Universidade Federal da Paraíba

Programa de Pós Graduação em Biotecnologia

Elton José Ferreira Chaves

Obtenção de Inibidores da Subunidade Catalítica da Ricina, RTA, Usando Métodos de Modelagem Molecular

João Pessoa, Paraíba

2021

Elton José Ferreira Chaves

Obtenção de Inibidores da Subunidade Catalítica da Ricina, RTA, Usando Métodos de Modelagem Molecular

Tese de Doutorado apresentada ao Programa de Pós-graduação em Biotecnologia da Universidade Federal da Paraíba, como requisito para a obtenção do título de Doutor em Biotecnologia.

Universidade Federal da Paraíba

Programa de Pós Graduação em Biotecnologia

Orientador: Dr. Demetrius Antônio Machado de Araújo

Coorientador: Dr. Gerd Bruno Rocha

João Pessoa, Paraíba

2021

Catalogação na publicação
Seção de Catalogação e Classificação

C512o Chaves, Elton José Ferreira.
Obtenção de inibidores da subunidade catalítica da ricina, RTA, usando métodos de modelagem molecular / Elton José Ferreira Chaves. - João Pessoa, 2021.
145 f. : il.

Orientação: Demetrius Antônio Machado Araújo.
Coorientação: Gerd Bruno Rocha.
Tese (Doutorado) - UFPB/CBiotec.

1. Biotecnologia. 2. Ricina. 3. RTA - Triagem virtual. 4. Dinâmica molecular. I. Araújo, Demetrius Antônio Machado. II. Rocha, Gerd Bruno. III. Título.

UFPB/BC CDU 60(043)

Elaborado por ANNA REGINA DA SILVA RIBEIRO - CRB-15/024

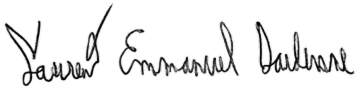
Elton José Ferreira Chaves

Obtenção de Inibidores da Subunidade Catalítica da Ricina, RTA, Usando Métodos de Modelagem Molecular

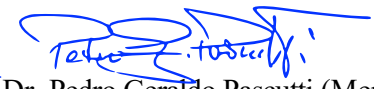
Tese de Doutorado apresentada ao Programa de Pós-graduação em Biotecnologia da Universidade Federal da Paraíba, como requisito para a obtenção do título de Doutor em Biotecnologia.

Trabalho aprovado. João Pessoa, Paraíba, 2021

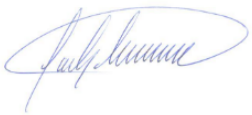
Prof. Dr. Marcus Tullius Scotti (Membro externo ao programa)



Prof. Dr. Laurent Emmanuel Dardenne (Membro externo a Instituição)



Prof. Dr. Pedro Geraldo Pascutti (Membro externo a Instituição)



Prof. Dr. Carlos Henrique da Silveira (Membro externo a Instituição)

João Pessoa, Paraíba

2021

Agradecimentos

À Universidade Federal da Paraíba (UFPB);

À CAPES, pela concessão da bolsa e auxílio (AUXPE 1375/2014);

Ao CNPq, FAPESQ e FACEPE;

Aos centros de computação científica CENAPAD e Santos Dumont;

Aos meus pais, irmãos e amigos pelo apoio durante mais esta jornada acadêmica;

À equipe do Laboratório de Biotecnologia Celular e Molecular da UFPB;

À equipe do Laboratório de Química Quântica Computacional da UFPB;

Aos meus orientadores, Prof. Demetrius Araújo e Prof. Gerd Bruno Rocha, por todo conhecimento passado ao longo dos anos;

À banca examinadora: Prof. Carlos Silveira, Prof. Laurent Dardenne, Prof. Marcus Scotti e Prof. Pedro Pascutti;

À todos que contribuíram direta ou indiretamente para a realização deste trabalho, muito obrigado!

"Desde que me cansei de procurar, aprendi a encontrar.
Desde que o vento me opôs resistência,
velejo com todos os ventos."
F. Nietzsche

Resumo

A ricina é uma potente toxina presente na semente da mamona (*Ricinus Communis*) e consiste em duas subunidades, RTA e RTB. Devido a sua toxicidade, a ricina tem alarmado autoridades mundiais por seu potencial uso como arma química. A ricina também impacta o agronegócio da mamona, devido ao risco de intoxicação animal e humana. Ao longo dos anos, grupos de pesquisa propuseram bloqueadores para o sítio ativo da RTA, a cadeia catalítica. Apesar de tais esforços, ainda não existe uma contramedida eficaz contra o envenenamento por ricina. O estudo computacional realizado no presente trabalho renova a discussão sobre pequenas moléculas com potencial de inibir a atividade catalítica dessa toxina. Aqui, um protocolo de triagem virtual baseado em estrutura capaz de discernir inibidores de RTA ativos de inativos foi empregado para prospectar dentre 2 milhões de compostos do banco de dados ZINC, novos ligantes com alto potencial de atracar ao sítio ativo da RTA. Além disso, um novo método de pontuação baseado em perfis de força de desencaixe do ligante e cálculos quânticos semi-empíricos forneceram uma nova perspectiva sobre o *rescore* das poses de acoplamento. Em suma, as etapas de filtragem apontaram para 7 compostos principais, sendo o SCF00–451 o candidato promissor para inibir a atividade tóxica da RTA.

Palavras-chave: Ricina, RTA, triagem virtual, dinâmica molecular

Abstract

Ricin is a potent toxin isolated from the castor bean plant and consists of two subunits, RTA and RTB. Because of its toxicity, ricin has alarm world authorities for its potential use as a chemical weapon. Ricin also impacts castor bean agribusiness, given the risk of animal and human poisoning. Over the years, many groups attempted to proposed small-molecules that bind to the RTA active site, the catalytic chain. Despite such efforts, there is still no effective countermeasure against ricin poisoning. The computational study carried out in the present work renewing the discussion about small-molecules that may inhibit this toxin. Here, a structure-based virtual screening protocol capable of discerning active RTA inhibitors from inactive ones was employed to screening more than 2 million compounds from the ZINC database to find novel scaffolds that strongly bind into the active site of the RTA. Besides, a novel score method based on ligand undocking force profiles and semi-empirical quantum chemical calculations provided insights into the rescore of docking poses. Summing up, the filtering steps pointed out to 7 main compounds, with the SCF00–451 as a promising candidate to inhibit the killing activity of such potent phytotoxin.

Keywords: Ricin, RTA, virtual screening, molecular dynamics

Lista de abreviaturas e siglas

ADME	Absorção, Distribuição, Metabolismo e Excreção
AZT	3-azido-3-deoxitimidina
BFA	Brefeldina A
CADD	*Computed-Aided Drug Design*
CTD	*C-terminal domain*
DDP	2,5-diamino-4,6-dihidroxipirimidina
DM	Dinâmica Molecular
FEP	Free Energy Pertubation
FMP	Formicina 5'-monofosfato
HMG	Hidroximetilglutaril
PDB	*Protein Data Bank*
PTA	Ácido pteróico
QSAR	*Quantitative Strucuture–Activity Relationship*
RE	Retículo Endoplasmático
RIP	*Ribosome Inactivating Protein*
RMN	Ressonância Magnética Nuclear
RTA	*Ricin Toxin A*
RTB	*Ricin Toxin B*
RMSD	*Root Mean Square Deviation*
SMD	*Steered Molecular Dynamics*
SCOP	*Classification of Proteins*
SRL	*Sarcin-Ricin Loop*
TGN	*Trans-Golgi Network*
TV	Triagem Virtual

TVBEA	Triagem Virtual Baseada na Estrutura do Alvo
TVBEL	Triagem Virtual Baseada na Estrutura do Ligante
TVH	Triagem Virtual Hierárquica
TVR	Transporte Vesicular Retrógrado

Lista de ilustrações

Figura 1 – Estrutura 3D da ricina. 9
Figura 2 – Estrutura 3D da RTA complexada com o ligante FMP. 10
Figura 3 – Mecanismo de catálise da RTA. 11
Figura 4 – Reconhecimento da RTA à região *P stalk* do ribossomos. 12
Figura 5 – Modelo de ensaio *in vitro* para mensurar a atividade catalítica da RTA. 13
Figura 6 – Estrutura cristalográfica do complexo RTA–DDP e conformações do resíduo Y80. 20
Figura 7 – Modelo farmacofórico do ligante 19M. 28
Figura 8 – Perfil energético de uma proteína (funil rugoso) 30
Figura 9 – Passos de uma simulação de DM. 33
Figura 10 – Condições específicas de uma simulação DM para eliminar o efeito de borda e otimizar o tempo de cálculo. 34
Figura 11 – Representação 3D de confórmeros obtidos com o programa OMEGA. . 43
Figura 12 – Fluxograma referente as etapas de preparo da biblioteca de ligantes, preparo e validação do modelo de consulta e triagem farmacofórica. . 47
Figura 13 – Fluxograma da triagem virtual. 51
Figura 14 – Fluxograma referente as análises pós-atracamento. 54
Figura 15 – X-ray structure of the RTA subunit (PDB code 3RTI). 63
Figura 16 – ROCS's shape query obtained from the ligand in 4HUP structure. . . 67
Figura 17 – llustration of the pulling direction employed in SMD simulations. . . . 71
Figura 18 – Chemical structures and ZINC codes of the selected compounds to post-docking analysis. 72
Figura 19 – Steered MD force profiles using $k = 2\ kcal \cdot mol^{-1}$Å^{-2}, and $v = 0.05$ Å/ps. 74
Figura 20 – Steered MD force profiles using $k = 10\ kcal \cdot mol^{-1}$Å^{-2}, and $v = 0.05$ Å/ps. 76
Figura 21 – RMSD analysis of the equilibrium MD trajectory of RTA–SCF00–451 complex. 80
Figura 22 – Interaction energy analysis between SCF00–451 and the RTA chain along the last 5ns of MD simulation. 81
Figura 23 – The binding pose of the SCF00–451 compound into the RTA binding pocket in early (A) and final (B) stages of the MD simulation. 82
Figura 24 – RMSD analysis of the equilibrium MD trajectory of the RTA–SCF00–332 complex. 83
Figura 25 – Interaction energy analysis between SCF00–332 and RTA chain over the last 5 ns of MD simulation. 84

Figura 26 – The binding pose of the SCF00–332 into the RTA binding pocket at the final stage of MD simulation. 85
Figura 27 – RMSD analysis of the equilibrium MD trajectory of RTA–SCF00–496 and –499 complexes. 86
Figura 28 – Interaction energy analysis of the SCF00–499 compound into the binding pocket of the RTA over the last 5 ns of MD simulation. 87
Figura 29 – The binding pose of the SCF00–496 (**A**) and –499 (**B**) complexes in the RTA binding pocket at the final stage of MD simulation. 88
Figura 30 – ROC curve analysis employed to each one score function available in the ROCS software. 92
Figura 31 – Illustration showing the main steps performed in a cross-docking study. 93
Figura 32 – Each value corresponds to the RMSD of the lowest-energy docking pose for each RTA structure available in the PDB. 94
Figura 33 – Predicted poses using the ASP score and 1BR5 structure. Native and predicted poses are depicted in cyan and green colors, respectively. . . 95
Figura 34 – ROC curve analysis when ASP score was used. 96
Figura 35 – Trajectory analysis of the RTA–SCF00-06 complex. 97
Figura 36 – Trajectory analysis of the RTA–SCF00-15 complex. 97
Figura 37 – Trajectory analysis of the RTA–SCF00-63 complex. 98
Figura 38 – Trajectory analysis of the RTA–SCF00-66 complex. 98
Figura 39 – Trajectory analysis of the RTA–SCF00-86 complex. 99
Figura 40 – Trajectory analysis of the RTA–SCF00-95 complex. 99
Figura 41 – Trajectory analysis of the RTA–SCF00-96 complex. 100
Figura 42 – Trajectory analysis of the RTA–SCF00-111 complex. 100
Figura 43 – Trajectory analysis of the RTA–SCF00-194 complex. 101
Figura 44 – Trajectory analysis of the RTA–SCF00-194 complex. 101
Figura 45 – Trajectory analysis of the RTA–SCF00-201 complex. 102
Figura 46 – Trajectory analysis of the RTA–SCF00-229 complex. 102
Figura 47 – Trajectory analysis of the RTA–SCF00-238 complex. 103
Figura 48 – Trajectory analysis of the RTA–SCF00-245 complex. 103
Figura 49 – Trajectory analysis of the RTA–SCF00-252 complex. 104
Figura 50 – Trajectory analysis of the RTA–SCF00-268 complex. 104
Figura 51 – Trajectory analysis of the RTA–SCF00-271 complex. 105
Figura 52 – Trajectory analysis of the RTA–SCF00-277 complex. 105
Figura 53 – Trajectory analysis of the RTA–SCF00-325 complex. 106
Figura 54 – Trajectory analysis of the RTA–SCF00-341 complex. 106
Figura 55 – Trajectory analysis of the RTA–SCF00-341 complex. 107
Figura 56 – Trajectory analysis of the RTA–SCF00-426 complex. 107
Figura 57 – Trajectory analysis of the RTA–SCF00-511 complex. 108
Figura 58 – Trajectory analysis of the RTA–SCF00-514 complex. 108

Figura 59 – Trajectory analysis of the RTA–SCF00-531 complex. 109
Figura 60 – Trajectory analysis of the RTA–SCF00-533 complex. 109
Figura 61 – Trajectory analysis of the RTA–SCF00-540 complex. 110

Lista de tabelas

Tabela 1 – Moléculas com atividade anti-ricina por modulação do TVR. 15
Tabela 2 – Derivados de purina com estrutura do complexo RTA–Ligante depositada no PDB. 17
Tabela 3 – Derivados de pterina com estrutura do complexo RTA–Ligante depositada no PDB. 18
Tabela 4 – Derivados de pirimidina com estrutura do complexo RTA–Ligante depositada no PDB. 19
Tabela 5 – List of Main X-Ray Structures and Experimental IC_{50} (μM) data of RTA-Ligand Complexes Available in the Protein Data Bank. 64
Tabela 6 – Summary of the Calculated Results. All Numerical Data Presented in the 2^{nd} and 3^{rd} columns are in pN; 4^{th} and 5^{th} columns are in $kcal \cdot mol^{-1}$. 73
Tabela 7 – Hydrogen bond calculations between RTA chain and best-ranked ligands. 79
Tabela 8 – List of RTA Inhibitors Retrieved From ChEMBL Database by Using the Keyword “Ricin”. 90
Tabela 9 – Summary of the Theoretical Results. All Numerical Data Presented in the 2^{nd} and 3^{th} columns are in pN; 4^{th} and 5^{th} columns are in $kcal \cdot mol^{-1}$. 91

Sumário

Lista de abreviaturas e siglas . . . . . **xiii**

Lista de ilustrações . . . . . **xv**

Lista de tabelas . . . . . **xviii**

1 INTRODUÇÃO **3**

2 REVISÃO BIBLIOGRÁFICA **7**

2.1 Ricina: Estrutura e Função **8**

2.2 Contramedidas para Intoxicação por Ricina **13**

2.2.1 Bloqueadores do Transporte Vesicular Retrógrado 14

2.2.2 RiVax 15

2.2.3 Bloqueadores do Sítio Ativo da RTA 16

2.2.3.1 Derivados de purina 16

2.2.3.2 Derivados de pterina 17

2.2.3.3 Derivados de pirimidina 18

2.2.4 Inibidores putativos do sítio ativo da RTA 19

2.3 Triagem Virtual **21**

2.4 Planejamento Racional de Compostos Bioativos **22**

2.4.1 Triagem Virtual Baseada na Estrutura do Ligante 22

2.4.2 Triagem Virtual Baseada na Estrutura do Alvo 23

2.5 Reconhecimento Molecular (Receptor–Ligante) **24**

2.6 *Docking* Receptor–Ligante **25**

2.7 Triagem Farmacofórica Baseada na Estrutura do Alvo **27**

2.8 Dinâmica Molecular de Proteínas **29**

2.8.1 Simulação de Dinâmica Molecular 30

2.9 Dinâmica Molecular de Caminho Induzido **36**

2.10 Métodos de Química Quântica **38**

3 OBJETIVOS **39**

3.1 Objetivo Geral **39**

3.2 Objetivos Específicos **39**

4 METODOLOGIA **41**

4.1 Preparo da Biblioteca de Triagem **42**

4.2 Preparo da Biblioteca de Validação **44**

4.3 Triagem Farmacofórica Baseada na Estrutura . . . 46
4.4 Atracamento Molecular . . . 48
4.4.1 Validação do protocolo de atracamento molecular . . . 48
4.4.1.1 Preparo dos receptores e ligantes . . . 48
4.4.1.2 1º Validação . . . 49
4.4.1.3 2º Validação . . . 50
4.4.1.4 Triagem virtual baseada em *docking* receptor–ligante . . . 50
4.5 Análises Pós-atracamento . . . 52
4.5.1 Inspeção por ligações de hidrogênio . . . 52
4.5.2 Dinâmica molecular de caminho induzido . . . 55
4.5.3 Análises de trajetória . . . 56

5 RESULTADOS E DISCUSSÃO . . . 59
5.1 Discovery of RTA Ricin Subunit Inhibitors: A Computational Study Using PM7 Quantum Chemical Method and Steered Molecular Dynamics . . . 60
5.1.1 Abstract . . . 60
5.1.2 Graphical Abstract . . . 61
5.1.3 Introduction . . . 62
5.1.4 Theoretical Methods . . . 66
5.1.4.1 Virtual Screening Library Preparation . . . 66
5.1.4.2 Structure-Based Pharmacophore Screening . . . 66
5.1.4.3 Molecular Docking . . . 67
5.1.4.4 Hydrogen Bonds Inspection . . . 68
5.1.4.5 Molecular Dynamics Simulations . . . 68
5.1.4.6 Post-Docking Filtering by Steered MD Simulations and PM7 . . . 69
5.1.5 Results . . . 72
5.1.5.1 SMD Force Profiles . . . 74
5.1.6 SMD Force Profiles Using a Different Spring Constant . . . 76
5.1.6.1 Semiempirical Calculations . . . 77
5.1.6.2 Rescoring by Using SMD Force Profiles and $\Delta H_{binding}$ . . . 77
5.1.6.3 MD Trajectory Analysis of the Best-Ranked Compounds . . . 78
5.1.6.4 Compound SCF00–451 . . . 80
5.1.6.5 Compound SCF00–332 . . . 83
5.1.6.6 Compounds SCF00–496 and SCF00–499 . . . 85
5.1.7 Conclusion . . . 88
5.1.8 Material Suplementar do Artigo . . . 89

REFERÊNCIAS . . . 111

REFERÊNCIAS DO ARTIGO . . . 119

REFERÊNCIAS DO MATERIAL SUPLEMENTAR DO ARTIGO 125

1 Introdução

A ricina é uma proteína citotóxica produzida na semente da *Ricinus communis* [1]. No Brasil, esta espécie é conhecida como mamoneira ou ricínio, e pode ser encontrada em qualquer parte do território. Seu fruto, a mamona, tem ganhado atenção da indústria, mídia e governos, devido às propriedades de seus constituintes, especialmente a ricina e o óleo. Aproximadamente 90% dos ácidos graxos extraídos da mamona compreendem ao ácido ricinoleico, que por sua vez apresenta propriedades físico-químicas importantes para indústria – tanto química como de combustíveis –, devido à presença de insaturação e hidroxilas em sua estrutura química [2]. Além disso, durante a extração do óleo são gerados alguns coprodutos, sendo os principais a casca do fruto e a torta (produto do processamento da mamona). Neste aspecto, a indústria de fertilizantes vem utilizando-os devido ao alto teor de nitrogênio e efeitos nematicidas. Uma outra rota seria a ração animal, devido ao alto teor de fibras e proteínas, no entanto, isto não ocorre devido à presença da ricina, uma potente citotoxina [3].

Proteínas tóxicas estão presentes em várias espécies. Sugere-se que elas evoluíram para desempenhar um papel específico na defesa contra animais, patógenos e insetos, conferindo vantagem em um nicho específico [4]. Neste aspecto, a ricina corresponde a proteína de origem natural mais letal já conhecida [5]. Em humanos, se inalada, ingerida ou injetada, a ricina pode ser letal [6]. Somado a isso, esta toxina vem preocupando as autoridades mundiais devido ao potencial uso como arma química, principalmente por grupos terroristas e ativistas [7]. Portanto, existe um interesse biotecnológico em propor contramedidas para erradicar os efeitos de sua intoxicação.

No que se refere a estrutura, a ricina é composta por duas subunidades, denominadas RTA e RTB, ligadas entre si por uma única ligação dissulfeto. A RTA é a subunidade catalítica, uma rRNA N-glicosidase, responsável por depurinar uma adenina conservada na região *Sarcin-Ricin Loop* (SRL), localizado na subunidade ribossomal 28S. Já a RTB, uma lectina, é responsável pela internalização do complexo RTA–RTB nas células–alvo pelo reconhecimento via açúcares da superfície celular. Uma vez internalizada, a ricina segue via Transporte Vesicular Retrógrado (TVR) até o Retículo Endoplasmático (RE), onde uma isomerase específica cliva a ligação dissulfeto entre as duas subunidades e a RTA é então translocada ao citosol, onde ataca os ribossomos. Com isso, a síntese de proteínas da célula é comprometida, o que leva a uma série de complicações metabólicas e consequentemente à morte celular.

Desde meados da década de oitenta, grupos de pesquisa vêm propondo contramedidas para diminuir e/ou erradicar os efeitos da intoxicação pela ricina, neste aspecto, temos: (*i*) bloqueadores do TVR, (*ii*) anticorpos, e (*iii*) bloqueadores do sítio ativo da RTA. Para este último, ligantes antagonistas têm sido inicialmente identificados por meio de métodos *in silico*, tal como o *docking* receptor–ligante e posteriormente validados por ensaios *in vitro*, cristalografia e difração de raios–X. Alguns autores sugerem que o grande desafio dessa perspectiva é encontrar uma pequena molécula que supere a barreira energética das interações da interface RTA–SRL.

O método de *docking* receptor–ligante se mostrou como uma estratégia racional eficiente para a ricina, uma vez que esta abordagem foi capaz de predizer poses bioativas de pequenas moléculas no sítio ativo da RTA, entretanto, o processo de ranqueamento para definir um ligante "muito potente" de um "fraco", não tem mostrado a mesma eficiência [8].

Predizer a afinidade de ligação em sistemas biológicos, principalmente receptor–ligante, é um campo de forte discussão na literatura, e ao mesmo tempo uma corrida tecnológica para a indústria farmacêutica e empresas de desenvolvimento de *softwares*. Os métodos resultantes dessa intensa discussão estão divididos em dois grupos: (*i*) métodos baseados na estrutura do alvo; e (*ii*) métodos baseados na estrutura do ligante. Ambos estão inclusos numa seção maior, denominada como Desenho de Drogas Guiado por Computador (CADD, do inglês *Computed-Aided Drug Design*) [9, 10]. Até o momento não existe um método melhor para tal – predizer a afinidade de um ligante em relação a um sítio receptor –, a literatura diz que validações prévias, assim como a aplicação paralela ou hierárquica desses métodos pode aumentar consideravelmente a identificação de moléculas ativas [11]. Somado a isso, a abordagem por simulação computacional, tal como a Dinâmica Molecular (DM), pode ser empregada para fins de refinamento [9].

Como exemplo, cito alguns resultados de um trabalho anterior [8], que por sua vez, sustentam alguns aspectos metodológicos desta tese. Em linhas gerais, foi possível constatar que o acoplamento entre um protocolo de *docking* receptor–ligante e DM é capaz de predizer a afinidade de ligantes antagonistas para a RTA. O algoritmo de busca do *docking* foi eficiente em predizer as poses cristalográficas de seis inibidores, entretanto, a afinidade de ligação obtida por este método apresentou baixa correlação com os dados experimentais de IC_{50}, com $R^2 = 0,383$. Por outro lado, ao realizar o *rescore* dessas poses por meio do método de Dinâmica Molecular de Caminho Induzido (SMD, do inglês *Steered Molecular Dynamics*), uma ótima correlação foi alcançada, com $R^2 > 0,9$ [8]. Outros autores obtiveram uma conclusão similar ao utilizar este tipo de abordagem para outros alvos, e na revisão da referência [12] é apresentado uma compilação didática desses resultados.

No mais, este procedimento de *rescore* das poses por meio de métodos baseados

em DM comumente demandam um alto custo computacional[1], logo, não são factíveis para avaliar uma biblioteca com milhões de ligantes, por exemplo. Em outras palavras, cabe a este tipo de refinamento – a dinamica molecular –, avaliar algumas dezenas de ligantes que foram previamente selecionados por métodos *in silico* factíveis de aplicação em larga escala, tal como o *docking* molecular[2]. Por fim, para fins de definição, entende-se por Triagem Virtual Hierárquica (TVH), o procedimento de avaliar o potencial terapêutico de milhares de moléculas por meio de diferentes métodos *in silico*. Podendo a aplicação destes métodos ocorrer de forma sequencial (como a descrita acima) ou paralela [11].

Tendo em vista a problemática da intoxicação por ricina, tanto na perspectiva industrial como da segurança pública, a presente tese teve por objetivo executar um protocolo de TVH para avaliar o potencial antagonista dentre ∼2 milhões de ligantes para o sítio ativo da RTA.

Antes da apresentação e discussão dos resultados, no capítulo 2, a seguir, será apresentado um aprofundamento sobre os temas Ricina e Triagem Virtual. Para este último, será discutido em mais detalhes os métodos utilizados na tese, tal como a triagem por similaridade farmacofórica, *docking* receptor–ligante e refinamento por DM. Em seguida, nos capítulo 3 e 4, os objetivos e metodologia, respectivamente. No capítulo 5 será apresentado os resultados no formato de artigo científico, e por último, no capítulo 6, as conclusões da tese.

[1] O custo refere-se ao tempo de processamento no computador.
[2] Os métodos baseados na estrutura do ligante também são utilizados nesta etapa.

2 Revisão Bibliográfica

Este capítulo tem como objetivo revisar aspectos importantes sobre a bioquímica da ricina e as contramedidas até então estudadas para conter os danos causados por sua intoxicação. Portanto, inicialmente será descrito algumas características estruturais desta toxina, tal como subunidades constituintes, conteúdo de estrutura secundária, domínios, sítio ativo e mecanismo de catálise. Em seguida, será abordado as características funcionais do processo de intoxicação, o qual envolve reconhecimento celular, endocitose, transporte retrógrado até o RE, translocação ao citosol e reconhecimento do ribossomos. A revisão desses aspectos bioquímicos darão suporte ao entendimento das contramedidas até então estudadas, são elas: (*i*) bloqueadores do TVR; (*ii*) vacinas; e (*iii*) bloqueadores do sítio ativo da RTA.

Por fim, será revisado os pressupostos teóricos do planejamento racional de fármacos guiado por computador, o qual tem por objetivo predizer a afinidade de um ligante em relação a um bioreceptor. A este respeito, será revisto alguns conceitos sobre o processo de reconhecimento receptor–ligante e dinâmica das interações intermoleculares em sistemas biológicos e o seu impacto na predição da afinidade em sistemas receptor–ligante.

2.1 Ricina: Estrutura e Função

A ricina pertence à família de proteínas inativadoras de ribossomos (RIP, do inglês, *Ribosome Inactivating Protein*), denominadas assim por possuírem a capacidade de catalizar a hidrólise da ligação glicosídica de uma adenina conservada no motivo GAGA da região SRL de ribossomos eucarióticos. Atualmente esta família de proteínas é dividida em dois grupos [13]. O primeiro consiste nas RIP–tipo1, que apresentam apenas uma única subunidade, denominada como "A", e com atividade N-glicosídica. Já as RIP–tipo2 apresentam essa "subunidade A" ligada covalentemente a uma lectina, ou "subunidade B", que por sua vez contém sítios de ligação a carboidratos [14].

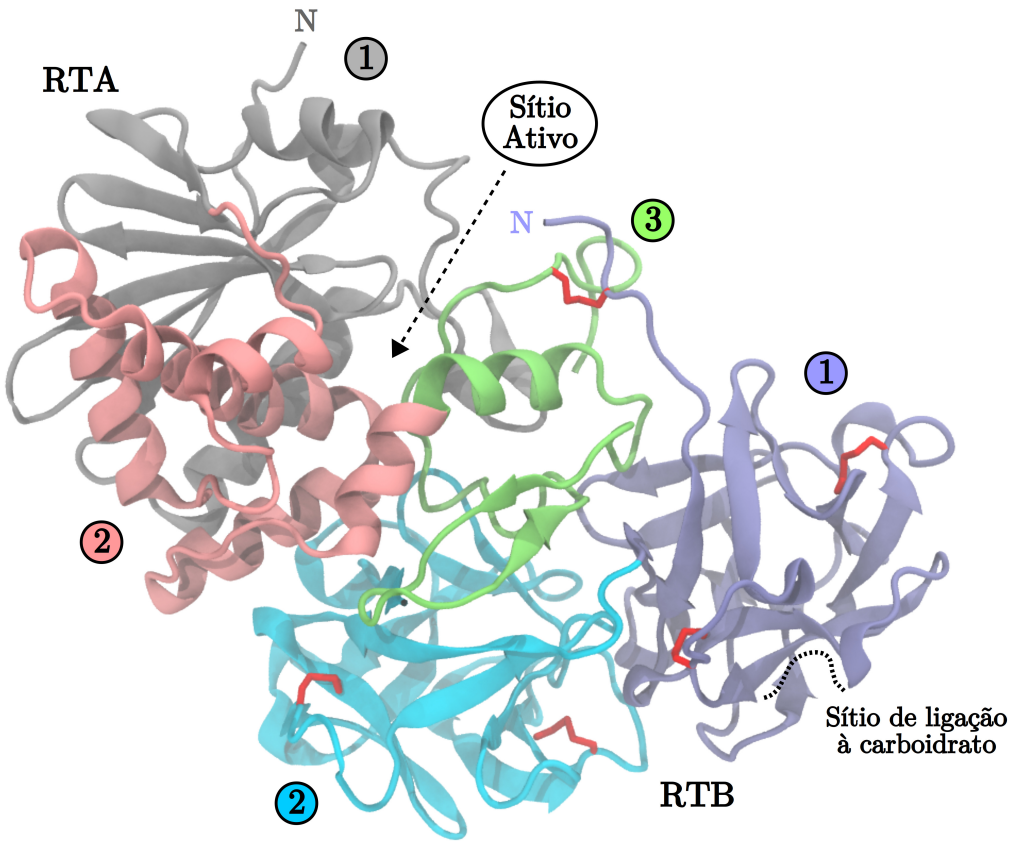
A ricina é uma RIP–tipo2 e as duas subunidades constituintes são denominadas como RTA e RTB, ligadas entre si por uma única ligação dissulfeto (Figura 1). A estrutura tridimensional da ricina foi elucidada por meio de cristalografia e difração de raios–X, inicialmente numa resolução de 2.8 Å [15] e refinada para 2.5 Å [16]. Posteriormente, estudos envolvendo clonagem e expressão heteróloga possibilitaram a elucidação da RTA isolada, primeiramente em resolução de 2.5 Å [17] e posteriormente a 2.3 Å [18] e 1.8 Å [19], respectivamente.

A RTA – subunidade catalítica – apresenta 267 resíduos de aminoácidos e de acordo com a base de dados de classificação de proteínas, o SCOP (do inglês, *Structure Classification of Proteins*) [20], é uma proteína do tipo α+β, enovelada por três domínios: (*i*) domínio 1, resíduos 1–117; (*ii*) domínio 2, resíduos 118–210; (*iii*) domínio 3, resíduos 211–267. O domínio 3 da RTA tem por função primária interagir com a RTB, já os outros dois contribuem para formação do sítio ativo com ao menos um resíduo catalítico (Figura 1). Quanto a RTB, esta contém 262 resíduos de aminoácidos e segue registrada no SCOP como *All*-Beta, e enovelada em dois domínios homólogos: (*i*) domínio 1, resíduos 1–135; (*ii*) domínio 2, resíduos 136–262. Cada domínio apresenta um par de ligações dissulfeto e um sítio de ligação à galactose (Figura 1) [15, 16].

Quanto ao mecanismo de catálise da RTA, os resíduos catalíticos foram inicialmente elucidados a partir de ensaios de mutação sítio–dirigida. O incentivo para tal emergiu de um trabalho comparativo da RTA com seus respectivos homólogos, o qual evidenciou resíduos invariantes, até então suspeitos de serem componentes do sítio ativo dessas N-glicosidases, são eles: Y80, Y123, E177, R180 e W211 [21].

Ready, Kim e Robertus [22] inseriram as mutações Y80F e Y123F na RTA e constataram um decréscimo no K_{cat} de $\sim$15 e 7 vezes, respectivamente, enquanto para as mutações Y80S e Y123S, foi observado um decréscimo de 160 e 70 vezes [21], respectivamente. Já para as mutações E177Q e R180Q foi observado um decréscimo no K_{cat} de $\sim$1000 e 200 vezes, respectivamente. Assim, constataram que E177 e R180 são importantes para a catálise da reação, enquanto os demais estão relacionados com reconhecimento e atracamento do substrato [21, 23, 24]. Esta hipótese foi confirmada por cristalografia e

Figura 1 – Estrutura 3D da ricina.



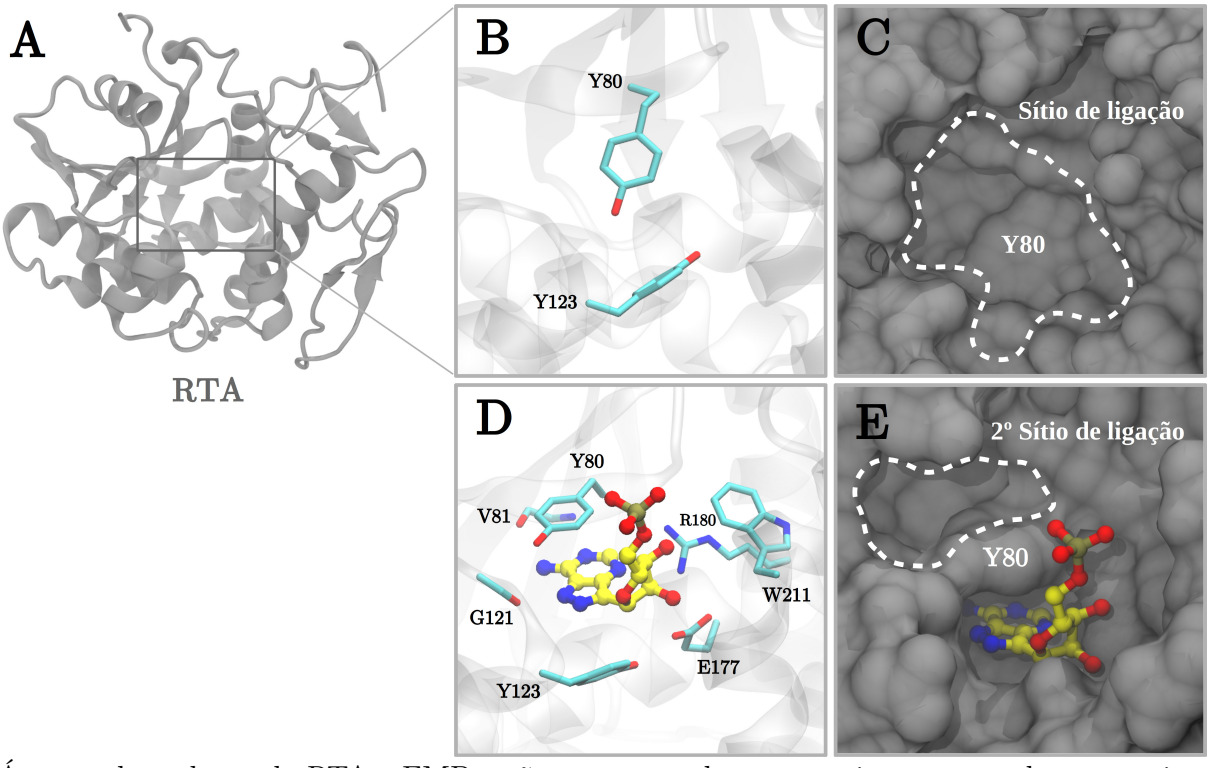
As cores destacam os diferentes domínios estruturais das subunidades RTA e RTB. Na parte superior temos a subunidade RTA, e nas cores cinza, rosa e verde, os domínios 1, 2 e 3, respectivamente. Na parte inferior está a RTB, e nas cores roxo e ciano, os domínios 1 e 2. As ligações dissulfeto estão destacadas em vermelho. Fonte: O autor; ilustração feita no programa VMD; PDB 2AAI.

difração de raios–X complexando a RTA com um análogo não hidrolisável da adenosina monofosfato, a formicina 5'-monofosfato (FMP) [25]. Quando comparado com a estrutura sem ligante, a cadeia lateral de Y80 rotaciona para acomodar o FMP, que interage via π-stacking com os resíduos Y80 e Y123, e faz ligações de hidrogênio com V81, G121, R180, E177 e E208 (Figura 2). No mesmo estudo, complexaram a RTA com o dinucleotídeo adenil-3'-5'-guanosil (ApG) e obtiveram resultados similares [25].

Essa rotação da cadeia lateral de Y80 define a RTA em duas conformações: (1) "fechada" (Figura 2B,C); e (2) "aberta" (Figura 2D,E) [26]. A conformação "aberta" expõe um segundo sítio de ligação, ao qual sabe-se está envolvido no reconhecimento da guanina do motivo GAGA da região SRL. Isto foi confirmado a partir de um estudo cristalográfico da RTA complexada com um tetranucleotídeo cíclico, similar ao motivo GAGA. Em suma, foi observado uma rede de interações π-stacking entre guanina, Y80, adenina e Y123. Somado a isso, este ligante apresentou dados experimentais de inibição da atividade catalítica da RTA na faixa de nM, porém, devido às péssimas propriedades

farmacocinéticas, estudos posteriores de otimização não foram realizados [27].

Figura 2 – Estrutura 3D da RTA complexada com o ligante FMP.



Átomos de carbono da RTA e FMP estão representados na cor ciano e amarelo, respectivamente. A área destacada (linha branca tracejada) correspodne ao sítio de ligação da RTA (**A**) e as ilustrações seguintes (B, C, D e E) correspondem a uma aproximação dessa área; Na ausência de substrato, Y80 adota uma conformação "fechada" (**B**); A representação por superfície da RTA na conformação "fechada" expõe um único sítio de ligação (**C**); Na presença de substrato similar à adenina – substrato natural –, Y80 rotaciona para acomodar o ligante (**D**); A representação por superfície da RTA na presença de substrato expõe um segundo sítio de ligação, decorrente da rotação de Y80. Esta conformação é denominada como "aberta" (**E**). Fonte: O autor; ilustração feita no programa VMD; PDB 2AAI, 3RTI.

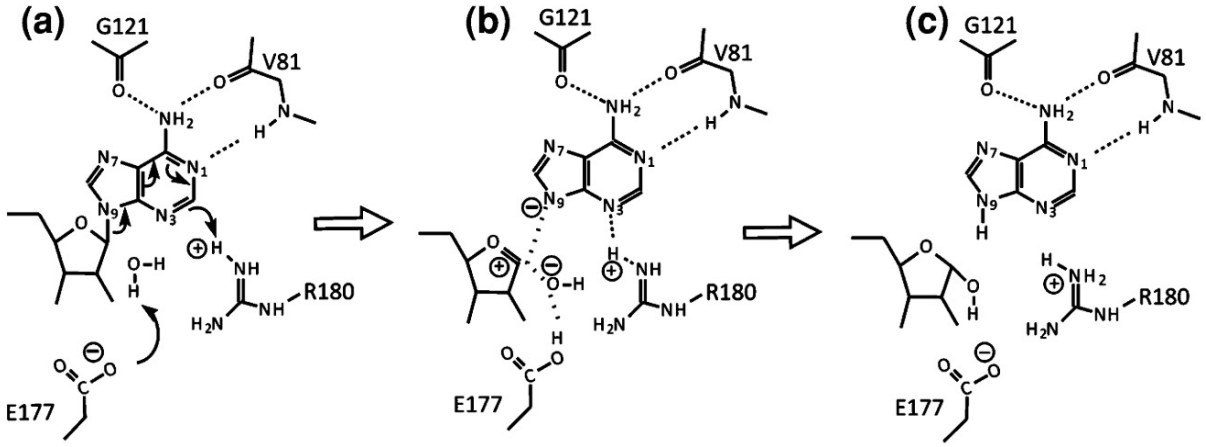
Tendo como base essas informações, Kim e Robertus[21] propuseram um mecanismo de catálise de três passos para o sítio ativo da RTA:

a) A cadeia lateral de E177 polariza uma molécula de água do sítio ativo, o que resulta em um íon hidróxido, que por sua vez estabiliza o centro positivo na posição C1 da ribose (Figura 3A);

b) R180 promove a clivagem da ligação adenina–ribose pela protonação do N3 na adenina, o que resulta na deslocalização de elétrons do anel (Figura 3B);

c) A deslocalização de elétrons do passo anterior promove a clivagem da ligação glicosídica entre C1'–N9, produzindo o íon oxocarbenium.

Figura 3 – Mecanismo de catálise da RTA.



Fonte: Wahome, Robertus e Mantis[26] (2011, p. 185).

Em relação a RTB, esta subunidade é responsável por reconhecer e internalizar o complexo RTA–RTB nas células–alvo, por meio do reconhecimento específico de glicoproteínas na superfície celular, principalmente as que contém N-acetilgalactosamina ou galactose. Quanto a isso, uma vez que resíduos de galactose são abundantes na superfície de qualquer tipo celular, o reconhecimento da ricina nestas células se torna muito sensível, e uma vez internalizada, a holotoxina pode seguir por três vias: (*i*) exocitose; (*ii*) degradação via lisossomos; ou (*iii*) transporte vesicular retrógrado, passando pelo complexo golgiense e depois ao RE [6].

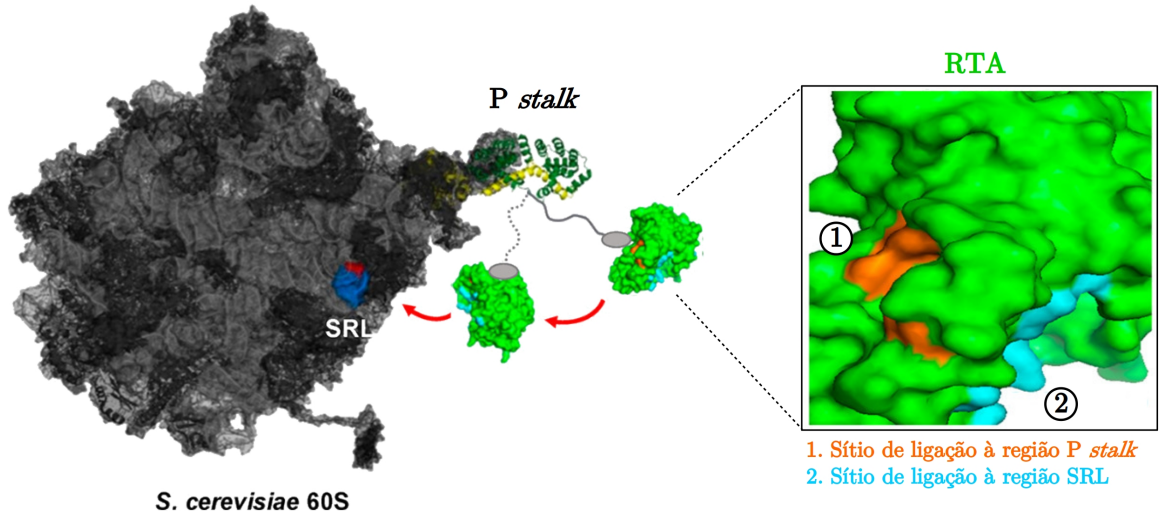
Sabe-se que apenas 5% da ricina internalizada chega ao RE, e lá, a ligação dissulfeto entre as subunidades é reduzida por uma isomerase específica[1] [28], e a RTA é então entendida pelo RE como uma proteína mal formada, e assim, é capturada e translocada ao citosol, onde deveria ser ubiquitinizada e depois degradada, porém, isto não ocorre. Isso porque a RTA tem baixa incidência de resíduos de lisina, o que é essencial para o endereçamento ao proteassoma, e uma vez no citosol, esta subunidade é capaz de atacar com grande eficiência a região SRL do ribossomos [29].

Apesar desta região – o SRL – ser altamente conservada nos três domínios da vida, apenas ribossomos eucarióticos são depurinados de forma eficiente pela RTA. Tal especificidade ocorre devido à dependência de interações intermoleculares com proteínas unicamente eucarióticas [5]. No caso da ricina, o atracamento da RTA com a região *P stalk* do ribossomo é crucial para o direcionamento à região SRL (Figure 4) [30].

O *P stalk* corresponde a um complexo pentamérico que contém as proteínas P1 e

[1] Essa isomerase se liga na RTB e é capaz de reduzir ou formar a ligação dissulfeto entre as subunidades. Em eucariotos existem dois sistemas de troca tiol-dissulfeto: Tioredoxina (Trx)/Tioredoxina redutase (TrxH) e glutaredoxina e glutationa redutase. A PDI sozinha é pouco eficiente em reduzir a ricina em suas subunidades constituintes. Por outro lado, quando a PDI e Trx estão na presença de TrxR e NADPH a ricina é reduzida

Figura 4 – Reconhecimento da RTA à região *P stalk* do ribossomos.



Fonte: Adaptado de Li et al.[38], (2020, p.1895).

