



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

MAURÍCIO BOFF DE ÁVILA

**Biologia de Sistemas Computacional Aplicada à Via Metabólica
do Chiquimato. Enfoque na Enzima 3-Desidroquinato Desidratase
(EC 4.2.1.10)**

PORTO ALEGRE
2017

MAURÍCIO BOFF DE ÁVILA

**Biologia de Sistemas Computacional Aplicada à Via Metabólica
do Chiquimato. Enfoque na Enzima 3-Desidroquinato Desidratase
(EC 4.2.1.10)**

Dissertação de Mestrado apresentada como
requisito para obtenção do grau de Mestre
pelo Programa de Pós-Graduação em
Biologia Celular e Molecular da Pontifícia
Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Dr. Walter Filgueira de Azevedo Jr.

PORTO ALEGRE
2017

Ficha Catalográfica

D278b de Ávila, Maurício Boff

Biologia de sistemas computacional aplicada à via metabólica do chiquimato. Enfoque na enzima 3-desidroquinato desidratase (EC 4.2.1.10) / Maurício Boff de Ávila . – 2017.

58 f.

Dissertação (Mestrado) – Programa de Pós-Graduação em Biologia Celular e Molecular, PUCRS.

Orientador: Prof. Dr. Walter Filgueira de Azevedo Jr.

1. Via metabólica do chiquimato. 2. 3-desidroquinato desidratase. 3. Biologia de sistemas computacional. 4. Aprendizado de máquina. 5. Docking. I. de Azevedo Jr, Walter Filgueira. II. Título.

*“Science and everyday life cannot
and should not be separated”.*

Rosalind Franklin

*“Cada pessoa deve trabalhar para o seu aperfeiçoamento e,
ao mesmo tempo, participar da responsabilidade
coletiva por toda a humanidade”.*

Marie Curie

AGRADECIMENTOS

Aos meus pais, Saleti e Renato, por todo o carinho e apoio durante toda a minha caminhada até aqui. Não tenho palavras para descrever o quanto sou grato por todos os sacrifícios que fizeram por mim sem pestanejar e por todo o incentivo nos momentos bons e ruins. Essa dissertação, com certeza, tem um pouco de vocês, por terem me ensinado o quão importante é o conhecimento e que sempre se deve querer aprender mais. Agradeço a vocês por tudo.

Ao meu orientador, Prof. Dr. Walter Filgueira de Azevedo Jr., por todo o conhecimento acadêmico-científico que me proporcionou ao longo desses quatro anos de convivência. Obrigado por mostrar que ciência é algo divertido e que, mesmo sem perder o foco e os resultados, pode ser tratada com extrema leveza. Obrigado pelas várias conversas sobre os mais variados assuntos e pela amizade, com certeza fizestes parte da minha formação profissional e pessoal.

Agradeço a todos os professores que fizeram parte da minha trajetória até aqui, em especial aos que pude trabalhar durante a graduação: Dra. Eva Chagas, Me. Berenice Rosito, Dra. Melissa Simões, Dra. Eliane Santarém e Dr. Júlio César Bicca-Marques. A paixão com a qual vocês trabalham diante nossos olhos, me fizeram encantar pela Biologia e pela arte de ser professor.

Aos meus sogros, Luciana e Régis, o meu agradecimento por todos os conselhos e ajuda dados desde sempre. A amizade de vocês, com certeza, foi uma base de apoio a quem sempre pude recorrer. Serei eternamente grato.

À minha esposa, Nathália, não posso dizer que agradeço por tudo, pois ainda sim seria pouco. Agradeço, em especial, pelo amor incondicional todos os dias, pelas madrugadas passadas em claro durante a escrita dessa dissertação e pela paciência durante os dias nem tão bons assim. A tua energia, positividade e garra me revigoram todos os dias para seguir em frente. Muito obrigado por me ensinar tanto nesses seis anos lado a lado, espero poder retribuir tudo isso. Te amo.

À CAPES pela bolsa concedida e ao PPGBCM pela oportunidade de realização do projeto.

RESUMO

Microrganismos, em geral, apresentam-se como os principais agentes de doenças em seres humanos. Dados do Ministério da Saúde Brasileiro demonstram doenças bacterianas como as principais causas de morte no país. Na terapêutica desses organismos, os antibióticos são considerados os métodos quimioterápicos de maior sucesso da medicina do século XXI, pois representam a primeira, e única, linha de combate contra doenças bacterianas. O desenvolvimento de novas drogas antibióticas torna-se cada vez mais necessário, uma vez que os índices de resistência bacteriana se tornam mais altos a cada ano. Nesse ponto, a rota metabólica do chiquimato é atraente a esse tipo de pesquisa, por ser considerada uma via essencial para a manutenção desses organismos no ambiente, além de estar ausente em animais. A via é responsável pela formação do corismato, precursor de aminoácidos aromáticos (Phe, Trp e Tyr), ácido fólico e ubiquinonas nos grupos de seres vivos onde está presente. A terceira reação da via biossintética do chiquimato é realizada pela enzima DHQD. Nesse passo é realizada a desidratação reversível da molécula DHQ visando transformá-la em 3-desidrochiquimato, reação foco desse estudo. Na busca por novos inibidores de DHQD foram realizadas simulações de *docking* de pequenos ligantes contra a estrutura tridimensional de uma proteína alvo, pois é um processo onde se visa encontrar, entre as possíveis orientações/conformações de um ligante no sítio ativo, aquela que apresenta a menor energia de ligação e, conseqüente, maior finidade. Além das simulações de *docking*, foram realizados métodos de Aprendizagem de Máquina na formulação de funções escores polinomiais, a partir de funções escores presentes no MVD, que fossem capazes de prever a afinidade entre proteína/ligante. Ao final de todas as simulações e testes realizados ao longo do projeto, chegamos à conclusão de que a equação *Polyscore56* apresentou-se como a mais hábil para prever a afinidade entre o sítio ativo de DHQD com compostos testados. Para esse polinômio os resultados de *test set* ($\rho = 0,900$; $p\text{-value} = 0,037$), AUC (74,686%), EF1 (540) e EF2 (159,23) foram, na maioria das categorias avaliadas, os melhores, confirmando as hipóteses formuladas sobre a equação e indicando-a para estudos posteriores com a enzima.

Palavras-Chave: DHQD. Aprendizado de Máquina. *Docking*.

ABSTRACT

Microorganisms, in general, are the major agents of disease in humans. Data from the Brazilian Ministry of Health show bacterial diseases as the main causes of death in the country. In the therapy of these organisms, antibiotics are considered the most successful chemotherapy methods of 21st century medicine, as they represent the first, and only, line of combat against bacterial diseases. The development of new antibiotic drugs is becoming increasingly necessary, as bacterial resistance rates become higher each year. At this point, shikimate pathway is attractive to this type of research, since it is considered an essential pathway for the maintenance of these organisms in the environment, besides being absent in animals. The pathway is responsible for the formation of chorismate, precursor of aromatic amino acids (Phe, Trp and Tyr), folic acid and ubiquinones in the groups of organisms that presents it. The third reaction of the shikimate biosynthetic pathway is performed by the enzyme DHQD. In this step the reversible dehydration of the DHQ molecule is performed aiming to transform into 3-dehydroshikimate, the focus reaction of this study. In the search for new DHQD inhibitors, docking simulations were performed against the three-dimensional structure of a target protein, since it is a process that seeks to find, among possible orientations/conformations of a ligand in the active site, the one that presents the lower binding energy and, consequently, greater affinity. In addition to the docking simulations, machine learning methods were used to formulate polynomial scoring functions, based on the MVD scoring functions, which were able to predict protein/binder affinity. At the end of all the simulations and tests carried out throughout the project, we conclude that the *PolScore56* equation was the most skilled to predict the affinity between the active site of DHQD and tested compounds. For this polynomial, the results of *test set* ($\rho = 0,900$; $p\text{-value} = 0,037$), AUC (74,686%), EF1 (540) and EF2 (159,23) were, in most of the categories evaluated, the best, confirming the formulated hypotheses on the equation and indicating it for further studies with the enzyme.

Keywords: DHQD. Machine Learning. Docking.

LISTA DE ABREVIATURAS

MS: Ministério da Saúde Brasileiro

OMS: Organização Mundial da Saúde

E4F: Eritrose-4-fosfato

FEP: Fosfoenolpiruvato

Phe: Fenilalanina

Trp: Triptofano

Tyr: Tirosina

DAHPS: 3-Deoxi-D-Arabino-Heptulosanato-7-Fosfato Sintase

DAH: 3-Deoxi-D-Arabino-Heptulosanato-7-Fosfato

DHQS: 3-Desidroquinase Sintase

DHQ: 3-Desidroquinato

NAD: Nicotinamida Dinucleotídeo

NADH: Nicotinamida Dinucleotídeo Hidreto

DHQD: 3-Desidroquinato Desidratase

DHQDI: 3-Desidroquinato Desidratase Tipo I

DHQDII: 3-Desidroquinato Desidratase Tipo II

DHS: 3-Desidrochiquimato

SDH: Chiquimato 5-Desidrogenase

NADPH: Nicotinamida Adenina Dinucleotídeo Fosfato Hidreto

SK: Chiquimato Quinase

S3P: Chiquimato-3-Fosfato

SKI: Chiquimato Quinase Tipo I

SKII: Chiquimato Quinase Tipo II

EPSPS: 5-Enolpiruvilchiquimato-3-Fosfato Sintase

EPSP: 5-enolpiruvilchiquimato-3-fosfato

CS: Corismato Sintase

FMN: Flavina Mononucleotídeo

CSB: Biologia de Sistemas Computacional

Ki: Constante Inibitória

PDB: *Protein Data Bank*

EA: Algoritmo Evolucionário

FE: Função Escore

FEE: Função Escore Empírica

RMSD: Desvio Médio Quadrático

VS: *Virtual Screening*

MVD: *Molegro Virtual Docker*

P29: Protocolo 29

ρ : Coeficiente de Correlação de Spearman

R^2 : Coeficiente de Correlação de Spearman ao Quadrado

DA: Acurácia do *Docking*

SAnDReS: *Statistical Analysis of Docking Results and Development of Scoring Functions*

SML: Aprendizado de Máquina

AUC: Área sobre a Curva

EF: Fator de Enriquecimento

ROC: Características Operacionais do Receptor

LISTA DE FIGURAS

Figura 1. Número de óbitos ocasionados por doenças infecciosas no Brasil no período entre 2006 e 2014.....	14
Figura 2. Fatores relacionados aos agentes antimicrobianos responsáveis com a efetividade do composto no combate aos patógenos.....	16
Figura 3. Via metabólica do chiquimato.....	18
Figura 4. Estrutura cristalográfica da enzima DAHPS(PHE) de <i>E. coli</i> (Difração de Raios X (2,8 Å)).....	20
Figura 5. Estrutura cristalográfica da enzima DHQS de <i>V. cholerae</i> (Difração de Raios X (2,5 Å)).....	21
Figura 6. Estrutura cristalográfica da enzima DHQDII de <i>M. tuberculosis</i> (Difração de Raios X (1,5 Å)).....	22
Figura 7. Caracterização do sítio ativo de DHQDII (PDB: 2Y71) complexada ao ligante ((1R,4S,5R)-1,4,5-trihydroxy-3-[(5-methyl-1-benzothiophen-2-yl)methoxy]cyclohex-2-ene-1-carboxylic acid).....	23
Figura 8. Mapa de interações ocorrentes no complexo proteína-ligante.....	24
Figura 9. Estrutura cristalográfica da enzima SDH de <i>E. coli</i> (Difração de Raios X (1,5 Å)).....	25
Figura 10. Estrutura cristalográfica da enzima SK de <i>M. tuberculosis</i> (Difração de Raios X (2,3 Å)).....	26
Figura 11. Estrutura cristalográfica da enzima EPSPS de <i>E. coli</i> (Difração de Raios X (1,5 Å)).....	27
Figura 12. Estrutura cristalográfica da enzima CS de <i>A. aeolicus</i> (Difração de Raios X (2,05 Å)).....	28

Figura 13. Etapas realizadas nos processos de Biologia de Sistemas e Computação Bioinspirada (Docking).....	33
Figura 14. Gráficos de dispersão das três equações que apresentaram os melhores resultados de test set em relação ao $\log(k_i)$	44
Figura 15. Curvas ROC para os dois melhores resultados de AUC.....	46

LISTA DE TABELAS

Tabela I. Resultados de <i>re-docking</i> com o valor de <i>RMSD</i> para Função Escore.....	40
Tabela II. Resultados de <i>Ensemble Docking</i> com o valor de <i>RMSD</i> para função escore.....	41
Tabela III. Correlação entre Funções Escores (MVD) e afinidade proteína/ligante ($\log(K_i)$).....	42
Tabela IV. Resultado de correlação para o <i>training set</i> e <i>test set</i>	43
Tabela V. Resultados e análises para AUC e EF.....	45

