environmental concentrations of the UV filter Benzophenone-

3. *Aquatic Toxicology*, v. 279, 107237, 2025.
<https://doi.org/10.1016/j.aquatox.2025.107237>.

SIERK, M. L.; ZHAO, Q.; RASTINEJAD, F. DNA deformability as a recognition feature in the RevErb response element. *Biochemistry*, v. 40, p. 12833–12843, 2001.

SIGRIST, C. J. A. et al. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research*, v. 38, Database issue, p. D161-D166, 2010. <https://doi.org/10.1093/nar/gkp885>.

SOIN, T.; SMAGGHE, G. Endocrine disruption in aquatic insects: a review. *Ecotoxicology*, v. 16, n. 1, p. 83-93, 2007. <https://doi.org/10.1007/s10646-006-0118-9>.

SOUZA, A. et al. Cytotoxicity evaluation of Amoxicillin and Potassium Clavulanate in *Perna perna* mussels. *Ecotoxicology and Environmental Contamination*, v. 11, n. 1, p. 21-26, 2016.
<https://doi.org/10.5132/eec.2016.01.04>.

SOUZA, Paulo C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods*, v. 18, p. 382-388, 2021. DOI: 10.1038/s41592-021-01098-3.

SPYRAKIS, F. et al. (2012). *Molecular Dynamics: a Tool to Understand Nuclear Receptors*. In P. Cozzini & G. E. Kellogg (Eds.), *Computational Approaches to Nuclear Receptors* (RSC Drug Discovery Series No. 30). Royal Society of Chemistry. <https://doi.org/10.1039/9781849735375>

SULLI, G. et al. Pharmacological activation of REV-ERBs is lethal in cancer and oncogene induced senescence. *Nature*, v. 553, n. 7688, p. 351–355, 18 jan. 2018.

SUZEK, Baris E. et al. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, v. 23, n. 10, p. 1282-1288, 2007. DOI: 10.1093/bioinformatics/btm098.

TAM, Justin Z. et al. Analysis of Protein-Protein Interactions for Intermolecular Bond Prediction. *Molecules*, v. 27, n. 19, 6178, 2022. DOI: 10.3390/molecules27196178.

TAUBENHEIM, J.; KORTMANN, C.; FRAUNE, S. Function and Evolution of Nuclear Receptors in Environmental-Dependent Postembryonic Development. *Frontiers in Cell and Developmental Biology*, v. 9, 653792, 2021.
<https://doi.org/10.3389/fcell.2021.653792>.

TAYLOR, Elias; HEYLAND, Andreas. Evolution of non-genomic nuclear receptor function. *Molecular and Cellular Endocrinology*, v. 539, p. 111468, 2022. DOI: 10.1016/j.mce.2021.111468.

URUSHITANI, Hiroshi et al. Cloning and characterization of the retinoic acid receptor-like protein in the rock shell, *Thais clavigera*. *Aquatic Toxicology*, v. 142-143, p. 403-413, 2013. DOI: 10.1016/j.aquatox.2013.09.008.

VASILACHI, I. C. et al. Occurrence and fate of emerging pollutants in water environment and options for their removal. *Water*, v. 13, n. 2, p. 181, 2021. <https://doi.org/10.3390/w13020181>.

VOGELER, S. et al. The nuclear receptor gene family in the Pacific oyster, *Crassostrea gigas*, contains a novel subfamily group. *BMC Genomics*, v. 15, 369, 2014. <https://doi.org/10.1186/1471-2164-15-369>.

VOGELER, Susanne et al. Cloning retinoid and peroxisome proliferator-activated nuclear receptors of the Pacific oyster and in silico binding to environmental chemicals. *PLoS ONE*, v. 12, n. 4, e0176024, 2017. DOI: 10.1371/journal.pone.0176024.

WEIKUM, E. R.; LIU, X.; ORTLUND, E. A. The nuclear receptor superfamily: A structural perspective. *Protein Science*, v. 27, n. 11, p. 1876-1892, 2018. <https://doi.org/10.1002/pro.3496>.

WELTER, V. et al. Recherches sur les ecdystéroïdes présents dans les cocons de la sangsue *Hirudo medicinalis* au cours de l'embryogenèse. *International Journal of Invertebrate Reproduction and Development*, v. 9, n. 3, p. 321-331, 1986.

WNUK, Agnieszka et al. Benzophenone-3 Impairs Autophagy, Alters Epigenetic Status, and Disrupts Retinoid X Receptor Signaling in Apoptotic Neuronal Cells. *Molecular Neurobiology*, v. 55, p. 5059-5074, 2018. DOI: 10.1007/s12035-017-0704-2.