P2 atracadas a proteína uL10 – ou P0 –, na forma de dois heterodímeros. Estudos iniciais mostraram que ribossomos mutantes, sem as proteínas P1 e P2, interagem fracamente com a RTA [31, 32]. Posteriormente foi verificado que a RTA interage com a região C-terminal da proteína P2, especificamente com os seis resíduos finais, são eles: GFGLFD [33, 34, 35]. Recentemente, estudos de mutação sítio-dirigida e cristalografia de raios–X propiciaram um melhor entendimento das interações intermoleculares do complexo RTA-P2/CTD (CTD, do inglês *C-terminal domain*), o qual evidenciou interações eletrostáticas e hidrofóbicas cruciais para o ancoramento da RTA nesta região do ribossomos. Em suma, foi visto que o resíduo R235 da RTA forma duas ligações de hidrogênio com os dois resíduos finais de P2, são eles: F114 e D115. Somado a isso, foi constatado que alguns resíduos hidrofóbicos também são importantes para o ancoramento na proteína P2, são eles: L232, Y183 e F240, sendo o primeiro o mais crítico [36, 37]. Em suma, a RTA se liga na região C-terminal da proteína P2 do *P stalk*, que por sua vez apresenta uma "dobradiça" flexível que orienta a RTA em direção ao SRL por meio de uma mudança conformacional.

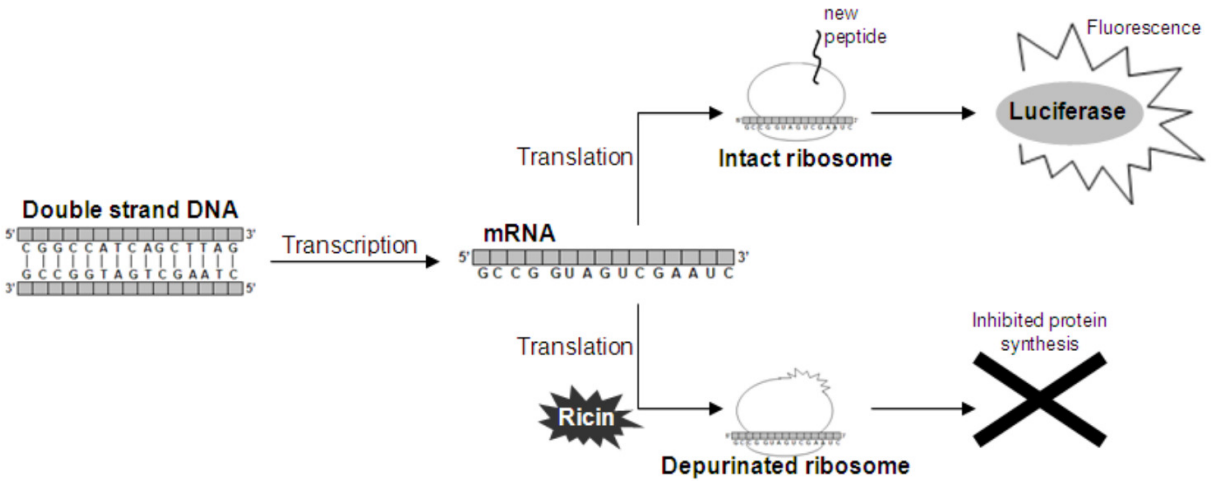
Por último, a descrição desta via leva à entender que a toxicidade da ricina está diretamente relacionada com a depurinação e consequente impedimento da síntese de proteínas. No entanto, alguns estudos indicam que o efeito tóxico da ricina no metabolismo é multifatorial e envolve inúmeras vias inflamatórias que levam à morte celular [5].

2.2 Contramedidas para Intoxicação por Ricina

Até o momento não existe uma contramedida clínica aprovada para intoxicação por ricina [39]. Estudos pré-clínicos têm sido conduzidos utilizando abordagens diversas, sendo as principais: (*i*) bloqueadores do TVR; (*ii*) anticorpos; e (*iii*) bloqueadores do sítio ativo da RTA. Porém, antes de adentrar no referencial teórico desses três itens, abordemos alguns aspectos metodológicos da análise *in vitro* para mensuração da toxicidade desta potente enzima.

A habilidade da RTA em inibir a síntese de proteínas pode ser monitorada em um único passo de leitura durante o processo de transcrição → tradução, por meio da detecção de luminescência oriunda da reação da luciferase (Figura 5).

Figura 5 – Modelo de ensaio *in vitro* para mensurar a atividade catalítica da RTA.



Fonte: Vieira e Tavares Machado[40] (2012, p. 292).

Neste ensaio, são realizadas as seguintes rotinas: (*i*) uma sequência de DNA é transcrita em mRNA; (*ii*) o mRNA recém transcrito é traduzido em proteína por um lisado de células que por sua vez contém ribossomos, entre outros componentes necessários para tradução. Uma vez que a proteína traduzida é a luciferase, a luminescência gerada serve como indicador da síntese de proteínas, e na ausência dela, indica que esta foi inibida. A inibição da síntese de luciferase pela ricina foi alcançada administrando a dose mínima numa concentração de 30.2 pM [40]. Desde então esta técnica vem sendo utilizada para mensurar a eficiência de bloqueadores da RTA, e comumente citada por ensaio *cell–free*.

2.2.1 Bloqueadores do Transporte Vesicular Retrógrado

Após a internalização da ricina na célula-alvo, esta pode seguir pelo complexo golgiense até o RE por uma via intracelular denominada transporte vesicular retrógrado (Seção 2.1). Devido a importância desta rota para a toxicidade da ricina, alguns estudos de prospecção por moléculas anti–ricina tiveram como alvo esta via. Na Tabela 1 estão apresentadas as estruturas químicas das moléculas que demonstraram atividade anti-ricina por modulação do TVR.

Yoshida et al. [41] avaliaram o potencial anti–ricina da brefeldina A (BFA), um anti-fúngico conhecido por inibir o transporte anterógrado de proteínas para o complexo golgiense. Os autores constataram que BFA inibiu a citoxocidade da ricina nas linhagens NRK, CHO e *Vero cells*, entretanto, um estudo posterior mostrou que em outras circunstâncias essa molécula é capaz de aumentar a toxicidade da ricina [42].

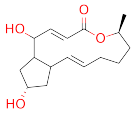
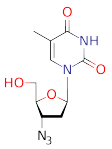
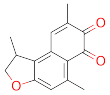
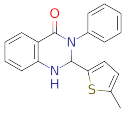
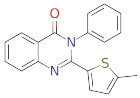
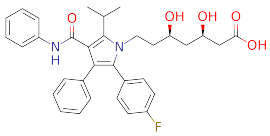
Wellner, Pless e Thompson [43] avaliaram a capacidade protetora do composto 3-azido-3-deoxitimidina (AZT) nas linhas CHO e *Vero cells* quando expostas à ricina. Neste estudo, os autores constaram que o AZT exerce proteção às células por interferir no processo de translocação da ricina.

Nambiar, Murugesan e Wu [44] avaliaram o potencial anti-ricina da Mansonona D na linhagem celular *Vero* e constataram um efeito protetor dose–dependente, no entanto, tal proteção é mediada por uma dramática alteração morfológica do complexo golgiense.

Stechmann et al. [45] realizaram uma triagem usando a abordagem de *High-Throughput Screening* para uma biblioteca com mais de 16.000 moléculas e encontraram dois compostos com alto potencial de inibir a citotoxicidade da ricina em linhagem celular Hella, são eles: Retro-1 e Retro-2. Ambas atuam de forma seletiva na interface entre a rede trans-golgi (TGN, do inglês *Trans-Golgi Network*) e o endossomo primário, sem alterar a morfologia do Golgi. Além disso, ensaios *in vivo* mostraram que apenas a Retro-2 é capaz de proteger contra uma dose letal de ricina.

Yu et al. [46] conduziram um estudo baseado em similaridade farmacofórica utilizando inibidores do transporte vesicular retrógrado, tal como a Retro-2 e seu análogo (o composto DA2MT, que também apresentou efeito protetor em ensaios com célula). Em suma, este estudo forneceu novas perspectivas sobre a identificação e otimização de compostos anti-ricina. Além disso, Bassik et al.[47] demonstraram que o transporte da ricina ao RE pode ser bloqueado *in vitro* pela atorvastatina, um inibidor da hidroximetilglutaril (HMG)–CoA redutase.

Tabela 1 – Moléculas com atividade anti-ricina por modulação do TVR.

Inibidor	Estrutura Química	Referência
Brefeldina A		[41]
3-azido-3-deoxitimidina		[43]
Mansonona D		[44]
Retro-2		[45]
DA2MT		[46]
Atorvastatina		[47]

2.2.2 RiVax

Os estudos conduzidos por Smallshaw e colaboradores em 2002 [48] e 2005 [49] sugerem a RiVax® como uma contramedida eficaz para combater a intoxicação por ricina. Esta, por sua vez, corresponde a uma forma inativa da RTA, tendo em sua estrutura as mutações V76M e Y80A. A primeira está relacionada com capacidade da RTA em induzir a "Síndrome de Extravasamento Capilar Sistêmico" e a última, e não menos importante, está relacionada com o reconhecimento e atracamento do substrato no sítio ativo da RTA (Seção 2.1).

De modo geral, esta contramedida tem por objetivo preparar o sistema imunológico para reconhecer e atacar a forma selvagem da ricina se esta vir a surgir no organismo. Vitetta et al. [50] realizaram a vacinação de um conjunto de indivíduos e demonstraram que a RiVax® é capaz de induzir a produção de anticorpos anti-ricina, no entanto, sua duração após três aplicações foi abaixo do ideal. Atualmente o estudo segue na fase 1B de ensaios clínicos[2].

[2] Para mais informações, acesse: https://clinicaltrials.gov/ct2/show/NCT00812071

2.2.3 Bloqueadores do Sítio Ativo da RTA

Uma vez que o substrato natural da RTA está inserido na região SRL do ribossomos, alguns esforços para síntese de ligantes antagonistas da RTA tiveram como molde esta região. Hesselberth et al. [51] avaliaram 31 aptâmeros de RNA que continham toda sequência e estrutura necessária para interagir com a RTA. Um dos aptâmeros testados foi capaz de inibir a atividade catalítica da RTA, com $K_D = 7.3\ nM$. Entretanto, apesar da alta afinidade, alguns autores sugerem que tais inibidores não são viáveis para o uso como antídoto, principalmente pela baixa permeabilidade na bicamada lipídica[3] [52]. Sendo assim, sugere-se que moléculas orgânicas menores podem desempenhar um papel mais eficaz. Neste aspecto, atualmente existem três classes de bloqueadores, descritas à seguir.

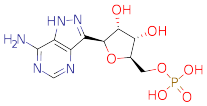
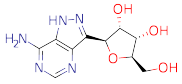
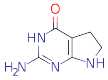
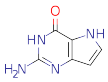
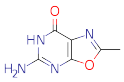
2.2.3.1 Derivados de purina

Esta classe de derivados mimetiza as interações intermoleculares do substrato natural e têm como plataforma química as bases nitrogenadas, adenina e guanina. Apesar de não terem apresentado atividade antagonista significativa, os estudos de cristalografia dos complexos RTA–Purina proveram um melhor entendimento do sítio ativo e mecanismo de catálise da RTA [26].

Os estudos com derivados baseados em adenina foram apresentados na Seção 2.1, é o caso dos ligantes FMP e ApG. Quanto aos baseados em guanina, estes foram propostos inicialmente como plataforma química de inibição da RTA por Yan et al. [53], utilizando de métodos computacionais baseados em cálculos de mecânica molecular e mecânica quântica (semi-empírico e *ab initio*). Posteriormente, Miller et al. [54] retornaram a esta ideia, porém, utilizaram do método de *docking* receptor–ligante para planejar a síntese de 11 ligantes protótipo baseados em guanina. Ensaios de validação *in vitro* destes compostos mostraram uma capacidade protetora moderada, com valores de IC_{50} no intervalo de 2.800 a 400 μM [54]. No mesmo estudo, a análise estrutural do cristal RTA–Ligante mostrou que estes derivados se ligam no sítio ativo em sua forma "aberta" e os anéis heterocíclicos compartilham algumas das interações intermoleculares do FMP e adenina com resíduos do sítio ativo, tal como: (*i*) ligações de hidrogênio entre o oxigênio da cadeia principal de V81 e G121 e a amina exocíclica do ligante; e (*ii*) ligação de hidrogênio entre a amida da cadeia principal de V81 e o átomo N3 do ligante. Na Tabela 2 é apresentado um sumário de resultados dos derivados de purina com estrutura do complexo RTA-Ligante resolvida e depositada no PDB.

[3] No mesmo estudo, os autores sugerem que estes ligantes podem ter um melhor aproveitamento como nanosensores.

Tabela 2 – Derivados de purina com estrutura do complexo RTA–Ligante depositada no PDB.

Inibidor	Estrutura	IC_{50} (μM)	PDB	Ref.
FMP		43*	1FMP	[25]
Formicina		-	1IFU	[19]
Adenina		-	1IFS, 2P8N	[19, 55]
7-deazaguanina		2.800	1IL3	[54]
9-deazaguanina		1.400	1IL4	[54]
8-metil-9-oxoguanina		400	1IL9	[54]

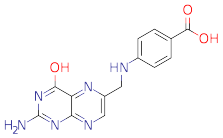
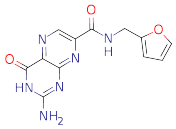
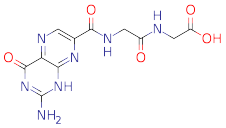
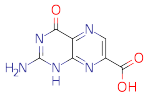
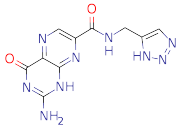
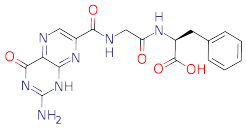
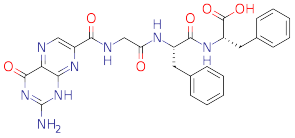
*: Resultado experimental de K_i (nM);

2.2.3.2 Derivados de pterina

Yan et al. [56] conduziram um estudo de *docking* receptor–ligante para a forma "aberta da RTA" e observaram que o ácido pteróico (PTA), um derivado da pterina, apresentava alto grau de complementariedade com sítio ativo da RTA. Por conseguinte, ensaios de validação *in vitro* mostraram que este ligante inibia a atividade da RTA com $IC_{50} \approx 600\mu M$ [56], porém, apresentava problemas quanto a solubilidade. No mesmo estudo, a análise estrutural do cristal RTA–PTA confirmou o que tinha sido previsto pelo *docking*. O grupo pterina interage de modo similar ao FMP no sítio ativo da RTA, o que inclui: (i) interações do tipo $\pi - \pi$ *stacking* com os resíduos Y80 e Y123; (ii) ligações de hidrogênio entre os resíduos V81, G121 e a amina exocíclica do ligante; (iii) ligações de hidrogênio com os resíduos R180 e N78 [56].

Devido a sua habilidade de interagir com os resíduos do sítio ativo, a pterina tornou-se uma das principais plataformas para o desenho de novos ligantes. Esforços seguintes tiveram como objetivo o melhoramento da solubilidade, como também o aumento da potência por meio da inserção de um grupo R que ocupasse o segundo sítio (Tabela 3).

Tabela 3 – Derivados de pterina com estrutura do complexo RTA–Ligante depositada no PDB.

Inibidor	Estrutura Química	IC_{50} (μM)	PDB	Referência
Ácido pteróico		600	1BR6	[56]
JP3		380	3PX9	[57]
19J		300	4HV7	[58]
JP2		230	3PX8	[57]
0RB		70	4ESI	[59]
RS8		20	4HUO	[58]
19M		15	4HUP	[58]

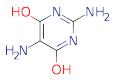
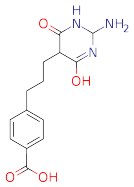
2.2.3.3 Derivados de pirimidina

Miller et al. [54] observaram que o 2,5-diamino-4,6-dihidroxipirimidina (DDP), um derivado de pirimidina, apresentava *in silico*, interações intermoleculares similares ao PTA. Entretanto, a análise estrutural do cristal RTA–DDP mostrou que este ligante interagia de forma hidrofóbica com a cadeia lateral de Y80, impedindo a sua rotação, e portanto, mantendo a RTA em sua conformação "fechada", incapaz de reconhecer o substrato específico (Figura 6A). A princípio este achado pareceu animador, porém, no mesmo estudo, os autores constataram que o DDP era um antagonista fraco, com IC_{50} na faixa de milimolar.

Bai, Monzingo e Robertus [60] realizaram um procedimento de triagem virtual baseado em *docking* receptor–ligante numa base de dados contendo 48.500 ligantes e identificaram o composto PBA como inibidor antagonista da RTA, com $IC_{50} = 270\mu M$. A análise estrutural do cristal RTA–PBA mostrou que este ligante interage no sítio ativo em sua forma "aberta", assumindo uma conformação similar aos ligantes baseados em purina

e pterina. A grande vantagem desta classe de inibidores está relacionada a solubilidade. Os autores estimaram que o PBA é cerca de 10 vezes mais solúvel que o PTA, e portanto, um bom candidato para otimizações.

Tabela 4 – Derivados de pirimidina com estrutura do complexo RTA–Ligante depositada no PDB.

Inibidor	Estrutura Química	IC_{50} (μM)	PDB	Referência
DDP		2.200	1IL5	[54]
EJ5		270	3EJ5	[60]

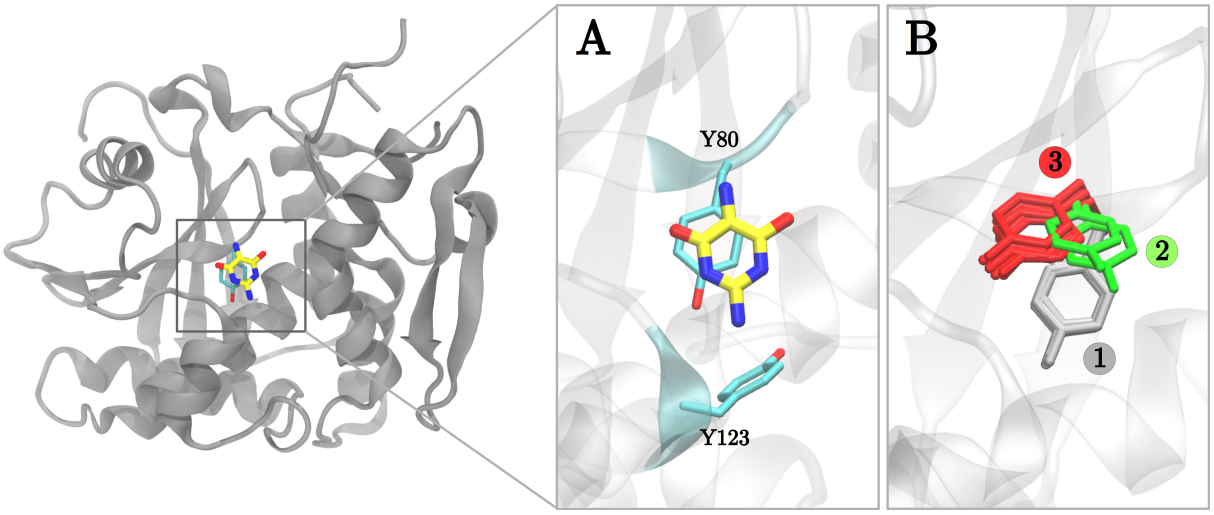
2.2.4 Inibidores putativos do sítio ativo da RTA

Pang et al. [61] exploraram a hipótese de estabilização do resíduo Y80 na conformação "fechada" (Figura 6A). Aqui, os autores avaliaram 13 estruturas de complexos RTA–Ligante e concluíram que o resíduo Y80 ocorria em 3 conformações distintas (Figura 6B). Indicaram a conformação 3 como sendo a ativa, necessária para catálise, logo, as moléculas que se ligam nesta conformação são capazes de inibir a RTA por competição direta com a adenina do SRL. No entanto, de acordo com os autores, essa estratégia envolve o grande desafio de superar as interações eletrostáticas da interface RTA–SRL, e por isso, atentaram-se em estabilizar Y80 na conformação 2, "não–ligante". Em linhas gerais, realizaram um estudo de *docking* receptor–ligante para uma base de dados contendo ∼200.000 ligantes, seguido de validação *in vitro*, e constataram que três compostos apresentavam percentual de inibição da atividade da RTA em mais de 30%. Por fim, sugerem que essa capacidade protetora ocorre pela estabilização de Y80 na conformação "não–ligante", porém, estudos de validação estrutural (cristalografia de raios–X) não foram realizados.

Wahome et al. [62] empregaram a técnica de HTS em linhagem celular *Vero* para avaliar o potencial anti–ricina de 80.000 compostos. Destes, 20 foram capazes de reduzir os efeitos citotóxicos da ricina em mais de 50%, sendo os compostos 1 e 8 os mais eficazes. No mesmo estudo, os autores conduziram o ensaio de tradução *in vitro* para avaliar se estes ligantes exerciam sua atividade diretamente na RTA, entretanto, o composto 1 se mostrou um fraco inibidor, com $IC_{50} = 1mM$. Em contraste, o composto 8 apresentou resultados de inibição mais animadores, com $IC_{50} = 30\mu M$. Ainda nesse estudo, um modelo teórico de interação destes compostos com a RTA foi proposto por meio do método de *docking*

receptor–ligante, onde sugerem que a conformação "fechada" é a mais provável, porém, a validação por difração de raios–X do complexo RTA–Ligante não foram realizados.

Figura 6 – Estrutura cristalográfica do complexo RTA–DDP e conformações do resíduo Y80.



Diferentes conformações de Y80 estão representadas em cinza, verde e vermelho. As estruturas cristalográficas que representam a conformação 1 são: 1IFT, 1RTC, 1IL5, 2AAI; conformação 2: 1IFS e 1APG; e conformação 3: 1BR5, 1BR6, 1IL3, 1IL4, 1IL9, 1IFU e 1FMP. Fonte: O autor; ilustração feita com o programa VMD.

2.3 Triagem Virtual

A triagem virtual no contexto da descoberta de novos fármacos, corresponde ao uso de métodos teóricos/computacionais para avaliar dentre milhares de moléculas, àquelas com potencial de atividade biológica desejada. Por conseguinte, as moléculas selecionadas são encaminhadas para validação experimental, e uma vez confirmado o seu potencial, esta pode seguir para os testes clínicos [63].

Na seção 2.2, vimos que alguns grupos de pesquisa utilizaram a informação estrutural do sítio ativo da RTA acoplado a métodos computacionais, tal como o *docking* receptor–ligante, para avaliar a complementariedade química e afinidade de ligantes protótipo. Alguns obtiveram sucesso, outros não. E mesmo para os bem sucedidos, em alguma fase posterior, tais compostos apresentaram problemas externos à interação proteína–ligante, tal como propriedades farmacocinéticas ruins. Além disso, abordagens em larga escala como o HTS também foram utilizadas para avaliar a capacidade protetora de milhares de moléculas contra ação da ricina. No entanto, para a pequena parte que apresentou especificidade de interação com o sítio ativo da RTA, os dados experimentais não foram animadores.

Devido a importância dos métodos computacionais para o planejamento racional de inibidores de RTA, nesta seção será apresentado uma revisão sobre estes métodos, que atualmente seguem agrupados em duas perspectivas: planejamento racional de compostos bioativos (*i*) baseado na estrutura do ligante; e (*ii*) baseado na estrutura do alvo. Inicialmente será apresentado os pressupostos teóricos de ambos, no entanto, os métodos dependentes da estrutura do alvo serão discutidos com mais detalhes.

2.4 Planejamento Racional de Compostos Bioativos

Desde a descoberta de um novo princípio ativo até estar disponível para os pacientes, um fármaco passa por um longo e custoso processo de pesquisa e desenvolvimento. Em média, isso pode levar de 10 a 15 anos e um gasto de até US$ 1,5 bilhão para cada novo fármaco desenvolvido. Nos estágios iniciais deste processo realiza-se triagens em larga escala envolvendo desde dezenas até milhares de moléculas. Aqui, a indústria farmacêutica comumente adota a abordagem experimental do HTS para avaliar possíveis ligantes contra alvos terapêuticos relevantes, e os que atingem o potencial de atividade desejado são considerados como "líderes" [64].

Porém, a literatura da área sugere que o uso majoritário do HTS confere algumas desvantagens. Além do alto custo, este método não oferece uma certeza quanto ao mecanismo de ação do ligante, o que dificulta o processo de otimização de propriedades farmacofóricas, por exemplo [64]. Até meados da década de oitenta, medicamentos eram descobertos dessa forma, sem conhecimento dos mecanismos de ação e/ou interação da molécula ligante. Esse "gargalo" fez emergir o emprego de métodos teóricos/computacionais capazes de utilizar a informação química relacionada a um alvo terapêutico para avaliar dentre milhares de ligantes, àqueles com potencial de afinidade promissor [63, 65]. O conceito de afinidade será discutido mais a frente, no momento, é importante entender apenas que a abordagem computacional é capaz de reduzir consideravelmente o custo e tempo para obtenção de resultados.

Ao longo dos anos, com o avanço das técnicas nas áreas da biologia molecular e química, foi possível a identificação e resolução de alvos moleculares relacionados a diversas doenças, o que impulsionou ainda mais o uso desses métodos teóricos, agora denominados como "Planejamento Racional de Fármacos Guiado por Computador". A notoriedade surgiu com os relatos de sucesso, e atualmente, é parte central para o desenvolvimento de um novo agente terapêutico, sendo usado como alternativa ou um complemento ao HTS [64]. O processo de avaliar milhares de ligantes utilizando métodos computacionais é denominado de Triagem Virtual (TV). Sabendo disso, podemos passar adiante e explorar alguns conceitos e técnicas dos dois principais grupos de métodos utilizados em TV, são eles: (*i*) Triagem Virtual Baseada na Estrutura do Ligante (TVBEL); e (*ii*) Triagem Virtual Baseada na Estrutura do Alvo (TVBEA).

2.4.1 Triagem Virtual Baseada na Estrutura do Ligante

Na TVBEL, leva-se em conta a informação química de moléculas ativas e/ou inativas de um alvo terapêutico como um "modelo de consulta" para realização do processo de triagem. Em outras palavras, esta abordagem utiliza o pressuposto de que moléculas com estrutura química similar à compostos bioativos ou não, podem apresentar também atividade biológica similar. Talvez o método mais utilizado por este tipo de abordagem seja

a Relação Quantitativa Estrutura–Atividade (QSAR, do inglês *Quantitative Strucuture–Activity Relationship*). Este método utiliza da informação de descritores físico–químicos de ligantes ativos e inativos – específicos para um alvo ou não – assim como suas respectivas atividades biológicas para sistematizar um modelo estatístico de regressão. Outros métodos TVBEL incluem: (*i*) triagem por similaridade estrutural (2D ou 3D) [66]; (*ii*) triagem por similaridade farmacofórica [67].

2.4.2 Triagem Virtual Baseada na Estrutura do Alvo

Já no TVBEA, utiliza-se a informação tridimensional da macromolécula receptora como critério de seleção dos compostos "líderes", tendo como base as interações intermoleculares inerentes ao reconhecimento molecular receptor–ligante.

O primeiro passo consiste em definir o alvo molecular a ser utilizado, o que normalmente é uma proteína vital ao processo biológico no qual objetiva-se bloquear ou ativar. A estrutura tridimensional desses bioreceptores pode ser obtida por meio de técnicas experimentais, tal como a difração de raios–X e a ressonância magnética nuclear (RMN), ou por métodos teóricos como a modelagem comparativa. Atualmente, o *Protein Data Bank* (PDB) [68] corresponde ao principal acervo público dessa informação, contendo ∼150.000 estruturas. Quanto aos ligantes, pode ser utilizado a estrutura de compostos já sintetizados no próprio laboratório, no entanto, o espaço químico amostrado por esta perspectiva pode ser pequeno. Em estudos de TV que almejam a inovação química para um determinado alvo molecular, é comum utilizar banco de ligantes contendo milhões de estruturas. Nesta perspectiva, existem empresas farmacêuticas e bancos de dados públicos, tal como o ZINC15 [69] (www.zinc15.docking.org), que dispõem gratuitamente a estrutura química de milhares de ligantes – sinteticamente viáveis – para estudos baseados em TV. Tendo isso – estrutura do alvo e ligantes –, a triagem ocorre de forma similar ao método tradicional utilizado no HTS [63]. A diferença agora é que tudo acontece *in silico*.

Por fim, estão inclusos no TVBEA os seguintes métodos: (*i*) triagem farmacofórica baseada na estrutura do alvo; (*ii*) *docking* receptor–ligante; (*iii*) dinâmica molecular; e (*iv*) planejamento *de novo*. Para fins de melhor entendimento dos resultados obtidos nesta tese, será priorizado nas próximas seções a explicação das técnicas dependentes do alvo dos itens *i*, *ii* e *iii*. No entanto, antes disso será revisado alguns pressupostos importantes quanto ao reconhecimento molecular receptor–ligante.

2.5 Reconhecimento Molecular (Receptor–Ligante)

A resposta biológica desencadeada por um fármaco pode ocorrer por interações específicas e/ou inespecíficas com uma macromolécula receptora. Isso porque algumas dessas respostas, tal como as de anestésicos em geral, dependem exclusivamente de propriedades físico–químicas, tal como o coeficiente de partição ($logP$) e pKa. Estes são conhecidos como fármacos "estruturalmente inespecíficos", e não são o foco desta tese. Do outro lado temos os fármacos "estruturalmente específicos", que dependem da interação com uma macromolécula biológica para desencadear uma resposta [70]. Como já citado, normalmente este bioreceptor é uma proteína, porém, existem ligantes que atuam em outras macromoléculas biológicas, como o DNA, por exemplo [71, 72]. Na estrutura desses receptores existem cavidades ou regiões específicas que estão diretamente relacionadas com a sua atividade biológica, e correspondem à região ao qual esse grupo de fármacos interage [63, 70]. A resposta desencadeada por essa interação pode vir a intensificar ou bloquear a efetividade biológica da macromolécula alvo. Aqui, eventos físico–químicos governam este reconhecimento e exprimem de modo quantitativo a afinidade do ligante pelo sítio receptor, tal como: (*i*) forças eletrostáticas; (*ii*) forças de dispersão; (*iii*) efeito hidrofóbico; (*ii*) ligações de hidrogênio; e (*iii*) ligações covalentes [63, 70]. Contudo, a resposta biológica relacionada a este reconhecimento depende também da flexibilidade conformacional de ambos. O paradigma atual afirma que a estrutura do bioreceptor pode existir em conformações similares e energeticamente equivalentes. Desta forma, o ligante, que pode ser flexível ou não, interage com alta afinidade em uma das conformações preexistentes do receptor [63, 70].

Em suma, complexos receptor–ligante de alta afinidade apresentam as seguintes características: (*i*) complementariedade estérica; (*ii*) alto índice de complementariedade tanto de grupos apolares como de grupos polares e carregados; e (*iii*) poucas ou nenhuma interação repulsiva [63, 70]. Em termos físico–químicos, o reconhecimento molecular é dirigido por efeitos entálpicos e entrópicos, sendo a energia livre de ligação de Gibbs (ΔG_{lig}) a constante que equaciona esses termos e expressa a afinidade. O ΔG_{lig} está diretamente relacionado a constante de inibição, K_i, que pode ser medida experimentalmente,

$$\Delta G_{lig} = \Delta H_{lig} - T\Delta S_{lig} = RT \ln K_i \tag{2.1}$$

sendo ΔH a variação de entalpia, T a temperatura – em Kelvins (K) –, ΔS a variação de entropia e R a constante universal dos gases.

O K_i pode ser obtido por meio de abordagens teóricas, tal como o método de perturbação de energia livre (FEP, do inglês *Free Energy Pertubation*), porém, por ser baseado em DM, seu uso em larga escala torna-se proibitivo para um número muito elevado de ligantes, devido ao alto custo computacional envolvido [63]. Dentre os métodos inclusos na TVBEA, o *docking* receptor–ligante torna-se atrativo para uso em larga escala

devido ao custo computacional razoável para determinação do K_i. Na seção seguinte será apresentado uma descrição mais detalhada deste método.

2.6 *Docking* Receptor–Ligante

A predição da afinidade de um ligante é estimada pelos algoritmos de *docking* em duas etapas principais: (*i*) investigação conformacional do ligante no sítio receptor; (*ii*) quantificação da afinidade do modo de ligação predito. No item *i*, os algoritmos de *docking* exploram o problema por meio de três abordagens:

- *docking* rígido;
- *docking* semi–flexível;
- *docking* flexível.

Na primeira citação, o algoritmo de *docking* "entende" os resíduos do sítio receptor e ligante como rígidos, sendo que para este último, é possível explorar graus de liberdade rotacionais e translacionais. Já na segunda citação, os resíduos do sítio ativo novamente são tidos como rígidos, enquanto o ligante é flexível. Aqui, além dos graus de liberdade rotacionais e translacionais do ligante, explora-se também os graus de liberdade de ligações torsionáveis. Quanto a terceira citação, ambos são flexíveis [63, 73].

Para fins de definição, diz-se que a investigação conformacional do ligante é realizada pelos "métodos de busca", enquanto a afinidade é calculada pela "função de avaliação". Para este último, o algoritmo deve ser capaz de sistematizar as interações intermoleculares receptor–ligante preditas pelo "método de busca" e retornar uma estimativa da afinidade de ligação, que pode ser em termos de energia ou pontuação (*score*). Essa função além de ser computacionalmente viável, deve ser capaz de discernir modos de ligação energeticamente favoráveis de um ligante na cavidade de um bioreceptor, assim como discernir entre um ligante ativo – com maior afinidade – de um inativo. No mais, esse procedimento de busca e avaliação ocorre recursivamente até que um critério de parada seja atingido [63, 73].

Pagadala, Syed e Tuszynski [74] afirmam que as energias absolutas associadas a interação receptor–ligante não é estimada com precisão pelos algoritmos de *docking* atuais. Estes autores sugerem que aspectos físico–químicos importantes, tal como efeitos do solvente, efeitos entrópicos e flexibilidade do receptor, por serem negligenciados, afetam diretamente a acurácia das predições. Apesar disso, com base em uma extensa revisão de aplicações desses algoritmos, os autores apontam alguns programas de *docking* como sendo "razoáveis" para esta tarefa – predizer a afinidade de um ligante –, dentre eles, citam o GOLD, que por sua vez, tem seu método de busca baseado num algoritmo genético.

Algoritmos deste tipo utilizam dos pressupostos da teoria da evolução darwinista para explorar o espaço conformacional do ligante na cavidade do bioreceptor. Basicamente, uma população de soluções do ligante é proposta (indivíduos) e avaliada quanto a adaptabilidade frente ao ambiente (sítio receptor). Em uma nova geração, outros indivíduos são produzidos por meio de recombinação (*cross-over* entre dois indívudos "pais") e mutações. A nova geração é então avaliada e em seguida submetida a mais um processo de "evolução". O término deste procedimento ocorre quando o total de ciclos de evolução é atingido ou quando os novos indivíduos de uma geração não obtém critérios de adaptabilidade melhores que indivíduos de gerações passadas. Por fim, as soluções "melhor adaptadas" são apresentadas ao usuário no *ranking* final [63, 73].

Bai et al. [75] utilizaram o GOLD como etapa final em um estudo de triagem virtual para a ricina. Em suma, foi avaliado a afinidade de 50.000 ligantes para conformação "aberta" da RTA. Destes, 306 foram identificados pelo GOLD como potenciais inibidores. No mesmo estudo, ensaios de tradução *in vitro* mostraram que 6 ligantes apresentavam capacidade de inibir a RTA com IC_{50} entre 0.2–2.0 mM. Somado a isso, destes 6 ligantes, 2 foram capazes de proteger uma linhagem celular (*Vero cells*) quando exposta à ricina.

Choudhry et al. [76] utilizaram o GOLD num estudo de triagem virtual para o alvo NS2B/NS3, uma serina-protease importante para replicação do Zika vírus. Em suma, os autores exploraram o potencial de interação de $\sim$100.000 compostos numa cavidade específica deste bioreceptor e encontraram 7 candidatos com alto potencial de interação. No entanto, ensaios de validação *in vitro* não foram realizados.

Yennamalli et al. [77] avaliaram a capacidade de ligantes interferirem na transição conformacional da proteína E, que é a responsável no processo de reconhecimento e fusão das membranas do vírus da dengue à célula hospedeira, portanto, importante no ciclo de vida do vírus. Os autores utilizaram os programas DOCK, GOLD e FlexX para avaliar o potencial de interação de uma biblioteca de ligantes num sítio de ligação desta proteína e identificaram 7 compostos. Ensaios posteriores de validação *in vitro* mostraram que um dos ligantes apresentava IC_{50} na faixa de μM.

2.7 Triagem Farmacofórica Baseada na Estrutura do Alvo

A resposta biológica associada ao reconhecimento molecular receptor–ligante, como já dito, é dependente das interações que a molécula ligante realiza no sítio receptor. Nesse aspecto, entende-se por farmacóforo, o conjunto de propriedades estéricas e eletrônicas necessárias para garantir as interações ideais com a macromolécula alvo, responsáveis por desencadear ou bloquear a resposta biológica . Na triagem virtual baseada em farmacóforos, sugere-se que ligantes com propriedades farmacofóricas similares à de compostos bioativos podem também apresentar atividade biológica equivalente [78].

Este padrão de interação pode ser obtido a partir de estruturas de complexos receptor–ligante resolvidos por métodos experimentais, tal como a difração de raios–X e RMN[4] e por seguinte transferido para um modelo 3D abstrato, que comporta a disposição dessas interações ao invés do grupo funcional propriamente dito. Os principais tipos de interação contidos neste modelo são:

- Aceptores/doadores de ligações de hidrogênio;
- Interações entre grupos carregados, cátion ou ânion;
- Contatos hidrofóbicos;
- Aromaticidade;
- Volume.

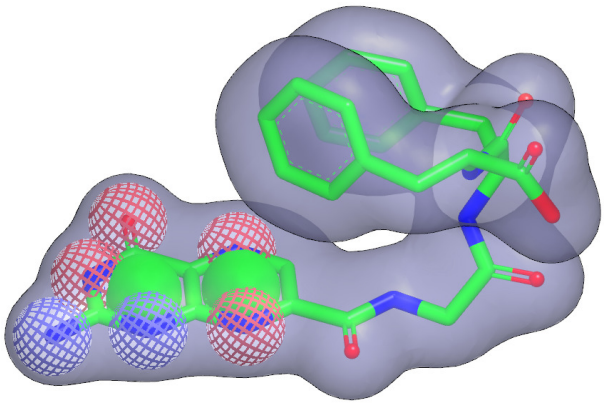
O processo de triagem ocorre agora pela sobreposição entre o modelo farmacofórico de consulta – extraído a partir dos dados experimentais de difração de raios–X ou RMN – e o modelo farmacofórico dos diferentes ligantes a serem avaliados. E de forma similar ao *docking*, um *score* é estimado, porém, ao invés da afinidade, quantifica-se a similaridade para cada sobreposição realizada. Portanto, se o modelo farmacofórico utilizado como consulta compreende a um ligante com potencia na faixa de nM, espera-se que compostos muito similares a este também apresentem atividade biológica similar [78].

Para fins de validação, comumente utiliza-se um conjunto de moléculas ativas e inativas (biblioteca de validação) para testar a capacidade de discernimento do modelo de consulta. Somado a isso, análises estatísticas devem ser empregadas para quantificar a capacidade preditiva, como por exemplo: (i) especificidade; (ii) sensitividade; e (iii) área sobre a Curva ROC. Dependendo dos resultados, refinamentos podem ser propostos no modelo de consulta, de forma a otimizar o seu discernimento em identificar moléculas ativas das inativas [78].

[4] O *Protein Data Bank* mais uma vez corresponde ao principal acervo digital dessa informação, contendo milhares de estruturas 3D de alvos terapêuticos complexados a ligantes.

Como exemplo, na Figura 7 é apresentado um modelo farmacofórico do ligante 19M no sítio ativo da RTA. Este derivado da pterina corresponde a um dos inibidores mais potentes até então descritos, com $IC_{50} = 15\mu M$ e estrutura cristalográfica resolvida e depositada no PDB (Tabela 3). Aqui, as interações que o grupo pterina realiza no sítio ativo da RTA foram mantidas no modelo, são elas: (*i*) doador de ligações de hidrogênio com o resíduo G121; (*ii*) aceptor de ligações de hidrogênio com os resíduos V81, Y80, Y123; e (*iii*) dois anéis aromáticos, que por sua vez interagem com os resíduos Y80 e Y123. Quanto à cadeia lateral deste derivado, ao invés dos grupos funcionais, foram mantidas as propriedades de volume e forma.

Figura 7 – Modelo farmacofórico do ligante 19M.



As esferas em verde, azul e vermelho ilustram a presença de anel aromático, doador e aceptor de ligações de hidrogênio, respectivamente. Em cinza é ilustrado o volume. Fonte: O autor; ilustração feita com o programa vROCS (versão 3.3.0.3; OpenEye Scientific®).

Temml et al. [79] aplicaram um protocolo de triagem virtual baseado na comparação de farmacofóros para encontrar ligantes seletivos dos receptores nucleares LXRα e LXRβ, envolvidos com o metabolismo do colesterol e quadros graves da doença aterosclerose. Nesse estudo, os autores discutem que a ativação destes receptores tanto por ligantes endógenos como sintéticos, aumenta o efluxo do colesterol das células, que por conseguinte é metabolizado até a sua completa eliminação, melhorando quadros graves da doença. Em linhas gerais, nesse estudo foi avaliado a similaridade farmacofórica de ~200.000 moléculas contra um modelo de consulta obtido a partir de três estruturas cristalográficas proteína–ligante de ambos receptores e encontraram 2 compostos com capacidade de ativar ambos os receptores.

Quanto a estudos de triagem utilizando esse tipo de abordagem para a ricina, até o momento nenhum trabalho foi realizado.

2.8 Dinâmica Molecular de Proteínas

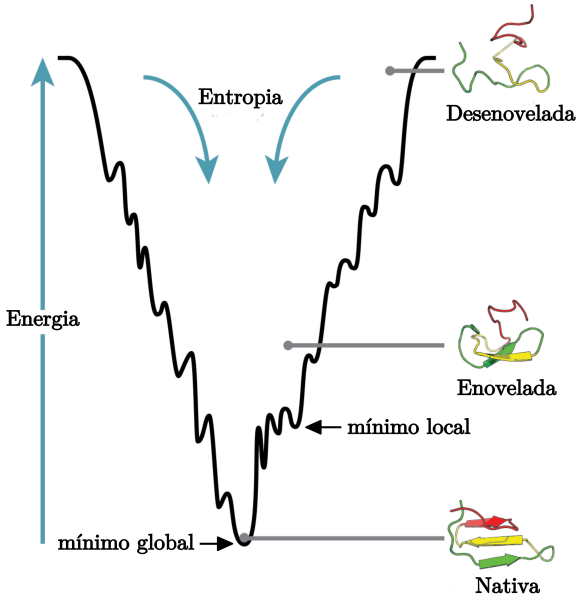
O conhecimento da estrutura tridimensional de proteínas em escala atômica é um dos fatores fundamentais para o entendimento de sua função biológica. Com o advento da determinação das estruturas da mioglobina e hemoglobina por Kendrew [80, 81] e Perutz [82], respectivamente, os estudos quanto a relação estrutura–função dos mais variados tipos de proteínas ganharam notoriedade e seguem em ascensão até então. Porém, assumir a interpretação das posições atômicas obtidas por essas técnicas pode levar ao falso entendimento de que a função está diretamente relacionada a uma única estrutura nativa. Isso porque flutuações estruturais de um pequeno conjunto de átomos, tal como a rotação da cadeia lateral de um resíduo até flutuações de maior amplitude, que envolvam motivos ou até mesmo domínios da proteína, podem estar relacionados a sua função. Apesar de métodos experimentais como a difração de raios–X e o RMN conseguirem acessar a informação estrutural de uma proteína em estados diferentes – com e sem ligante, por exemplo –, a dinâmica do processo que interliga ambos os estados é medida de forma indireta [63, 83, 84].

Nesse aspecto, a topografia que reúne toda a superfície de energia dos microestados acessíveis de uma proteína é denominada de perfil energético, e para maioria das estruturas proteicas este perfil assemelha-se a um "funil rugoso", como ilustrado na Figura 8A. Nessa representação, os poços de mínimo na superfície de energia caracterizam diferentes estados da proteína, separados entre si por uma barreira energética denominada de estado de transição. A altura dessa barreira determina a taxa de acessibilidade entre os estados, podendo envolver flutuações mais rápidas, que ocorrem no intervalo de tempo de ps à ns, até flutuações de maior amplitude, tal como movimentos coletivos (μs) e lentos (ms). Além disso, este perfil é dependente de certas condições, tal como temperatura, pressão, tipo de solvente e pH. Portanto, a manipulação destes parâmetros têm implicações diretas na população dos microestados, assim como na cinética de interconversão entre eles [83, 84].

Em condições fisiológicas, a estrutura desnaturada de uma proteína, ou seja, sem função biológica específica, é caracterizada por um alto grau de interações com o solvente, e portanto, uma alta entropia conformacional. Na Figura 8A, esse estado encontra-se no topo do funil. Partindo daí, à medida que caminhamos pela superfície de energia, passamos por diferentes poços de mínimo e seus respectivos estados de transição, até chegar na região mais afunilada (ou mínimo global), onde encontramos os microestados que compreendem à estrutura nativa. Nessa região, as interações intramoleculares encontram-se otimizadas, isto é, os resíduos da cadeia polipeptídica que antes interagiam amplamente com o solvente, agora o fazem entre as diferentes cadeias da proteína. Aqui, as interações que contribuem para estabilização da estrutura nativa são, além das ligações dissulfeto, interações como ligações de hidrogênio e o tão importante efeito hidrofóbico. Somado a isso, essa região também caracteriza-se pela ocorrência de flutuações rápidas (ps à ns) na estrutura, o

que por sua vez proporciona um grande número de subestados intimamente relacionados (Figura 8B). Dessa forma, a DM de proteínas compreende então a esse processo de visitação aos diferentes estados de uma proteína em seu perfil energético, levando em conta a quarta dimensão, o tempo [83, 84].