SUMÁRIO

1 INTRODUÇÃO	15
1.1 ÍNDICES DE DOENÇAS INFECCIOSAS NO BRASIL E NO MUNDO	15
1.2 ANTIBIÓTICOS E RESISTÊNCIA BACTERIANA	16
1.2.1 Mecanismos de Ação do Antibióticos	16
1.2.2 Resistência Bacteriana à Antibióticos.....	17
1.3 VIA METABÓLICA DO CHIQUIMATO	18
1.3.1 Deoxi-D-Arabino-Heptulosanato-7-Fosfato Sintase (DAHPS) (EC 2.5.1.54)	20
1.3.2 3-Desidroquinase Sintase (DHQS) (EC 4.2.3.4)	21
1.3.3 3-Desidroquinato Desidratase (DHQD) (EC 4.2.1.10).....	22
1.3.4 Chiquimato 5-Desidrogenase (SDH) (EC 1.1.1.25)	25
1.3.5 Chiquimato Quinase (SK) (EC 2.7.1.71)	26
1.3.6 5-Enolpiruvilchiquimato-3-Fosfato Sintase (EPSPS) (EC 2.5.1.19)	27
1.3.7 Corismato Sintase (CS) (EC 4.2.3.5)	28
1.4 BIOLOGIA DE SISTEMAS COMPUTACIONAL E COMPUTAÇÃO BIOINSPIRADA	29
1.5 <i>DOCKING</i> MOLECULAR.....	32
2 JUSTIFICATIVA.....	35
3. OBJETIVO GERAL E OBJETIVOS ESPECÍFICOS.....	37
3.1 OBJETIVO GERAL	37
3.2 OBJETIVOS ESPECÍFICOS.....	37
4. ARTIGOS CIENTÍFICOS	38
4.1 ARTIGO PRINCIPAL (DADOS DA PESQUISA).....	38
4.2 ARTIGOS DESENVOLVIDOS DURANTE O PERÍODO DO MESTRADO	39

5. RESULTADOS E DISCUSSÃO	41
5.1 <i>PRE-DOCKING</i> E <i>RE-DOCKING</i>	
5.2 <i>ENSEMBLE DOCKING</i>	42
5.3 FUNÇÕES ESCORE	42
5.4 FUNÇÕES ESCORE POLINOMIAIS	43
5.5 RECONHECIMENTO DE VERDADEIROS E FALSOS LIGANTES	45
6. CONSIDERAÇÕES FINAIS	48
7. REFERÊNCIAS.....	49
ANEXO A.....	57

1 INTRODUÇÃO

1.1 Índices de doenças infecciosas no Brasil e no Mundo

Microrganismos, em geral, apresentam-se como os principais agentes de doenças ocasionadas em seres humanos. Dados de 2016 (Dezembro) do Ministério da Saúde Brasileiro (2016) (Figura 1) demonstram que entre as quatro doenças infecciosas e parasitárias que mais causaram óbitos no Brasil, as doenças bacterianas figuram na terceira e quarta posição (Tuberculose; Diarreia/Gastroenterite), representando mais da metade das mortes, por esse tipo de patógeno, no ano. Em relação ao número de mortes de crianças, até 4 anos de idade, vinculadas às doenças gerais, o quadro é pior quando são consideradas doenças bacterianas. Infecções respiratórias, sepse em recém-nascidos, diarreia e meningite figuram entre as dez principais causas de morte. Quando somados, o número de óbitos das quatro patologias é maior do que a principal causa, anomalias congênitas (Organização Mundial da Saúde (OMS), 2016).

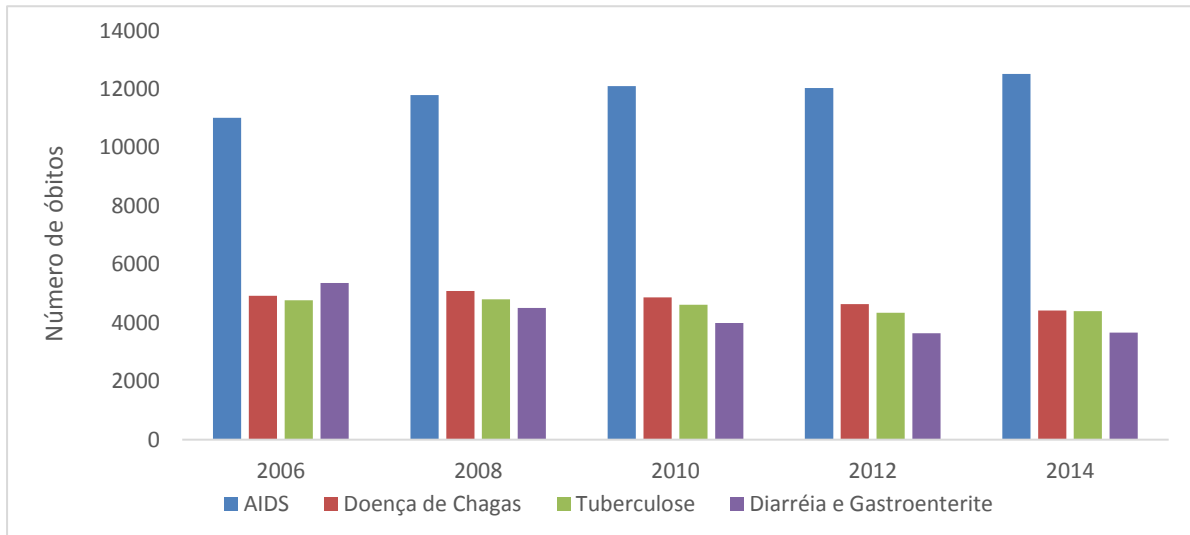


Figura 1. Número de óbitos ocasionados por doenças infecciosas no Brasil no período entre 2006 e 2014. (Dados: Ministério da Saúde).

Em níveis mundiais faltam dados atuais sobre a quantidade de mortes relacionadas à cada doença, porém dados da década passada (2000 – 2012 para causas de morte em geral) trazem doenças infecciosas respiratórias e diarreia entre as dez doenças com maior número de óbitos no mundo (OMS, 2012). É importante

lembrar que infecções respiratórias e diarreias podem ter vários patógenos como seus agentes, porém não há dados percentuais que relacionem o número de mortes ao microrganismo causador.

1.2 Antibióticos e Resistência Bacteriana

1.2.1 Mecanismos de Ação do Antibióticos

Os antibióticos são considerados os métodos quimioterápicos de maior sucesso da medicina do século XXI (Venter, *et al.* 2017), pois representam a primeira, e única, linha de combate contra doenças infecciosas ocasionadas por bactérias. São considerados os únicos agentes farmacológicos que não atingem somente o paciente que os ingere, mas, também, no ambiente dos hospitais e postos de saúde ao interferir nas populações bacterianas por meio da seleção de cepas resistentes/modificadas aos quimioterápicos (ANVISA). A utilização clínica de agentes antimicrobianos apresenta-se de duas formas: profilática e terapêutica. A ação profilática é utilizada com o objetivo de prevenir uma infecção inicial (tratamento primário), ou na prevenção de infecção recorrente ou reativada (tratamento secundário) (Enzler, *et al.* 2011). O segundo modo de tratamento, terapêutico, é direcionado para infecções bacterianas agudas, tendo como objetivo a eliminação do patógeno e, conseqüente, término da infecção e danificação do tecido contaminado (Leekha, *et al.* 2011).

Os agentes antimicrobianos possuem um foco molecular de atividade no interior da célula, voltando-se para vias metabólicas essenciais ou na desestabilização, direta ou indireta, de importantes estruturas celulares, como a membrana plasmática, parede celular e ácidos nucleicos (DNA e RNA). Atualmente, existem treze classes de antibióticos arranjadas em cinco grupos com base no seu mecanismo de ação na célula bacteriana (ANVISA), são os seguintes: impedimento da síntese de peptidoglicanos da parede celular; ruptura da membrana plasmática; impedimento da síntese proteica (o composto pode se ligar ao RNAr, ou ao RNAt); desestabilização indireta do DNA (pela formação de radicais livres), ou bloqueio da transcrição (impede o correto funcionamento da RNA-polimerase); inibição competitiva de enzimas presentes na via de síntese do ácido fólico e de purinas. A atividade microbiana dos quimioterápicos depende diretamente de quatro fatores: concentração ideal no local de infecção, adentrar o citoplasma da célula (passando

pela parede celular e membrana plasmática), afinidade pelo alvo de atividade do composto (citados acima) e tempo de permanência na célula-alvo (ANVISA) (Figura 2).

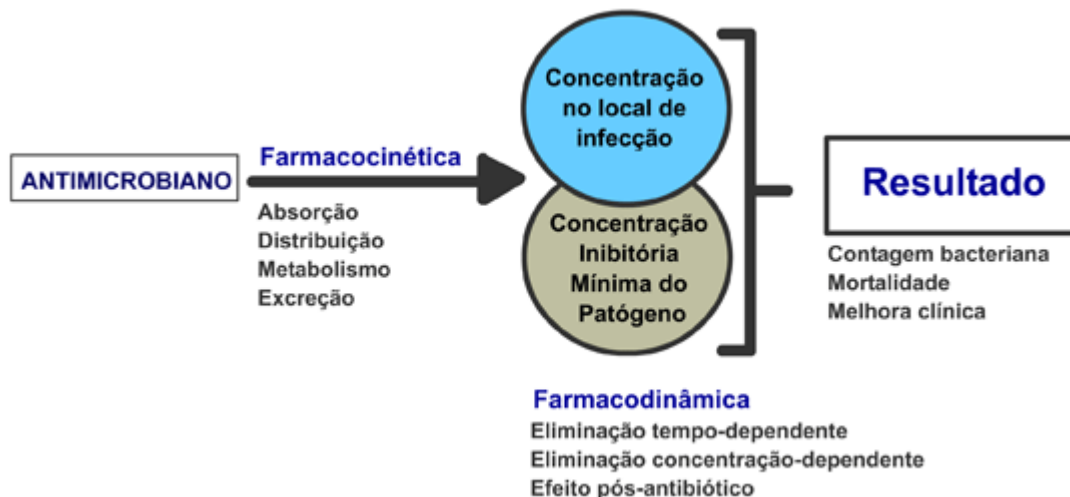


Figura 2. Fatores relacionados aos agentes antimicrobianos responsáveis com a efetividade do composto no combate aos patógenos. (Fonte: ANVISA).

1.2.2 Resistência Bacteriana à Antibióticos

Pode-se considerar a resistência bacteriana como uma rede molecular interligada que confere proteção contra os diversos compostos presentes nos antimicrobianos (Wright, 2007) que começou a ser observada a partir da descoberta da penicilina (1940) e sua utilização desenfreada, sem controle. Os mecanismos de resistência mais comuns são: efluxo dos compostos antimicrobianos por porinas presentes na membrana plasmática e parede celular; mutações gênicas que alteram o sítio de ligação de algumas enzimas específicas, diminuindo a afinidade pelos quimioterápicos. Certos grupos bacterianos podem apresentar casos mais específicos de resistência, como por exemplo: bactérias gram-negativas apresentam outra membrana externa (lipopolissacarídeo-fosfolipídica) formando uma barreira física contra a entrada dos compostos; produção de enzimas que destroem os compostos químicos fazendo com que percam a sua atividade antimicrobiana, caso das β -lactamases (destroem o anel β -lactâmico presente na penicilina e em seus derivados) e aminoglicosídeo quinase (grupo de enzimas que inativam compostos antibióticos através da fosforilação em região específica). Wright, 2005 salienta que todos os mecanismos de resistência bacteriana requerem uma reprogramação genética do

organismo e dispõem de alta quantidade de energia para a produção de enzimas de resistência, ou para a reorganização ocorrida em estruturas/moléculas fundamentais da célula (membrana plasmática, parede celular, ribossomo, RNAr, etc.).

A Organização Mundial da Saúde (OMS, 2016) classifica a resistência bacteriana como um dos principais problemas de saúde mundial na atualidade, pois esse mecanismo de defesa gera dificuldades no tratamento de certas doenças, podendo ocasionar, em alguns casos, a morte dos pacientes, além do aumento de gastos com hospitais e saúde pública em geral. Estima-se que na União Europeia vinte e cinco mil pessoas morrem a cada ano em decorrência de infecções ocasionadas por bactérias multirresistentes, gerando um gasto de, aproximadamente, 1,5 bilhões de dólares anualmente. Estima-se que se providências não forem tomadas para a diminuição de cepas resistentes aos compostos utilizados atualmente, nos próximos trinta e cinco anos, 300.000 pessoas morrerão em decorrência de infecções por bactérias resistentes à antibióticos (EMA-UE, 2016).

1.3 Via Metabólica do Chiquimato

A rota metabólica do ácido chiquímico está presente em larga variedade de organismos, como bactérias, plantas e protistas (Apicomplexa) (Bentley & Haslam 1990). A via metabólica é constituída de sete passos, apresentando uma grande diversidade na função enzimática dos três grupos em que ocorre. Em bactérias, foco desse estudo, as sete reações enzimáticas são realizadas por sete enzimas monofuncionais codificadas por genes diferentes (Millar & Coggins, 1986) (Figura 3). As plantas apresentam um complexo bifuncional que realiza as atividades enzimáticas da terceira e quarta enzima (Herrmann, 1995). Nos fungos as enzimas que realizam a primeira e a última reação da via estão conservadas, porém as funções da segunda até a sexta enzima são realizadas por um complexo proteico multifuncional, chamado AROM (Hawkins & Smith, 1991).

Os substratos iniciais que dão início à rota são a Eritrose-4-fosfato (E4F) e o Fosfoenolpiruvato (FEP), ambos provindos de vias catabólicas de carboidratos. Roberts *et al.* 2002 demonstrou a importância dessa via metabólica comprovando que 20% dos “carbonos” produzidos no catabolismo de carboidratos são direcionados para

a utilização nessa via. O objetivo final das reações é a formação do corismato, precursor de aminoácidos aromáticos (Fenilalanina (Phe), Triptofano (Trp) e Tirosina (Tyr)) ácido fólico e ubiquinonas nos grupos de seres vivos onde está presente.