WORLD HEALTH ORGANIZATION; INTERNATIONAL PROGRAMME ON CHEMICAL SAFETY. Biomarkers and risk assessment: concepts and principles. Geneva: World Health Organization, 1993. Available at: <https://iris.who.int/handle/10665/39037>.

XU, Jinrui; ZHANG, Yang. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, v. 26, n. 7, p. 889-895, 2010. DOI: 10.1093/bioinformatics/btq066.

YU, Binhai; PETTITT, B. Montgomery; IWAHARA, Junji. Dynamics of Ionic Interactions at Protein-Nucleic Acid Interfaces. *Accounts of Chemical Research*, v. 53, n. 9, p. 1802-1810, 2020. DOI: 10.1021/acs.accounts.0c00212.

ZHANG, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, v. 490, n. 7418, p. 49-54, 2012. <https://doi.org/10.1038/nature11413>.

ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, v. 57, n. 4, p. 702-710, 2004. <https://doi.org/10.1002/prot.20264>.

APÊNDICE A – PARÂMETROS UTILIZADOS NAS ANÁLISES DE DINÂMICA MOLECULAR

Parâmetros para todas as simulações <i>coarse-grained</i> .					
	<i>Hs RXR/Hs PPAR</i>	<i>Hs RXR/Hs RAR</i>	<i>Mg RXR/Mg PPAR</i>	<i>Mg RXR/Mg RAR</i>	<i>Mg RXR/Mg NR1P10</i>
Cutoff-scheme	Verlet	Verlet	Verlet	Verlet	Verlet
Buffer tolerance [kJ mol ⁻¹]	0,005	0,005	0,005	0,005	0,005
Coulomb type	Reaction-field	Reaction-field	Reaction-field	Reaction-field	Reaction-field
cutoff [nm]	1,1	1,1	1,1	1,1	1,1
Epsilon-r	15; 2,5	15; 2,5	15; 2,5	15; 2,5	15; 2,5
Van-der-Waals type	cutoff	cutoff	cutoff	cutoff	cutoff
cutoff [nm]	1,1	1,1	1,1	1,1	1,1
Thermostat	V-rescale	V-rescale	V-rescale	V-rescale	V-rescale
Tau-T	1,0	1,0	1,0	1,0	1,0
Reference T [K]	300	300	300	300	300
Barostat	Parrinello-Rahman	Parrinello-Rahman	Parrinello-Rahman	Parrinello-Rahman	Parrinello-Rahman
Tau-p	12,0	12,0	12,0	12,0	12,0
Reference p [bar]	1,0	1,0	1,0	1,0	1,0
Integrador da minimização	Steepest-descent	Steepest-descent	Steepest-descent	Steepest-descent	Steepest-descent
Número de passos	40000	40000	40000	40000	40000
Integrador da equibração	Stochastic dynamics	Stochastic dynamics	Stochastic dynamics	Stochastic dynamics	Stochastic dynamics
Delta-t [ps]	0,02	0,02	0,02	0,02	0,02
Número de passos	5x10 ⁵	5x10 ⁵	5x10 ⁵	5x10 ⁵	5x10 ⁵
Tempo total [ns]	10	10	10	10	10
Integrador da produção	Stochastic dynamics	Stochastic dynamics	Stochastic dynamics	Stochastic dynamics	Stochastic dynamics
Delta-t [ps]	0,02	0,02	0,02	0,02	0,02
Número de passos	5x10 ⁸	5x10 ⁸	5x10 ⁸	5x10 ⁸	5x10 ⁸
Tempo total [μs]	10	10	10	10	10
Número de réplicas	3	3	3	3	3