Figura 8 – Perfil energético de uma proteína (funil rugoso)



Fonte: Adaptado de <https://en.wikipedia.org/wiki/File:Folding_funnel_schematic.svg>

Nas Seções 2.5 e 2.6 foram discutidos aspectos importantes do reconhecimento molecular receptor–ligante e a abordagem teórica do *docking* para mensurar a afinidade. Para este último, um dos problemas relacionados a acurácia na predição da afinidade é a não consideração da flexibilidade do receptor em prol do ganho computacional para aplicação em larga escala. Nesse aspecto, métodos de simulação computacional baseados em DM podem ser utilizados para refinar o processo de busca e avaliação dos algoritmos de *docking*, como também para explorar os diferentes subestados de um sistema, de forma a avaliar a estabilidade do complexo receptor–ligante, ou até mesmo realizar cálculos teóricos mais robustos da energia livre de ligação (*rescore*).

Portanto, tendo em vista a importância da dinâmica de proteínas para uma maior reliabilidade da interação receptor–ligante, na próxima seção será abordado alguns aspectos teóricos da simulação de DM.

2.8.1 Simulação de Dinâmica Molecular

Esta técnica de simulação utiliza dos pressupostos da mecânica clássica para determinar o movimento dos átomos em função do tempo. Para isso é necessário ter uma configuração inicial do sistema de interesse, o que por sua vez corresponde as coordenadas

espaciais dos átomos[5]. Além disso, considera-se também a disposição desses átomos em uma "caixa de simulação", que por sua vez pode conter moléculas de um solvente específico, tal como a água, por exemplo. Tendo isso, a propagação das posições atômicas ao longo do tempo depende agora do conhecimento das velocidades iniciais e dos potenciais de interação desses átomos. Quanto a velocidade, esta pode ser atribuída por meio da distribuição de Maxwell–Boltzmann em que a energia cinética é determinada pela temperatura,

$$\frac{3}{2}Nk_BT = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \tag{2.2}$$

onde N é o número de átomos, k_B é a constante de Boltzmann, T é a temperatura e as variáveis m_i e v_i compreendem as massas e a velocidade escalar de cada átomo i, respectivamente [63, 84].

Quanto ao potencial de interação, este é responsável por descrever as forças atuantes em cada átomo e consequentemente, o comportamento molecular ao longo do tempo. Este potencial seria melhor descrito pela resolução da equação de Schrödinger dependente do tempo, ou melhor dizendo, pela mecânica quântica. No entanto, equacionar sistemas biológicos na luz da mecânica quântica ainda demanda um algo custo computacional, logo, aproximações são admitidas de forma a simplificar o modelo físico. Aqui, potenciais efetivos clássicos regem as interações intra- e intermoleculares entre átomos individuais do sistema (*full–atom*) ou entre grupos específicos de átomos (*united–atoms* ou *coarse-grained*). Nesse aspecto, o conjunto de funções matemáticas e parâmetros que reproduzem esses potenciais efetivos responsáveis por descrever o comportamento molecular são denominados de "campos de força". Atualmente existem campos de força para todos os tipos de macromoléculas biológicas, e no caso das proteínas (objeto de estudo desta tese), os mais utilizados são: AMBER, CHARMM, GROMOS e o OPLS [63, 84].

Dada as condições iniciais – posição dos átomos e suas respectivas velocidades e potenciais de interação –, a evolução temporal ocorre por meio da mensuração da força atuante sobre cada átomo, de acordo com a equação abaixo,

$$F_i(t) = -\nabla V^{total} \tag{2.3}$$

sendo,

$$V^{total} = V^{intra} + V^{inter} \tag{2.4}$$

onde,

$$V^{intra} = \underbrace{\frac{1}{2}\sum K_r(r - r_0)^2}_{ligação} + \underbrace{\frac{1}{2}\sum K_\theta(\theta - \theta_0)^2}_{angular} + \underbrace{\frac{1}{2}\sum K_\varphi[1 + cos(n\varphi - \delta)]}_{diedro} \tag{2.5}$$

[5] Assim como no *docking* receptor–ligante, a estrutura da proteína pode ser obtida por técnicas experimentais como a difração de raios–X e o RMN, como também por técnicas de modelagem comparativa. Para as que são resolvidas experimentalmente, o *Protein Data Bank* mais uma vez corresponde ao principal acervo dessa informação.

e

$$V^{inter} = \underbrace{\sum_{i,j}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]}_{Lennard\text{-}Jones} + \underbrace{\sum_{i,j}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}}_{Coulomb} \tag{2.6}$$

Na Equação 2.5 é apresentado os potenciais que descrevem interações entre átomos ligados, sendo o primeiro e segundo termo descritores do potencial harmônico de estiramento entre pares de átomos e deformação angular, respectivamente, enquanto o terceiro descreve o potencial de torção diedral. Já na Equação 2.6 é descrito os potenciais de interação entre átomos não ligados, sendo o primeiro termo um descritor de interações repulsivas e dispersivas enquanto o segundo descreve as interações eletrostáticas. Parâmetros importantes como a distância de equilíbrio entre dois átomos, carga parcial ($q_{i,j}$) ou a força elástica de estiramento (K_r), são calibrados de forma a reproduzir o comportamento do sistema em estudo. Nesse aspecto, cada campo de força utiliza uma metodologia específica para calibração desses parâmetros, mas no geral eles estão bem estabelecidos na literatura para proteínas e outras biomoléculas. [63, 84].

De acordo com a segunda lei de Newton, é possível calcular as acelerações $\vec{a}$ de cada átomo por meio dos potenciais de interação (Eq. 2.4) e da massa (m), utilizando a Equação,

$$-\nabla\vec{V}^{total}(t) = m\vec{a}_i(t) \tag{2.7}$$

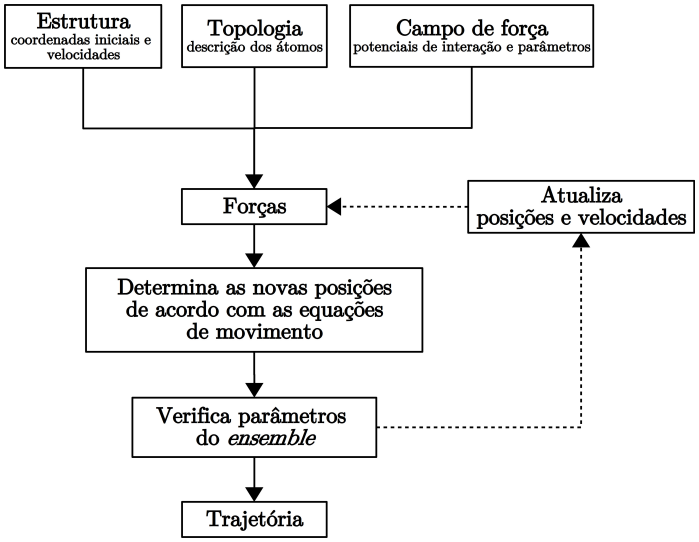
Por meio das configurações iniciais, ou seja, coordenadas, velocidades e acelerações para cada átomo do sistema, o próximo passo consiste em determinar as novas posições em um instante de tempo seguinte ($t + \delta t$). Para isso utiliza-se algoritmos específicos, responsáveis por integrar as equações de movimento, tal como o *Velocity-Verlet* ou *Leap-frog*, e uma vez obtida as novas coordenadas, velocidades e acelerações, o ciclo é reiniciado. No mais, a repetição deste procedimento compõe a trajetória dos átomos ao longo do tempo [63, 84].

Após essa descrição introdutória sobre a simulação de DM, outros aspectos importantes precisam ser considerados antes da execução, são eles: (*i*) *ensemble*; (*ii*) condições periódicas de contorno; e (*iii*) raio de corte para interações de longo alcance [63, 84].

Sobre o item *i*, uma vez que a DM consiste em realizar uma ampla amostragem dos microestados do sistema, conjuntos específicos de parâmetros, tal como temperatura, pressão, energia, solvente, definem diferentes configurações do sistema. No caso de sistemas biológicos, os processos químicos são estudados experimentalmente sob pressão (P) e temperatura (T) constantes, portanto, é importante que as condições de simulação levem em conta este aspecto. Para isso, é possível acoplar um barostato e termostato às equações de movimento, de forma a incluir a contribuição destes parâmetros para evolução temporal do sistema. Além disso, para fins estatísticos, é necessário que o número de

moléculas (N) seja constante também. Na Figura 9 é apresentado um esquema que ilustra os passos de uma simulação de dinâmica molecular [63, 84].

Figura 9 – Passos de uma simulação de DM.



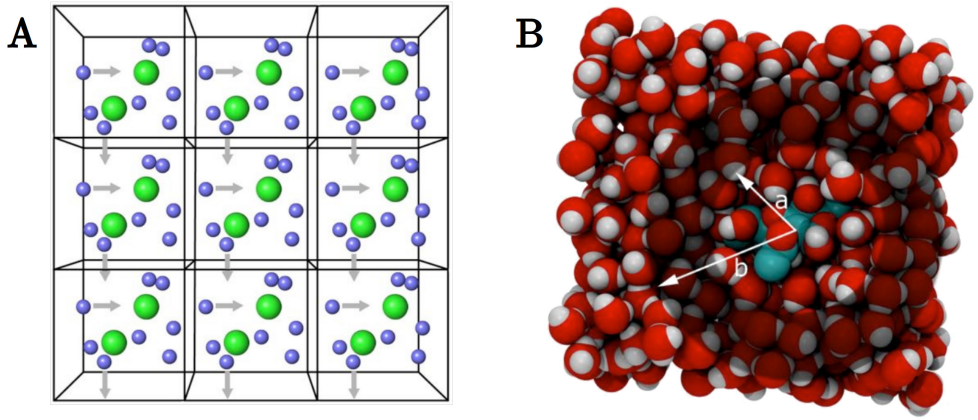
Fonte: O autor.

Quanto ao item *ii*, no início desta seção foi dito que para simular uma macromolécula por DM é necessário incluí-la em uma "caixa de simulação". Partindo dessa informação, à depender dos recursos computacionais do laboratório, o total de átomos simulados em uma DM pode ser bem menor do que aquele realizado com amostras em laboratório. Além disso, a caixa de simulação impõe uma distinção entre o número médio de interações das moléculas próximas aos limites da caixa e aquelas localizadas na região mais interna, podendo acarretar em mensurações inapropriadas, que não se aproximam de condições experimentais. Nesse aspecto, utiliza-se das condições periódicas de contorno para eliminar esse efeito de superfície. Aqui, a caixa de simulação é envolta por cópias idênticas em todas as direções, infinitamente, formando um sistema periódico. Desta forma, a face externa da caixa que antes estava em contato com o vácuo, agora faz contato com a face oposta. Em outras palavras, caso uma molécula de água saia da caixa central, por exemplo, uma de suas imagens entrará pela face oposta, assim, além de manter o número de moléculas constante, o sistema experiencia condições aproximadas de um ensaio em bancada (Figura 10A) [63, 84].

Já o item *iii* diz respeito ao cálculo das interações eletrostáticas e de van der Waals pra todos os átomos do sistema, o que por sua vez compreende a uma das etapas mais custosas da simulação de DM. De acordo com os potenciais que descrevem estas interações, uma vez que a distância entre dois átomos aumenta, o potencial de interação entre eles diminui. Dessa forma, é possível economizar no custo computacional atribuindo um raio de corte para uma determinada distância, assim, interações com distâncias acima deste valor não são calculadas, ou consideradas em grupos de carga, num campo de reação ou

em sistemas periódicos como PME (Figura 10B) [63, 84].

Figura 10 – Condições específicas de uma simulação DM para eliminar o efeito de borda e otimizar o tempo de cálculo.



(**A**) Condições periódicas de contorno; (**B**) Raio de corte. Fonte: Verli[73] (2014, p. 180, 182)

Uma vez finalizada a simulação de DM, dispomos de um conjunto de coordenadas que representam a evolução temporal do sistema (trajetória). Com essa informação é possível então realizar uma série de análises, tal como médias e frequências de alguma propriedade estrutural ao longo da trajetória, sendo as principais: (*i*) raiz do deslocamento médio quadrático (RMSD, do inglês *Root Mean Square Deviation*); (*ii*) conteúdo de estrutura secundária; (*iii*) solvatação, entre outras. Além disso, parâmetros geométricos e energéticos também podem ser extraídos, tal como distâncias entre átomos, ângulos e até mesmo a energia de interação entre grupos de átomos, tal como interações atrativas e repulsivas entre um fármaco e o sítio de ligação no bioreceptor, por exemplo [63, 84].

De modo geral, as possibilidades de análise são inúmeras, e o emprego de algumas está intrinsecamente relacionado a natureza do problema em estudo, no entanto, é seguro afirmar que uma análise consenso para qualquer estudo de DM é o RMSD. Essa análise indica o quanto uma estrutura se modifica ao longo de uma simulação em relação à estrutura de partida (cristalográfica, por exemplo). No caso de simulações DM de complexos receptor–ligante, pelo RMSD é possível inferir hipóteses quanto afinidade de um ligante por meio de sua estabilidade no sítio receptor. Por esta perspectiva, ligantes inativos comumente apresentam um perfil de RMSD ruidoso (instável), diferente dos ativos [85].

Por outro lado, para estimativas de propriedades macroscópicas, tal como o ΔG_{lig}, seriam necessárias trajetórias envolvendo escalas de tempo de μs a s, o que é equivalente a obter todos os microestados da superfície de energia (Figura 8). Porém, este tipo de procedimento é incompatível com a capacidade computacional da maioria dos grupos de pesquisa. Neste caso, é necessário o uso de outros métodos de simulação – também baseados em DM – para aumentar a amostragem ou métodos empíricos como MM/PBSA

ou MM/GBSA [63, 84]. Na próxima seção será discutido alguns aspectos sobre estimativas do ΔG utilizando um desses métodos.

2.9 Dinâmica Molecular de Caminho Induzido

A variação de energia livre (ΔG) entre dois estados de um sistema é igual ao trabalho (W) para transformar um estado (A) em outro (B), desde que este seja feito por meio de um processo reversível sob condições de temperatura e pressão constantes, como mostra a Equação 2.8,

$$\Delta G = W_{rev} = \int_{\xi_0}^{\xi_1} \vec{F} \cdot d\vec{\xi} \tag{2.8}$$

onde $\vec{F}$ é a força e $\vec{\xi}$ é a coordenada de reação (ou o "caminho" que leva de um estado a outro). Nestes termos é possível adaptar a simulação de DM para amostrar $\vec{F}$ ao longo da coordenada de reação $\vec{\xi}$, que é o caso da simulação de SMD [86, 87].

Nessa técnica é aplicado uma força em um grupo de átomos do sistema de modo a acelerar um determinado processo, tal como o desencaixe de um ligante do seu sítio de ligação. A força externa ($\vec{F}_{ext}$) que esses átomos experienciam tem a forma da Equação 2.9,

$$\vec{F}_{ext} = k(\vec{v} \cdot t - \vec{\Delta\xi}) \tag{2.9}$$

onde k é uma constante que representa a elasticidade de uma mola, $\vec{v}$ a velocidade de aplicação da força e $\vec{\Delta\xi}$ é a variação de deslocamento do grupo de átomos em relação à posição inicial. Os valores de k e $\vec{v}$ por serem constantes podem ser calibrados, como também a direção ao qual a força é aplicada [84, 86, 87, 88].

Na Equação 2.9, $\vec{F}_{ext}$ se torna interessante para descrição de uma transição de estado por estar a cada instante de tempo sendo modulado pela resistência que o meio exerce sobre o deslocamento dos átomos. Por exemplo, no processo de desacoplamento de um ligante do sítio receptor para o solvente, no instante inicial da simulação SMD, o ligante encontra-se estável devido às fortes interações intermoleculares com os resíduos circundantes, logo, a variação de deslocamento $\vec{\Delta\xi}$ é pequena, dessa forma, o produto de $\vec{v}t$ garante o aumento de $\vec{F}_{ext}$ até que essas resistências sejam vencidas. Feito isso, $\vec{\Delta\xi}$ aumenta e como consequência, $\vec{F}_{ext}$ diminui. Por conseguinte, o trabalho (W) acumulado pelo deslocamento do ligante ao longo da coordenada de reação pode ser calculado pela Equação 2.8 [84, 86, 87, 88].

Para um conjunto de trabalhos (W) obtidos por meio de simulações SMD utilizando configurações iniciais distintas do sistema, a relação com o ΔG pode então ser feita pela igualdade de Jarzynski, de acordo com a Equação 2.10,

$$e^{-\beta\Delta G} = \left\langle e^{-\beta W} \right\rangle_{\infty} \tag{2.10}$$

Em linhas gerais, a Equação 2.10 relaciona o ΔG com a média exponencial de um conjunto de trabalhos (W), realizados para transformar um estado do sistema em outro.

Além disso, a acurácia do método está diretamente relacionada com o total de replicatas (o que justifica o ∞ no termo a direita da igualdade) e a proximidade às condições de reversibilidade. Para este último, uma vez que a transformação do estado A em B ocorre rapidamente, ou seja, em condições fora do equilíbrio (ou irreversíveis), a dispersão dos valores de W será maior e consequentemente, maior será o erro de $\langle W \rangle$. Por outro lado, numa transformação infinitamente lenta, ou seja, em condições de reversibilidade, a dispersão dos dados numéricos de W será bem menor e o ΔG será igual a $\langle W \rangle$, o que seria equivalente ao próprio W_{rev} da Equação 2.8 [89].

2.10 Métodos de Química Quântica

Até aqui, a descrição teórica para compreender e prever as interações não-covalentes em sistemas químicos foram baseadas em potenciais efetivos clássicos, desconsiderando os efeitos de reatividade, e portanto, as propriedades eletrônicas do sistema. Esta tarefa agora é dada aos métodos químico-quânticos, que por sua vez buscam descrever o sistema molecular à luz da mecânica quântica, mais especificamente pela solução da equação de Schrödinger [90]. No entanto, a solução desta equação não é possível para sistemas químicos complexos, tal como um ligante/peptídeo complexado a uma proteína, devido a impossibilidade de solução analítica. Esse "gargalo" analítico motivou grupos de pesquisa a formular e implementar procedimentos numéricos aproximados, são eles: (*i*) métodos *ab initio*; (*ii*) método da teoria do funcional da densidade (DFT, do inglês *Density Functional Theory*) [91]; e os (*iii*) métodos semi-empíricos (SE) [92].

Dentre os métodos supracitados, os do item *iii* tornam-se atrativos para análise de sistemas biológicos devido ao baixo custo computacional associado. Esta classe de métodos inseriu mudanças no procedimento de cálculo para solucionar a equação de Hartree-Fock, tal como a inclusão de parâmetros e observações experimentais, ou por manipulações para definir certas integrais igual a zero [92]. No mais, os métodos semi-empíricos modernos diferem entre si pelas aproximações adicionais que são feitas ao avaliar integrais de um e dois elétrons e pela filosofia de parametrização [92].

Dentre os diferentes métodos SE, são eles: RM1 [93], PM6 [94], PM6-DH+ [95], PM6-D3H4 [96] e PM7 [97]; o PM7 aparece como uma abordagem promissora para derivar os valores de entalpia em sistemas biológicos. Em comparação com outros métodos, principalmente variações do PM6, o PM7 obteve boa concordância com resultados experimentais obtidos em ensaios biológicos, com coeficientes de determinação de até 0.98 [98, 99, 100, 101].

3 Objetivos

3.1 Objetivo Geral

Executar uma abordagem computacional capaz de minimizar as limitações atuais para avaliação *in silico* de novos inibidores para a subunidade catalítica da Ricina, a RTA.

3.2 Objetivos Específicos

a) Estabelecer *in silico*, uma biblioteca de ligantes candidatos por meio da base de dados ZINC15;

b) Determinar *in silico*, a capacidade preditiva de um modelo farmacofórico em discernir ligantes de RTA ativos de inativos;

c) Determinar *in silico*, a capacidade preditiva de um protocolo de *docking* receptor–ligante em discernir ligantes de RTA ativos de inativos;

d) Determinar *in silico*, por meio dos métodos citados nos itens b) e c), a afinidade relativa em relação ao sítio ativo da RTA da biblioteca de ligantes obtida no item a);

e) Determinar *in silico*, por meio do método de Dinâmica Molecular de Caminho Induzido, o *rescore* dos ligantes melhor classificados no item c);

f) Estabelecer um *ranking* de ligantes candidatos.

4 Metodologia

Este capítulo tem como objetivo apresentar os detalhes metodológicos de preparo e execução das abordagens teóricas propostas no Capítulo 3. Inicialmente serão apresentados os procedimentos de preparo das bibliotecas de ligantes, tanto de triagem como de validação (Seções 4.1 e 4.2), para em seguida, adentrar na descrição de execução dos métodos. Nesta segunda parte, com exceção dos métodos de simulação computacional, será descrito como foi realizada a validação de cada método, e em seguida, a sua execução na biblioteca de triagem. Quanto aos aspectos de validação dos métodos de simulação computacional, especificamente a Dinâmica Molecular de Caminho Induzido, no Capítulo 1 foi feito um breve resumo, no entanto, para aprofundamentos, recomenda-se a leitura da referência [8].

4.1 Preparo da Biblioteca de Triagem

A base de dados ZINC15 [69] contém mais de 1 bilhão de estruturas químicas prontas para uso em estudos de TV. No entanto, para esta tese foi utilizado um subconjunto específico dessa base de dados. Somado a isso, o ZINC15 contém um sistema de busca intitulado como *tranches*, o qual possibilita especificar critérios sobre as moléculas que se deseja incluir na triagem, tal como: (*i*) propriedades físico-químicas; (*ii*) toxicidade; e (*iii*) disponibilidade de compra. Portanto, para montagem da biblioteca de triagem foi utilizado esse sistema de busca de acordo com os seguintes critérios:

a) **Representação**: 3D;

b) **Reatividade**: padrão;

c) **Carga**: neutro;

d) **Comprabilidade**: em estoque;

e) **pH**: tautômeros predominantes em *pH* fisiológico (7.4);

f) **Subconjunto**: *Lead-like*;

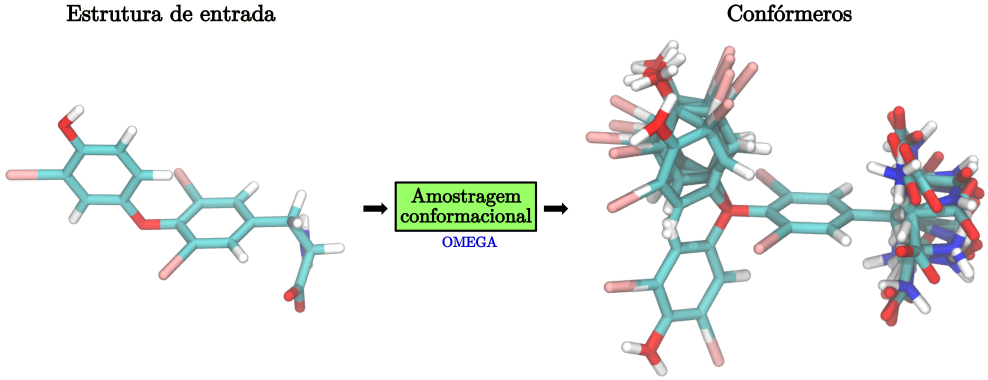
g) **Formato de arquivo**: MOL2

Quanto ao item **Representação**, a opção "3D" possibilita o acesso às coordenadas *xyz* das moléculas, enquanto a **Reatividade** diz respeito a presença de grupos químicos relacionados com toxicidade. Nesse contexto, a opção "padrão" permite recuperar apenas os compostos com baixa incidência desses grupos. O item **Carga** refere-se à carga formal das moléculas. Aqui, a opção "neutro" foi escolhida baseado no pressuposto de que compostos carregados têm baixa permeabilidade à bicamada lipídica, o que corresponde a uma propriedade farmacocinética ruim. Somado a isso, o **pH** tem uma contribuição direta na forma tautomérica das estruturas, portanto, foi priorizado os tautômeros dominantes em *pH* fisiológico. Para o item **Subconjunto**, a opção "*Lead-like*" dispõe de um grupo de moléculas com propriedades físico-químicas preditas de *logP* e peso molecular nos intervalos $[-1, 3.5]$ e $[300, 350]$, respectivamente. De acordo com Oprea et al.[102], este tipo de molécula é interessante para estudos em química medicinal por apresentarem estrutura química simples, alto grau de patenteabilidade e bons aspectos de Absorção, Distribuição, Metabolismo e Excreção (ADME). Quanto ao item **Formato de arquivo**, o "MOL2" foi escolhido por ser o formato de entrada para os programas utilizados nas etapas seguintes. Por meio destes parâmetros foi possível resgatar um total de 2.261.576 moléculas.

Em seguida foi realizado um passo de filtragem por moléculas indesejáveis dentro do conjunto obtido no passo anterior, utilizando os requisitos propostos por Oprea et al. [102], que por sua vez estão inclusos no programa FILTER (versão 3.0.0.1; OpenEye Scientific®) por meio do parâmetro `-filter lead`. Ao término desta etapa foram retirados ∼750.000 compostos, enquanto as remanescentes (∼1.500.000) foram encaminhadas ao protocolo de

amostragem conformacional usando o programa OMEGA [103] (versão 3.0.0.1; OpenEye Scientific®). De modo geral, esse último passo teve por objetivo encontrar ao máximo 50 confórmeros de baixa energia para cada uma das moléculas (Figura 11). Fora este requisito, os demais parâmetros de configuração foram mantidos em modo padrão. A motivação para obtenção desses confórmeros emerge do fato de que uma das etapas da triagem é baseada na comparação de farmacóforos (Seção 2.7), sendo a amostragem conformacional um passo importante para uma melhor reliabilidade da análise.

Figura 11 – Representação 3D de confórmeros obtidos com o programa OMEGA.



O programa OMEGA recebe como entrada a conformação tridimensional da molécula de interesse, e após determinação dos parâmetros de amostragem, retorna ao usuário os confórmeros possíveis, que por sua vez obedecem um critério de mínimo de energia. Fonte: O autor. Ilustração feita no programa VMD.

A biblioteca de ligantes obtida corresponde a um grupo seleto de moléculas com propriedades físico-químicas definidas e espaço conformacional "levemente" amostrado. Mais a frente, estes confórmeros serão analisados sequencialmente quanto ao seu potencial de complementariedade química ao sítio ativo da RTA. Na próxima seção será descrito os procedimentos de preparo da biblioteca de ligantes utilizada para validação dos métodos de triagem.

4.2 Preparo da Biblioteca de Validação

De forma a garantir o mínimo de resultados falso-positivos, antes de dar início à cascata de filtros na biblioteca de triagem (Seção 4.1), foi realizado uma série de validações usando um segundo conjunto de moléculas, obtida de acordo com os passos descritos à seguir:

a) A base de dados do ChEMBL (www.ebi.ac.uk/chembl) foi acessada a fim de resgatar informações sobre dados experimentais e estrutura química de ligantes ativos e inativos para a subunidade RTA da ricina. Utilizando a palavra-chave "*ricin*" como entrada na área de busca relacionada à "alvos terapêuticos", a base de dados retornou 68 moléculas (atualizado em 29/09/2018), sendo resgatado os dados experimentais de IC_{50} e estrutura química no formato SMILES apenas daqueles com atividade anti-RTA mensurada pelo mesmo tipo de ensaio *in vitro* – como proposto por Cerqueira et al. [9] –, totalizando 58. Em seguida, dessas 58 estruturas, foi constatado que 27 apresentam valores de $IC_{50} \leq 100\mu M$, e foram considerados por este estudo como ativas, e as demais como inativas ($IC_{50} > 100\mu M$).

b) Foi realizado uma busca no serviço web DUD-E [104] (www.dude.docking.org), tendo como "entrada" a lista de moléculas "ativas" obtida na etapa anterior. O DUD-E retorna ao usuário um conjunto de moléculas no mesmo formato de entrada, SMILES, que por sua vez não contém a informação 3D da molécula, necessária para a próxima etapa. Portanto, foi preciso fazer a conversão SMILES $\rightarrow$ MOL2 utilizando o programa OpenBabel [105] (versão 3.4.1).

   Em linhas gerais, este passo teve por objetivo resgatar moléculas com aspectos físico-químicos similares aos ligantes ativos para a RTA, tal como: (*i*) ligações torsionáveis; (*ii*) coeficiente de partição ($logP$); (*iii*) peso molecular; (*iv*) aceptores de ligação de hidrogênio; e (*v*) doadores de ligação de hidrogênio. Ao todo foram obtidos 2.600 estruturas, consideradas nesta nova biblioteca como inativos putativos (ou *decoys*).

c) Ligantes ativos, inativos e *decoys* foram embaralhados em um único arquivo múltiplo no formato MOL2, compondo assim, a biblioteca de validação.

d) Similar à etapa 3 da Seção 4.1, foi realizado também uma amostragem conformacional para os compostos desta biblioteca, utilizando o programa OMEGA [103] (versão 3.0.0.1; OpenEye Scientific®).

Esta segunda biblioteca também contém um conjunto bem definido de moléculas. Destas, algumas dezenas apresentam atividade antagonista para a RTA (ligantes ativos); outras dezenas não apresentam atividade significativa (inativos), ou seja, é necessário

a administração de concentrações elevadas para atingir a atividade desejada; e outras centenas não apresentam qualquer resquício de atividade, e são estruturalmente dissimilares às ativas e inativas (*decoys*). Em suma, a presença de ligantes inativos e *decoys* funcionam como um "ruído" para as análises posteriores de validação, que por sua vez devem ser capazes de discretizar a informação química dos ligantes ativos a partir de um *score*, ranqueando-os nas primeiras posições.

4.3 Triagem Farmacofórica Baseada na Estrutura

O ligante 19M (PDB ID 4HUP) corresponde ao composto que apresentou melhor atividade antagonista para a RTA e tem estrutura cristalográfica do complexo resolvida e publicada no PDB (rever Tabela 3). Dessa forma, como descrito na Seção 2.7, a fim de aumentar a chance de encontrar compostos seletivos para o sítio ativo da RTA, a conformação deste ligante da estrutura cristalográfica 4HUP foi utilizada como modelo de consulta para realização da triagem farmacofórica.

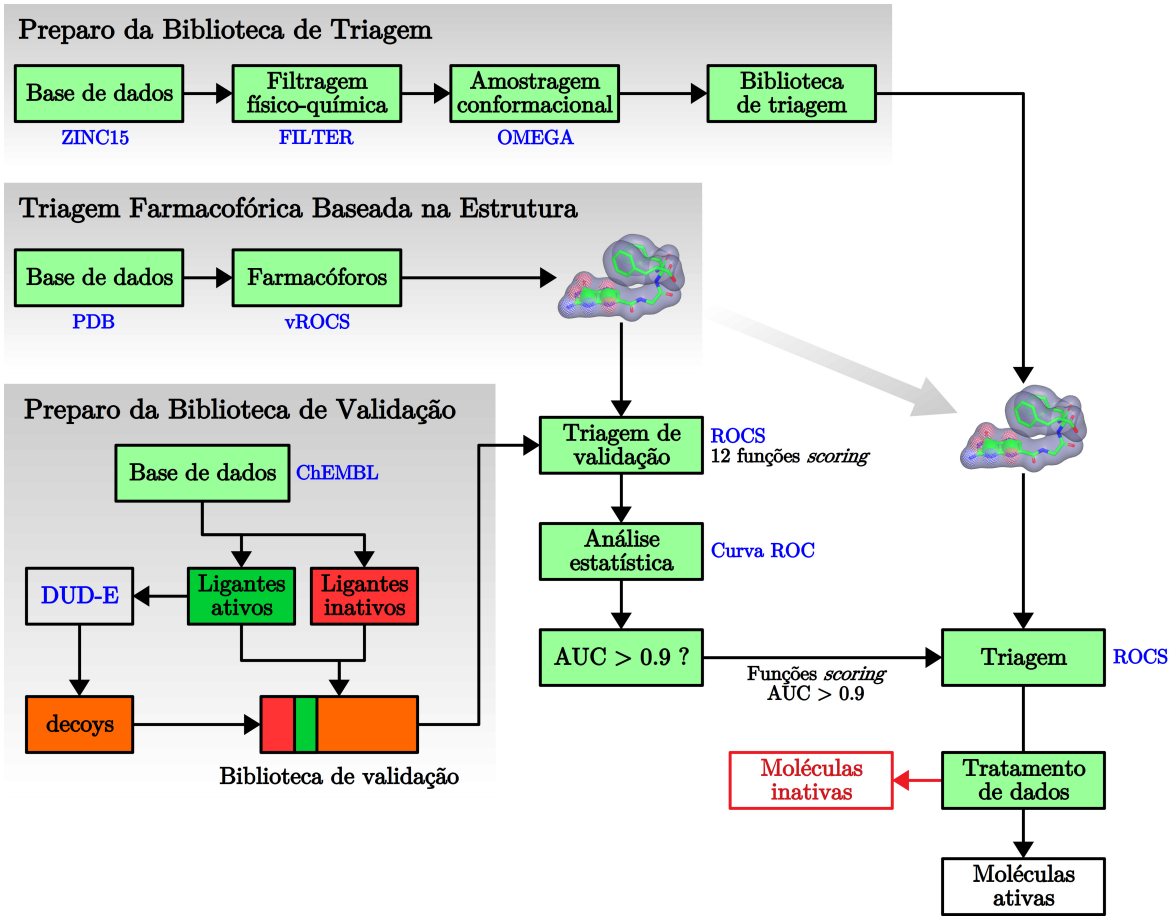
Para montagem do modelo foi utilizado o programa vROCS (versão 3.3.0.3; OpenEye Scientific®) em configuração padrão, com exceção apenas da exclusão de alguns farmacóforos do modelo inicial gerado pelo programa. Em linhas gerais, foi priorizado os farmacofóros relacionados ao grupo pterina e apenas o volume e forma molecular da cadeia lateral do ligante (rever Figura 16). Por conseguinte, a fim de avaliar a eficiência deste modelo de consulta em discernir ligantes ativos de inativos, foi realizado um teste com a biblioteca de validação. Para isso foi utilizado o programa ROCS (versão 3.3.0.3; OpenEye Scientific®), que por sua vez contém 12 funções *score*.

De modo geral, os confórmeros desta biblioteca foram alinhados ao molde de consulta, e para cada alinhamento foi gerado um *score*, este, por sua vez, obedece a um intervalo numérico com limite inferior e superior. Em suma, um *score* próximo ao limite superior corresponde a um alto grau de similaridade com o molde de referência, o que sugere uma atividade biológica similar. Sabendo disso, foi então realizado um procedimento único para cada uma das 12 funções, e em seguida, a análise estatística da curva ROC foi empregada para avaliar a acurácia dos resultados obtidos, tendo como valor de corte, AUC > 0.9. Para este último foi utilizado o serviço web ScreeningExplorer [106] (www.stats.drugdesign.fr) configurado em modo padrão.

As funções de pontuação que atenderam o critério de AUC foram utilizadas na biblioteca de triagem. Ao término de cada procedimento, os dados numéricos foram ordenados e em seguida extraído os 10.000 melhores resultados, e o que fora tido como resultado consenso entre as funções, foi excluído, permanecendo apenas moléculas únicas.

Na Figura 12 é apresentado um resumo das etapas até então descritas, incluindo as Seções 4.1 e 4.2.

Figura 12 – Fluxograma referente as etapas de preparo da biblioteca de ligantes, preparo e validação do modelo de consulta e triagem farmacofórica.



Fonte: O autor.

4.4 Atracamento Molecular

Os ligantes remanescentes ao filtro anterior (Seção 4.3) foram encaminhados ao passo de *docking* receptor–ligante. Aqui, foi realizado também etapas de validação antes da execução na biblioteca de triagem, a fim de avaliar a eficiência da abordagem em predizer poses bioativas, como também de discernir ligantes ativos de inativos. Nas seções à seguir é apresentada uma explicação sucinta dos procedimentos de validação e triagem utilizados nesta abordagem.

4.4.1 Validação do protocolo de atracamento molecular

Para execução de um procedimento de *docking* receptor–ligante é necessário ter a estrutura 3D do receptor e ligante. Para este último temos dois arquivos múltiplos com centenas de estruturas 3D – bibliotecas de triagem e validação, respectivamente –, faltando definir apenas a estrutura do receptor.

Estudos anteriores utilizaram como receptor a RTA da estrutura cristalográfica 1BR5, porém, como apresentado nas Tabelas 2 e 3, o PDB dispõe de outros complexos RTA–Ligante, que por sua vez, apresentam conformações diferentes dos resíduos do sítio ativo. Somado a isso, para esta tese, todos os procedimentos de *docking* foram realizados com o programa GOLD, que contém 4 funções *score*, são elas: (*i*) GoldScore; (*ii*) ChemScore; (*iii*) ChemPL; e (*iv*) ASP. Portanto, qual dessas estruturas do PDB e função *score* seriam ideais para um estudo de triagem virtual? Nesta perspectiva, nenhum trabalho foi realizado. A fim de responder estes questionamentos, foram conduzidas duas etapas de validação:

- Definir um cenário "ótimo" de parâmetros, o que inclui: (*i*) definir a estrutura do receptor, dentre as 14 disponíveis no *Protein Data Bank*; (*ii*) definir o conjunto de parâmetros à ser utilizado no GOLD.
- Avaliar se este "cenário ótimo" é eficiente em discernir ligantes de RTA ativos de inativos.

No entanto, antes de adentrar nos detalhes teórico–práticos destas etapas, é necessário realizar os procedimentos de preparo dos receptores e ligantes que serão utilizados.

4.4.1.1 Preparo dos receptores e ligantes

A etapa de preparação consiste basicamente em duas etapas sequenciais: (*i*) obtenção das estruturas; e (*ii*) inspeção/correção de possíveis problemas como também adaptação dessas estruturas a uma condição de referência, tal como a adição de hidrogênios nos resíduos da proteína considerando o estado de protonação para um pH específico. À seguir é apresentado uma explicação detalhada dos procedimentos realizados nesta etapa.

a) Foram resgatadas do *Protein Data Bank* as seguintes estruturas cristalográficas: 4HV7, 4MX1, 3PX8, 3PX9, 3EJ5, 4HV3, 1IL9, 1BR6, 1IL4, 1BR5 e 1IL3. Cada uma das estruturas supracitadas contém um ligante complexado ao sítio ativo da RTA, são eles: 19J, 1MX, JP2, JP3, EJ5, 19L, MOG, PT1, 9DG, NEO, 7DG;

b) Os receptores citados no item "a" foram inspecionados por *gaps* na estrutura principal assim como ausência de átomos nos resíduos que compõe o sítio ativo. Em seguida, as moléculas de água foram removidas e as estruturas resultantes foram alinhadas em função da estrutura principal da proteína (átomos `CA C N O`), utilizando o programa Multiseq [107] (versão 3.1, incluso como *plug-in* no programa VMD);

c) Para cada uma das estruturas alinhadas da etapa anterior foi extraído a estrutura do ligante para um arquivo individual no formato MOL2. Aqui, definimos um terceiro conjunto de ligantes, utilizados apenas para fins de validação do protocolo de *docking*;

d) As estruturas obtidas na etapa "b" foram encaminhadas ao serviço web PDB2PQR [108] (www.nbcr-222.ucsd.edu) a fim de estimar o estado de protonação dos resíduos de Arginina, Glutamato, Aspartato e Histidina em pH 7.4. Ao término desta rotina, a estrutura resultante contém os resíduos com hidrogênios devidamente adicionados ao estado de protonação indicado. Este procedimento de adição de hidrogênios considerando um pH de referência foi realizado também para os ligantes, porém, foi utilizado o programa OpenBabel [105] (versão 3.4.1);

e) No programa GOLD, o sítio ativo de cada um dos receptores alinhados foi definido pela função "*Point*", o qual permite ao usuário definir a região de busca (*box grid*) à "mão-livre". Desta forma, fora escolhido um ponto – de coordenadas xyz – próxima a região entre Y80 e Y123 (Figura 2). Somado a isso, todos os resíduos no raio de 10 Å deste ponto foram tidos como ativos.

4.4.1.2 1º Validação

Foi realizado um procedimento denominado *cross-docking*, o qual tem por objetivo avaliar a eficiência do algoritmo de *docking* em reproduzir a pose bioativa de ligantes co-cristalizados com o alvo terapêutico em questão. Para isso, utiliza-se: (i) a estrutura cristalográfica alinhada de receptores co-cristalizados com ligantes; e (ii) a conformação bioativa desses ligantes co-cristalizados. Em suma, o procedimento consiste em atracar todos os ligantes para cada um dos receptores, e por seguinte, avaliar se a pose predita é equivalente à pose bioativa por meio da análise de RMSD.

Para execução desse procedimento, os parâmetros do algoritmo genético implementados no GOLD foram mantidos em modo padrão, com exceção de um, o total de execuções

do algoritmo genético. Por padrão, o GOLD realiza 10 execuções, sempre partindo de uma estrutura inicial diferente. Somado a isso, o parâmetro *autoscale* controla o total de operações realizados a cada ciclo, podendo assumir um valor no intervalo]0,5]. Quanto a isso, valores próximos ao limite superior correspondem a uma investigação conformacional exaustiva, enquanto o limite inferior, uma busca não-exaustiva. Executar um procedimento exaustivo ou não está diretamente relacionado com o tempo de processamento para encontrar uma pose ótima do ligante no sítio receptor. Neste aspecto, espera-se que um maior número de operações seja capaz de investigar melhor o problema, no entanto, dependendo da quantidade de ligantes e recursos computacionais disponíveis, optar por este tipo de abordagem pode ser impraticável. Portanto, a relação tempo vs. acurácia pode ser avaliada a fim de definir o melhor cenário para os recursos computacionais disponíveis. Por padrão, o programa GOLD define o valor 0.3 como sendo ideal para estudos de triagem virtual. Dessa forma, foram avaliados dois cenários: (*i*) total de ciclos = 10; e (*ii*) total de ciclos = 30. Ambos os cenários foram aplicados para cada uma das funções *score* do GOLD, são elas: (*i*) ASP; (*ii*) ChemScore; (*iii*) ChemPL; e (*iv*) GoldScore.

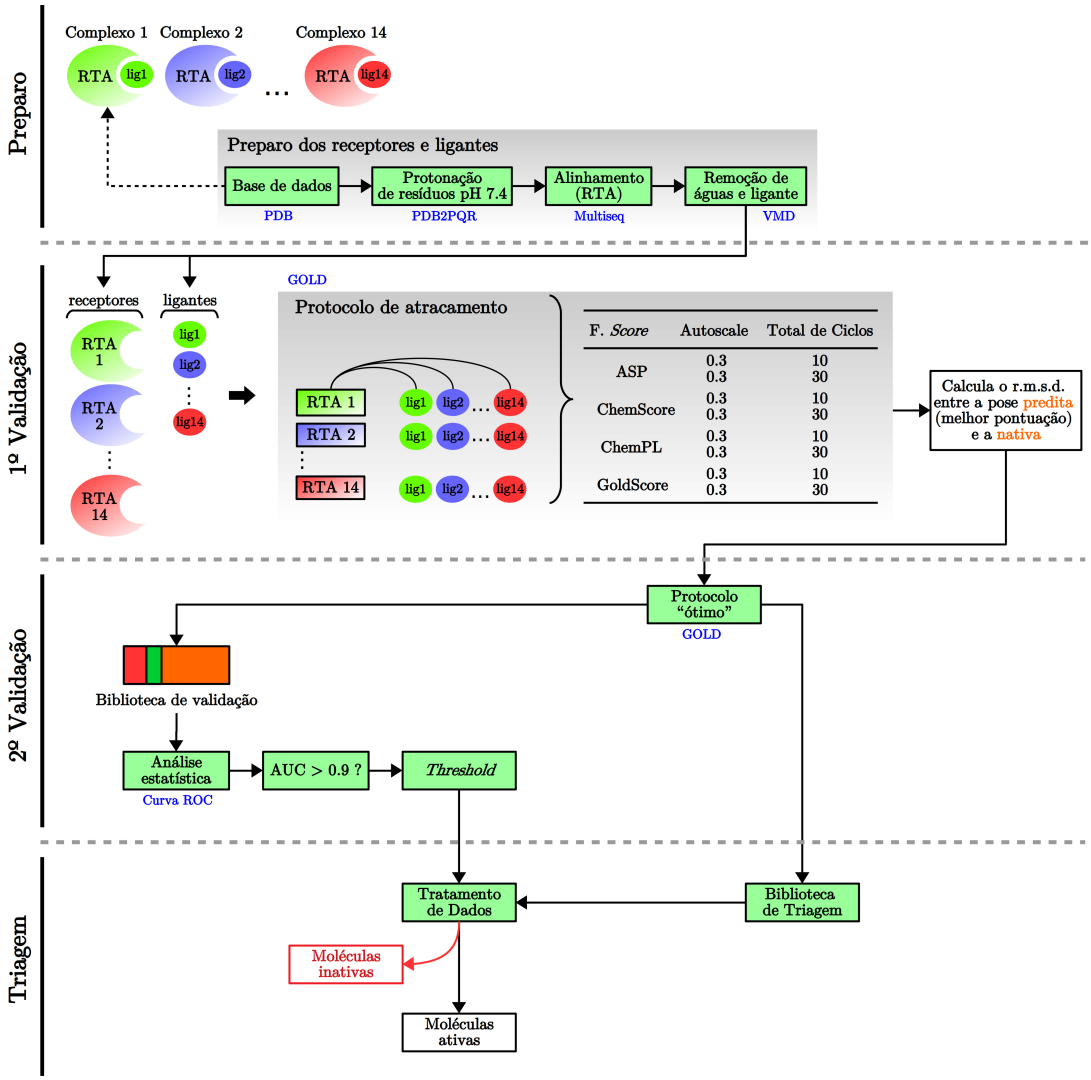
4.4.1.3 2º Validação

Uma vez avaliado o melhor conjunto de parâmetros para o estudo de triagem virtual, o próximo passo consistiu em avaliar a capacidade deste em discernir ligantes de RTA ativos de inativos. Aqui, os ligantes da biblioteca de validação foram atracados no sítio ativo da RTA e o *score* para cada pose predita foi utilizado na análise estatística da curva ROC.