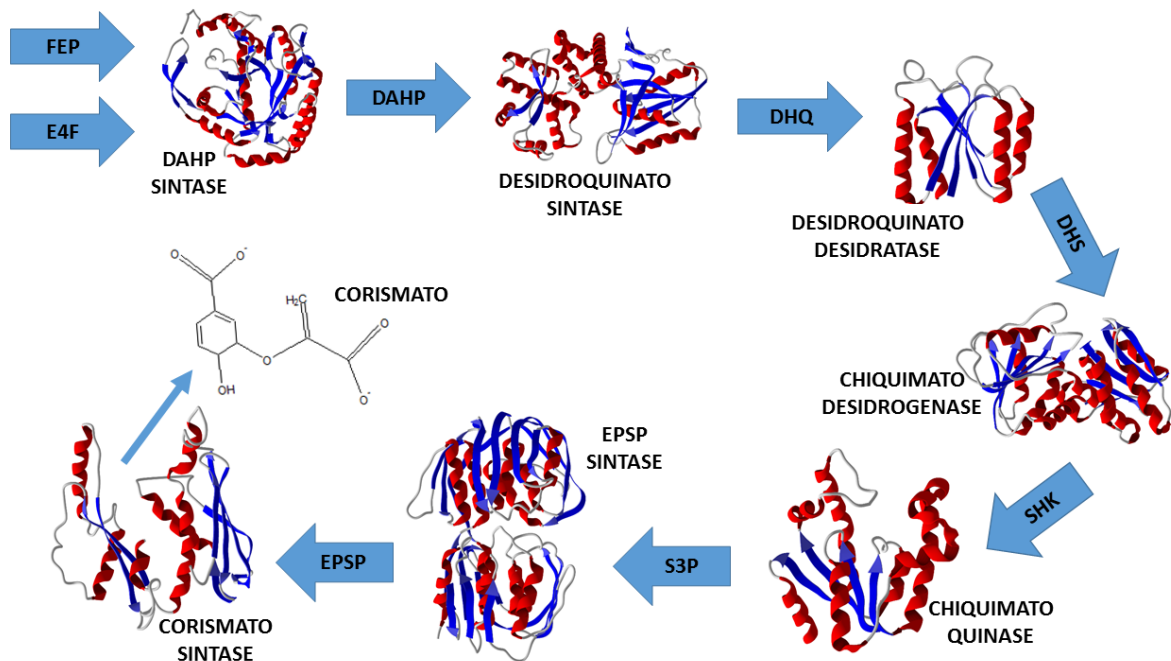


Figura 3. Via metabólica do chiquimato em bactérias. (Abreviações: FEP, fosfoenolpiruvato; E4F, eritrose-4-fosfato; DAHP, 3-deoxi-D-arabino-heptulosanato-7-fosfato; DHQ, 3-desidroquinato; DHS, 3-desidrochiquimato; SHK, chiquimato; S3P, chiquimato-3-fosfato; EPSP, 5-enolpiruvilchiquimato-3-fosfato).

Em experimento realizado por Maeda & Dudareva, 2013 foi possível comprovar a essencialidade da rota para bactérias, uma vez que mutações induzidas nos genes das enzimas presentes na via inibiram o crescimento bacteriano. A rota já é utilizada como alvo de um dos principais herbicidas presentes no mercado, o glifosato. Inibidor competitivo da sexta (penúltima) enzima da rota, liga-se ao sítio ativo da enzima impedindo que ocorra a reação, provocando, assim, a morte da planta. Existem alguns resultados (Roberts *et al.* 1998) que trazem sucesso na utilização de glifosato como tratamento quimioterápico terapêutico em ratos infectados com *Toxoplasma gondii* e *Plasmodium* sp., mas ainda há necessidade de mais estudos para comprovar a eficácia contra os dois patógenos. Uma vez que a via é considerada essencial para a manutenção desses organismos no ambiente, existem indícios que qualquer uma das enzimas participantes seja um promissor alvo para o desenho de

novos fármacos, apresentando-se como uma saída para a resistência bacteriana. Além disso, outro bom motivo para o estudo dessas enzimas, é a ausência da rota em humanos, o que diminui as chances de efeitos colaterais de quimioterápicos que objetivam a interrupção da produção de corismato.

1.3.1 Deoxi-D-Arabino-Heptulosanato-7-Fosfato Sintase (DAHPS) (EC 2.5.1.54)

DAHPS (Figura 4) é a primeira enzima atuante na rota do ácido chiquímico. A reação mediada pela enzima é a condensação de FEP e E4F gerando como produto o composto 3-deoxi-D-arabino-heptulosanato-7-fosfato (DAHP). A reação intermediada pela enzima é do tipo condensação aldol gerando como produto final uma cadeia cetocarbônica ácida de sete carbonos.

Durante a reação, um fosfato inorgânico (Pi) é retirado da molécula de FEP (DAHP) (Mir *et al.* 2013). A DAHPS pode apresentar-se em três isoformas distintas: DAHPS(PHE), um homotetrâmero codificado pelo gene *aroG*, é a forma mais abundante representando $\approx 80\%$ de DAHPS presente na célula; DAHPS(TYR) e DAHPS(TRP) são homodímeros codificados, respectivamente, pelos genes *aroF* e *aroH* representando $\approx 20\%$ e $\approx 1\%$ da atividade enzimática. Essas variadas formas indicam um mecanismo de inibição por retroalimentação pelos aminoácidos que são reconhecidos pelos seus sítios alostéricos. As cadeias polipeptídicas das três isoformas apresentam um número similar de aminoácidos demonstrando um alto grau de identidade entre as sequências. DAHPS(PHE) apresenta 350 resíduos, DAHPS(TYR) 356 resíduos e DAHPS(TRP) 348 resíduos, com 41% dos resíduos idênticos (Shumilin *et al.* 1999).

Esse alto grau de similaridade sugere a comum origem evolucionária para as três isoenzimas. Esse processo ocorreu, como suposto por Wallace & Pittard, 1967, através de duplicação gênica seguida por mutações que atuaram na formação de proteínas sensíveis a diferentes mecanismos de controle alostérico. DAHPS é uma metaloenzima dependente de íons divalentes, não estando bem definido se são necessários na atividade catalítica, ou na manutenção da estrutura proteica. Os íons, documentados *in vitro*, são: Mn^{2+} , Fe^{2+} , Cd^{2+} , Co^{2+} , Cu^{2+} , Ni^{2+} , Zn^{2+} e Ca^{2+} .

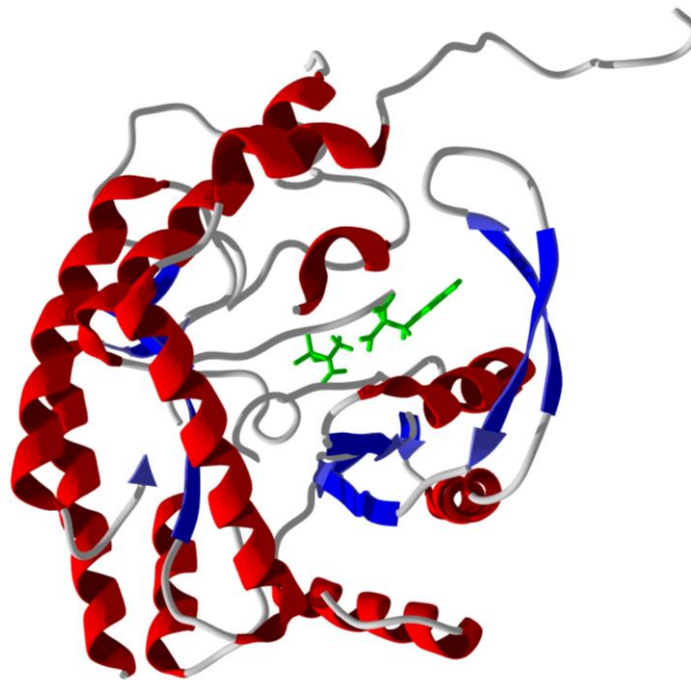


Figura 4. Estrutura cristalográfica da enzima DAHPS(PHE) de *E. coli* (Difração de Raios X (2,8 Å)). Em vermelho estão representadas as oito α -hélices e em azul as oito fitas- β . Em cinza estão representados os *loops*. Em verde à esquerda encontra-se FEP e à direita PHE. Código de acesso PDB: 1KFL (cadeia A) (Shumilin *et al.* 2002).

1.3.2 3-Desidroquinase Sintase (DHQS) (EC 4.2.3.4)

DHQS (Figura 5) é a segunda enzima presente na via metabólica do ácido chiquímico. Ela catalisa a reação de conversão de DAHP em 3-Desidroquinato (DHQ) num processo de cinco reações onde há a formação de um anel carbocíclico que será utilizado na produção do anel benzênico presente nos aminoácidos aromáticos (Chandran & Frost, 2001). O processo é dependente de Nicotinamida Dinucleotídeo (NAD) para a realização de suas reações. A primeira das cinco reações realizadas no sítio ativo da proteína é a redução de NAD^+ para Nicotinamida Dinucleotídeo Hidreto (NADH). Após as reações intermediárias, onde ocorre a transformação da molécula de DAHP em DHQ, o NAD é reoxidado, processo essencial para que a DHQ seja liberada do sítio de ligação com a proteína. As cinco reações mediadas pela enzima têm como objetivo a eliminação do grupamento fosfato da molécula DAHP para a formação de DHQ (Herrmann & Weaver, 1999). DHQS possui 362 resíduos em sua cadeia polipeptídica formando um homodímero. É uma metaloenzima, apresentando afinidade com dois cátions divalentes, Co^{2+} e Zn^{2+} , presentes e atuantes diretos no sítio ativo da proteína (Chandran *et al.* 2001).

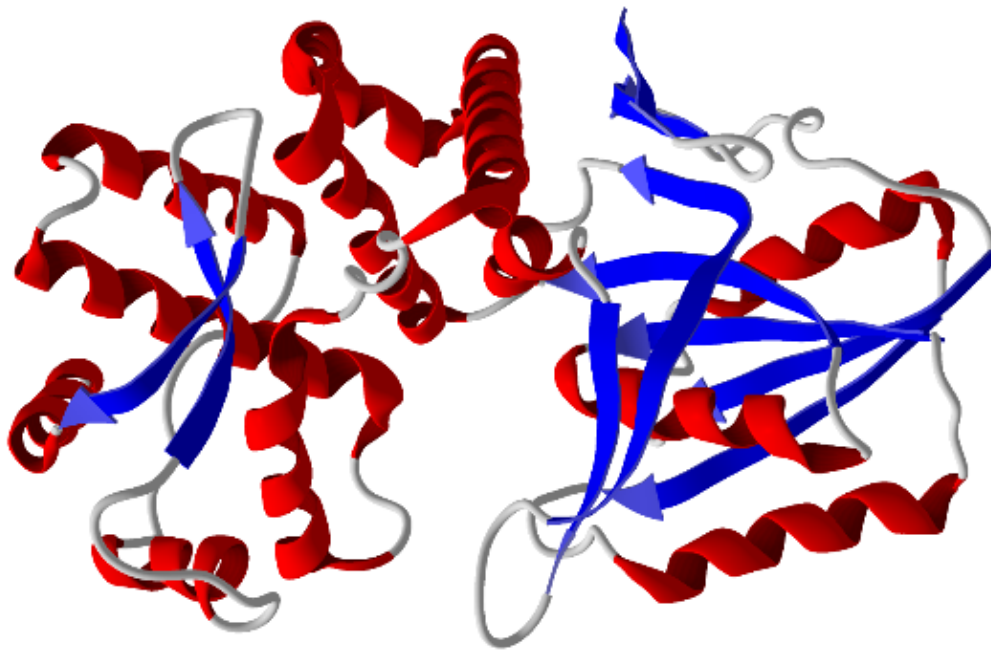


Figura 5. Estrutura cristalográfica da enzima DHQS de *V. cholerae* (Difração de Raios X (2,5 Å)). Em vermelho estão representadas as dez α -hélices e em azul as dez fitas- β . Em cinza estão representados os *loops*. Código de acesso PDB: 3OKF (cadeia A) (Minasov *et al.* 2010).

1.3.3 3-Desidroquinato Desidratase (DHQD) (EC 4.2.1.10)

A terceira reação da via biossintética do chiquimato é realizada pela enzima DHQD. Nesse passo é realizada a desidratação reversível da molécula DHQ visando transformá-la em 3-desidrochiquimato (DHS). A enzima também participa da rota do quinato, via responsável por transformar o ácido quínico em diversos produtos, como o acetil-CoA e succinil-CoA, participantes do Ciclo de Krebs (Roszak *et al.* 2002). Grande parte das espécies bacterianas apresenta duas isoformas da enzima, Tipo I (DHQDI) e Tipo II (DHQDII). As duas enzimas diferem em estrutura, mas catalisam o mesmo tipo de reação, desidratação do C2 presente no anel carbocíclico inserindo a primeira ligação dupla para futura formação do anel aromático. DHQDI apresenta estrutura de barril com oito fitas- β circundadas por oito α -hélices, em contraste, DHQDII é formada por uma folha- β (cinco fitas- β) flanqueadas por 4 α -hélices (Figura 6) (Gourley *et al.* 1999). As duas isoformas apresentam, em média, 252 resíduos, formando um homodímero.