APÊNDICE B – MÉTRICAS DE QUALIDADE DA MODELAGEM TRIDIMENSIONAL DE RECEPTORES NUCLEARES DE *H. SAPIENS*

Receptor nuclear	Átomos	pLDDT					pTM	ipTM	Porção Desordenada
		Média	Muito Alto	Alto	Médio	Baixo			
NR0B1	4689	70,44	17,2	47,1	13,1	22,6	0,41	0,35	15,0%
NR0B2	4471	76,64	30,3	41	14,6	14,1	0,71	0,73	9,0%
NR1A1	5461	81,01	39,2	40	14	6,8	0,68	0,69	5,0%
NR1A2	5457	81,21	40	39,2	14,2	6,5	0,67	0,68	5,0%
NR1A2	5247	81,44	42,7	36,4	13	7,9	0,73	0,77	5,0%
NR1B2	5260	81,58	44,3	35,5	12,1	8,1	0,73	0,78	6,0%
NR1B3	5244	81,87	45,9	34,1	11,7	8,3	0,73	0,8	5,0%
NR1C1	5521	79,91	40,6	35	15,3	9,2	0,76	0,8	4,0%
NR1C2	5559	79,1	38,2	35	17,7	9	0,75	0,78	5,0%
NR1C3	5545	78,2	36,9	34,5	18,1	10,6	0,75	0,8	2,0%
NR1D1	6369	66,98	24,8	31,7	14,2	29,3	0,56	0,67	21,0%
NR1D2	6414	67,34	23,1	34,6	15,8	26,6	0,58	0,66	20,0%
NR1F1	6131	71,11	20,4	45,7	13,7	20,3	0,44	0,32	17,0%
NR1F2	6122	71,77	20,2	44,4	16,5	18,9	0,4	0,31	15,0%
NR1F3	6383	71,91	33,7	31,3	12,1	22,9	0,41	0,48	20,0%
NR1H2	5571	79,08	37,1	38,7	13,7	10,5	0,66	0,65	10,0%
NR1H3	5449	78	31,7	41,4	17,5	9,4	0,65	0,64	7,0%
NR1H4	5459	76,33	29,1	42,7	15,7	12,5	0,64	0,65	8,0%
NR1I1	5787	77,72	37,9	37,3	11,6	13,2	0,64	0,64	12,0%
NR1I2	5783	79,8	36,6	40,2	14,8	8,4	0,7	0,72	7,0%
NR1I3	5339	79,3	33,2	41,2	19,1	6,6	0,73	0,79	2,0%
NR2A1	5151	79,26	33,3	42,6	15,9	8,2	0,69	0,71	2,0%
NR2A2	5145	78,92	33,6	42,7	14,7	8,9	0,68	0,7	3,0%
NR2B1	5190	78,35	35,4	37,1	17,9	9,5	0,65	0,63	3,0%
NR2B2	5160	74,92	28,4	41,3	14	16,4	0,63	0,61	10,0%
NR2B3	5176	76,13	28,2	43,7	14,5	13,7	0,64	0,61	6,0%
NR2C1	6370	71,32	30,7	34,8	12,3	22,2	0,6	0,67	21,0%
NR2C2	6247	72,43	27,2	38,5	14	20,3	0,61	0,64	16,0%
NR2E1	5482	73,07	24,8	43,2	13	19	0,62	0,65	10,0%
NR2E3	5426	76,98	40,8	31,6	11	16,7	0,64	0,68	12,0%
NR2F1	5181	78,06	33,1	41,1	14,9	11	0,67	0,67	5,0%
NR2F2	5183	78,55	34,6	40	14,6	10,9	0,67	0,67	6,0%
NR2F6	5178	77,92	34,9	40,2	12,6	12,2	0,65	0,64	8,0%
NR3A1	5508	78,61	40,7	34,6	11,8	12,9	0,66	0,66	10,0%
NR3A2	5382	77,88	34,9	39,2	13,8	12,1	0,66	0,65	9,0%
NR3B1	5244	79,78	43,7	33,1	11,8	11,4	0,67	0,67	9,0%
NR3B2	5260	79,86	40,9	37,5	10,6	10,9	0,68	0,69	8,0%
NR3B3	5242	80,43	43,1	35,5	11,1	10,2	0,68	0,69	8,0%
NR3C1	5316	73,85	18,3	49,5	17,9	14,3	0,41	0,19	8,0%
NR3C2	5496	72,34	19,3	47,2	16,1	17,4	0,37	0,15	9,0%
NR3C3	5390	73,28	23,5	44,4	16,8	15,3	0,43	0,2	10,0%
NR3C4	5355	74,13	20,3	48,5	17	14,3	0,42	0,19	8,0%
NR4A1	5195	79,9	38,5	39,2	13	9,4	0,68	0,7	6,0%
NR4A2	5248	80,15	40,5	37,6	12,1	9,9	0,71	0,76	6,0%
NR4A3	5226	78,71	34,7	41,6	12,8	10,9	0,64	0,63	7,0%
NR5A1	6124	73,92	33,5	34,5	13,1	18,9	0,61	0,66	14,0%
NR5A2	6252	73,41	34,1	33,7	12,4	19,9	0,58	0,61	18,0%
NR6A1	6012	71,31	14,6	52	12,6	20,8	0,41	0,33	13,0%