4.4.1.4 Triagem virtual baseada em *docking* receptor–ligante

Seguinte às etapas de validação, o "cenário ótimo" de atracamento foi então aplicado na biblioteca de triagem. Em suma, foi realizado um procedimento recursivo de atracamentos, similar às etapas de validação, e ao término deste, foi utilizado um *threshold*, estimado na análise estatística da curva ROC (Seção 4.4.1.3) para selecionar um conjunto ainda menor de moléculas. Na Figura 13 é apresentado um resumo das etapas descritas nesta seção.

Figura 13 – Fluxograma da triagem virtual.



Fonte: O autor.

4.5 Análises Pós-atracamento

A fim de reduzir ainda mais o total de moléculas a ser avaliado *in vitro*, iremos caminhar em um nível mais descritivo da interação proteína–ligante, de forma a selecionar aquelas com maior complementariedade química ao sítio ativo da RTA.

Como dito no referencial teórico desta tese, ligantes antagonistas e co-cristalizados com a RTA apresentaram ligações de hidrogênio com o átomo de oxigênio da cadeia principal de V81 e G121. Portanto, uma primeira análise consistiu na busca por este padrão de interação intermolecular. Em seguida, os complexos RTA–Ligante remanescentes foram preparados e encaminhados ao protocolo de simulação de DM, a fim de avaliar a estabilidade dessa interação em função do tempo. Posteriormente, os complexos remanescentes foram encaminhados à outros métodos de simulação computacional para determinação quanto a variações da energia livre. Nas seções à seguir será apresentado o delineamento teórico dessas etapas.

4.5.1 Inspeção por ligações de hidrogênio

Um *script* de autoria própria foi utilizado para realizar a inspeção quanto ao total e especificidade das ligações de hidrogênio entre o ligante e a RTA de cada complexo obtido na etapa anterior (Seção 4.4.1.4). Os ligantes que obtiveram um total de ligações de hidrogênio maior ou igual a 2, incluindo interações com o oxigênio da cadeia principal dos resíduos V81 e G121, foram selecionados. Este critério de análise foi conduzido em três etapas sequenciais:

a) O *script* recebeu como entrada todos os complexos proteína–ligante obtidos na etapa anterior, e as poses que atingiram o critério pré-estabelecido foram selecionadas para a etapa seguinte.

b) Foi aplicado um modelo de solvente implícito aos complexos remanescentes e em seguida realizado 1.000 passos de minimização de energia apenas para os átomos do ligante, utilizando o programa NAMD [109 (versão 2.13). Em seguida, as ligações de hidrogênio foram contabilizadas, e os complexos que apresentaram o critério pré-estabelecido foram selecionados para a etapa seguinte.

c) Os complexos remanescentes foram encaminhados ao protocolo de dinâmica molecular. Para isso, proteína e ligante foram preparados com os campos de força FF99SB [110 e GAFF [111], respectivamente. Somado a isso, para o ligante, as cargas parciais dos átomos foram estimadas de acordo com o método AM1-BCC, por meio do programa Antechamber, incluso no pacote Ambertools. Em seguida, os complexos foram inseridos em uma caixa cúbica de 20 Å de borda, contendo águas do tipo TIP3P e íons. Para execução das simulações foi utilizado o programa NAMD, configurado com os seguintes **parâmetros**:

(*i*) condições períodicas de contorno; (*ii*) restrição para ligações covalentes envolvendo átomos de hidrogênio, como também para ângulos `HOH` e distância de ligação `OH`, por meio do algoritmo SHAKE; (*iii*) tempo de integração de 2 fs; (*iv*) raio de corte de 9 Å para interações eletrostáticas.

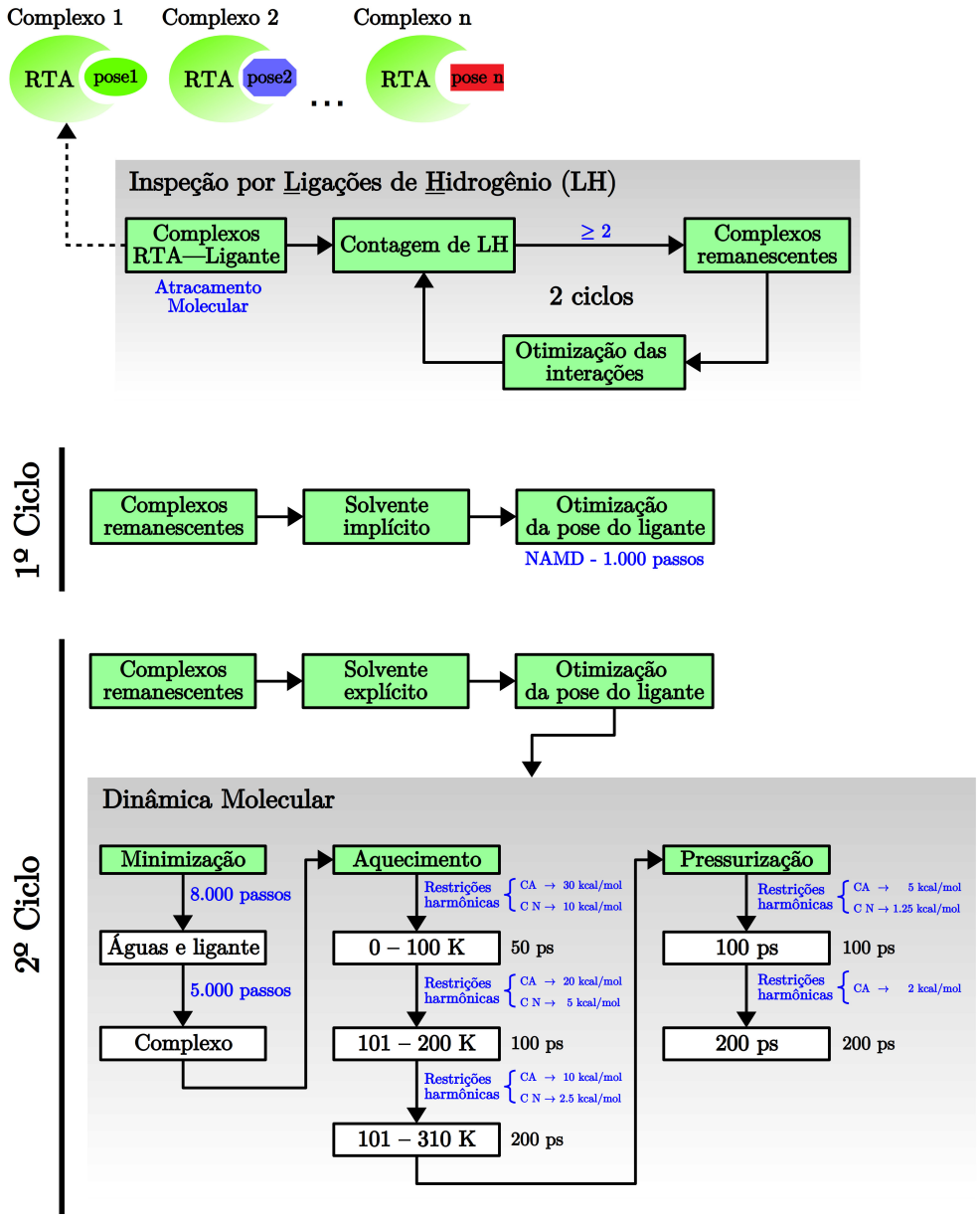
A geometria de cada complexo contém informações quanto a disposição espacial de átomos de água, íons, proteína (RTA) e ligante. No entanto, esta configuração inicial apresenta interações desfavoráveis, principalmente entre o solvente e o complexo RTA–Ligante. Portanto, a fim de otimizar tais interações, foram realizadas 3 etapas sequenciais de simulação no programa NAMD [109]: (*i*) minimização; (*ii*) aquecimento; e (*iii*) pressurização, descritas à seguir.

Minimização → Essa etapa foi conduzida em dois passos sequenciais: (*i*) as coordenadas xyz dos átomos da proteína foram fixados, enquanto ligante, águas e íons passaram por 5.000 passos de minimização de energia; (*ii*) a restrição implementada na proteína foi retirada e todos os átomos do sistema passaram por mais 8.000 passos de minimização.

Aquecimento → A geometria obtida ao término da minimização passou por mais três etapas sequenciais de simulação, incluindo restrições harmônicas nos átomos da estrutura principal da RTA e aumento de temperatura em cada uma das etapas. Este aquecimento ocorreu no intervalo de 0 a 310 K, da seguinte forma: (*i*) os átomos `CA` receberam uma restrição harmônica na ordem de 30 $kcal \cdot mol^{-1}$, enquanto os átomos `C N` receberam uma restrição harmônica de 10 $kcal \cdot mol^{-1}$, ao mesmo tempo em que o sistema era gradualmente aquecido de 0 a 100 K, durante 50 ps; em seguida, (*ii*) esse mesmo conjunto de átomos foi flexibilizado implementando restrições harmônicas de ordem menor, 20 e 5 $kcal \cdot mol^{-1}$, respectivamente, enquanto o sistema era aquecido de 101 a 200 K, durante 100 ps; por fim, (*iii*) a cadeia principal foi flexibilizada ainda mais por meio de restrições harmônicas ainda menores, dessa vez na ordem de 10 e 2.5 $kcal \cdot mol^{-1}$, respectivamente, sendo o sistema aquecido de 201 a 310 K, durante 200 ps.

Pressurização → A geometria obtida ao término do aquecimento foi submetida a mais duas etapas sequenciais, desta vez, incluindo controle da temperatura à 310 K por meio do termostato de Langevin e pressão constante de 1 atm, usando o barostato de Berendsen. Com isso, a simulação prosseguiu da seguinte forma: (*i*) inicialmente foi implementado restrições harmônicas de 5 e 1.25 $kcal \cdot mol^{-1}$ nos átomos da cadeia principal da RTA – similar à etapa de aquecimento –, durante 100 ps; em seguida, (*ii*) as restrições foram retiradas completamente para os átomos `C N` e flexibilizada ainda mais para os átomos `CA`, que receberam apenas uma restrição na ordem de 2 $kcal \cdot mol^{-1}$, durante 200 ps. Ao término desta etapa as ligações de hidrogênio foram contabilizadas, e os complexos que atingiram o critério pré-estabelecido foram selecionadas para a etapa seguinte.

Figura 14 – Fluxograma referente as análises pós-atracamento.



Fonte: O autor.

4.5.2 Dinâmica molecular de caminho induzido

A afinidade dos ligantes selecionados foi estimada utilizando os perfis de força de dissociação obtidos por meio da Dinâmica Molecular de Caminho Induzido, são eles: (*i*) força média $\langle F \rangle$; (*ii*) força de ruptura $\langle F_{max} \rangle$; e (*iii*) média do trabalho de não–equilíbrio $\langle W \rangle$. A obtenção desses perfis ocorreu em 2 etapas sequenciais, descritas a seguir:

a) Os complexos remanescentes foram encaminhados a uma última etapa de simulação, a equilibração. Aqui, os complexos passaram por um tempo maior de simulação; 10 ns, sem restrições na cadeira principal da RTA, e com temperatura e pressão mantidas em torno de $310K$ e $1atm$, respectivamente. Somado a isso, foram mantidos os mesmos parâmetros de simulação apresentados no item c) da Seção 4.5.1.

b) Para a DM de caminho induzido, os átomos fictícios foram inseridos de acordo com as coordenadas cartesianas de um átomo em posição central na estrutura do ligante. Além disso, foram implementados três vieses: (*i*) restrições harmônicas de 1 $kcal \cdot mol^{-1}$ nos átomos de Cα da RTA, de forma a limitar movimentos abruptos na subunidade; e (*ii*) cilindro de restrição harmônica com 6 Å de diâmetro, posicionado ao longo do vetor de saída. Este último viés impede a interação do ligante com resíduos na superfície da RTA, garantindo que o processo de dissociação ocorra apenas ao longo do vetor de saída. Quanto as constantes de força da mola (k) e velocidade (v), foram avaliados dois cenários: (*i*) $k = 2\ kcal \cdot mol^{-1}$ e $v = 0.05$Å$/ps$; (*ii*) $k = 10\ kcal \cdot mol^{-1}$ e $v = 0.05$Å$/ps$. Ademais, os parâmetros de simulação foram os mesmos da etapa "c" da Seção 4.5.1, com exceção apenas das restrições harmônicas citadas anteriormente. No mais, foram realizadas 5 réplicas de 600 ps para cada complexo.

4.5.3 Análises de trajetória

A complementariedade química do ligante em relação ao sítio ativo do receptor define o seu tempo de residência no sítio, que por sua vez está relacionado com o potencial antagonista/agonista. Tal interdependência envolve alguns eventos físico-químicos, tal como o efeito hidrofóbico, complementariedade eletrostática e ligações de hidrogênio. Portanto, a fim de avaliar tais propriedades, as trajetórias de equilibração obtidas na seção anterior foram submetidas às seguintes análises: (*i*) estabilidade conformacional da cadeia principal da RTA e ligante, respectivamente; e (*ii*) perfil de interações intermoleculares do ligante em relação à resíduos do sítio, o que inclui, energia de interação e ligações de hidrogênio. Segue abaixo uma explanação dos procedimentos teóricos utilizados para obtenção desses perfis.

a) **Estabilidade conformacional:** Variações da posição atômica nas trajetórias de DM foram estimadas por meio do RMSD. Para isso, sendo v e w dois microestados de uma proteína ou pequena molécula, o RMSD relativo à cadeia principal pode ser obtido utilizando a Equação 4.1.

$$RMSD_{(v,w)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}, \qquad (4.1)$$

sendo n o total de átomos nos dois microestados.

Além disso, de forma a reduzir o ruído inerente a movimentos translacionais e rotacionais ao longo da simulação, a trajetória foi alinhada em função dos átomos da cadeia principal do primeiro quadro da trajetória. Desta forma, é possível estimar em função do tempo, mudanças conformacionais ao longo da estrutura. Além disso, os dados obtidos podem ser utilizados para realização de outras análises descritivas, tal como média, desvios e frequências.

b) **Energia de interação:** Parte da contribuição energética à afinidade de um ligante para com o sítio de ligação em um receptor é dada por interações intermoleculares, que por sua vez contribuem para estabilidade conformacional do ligante no sítio. Neste aspecto, é possível avaliar ao longo de uma trajetória de DM, a energia de interação de um conjunto de átomos em relação ao ambiente que o circunda. Essa energia de interação E_{total} é dada pela soma das contribuições eletrostáticas E_{elec} de curto e longo alcance (Equação 4.2).

$$E_{elec} = k\frac{q_i q_j}{r_{ij}}\left(1 - \frac{r_{ij}^2}{r_{cut}^2}\right), \qquad (4.2)$$

sendo q a carga formal para dois conjuntos de átomos i e j, respectivamente; r_{ij} a distância entre essas cargas, e r_{cut} a distância mínima (*cutoff distance*).

Por esta análise é possível inferir hipóteses quanto às interações favoráveis ou desfavoráveis à estabilidade do ligante no sítio. Aqui, entende-se por favoráveis, as interações atrativas ou forças restauradoras, e desfavoráveis, àquelas de caráter repulsivo ou desestabilizadoras. Para esta tese, foi avaliado a energia de interação entre os átomos do ligante e os resíduos do sítio ativo que realizaram qualquer tipo de contato num raio de 5 Å com a estrutura do ligante, e ao longo dos últimos 5 ns de trajetória. Somado a isso, assim como na etapa anterior, os dados obtidos foram utilizados para realização de análises descritivas, tal como médias e desvios.

c) **Ligações de hidrogênio:** Esse tipo de força intermolecular ocorre quando um átomo de hidrogênio ligado a um átomo fortemente eletronegativo existe nas proximidades de um outro átomo eletronegativo com um par de elétrons isolado. As ligações de hidrogênio são geralmente mais fortes que as forças dipolo-dipolo e de dispersão comuns, no entanto, são mais fracas que as ligações covalentes e iônicas. Em sistemas proteína–ligante, este tipo de força intermolecular confere grande importância para a estabilidade conformacional e consequente tempo de residência do ligante no sítio.

   Portanto, para os últimos 5 ns da trajetória de DM, foi avaliado a frequência de ligações de hidrogênio entre o ligante e a RTA. Para isso, uma vez que este tipo de interação tem uma dependência direta quanto a distância e ângulo de ligação, foi considerado os valores de 3.5 Å e 30°, respectivamente.

5 Resultados e Discussão

Os resultados alcançados foram suficientes para compor o seguinte artigo científico: "*Discovery of RTA Ricin Subunit Inhibitors: A Computational Study Using PM7 Quantum Chemical Method and Steered Molecular Dynamics*". A seguir será apresentado o texto referente a versão de aceite do manuscrito, publicado no periódico *Journal of Biomolecular Structure & Dynamics*[1].

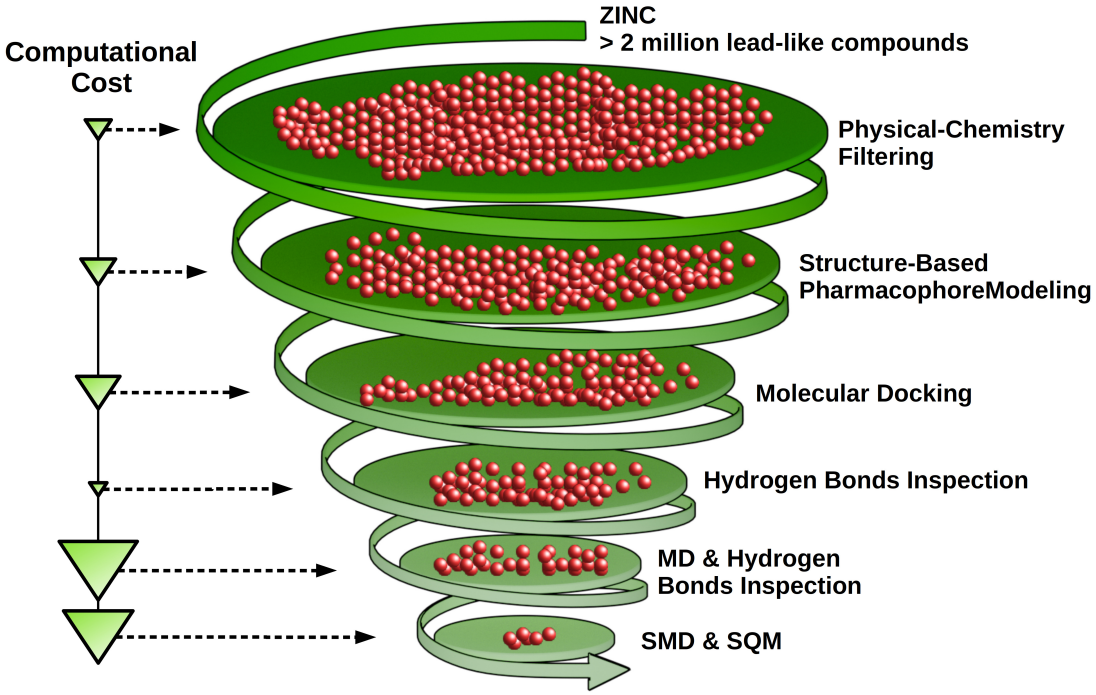
[1] DOI: https://doi.org/10.1080/07391102.2021.1878058

5.1 Discovery of RTA Ricin Subunit Inhibitors: A Computational Study Using PM7 Quantum Chemical Method and Steered Molecular Dynamics

5.1.1 Abstract

Ricin is a potent toxin derived from the castor bean plant and comprises two subunits, RTA and RTB. Because of its cytotoxicity, ricin has alarmed world authorities for its potential use as a chemical weapon. Ricin also affects castor bean agribusiness, given the risk of animal and human poisoning. Over the years, many groups attempted to propose small-molecules that bind to the RTA active site, the catalytic chain. Despite such efforts, there is still no effective countermeasure against ricin poisoning. The computational study carried out in the present work renews the discussion about small-molecules that may inhibit this toxin. Here, a structure-based virtual screening protocol capable of discerning active RTA inhibitors from inactive ones was performed to screen over 2 million compounds from the ZINC database to find novel scaffolds that strongly bind into the active site of the RTA. Besides, a novel score method based on ligand undocking force profiles and semi-empirical quantum chemical calculations provided insights into the rescore of docking poses. Summing up, the filtering steps pointed out seven main compounds, with the SCF00–451 as a promising candidate to inhibit the killing activity of such potent phytotoxin.

5.1.2 Graphical Abstract



5.1.3 Introduction

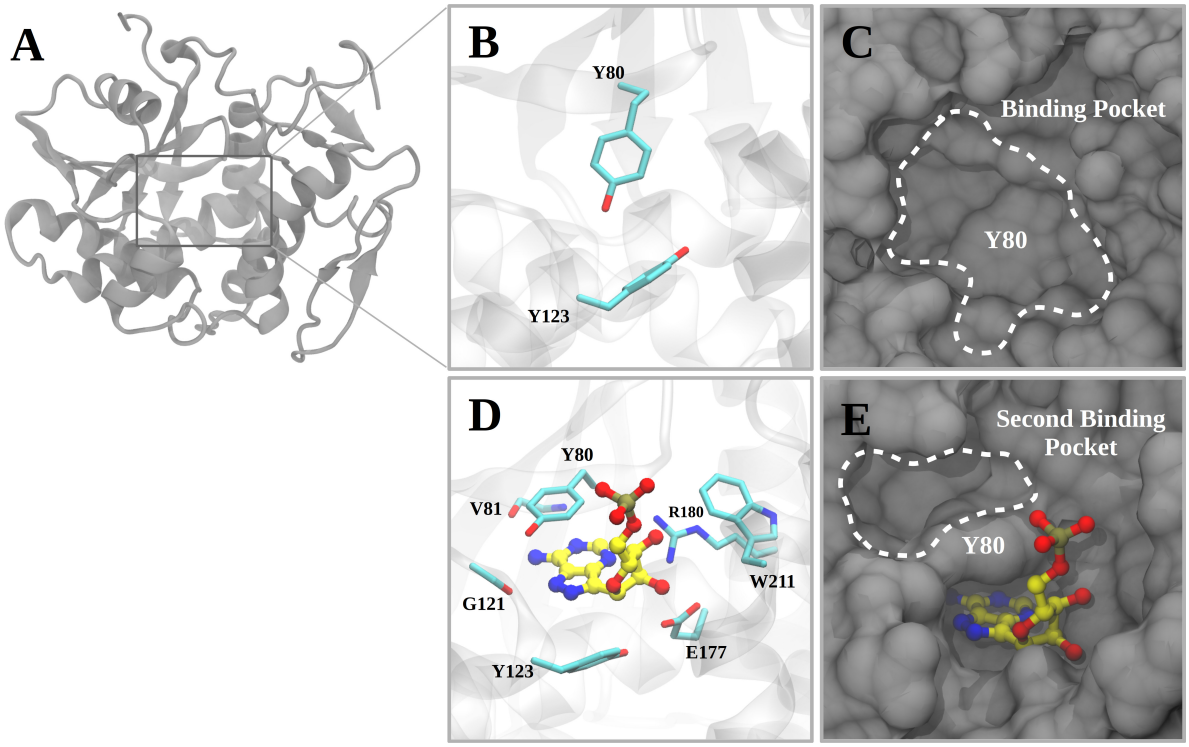
Ricin is a highly potent toxic protein derived from the castor bean plant (*Ricinus communis*) and consists of two sub-units, Ricin Toxin A (RTA) and Ricin Toxin B (RTB) that bind covalently to each other by one disulfide bond. Ricin recognizes eukaryotic cells via RTB lectin, which in turn binds to galactose residues on the cell surface and then traffics it into the endoplasmic reticulum, where these two subunits are separated by disulfide bond reduction. Then, RTA translocates into the cytoplasm, where a fraction of it escapes from degradation by proteassomes, and depurinates the second adenine residue (A^{4328}) located in sarcin–ricin RNA loop (SRL) in the 28S eukaryotic ribosomal RNA. This depurination is irreversible and interrupts the protein synthesis, leading the cell to death.[112] In humans, ricin has a fatal oral dose ranging between 1–20 mg/kg, depending on the mode of administration and ranks it among the most toxic known substances.[113]

Additionally, ricin has worried world authorities due to its potential use as a chemical weapon. Contributes to this concern its high availability and ease of preparation.[114] Ricin also impacts castor bean agribusiness. By-products from castor oil production can contain lethal amounts of ricin. In addition to the risk of animal and human poisoning, ricin has made these by-products difficult to reuse and recycle.[115] So, there is considerable interest in identifying countermeasures to act as antidotes against ricin poisoning.

Along the years, several X-ray structures of the ricin in apo form as well as in complex with its native substrate has been solved.[116, 117, 118, 119] In this sense, the key structural information about the RTA chain are: (*i*) active site is composed of Y80, Y123, E177, R180, and W211 (Figure 15); (*ii*) in the presence of substrate, the side-chain of Y80 adopts an "open" conformation, which in turn forms π–stacking interactions with Y123 as well as hydrogen bonds with V81, G121, R180, E177, and E208.[120]; (*iii*) open conformation exposes a second binding pocket, which is known to accommodate guanine residue of the GAGA ribosomal target sequence[121] (Figure 15D,E); (*iv*) in the absence of substrate, the side chain of Y80 does not rotate and adopts a "closed" conformation (Figure 15B,C).

Efforts to inhibit ricin have emerged from many paths, and one of them comprises to the virtual screening (VS) of small-molecules capable to bind strongly to the RTA active site and retard the depurination event.[122, 123] This process commonly occurs combining a molecular docking procedure to screening large ligand databases targeting to RTA active site in its "open form" (Figure 15D,E), followed by *in vitro* validation techniques, such as the luciferase luminescent assay (also called as "cell–free based assay"). Some of these efforts had the X-ray structure solved and deposited in the Protein Data Bank (PDB), and the PDB codes are presented in the Table 5. Additional experimental data related to RTA inhibitors are presented in the Table S1.

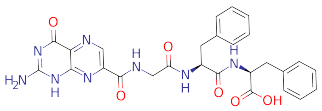
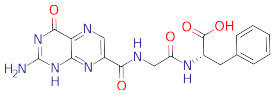
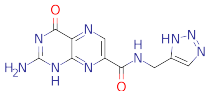
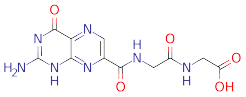
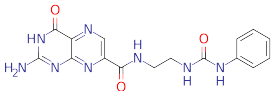
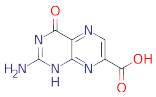
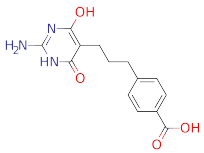
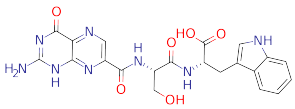
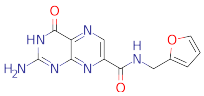
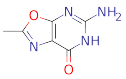
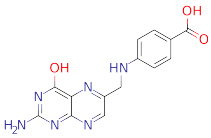
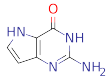
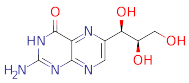
Figura 15 – X-ray structure of the RTA subunit (PDB code 3RTI).



X-ray structure of the RTA subunit (PDB code 3RTI). Carbon atoms of the RTA and formycin monophosphate - FMP (ligand) are depicted in cyan and yellow colors, respectively. In addition, the highlighted area corresponds to the location of RTA binding pocket **(A)** and the following illustrations (B, C, D, and E) comprises to an approximation of such location; In the absence of substrate, Y80 adopts a "closed" conformation **(B)**; The surface representation of the RTA in "closed" conformation presents a single binding pocket **(C)**; In the presence of substrate, Y80 adopts an "open" conformation. FMP binding mode in the RTA active site resembles the binding mode of adenine – a natural substrate – and exposes a second binding pocket **(D)**; The surface representation of RTA in the presence of substrate presents two binding pockets **(E)**; Dashed white lines highlight the RTA binding pocket in its closed and open conformation, respectively.

Although some of these blockers (Table 5 and Table S1) present IC_{50} in the micromolar range, such compounds have been show to be insoluble and/or toxic in cell-based assays.[114] Other countermeasures for ricin poisoning include vaccines and inhibitors of cellular trafficking[124, 125, 126, 127], however, vaccination carries many practical challenges, and the target for cellular trafficking inhibitors is not completely understood, which, in turn, complicates rational drug design.[128] Overall, with the increasing of screening libraries database, such as the ZINC15[129, 130], which has more than 1 billion compounds available for virtual screening (VS) campaigns, highlights the idea that VS methods can be still a promising alternative to find a competitive inhibitor for the RTA active site, although it must be thoroughly evaluated.[131, 132] In this sense, Botelho et al.[133][133] evaluated the PubChem database through VS approach, considering ligand- and structure-based methods and found three compounds with potential to inhibit the RTA active site.

Tabela 5 – List of Main X-Ray Structures and Experimental IC_{50} (μM) data of RTA-Ligand Complexes Available in the Protein Data Bank.

PDB ID	Chemical Structure	IC_{50}	Reference
4HUP		15	[128]
4HUO		20	[128]
4ESI		70	[134]
4HV7		115	[128]
4MX1		209	[135]
3PX8		230	[136]
3EJ5		270	[123]
4HV3		300	[128]
3PX9		380	[136]
1IL9		400	[137]
1BR6		600	[138]
1IL4		1400	[137]
1BR5		2000	[138]
1IL3		2800	[137]

Recently, in order to improve a protocol to discern RTA active inhibitors from inactive ones, Chaves and co-workers[139] performed a validation study by using re-docking and Steered Molecular Dynamics (Steered MD; SMD) simulations to assess the relative binding affinity of the six most potent pterin-based RTA inhibitors. Steered MD is a special type of MD simulation which applies a time-dependent external force on parts of the system, in order to verify how it responds to such forces, similar to the Atomic Force Microscopy (AFM) experiment. In contrast to AFM, the force probe employed by Steered MD simulation provides information at the atomic level resolution, which can be a crucial complement to the AFM experiment.[140, 141] Furthermore, in the drug discovery process, such a method can be used to investigate the relative binding affinity to distinguish active inhibitors from inactive ones[142], as also to elucidate specific molecular mechanisms responsible for favoring binding or uncoupling.[143, 144, 145, 146] In their validation study, Chaves and co-workers[139] showed that the searching algorithm of the Vina program[147] is able to predict the native pose of pterin-based inhibitors into the RTA binding pocket, however, its scoring function achieved a poor performance to rank such compounds according to experimental data. In this way, it was employed a rescore process based on ligand undocking force/energy profiles provided by replicas of SMD simulations. Their results presented strong correlations with experimental data ($R^2 > 0.9$). The authors proposed that such an approach can be used as a final step in the VS protocol to correctly rank active RTA inhibitors from inactive ones.

Many score functions in docking programs attempt to express binding energy, but in general, such functions present poor performance to ranking poses according to experimental data. In the final stages of VS campaigns, an incorrect ligand ranking results in false-positive predictions, which is not ideal for *in vitro* validation assays. Overall, this poor performance is due to the inability of considering quantum effects such as charge transfer between protein and ligand, for example.[148] In addition, although affinity in protein-ligand complexes is influenced by entropic and enthalpic terms, the ligand specificity to a respective binding pocket is primarily driven by enthalpy.[149, 148] In other words, selective ligands for a given receptor have a favorable binding enthalpy value resulting from specific intermolecular interactions. Therefore, screening strategies that search for more selective ligands could be focused on finding compounds with the most favorable binding enthalpy values. In this way, semiempirical quantum chemical methods (SQM) has been employed in the virtual screening scenario for such purpose. Among some types of SQM (RM1[150], PM6[151], PM6-DH+[152], PM6-D3H4[153], and PM7[154]), PM7 appears as a promising approach to derive the enthalpy values in protein–ligand complexes. In comparison with other predictive methods, mainly variations of PM6, PM7 achieved a good agreement with experimental data, with Pearson correlations up to 0.98.[155, 156, 157].

To date, no VS studies have been reported by using the SMD and SQM approaches as a filtering step to screen RTA inhibitors. Therefore, in this work, we coupled such

approaches in a VS pipeline, as follows: (*i*) VS library preparation; (*ii*) exclusion of undesirable compounds according to lead-like rules; (*iii*) structure-based pharmacophore screening; (*iv*) molecular docking combined with hydrogen bonds inspection; (*v*) a consensus score considering the ligand undocking procedure through the SMD simulations and SQM calculations binding enthalpy by using PM7. The whole VS cascade was used to screen more than 2 million compounds from the ZINC database, which in turn provided us seven novel scaffolds with high potential to inhibit the active site of the RTA.

5.1.4 Theoretical Methods

The lead-like subset of the ZINC database[129] was used in this VS study. According to Oprea and co-workers[158], lead compounds show interesting and relevant features, such like: (*i*) simple chemical attributes, in order to be amenable in future optimization efforts; (*ii*) favorable patent situation; and (*iii*) good ADME (absorption, distribution, metabolism, and excretion). In general, the hierarchical VS cascade was performed in six sequential steps, and more details will be presented in the following subsections as well in the supporting material.

5.1.4.1 Virtual Screening Library Preparation

The VS library used in the present study was retrieved from the ZINC database[129] by using the following options: (*i*) **Representation**: 3D; (*ii*) **Reactivity**: Standard; (*iii*) **Purchasability**: In Stock; (*iv*) **pH**: Reference Models (forms dominated at pH 7.4); (*v*) **Charge**: Neutral; and we retrieved an amount of 2.261.576 compounds. Then, a filtering step was performed in order to exclude undesirable compounds according to the lead filter proposed by Oprea[158], which in turn is available in the FILTER program (version 3.0.0.1; OpenEye Scientific). For the remaining compounds (1.484.477) we also performed a multi conformational searching by using OMEGA[159] program (version 3.0.0.1; OpenEye Scientific) that was set to find 50 low energy conformers for each compound in the VS library.

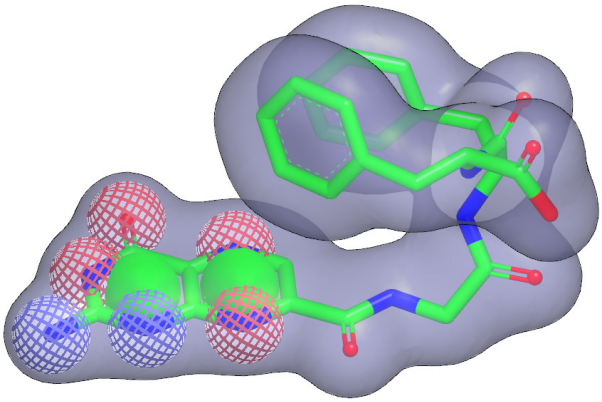
5.1.4.2 Structure-Based Pharmacophore Screening

To increase the chance to find an RTA selective hit, a shape-based overlay screening was performed by using ROCS[160] software (version 3.2.2.2; OpenEye Scientific). In this part, the interaction pattern of the 19M ligand into the RTA active site was used to build a shape query (Figure 16). Then, this shape query was compared to the shapes of VS library using a Gaussian representation of molecular shape and, the molecules that show the best fitting are expected to be most likely actives on the RTA binding pocket. Furthermore, there are 13 score functions in the ROCS program, therefore, we carried out a validation

step using the active RTA inhibitors and inactive ones (supplementary material, Table S1), as well as a set of decoys provided by DUD-E[161] web service (www.dude.docking.org).

The statistical analysis of the ROC curve was used to evaluate the performance of these score functions in such a validation library. In summary, six score functions presented AUC > 0.9: TanimotoCombo, ComboScore, ColorTanimoto, RefColorTversky and RefTverskyCombo (supplementary material, Figure S1). For more details on how the validation library was built as well as ROC curve results, see the supporting information document. Then, the ROCS search in the VS library reported the 10.000 best-ranked compounds for each one score function and after redundancy exclusion, we have been found 25.600 compounds.

Figura 16 – ROCS's shape query obtained from the ligand in 4HUP structure.



ROCS's shape query was obtained from the ligand in the 4HUP structure. The green spheres illustrate ring features, red spheres correspond to hydrogen bond acceptors, and blue spheres to hydrogen bond donors. The main interaction pattern between the binding pocket of RTA and 19M considers: (*i*) hydrogen bond donors with the backbone carbonyls of G121 and V81 (blue spheres); (*ii*) hydrogen bond acceptors with the backbone nitrogen atom of V81, Y123, and the side-chain of R180 (red spheres); (*iii*) π–stacking network between the aromatic rings of Y80 and Y123 (green spheres).

5.1.4.3 Molecular Docking

The fished compounds from the previous stage were submitted to molecular docking protocol using GOLD[162]. A recent review showed that GOLD has a high degree of accuracy and provides great potential to retrieve inhibitor hits in VS studies.[163] In addition, as we quoted in the Table 5, there are fourteen RTA structures available in the PDB and GOLD software have multiple score functions (GoldScore, ChemScore, ASP, and ChemPL), therefore, which of these score functions and RTA receptors would be the best choice for a molecular docking study? As far as we know, no studies have been reported performing similar strategies. Thereby, analogous to the previous section, we performed validation steps in order to answer this question and more details are available in the supporting information. In summary, our semi-flexible docking protocol has achieved a high degree of

hits when using the 1BR5 receptor combined with the ASP score (supplementary material, Figure S2). We observed that most poses with high scores were close to the experimental pose in X-ray structure (supplementary material, Figures S3, S4). Further, an enrichment plot showed that ASP score had a good performance to discern active RTA inhibitors from inactive ones, with AUC > 0.9 (supplementary material, Figure S4). Then, the GOLD search in VS library reported ASP scores in the range of 19 to 55, however, we retrieved only poses with ASP score > 45, which comprise an amount of 1.357 poses.

5.1.4.4 Hydrogen Bonds Inspection

The inhibitors that were co-crystallized with the RTA chain showed some structural similarities. The most important corresponds to functional groups that form hydrogen bonds to the backbone carbonyls of V81 and/or G121. Therefore, in this step, we tried to select docking poses that did more than one hydrogen bond with such residues. A homemade script that uses the Hbonds plugin included in the VMD program[164] was used to perform this search. The hydrogen bond counting was performed in two sequential steps: (*i*) between ligand pose outputted from molecular docking protocol and RTA; here, we retrieved 453 poses; (*ii*) further, we applied an implicit solvent model to each RTA–ligand complex and then performed 1.000 steps of geometry optimization only for the ligand atoms by using the conjugate gradient algorithm included in the NAMD software[165]; at the end of this step, hydrogen bonds were recounted (using the same criteria) and 74 novel poses were selected and submitted to molecular dynamics simulations.

5.1.4.5 Molecular Dynamics Simulations

MD simulations were performed in order to equilibrate all RTA–ligand complexes. In this step, RTA and ligands were parameterized with FF14SB[166] and GAFF[167] force fields, respectively, and the partial atomic charges of the ligands were calculated according to the AM1–BCC method using the Antechamber program[168] included in AmberTools. Then, all complexes were inserted in a cubic water box of 20 Å containing TIP3P waters and ions Cl^-. For MD simulations we used NAMD program[165] with the following configuration parameters: (*i*) periodic boundary conditions; (*ii*) restriction of vibration of covalent bonds involving hydrogen atoms, HOH angles, and the OH bond distance of TIP3P water molecules; (*iii*) time steps equal to 2fs; (*iv*) electrostatic interaction cutoff of 9Å for all steps of the simulations. The starting geometries were submitted to minimization, heating, and pressurization steps. Then, the last frame of each complex was extracted and hydrogen bonds were recounted, using the same criteria of the previous section (more than one hydrogen bond into the binding pocket, including hydrogen bonds with V81 and/or G121). Here, we retrieved 30 poses that were submitted to equilibration step during 10 ns with constant pressure (1.0 atm) and temperature (310 K); the coordinates were stored every 2 ps.

We carried out the analysis for MD trajectories by calculating RMSD and other quantities, such as: (*i*) contact analysis to detected residues that surrounding ligand during the MD simulation; (*ii*) interaction energy calculations between residues selected in step (*i*) and ligand; and (*iii*) hydrogen bond calculations. Such analysis was performed using the CPPTRAJ software[169], also included in AmberTools.

5.1.4.6 Post-Docking Filtering by Steered MD Simulations and PM7

SMD force profiles such as mean force $\langle F \rangle$, rupture force $\langle F_{max} \rangle$, and mean of non-equilibrium work $\langle W \rangle$ were estimated by replicas of independent SMD simulations. Here, a specific atom of the ligand (SMDAtom) is connected to a dummy atom by a spring with spring constant k. This dummy atom moving over the pulling direction (or reaction coordinate) with a constant velocity v, which in turn also moving the SMDAtom over the same pulling direction. Eq. 5.1 summarizes this biased condition to move the dummy atom in function of time:

$$F_{ext} = k(vt - \Delta x_i) \tag{5.1}$$

where F_{ext} is the external force, k is the spring constant, v is the constant velocity, t is the time and $\Delta x_i = x_i - x_{i,0}$ is the displacement of atom i from its start position. The work W done by the force on the isobaric-isothermal ensemble can be calculated using Equation 5.2, which in turn is linked to the free energy (ΔG) by Jarzynski's equality[170, 171, 172] (Eq. 5.2):

$$W = \int_0^t F_{ext} \cdot dt \tag{5.2}$$

$$e^{-\beta \Delta G} = \left\langle e^{-\beta W} \right\rangle_{\infty} \tag{5.3}$$

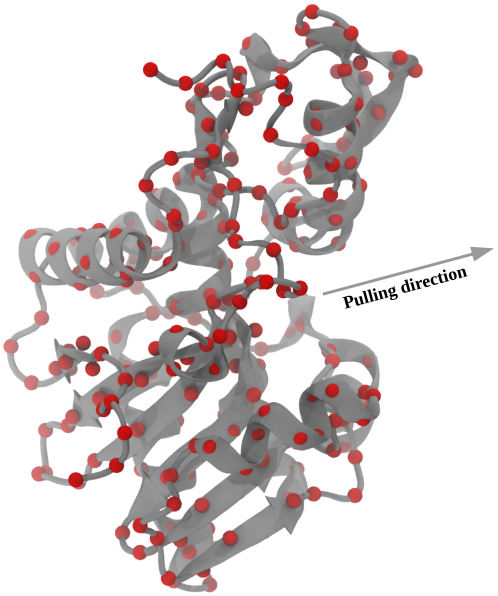
where $\beta = \dfrac{1}{k_B T}$, in which k_B is the Boltzmann constant and T is the absolute temperature; W is the work, given by Eq. 5.2; ΔG is the difference in Gibbs free energy between the thermodynamic states of the bound and unbound systems; and $\langle ... \rangle_{\infty}$ denotes the mean average work for infinite times of switching processes between the two thermodynamic states. We did not perform the block-average analysis; thereby, we got the free energy difference as $\Delta G = \langle W \rangle_n$, with $n = 5$.

The last frame of equilibration step was used as starting point to perform the ligand undocking simulations. NAMD program[109] was used with the same parameters of the equilibration step, except for both the biased conditions to pulled out ligands from the binding pocket and the SMD parameters, which were defined as: (*i*) pulling direction (or reaction coordinate, see Figure 17); (*ii*) constant velocity, v; and (*iii*) spring constant, k. In summary, SMD post-docking step was carry out in two scenarios: (*i*) $k = 2$ $kcal \cdot mol^{-1}$Å^{-2} and $v = 0.05$ Å/ps; (*ii*) $k = 10$ $kcal \cdot mol^{-1}$Å^{-2} and $v = 0.05$ Å/ps.

As previously mentioned, SMD simulation corresponds to a computational technique that mimics the AFM experiment, however, there are some limitations that may cause a divergence of results between the simulation and experiment. Such factors include the choice of "pulling path", spring hardness and pulling velocity. Regards to reaction coordinate, if it is chosen arbitrarily, likely, the vector does not represent the experimental condition. In physicochemistry terms, the pathway to egress or enter occurs in ways where the energy barrier is the smallest possible, in other words, the best "pulling path" corresponds to the path with less steric hindrances. In the validation study carried out by Chaves and co-workers[139], the SMD parameters combined with a reaction coordinate similar to that presented in this study showed a high correlation with experimental data, which in turn led us to conclude that the suggested "pulling path" is not a factor for such divergence of results (simulation vs. experiment). Regards to velocity constant, Jarzinski's Equality tells us that the relative Gibbs free energy between two thermodynamic states can be measured more accurately if the sampling process occurs slowly, that is, under conditions close to equilibrium. In this case, a low pulling velocity is used, which results in trajectories on the time scale of nanoseconds. However, in the drug discovery process, performing SMD simulations in a condition close to equilibrium is a computational bottleneck due to the cost involved. Therefore, calibration procedures with different velocity constants are strongly recommended in order to obtain a good balance between computational cost and correlation with experimental data. Chaves and co-workers tested sets with different spring force and velocity constants, respectively, and showed that both computational cost and the correlation with experimental data decrease as a velocity increment [139].