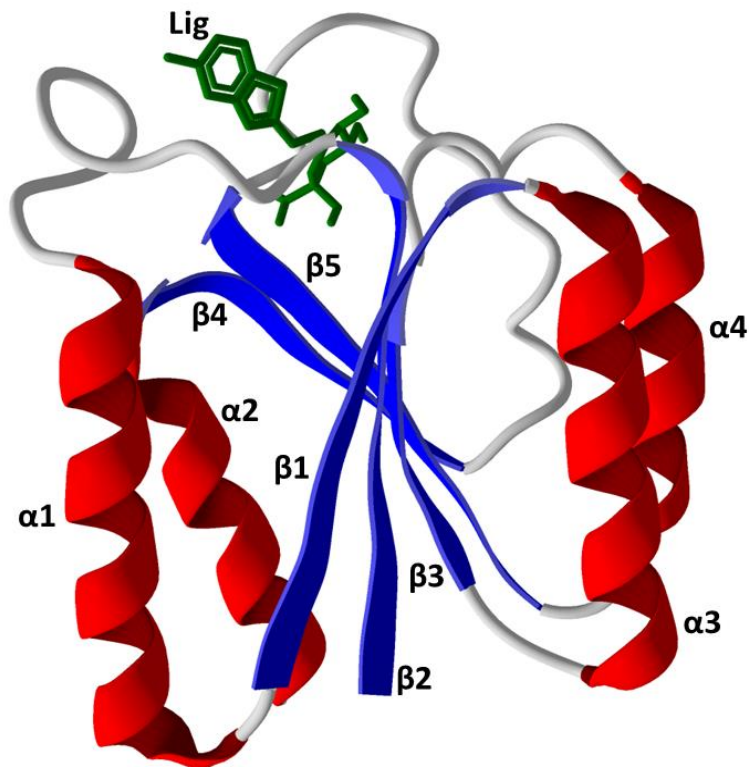


Figura 6. Estrutura cristalográfica da enzima DHQDII de *M. tuberculosis* (Difração de Raios X (1,5 Å)). Em vermelho estão representadas as quatro α -hélices e em azul as cinco fitas- β . Em cinza estão representados os *loops*. Código de acesso PDB: 2Y71 (cadeia A) (Tizon *et al.* 2011). (Lig: ligante; α 1: Thr28 – Gly44; α 2: Ile130 – His143; α 3: Ser54 – Ala68; α 4: Ser83 – Glu92; β 1: Lys46 – Asp53; β 2: Leu3 – Pro11; β 3: Pro71 – Asn75; β 4: Ala95 – Ser103; β 5: Thr122 – Gly129).

DHQDI está associada exclusivamente com a biossíntese do corismato estando presente em fungos, plantas e algumas bactérias. Catalisa a desidratação por eliminação *syn* necessitando da utilização de um intermediário imina (Harris, *et al.* 1993). Já DQHDII atua nas duas rotas, chiquimato (bactérias) e quinato (fungos), catalisando a desidratação através de eliminação *anti* com a utilização de enolato como intermediário (Harris, *et al.* 1996). As duas isoformas apresentam alto grau de conservação de alguns aminoácidos essenciais para reações ocorrentes no sítio ativo (Mir *et al.* 2013). Os resíduos conservados são Lisina, Histidina e Glutamato. Como foram encontrados estudos caracterizando o sítio ativo somente para DHQDII, essa foi a isoforma da proteína utilizada no estudo. Para essa isoenzima os resíduos encontrados no sítio ativo foram: Asn75; His81; His101; Ile102; Ser103; Arg112 (Figura 7).

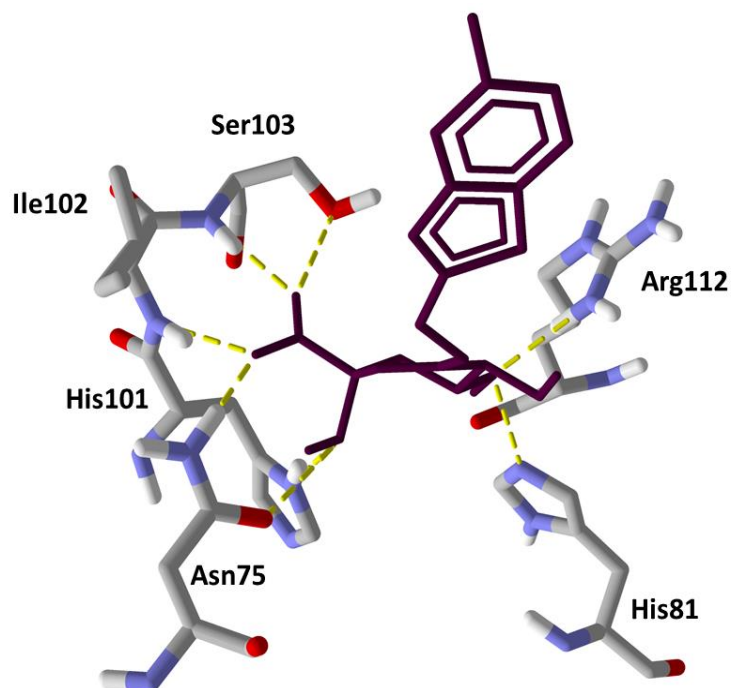


Figura 7. Caracterização do sítio ativo de DHQDII complexada ao ligante ((1R,4S,5R)-1,4,5-trihydroxy-3-[(5-methyl-1-benzothiophen-2-yl)methoxy]cyclohex-2-ene-1-carboxylic acid). Código de acesso PDB: 2Y71 (Tizon *et al.* 2011). (Púrpura: Ligante; vermelho: oxigênio; azul: nitrogênio; cinza escuro: carbono; cinza claro: hidrogênios).

A figura 8 traz o mapa de interações ocorridas entre a proteína e o ligante, gerado pelo *software* LigPlot⁺ (Laskowski & Swindells, 2011). Pode-se perceber que, além dos resíduos que formam ligações de hidrogênio com o ligante, ocorrem ligações hidrofóbicas (van der Waals) entre outros resíduos e os átomos formadores do inibidor. É importante o conhecimento desse tipo de interação entre o complexo proteína-receptor, uma vez que os contatos hidrofóbicos agem no reconhecimento molecular e estabilização do complexo. Há uma relação direta entre o número de contatos de van der Waals e o sucesso do inibidor testado, pois como a maioria dos ligantes são competitivos pelo sítio ativo da enzima, quanto mais estável for a relação entre inibidor e enzima, menor será a chance de o complexo desestruturar-se. Os resíduos participantes dos contatos apolares são: Val105, Gly78, Gly77, Tyr24, Asp88, Asn12, Arg15 e Glu92.

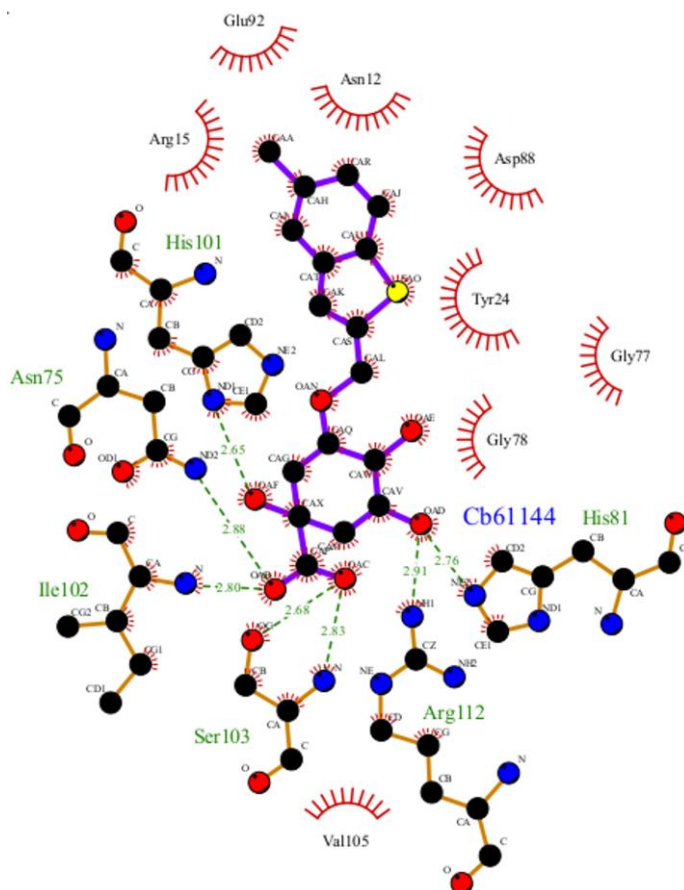


Figura 8. Mapa de interações ocorrentes no complexo proteína-ligante. Código de acesso PDB: 2Y71 (Tizon *et al.* 2011). (Púrpura: ligante (Cb61144); Amarelo: DHQD; Pontilhado verde: ligações de hidrogênio; Meios círculos com riscos vermelhos: contatos hidrofóbicos (van der Waals)).

1.3.4 Chiquimato 5-Desidrogenase (SDH) (EC 1.1.1.25)

A quarta reação da rota metabólica é mediada pela enzima SDH (Figura 9). A molécula catalisa a redução dependente de NADPH da molécula de DHS para chiquimato (SHK) (Peek *et al.* 2014). A enzima pode estar presente na célula como um monômero ou dímero, apresentando duas diferentes formas, aberta e fechada, sendo a última considerada a catalítica (Avitia-Domínguez *et al.* 2014). Peek e colaboradores, 2014 descrevem a existência de cinco enzimas (AroE, Ydib, RifI, SdhL e Ael1) muito similares (20% - 50% de identidade) formando uma família de proteínas, família SDH. No entanto, somente SDH (AroE) foi identificada com propriedades físico-químicas para realizar as reações na rota aqui trabalhada. A enzima apresenta uma alta variabilidade de formas, sendo, essa característica, diretamente relacionada com o organismo analisado. Em *Staphylococcus pidermidis* e *Thermotoga marítima* está presente como monômero, já em *Mycobacterium tuberculosis* e *Thermus*

thermophilus há a formação de dímeros. São encontrados, geralmente, 275 resíduos totais e dois sítios catalíticos, um para DHS e outro para a ligação de NADPH.

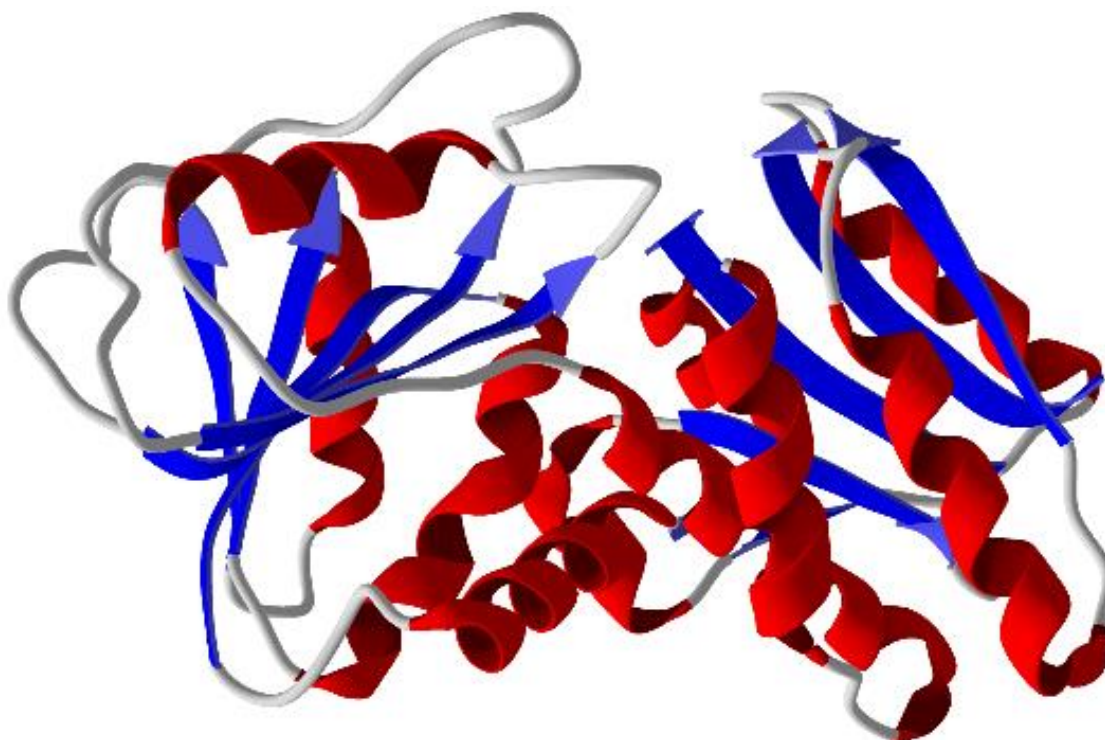


Figura 9. Estrutura cristalográfica da enzima SDH de *E. coli* (Difração de Raios X (1,5 Å)). Em vermelho estão representadas as dez α -hélices e em azul as onze fitas- β . Em cinza estão representados os *loops*. Código de acesso PDB: 1NYT (Cadeia A) (Michel *et al.* 2003).