**APÊNDICE C – MÉTRICAS DE QUALIDADE DA MODELAGEM
TRIDIMENSIONAL DE RECEPTORES NUCLEARES DE *M. GIGAS***

Receptor nuclear	Átomos	pLDDT					pTM	ipTM	Porção desordenada
		Média	Muito Alto	Alto	Médio	Baixo			
2DBD Delta	9813	53,91	3,3	34,4	13,7	48,6	0,33	0,31	35,0%
2DBD Gamma	6716	74,97	30,2	39,9	13,3	16,6	0,58	0,71	13,0%
NR0B	4296	79,81	35,3	43,6	12,2	8,9	0,7	0,75	5,0%
NR1A	5298	76,58	29,2	41	18	11,7	0,63	0,64	8,0%
NR1B	5311	76,09	27,1	44,2	15,8	13	0,69	0,72	11,0%
NR1CDEF Alpha	7359	62,43	11,7	38,4	14	35,9	0,36	0,25	26,0%
NR1CDEF Beta	7929	64,28	19,4	34,5	10,4	35,7	0,52	0,65	31,0%
NR1C	5164	77,2	27,3	44,5	19,1	9,1	0,72	0,77	3,0%
NR1D	5418	75,89	26,9	43,4	16,8	13	0,64	0,63	5,0%
NR1E	5424	79,78	37,3	39,9	13,4	9,4	0,74	0,79	7,0%
NR1F	5865	74,64	30,5	41,3	10,8	17,5	0,51	0,5	14,0%
NR1H	5589	76,64	33,8	37,6	15,5	13,1	0,65	0,67	10,0%
NR1J Alpha	5650	75,26	30,8	35	21,6	12,6	0,63	0,66	10,0%
NR1J Beta	5397	75,86	27,3	40,6	20,6	11,5	0,64	0,66	7,0%
NR1J Gamma	6356	72,51	21,4	43,5	18,5	16,5	0,7	0,74	11,0%
NR1P1	6349	75,57	35,1	35,4	13,2	16,3	0,65	0,69	16,0%
NR1P2	6253	75,41	35,5	34,6	13,3	16,6	0,64	0,68	14,0%
NR1P3	6531	73,46	35,5	31,3	14,8	18,5	0,64	0,72	17,0%
NR1P4	5974	76,49	38,7	29,1	18,6	13,6	0,71	0,82	13,0%
NR1P5	6636	69,7	23,6	34,8	18,9	22,8	0,59	0,68	20,0%
NR1P6	6655	71,83	29,6	35,3	13,6	21,6	0,61	0,69	20,0%
NR1P7	6450	70,81	24,1	38,6	16,3	21,1	0,64	0,75	9,0%
NR1P8	6423	74,55	31,2	37,9	13,8	17,1	0,65	0,69	15,0%
NR1P9	7396	66,25	20,7	35,6	14,9	28,8	0,58	0,67	25,0%
NR1P10	5723	76,81	30,3	40,4	19,3	10	0,67	0,67	8,0%
NR1P11	5613	78,05	27,6	47,3	17,7	7,5	0,72	0,77	5,0%
NR2A	5094	78,94	33,1	44,1	13,6	9,2	0,7	0,73	5,0%
NR2B	5082	79,28	39,5	36,1	14,1	10,3	0,66	0,65	5,0%
NR2CD	6201	72,61	27,5	37,4	16,2	19	0,63	0,66	18,0%
NR2E1	5392	72,54	16,2	50,9	16	16,8	0,59	0,58	9,0%
NR2E2	5482	74,07	30,3	39,9	12,4	17,3	0,63	0,68	14,0%
NR2E3	5476	74,1	37,6	32,1	10,1	20,3	0,6	0,67	20,0%
NR2E5	6356	68,52	35	27	8,6	29,4	0,56	0,69	29,0%
NR2F	5271	77,36	32,4	41,2	14,6	11,7	0,66	0,65	8,0%
NR3A	5233	79,07	37,4	40,4	11,9	10,3	0,65	0,63	10,0%
NR3B	5279	78,6	36,4	41	11,6	11	0,66	0,65	9,0%
NR4A	5176	80,97	41,5	38,2	11,3	8,9	0,71	0,72	7,0%
NR5A	6332	71,22	23,3	41,7	12,5	22,5	0,55	0,53	22,0%
NR5B	5543	74,31	27,5	40,1	14,2	18,1	0,58	0,59	12,0%
NHR40	3153	76,49	31,7	38,2	18	12,1	0,54	0,26	5,0%
NHR41	4537	78,6	37,7	36,1	13,9	12,3	0,72	0,81	5,0%
NHR42	5250	78,03	36,8	38,6	12,9	11,7	0,67	0,67	10,0%
NHR43	8262	54,64	2,1	38,2	14,2	45,6	0,32	0,21	34,0%