An illustration of the pulling direction and SMDAtoms that we used in this study are depicted in Figure 17 and Figure 18, respectively. Furthermore, in order to ensure the ligand movement only over the reaction coordinate axis, some restrictions were employed: (i) harmonic constraints with a constant force of 1.0 $kcal \cdot mol^{-1}$Å^{-2} on the Cα atoms of the RTA in order to limit abrupt movements (Figure 17); and (ii) a harmonic constraint cylinder with a diameter of 6 Å.

Figura 17 – llustration of the pulling direction employed in SMD simulations.

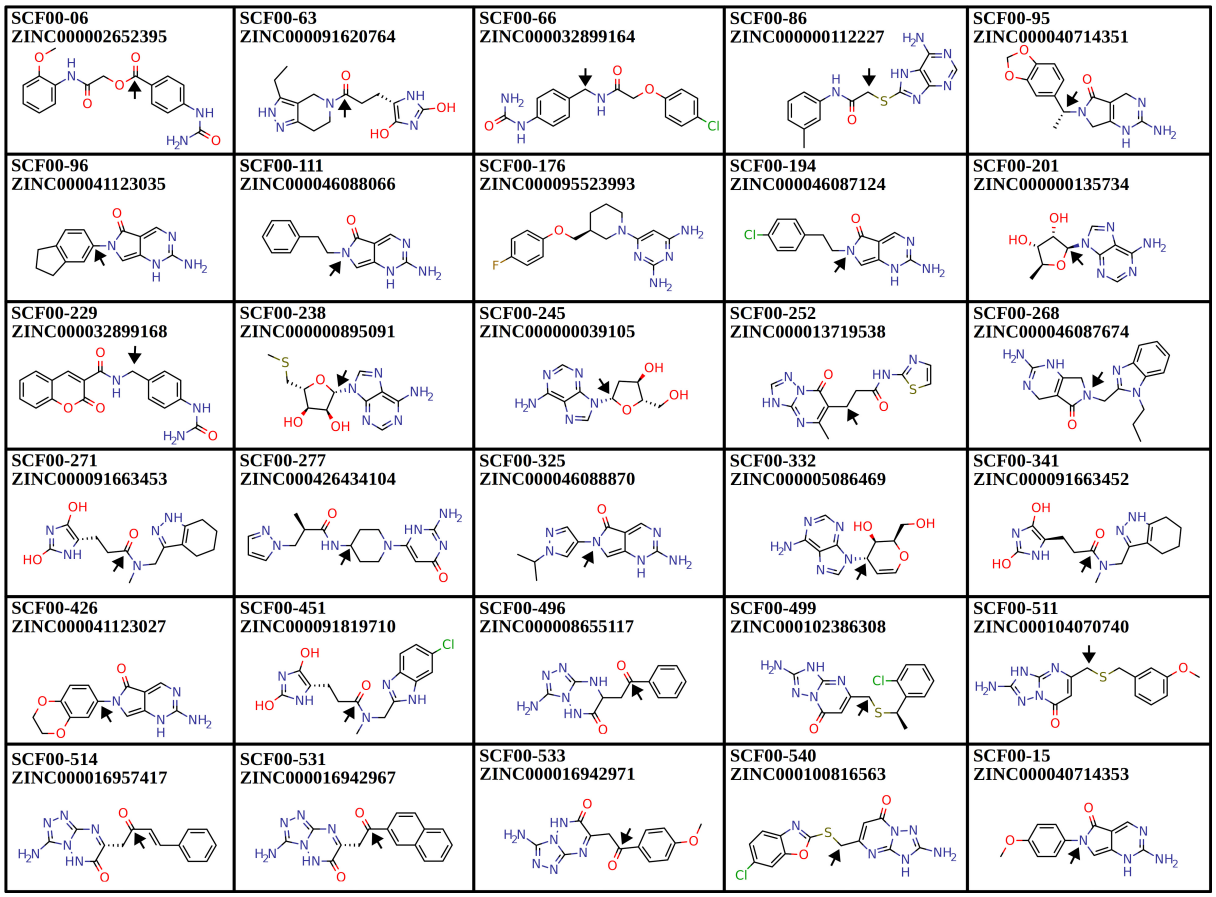


The red spheres comprise the Cα atoms of the RTA which were restrained over the whole SMD simulations.

As a final step, we performed single-point quantum chemistry calculations to measure relative quantities of the binding enthalpy ($\Delta H_{binding}$) for all simulated complexes, using PM7[154], a semiempirical quantum chemical method. The SQM parameters were set according to the study of Urquiza-Carvalho and co-workers[173]: (*i*) linear scaling algorithm MOZYME[174], available into MOPAC [175, 176] package; (*ii*) SCF convergence criteria in default configuration and cutoff radius of 9 Å; (*iii*) COSMO implicit solvent field with a relative permittivity of 78.4; (*iv*) and an effective solvent molecule radius of 1.3 Å. Such calculations were performed for different frames of the trajectory from the time range of 5 to 10 ns (every 200 ps); Eq. 5.4 was used to calculate the binding enthalpy, and averages, as well as standard deviations were calculated using the set of frames quoted previously (25 frames).

$$\Delta H_{binding} = \Delta H_f^{complex} - \left(\Delta H_f^{receptor} + \Delta H_f^{ligand}\right) \tag{5.4}$$

Figura 18 – Chemical structures and ZINC codes of the selected compounds to post-docking analysis.



The black arrow point to the SMDAtom used to pull the ligand out from the binding pocket.

5.1.5 Results

The cross-docking study showed that considering the 1BR5 structure combined with the ASP score was extremely useful to predict the co-crystallized ligands pose into the active site of the RTA (Figure S3). Moreover, the ASP score was able to discern active ligands from inactive ones in another validation test, with AUC > 0.9)(Figure S4).

The GOLD search in the VS library produced 1.357 poses with ASP score > 45. However, the ranking process requires a more selective filtering step due to the inefficiency of docking scoring to correctly rank poses according to the experimental data. Therefore, we carried out post-docking steps to increase the chance to find a selective hit.

Once the most potent RTA inhibitors have shown hydrogen bonds with G121 and V81 (such residues play a significant role in the recognition of natural substrate), thereby, the first post-docking step was based in the search of hydrogen bonds between the pose of the ligand and such residues.

The hydrogen bond counting step performed in both sets of MD simulations

(geometry optimizations and the initial steps of the MD simulation) provided us an amount of 30 poses. Such poses resemble the binding mode of the natural substrate, adenine, including hydrogen bonds with V81, G121, R180, E177, and π-stacking interactions with Y80. Figure 18 presents the chemical structures of such compounds. Besides, in the following subsections, we will quote these molecules by the prefix "SCF00"instead of ZINC codes.

Then, we carried out a final post-docking step based on SMD simulations and binding enthalpy calculations, respectively. In the following sections, we will discuss such profiles for the 30 selected compounds and their relationship with the binding affinities. Table 6 presents a summary of the theoretical results that will be recurrently quoted throughout the following subsections.

Tabela 6 – Summary of the Calculated Results. All Numerical Data Presented in the 2nd and 3rd columns are in pN; 4th and 5th columns are in $kcal \cdot mol^{-1}$.

ID	$\langle F \rangle_{max}$	$\langle F_{max} \rangle$	$\langle W \rangle$	$\Delta H_{binding}$	$E_{docking}$	Rescore
SCF00–451	1007.88 ± 34.81	1142.53 ± 73.08	113.02 ± 12.82	-67.62 ± 4.46	48.50	3.00
SCF00–332	969.38 ± 271.66	1081.76 ± 156.62	94.51 ± 14.70	-59.21 ± 4.41	47.28	2.48
SCF00-238	798.25 ± 38.20	903.62 ± 30.49	96.41 ± 6.71	-62.53 ± 4.59	46.61	2.46
SCF00–499	765.34 ± 36.07	887.18 ± 72.65	82.48 ± 13.98	-65.53 ± 3.76	49.56	2.43
SCF00-63	923.03 ± 20.43	1041.82 ± 83.04	98.66 ± 11.69	-56.65 ± 4.08	45.44	2.39
SCF00-271	818.57 ± 303.89	940.87 ± 164.40	88.59 ± 17.55	-59.54 ± 4.76	46.90	2.31
SCF00–496	726.83 ± 45.94	823.45 ± 45.79	73.57 ± 8.88	-62.41 ± 3.32	49.49	2.18
SCF00-341	771.19 ± 57.60	857.37 ± 68.15	77.21 ± 4.69	-54.69 ± 6.51	47.35	1.97
SCF00-194	716.71 ± 42.85	880.77 ± 99.82	77.51 ± 8.76	-53.02 ± 4.66	46.36	1.93
SCF00-86	739.85 ± 77.17	846.77 ± 47.17	72.36 ± 4.14	-52.74 ± 5.18	45.64	1.85
SCF00-511	539.69 ± 59.60	632.73 ± 41.23	61.89 ± 9.00	-59.92 ± 5.10	50.00	1.82
SCF00-245	650.66 ± 74.26	761.81 ± 54.06	72.49 ± 11.35	-53.84 ± 3.81	46.67	1.81
SCF00-268	520.06 ± 187.65	652.02 ± 77.10	70.35 ± 9.31	-56.06 ± 4.65	46.80	1.78
SCF00-533	657.62 ± 154.59	821.00 ± 71.16	59.46 ± 10.45	-53.62 ± 5.74	52.02	1.74
SCF00-514	680.01 ± 18.75	764.44 ± 25.84	58.05 ± 3.70	-54,68 ± 6.30	50.14	1.72
SCF00-277	699.79 ± 144.70	786.81 ± 131.13	66.59 ± 7.25	-51.84 ± 5.38	46.98	1.71
SCF00-531	619.66 ± 25.07	702.02 ± 47.42	57.43 ± 4.47	-55.96 ± 6.33	51.83	1.70
SCF00-540	468.32 ± 107.21	579.61 ± 34.27	59.67 ± 8.16	-57.40 ± 5.49	53.06	1.67
SCF00-66	513.70 ± 44.70	609.26 ± 42.15	56.46 ± 4.61	-57.15 ± 7.24	45.45	1.66
SCF00-201	563.26 ± 67.81	710.42 ± 78.10	69.09 ± 14.56	-51.37 ± 3.13	46.42	1.65
SCF00-229	578.31 ± 52.13	695.78 ± 47.46	70.38 ± 4.51	-51.11 ± 4.86	46.56	1.64
SCF00-252	489.21 ± 67.86	593.48 ± 19.49	59.22 ± 3.62	-56.10 ± 5.61	46.71	1.63
SCF00-96	583.40 ± 51.86	690.49 ± 68.04	56.99 ± 9.67	-54.08 ± 5.00	45.68	1.62
SCF00-325	637.50 ± 142.97	826.77 ± 132.12	58.39 ± 10.29	-48.20 ± 4.09	47.25	1.54
SCF00-95	518.00 ± 106.52	615.08 ± 68.41	54.16 ± 2.47	-53.37 ± 3.03	45.68	1.51
SCF00-111	580.32 ± 42.19	632.73 ± 41.23	61.89 ± 9.00	-49.33 ± 5.01	45.77	1.45
SCF00-06	690.51 ± 55.14	753.75 ± 40.58	60.83 ± 4.43	-46.12 ± 6.87	45.03	1.43
SCF00-426	632.73 ± 52.99	701.88 ± 28.47	53.77 ± 2.51	-48.29 ± 3.97	48.26	1.40
SCF00-15	741.80 ± 84.89	839.01 ± 57.66	73.25 ± 8.85	-39.72 ± 3.49	45.09	1.38
SCF00-176	-	-	-	-39.52 ± 4.74	46.24	0.00

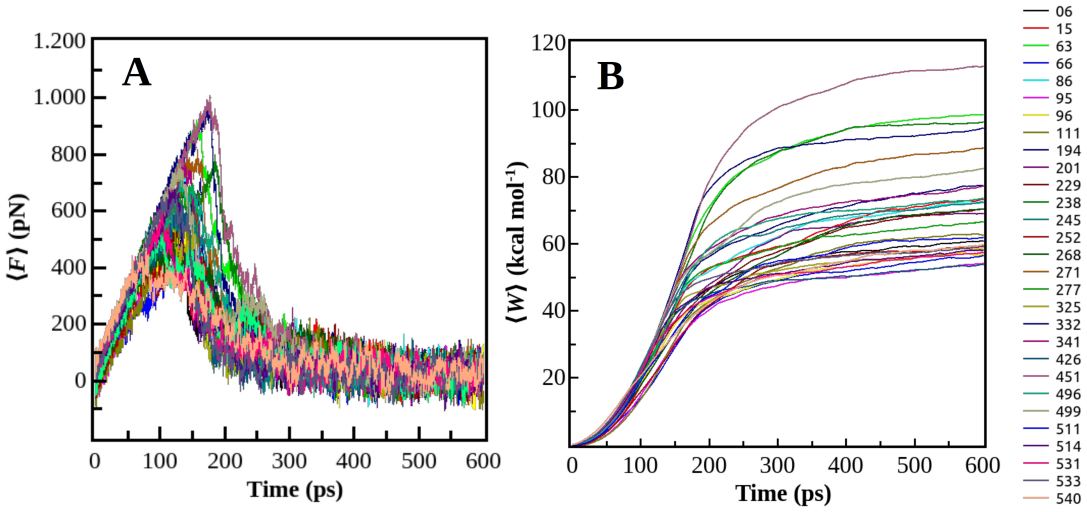
5.1.5.1 SMD Force Profiles

The last frame of the equilibration MD trajectory was used as the start point for the replicas of SMD simulations. In the Figure 19 is presented the force/energy profiles. The SCF–176 compound has been shown unstable into the binding pocket at the point when hydrogen bonds with V81 and G121 were broken, thereby, we do not perform undocking simulations for it.

The SMD profiles showed the same pattern of other studies that used such procedure in protein–ligand complexes to determine the relative binding affinity of ligands. [177, 178, 179] Summing up, the force increases until a limit, $\langle F \rangle_{max}$, and then decreases to lower values (Figure 19A). The force increment occurs due to non-covalent interactions between ligand and active residues of the binding pocket. The force profile after this threshold corresponds to the stage that such non-covalent interactions are broken and ligand walks out of the active site by the pulling pathway. If no one hindrance will impose on the ligand, the force decreases. It is expected that strong ligands show high values of $\langle F \rangle_{max}$, consequently, it is also expected high energy of $\langle W \rangle$, once the forces are integrated by Eq. 5.2.

The SMD force profiles pointed out SCF00–451 as a strong ligand due the high values of $\langle F_{max} \rangle$ and $\langle W \rangle$; $\langle F_{max} \rangle = 1142.53$ pN, and $\langle W \rangle = 113.02\ kcal{\cdot}mol^{-1}$, respectively (Table 6; Figure 19B). Next, SCF00–332, SCF00–63, SCF00–271, and SCF00–238 also presented high values of $\langle F_{max} \rangle$, with $\langle F_{max} \rangle$ = 1081.76, 1041.82, 940.87, and 903.62 pN, respectively (Table 6). Regards to the non-equilibrium work $\langle W \rangle$, from above quoted ligands at least three showed $\langle W \rangle > 90\ kcal \cdot mol^{-1}$: SCF00-63, SCF00–238, and SCF00–332, with $\langle W \rangle$ = 98.66, 96.41, and 94.51 $kcal \cdot mol^{-1}$, respectively (Table 6, Figure 19B).

Figura 19 – Steered MD force profiles using $k = 2\ kcal \cdot mol^{-1}$Å^{-2}, and $v = 0.05$ Å/ps.



Mean force profile $\langle F \rangle$ as a function of time (**A**); Mean of non-equilibrium work profile $\langle W \rangle$ as a function of time (**B**).

At atomic level, intermolecular interactions play a key role in protein-ligand complexes, being the most relevant ones: (*i*) hydrogen bonds, (*ii*) electrostatic, and (*iii*) Van der Waals interactions. Ruiz-Carmona and co-workers[180] pointed out that hydrogen bonds considerably contribute to both stability and binding affinity of protein-ligand complexes, and breaking such interactions, through undocking procedures, can provide accurate and meaningful information about the binding ability of the compounds under study.

In our previously validation work, by using a similar pulling direction for the 19M ligand and the same biased conditions of undocking simulations employed here, we showed that such ligand had a rupture force and non-equilibrium work of $\langle F_{max} \rangle = 1056.19\ pN$, and $\langle W \rangle = 95.01\ kcal \cdot mol^{-1}$, respectively.[139] Additionally, such a compound showed a network of hydrogen bonds into the RTA binding pocket, including V81 and G121. In the present study, SCF00–451 showed a similar network of hydrogen bonds and higher values of $\langle F_{max} \rangle$ and $\langle W \rangle$, which lead us to suggest that such compound might present similar or better antagonist activity than 19M ligand. Moreover, the following compounds presented high values of $\langle F_{max} \rangle$ and $\langle W \rangle$: SCF00–332, SCF00-63, SCF00-271, SCF00-238, and a similar intermolecular interactions network into the binding pocket of the RTA.

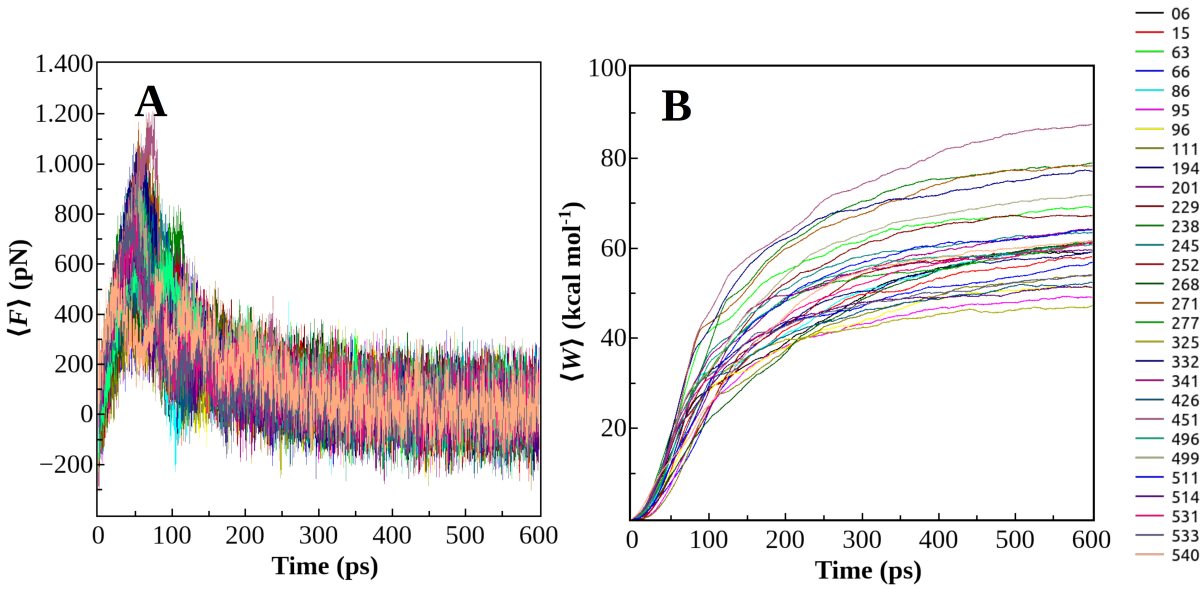
5.1.6 SMD Force Profiles Using a Different Spring Constant

We carried out a second scenario for the ligand undocking simulations. Here, we attempt to evaluate the impact of spring hardness on the SMD profiles, using $k = 10$ $kcal \cdot mol^{-1}$Å^{-2} and $v = 0.05$ Å/ps.

The undocking procedure using such biased conditions showed that inhibitors were pulled out faster from the binding pocket, however, $\langle F \rangle$ was dominated by noises (Figure 20); we also observed an increase in the standard deviation (supplementary material, Table S2). Moreover, $\langle F \rangle$ shows the same pattern of the force profile presented in the previous section, which means that force increases until a limit, $\langle F \rangle_{max}$, and, then, decreases (Figure 20A).

Although the high noise incidence, SCF00–451 also showed the highest values of $\langle F_{max} \rangle$ and $\langle W \rangle$, with $\langle F_{max} \rangle = 1518.27$ pN, and $\langle W \rangle = 87.53$ $kcal \cdot mol^{-1}$, respectively (Table S2, Figure 20B). Sequentially, SCF00–271, SCF00–332, and SCF00–63 have also shown high values of $\langle F_{max} \rangle$, with $\langle F_{max} \rangle = 1445.41$, 1440.37, and 1327.70 pN, respectively (Table S2). Regards to the non-equilibrium work, the following complexes showed $\langle W \rangle >$ 70 $kcal \cdot mol^{-1}$: SCF00–238, SCF00–271, SCF00–332, and SCF00–499, with $\langle W \rangle = 78.9$, 78.29, 77.36, and 71.85 $kcal \cdot mol^{-1}$, respectively (Table S2, Figure 20B).

Figura 20 – Steered MD force profiles using $k = 10$ $kcal \cdot mol^{-1}$Å^{-2}, and $v = 0.05$ Å/ps.



Mean force profile $\langle F \rangle$ as a function of time (**A**); Mean of non-equilibrium work profile $\langle W \rangle$ as a function of time (**B**).

5.1.6.1 Semiempirical Calculations

The binding enthalpy calculation using PM7 method provided us an additional guidance to confirm the leader compounds. A set of 25 geometries got from the last 5ns of the equilibrium MD stage were used to calculate binding enthalpies. In summary, we obtained energy values in the range of -39.52 to -67.62 $kcal \cdot mol^{-1}$; likewise SMD simulations, this approach pointed out SCF00–451 as the best hit, with $\Delta H_{binding} = -67.62$ $kcal \cdot mol^{-1}$ (Table 6). The following compounds also showed favorable binding enthalpies: SCF00–499, SCF00–238, SCF00–496, SCF00–511, SCF00–271, and SCF00–332, with $\Delta H_{binding} = -65.53, -62.53, -62.41, -59.92, -59.54$, and -59.21 $kcal \cdot mol^{-1}$, respectively (Table 6).

Rocha and Sant'Anna[149] carried out a computational study to find inhibitors for tyrosine phosphatases, a key family of enzymes that play an important role in cancer therapies. The authors conducted this study by using molecular docking and SQM (PM7) in order to calculate binding enthalpies of known inhibitors. In general, these authors showed that the binding enthalpy calculations using the PM7 method presented a correlation of 0.7 and a confidence level of 99% to identify active inhibitors.

Large negative values of binding enthalpy comprise a high degree of intermolecular specificity between the ligand and protein, thus contributing to the binding affinity. In another way, according to Jarzynski's equality, a relative measurement of the Gibbs free energy can be accessed by sampling simulations in non-equilibrium conditions. In the present study, five replicas of undocking simulations in non-equilibrium condition showed that some of the above quoted compounds presented high values of $\langle W \rangle$. Therefore, two different approaches pointed to the same set of ligands, which lead us to suggests that such compounds interact strongly with the binding pocket of the RTA, which can thus act as a potential blocker of its activity.

5.1.6.2 Rescoring by Using SMD Force Profiles and $\Delta H_{binding}$

In order to improve the ranking process, we have been proposed a rescoring procedure composed of SMD force profiles and $\Delta H_{binding}$. The score calculation was performed by the summation of normalized data using the following profiles: $\langle F_{max} \rangle$, $\langle W \rangle$, and $\Delta H_{binding}$, respectively. $\langle F \rangle_{max}$ was not used in such calculation because the maximum force peaks occur at different times in individual simulations; thus, the mean is not calculated from the maximum force data.[179] Table 6 presents all results of this procedure.

According to the rescoring data presented in Table 6, the following compounds were pointed out as top hits: SCF00–451 (3.00), SCF00–332 (2.48), SCF00–238 (2.46), SCF00–499 (2.43), SCF00–63 (2.39), SCF00–271 (2.31), and SCF00–496 (2.18).

The rescoring calculation using the SMD force profiles performed with spring constant of $k = 10\ kcal \cdot mol^{-1}\text{Å}^{-2}$ also pointed out SCF00–451 as the best hit, and the same group of compounds were ranked in top positions (Table S2).

5.1.6.3 MD Trajectory Analysis of the Best-Ranked Compounds

In this section we will discuss the dynamics of the equilibration step of the following RTA–ligand complexes: SCF00–451, SCF00–332, SCF00–496, and SCF00–499. We have selected such complexes mainly due to the high values of rescoring and different chemical scaffolds.

It is important to consider the conformational changes that each ligand induce in the protein backbone, however, in the present study, we have not been found significant abrupt changes in the RTA backbone. Therefore, our discussion will be focused mainly on non-covalent interactions between the ligand and receptor. In this sense, hydrogen bonds of the last 5ns were quantified and results are presented in the Table 7.

Tabela 7 – Hydrogen bond calculations between RTA chain and best-ranked ligands.

ID	Acceptor	DonorH	Donor	%	Average Distance (Å)
SCF00-451	G121(O)	H13	N3	87.33	2.83
	N1	H	V81(N)	69.13	2.90
	V81(O)	H15	O2	65.33	2.79
	O1	HH12	R180(NH1)	31.99	2.89
	O	H	Y123(N)	26.15	2.90
	A79(O)	H14	O1	14.59	2.86
	O1	HH22	R180(NH2)	10.40	2.90
	N	HH22	R180(NH2)	2.88	2.89
	E208(O)	H12	N2	1.72	2.88
SCF00-332	E177(OE2)	H11	O1	99.76	2.62
	E177(OE1)	H12	O2	92.64	2.68
	N	HH22	R180(NH2)	48.62	2.91
	N1	H	V81(N)	34.55	2.92
	N	HH12	R180(NH1)	22.47	2.92
	G121(O)	H9	N4	19.23	2.86
	G121(O)	H10	N4	10.36	2.83
	G120(O)	H9	N4	10.20	2.87
	O1	HH12	R180(NH1)	8.20	2.84
	E208(O)	H12	O2	3.76	2.71
	E177(OE2)	H12	O2	2.92	2.88
	V81(O)	H10	N4	1.92	2.90
SCF00-496	N1	H	V81(N)	67.49	2.89
	G121(O)	H9	N5	59.86	2.87
	G121(O)	H7	N4	58.50	2.87
	N2	HH12	R180(NH1)	41.66	2.91
	O	H	Y123(N)	36.55	2.88
	V81(O)	H8	N5	35.27	2.89
	O1	HH22	R180(NH2)	9.04	2.86
SCF00-499	O	HH12	R180(NH1)	74.41	2.84
	G121(O)	H11	N2	62.10	2.83
	V81(O)	H12	N4	53.94	2.85
	O	HH22	R180(NH2)	27.55	2.87
	V81(0)	H13	N4	21.07	2.86
	N1	H	N	20.87	2.94
	O	HG	S176(OG)	11.60	2.79
	G121(O)	H13	N4	9.00	2.88
	G121(O)	H12	N4	3.92	2.88

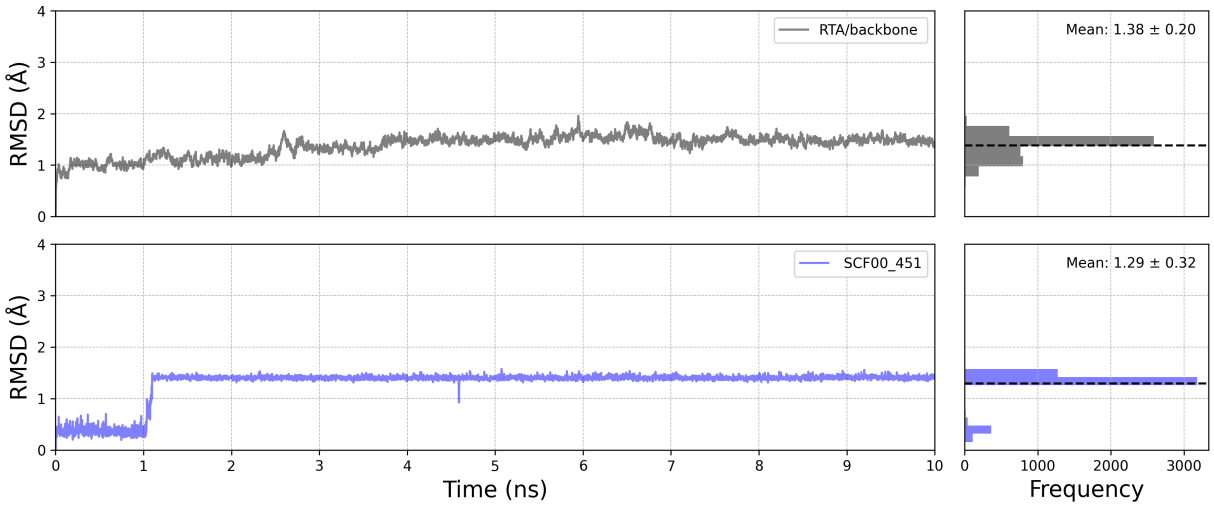
2^{nd} and 3^{rd} columns contain an atom of the ligand or residue that acting as acceptor/donor of hydrogen bonds, respectively. 5^{th} and 6^{th} columns correspond to the occurrence and average distance of such interaction in 2.500 frames of the MD simulation (5 to 10ns), respectively.

5.1.6.4 Compound SCF00–451

For this compound, we observed a binding pose that resembles the natural substrate, adenine, along the whole MD trajectory, which includes hydrogen bonds with G121, V81, R180, Y123. Furthermore, uncommon hydrogen bonds with A79 and E208 were also detected (Table 7).

Regarding the dynamics of this compound into the RTA active site, the first nanosecond presents few conformers with low RMSD values. Then, occurs a conformational change which, in turn, goes until the end of the simulation (Figure 21). This second conformation has also been filled by a high degree of low RMSD values, which lead us to suggest that achieved a local minimum. The frequencies analysis provided by histogram showed the presence of two clusters, each one comprises different conformations of SCF00–451 into the binding pocket (Figure 21).

Figura 21 – RMSD analysis of the equilibrium MD trajectory of RTA–SCF00–451 complex.

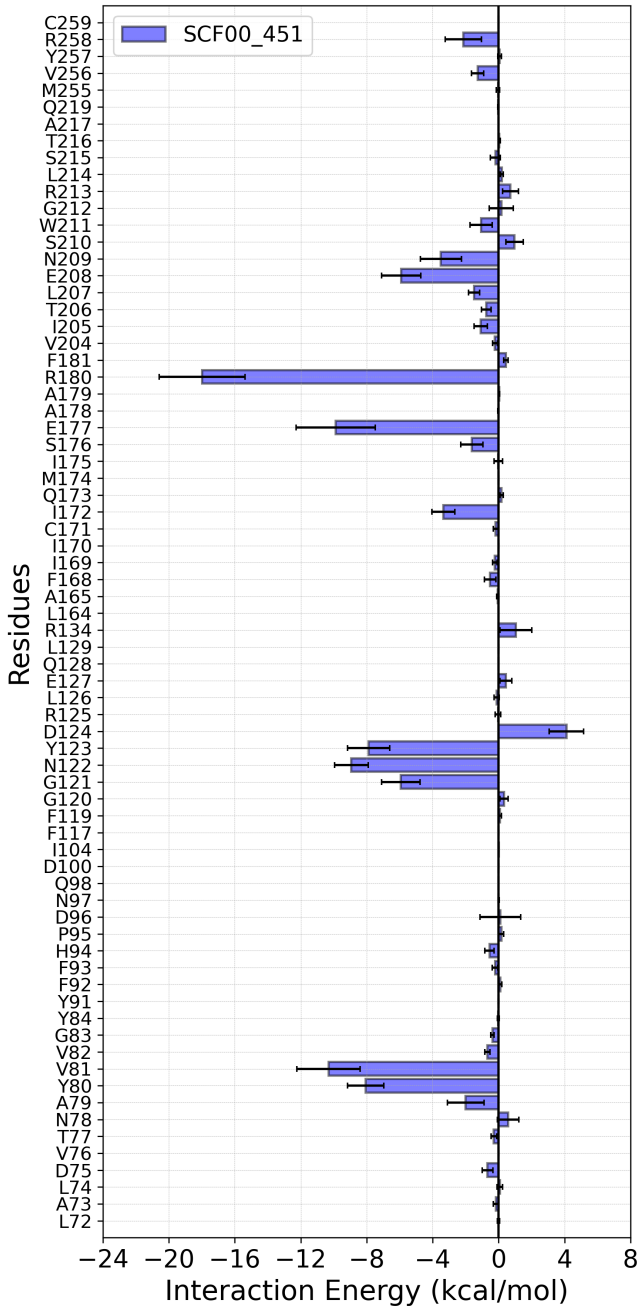


In top position is plotted the RMSD considering only backbone atoms of the RTA. In bottom position, the same analysis, however, considering only heavy atoms of SCF00–451.

Furthermore, the interaction energy calculations showed an amount of 37 intermolecular interactions; 9 were repulsive (the most prominent were D96, D124, R134, S210, G212, and R213), and 28 were attractive ones (the most prominent were Y80, V81, G121, N122, Y123, E177, R180, and E208) (Figure 22).

In the early stage of MD simulation (0–1ns) the double ring moiety of SCF00–451 appears in an unfavorable energy conformation, maybe due to a repulsive interaction between N2 and side chain of R180, which in turn forms a hydrogen bond with a surrounding water molecule (Figure 23A). The other moiety appears forming strong hydrogen bonds with G121 and V81, with occurrences of 87.33% and 69.13% (and 65.33%), respectively (Table 7).

Figura 22 – Interaction energy analysis between SCF00–451 and the RTA chain along the last 5ns of MD simulation.



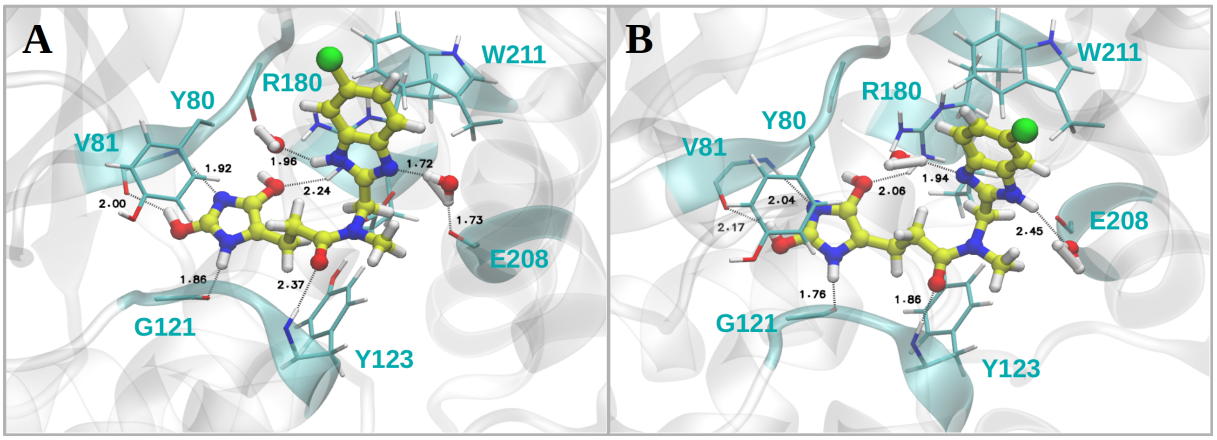
Blue bars comprise the average interaction energy between residues and ligand.

In the second stage (after 1ns), the double ring moiety rotates and the system likely converges to a lower energy level, forming two hydrogen bonds: (*i*) between the nitrogen (N) of the double ring in the ligand (acceptor) and the side chain of R180 (donor); (*ii*) between the nitrogen (N2) of the double ring in the ligand (donor) and the backbone oxygen of E208 (acceptor) (Figure 23B); however, the occurrences of these last ones are quite low, 2.88% and 1.72%, respectively (Table 7). In this case, we suggest that the lowest values may have occurred due to a hydrogen bond network between such atoms

and surrounding water molecules.

In summary, these intermolecular interactions contributed to its conformational rigidity into the pocket, including the whole network of hydrogen bonds and attractive interactions, mainly with G121 and V81 in the deep part of the pocket, as well as with E177 and R180. Additionally, such interactions have a direct contribution in the large negative value of $\Delta H_{binding}$, and the high values of $\langle F_{max} \rangle$ and $\langle W \rangle$ (Table 6).

Figura 23 – The binding pose of the SCF00–451 compound into the RTA binding pocket in early (A) and final (B) stages of the MD simulation.



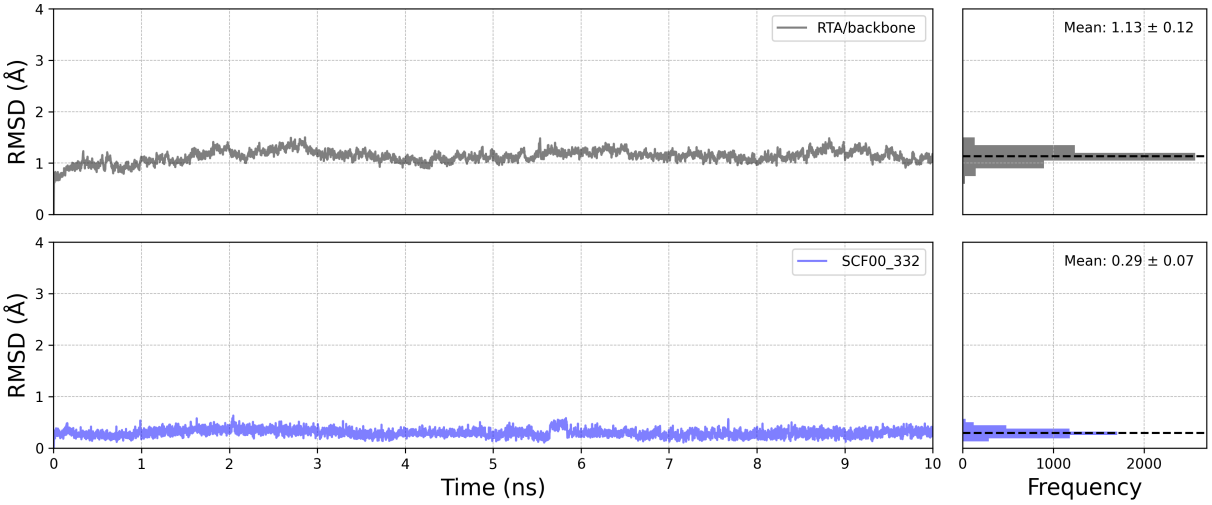
Carbon atoms of the RTA and SCF00–451 are depicted in cyan and yellow colors, respectively. Black dashed lines correspond to the hydrogen bonds.

5.1.6.5 Compound SCF00–332

Here, it was also observed a network of hydrogen bonds that resemble the natural substrate, adenine, including hydrogen bonds with V81, G121, and R180. Furthermore, uncommon hydrogen bonds with G120, E177, and E208 were also detected along the MD simulation (Figure 24).

In contrast to SCF00–451, such compound has filled by only one group of conformers (Figure 24). The most prominent hydrogen bond occurs between the hydroxyl groups of the ligand and side-chain oxygen atoms (OE2 and OE1) of E177, with high occurrences, 99.76% and 92.64%, respectively (Table 7). In comparison to SCF00–451, such compound has shown lower occurrences of hydrogen bonds with G121 and V81, moreover, we also observed a non-specificity of hydrogen bonds with such residues. We suggest that the strong interactions with E177 promote displacement of the ligand core from the deep part of the binding pocket, where are located G121 and V81, respectively, which in turn might explain the lower occurrences of hydrogen bonds with such residues (G121 and V81).

Figura 24 – RMSD analysis of the equilibrium MD trajectory of the RTA–SCF00–332 complex.



In top position is plotted the RMSD considering only backbone atoms of the RTA. In the bottom position, the same analysis, however, considering only heavy atoms of SCF00–332.

Interaction energy calculations showed an amount of 34 interactions; 7 were repulsive (the most prominent was R134), and 27 attractive (the most prominent were Y80, E177, and R180) (Figure 25). We also highlight the π-stacking interaction between Y80 and the double ring moiety of SCF00–332, contributing to the stability of such compound into the binding pocket (Figure 26).

The trajectory analysis indicated that E177 plays a key role in the binding affinity for this compound, followed by a network of hydrogen bonds with V81, G121, and R180. We suggests that such interactions are directly related to the achieved values of $\Delta H_{binding}$,

$\langle F_{max} \rangle$ and $\langle W \rangle$.

Figura 25 – Interaction energy analysis between SCF00–332 and RTA chain over the last 5 ns of MD simulation.

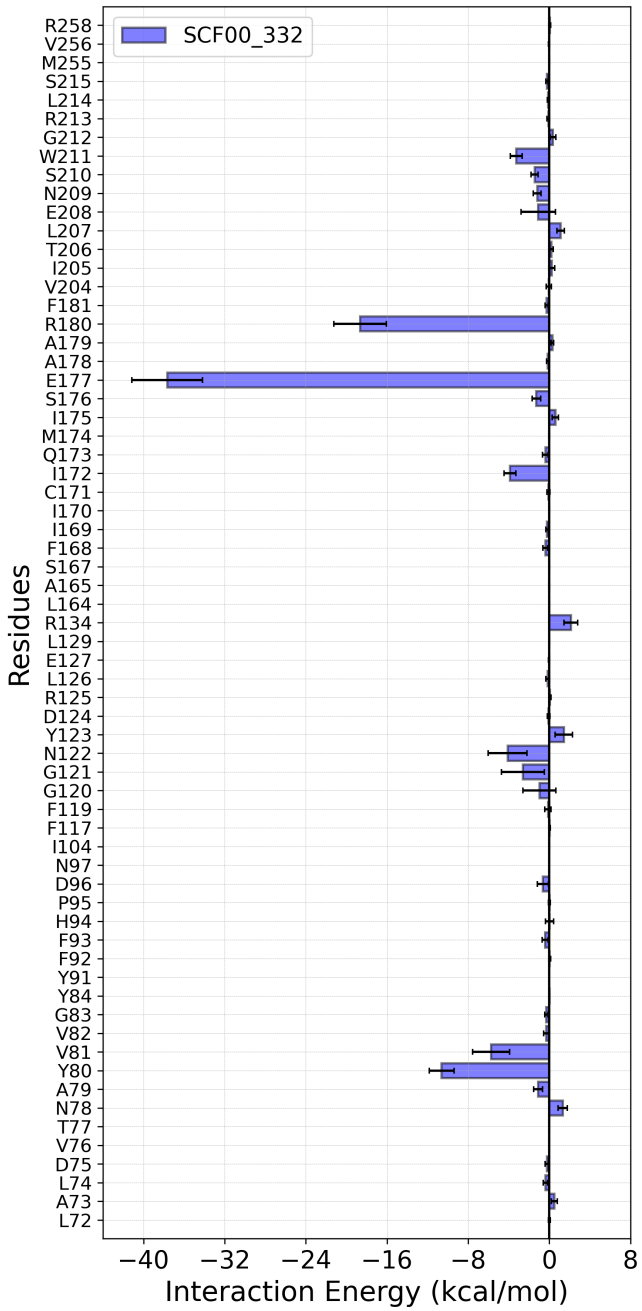
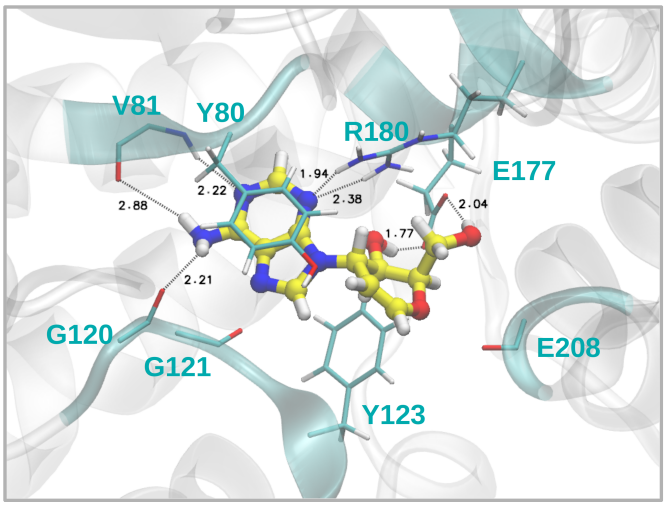


Figura 26 – The binding pose of the SCF00–332 into the RTA binding pocket at the final stage of MD simulation.