1.3.5 Chiquimato Quinase (SK) (EC 2.7.1.71)

O quinto passo da via metabólica do chiquimato é a reação de fosforilação da molécula de chiquimato convertendo-a em chiquimato-3-fosfato (S3P). Essa reação é intermediada pela enzima SK (Figura 10) que transfere um fosfato inorgânico (Pi) do doador (adenosina trifosfato) para o terceiro carbono do anel carbocíclico do chiquimato. Em *E. coli*, essa enzima apresenta-se em duas isoformas codificadas por genes diferentes, *aro-K*: Chiquimato Quinase Tipo I (SKI) e *aro-L*: Chiquimato Quinase Tipo II (SKII). SKI apresenta um menor grau de afinidade (K_m) com o substrato, em média de 20mM, enquanto SKII apresenta K_m de 200 μ M (Whipp *et al.* 1995), demonstrando que a segunda isoforma está diretamente envolvida na rota de produção de compostos aromáticos. A enzima pertence à classe das quinases nucleosídicas monofosfato possuindo três domínios conservados em sua estrutura: CORE, onde estão presentes resíduos conservados que interagem com o fosfato; LID, realiza uma mudança conformacional sobre o sítio de ligação da enzima após o

substrato estar interagindo; NMP, responsável pelo reconhecimento e ligação ao substrato específico, no caso, chiquimato (DIAS *et al.* 2007 *apud* Gu *et al.* 2002).

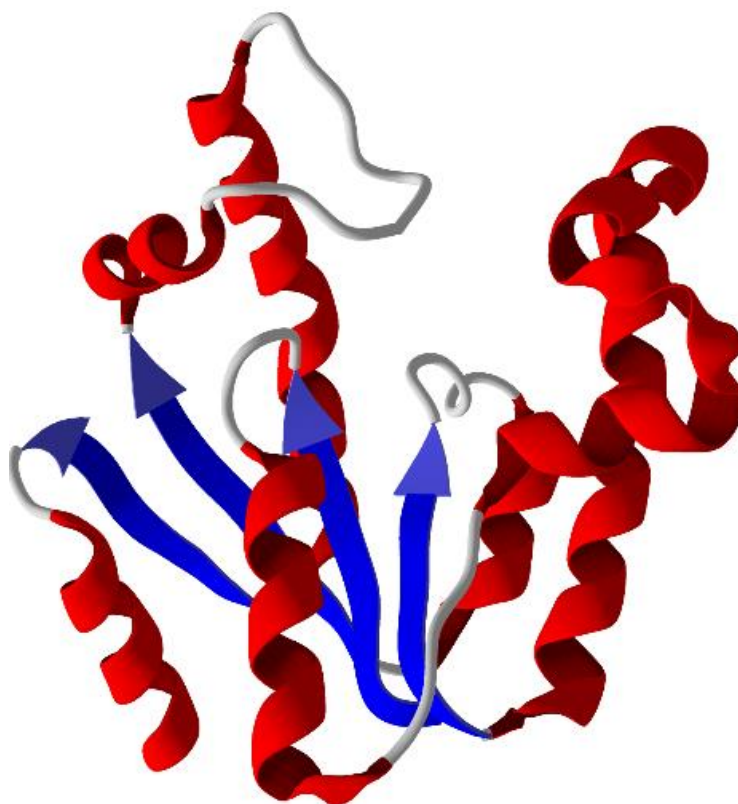


Figura 10. Estrutura cristalográfica da enzima SK de *M. tuberculosis* (Difração de Raios X (2,3 Å)). Em vermelho estão representadas as seis α -hélices e em azul as quatro fitas- β . Em cinza estão representados os *loops*. Código de acesso PDB: 1WE2 (Pereira *et al.* 2005).

1.3.6 5-Enolpiruvilchiquimato-3-Fosfato Sintase (EPSPS) (EC 2.5.1.19)

O sexto passo da via do ácido chiquímico é intermediado pela enzima EPSPS (Figura 11). Durante a reação ocorre a transferência de enolpiruvil, a partir de FEP, à posição 5-hidroxi da molécula de S3P convertendo-a em 5-enolpiruvilchiquimato-3-fosfato (EPSP). Durante as reações de conversão a enzima sofre uma mudança conformacional no momento em que se liga ao C3F, partindo do estado “aberto” para “fechado”. Essa modificação permite a ligação de FEP ao sítio ativo da proteína, formando uma interface entre o C-terminal e o N-terminal (Funke *et al.* 2009), possibilitando as reações de transferência ocorridas no sítio de ligação. A estrutura tridimensional da enzima foi elucidada pela primeira vez por Stallings *et al.* 1991, para *E. coli*, onde observou-se a presença de 427 resíduos de aminoácidos formando um dímero. Para esta enzima já são conhecidos resultados satisfatórios para inibição e parada da rota em plantas. O herbicida glifosato é considerado um inibidor competitivo

(Maeda *et al.* 2013), pois concorre diretamente com o segundo substrato da enzima, FEP, ocupando o sítio ativo, o que impede o acoplamento do ligante natural.

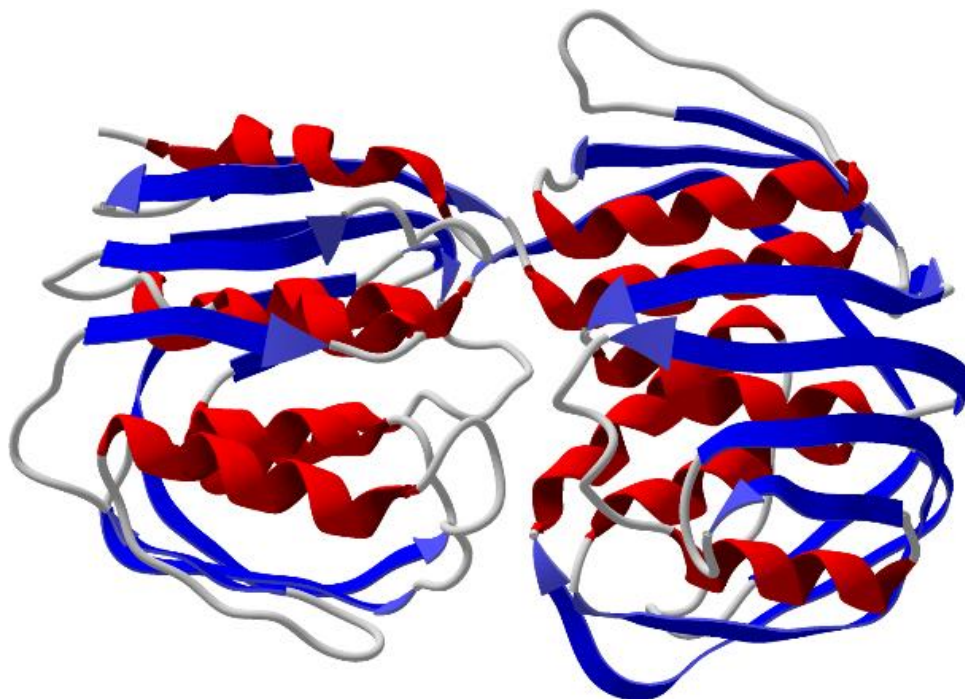


Figura 11. Estrutura cristalográfica da enzima EPSPS de *E. coli* (Difração de Raios X (1,5 Å)). Em vermelho estão representadas as doze α -hélices e em azul as dezenove fitas- β . Em cinza estão representados os *loops*. Código de acesso PDB: 1G6S (Schonbrunn *et al.* 2001).

1.3.7 Corismato Sintase (CS) (EC 4.2.3.5)

CS (Figura 12) é a última e sétima enzima da via de produção do chiquimato. Ela catalisa a reação de conversão de EPSP em corismato, que serve de precursor dos aminoácidos aromáticos, além de outros diversos compostos (Weaver & Herrmann, 1997). A enzima catalisa a reação de eliminação 1,4-trans retirando o fosfato da molécula de EPSP, além de introduzir a segunda ligação no anel aromático presente no corismato (Thomas *et al.* 2003). Para a realização da reação é necessária a ligação de flavina mononucleotídeo (FMN) em um segundo sítio ativo da enzima. CS pode diferir de um organismo para outro, em *E. coli* ela apresenta-se com uma enzima monofuncional dependente de FMN, em *Neurospora crassa* e *Saccharomyces cerevisiae* há um sítio de ligação a mais, onde NADPH e FMN são ligados. Nesses últimos dois organismos a enzima é bifuncional (Ahn *et al.* 2003). Quevillon-Cheruel *et al.* 2003, demonstraram que a enzima apresenta, aproximadamente, 370 resíduos de aminoácidos formando um homotetrâmero.

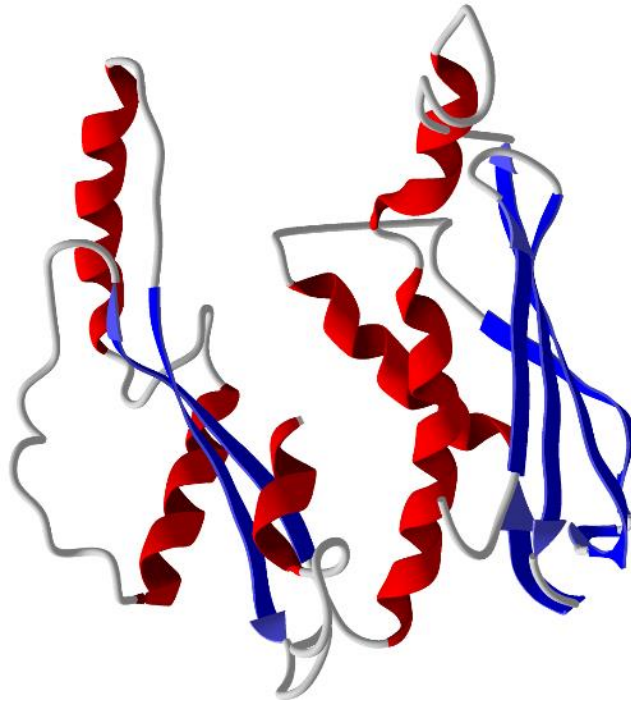


Figura 12. Estrutura cristalográfica da enzima CS de *A. aeolicus* (Difração de Raios X (2,05 Å)). Em vermelho estão representadas as seis α -hélices e em azul as sete fitas- β . Em cinza estão representados os *loops*. Código de acesso PDB: 1Q1L (Viola *et al.* 2003).

1.4 Biologia de Sistemas Computacional e Computação Bioinspirada

Durante anos, os avanços da biologia celular e molecular trouxeram grandes quantidades de dados brutos a respeito das moléculas participantes de importantes processos celulares. Sequências de ácidos nucleicos (DNA e RNA) e proteínas; éxons e íntrons de determinados genes; estruturas secundárias de proteínas e muitos outros dados brutos estão disponíveis em banco de dados (GenBank (Benson *et al.* 1999) e *Protein Data Bank* (PDB) (Berman *et al.* 2000)) prontos para serem conectados e traçarem as relações ocorrentes no interior das células. Kitano, 2002 afirma que a biologia molecular foi capaz de descobrir uma enormidade de fatos biológicos, porém esses dados/fatos sozinhos não são capazes de elucidar, ou dar a correta interpretação, dos sistemas biológicos presentes nos diversos níveis de organização (células, tecidos, órgãos, ecossistemas, etc.). Na tentativa de uma melhor compreensão dos sistemas biológicos e de uma análise holística sobre a célula, o campo da Biologia de Sistemas apresenta-se como uma saída interessante ao problema apresentado.

A Biologia de Sistemas Computacional (CSB) é um ramo da Biologia de Sistemas que se utiliza de técnicas computacionais na tentativa de simular sistemas

biológicos para sua melhor compreensão. CSB é considerada como “Nova Biologia”, pois integra uma série de áreas do conhecimento – biologia, bioinformática, computação científica, engenharias, física, química, entre outras – que realizam uma análise a nível de “sistema”, buscando a solução de problemas biológicos críticos. Kitano, 2002 apresenta dois modos de utilização da CSB na resolução de problemas biológicos: *data-mining*, consiste na identificação de padrões de resultado presentes em grandes bases de dados (ver de Ávila & de Azevedo, 2014); *análises baseadas em simulações*, realização de testes das hipóteses geradas pela *data mining*, utilizando, previamente, análises *in silico* para, posteriormente, realizarem-se experimentos *in vitro* e *in vivo*.

Neste trabalho optamos por utilizar as duas abordagens de CSB citadas acima, pois acreditamos que, assim, tornamos a abrangência do trabalho mais completa. Nosso *data mining*, foi realizado a partir da seleção de estruturas cristalográficas (Raios X) da enzima DHQD que apresentavam fatores de afinidade (constante inibitória (K_i)) entre pequenas moléculas (ligante natural e/ou inibidor) e a enzima em questão. Os dados foram retirados de duas bases de dados on-line, PDB e *BindingDB* (Liu *et al.* 2001). Após essa “filtragem bruta” inicial, foi realizada uma “filtragem fina”, onde foram encontrados os padrões de dados descritos acima por Kitano, 2002. Os critérios de seleção foram: presença de constante de inibição (K_i) e presença de dez moléculas de água em uma esfera teórica centrada no sítio da enzima, com 10 Å de raio.