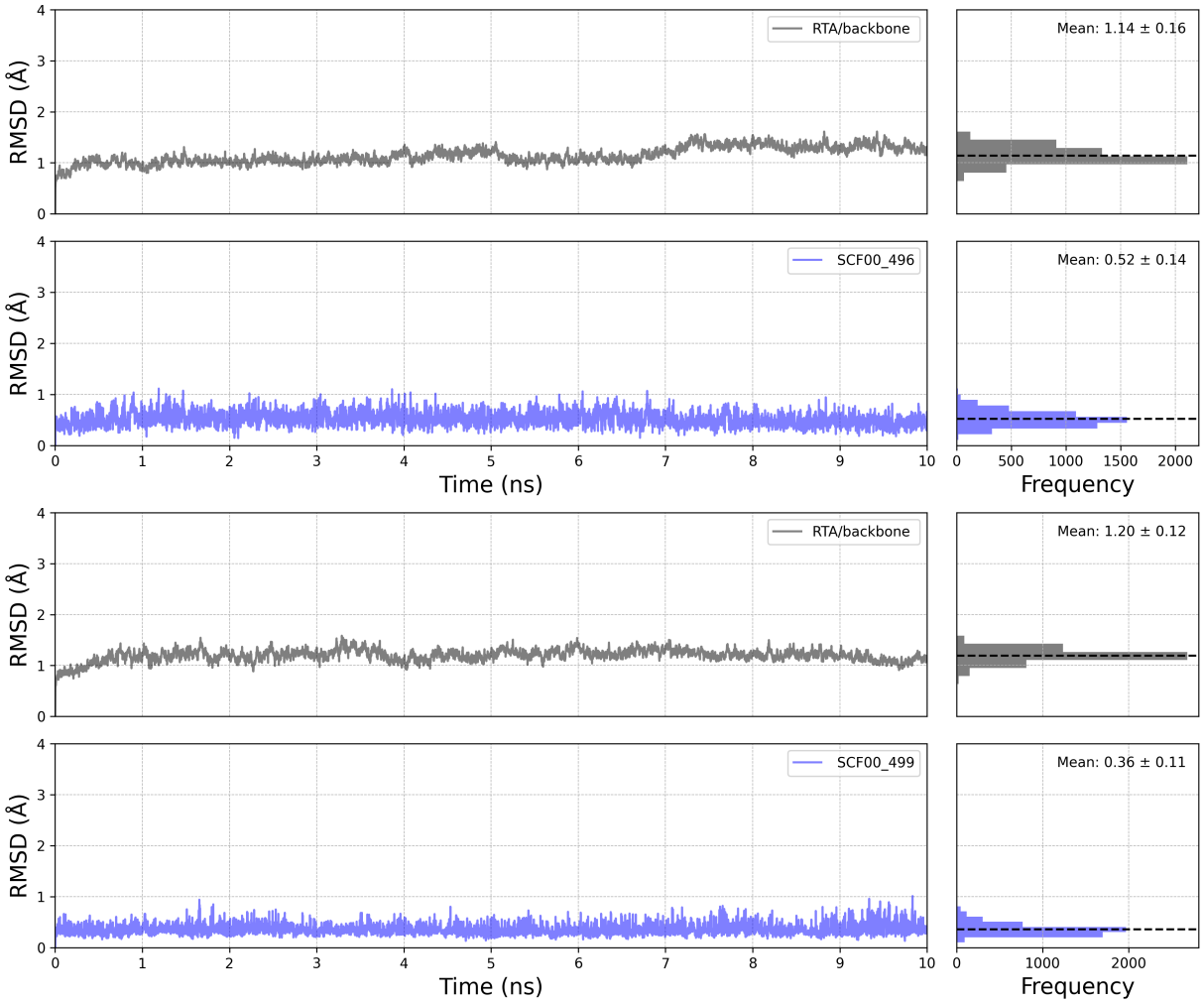


5.1.6.6 Compounds SCF00–496 and SCF00–499

Such compounds (SCF00–496 and –499) also presented a binding pose similar to the natural substrate, adenine, including hydrogen bonds with G121, V81, and R180. In contrast to SCF00–451 and SCF00–332, SCF00–496 showed a mild instability, however, without abrupt changes (Figure 27), and it did not present strong interactions with E177. However, hydrogen bonds with V81, G121, and R180 were detected (Table 7).

Interaction energy calculations showed 35 interactions; 11 were repulsive (the most prominent were R134 and R213), and 26 were attractive (the most prominent were Y80, V81, G121, N122, Y123, and R180) (Figure 28). Additionally, the binding mode along the MD trajectory showed a π-stacking interaction with Y80 (Figure 29).

Figura 27 – RMSD analysis of the equilibrium MD trajectory of RTA–SCF00–496 and –499 complexes.



In top position is plotted the RMSD considering only backbone atoms of the RTA. In the bottom position, the same analysis, however, considering only heavy atoms of SCF00–496 and –499, respectively.

About the SCF00–499, such compound has shown be stable (Figure 27) throughout MD simulation. Interaction energy calculations showed 37 interactions; 11 were repulsive (the most prominent was A79), and 26 attractive (the most prominent were Y80, V81, N122, S176, E177, and R180). Similarly to SCF00–496, SCF00–499 did not show a hydrogen bond with E177, however, interaction energy calculations showed the presence of attractive forces between such residue and SCF00–499.

Some studies suggest that the ligand residence time into the binding pocket is directly proportional to the SMD force profile.[181] Repulsive or unstable interactions increase the binding entropy, which consequently promote an easy egress from the pocket, therefore it is expected that compounds with weak SMD force profiles will show a minor residence time. Thereby, we suggest that the non-covalent interaction with E177 as well the non-specificity of hydrogen bonds with G121 and V81, might be the main cause of

instabilities into the binding pocket of the RTA.

Figura 28 – Interaction energy analysis of the SCF00–499 compound into the binding pocket of the RTA over the last 5 ns of MD simulation.

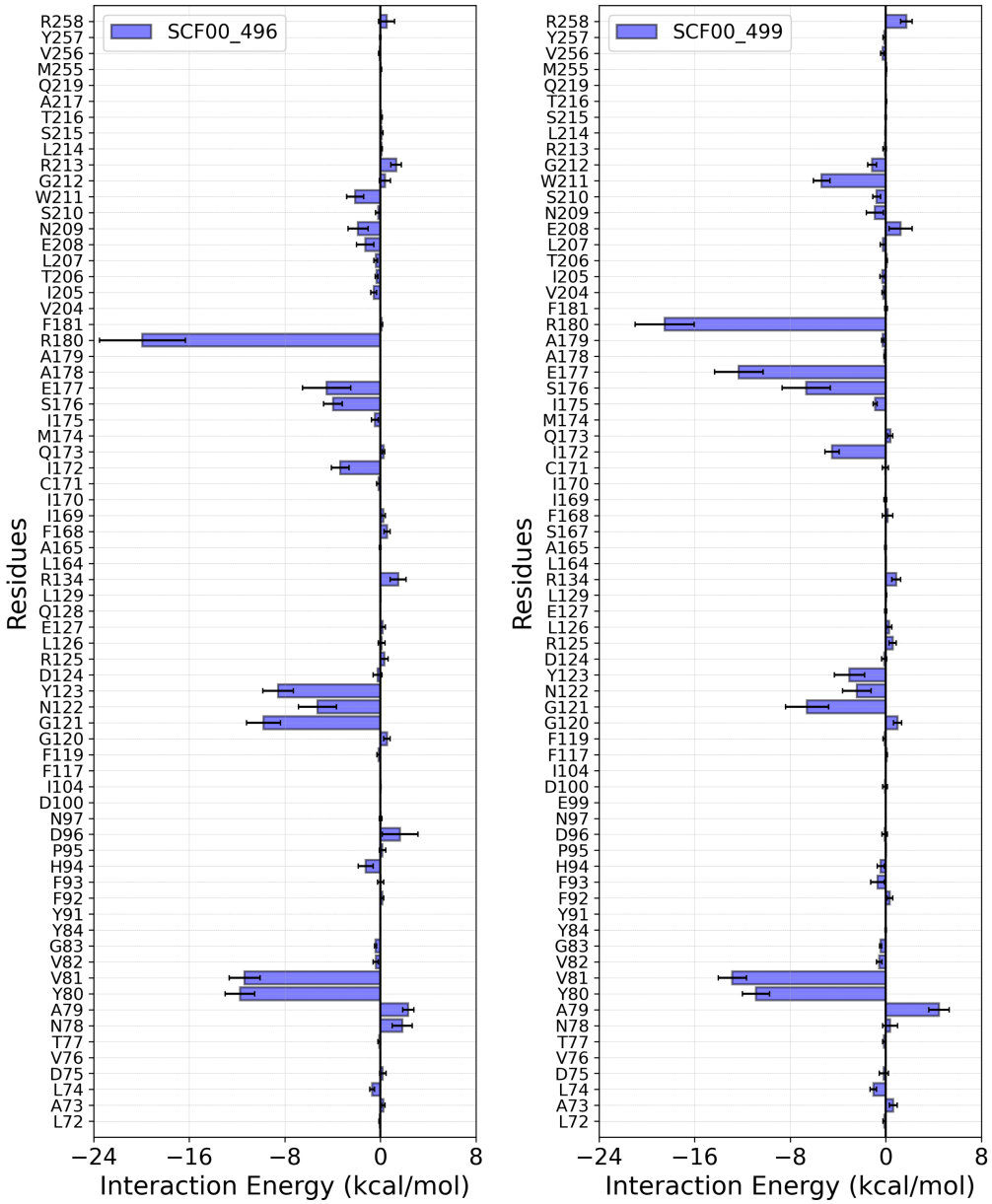
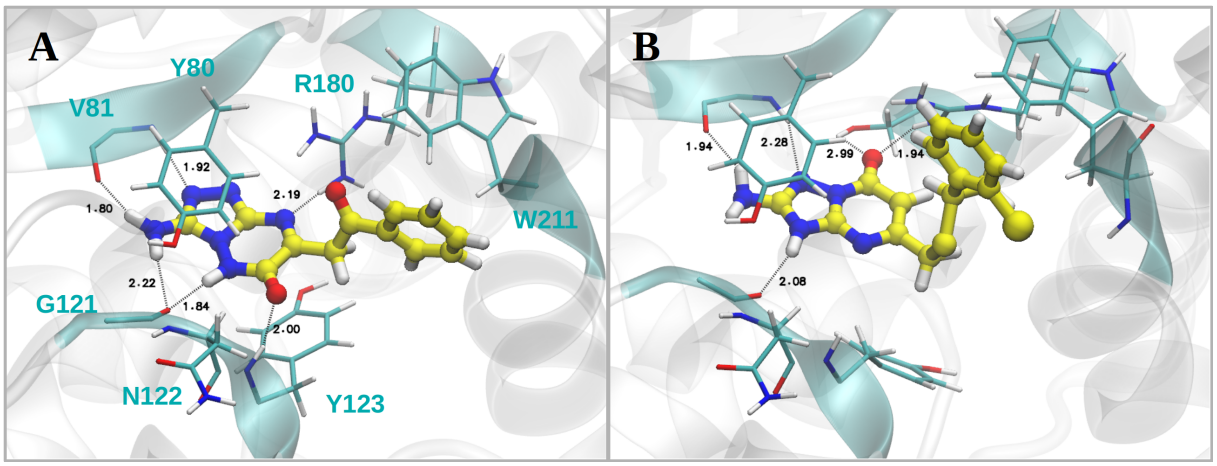


Figura 29 – The binding pose of the SCF00–496 (**A**) and –499 (**B**) complexes in the RTA binding pocket at the final stage of MD simulation.



5.1.7 Conclusion

Finding a selective inhibitor for the catalytic chain of ricin toxin, RTA, has been a hard task. In this work, the initial filtering steps of the VS pipeline were capable to discern the active RTA inhibitors from inactive ones, reaching AUC > 0.9, and then, were used to screen over 2 million compounds from the ZINC database. We identified twenty-nine compounds. Then, as main finding, the MD, SMD and SQM calculations pointed out to SCF00-451, SCF00-332, SCF238, SCF00-499, SCF00-63, SCF00-271 and SCF00-496 compounds as the most promising. Compared to other anti-RTA ligands, the SMD force profiles showed that such compounds, mainly the SCF00-451, with their small number of atoms, can strongly bind into the RTA active site. The SQM calculations corroborate these findings, showing that such compounds have large negative values of binding enthalpy. Our findings are attractive from the perspective of rational drug design for the ricin problem, since these best hits are lead-like compounds, which, in turn, ensuring the possibility of optimizations. Therefore, we believe that such results encourage further experimental evaluation and provide a new beginning for the eradication of ricin poisoning.

5.1.8 Material Suplementar do Artigo

In order to ensure the decrease of false-positive results in our virtual screening protocol, we performed a series of validation steps.

Validation Library. We accessed the ChEMBL database in order to retrieve information about experimental data of active and inactive compounds for the RTA subunit. Using the keyword "Ricin" we found 58 compounds (updated in 2018/09/29), of these, 26 showed IC_{50} lower than 100 μM and were considered as actives and the remaining as inactive ones (Table 8). Moreover, such compounds had the protective ability estimated by the same type of experimental assay, such like proposed by Cerqueira and co-workers.[182] Further, we performed a decoys search into the DUD-E[183] web service (www.dude.docking.org) using the list of active compounds as input, and then we submitted the whole validation library to the multiconformacional search by using OMEGA[184] program (version 3.0.0.1; OpenEye Scientific), that was set to find 50 low energy conformers for each compound.

Validation of the Pharmacophore Modeling Protocol. The alignment of validation library against the shape query was performed by using a GUI version of ROCS program, termed vROCS (version 3.2.2.2; OpenEye Scientific). The statistical analysis of ROC curve was used to evaluate the accuracy of 13 ROCS's score functions, and 6 score methods presented AUC > 0.9 (Figure 30).

Validation of the Molecular Docking Protocol. We carried out a cross-docking study by using GOLD software[185] and its scoring functions (ChemScore, PLP, ASP, and GoldScore) for the following RTA receptors and their ligands: 4HUP[186], 4HUO[186], 4ESI[187], 4HV7[186], 4MX1[188], 3PX8[189], 3PX9[189], 3EJ5[190], 4HV3[186], 3PX9[189], 1IL9[191], 1BR6[192], 1IL4[191], 1BR5[192], and 1IL3[191]. First, all receptor structures were checked and, native ligand and waters were removed. Then, such structures were submitted to PDB2PQR web service[193] to predict residue protonation at pH 7.4, and finally, hydrogen atoms were properly added. Furthermore, receptors were aligned in order to perform the cross-docking study, then, the active site was set by "Point"option, which in turn was defined between residues Y80 and Y123; all residues within 10 Å of this point were considered as active ones. In addition, we used GOLD's score functions in default mode. A summary of the cross-docking steps is depicted in the Figure 31

In the Figure 32 is presented the RMSD values between alignment of the best ranked predict poses and native ones. Here, we conclude that our docking protocol setting with ASP score combined with 1BR5 receptor achieves a good performance to predict native poses, showing a RMSD average of 1.82 Å (Figure 32). Furthermore, we observed that most poses with the high scores did agree with experimental pose of the co-crystallized (Figure 33). In other words, such results indicate that 1BR5 receptor accepts ligands with different features of binding, spanning from small to large ligands, therefore the

combination of ASP score with 1BR5 receptor it is the best scenario to predict bioactive pose of ligands in a virtual screening. Then, it was evaluated ASP score performance to discern active inhibitors from inactive ones. Here, we applied the same molecular docking protocol in the validation library (without conformers). Figure 34 presents the ROC curve of such analysis, which in turn showed a good performance, with AUC > 0.9.

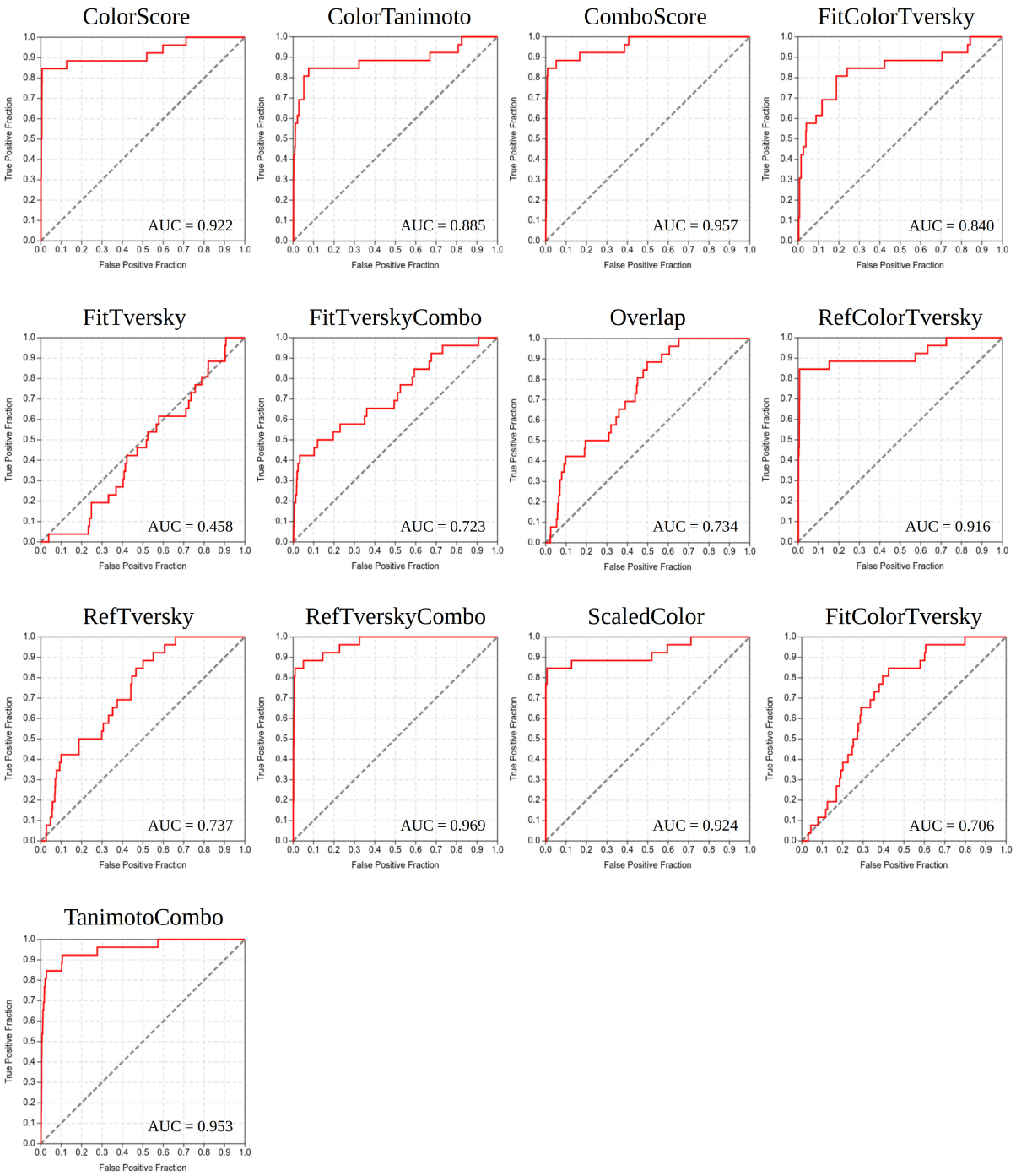
Tabela 8 – List of RTA Inhibitors Retrieved From ChEMBL Database by Using the Keyword "Ricin".

ID	IC_{50} (μM)	**Ref.**	**ID**	IC_{50} (μM)	**Ref.**
2314138	6	[186]	3098908	128	[188]
2089319	15	[187]	2086423	150	[187]
2314130	15	[186]	2089310	150	[187]
2314127	19	[186]	3098906	160	[188]
2314137	20	[186]	2089309	180	[187]
2314140	21	[186]	3098917	209	[188]
2314134	22	[186]	1738820	230	[187]
2314131	25	[186]	1738821	380	[187]
3098919	28	[188]	175367	400	[191]
3098913	29	[188]	2089308	400	[187]
2089314	30	[187]	2089313	400	[187]
2314132	35	[186]	2089316	415	[187]
2314133	35	[186]	2089317	430	[187]
3098915	38	[188]	3098918	449	[188]
3098916	39	[188]	2089312	500	[187]
2314139	40	[186]	2089315	500	[187]
3098911	44	[188]	2089328	500	[187]
3098912	47	[188]	178006	560	[191]
3098914	49	[188]	341824	600	[191]
3098909	52	[188]	367184	600	[191]
2314136	55	[186]	173527	700	[191]
2314135	60	[186]	219568	900	[191]
3098907	68	[188]	367746	1400	[191]
2089311	70	[187]	53653	2000	[191]
3098910	76	[188]	177420	2100	[191]
2314128	90	[186]	175083	2200	[191]
2311565	100	[186]	219544	2800	[191]
2089318	110	[187]	1424	3600	[191]
2314129	115	[186]	-	-	-

Tabela 9 – Summary of the Theoretical Results. All Numerical Data Presented in the 2nd and 3th columns are in pN; 4th and 5th columns are in $kcal \cdot mol^{-1}$.

ID	$\langle F \rangle_{max}$	$\langle F_{max} \rangle$	$\langle W \rangle$	$\Delta H_{binding}$	Rescore
SCF00-451	1208.87 ± 90.96	1518.27 ± 118.10	87.53 ± 13.74	-67.62 ± 4.46	3.00
SCF00-271	1180.54 ± 237.08	1445.41 ± 64.08	78.29 ± 11.47	-59.54 ± 4.76	2.55
SCF00-499	969.36 ± 192.48	1216.64 ± 53.39	71.85 ± 8.49	-65.53 ± 3.76	2.54
SCF00-238	922.66 ± 306.85	1248.16 ± 126.18	78.90 ± 10.61	-62.53 ± 4.59	2.54
SCF00-332	1081.13 ± 149.17	1440.37 ± 137.59	77.36 ± 14.01	-59.21 ± 4.41	2.53
SCF00-496	926.97 ± 40.14	1253.62 ± 106.26	60.81 ± 6.51	-62.41 ± 3.32	2.33
SCF00-63	1082.61 ± 43.72	1327.70 ± 97.73	69.26 ± 10.60	-56.65 ± 4.08	2.27
SCF00-511	738.86 ± 236.77	1068.89 ± 120.01	64.14 ± 10.46	-59.92 ± 5.10	2.16
SCF00-341	893.79 ± 161.77	1168.52 ± 66.16	64.26 ± 6.23	-54.69 ± 6.51	2.04
SCF00-245	900.59 ± 114.27	1182.21 ± 122.65	63.51 ± 7.26	-53.84 ± 3.81	2.01
SCF00-86	952.10 ± 150.77	1256.00 ± 81.86	61.44 ± 9.46	-52.74 ± 5.18	1.99
SCF00-66	762.98 ± 84.39	1081.44 ± 119.40	56.82 ± 9.96	-57.15 ± 7.24	1.98
SCF00-531	836.18 ± 150.33	1041.22 ± 99.87	61.34 ± 5.53	-55.96 ± 6.33	1.97
SCF00-540	636.80 ± 116.51	930.83 ± 34.76	61.70 ± 11.35	-57.40 ± 5.49	1.95
SCF00-194	972.43 ± 91.05	1197.28 ± 95.95	59.17 ± 5.83	-53.02 ± 4.66	1.94
SCF00-252	663.12 ± 221.70	966.63 ± 89.22	61.16 ± 7.92	-56.10 ± 5.61	1.92
SCF00-229	789.42 ± 151.72	1131.35 ± 113.36	67.33 ± 12.50	-51.11 ± 4.86	1.92
SCF00-268	607.95 ± 109.81	941.68 ± 115.96	61.62 ± 8.98	-56.06 ± 4.65	1.91
SCF00-277	928.22 ± 161.57	1140.02 ± 116.48	61.73 ± 5.95	-51.84 ± 5.38	1.89
SCF00-533	883.29 ± 127.19	1132.55 ± 94.30	54.11 ± 7.20	-53.62 ± 5.74	1.86
SCF00-201	758.15 ± 378.78	1144.31 ± 141.28	59.69 ± 4.42	-51.37 ± 3.13	1.85
SCF00-514	876.73 ± 210.60	1087.21 ± 73.88	51.57 ± 6.45	-54,68 ± 6.30	1.84
SCF00-96	756.71 ± 187.07	1003.15 ± 72.59	51.48 ± 3.68	-54.08 ± 5.00	1.76
SCF00-95	715.28 ± 156.10	1010.06 ± 120.83	49.33 ± 4.77	-53.37 ± 3.03	1.72
SCF00-111	850.64 ± 98.61	1046.36 ± 79.63	58.08 ± 10.29	-49.33 ± 5.01	1.70
SCF00-06	768.88 ± 159.86	1116.65 ± 34.78	59.18 ± 4.93	-46.12 ± 6.87	1.64
SCF00-325	860.90 ± 165.54	1166.15 ± 118.36	47.15 ± 8.37	-48.20 ± 4.09	1.61
SCF00-426	739.37 ± 115.59	974.69 ± 67.69	52.44 ± 4.40	-48.29 ± 3.97	1.55
SCF00-15	933.43 ± 158.81	1110.07 ± 60.32	58.08 ± 10.29	-39.72 ± 3.49	1.40
SCF00-176	-	-	-	-39.52 ± 4.74	0.00

Figura 30 – ROC curve analysis employed to each one score function available in the ROCS software.



In the bottom right of every plot is presented the AUC value.

Figura 31 – Illustration showing the main steps performed in a cross-docking study.

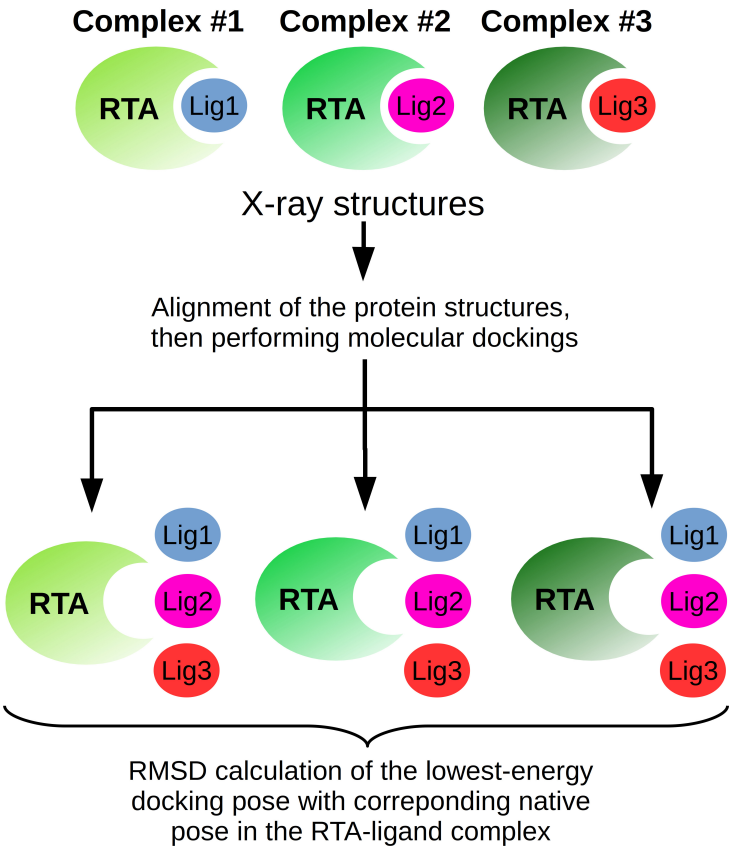


Figura 32 – Each value corresponds to the RMSD of the lowest-energy docking pose for each RTA structure available in the PDB.

A

	4ESI	3EJ5	3PX8	3PX9	1IL9	1BR5	1BR6	4HUO	4MX1-A	4MX1-B	1IL3	1IL4	4HV3	4HV7-A	4HV7-B	4HUP
4ESI	**2.44**	2.41	2.57	2.24	2.30	2.67	2.69	2.65	2.59	2.39	2.19	2.76	2.26	2.59	2.41	1.71
3EJ5	4.61	**4.52**	2.78	3.46	4.29	2.07	2.19	1.61	2.67	2.05	2.04	2.14	3.29	1.93	2.50	4.24
3PX8	0.22	0.22	**0.35**	5.62	0.90	0.57	0.62	0.50	0.29	0.47	0.54	0.33	5.56	0.32	0.25	0.34
3PX9	1.96	2.30	1.96	**3.28**	7.04	0.75	1.18	0.39	2.22	1.61	1.05	3.40	2.89	1.78	2.45	1.86
1IL9	3.45	3.23	3.39	3.71	**2.84**	2.96	4.47	3.30	3.12	4.42	3.20	2.83	3.41	3.00	3.27	3.08
1BR5	0.66	1.58	0.85	3.26	1.29	**0.55**	0.57	1.41	0.80	0.60	1.20	1.08	3.70	0.63	0.61	0.62
1BR6	2.13	0.60	4.85	4.64	1.87	1.88	**1.88**	1.40	2.04	1.88	2.41	1.73	2.60	1.55	2.16	1.30
4HUO	2.12	2.03	2.00	1.87	1.92	0.76	1.36	**1.01**	2.06	5.28	2.24	5.60	2.40	5.54	1.83	0.51
4MX1	1.05	5.85	1.69	1.81	4.45	0.40	7.05	0.44	**1.32**	**1.10**	5.51	0.84	2.55	0.82	1.17	6.63
1IL3	2.87	2.71	2.97	2.94	0.84	2.72	2.63	2.65	3.03	2.80	**0.26**	0.37	4.20	2.59	2.69	2.68
1IL4	3.90	2.62	3.94	3.83	0.75	2.64	3.68	3.57	2.63	3.79	0.45	**0.38**	4.02	2.64	3.54	3.81
4HV3	1.97	2.10	2.62	1.98	2.09	2.10	2.99	2.79	2.62	2.12	2.13	2.19	**1.97**	3.39	2.68	3.10
4HV7	3.14	3.10	2.26	1.99	2.90	1.60	5.65	3.20	1.82	5.86	6.48	6.00	1.56	**7.16**	**1.90**	1.96
4HUP	4.39	5.12	4.35	3.10	7.47	3.86	4.53	4.53	3.55	7.22	5.14	4.91	5.13	3.70	7.89	**2.86**
Average	2.49	2.74	2.61	3.12	2.93	1.82	2.96	2.10	2.20	2.97	2.49	2.47	3.25	2.69	2.53	2.48
Standard Deviation	1.33	1.58	1.26	1.12	2.16	1.09	1.92	1.31	0.90	2.05	1.96	1.92	1.17	1.88	1.80	1.70

B

	4ESI	3EJ5	3PX8	3PX9	1IL9	1BR5	1BR6	4HUO	4MX1-A	4MX1-B	1IL3	1IL4	4HV3	4HV7-A	4HV7-B	4HUP
4ESI	**1.57**	7.39	1.39	1.32	2.51	2.52	1.30	1.04	2.69	1.67	2.67	2.67	1.58	1.99	1.54	1.66
3EJ5	7.17	**4.32**	7.11	7.15	4.03	4.44	2.11	4.35	7.15	7.28	4.43	4.36	7.24	7.13	7.20	7.19
3PX8	1.01	0.31	**0.26**	5.57	0.92	0.30	0.35	0.30	0.26	0.38	0.46	0.33	0.53	0.35	0.34	0.24
3PX9	3.02	9.98	3.61	**9.95**	7.85	2.09	7.48	7.17	7.85	6.67	7.17	7.76	3.36	7.77	3.43	3.77
1IL9	4.71	4.93	4.27	5.02	**0.23**	3.00	4.68	4.76	4.37	4.03	4.09	4.20	4.22	4.77	3.89	4.77
1BR5	1.58	0.95	1.01	3.61	5.91	**1.54**	3.00	2.92	0.96	2.81	1.47	2.96	0.65	6.91	1.79	3.01
1BR6	7.86	8.14	6.59	8.09	6.79	4.89	**1.22**	1.00	4.95	7.88	1.30	6.59	4.54	7.93	4.70	7.91
4HUO	3.19	5.47	4.26	7.05	7.57	3.85	0.65	**5.01**	2.39	3.83	4.15	6.43	2.08	4.95	3.52	3.26
4MX1	6.62	9.03	6.89	8.21	6.83	9.29	9.07	9.02	**7.43**	**9.08**	8.93	9.18	8.74	9.03	7.44	9.37
1IL3	4.51	4.33	4.46	4.23	0.75	2.75	4.16	4.55	4.51	4.50	**0.40**	0.33	4.31	4.33	4.27	4.54
1IL4	3.66	3.63	3.67	3.63	0.84	3.45	3.42	3.64	3.69	3.78	0.59	**0.42**	3.61	3.39	3.63	3.65
4HV3	2.63	1.80	1.69	2.37	2.23	3.22	1.95	1.69	2.72	2.75	2.12	1.84	**2.74**	1.71	2.72	3.17
4HV7	4.49	6.15	2.61	4.52	6.00	8.14	2.87	1.47	3.88	2.94	3.20	8.89	6.77	**2.53**	**7.59**	2.26
4HUP	6.34	3.30	8.15	10.63	7.56	5.39	2.58	8.25	3.47	8.04	9.82	8.25	3.51	8.25	2.71	**3.99**
Average	4.17	4.98	4.00	5.81	4.29	3.92	3.20	3.94	4.02	4.69	3.63	4.59	3.85	5.08	3.91	4.20
Standard Deviation	2.19	2.94	2.47	2.79	2.94	2.43	2.50	2.77	2.28	2.65	3.08	3.26	2.42	2.80	2.21	2.48

C

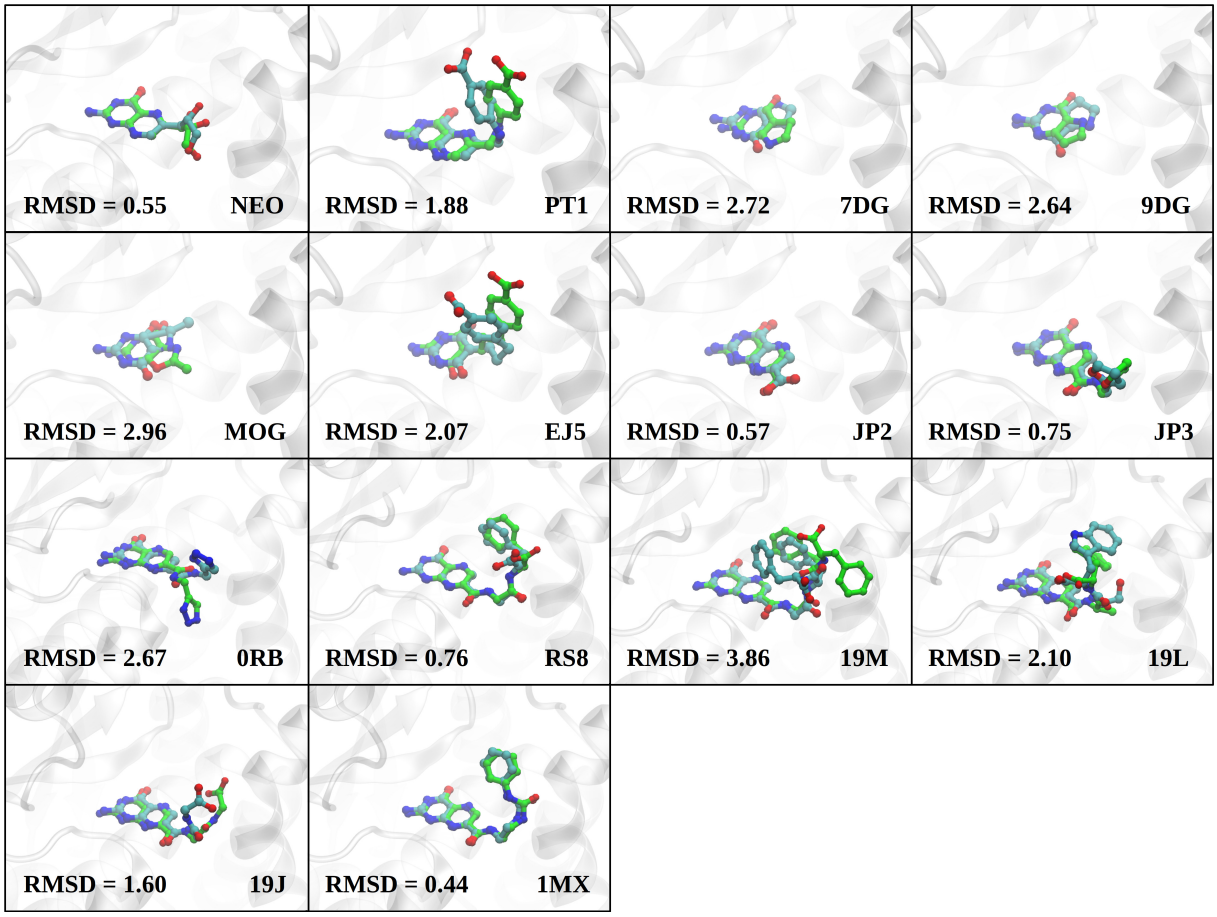
	4ESI	3EJ5	3PX8	3PX9	1IL9	1BR5	1BR6	4HUO	4MX1-A	4MX1-B	1IL3	1IL4	4HV3	4HV7-A	4HV7-B	4HUP
4ESI	**7.40**	1.40	2.15	8.33	7.80	2.16	5.89	2.19	7.71	2.13	7.26	8.67	0.73	2.23	1.90	1.69
3EJ5	7.27	**2.12**	7.18	7.65	7.59	1.84	2.79	2.67	4.15	2.05	4.38	7.55	6.22	2.04	6.12	4.67
3PX8	0.28	0.39	**0.27**	9.26	5.75	0.38	0.38	0.42	0.31	0.28	5.08	5.06	0.34	0.24	0.24	0.16
3PX9	7.83	7.24	7.67	**7.99**	8.00	8.16	8.08	7.70	8.14	7.81	7.41	7.42	8.14	7.76	7.95	7.94
1IL9	8.31	2.89	4.98	9.01	**8.84**	7.56	4.84	7.37	7.94	4.92	5.65	4.66	10.38	7.41	7.74	4.91
1BR5	8.72	6.03	3.16	8.71	5.83	**7.37**	6.70	6.11	8.25	6.17	6.31	5.86	9.30	5.68	7.10	6.73
1BR6	7.24	7.57	7.25	7.36	7.28	7.35	**7.33**	7.73	7.99	7.80	7.24	7.35	2.59	7.96	7.38	4.99
4HUO	10.65	1.56	1.84	10.45	7.35	1.00	8.12	**7.21**	8.17	7.72	10.79	9.31	10.35	7.79	8.07	8.78
4MX1	7.64	8.49	9.39	7.60	9.08	8.90	8.35	9.31	**10.55**	**7.14**	8.24	8.42	8.78	9.41	7.34	8.92
1IL3	4.43	4.56	4.92	4.33	0.69	3.87	4.21	4.78	4.45	4.52	**0.28**	0.27	4.76	4.43	4.83	4.28
1IL4	3.77	3.73	3.77	3.80	0.79	3.40	3.75	3.69	3.77	3.83	3.42	**4.44**	8.53	3.71	3.75	3.74
4HV3	1.86	1.56	2.58	10.22	8.76	1.96	1.66	1.64	1.55	2.92	6.63	8.83	**2.45**	1.60	1.58	1.56
4HV7	8.99	8.83	3.25	8.33	9.49	9.12	8.40	8.80	1.86	7.94	8.47	7.95	8.48	**1.89**	**9.08**	9.65
4HUP	8.11	8.48	8.39	3.56	7.78	9.39	10.20	7.33	6.85	8.39	8.75	8.49	9.03	8.10	8.09	**8.57**
Average	6.61	4.63	4.77	7.62	6.79	5.17	5.76	5.50	5.84	5.26	6.42	6.73	6.43	5.02	5.80	5.47
Standard Deviation	2.92	3.06	2.79	2.21	2.78	3.36	2.90	2.89	3.12	2.68	2.60	2.46	3.57	3.06	2.85	3.05

D

	4ESI	3EJ5	3PX8	3PX9	1IL9	1BR5	1BR6	4HUO	4MX1-A	4MX1-B	1IL3	1IL4	4HV3	4HV7-A	4HV7-B	4HUP
4ESI	**2.10**	2.09	1.50	1.38	2.37	2.70	2.44	1.87	2.13	2.08	2.54	2.62	1.54	2.57	2.13	1.89
3EJ5	3.29	**3.25**	3.50	5.90	11.09	2.23	2.34	4.31	3.50	2.28	2.33	6.73	3.58	2.19	3.47	4.47
3PX8	0.37	0.60	**0.67**	0.61	0.65	0.42	1.03	0.87	0.85	0.94	0.74	0.69	0.68	0.85	0.76	0.84
3PX9	7.45	2.57	7.39	**7.31**	10.71	1.43	1.83	1.81	7.39	2.33	3.19	2.89	7.40	1.91	7.29	1.81
1IL9	3.80	3.71	3.99	4.01	**3.84**	3.97	4.25	4.15	3.74	4.16	3.89	3.90	3.83	3.21	3.83	2.80
1BR5	1.49	1.92	2.16	1.54	11.60	**1.74**	1.68	1.04	0.99	1.39	1.32	5.41	1.39	1.40	1.40	1.41
1BR6	3.94	1.55	6.60	6.47	10.70	1.38	**1.78**	1.53	6.44	1.31	3.72	3.55	6.39	1.25	6.54	1.33
4HUO	7.82	1.87	8.69	8.62	7.88	0.81	1.51	**2.92**	7.76	1.40	7.41	8.89	1.66	6.88	7.68	0.67
4MX1	8.13	6.34	8.14	7.71	10.23	3.72	3.64	7.52	**8.25**	**6.50**	6.21	4.05	8.37	0.93	8.13	7.36
1IL3	2.77	2.70	2.78	2.71	3.37	2.76	2.77	2.74	2.78	2.80	**0.65**	0.60	2.66	2.71	2.77	2.81
1IL4	3.56	3.53	3.62	3.62	3.38	3.53	3.50	3.49	3.61	3.58	0.53	**0.30**	3.50	3.51	3.55	3.51
4HV3	3.60	3.39	3.35	3.72	3.12	3.78	2.22	2.30	2.16	2.08	3.45	3.77	**2.15**	2.07	6.03	3.64
4HV7	2.96	7.42	5.17	2.24	6.85	3.27	2.70	4.04	7.90	3.46	6.65	9.25	5.01	**3.44**	**2.47**	3.37
4HUP	8.59	7.59	9.00	8.42	8.55	7.05	4.12	3.88	5.44	3.93	6.15	6.59	8.24	6.45	5.24	**6.64**
Average	4.28	3.47	4.75	4.59	6.74	2.77	2.56	3.03	4.49	2.73	3.48	4.23	4.03	2.81	4.38	3.04
Standard Deviation	2.63	2.17	2.76	2.77	3.83	1.69	1.00	1.74	2.65	1.49	2.34	2.86	2.64	1.85	2.43	2.03

ASP score (**A**); ChemPL (**B**); Chemscore (**C**); and Goldscore (**D**).

Figura 33 – Predicted poses using the ASP score and 1BR5 structure. Native and predicted poses are depicted in cyan and green colors, respectively.



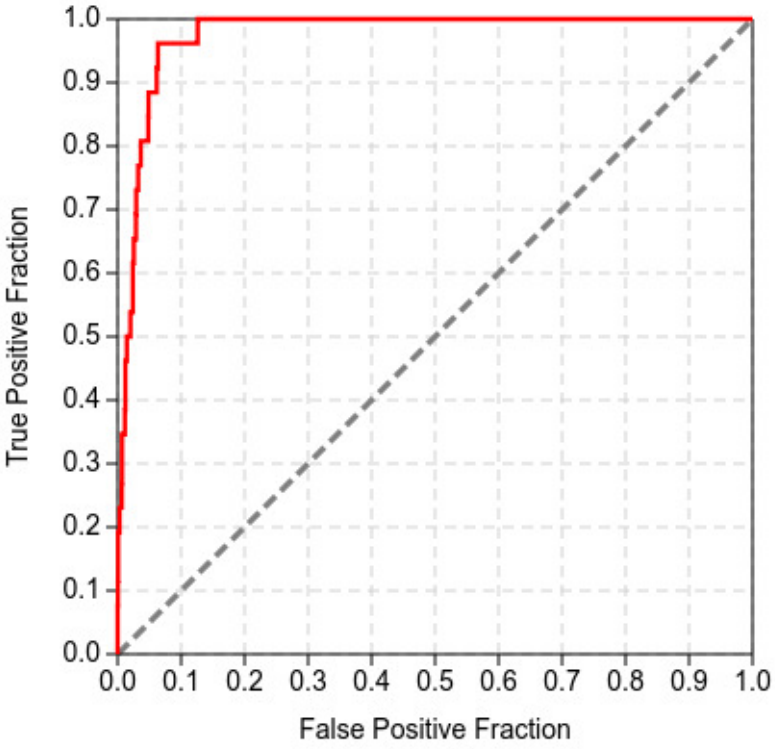


Figura 34 – ROC curve analysis when ASP score was used.

Figura 35 – Trajectory analysis of the RTA–SCF00-06 complex.

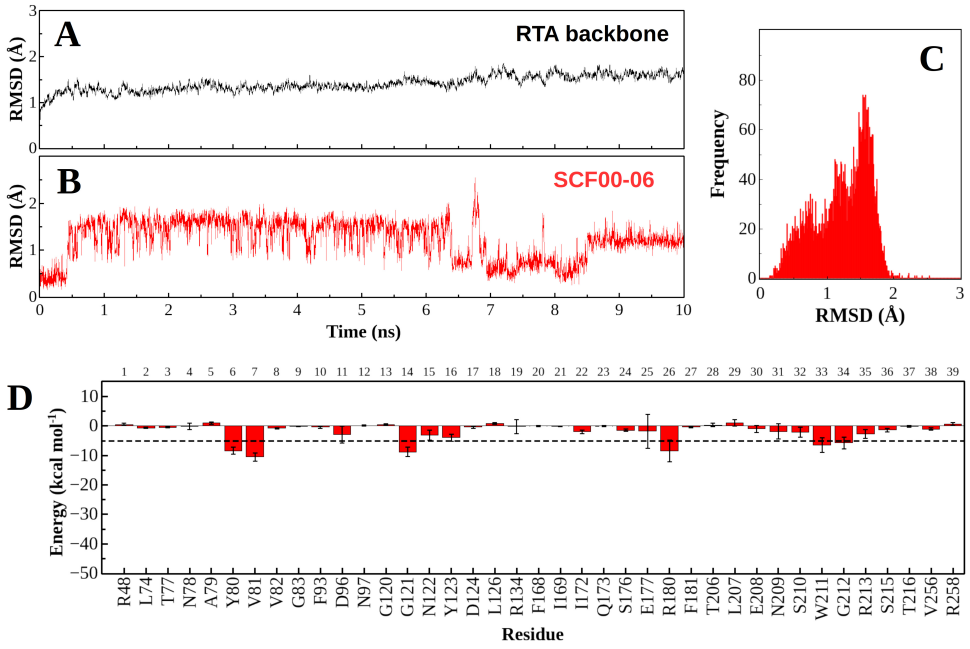


Figura 36 – Trajectory analysis of the RTA–SCF00-15 complex.