A segunda abordagem, *análises baseadas em simulações*, foi realizada no sistema de *docking* molecular, que visa o estudo entre a interação proteína-ligante. Essa abordagem encaixa-se, também, no conceito de Computação Bioinspirada que visa o desenvolvimento de algoritmos para a solução de problemas de otimização, podendo esses, serem biológicos. O modelo que melhor se encaixa para a otimização do sistema proteína-ligante é o Algoritmo Evolucionário (EA) (Thomsen & Christensen, 2006). EA's usam as ideias da evolução com o objetivo de gerar abordagens computacionais para tratar problemas de otimização e biológicos. De forma geral, EA's podem ser classificados como algoritmos estocásticos, classe em que se enquadram também Monte Carlo, *Tabu Search* e *Swarm Optimization* (Floreano & Mattiussi, 2008). A ideia evolutiva utilizada nesse tipo de algoritmo serve para “selecionar” as melhores posições geradas ao longo do processo. O sistema de

avaliação utilizado é a energia gerada pelo sistema proteína-ligante, quanto menor for a energia, “mais adaptada” aquela posição está ao sistema. Ao longo da simulação, as posições estão suscetíveis a modificações, como se ocorressem mutações, o que pode melhorar o *fitness* do sistema, ou piorá-lo. Caso haja uma piora (aumento) na energia gerada, a posição é eliminada, se for observada uma melhora (diminuição) a posição é mantida na simulação.

A utilização dos EA's destaca a função como critério de seleção de indivíduos mais bem adaptados. Na análise de resultados da simulação de *docking* molecular, consideramos que os melhores resultados são os de menor energia, assim na aplicação de EA's ao problema de *docking* molecular, o tipo de função a ser utilizada na seleção de melhores resultados será a Função Escore (FE). As FE's são divididas em três grandes famílias. A primeira família é formada por funções baseadas em campos de força, como a FE do programa DOCK (Meng *et al.*, 1992), que se baseia no campo de força AMBER (Weiner *et al.*, 1984). A segunda família é formada por Funções Escores Empíricas (FEE's), originalmente propostas por Böhm (Böhm, 1994; 1998), onde os termos de uma função recebem peso, de forma a concordar com afinidades determinadas experimentalmente. A terceira família são funções baseadas em conhecimento, que usam funções de energia potencial derivadas de estruturas obtidas experimentalmente (Tanaka & Scheraga, 1976; Sippl, 1990).

Nosso foco inicial foi a implementação de FEE's, que são relativamente mais rápidas de calcular que as outras citadas acima. Testamos a adequação de três tipos de FEE's, aqui chamadas de funções escores lineares, não-lineares e não-lineares mistas.

A função escore empírica linear ($\Delta G1$) tem a seguinte forma geral,

$$\Delta G1 = \sum_j w_j \Delta g_j$$

Equação 1. Função escore empírica linear, onde w_j é o peso de cada termo energético Δg_j .

Os pesos relativos de cada termo da somatória acima, são obtidos a partir de ajuste contra um conjunto de estruturas, para as quais há informação estrutural e de afinidade disponíveis nas bases de dados MOAD (*Mother Of All Databases*) (Hu *et al.*, 2005), *BindingDB* e *PDBbind* (Wang *et al.*, 2004). São testados diferentes termos relevantes para interações intermoleculares, tais como ligações de hidrogênio, área

de contato, área de contato polar e hidrofóbica, área acessível ao solvente, interação eletrostática entre outros possíveis (de Azevedo & Dias, 2008).

A FEE não-linear (ΔG_2) admite termos de potenciais mais altas, como $(\Delta g_j)^2$ ou outros expoentes. Os termos serão os mesmos descritos da ΔG_1 , com a diferença do expoente de cada termo. A FEE não-linear mista (ΔG_3) adiciona termos mistos. Por exemplo, as interações de ligação de hidrogênio são representadas pelo termo Δg_1 e as interações de van der Waals por Δg_2 , teremos um termo misto $(\Delta g_1 \Delta g_2)$ na somatória da equação 1, além dos termos lineares da ΔG_1 e não-lineares, já citados para a ΔG_2 .

Os três tipos de famílias de FEE's são ajustadas por classe enzimática, sendo selecionada aquela que apresentar melhor concordância com dados experimentais, usando-se critérios estatísticos, como o coeficiente de correlação de Spearman (Zar, 1972), entre a afinidade prevista pela FEE e a experimental. Teremos dois tipos de bases de dados de afinidade, uma será o conjunto treino (*training set*) com 90 % dos dados experimentais disponíveis sobre afinidade, usados para obter os pesos da equação 1. A segunda base de dados será formada por 10 % dos dados experimentais disponíveis, que não serão usados para obtermos os pesos da equação 1, este conjunto é chamado de conjunto teste (*test set*). Tal abordagem normalmente é usada na calibragem das FEE's (de Azevedo & Dias, 2008).

1.5 Docking Molecular

A simulação de *docking* de pequenos ligantes contra a estrutura tridimensional de uma proteína alvo é um processo onde se visa buscar, entre as possíveis orientações/conformações, de um ligante no sítio ativo de uma proteína, aquela que apresenta a menor energia de ligação (Mitrasinovic, 2012, 2013; Huang *et al.* 2010). Assim, podemos dividir a simulação de *docking* molecular entre duas etapas principais. Numa primeira etapa são geradas posições possíveis para o ligante, chamadas *poses*. Numa segunda etapa, a energia de cada posição é avaliada, a partir de uma FE.

Considera-se como condição desejável para a aplicação da metodologia de *docking* molecular, a existência da estrutura tridimensional de uma proteína complexada com um ligante. Para a utilização das estruturas das enzimas da rota

metabólica utilizada neste projeto, considerou-se a existência de dados experimentais de afinidade das enzimas com pequenas moléculas e, também, a presença de estruturas cristalográficas determinadas. Tal complexo será usado para validação do protocolo de *docking*. De uma forma geral, essa validação é chamada de *re-docking*, e o seu sucesso pode ser avaliado pelo cálculo do desvio médio quadrático (*Root-Mean Square Deviation* (RMSD)), indicado na equação 2 abaixo:

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}$$

Equação 2. Representação do cálculo de RMSD. w_{ix} , w_{iy} e w_{iz} : posição cristalográfica do ligante. v_{ix} , v_{iy} e v_{iz} : posição do ligante obtida através da simulação de *Docking* (*pose*).

Um valor de RMSD considerado ótimo para as simulações *docking* deve ser encontrado na faixa de 0-2 Å. No entanto, simulações de *docking* com ligantes que apresentam grande número de ângulos de torção, podem apresentar RMSD um pouco maior (< 3,0 Å) e ainda serem aceitáveis.

Todas as enzimas que serão utilizadas neste trabalho apresentam mais de uma estrutura cristalográfica complexada ao ligante depositada no PDB. Dessa forma podemos estender a validação do protocolo estabelecido pelo *re-docking*. Esse processo de validação utilizando outras estruturas é chamado de *ensemble docking*. Uma vez identificado o melhor protocolo de *docking*, este pode ser utilizado para vasculhar uma base de dados de pequenos ligantes, como as disponíveis no ZINC (Irwin & Shoichet, 2005), pois se acredita que aquele protocolo terá a capacidade de gerar a posição mais próxima de uma possível estrutura cristalográfica do complexo. Tal processo é chamado *Virtual Screening* (Cerqueira *et al.* 2015). Para todas as simulações de *docking* presentes no projeto, usaremos o programa *Molegro Virtual Docker* (MVD) (Thomsen & Christensen, 2006). As etapas de todo o processo descrito acima, até o objetivo final, a previsão da afinidade entre o sítio ativo da enzima e uma molécula ligante, estão resumidas na figura 13.

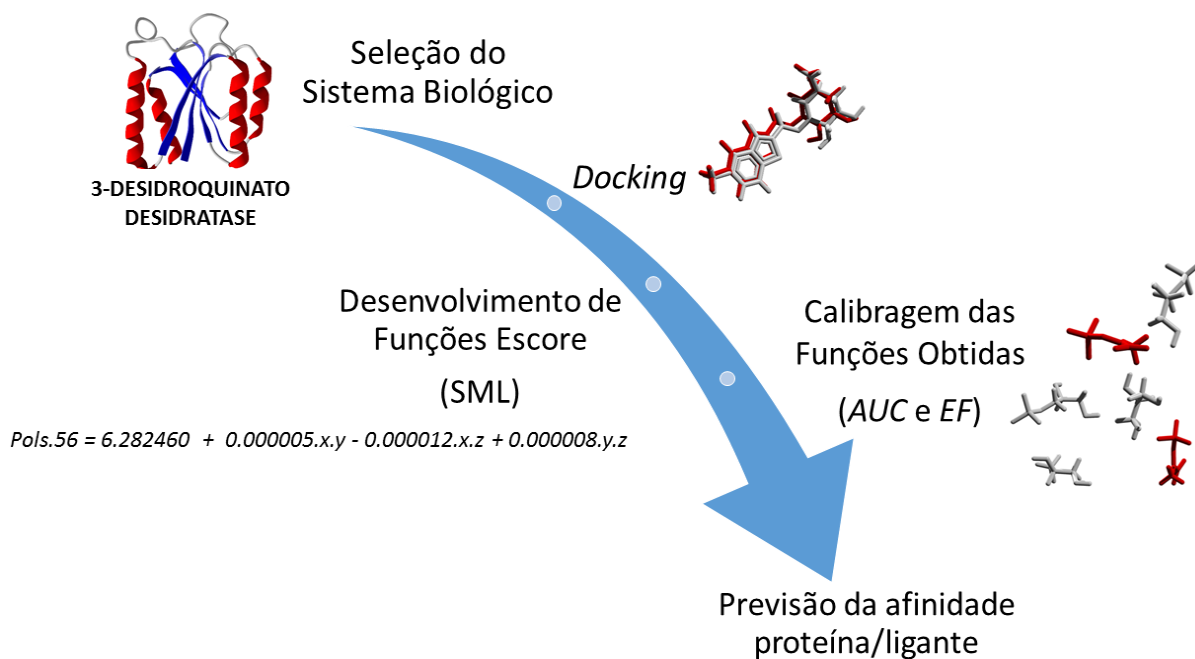


Figura 13. Etapas realizadas nos processos de Biologia de Sistemas e Computação Bioinspirada (*Docking*). A *Seleção* do sistema a ser trabalhado é o que dá início ao projeto; ao *Docking* pertencem todos os processos de seleção, testes e identificação das melhores FE's; em *SML* são formadas novas funções (polinômios) a partir das funções selecionadas no processo anterior; a *Calibragem* é a fase onde verifica-se se as novas funções formadas são capazes de reconhecer ligantes verdadeiros.

2 JUSTIFICATIVA

Doenças infecciosas estão entre as doenças que mais causam mortes no mundo atualmente. Entre elas, as doenças causadas por bactérias figuram entre as primeiras posições em rankings de doenças com maior número de óbitos, demonstrando a grande influência que esses patógenos exercem na saúde pública do Brasil e do Mundo. Uma das problemáticas que torna ainda mais difícil o combate a doenças bacterianas, é a baixa diversidade de compostos quimioterápicos para tratamentos. Com a descoberta da penicilina, em 1940, e nos vinte anos posteriores (1940 – 1960) existiu a chamada “época de ouro” dos antibióticos, onde ocorreu uma grande descoberta de novos compostos com ótima eficácia em tratamentos terapêuticos e profiláticos. Porém, com o uso negligente desses medicamentos pelos médicos, além de muitos pacientes não seguirem o tratamento adequado, encerrando-o antes do prazo solicitado pelo profissional da saúde, deu-se início ao processo de seleção de cepas bacterianas resistentes a certos tipos de compostos, o que trouxe maiores dificuldades no controle de patógenos. No ano de 2016, a OMS trouxe pesquisas alarmantes sobre a descoberta de novas cepas multirresistentes à quimioterápicos. Em organismos como *Neisseria gonorrhoeae*, *Treponema pallidum* e *Chlamydia trachomatis*, causadoras de importantes Doenças Sexualmente Transmissíveis, foi identificada resistência aos tratamentos clássicos, gerando uma série de debates e elaboração de novas estratégias para combater esses “novos” microrganismos.

Em reuniões organizadas pela OMS em 2016 (*Global Action Plan on Antimicrobial Resistance*) foram debatidas diversas estratégias para o combate de cepas multirresistentes. A mais desafiadora delas, é o desenvolvimento de novos antibióticos, pois além de encontrar-se um composto com características farmacológicas específicas em meio a uma infinidade de outros compostos, há uma série de leis de biossegurança e testes clínicos que devem ser respeitadas antes do fármaco chegar ao mercado.

No entanto, existem algumas fases do desenvolvimento de drogas antibacterianas que podem ser aceleradas e otimizadas sem ocasionar prejuízos ao processo. A otimização no processo de seleção de compostos para a realização de testes *in vitro* e *in vivo*, a partir da “confecção” de FEE’s, ou do processo de *docking* molecular é uma alternativa “boa e barata” ao processo, pois são capazes de

identificar possíveis candidatos a testes posteriores em meio a uma infinidade de possibilidades.

Partindo da visão apresentada acima e das novas estratégias colocadas pela OMS à comunidade científica, podemos afirmar que o estudo apresentado traz contribuições para a resolução dessa problemática. Determinação de FE's otimizam e direcionam a busca por fármacos, diminuindo os custos e o tempo dedicado para testes *in vitro*. Além disso, o entendimento das bases moleculares ocorrentes na interação entre proteína e ligante permitem a elaboração de fármacos específicos para cada sítio ativo e, também, alostéricos, aumentando a possibilidade de sucesso da terapia.

3. Objetivo Geral e Objetivos Específicos

3.1 Objetivo Geral

Determinar as bases estruturais para as interações das enzimas com ligantes aplicados às enzimas presentes na rota metabólica do chiquimato, com foco na 3-Desidroquinato Desidratase.

3.2 Objetivos Específicos

1. Realizar simulações de *docking* molecular para identificar protocolos capazes de prever a estrutura cristalográfica do complexo proteína-ligante.
2. Realizar *docking* molecular para ligantes com afinidade conhecida para a enzima possibilitando a análise com funções escores empíricas.
3. Testar diferentes funções escores empíricas para predição de afinidade dos ligantes documentados para a enzima 3-desidroquinato desidratase da via metabólica do chiquimato.

4. Artigos Científicos

4.1 Artigo principal (dados da pesquisa)

Machine learning methods for development of scoring functions to predict 3-dehydroquinase dehydratase inhibition.

Artigo submetido ao periódico “*Archives of Biochemistry and Biophysics*” de responsabilidade da Elsevier.

Fator de Impacto: 2.807

Sítio do periódico: <https://www.journals.elsevier.com/archives-of-biochemistry-and-biophysics/>

This message was sent automatically. Please do not reply.

Ref: YABBI_2016_100

Title: Machine learning methods for development of scoring functions to predict 3-dehydroquinase dehydratase inhibition.

Journal: Archives of Biochemistry and Biophysics

Dear Professor. Filgueira de Azevedo Junior,

Thank you for submitting your manuscript for consideration for publication in Archives of Biochemistry and Biophysics. Your submission was **received** in good order.

To track the status of your manuscript, please log into EVISE® at: http://www.evise.com/evise/faces/pages/navigation/NavController.jspx?JRNL_ACR=YABBI and locate your submission under the header 'My Submissions with Journal' on your 'My Author Tasks' view.

Thank you for submitting your work to this journal.

Kind regards,

Archives of Biochemistry and Biophysics

Have questions or need assistance?

For further assistance, please visit our [Customer Support](#) site. Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about EVISE® via interactive tutorials. You can also talk 24/5 to our customer support team by phone and 24/7 by live chat and email.

Copyright © 2016 Elsevier B.V. | [Privacy Policy](#)

4.2 Artigos desenvolvidos durante o período do Mestrado

1. SAnDReS: a Computational Tool for Statistical Analysis of Docking Results and Development of Scoring Functions

Artigo publicado pelo periódico “*Combinatorial Chemistry and High Throughput Screening*” de responsabilidade da Bentham Science.

Fator de Impacto: 1.041

Sítio do periódico: <http://benthamscience.com/journals/combinatorial-chemistry-and-high-throughput-screening/>

Combinatorial Chemistry & High Throughput Screening, 2016, 19, 1-12

1

RESEARCH ARTICLE

SAnDReS a Computational Tool for Statistical Analysis of Docking Results and Development of Scoring Functions

Mariana Morrone Xavier¹, Gabriela Sehnem Heck¹, Mauricio Boff de Avila^{1,2}, Nayara Maria Bernhardt Levin^{1,2}, Val Oliveira Pinto¹, Nathália Lemes Carvalho¹, Walter Filgueira de Azevedo Jr.*^{1,2}

¹Laboratory of Computational Systems Biology, Faculty of Biosciences, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Av. Ipiranga, 6681, Porto Alegre-RS 90619-900, Brazil; ²Graduate Program in Cellular and Molecular Biology, Faculty of Bioscience, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Av. Ipiranga, 6681, Porto Alegre-RS 90619-900, Brazil

1. INTRODUCTION

The data explosion in the number of macromolecules deposited in the Protein Data Bank (PDB) [1-3] opens the possibility to investigate the correlation of these experimentally determined structures with functional information. This is a favorable scenario for application of computational systems biology approaches [4]. Such approaches can be used to develop mathematical models to predict ligand-binding affinity for a target protein. It is also possible to use these three-dimensional structures to study

drug targets. The use of structural information makes possible to apply virtual screening (VS) methodology to identify novel hits and guide future development of new drugs. The main method to investigate potential new hits for a target protein is the procedure of protein-ligand docking simulations [5-11].

Protein-ligand docking simulations employ scoring functions to evaluate ligand-binding energy [10]. For validation of scoring functions, it is common to investigate the correlation between the experimental binding affinity with scoring functions. This statistical analysis can be based on squared Pearson's (R^2) or Spearman's (ρ) correlation coefficients [12]. Analysis of scoring function performance can also be carried out using data sets with active and decoy ligands, as proposed in the directory of useful decoys, enhanced (DUD-E) [13].

*Address correspondence to this author at the Faculty of Biosciences - Pontifical Catholic University of Rio Grande do Sul (PUCRS), Av. Ipiranga, 6681, Porto Alegre-RS 90619-900, Brazil; Tel/Fax: ++55-51-3353-4529; E-mails: walter.junior@pucrs.br, walter@azevedolab.net
MMX, GSH, MBA, NMBL, and VOP contributed equally to this work and should be considered first authors.

2. Understanding the Structural Basis for Inhibition of Cyclin-Dependent Kinases. New Pieces in the Molecular Puzzle

Artigo publicado pelo periódico "Current Drug Targets" de responsabilidade da Bentham Science.

Fator de Impacto: 3.029

Sítio do periódico: <http://benthamscience.com/journals/current-drug-targets/>

Current Drug Targets, 2017, 18, 000-000

1

REVIEW ARTICLE

Understanding the Structural Basis for Inhibition of Cyclin-Dependent Kinases. New Pieces in the Molecular Puzzle

Nayara M. Bernhardt Levin^{a,b}, Val Oliveira Pintro^a, Mauricio Boff de Ávila^{a,b}, Bruna Boldrini de Mattos^a and Walter Filgueira de Azevedo Jr.^{a,b,*}

1. INTRODUCTION

Several protein kinases (enzyme classification (EC) 2.7.-.-) have been described as participating in intracellular regulatory pathways (reviews in [1-10]). Our focus in this paper is on cyclin-dependent kinases (CDKs) (EC 2.7.11.22). We could say that, from the enzymology point of view, CDKs are enzymes that catalyze the phosphor transfer to a protein, as indicated in the catalyzed reaction below,



CDKs are also known as serine-threonine kinases, due to the specificity of their substrates.

Functional studies were able to determine the significance of CDKs in the cell cycle progression [11, 12]. As functional studies progressed, new CDKs were identified and their biological roles established [13, 14]. Because of the importance of CDKs in the regulation of the cell division cycle, these enzymes have been the object of extensive investigation. Furthermore, due to their central role in cell cycle progression, CDKs have caught attention as a target for development of anticancer drugs. CDK inhibitors have also shown potential for the treatment of inflammatory disorders [15] and neurodegenerative diseases [16]. Screening studies aimed to identify CDK inhibitors were initially focused on starfish due to the abundance of CDK1/cyclin D in starfish oocytes [17], which allowed the identification of Roscovitine, a CDK inhibitor with half maximal inhibitory concentration (IC₅₀) in nanomolar range [18, 19].

In parallel with CDK functional studies, several research groups have tried to obtain x-ray diffracting crystals to de-

*Address correspondence to this author at the Laboratory of Computational Systems Biology, Faculty of Biosciences - Pontifical Catholic University of Rio Grande do Sul (PUCRS), Av. Ipiranga, 6681, Porto Alegre-RS 90619-900, Brazil; Tel/Fax: ??????????????; E-mail: walter.junior@puers.br

5. Resultados e Discussão

5.1 Pre-Docking e Re-Docking

Usando somente estruturas resolvidas por difração de Raios X, a estrutura com o código PDB: 2Y71 (1,5 Å) foi escolhida, por apresentar a melhor resolução, para a testagem dos protocolos desenvolvidos pelo coordenador do estudo (Anexo A), unindo as FE's e os algoritmos de busca contidos no *software* MVD. O processo de *re-docking* foi realizado com a utilização de trinta e dois protocolos, sendo que cada um gerou um total de 50 poses (posição do ligante segundo a equação utilizada). Em análise geral de todos os protocolos, o de número 29 (P29) apresentou o melhor resultado de correlação (ρ e R^2) e o menor valor de $RMSD < 2,0$ Å. Os fatores de correlação confirmaram os resultados do processo, uma vez que apresentaram valores elevados, 0,137 – 0,828 (ρ) e 0,005 – 0,749 (R^2) (Tabela I).

Tabela I. Resultados de re-docking com o valor de RMSD para função escore. *p-value* 1 relaciona-se com ρ ; *p-value* 2 relaciona-se com R^2 (Abreviações: ρ , Coeficiente de Spearman; R^2 , ρ^2).

Scoring function	RMSD(Å)	ρ	p-value1	R^2	p-value2
Plants Score	1,445	0,688	$3,29 \cdot 10^{-8}$	0,475	$3,1810^{-5}$
MolDock Score	1,443	0,828	$1,26 \cdot 10^{-13}$	0,716	$1,02 \cdot 10^{-2}$
Rerank Score	1,443	0,824	$2,04 \cdot 10^{-13}$	0,749	$5,19 \cdot 10^{-13}$
Interaction Score	1,443	0,811	$9,61 \cdot 10^{-13}$	0,677	$2,23 \cdot 10^{-3}$
Protein Score	1,443	0,811	$9,61 \cdot 10^{-13}$	0,677	$2,23 \cdot 10^{-10}$
Internal Score	2,174	0,441	$1,34 \cdot 10^{-3}$	0,223	$5,35 \cdot 10^{-1}$
Electro Score	3,439	0,283	$4,65 \cdot 10^{-2}$	0,015	$3,90 \cdot 10^{+2}$
Electro Long Score	3,334	0,643	$4,86 \cdot 10^{-7}$	0,409	$5,75 \cdot 10^{-4}$
H-Bond Score	0,668	0,805	$1,91 \cdot 10^{-12}$	0,741	$1,10 \cdot 10^{-12}$
LE1 Score	1,443	0,828	$1,26 \cdot 10^{-13}$	0,716	$1,02 \cdot 10^{-11}$
LE3 Score	1,443	0,824	$2,04 \cdot 10^{-13}$	0,749	$5,19 \cdot 10^{-13}$
Docking Score	1,445	0,688	$3,30 \cdot 10^{-8}$	0,475	$3,18 \cdot 10^{-5}$
Displaced Water Score	6,263	0,137	$3,41 \cdot 10^{-1}$	0,005	$6,28 \cdot 10^{+2}$

5.2 Ensemble Docking

O *ensemble* consiste na utilização do protocolo selecionado durante o *re-docking* (P29) em todas as estruturas do *data set* gerado a partir dos bancos de dados disponíveis (PDB e *BindingDB*) e do processo de *pré-docking*. Nessa análise o coeficiente de Spearman (ρ) variou entre -0,418 e 0,239, demonstrando, novamente, bons valores para o processo. As funções com o melhor resultado foram *Plants* e *Docking* ($\rho = -0,418$; $p\text{-value} = 0,0526$) (Tabela II). Após as análises de correlação o *software* SAnDReS (Xavier *et al.* 2016) calculou valores para a Acurácia do *Docking* (DA), capacidade do MVD, com determinado protocolo, de gerar poses com $RMSD < 2.0 \text{ \AA}$ (DA1) e $RMSD < 3.0 \text{ \AA}$ (DA2). Para DA1 foi encontrado um valor igual a 88,364% e para DA2, 88,636%, o que confirma que o P29 é capaz de gerar a maioria das poses com $RMSD < 2,0 \text{ \AA}$.

Tabela II. Resultados de Ensemble Docking com o valor de RMSD para função escore. $p\text{-value} 1$ relaciona-se com ρ ; $p\text{-value} 2$ relaciona-se com R^2 (Abreviações: ρ , Coeficiente de Spearman; R^2 , ρ^2).

Scoring function	RMSD(Å)	ρ	p-value1	R^2	p-value2
Plants Score	0,665	-0,418	$5,26 \cdot 10^{-2}$	0,189	$4,32 \cdot 10^{-2}$
MolDock Score	0,218	0,237	$2,89 \cdot 10^{-1}$	0,003	$8,21 \cdot 10^{-1}$
Re-rank Score	0,485	0,225	$3,13 \cdot 10^{-1}$	0,004	$7,78 \cdot 10^{-1}$
Interaction Score	1,0245	0,019	$9,34 \cdot 10^{-1}$	0,001	$8,66 \cdot 10^{-1}$
Protein Score	1,0245	0,019	$9,34 \cdot 10^{-1}$	0,001	$8,66 \cdot 10^{-1}$
Internal Score	0,665	-0,174	$4,37 \cdot 10^{-1}$	0,006	$7,34 \cdot 10^{-1}$
Electro Score	3,317	-0,081	$7,21 \cdot 10^{-1}$	0,226	$2,55 \cdot 10^{-2}$
Electro Long Score	3,631	-0,303	$1,70 \cdot 10^{-1}$	0,253	$1,69 \cdot 10^{-2}$
H-Bond Score	0,218	0,239	$2,84 \cdot 10^{-1}$	0,044	$3,48 \cdot 10^{-1}$
LE1 Score	3,631	0,126	$5,77 \cdot 10^{-1}$	0,017	$5,66 \cdot 10^{-1}$
LE3 Score	3,631	0,029	$8,99 \cdot 10^{-1}$	0,039	$3,76 \cdot 10^{-1}$
Docking Score	0,665	-0,418	$5,26 \cdot 10^{-2}$	0,189	$4,32 \cdot 10^{-2}$

5.3 Funções Escore

Nessa fase, busca-se relacionar as FE's utilizadas até o momento (P29) com valores experimentais de inibição ($\log(K_i)$). O objetivo desse estágio é identificar as funções que conseguem simular computacionalmente os dados obtidos de forma experimental e selecionar as melhores, a partir de ρ e $p\text{-value}$, para a formação de novas funções escores polinomiais. Antes das simulações, a estrutura PDB: 1GU1 foi

excluída do processo por apresentar dualidade de dados em diferentes bases (PDBbind, Binding MOAD e BindingDB) podendo gerar resultados não confiáveis. O melhor valor de correlação foi apresentado por *Plants* ($\rho = 0,590$; $p\text{-value} = 0,003$). $P\text{-values} < 0,05$ também foram encontrados para *Re-Rank*, *Protein*, *LE1*, *LE3 Score* e *Electro Long* (Tabela III), podendo qualquer uma dessas funções participar do polinômio futuramente formado.