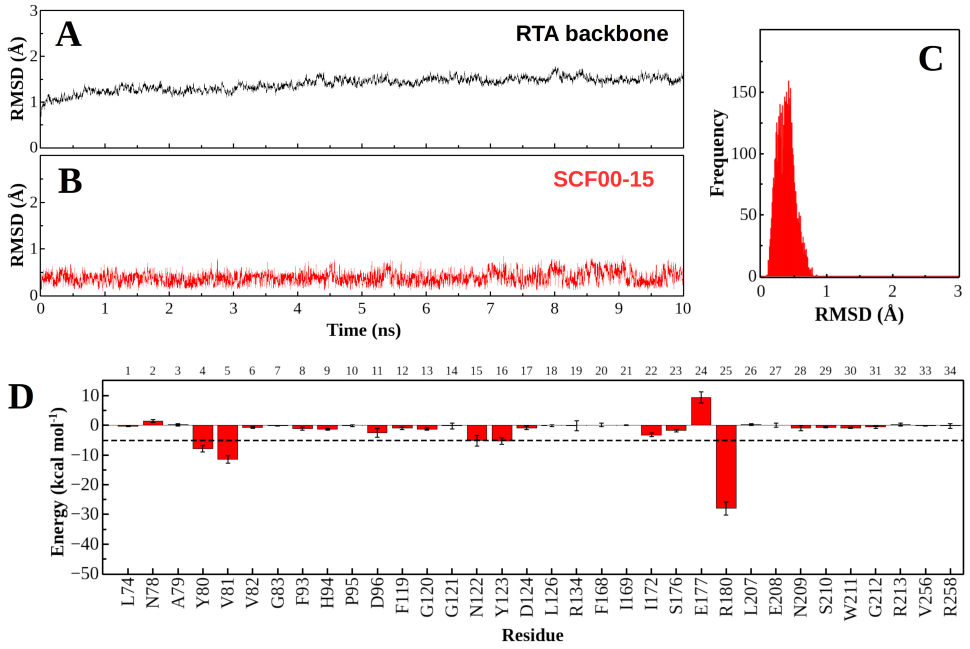


Figura 37 – Trajectory analysis of the RTA–SCF00-63 complex.

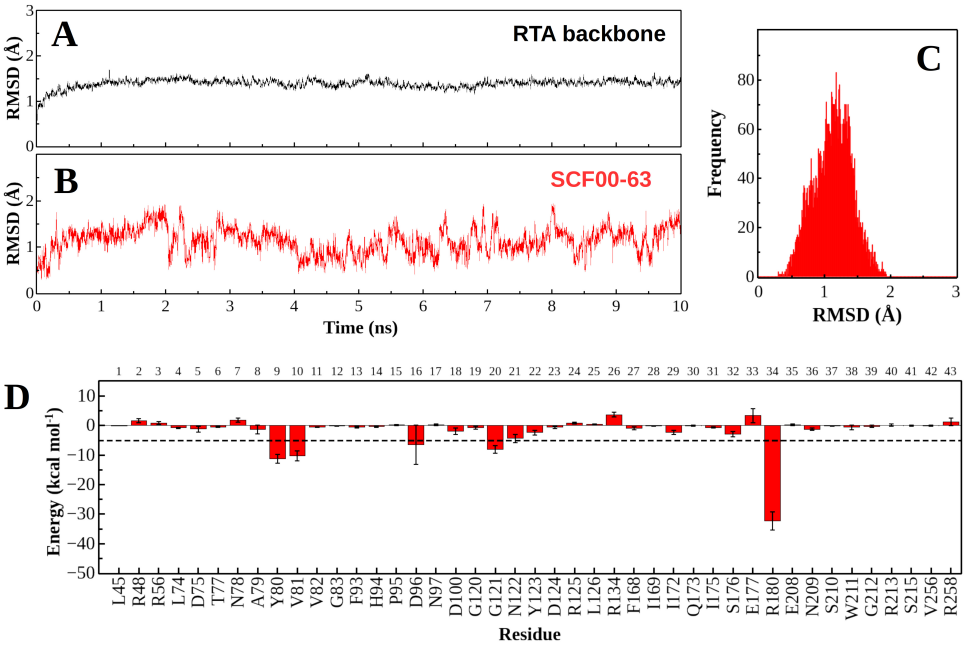


Figura 38 – Trajectory analysis of the RTA–SCF00-66 complex.

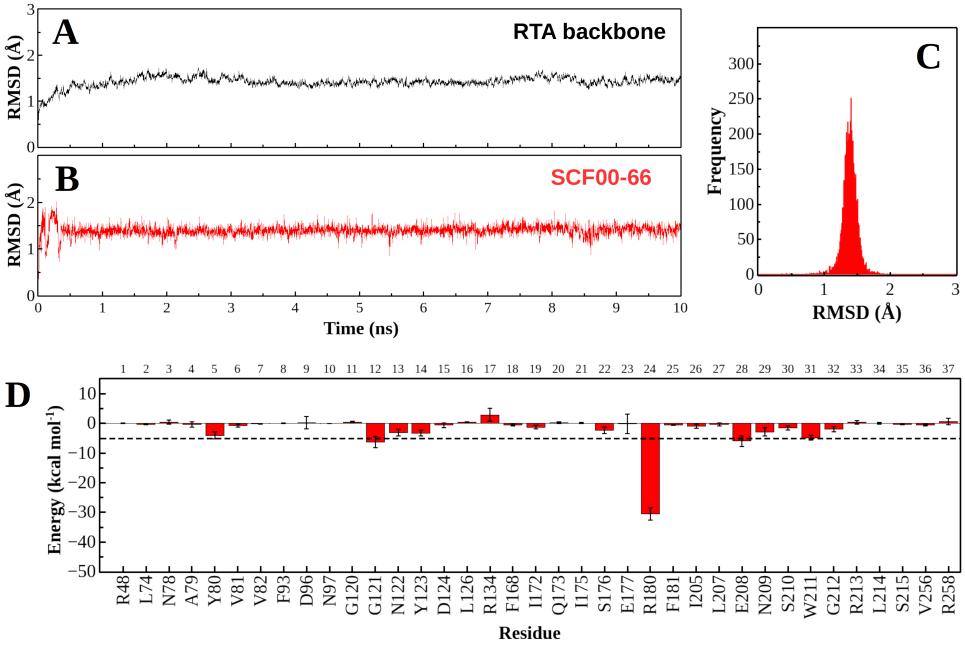


Figura 39 – Trajectory analysis of the RTA–SCF00-86 complex.

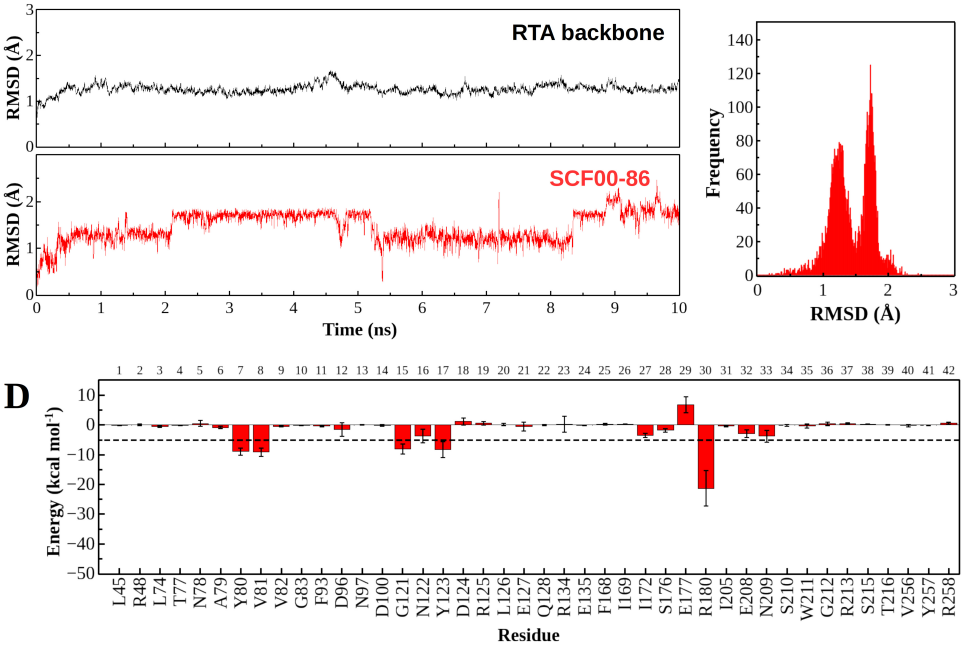


Figura 40 – Trajectory analysis of the RTA–SCF00-95 complex.

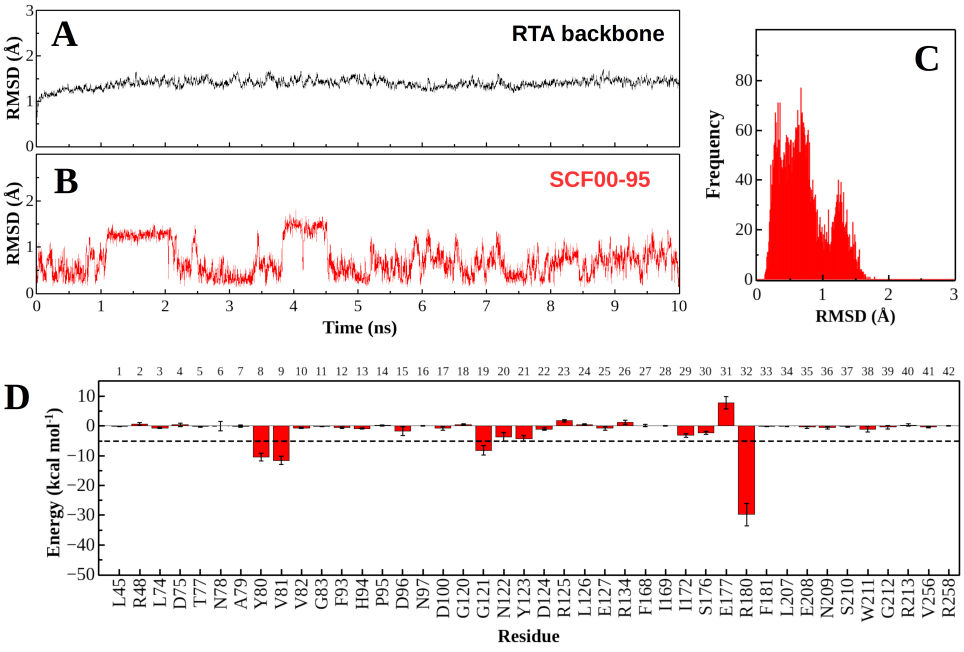


Figura 41 – Trajectory analysis of the RTA–SCF00-96 complex.

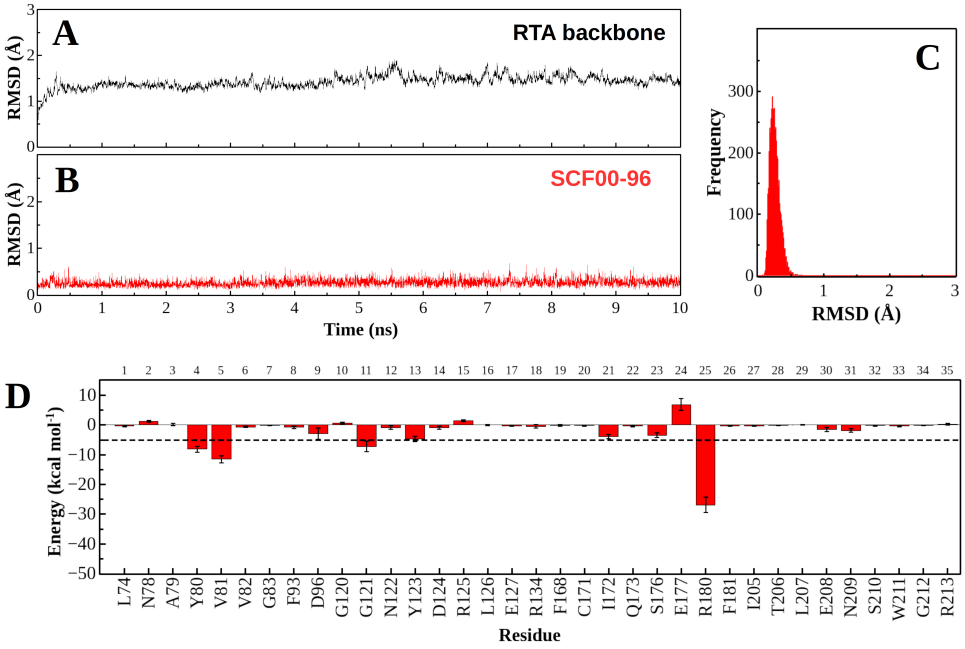


Figura 42 – Trajectory analysis of the RTA–SCF00-111 complex.

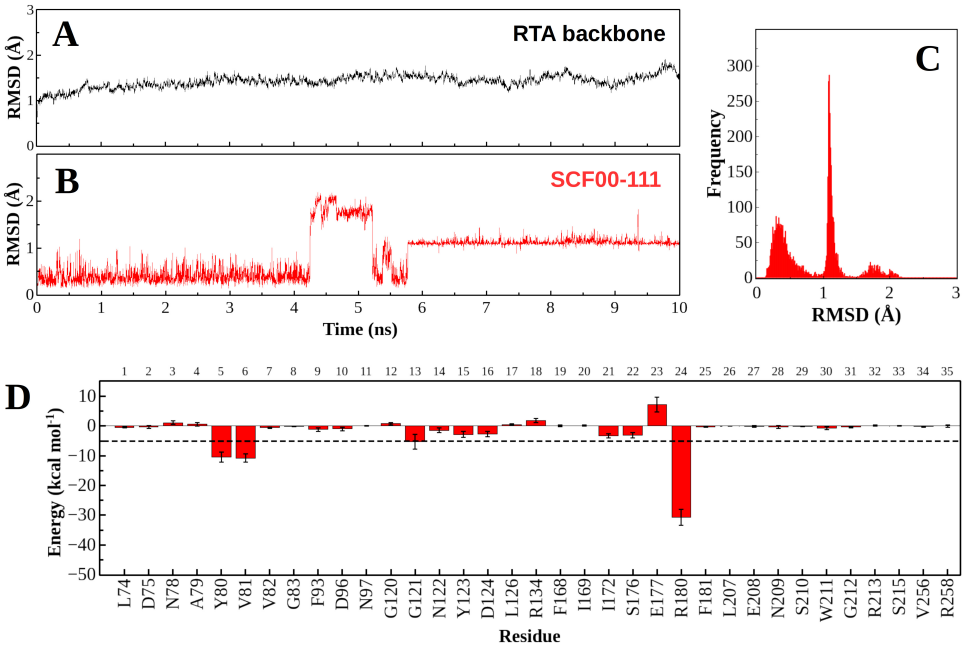


Figura 43 – Trajectory analysis of the RTA–SCF00-194 complex.

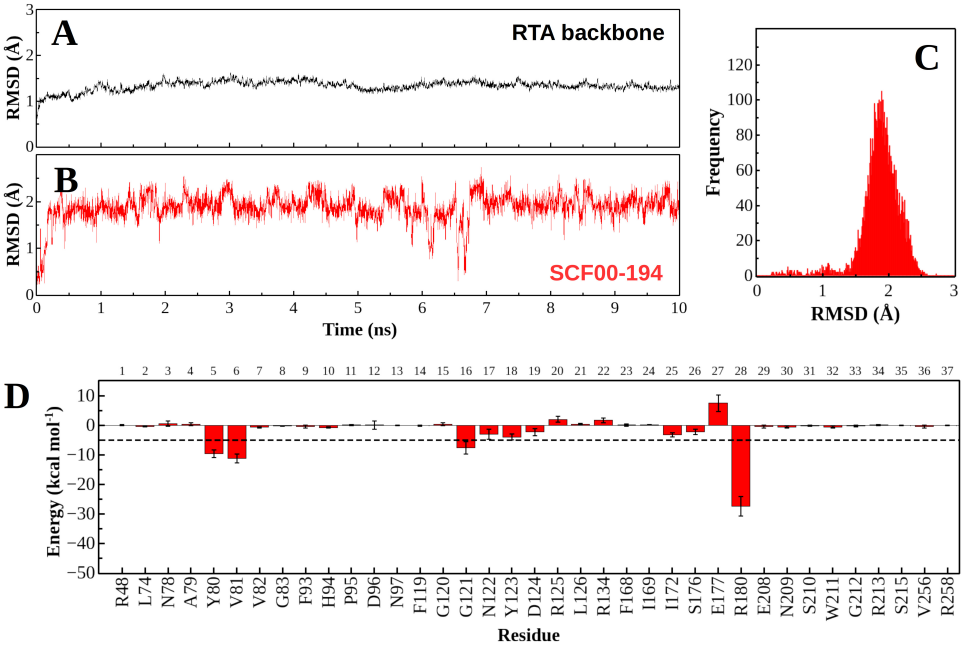


Figura 44 – Trajectory analysis of the RTA–SCF00-194 complex.

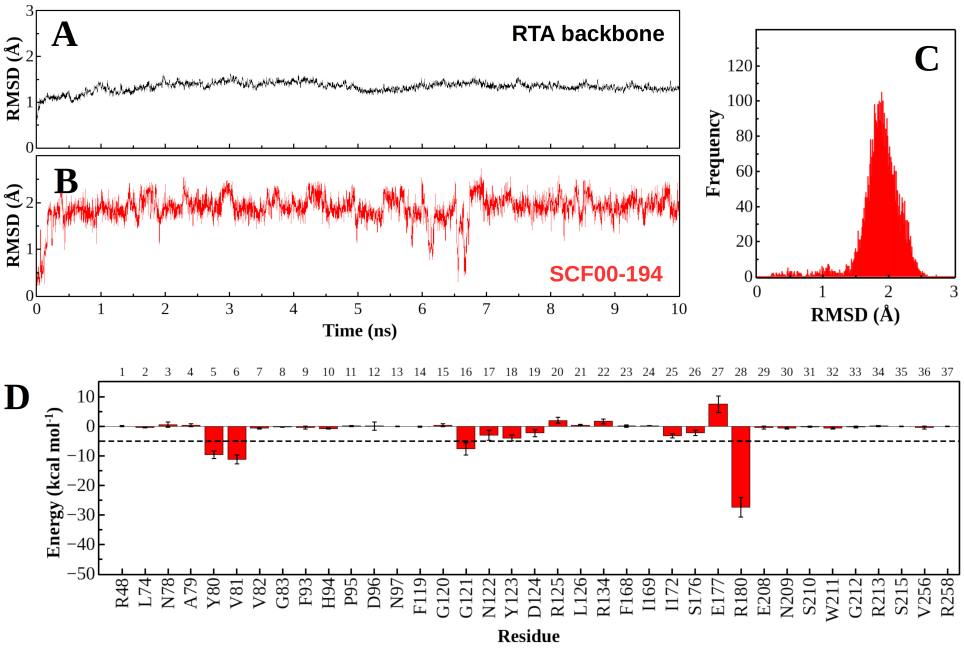


Figura 45 – Trajectory analysis of the RTA–SCF00-201 complex.

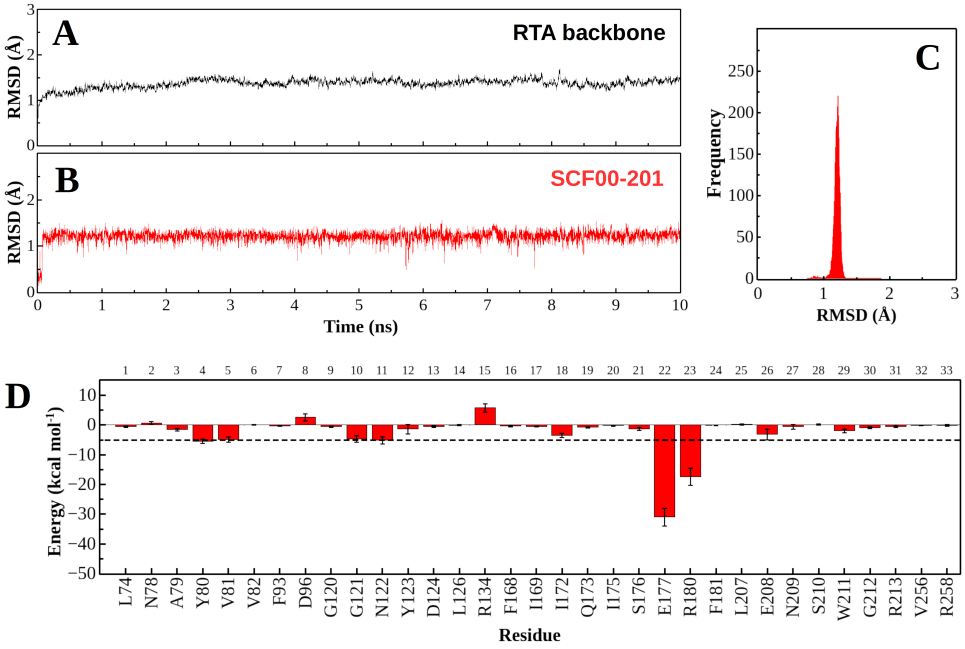


Figura 46 – Trajectory analysis of the RTA–SCF00-229 complex.

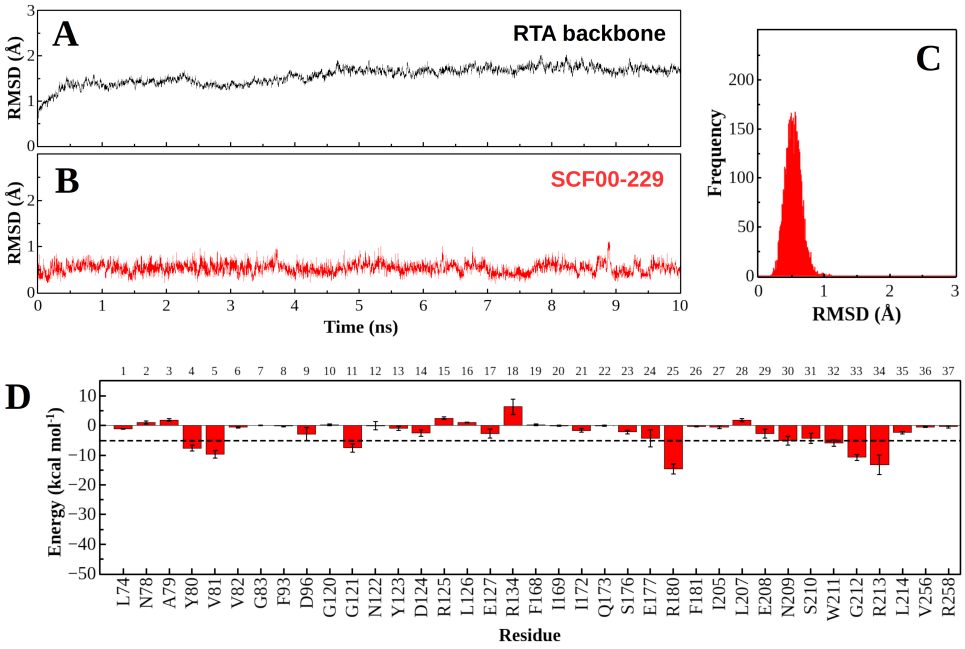


Figura 47 – Trajectory analysis of the RTA–SCF00-238 complex.

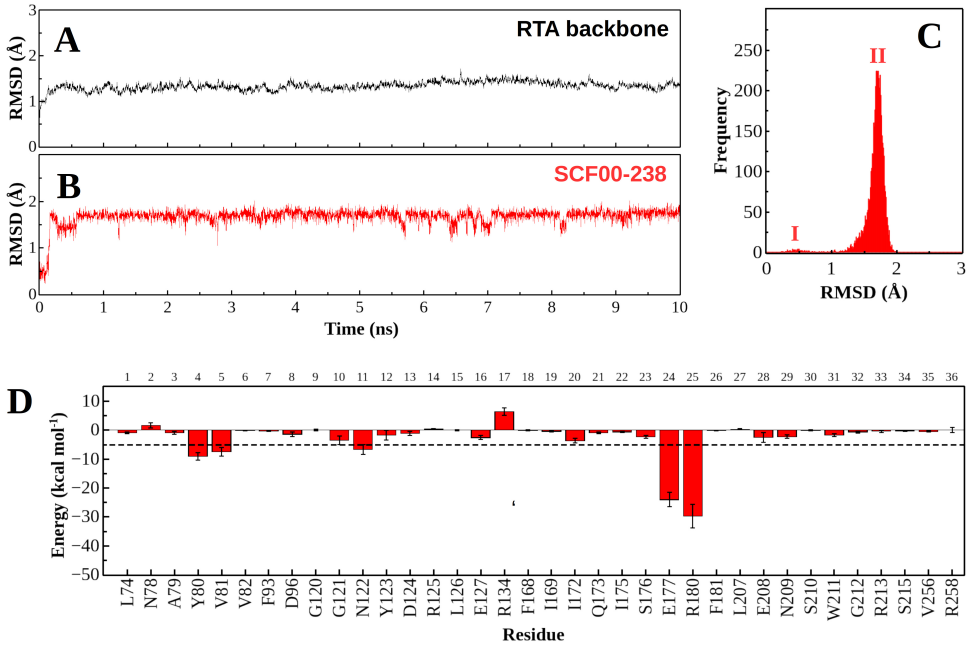


Figura 48 – Trajectory analysis of the RTA–SCF00-245 complex.

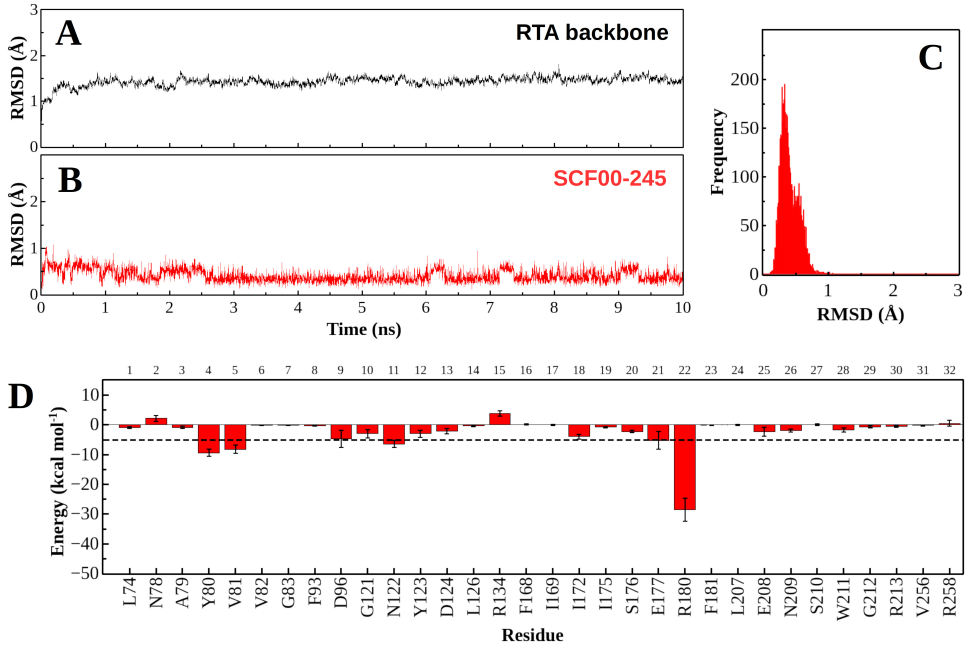


Figura 49 – Trajectory analysis of the RTA–SCF00-252 complex.

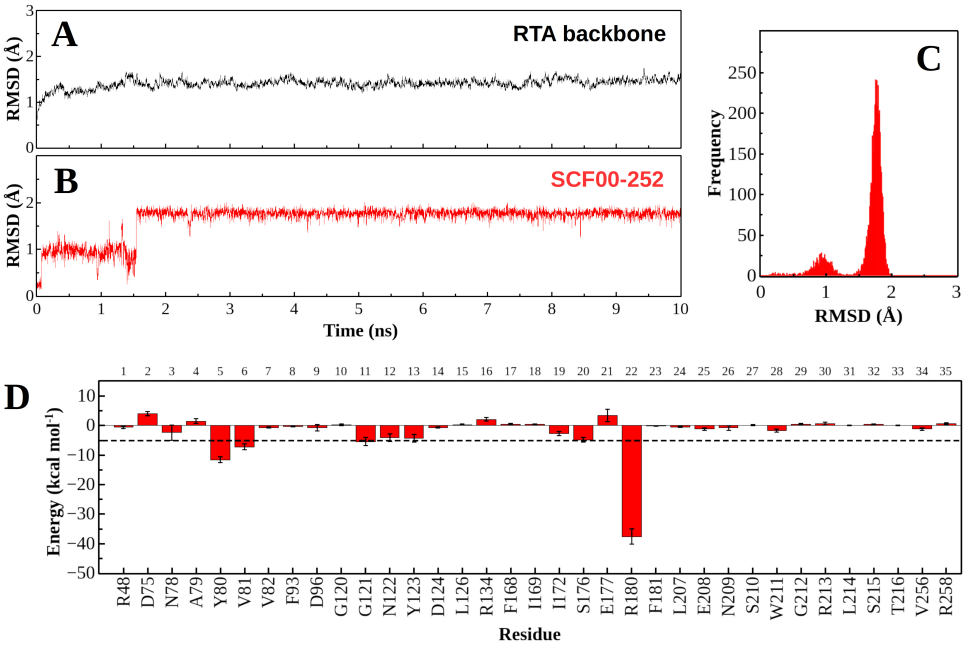


Figura 50 – Trajectory analysis of the RTA–SCF00-268 complex.

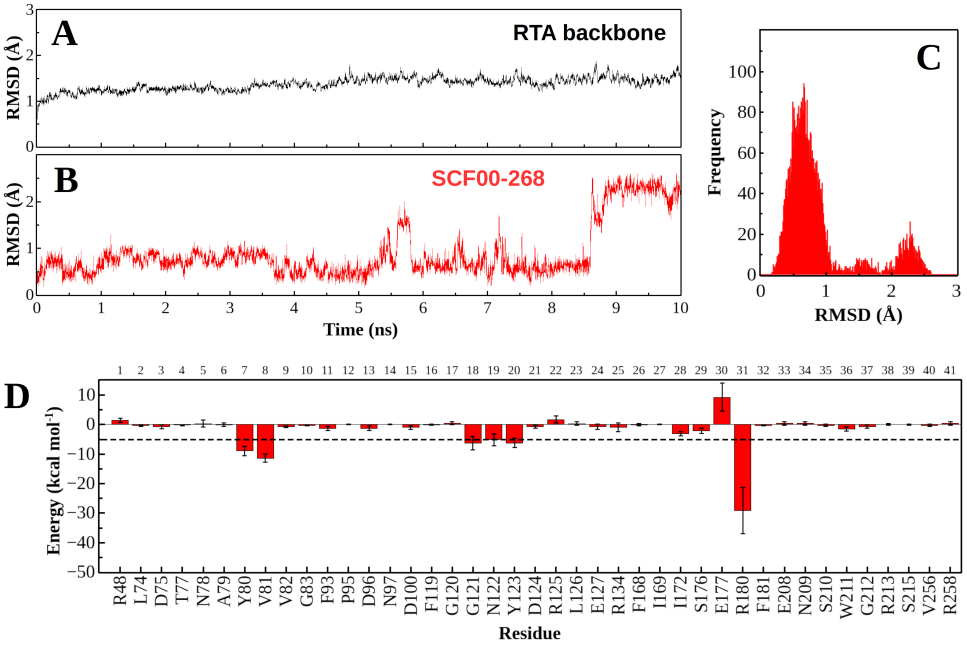


Figura 51 – Trajectory analysis of the RTA–SCF00-271 complex.

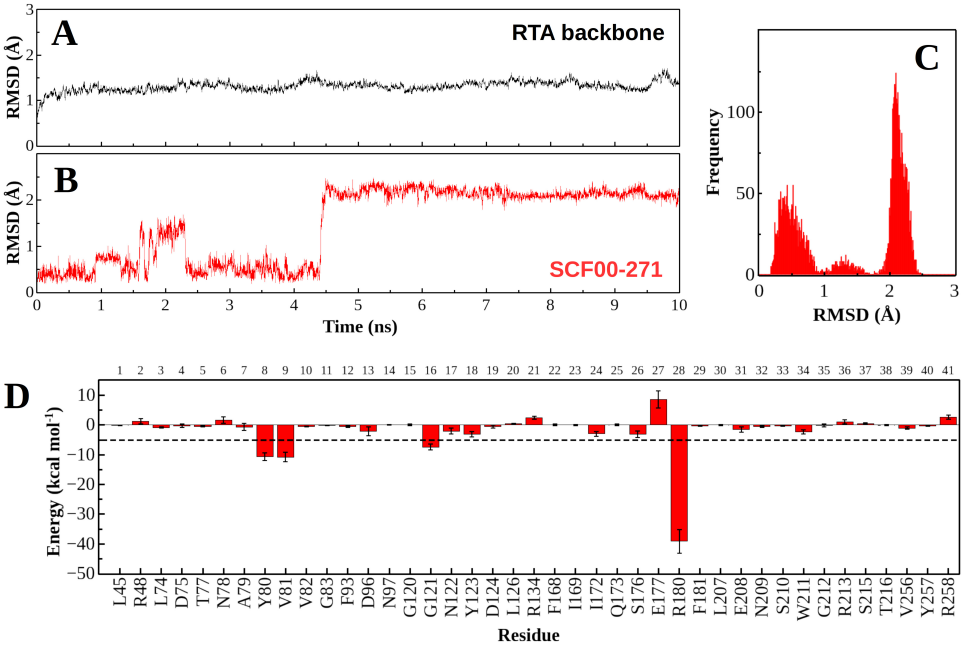


Figura 52 – Trajectory analysis of the RTA–SCF00-277 complex.

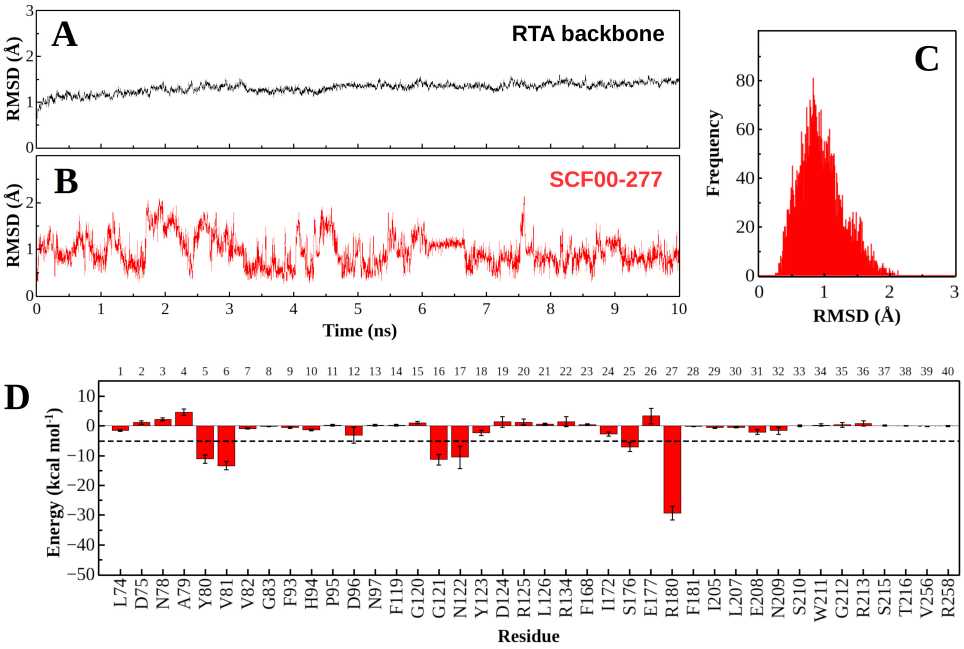


Figura 53 – Trajectory analysis of the RTA–SCF00-325 complex.

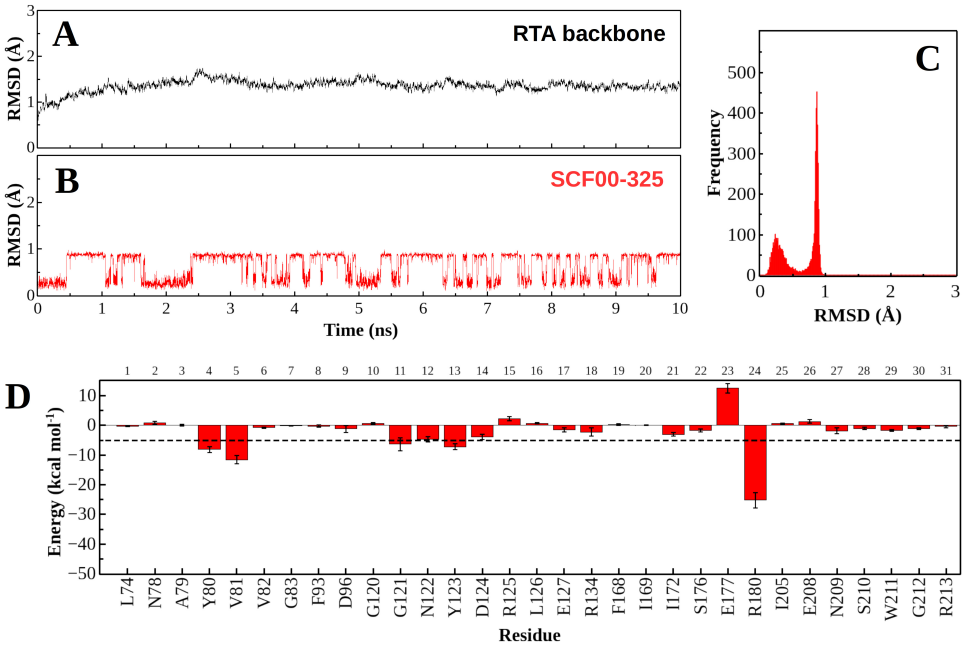


Figura 54 – Trajectory analysis of the RTA–SCF00-341 complex.

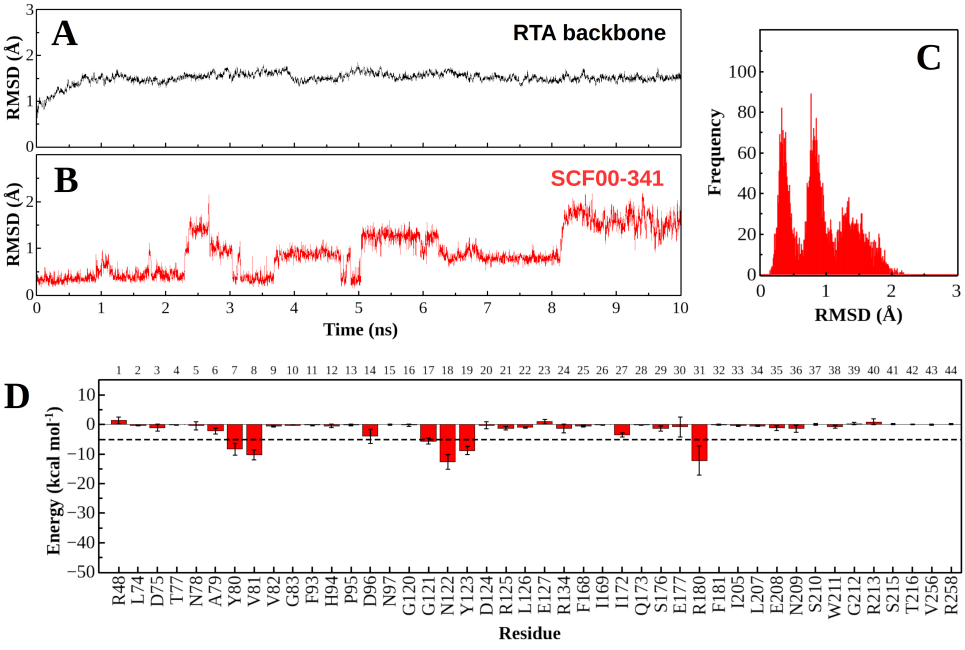


Figura 55 – Trajectory analysis of the RTA–SCF00-341 complex.

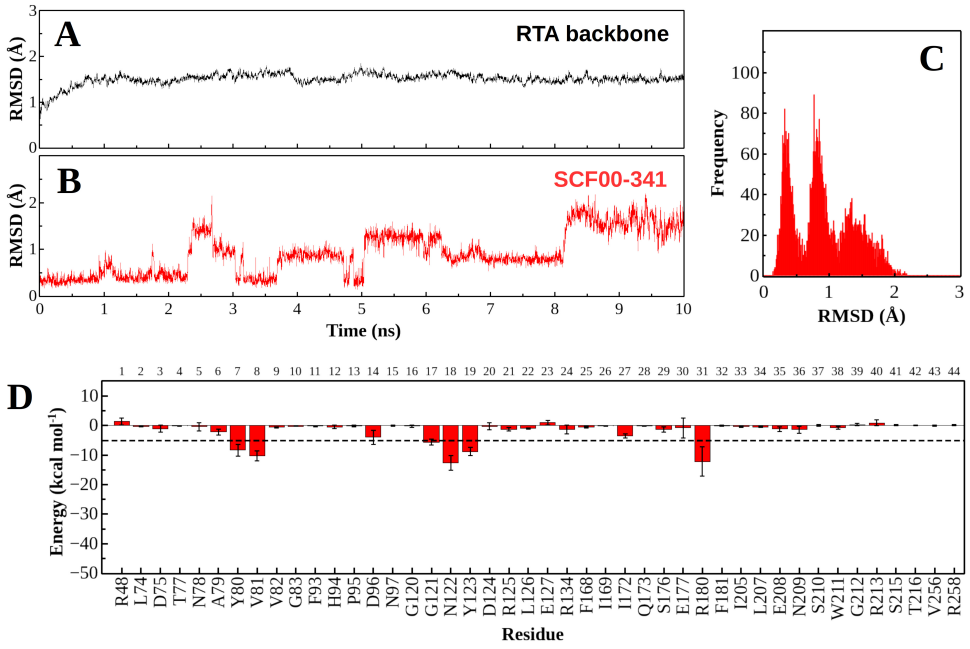


Figura 56 – Trajectory analysis of the RTA–SCF00-426 complex.

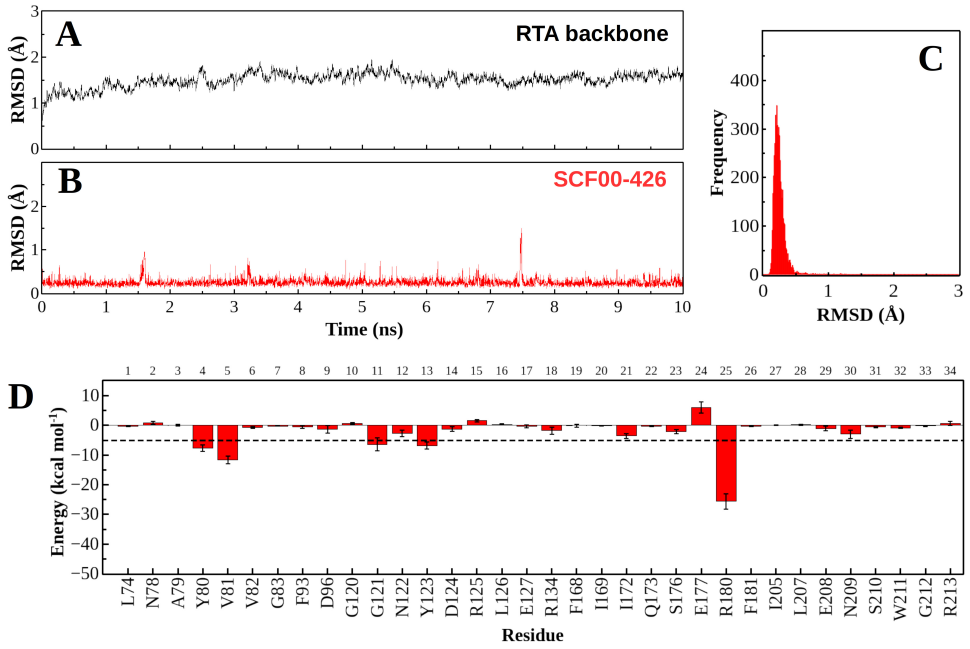


Figura 57 – Trajectory analysis of the RTA–SCF00-511 complex.

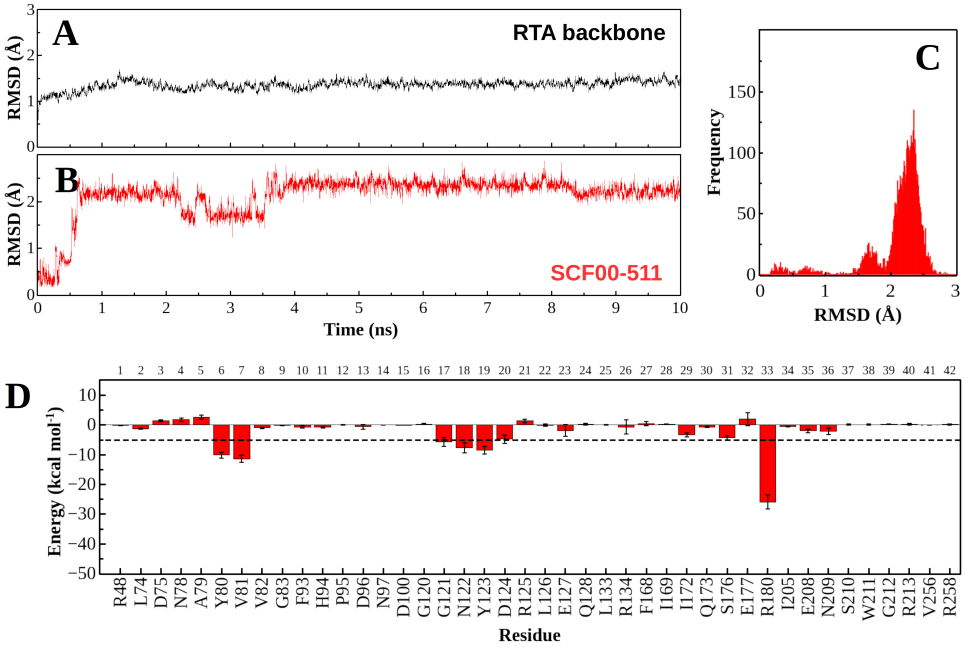


Figura 58 – Trajectory analysis of the RTA–SCF00-514 complex.

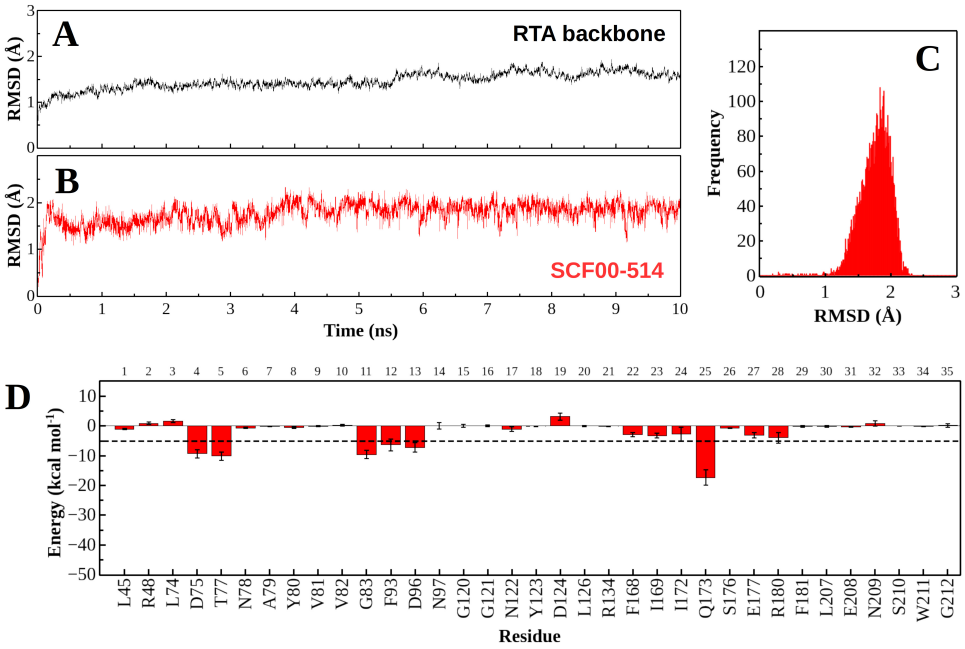


Figura 59 – Trajectory analysis of the RTA–SCF00-531 complex.

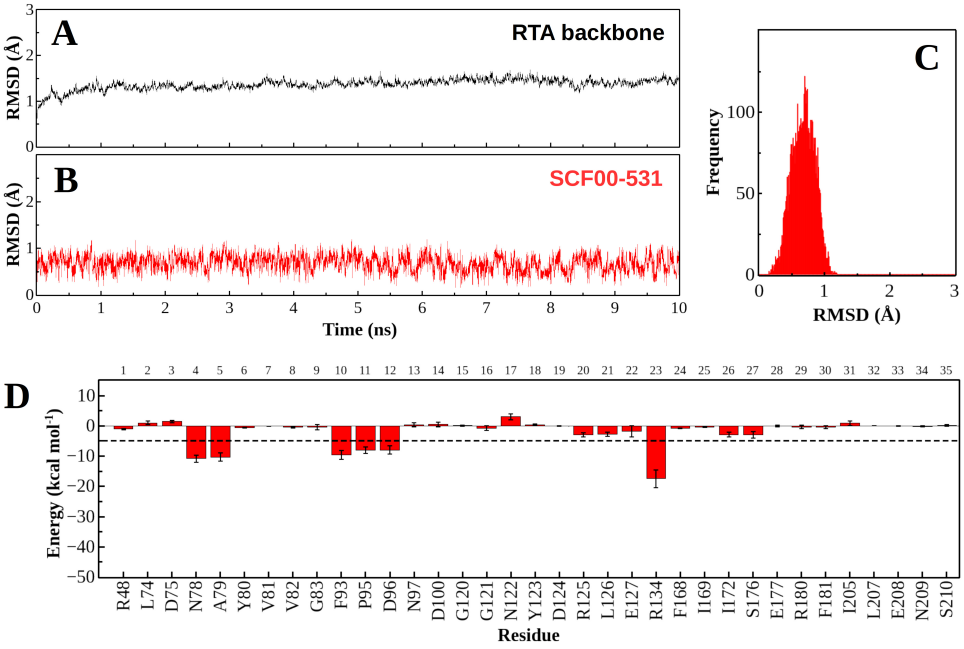


Figura 60 – Trajectory analysis of the RTA–SCF00-533 complex.

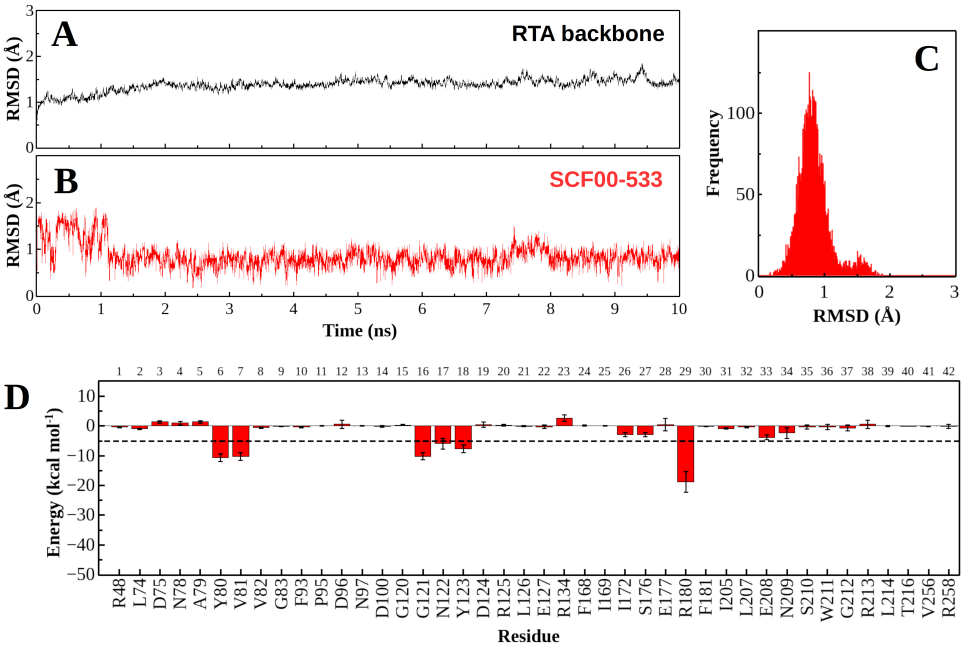
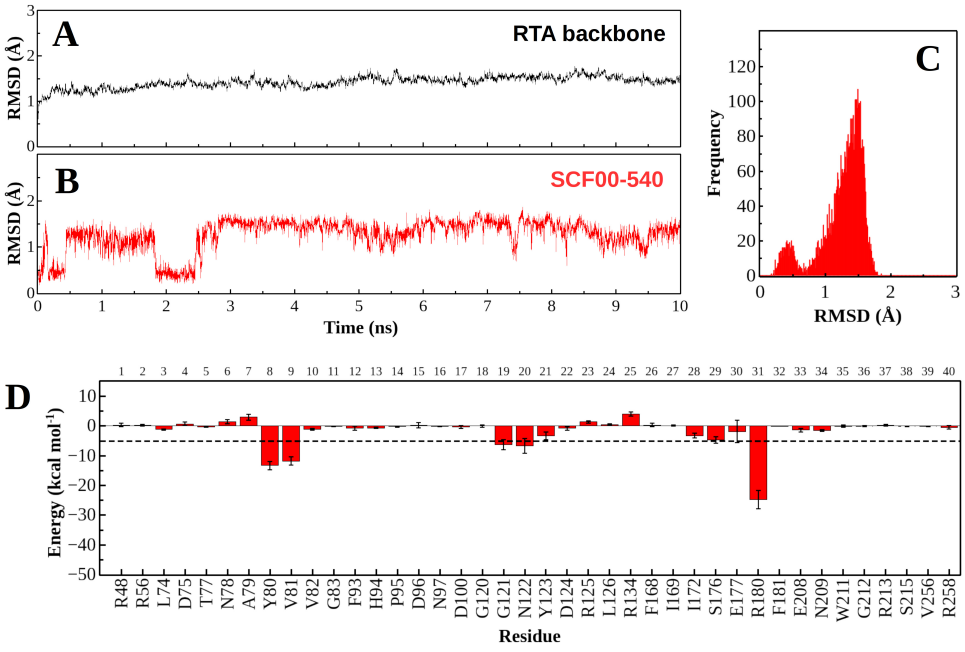


Figura 61 – Trajectory analysis of the RTA–SCF00-540 complex.



Referências

1 POLITO, L. et al. Ricin: An Ancient Story for a Timeless Plant Toxin. *Toxins*, v. 11, n. 6, p. 324, 2019.

2 OSORIO-GONZÁLEZ, C. S. et al. Production of Biodiesel from Castor Oil: A Review. *Energies*, v. 13, n. 10, p. 2467, 2020.

3 SEVERINO, L. S. et al. A Review on the Challenges for Increased Production of Castor. *Agronomy Journal*, v. 104, n. 4, p. 853–880, 2012.

4 DANG, L.; Van Damme, E. J. Toxic proteins in plants. *Phytochemistry*, v. 117, p. 51–64, 2015.

5 GRELA, P. et al. How Ricin Damages the Ribosome. *Toxins*, v. 11, n. 5, p. 241, 2019.

6 AUDI, J. et al. Ricin Poisoning. *JAMA*, v. 294, n. 18, p. 2342–2351, 2005.

7 REISLER, R. B.; SMITH, L. a. The need for continued development of ricin countermeasures. *Advances in preventive medicine*, v. 2012, p. 149737, 2012.

8 CHAVES, E. J. F. et al. Determining the Relative Binding Affinity of Ricin Toxin A Inhibitors by Using Molecular Docking and Nonequilibrium Work. *Journal of Chemical Information and Modeling*, v. 58, n. 6, p. 1205–1213, 2018.

9 CERQUEIRA, N. M. et al. Receptor-based virtual screening protocol for drug discovery. *Archives of Biochemistry and Biophysics*, v. 582, p. 56–67, 2015.

10 LAVECCHIA, A.; GIOVANNI, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Current Medicinal Chemistry*, v. 20, n. 23, p. 2839–2860, 2013.

11 KUMAR, A.; ZHANG, K. Y. Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*, v. 71, n. C, p. 26–37, 2015.

12 DO, P.-C.; LEE, E. H.; LE, L. T. Steered Molecular Dynamics Simulation in Rational Drug Design. *Journal of Chemical Information and Modeling*, 2018.

13 LAPADULA, W. J.; AYUB, M. J. Ribosome Inactivating Proteins from an evolutionary perspective. *Toxicon*, v. 136, p. 6–14, 2017.

14 STIRPE, F.; BARBIERI, L. Ribosome-inactivating proteins up to date. *FEBS Letters*, v. 195, n. 1-2, p. 1–8, 1986.

15 MONTFORT, W. et al. The three-dimensional structure of ricin at 2.8 a. *The Journal of biological chemistry*, v. 262, n. 11, p. 5398–403, 1987.

16 RUTENBER, E. et al. Crystallographic refinement of ricin to 2.5 Å. *Proteins: Structure, Function, and Genetics*, v. 10, n. 3, p. 240–250, 1991.

17 KATZIN, B. J.; COLLINS, E. J.; ROBERTUS, J. D. Structure of ricin A-chain at 2.5 Å. *Proteins: Structure, Function, and Genetics*, v. 10, n. 3, p. 251–259, 1991.

18 MLSNA, D. et al. Structure of recombinant ricin a chain at 2.3 Å. *Protein Science*, v. 2, n. 3, p. 429–435, 1993.

19 WESTON, S. a. et al. X-ray structure of recombinant ricin a-chain at 1.8 a resolution. *Journal of molecular biology*, v. 244, n. 4, p. 410–422, 1994.

20 ANDREEVA, A. et al. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Research*, v. 42, n. D1, p. D310–D314, 2014.

21 KIM, Y.; ROBERTUS, J. D. Analysis of several key active site residues of ricin A chain by mutagenesis and X-ray crystallography. *Protein Engineering, Design and Selection*, v. 5, n. 8, p. 775–779, 1992.

22 READY, M. P.; KIM, Y.; ROBERTUS, J. D. Site-directed mutagenesis of ricin A-chain and implications for the mechanism of action. *Proteins*, v. 10, n. 3, p. 270–8, 1991.

23 FRANKEL, A. et al. Role of arginine 180 and glutamic acid 177 of ricin toxin A chain in enzymatic inactivation of ribosomes. *Molecular and Cellular Biology*, v. 10, n. 12, p. 6257–6263, 1990.

24 KIM, Y. et al. Structure of a ricin mutant showing rescue of activity by a noncatalytic residue. *Biochemistry*, v. 31, n. 12, p. 3294–3296, 1992.

25 MONZINGO, A. F.; ROBERTUS, J. D. X-ray analysis of substrate analogs in the ricin A-chain active site. *Journal of molecular biology*, v. 227, n. 4, p. 1136–1145, 1992.

26 WAHOME, P. G.; ROBERTUS, J. D.; MANTIS, N. J. *Small-Molecule Inhibitors of Ricin and Shiga Toxins*. [S.l.: s.n.], 2011. v. 357. 179–207 p.

27 HO, M.-C. et al. Transition state analogues in structures of ricin and saporin ribosome-inactivating proteins. *Proceedings of the National Academy of Sciences*, v. 106, n. 48, p. 20276–20281, 2009.

28 SPOONER, R. A. et al. Protein disulphide-isomerase reduces ricin to its A and B chains in the endoplasmic reticulum. *Biochemical Journal*, v. 383, n. 2, p. 285–293, 2004.

29 SIMPSON, J. C. et al. Ricin A chain utilises the endoplasmic reticulum-associated protein degradation pathway to enter the cytosol of yeast. *FEBS Letters*, v. 459, n. 1, p. 80–84, 1999.

30 FAN, X. et al. Structural insights into the interaction of the ribosomal P stalk protein P2 with a type II ribosome-inactivating protein ricin. *Scientific Reports*, v. 6, n. jun, p. 1–10, 2016.

31 CHIOU, J.-C. et al. The ribosomal stalk is required for ribosome binding, depurination of the rRNA and cytotoxicity of ricin A chain in Saccharomyces cerevisiae. *Molecular Microbiology*, v. 70, n. 6, p. 1441–1452, 2008.

32 MAY, K. L. et al. The P1/P2 proteins of the human ribosomal stalk are required for ribosome binding and depurination by ricin in human cells. *FEBS Journal*, v. 279, n. 20, p. 3925–3936, 2012.

33 CHOI, A. K. et al. Structures of eukaryotic ribosomal stalk proteins and its complex with trichosanthin, and their implications in recruiting ribosome-inactivating proteins to the ribosomes. *Toxins*, v. 7, n. 3, p. 638–647, 2015.

34 SHI, W. W. et al. Crystal structure of ribosome-inactivating protein ricin a chain in complex with the C-terminal peptide of the ribosomal stalk protein P2. *Toxins*, v. 8, n. 10, 2016.

35 ZHOU, Y. et al. Leucine 232 and hydrophobic residues at the ribosomal P stalk binding site are critical for biological activity of ricin. *Bioscience Reports*, v. 39, n. 10, p. 1–15, 2019.

36 ZHOU, Y. et al. Ricin uses arginine 235 as an anchor residue to bind to P-proteins of the ribosomal stalk. *Scientific Reports*, v. 7, n. August 2016, p. 1–12, 2017.

37 LI, X. P.; KAHN, J. N.; TUMER, N. E. Peptide Mimics of the Ribosomal P Stalk Inhibit the Activity of Ricin A Chain by Preventing Ribosome Binding. *Toxins*, v. 10, n. 9, p. 1–13, 2018.

38 LI, X.-P. et al. Small Molecule Inhibitors Targeting the Interaction of Ricin Toxin A Subunit with Ribosomes. *ACS Infectious Diseases*, p. acsinfecdis.0c00127, 2020.

39 GAL, Y. et al. Treatments for Pulmonary Ricin Intoxication: Current Aspects and Future Prospects. *Toxins*, v. 9, n. 10, p. 311, 2017.

40 VIEIRA, K.; Tavares Machado, O. L. Approaches for the Detection of Toxic Compounds in Castor and Physic Nut Seeds and Cakes. In: . [S.l.: s.n.], 2012. ISBN 978-953-51-0910-5.

41 YOSHIDA, T. et al. Disruption of the Golgi apparatus by brefeldin A inhibits the cytotoxicity of ricin, modeccin, and Pseudomonas toxin. *Experimental Cell Research*, v. 192, n. 2, p. 389–395, 1991.

42 SANDVIG, K. et al. Ricin transport in brefeldin A-treated cells: correlation between Golgi structure and toxic effect. *The Journal of Cell Biology*, v. 115, n. 4, p. 971–981, 1991.

43 WELLNER, R. B.; PLESS, D. D.; THOMPSON, W. L. Characterization of 3'-azido-3'-deoxythymidine inhibition of ricin andpseudomonas exotoxin A toxicity in CHO and vero cells. *Journal of Cellular Physiology*, v. 159, n. 3, p. 495–505, 1994.

44 NAMBIAR, M. P.; MURUGESAN, R.; WU, H. C. Inhibition of the cytotoxicity of protein toxins by a novel plant metabolite, Mansonone-D. *Journal of Cellular Physiology*, v. 176, n. 1, p. 40–49, 1998.

45 STECHMANN, B. et al. Inhibition of Retrograde Transport Protects Mice from Lethal Ricin Challenge. *Cell*, v. 141, n. 2, p. 231–242, 2010.

46 YU, S. et al. Common Pharmacophore of Structurally Distinct Small-Molecule Inhibitors of Intracellular Retrograde Trafficking of Ribosome Inactivating Proteins. *Scientific Reports*, v. 3, n. 1, p. 3397, 2013.

47 BASSIK, M. C. et al. A Systematic Mammalian Genetic Interaction Map Reveals Pathways Underlying Ricin Susceptibility. *Cell*, v. 152, n. 4, p. 909–922, 2013.

48 SMALLSHAW, J. A novel recombinant vaccine which protects mice against ricin intoxication. *Vaccine*, v. 20, n. 27-28, p. 3422–3427, 2002.

49 SMALLSHAW, J. E. et al. Preclinical toxicity and efficacy testing of RiVax, a recombinant protein vaccine against ricin. *Vaccine*, v. 23, n. 39, p. 4775–4784, 2005.

50 VITETTA, E. S. et al. A pilot clinical trial of a recombinant ricin vaccine in normal humans. *Proceedings of the National Academy of Sciences*, v. 103, n. 7, p. 2268–2273, 2006.

51 HESSELBERTH, J. R. et al. In vitro selection of RNA molecules that inhibit the activity of ricin A-chain. *The Journal of biological chemistry*, v. 275, n. 7, p. 4937–4942, 2000.

52 ROBERTUS, J.; MONZINGO, A. The Structure of Ribosome Inactivating Proteins. *Mini-Reviews in Medicinal Chemistry*, v. 4, n. 5, p. 477–486, 2004.

53 YAN, X. et al. Recognition and interaction of small rings with the ricin A-chain binding site. *Proteins*, v. 31, n. 1, p. 33–41, 1998.

54 MILLER, D. J. et al. Structure-Based Design and Characterization of Novel Platforms for Ricin and Shiga Toxin Inhibition. *Journal of Medicinal Chemistry*, v. 45, n. 1, p. 90–98, 2002.

55 CARRA, J. H. et al. Fragment-based identification of determinants of conformational and spectroscopic change at the ricin active site. *BMC structural biology*, v. 7, p. 72, 2007.

56 YAN, X. et al. Structure-based identification of a ricin inhibitor. *Journal of Molecular Biology*, v. 266, n. 5, p. 1043–1049, 1997.

57 PRUET, J. M. et al. 7-Substituted pterins provide a new direction for ricin A chain inhibitors. *European Journal of Medicinal Chemistry*, v. 46, n. 9, p. 3608–3615, 2011.

58 SAITO, R. et al. Peptide-Conjugated Pterins as Inhibitors of Ricin Toxin A. *Journal of Medicinal Chemistry*, v. 56, n. 1, p. 320–329, 2013.

59 PRUET, J. M. et al. Optimized 5-Membered Heterocycle-Linked Pterins for the Inhibition of Ricin Toxin A. *ACS Medicinal Chemistry Letters*, v. 3, n. 7, p. 588–591, 2012.

60 BAI, Y.; MONZINGO, A. F.; ROBERTUS, J. D. The X-ray structure of ricin A chain with a novel inhibitor. *Archives of Biochemistry and Biophysics*, v. 483, n. 1, p. 23–28, 2009.

61 PANG, Y. P. et al. Small-molecule inhibitor leads of Ribosome-inactivating proteins developed using the doorstop approach. *PLoS ONE*, v. 6, n. 3, p. 1–15, 2011.

62 WAHOME, P. G. et al. Identification of small-molecule inhibitors of ricin and shiga toxin using a cell-based high-throughput screen. *Toxicon*, v. 56, n. 3, p. 313–323, 2010.

63 MORGON, N. H.; COUTINHO, K. R. *Métodos de química teórica e modelagem molecular*. 1. ed. [S.l.: s.n.], 2007. 540 p.

64 LIONTA, E. et al. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry*, v. 14, n. 16, p. 1923–1938, 2014.

65 KAR, S.; ROY, K. How far can virtual screening take us in drug discovery? *Expert Opinion on Drug Discovery*, v. 8, n. 3, p. 245–261, 2013.

66 STUMPFE, D.; BAJORATH, J. Similarity searching. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 1, n. 2, p. 260–282, 2011.

67 PIRHADI, S.; SHIRI, F.; GHASEMI, J. B. Methods and Applications of Structure Based Pharmacophores in Drug Discovery. *Current Topics in Medicinal Chemistry*, v. 13, n. 9, p. 1036–1047, 2013.

68 BERMAN, H. M. The Protein Data Bank. *Nucleic Acids Research*, v. 28, n. 1, p. 235–242, 2000.

69 STERLING, T.; IRWIN, J. J. ZINC 15–Ligand Discovery for Everyone. *Journal of chemical information and modeling*, v. 55, n. 11, p. 2324–37, 2015.

70 BARREIRO, E. J.; FRAGA, C. A. M. *Química Medicinal: As Bases Moleculares da Ação dos Fármacos*. 3. ed. [S.l.: s.n.], 2014. 608 p.

71 VIJAYAKRISHNAN, P.; Arul Antony, S.; VELMURUGAN, D. Structural data of DNA binding and molecular docking studies of dihydropyrimidinone transition metal complexes. *Data in Brief*, v. 19, p. 817–825, 2018.

72 Murali Krishna, P. et al. Synthesis, structural studies, molecular docking and DNA binding studies of 4N-substituted hydrazinecarbothioamides. *Journal of Molecular Structure*, v. 1175, p. 97–104, 2019.

73 VERLI, H. *Bioinformática da Biologia à Flexibilidade Molecular*. 1. ed. São Paulo: [s.n.], 2014. 282 p.

74 PAGADALA, N. S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. *Biophysical Reviews*, v. 9, n. 2, p. 91–102, 2017.

75 BAI, Y. et al. Identification of new classes of ricin toxin inhibitors by virtual screening. *Toxicon*, v. 56, n. 4, p. 526–534, 2010.

76 CHOUDHRY, H. et al. Zika Virus Targeting by Screening Inhibitors against NS2B/NS3 Protease. *BioMed Research International*, v. 2019, p. 1–11, 2019.

77 YENNAMALLI, R. et al. Identification of novel target sites and an inhibitor of the dengue virus E protein. *Journal of Computer-Aided Molecular Design*, v. 23, n. 6, p. 333–341, 2009.

78 KASERER, T. et al. Pharmacophore models and pharmacophore-based virtual screening: Concepts and applications exemplified on hydroxysteroid dehydrogenases. *Molecules*, v. 20, n. 12, p. 22799–22832, 2015.

79 TEMML, V. et al. Discovery of new liver X receptor agonists by pharmacophore modeling and shape-based virtual screening. *Journal of Chemical Information and Modeling*, v. 54, n. 2, p. 367–371, 2014.

80 KENDREW, J. C. et al. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, v. 181, n. 4610, p. 662–666, 1958.

81 KENDREW, J. C. et al. Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature*, v. 185, n. 4711, p. 422–427, 1960.

82 PERUTZ, M. F. et al. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*, v. 185, n. 4711, p. 416–422, 1960.

83 HENZLER-WILDMAN, K.; KERN, D. Dynamic personalities of proteins. *Nature*, v. 450, n. 7172, p. 964–972, 2007.

84 PAULO; SOUZA, C. T. de. *Modelagem molecular de receptores nucleares: estrutura, dinâmica e interação com ligantes*. 145 p. Tese (Doutorado) — Universidade de Campinas, 2013.

85 GUTERRES, H.; IM, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, v. 60, n. 4, p. 2189–2198, 2020.

86 GRUBMÜLLER, H.; HEYMANN, B.; TAVAN, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science*, v. 271, n. 9, p. 997–999, 1996.

87 ISRALEWITZ, B. et al. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics and Modelling*, v. 19, n. 1, p. 13–25, 2001.

88 COLIZZI, F. et al. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *Journal of the American Chemical Society*, v. 132, n. 21, p. 7361–7371, 2010.

89 PARK, S.; SCHULTEN, K. Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of Chemical Physics*, v. 120, n. 13, p. 5946, 2004.

90 ARODOLA, O. A.; SOLIMAN, M. E. Quantum mechanics implementation in drug-design workflows: does it really help? *Drug Design, Development and Therapy*, Volume 11, p. 2551–2564, 2017.

91 PUTZ, M. V.; MINGOS, D. M. P. (Ed.). *Applications of Density Functional Theory to Biological and Bioinorganic Chemistry*. [S.l.]: Springer Berlin Heidelberg, 2013.

92 CHRISTENSEN, A. S. et al. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chemical Reviews*, v. 116, n. 9, p. 5301–5337, 2016.

93 ROCHA, G. B. et al. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *Journal of Computational Chemistry*, v. 27, n. 10, p. 1101–1111, 2006.

94 STEWART, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, v. 13, n. 12, p. 1173–1213, 2007.

95 KORTH, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields. *Journal of Chemical Theory and Computation*, v. 6, n. 12, p. 3808–3816, 2010.

96 ŘEZÁČ, J.; HOBZA, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *Journal of Chemical Theory and Computation*, v. 8, n. 1, p. 141–151, 2012.

97 STEWART, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*, v. 19, n. 1, p. 1–32, 2013.

98 PENG, C. et al. Improving the accuracy of predicting protein–ligand binding-free energy with semiempirical quantum chemistry charge. *Future Medicinal Chemistry*, v. 11, n. 4, p. 303–321, 2019.

99 SULIMOV, A. V. et al. Combined docking with classical force field and quantum chemical semiempirical method PM7. *Advances in Bioinformatics*, v. 2017, p. 1–6, 2017.

100 HOSTAŠ, J.; ŘEZÁČ, J.; HOBZA, P. On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions. *Chemical Physics Letters*, v. 568-569, p. 161–166, 2013.

101 ROCHA-SANTOS, A. et al. Thermochemical and quantum descriptor calculations for gaining insight into ricin toxin a (RTA) inhibitors. *ACS Omega*, American Chemical Society (ACS), v. 6, n. 13, p. 8764–8777, mar. 2021. Disponível em: <https://doi.org/10.1021/acsomega.0c02588>.

102 OPREA, T. I. et al. Is There a Difference between Leads and Drugs? A Historical Perspective. *Journal of Chemical Information and Computer Sciences*, v. 41, n. 5, p. 1308–1315, 2001.

103 HAWKINS, P. C. D. et al. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, v. 50, n. 4, p. 572–584, 2010.

104 MYSINGER, M. M. et al. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry*, v. 55, n. 14, p. 6582–6594, 2012.

105 O'BOYLE, N. M. et al. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, v. 3, n. 1, p. 33, 2011.

106 EMPEREUR-MOT, C.; ZAGURY, J.-F.; MONTES, M. Screening Explorer–An Interactive Tool for the Analysis of Screening Results. *Journal of Chemical Information and Modeling*, v. 56, n. 12, p. 2281–2286, 2016.

107 ROBERTS, E. et al. MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, v. 7, p. 1–11, 2006.

108 DOLINSKY, T. J. et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, v. 35, n. Web Server, p. W522–W525, 2007.

109 PHILLIPS, J. C. et al. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, v. 26, n. 16, p. 1781–802, 2005.

110 HORNAK, V. et al. *Comparison of multiple amber force fields and development of improved protein backbone parameters*. 2006. 712–725 p.

111 WANG, J. et al. Development and testing of a general amber force field. *Journal of Computational Chemistry*, v. 25, n. 9, p. 1157–1174, 2004.

Referências do Artigo

112 GRELA, P. et al. How ricin damages the ribosome. *Toxins*, v. 11, n. 5, p. 241, apr 2019. ISSN 2072-6651.

113 POLITO, L. et al. Ricin: An Ancient Story for a Timeless Plant Toxin. *Toxins*, v. 11, n. 6, p. 324, jun 2019.

114 GAL, Y. et al. Treatments for Pulmonary Ricin Intoxication: Current Aspects and Future Prospects. *Toxins*, v. 9, n. 10, p. 311, oct 2017. ISSN 2072-6651.

115 SEVERINO, L. S. et al. A review on the challenges for increased production of castor. *Agronomy journal*, v. 104, n. 4, p. 853–880, 2012.

116 MONTFORT, W. et al. The three-dimensional structure of ricin at 2.8 a. *The Journal of biological chemistry*, v. 262, n. 11, p. 5398–403, apr 1987. ISSN 0021-9258.

117 RUTENBER, E. et al. Crystallographic refinement of ricin to 2.5 Å. *Proteins: Structure, Function, and Genetics*, v. 10, n. 3, p. 240–250, mar 1991.

118 MLSNA, D. et al. Structure of recombinant ricin a chain at 2.3 Å. *Protein Science*, v. 2, n. 3, p. 429–435, 1993.

119 WESTON, S. a. et al. X-ray structure of recombinant ricin a-chain at 1.8 a resolution. *Journal of molecular biology*, v. 244, n. 4, p. 410–422, 1994. ISSN 0022-2836.

120 MONZINGO, A. F.; ROBERTUS, J. D. X-ray analysis of substrate analogs in the ricin A-chain active site. *Journal of molecular biology*, v. 227, n. 4, p. 1136–1145, oct 1992. ISSN 00222836.

121 HO, M.-C. et al. Transition state analogues in structures of ricin and saporin ribosome-inactivating proteins. *Proceedings of the National Academy of Sciences*, v. 106, n. 48, p. 20276–20281, dec 2009.

122 JASHEWAY, K. et al. Structure-based design of ricin inhibitors. *Toxins*, v. 3, n. 10, p. 1233–1248, oct 2011.

123 BAI, Y.; MONZINGO, A. F.; ROBERTUS, J. D. The x-ray structure of ricin a chain with a novel inhibitor. *Archives of Biochemistry and Biophysics*, v. 483, n. 1, p. 23–28, mar 2009. ISSN 00039861.

124 NOY-PORAT, T. et al. Extended therapeutic window for post-exposure treatment of ricin intoxication conferred by the use of high-affinity antibodies. *Toxicon*, v. 127, p. 100–105, mar 2017.

125 NOY-PORAT, T. et al. Isolation of Anti-Ricin Protective Antibodies Exhibiting High Affinity from Immunized Non-Human Primates. *Toxins*, v. 8, n. 3, p. 64, mar 2016.

126 YERMAKOVA, A.; MANTIS, N. J. Protective immunity to ricin toxin conferred by antibodies against the toxin's binding subunit (RTB). *Vaccine*, v. 29, n. 45, p. 7925–7935, oct 2011.

127 NEAL, L. M. et al. A Monoclonal Immunoglobulin G Antibody Directed against an Immunodominant Linear Epitope on the Ricin A Chain Confers Systemic and Mucosal Immunity to Ricin. *Infection and Immunity*, v. 78, n. 1, p. 552–561, jan 2010.

128 SAITO, R. et al. Peptide-Conjugated Pterins as Inhibitors of Ricin Toxin A. *Journal of Medicinal Chemistry*, v. 56, n. 1, p. 320–329, jan 2013.

129 STERLING, T.; IRWIN, J. J. ZINC 15–Ligand Discovery for Everyone. *Journal of chemical information and modeling*, v. 55, n. 11, p. 2324–37, nov 2015.

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

131 SLIWOSKI, G. et al. Computational methods in drug discovery. *Pharmacol. Rev.*, v. 66, n. 1, p. 334–395, 2014.

132 GIMENO, A. et al. The light and dark sides of virtual screening: What is there to know? *International Journal of Molecular Sciences*, v. 20, n. 6, 2019.

133 BOTELHO, F. D. et al. Ligand-based virtual screening, molecular docking, molecular dynamics, and MM-PBSA calculations towards the identification of potential novel ricin inhibitors. *Toxins*, MDPI AG, v. 12, n. 12, p. 746, nov. 2020.

134 PRUET, J. M. et al. Optimized 5-Membered Heterocycle-Linked Pterins for the Inhibition of Ricin Toxin A. *ACS Medicinal Chemistry Letters*, v. 3, n. 7, p. 588–591, jul 2012.

135 WIGET, P. A. et al. Sulfur incorporation generally improves Ricin inhibition in pterin-appended glycine-phenylalanine dipeptide mimics. *Bioorganic & Medicinal Chemistry Letters*, v. 23, n. 24, p. 6799–6804, dec 2013.

136 PRUET, J. M. et al. 7-Substituted pterins provide a new direction for ricin A chain inhibitors. *European Journal of Medicinal Chemistry*, v. 46, n. 9, p. 3608–3615, sep 2011.

137 MILLER, D. J. et al. Structure-Based Design and Characterization of Novel Platforms for Ricin and Shiga Toxin Inhibition. *Journal of Medicinal Chemistry*, v. 45, n. 1, p. 90–98, jan 2002.

138 YAN, X. et al. Structure-based identification of a ricin inhibitor. *Journal of Molecular Biology*, v. 266, n. 5, p. 1043–1049, mar 1997.

139 CHAVES, E. J. F. et al. Determining the Relative Binding Affinity of Ricin Toxin A Inhibitors by Using Molecular Docking and Nonequilibrium Work. *Journal of Chemical Information and Modeling*, v. 58, n. 6, p. 1205–1213, jun 2018.

140 ISRALEWITZ, B.; GAO, M.; SCHULTEN, K. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, v. 11, n. 2, p. 224–230, apr 2001. ISSN 0959440X.

141 FLORIN, E.; MOY, V.; GAUB, H. Adhesion forces between individual ligand-receptor pairs. *Science*, v. 264, n. 5157, p. 415–417, apr 1994. ISSN 0036-8075.

142 COLIZZI, F. et al. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *Journal of the American Chemical Society*, v. 132, n. 21, p. 7361–7371, 2010. ISSN 00027863.

143 DO, P.-C.; LEE, E. H.; LE, L. T. Steered Molecular Dynamics Simulation in Rational Drug Design. *Journal of Chemical Information and Modeling*, p. acs.jcim.8b00261, jul 2018.

144 LO, R. et al. The reactivation of tabun-inhibited mutant AChE with Ortho-7: steered molecular dynamics and quantum chemical studies. *Molecular BioSystems*, Royal Society of Chemistry, v. 12, n. 4, p. 1224–1231, 2016. ISSN 1742-206X.

145 LÜDEMANN, S. K.; LOUNNAS, V.; WADE, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways 1 1Edited by J. Thornton. *Journal of Molecular Biology*, v. 303, n. 5, p. 813–830, nov 2000.

146 LÜDEMANN, S. K.; LOUNNAS, V.; WADE, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms 1 1Edited by J. Thornton. *Journal of Molecular Biology*, v. 303, n. 5, p. 797–811, nov 2000.

147 TROTT, O.; OLSON, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, p. NA–NA, 2009.

148 RYDE, U.; SÖDERHJELM, P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chemical Reviews*, v. 116, n. 9, p. 5520–5566, may 2016.

149 ROCHA, S. F. L. S.; SANT'ANNA, C. M. R. A procedure combining molecular docking and semiempirical method PM7 for identification of selective Shp2 inhibitors. *Biopolymers*, n. February, p. 1–8, 2019. ISSN 0006-3525.

150 ROCHA, G. B. et al. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *Journal of Computational Chemistry*, v. 27, n. 10, p. 1101–1111, jul 2006.

151 STEWART, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, v. 13, n. 12, p. 1173–1213, oct 2007.

152 KORTH, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields. *Journal of Chemical Theory and Computation*, v. 6, n. 12, p. 3808–3816, dec 2010.

153 ŘEZÁČ, J.; HOBZA, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *Journal of Chemical Theory and Computation*, v. 8, n. 1, p. 141–151, jan 2012.

154 STEWART, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*, v. 19, n. 1, p. 1–32, jan 2013. ISSN 1610-2940.

155 PENG, C. et al. Improving the accuracy of predicting protein–ligand binding-free energy with semiempirical quantum chemistry charge. *Future Medicinal Chemistry*, v. 11, n. 4, p. 303–321, feb 2019. ISSN 1756-8919.

156 SULIMOV, A. V. et al. Combined Docking with Classical Force Field and Quantum Chemical Semiempirical Method PM7. *Advances in Bioinformatics*, v. 2017, 2017. ISSN 16878035.

157 HOSTAŠ, J.; ŘEZÁČ, J.; HOBZA, P. On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions. *Chemical Physics Letters*, v. 568-569, p. 161–166, may 2013. ISSN 00092614.

158 OPREA, T. I. et al. Is There a Difference between Leads and Drugs? A Historical Perspective. *Journal of Chemical Information and Computer Sciences*, v. 41, n. 5, p. 1308–1315, sep 2001.

159 HAWKINS, P. C. D. et al. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, v. 50, n. 4, p. 572–584, apr 2010.

160 HAWKINS, P. C. D.; SKILLMAN, A. G.; NICHOLLS, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *Journal of Medicinal Chemistry*, v. 50, n. 1, p. 74–82, jan 2007.

161 MYSINGER, M. M. et al. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry*, v. 55, n. 14, p. 6582–6594, jul 2012.

162 JONES, G. et al. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, v. 267, n. 3, p. 727–748, apr 1997.

163 PAGADALA, N. S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. *Biophysical Reviews*, Biophysical Reviews, v. 9, n. 2, p. 91–102, 2017.

164 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, v. 14, n. 1, p. 33–38, feb 1996.

165 PHILLIPS, J. C. et al. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, v. 26, n. 16, p. 1781–802, dec 2005.

166 HORNAK, V. et al. *Comparison of multiple amber force fields and development of improved protein backbone parameters*. 2006. 712–725 p.

167 WANG, J. et al. Development and testing of a general amber force field. *Journal of Computational Chemistry*, v. 25, n. 9, p. 1157–1174, jul 2004.

168 WANG, J. et al. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, v. 25, n. 2, p. 247–260, oct 2006.

169 ROE, D. R.; CHEATHAM, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, v. 9, n. 7, p. 3084–3095, jul 2013.

170 PARK, S. et al. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics*, v. 119, n. 6, p. 3559, 2003.

171 LIPHARDT, J. et al. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, v. 296, n. 5574, p. 1832–1835, 2002. ISSN 00368075.

172 XIONG, H. et al. Free energy calculations with non-equilibrium methods: Applications of the Jarzynski relationship. *Theoretical Chemistry Accounts*, v. 116, n. 1-3, p. 338–346, 2006. ISSN 1432881X.

173 URQUIZA-CARVALHO, G. A.; FRAGOSO, W. D.; ROCHA, G. B. Assessment of semiempirical enthalpy of formation in solution as an effective energy function to discriminate native-like structures in protein decoy sets. *Journal of computational chemistry*, v. 37, n. 21, p. 1962–1972, 2016. ISSN 1096987X.

174 STEWART, J. J. P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *International Journal of Quantum Chemistry*, v. 58, n. 2, p. 133–146, 1996. ISSN 0020-7608.

175 STEWART, J. J. P. MOPAC: A semiempirical molecular orbital program. *Journal of Computer-Aided Molecular Design*, v. 4, n. 1, p. 1–103, mar 1990. ISSN 0920-654X.

176 MAIA, J. D. C. et al. GPU Linear Algebra Libraries and GPGPU Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations. *Journal of Chemical Theory and Computation*, v. 8, n. 9, p. 3072–3081, sep 2012.

177 MAI, B. K.; VIET, M. H.; LI, M. S. Top leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach. *Journal of Chemical Information and Modeling*, v. 50, n. 12, p. 2236–2247, 2010. ISSN 15499596.

178 RANDJELOVIĆ, J.; ERIĆ, S.; SAVIĆ, V. Computational study and peptide inhibitors design for the CDK9 – cyclin T1 complex. *Journal of Molecular Modeling*, v. 19, n. 4, p. 1711–1725, apr 2013.

179 NGO, S. T.; HUNG, H. M.; NGUYEN, M. T. Fast and accurate determination of the relative binding affinities of small compounds to HIV-1 protease using non-equilibrium work. *Journal of Computational Chemistry*, v. 37, n. 31, p. 2734–2742, dec 2016.

180 RUIZ-CARMONA, S. et al. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nature Chemistry*, Nature Publishing Group, v. 9, n. 3, p. 201–206, 2017. ISSN 17554349.

181 NIU, Y. et al. Computational study on the unbinding pathways of B-RAF inhibitors and its implication for the difference of residence time: insight from random acceleration and steered molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, v. 18, n. 7, p. 5622–5629, 2016.

Referências do Material Suplementar do Artigo

182 CERQUEIRA, N. M. et al. Receptor-based virtual screening protocol for drug discovery. *Archives of Biochemistry and Biophysics*, Elsevier Inc., v. 582, p. 56–67, 2015. ISSN 10960384.

183 MYSINGER, M. M. et al. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry*, v. 55, n. 14, p. 6582–6594, jul 2012. ISSN 0022-2623. Disponível em: <https://pubs.acs.org/doi/10.1021/jm300687e>.

184 HAWKINS, P. C. D. et al. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, v. 50, n. 4, p. 572–584, apr 2010. ISSN 1549-9596. Disponível em: <https://pubs.acs.org/doi/10.1021/ci100031x>.

185 JONES, G. et al. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, v. 267, n. 3, p. 727–748, apr 1997. ISSN 00222836. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0022283696908979>.

186 SAITO, R. et al. Peptide-Conjugated Pterins as Inhibitors of Ricin Toxin A. *Journal of Medicinal Chemistry*, American Chemical Society, v. 56, n. 1, p. 320–329, jan 2013. ISSN 0022-2623. Disponível em: <http://dx.doi.org/10.1021/jm3016393 http://pubs.acs.org/doi/10.1021/jm3016393>.

187 PRUET, J. M. et al. Optimized 5-Membered Heterocycle-Linked Pterins for the Inhibition of Ricin Toxin A. *ACS Medicinal Chemistry Letters*, v. 3, n. 7, p. 588–591, jul 2012. ISSN 1948-5875.

188 WIGET, P. A. et al. Sulfur incorporation generally improves Ricin inhibition in pterin-appended glycine-phenylalanine dipeptide mimics. *Bioorganic & Medicinal Chemistry Letters*, v. 23, n. 24, p. 6799–6804, dec 2013. ISSN 0960894X.

189 PRUET, J. M. et al. 7-Substituted pterins provide a new direction for ricin A chain inhibitors. *European Journal of Medicinal Chemistry*, v. 46, n. 9, p. 3608–3615, sep 2011. ISSN 02235234.

190 BAI, Y.; MONZINGO, A. F.; ROBERTUS, J. D. The x-ray structure of ricin a chain with a novel inhibitor. *Archives of Biochemistry and Biophysics*, v. 483, n. 1, p. 23–28, mar 2009. ISSN 00039861.

191 MILLER, D. J. et al. Structure-Based Design and Characterization of Novel Platforms for Ricin and Shiga Toxin Inhibition. *Journal of Medicinal Chemistry*, American Chemical Society, v. 45, n. 1, p. 90–98, jan 2002. ISSN 0022-2623.

192 YAN, X. et al. Structure-based identification of a ricin inhibitor. *Journal of Molecular Biology*, v. 266, n. 5, p. 1043–1049, mar 1997. ISSN 00222836.

193 DOLINSKY, T. J. et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, v. 35, n. Web Server, p. W522–W525, may 2007. ISSN 0305-1048.