Tabela III. Correlação entre funções escores (MVD) e afinidade proteína/ligante (log(Ki)). $p\text{-value} 1$ relaciona-se com ρ ; $p\text{-value} 2$ relaciona-se com R^2 . (Abreviações: ρ , Coeficiente de Spearman; R^2 , ρ^2)

Scoring function	ρ	p-value1	R^2	p-value2
<i>Plants Score</i>	0,590	$3,84 \cdot 10^{-3}$	0,393	$1,70 \cdot 10^0$
<i>MolDock Score</i>	-0,335	$1,27 \cdot 10^{-1}$	0,173	$5,40 \cdot 10^{+1}$
<i>Re-rank Score</i>	-0,534	$1,30 \cdot 10^{-2}$	0,383	$2,13 \cdot 10^0$
<i>Interaction Score</i>	0,355	$1,05 \cdot 10^{-1}$	0,179	$4,96 \cdot 10^{+1}$
<i>Cofactor Score</i>	-0,314	$1,54 \cdot 10^{-1}$	0,124	$1,08 \cdot 10^{+2}$
<i>Protein Score</i>	0,455	$3,31 \cdot 10^{-2}$	0,259	$1,55 \cdot 10^{+1}$
<i>Water Score</i>	0,286	$1,94 \cdot 10^{-1}$	0,007	$7,20 \cdot 10^{+2}$
<i>Internal Score</i>	0,208	$3,53 \cdot 10^{-1}$	0,128	$1,02 \cdot 10^{+2}$
<i>Electro Score</i>	-0,024	$9,14 \cdot 10^{-1}$	0,020	$5,35 \cdot 10^{+2}$
<i>Electro Long Score</i>	-0,467	$2,83 \cdot 10^{-2}$	0,375	$2,43 \cdot 10^0$
<i>H-Bond Score</i>	0,103	$6,47 \cdot 10^{-1}$	0,000	$9,90 \cdot 10^{+2}$
<i>LE1 Score</i>	-0,576	$5,03 \cdot 10^{-3}$	0,297	$8,71 \cdot 10^{+2}$
<i>LE3 Score</i>	-0,526	$1,19 \cdot 10^{-2}$	0,274	$1,24 \cdot 10^{+2}$

5.4 Funções Escore Polinomiais

Na formação de novas funções escores foram utilizados “Métodos de Aprendizado de Máquina” (SML), pois, assim, é possível formar e testar diferentes esquemas de equações utilizando, como base, as FE’s presentes no MVD. O software SAnDReS (Xavier *et al*, 2016) foi responsável por gerar 286 combinações de polinômios contendo três variáveis (FE’s originais). Para cada um dos 286 polinômios, foram geradas 511 possibilidades formando, assim, 146.146 equações diferentes de funções na tentativa de prever a afinidade entre o sítio ativo da enzima e o, possível, ligante. As equações que apresentaram maior valor de correlação foram as que continham *Re-Rank*, *LE1* e *Electro Long*, e entre elas os melhores resultados foram alcançados pelos polinômios 56, 164, 172, 181 e 416.

Dentre as cinco equações, 56, 172 e 181 foram as que apresentaram os melhores valores de correlação para o *test set* (56 ($\rho = 0,900$; $p\text{-value} = 0,037$), 172

($\rho = 0,943$; $p\text{-value} = 0,048$) e 181 ($\rho = 0,943$; $p\text{-value} = 0,048$) (Tabela IV). Abaixo estão as três equações geradas pelo método de regressão linear:

$$\text{Polyscore 56} = 6.282460 + 0.000005.x.y - 0.000012.x.z + 0.000008.y.z$$

$$\text{Polyscore 172} = -7.194693 + -0.002396.y + -0.000001.x.y + 0.000005.y.z + 0.000002.x^2$$

$$\text{Polyscore 181} = -7.263051 + -0.002478.y + 0.000000.x.y + 0.000002.x.z + 0.000002.x^2 + -0.000017.z^2$$

Os três polinômios utilizam *Re-rank* (x), *LE1* (y) e *Electro Long* (z) como variáveis.

Tabela IV. Resultado de correlação para o training set e test set. $p\text{-value}$ 1 relaciona-se com ρ . (Abreviações: ρ , Coeficiente de Spearman). (*Um ponto outlier (estrutura 2XD9) foi eliminado da análise).

Scoring function	ρ (training set)	$p\text{-value1}$ (training set)	ρ (test set)	$p\text{-value1}$ (test set)
Plants Score	0,629	$9,06.10^{-3}$	0,314	$5,44.10^{-1}$
MolDock Score	-0,199	$4,60.10^{-1}$	-0,943	$4,80.10^{-3}$
Re-rank Score	-0,259	$3,32.10^{-1}$	-0,771	$7,24.10^{-2}$
Interaction Score	0,406	$1,18.10^{-1}$	0,029	$9,57.10^{-1}$
Co-factor Score	-0,277	$3,98.10^{-1}$	-0,543	$2,65.10^{-1}$
Protein Score	0,427	$9,90.10^{-2}$	0,771	$7,24.10^{-2}$
Water Score	0,174	$5,2.10^2$	0,371	$4,68.10^{-1}$
Internal Score	0,093	$7,33.10^2$	0,657	$1,56.10^{-1}$
Electro Score	-0,180	$5,06.10^{-1}$	0,257	$6,22.10^{-1}$
Electro Long Score	-0,449	$8,09.10^{-1}$	-0,086	$8,71.10^{-1}$
H-Bond Score	0,068	$8,03.10^{-1}$	0,257	$6,22.10^{-1}$
LE1 Score	-0,610	$1,22.10^{-1}$	-0,771	$7,24.10^{-2}$
LE3 Score	-0,711	$2,00.10^{-3}$	-0,200	$7,04.10^{-1}$
Polyscore#0000416	0,817	$1,10.10^{-4}$	0,829	$4,15.10^{-2}$
Polyscore#0000164	0,798	$2,10.10^{-4}$	0,829	$4,15.10^{-2}$
Polyscore#0000181	0,776	$4,08.10^{-4}$	0,943	$4,80.10^{-4}$
Polyscore#0000172	0,773	$4,44.10^{-4}$	0,943	$4,80.10^{-4}$
Polyscore#0000056*	0,763	$5,87.10^{-4}$	0,900	$3,70.10^{-2}$

Analisando as variáveis utilizadas na formação dos polinômios, pode-se perceber a importante participação das interações eletrostáticas na definição da afinidade de ligação. Há um termo explícito nas ligações de longo alcance (*Electro Long*) e termos implícitos em *Re-Rank* e *LE1* (Xavier et al. 2016). Uma análise do sítio ativo da enzima demonstra a participação dos resíduos Arg 15, His 81, Asp 88, Glu 92, His 101, Arg 108, e Arg 112 na maioria das ligações entre sítio/ligante. Tais resíduos carregados, estão, muito possivelmente, diretamente relacionados com a

escolha de uma equação que priorize as relações eletrostáticas para prever a afinidade proteína/ligante. A figura 14 traz os gráficos de dispersão para as três equações que apresentaram os melhores resultados (181, 172 e 56).

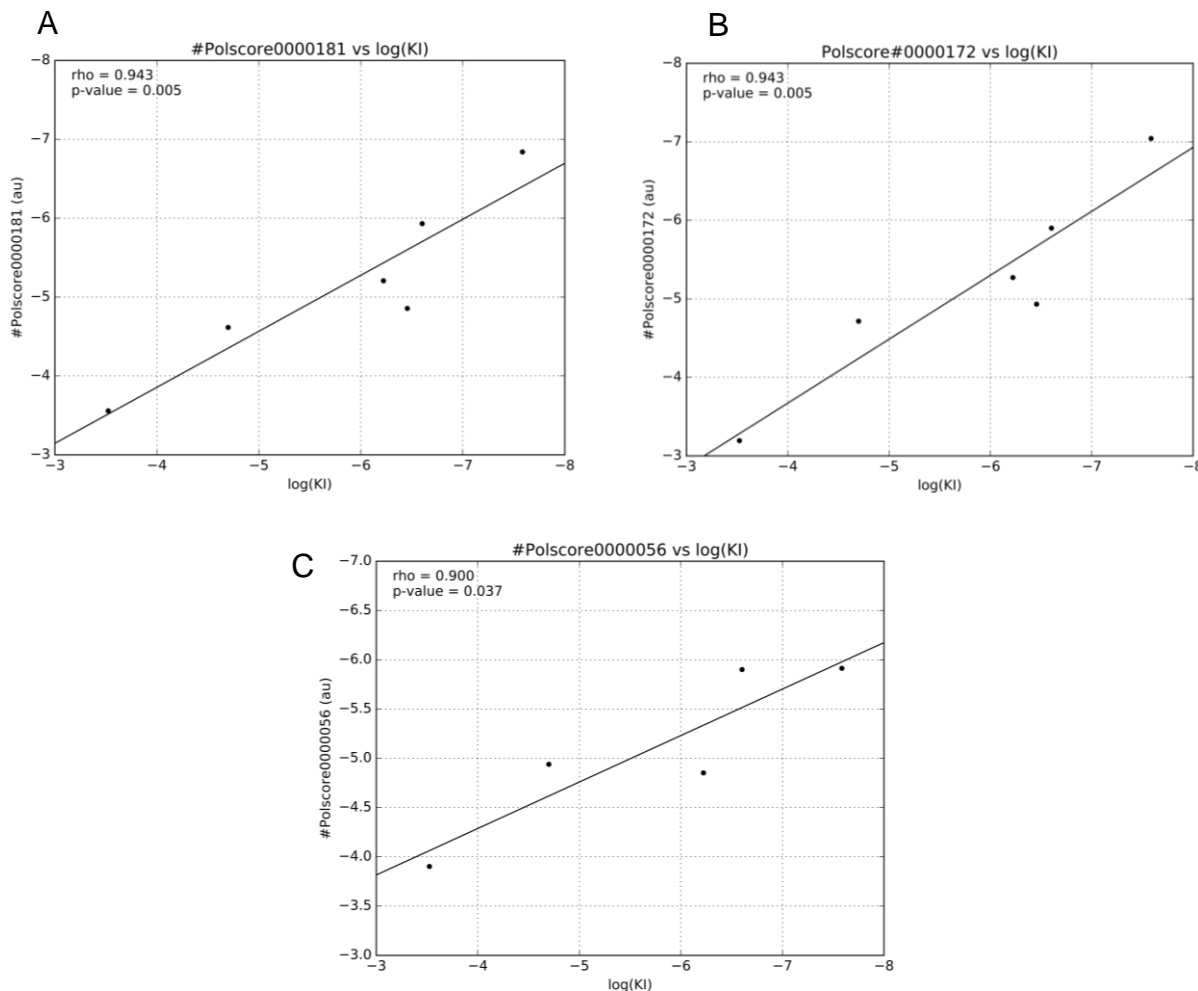


Figura 14. Gráficos de dispersão das três equações que apresentaram os melhores resultados de *test set* em relação ao $\log(ki)$. Cada ponto no gráfico representa o $\log(KI)$ calculado pela equação para cada estrutura da enzima. A) *Polscore0000181*; B) *Polscore0000172*; C) *Polscore0000056*.

5.5 Reconhecimento de verdadeiros e falsos ligantes

A última fase de testes é caracterizada por definir a capacidade dos polinômios formados em distinguir moléculas que se sabe terem afinidade (verdadeiros) pelo sítio ativo da enzima, de moléculas que, também, são conhecidas por não terem afinidade (falsos) pelo sítio de ligação. Nesse intuito foi construído um *data set* com vinte e dois ligantes verdadeiros, extraídos das estruturas cristalográficas, e cento e noventa e oito falsos, escolhidos aleatoriamente de um grupo de moléculas montado por Kumar & Siddiqi, 2010. O VS foi realizado com os duzentos e vinte ligantes do *data set*

construído contra a estrutura PDB: 2Y71, escolhida no *re-docking*. Perante os resultados obtidos com a *Área sobre a Curva* (AUC), capacidade de reconhecimento dos ligantes positivos, e do *Fator de Enriquecimento* (EF), porcentagem de ligantes positivos reconhecidos em 1% (EF1), 2% (EF2), 5% (EF5), 10% (EF10) e 20% (EF20) da amostra total do conjunto de ligantes, a melhor função polinomial foi escolhida. Os resultados de AUC podem gerar gráficos do tipo *ROC* (Características Operacionais do Receptor) que, também, está relacionado com a capacidade das equações (polinômios e FE's) de reconhecer positivos e falsos ligantes.

Os maiores valores para AUC foram encontrados para *H-Bond* (75,678%) e *Polyscore 56* (74,686%), para EF1 e EF2, valores com maior relevância, a equação 56 apresentou os melhores resultados, EF1 = 540 e EF2 = 159,23 (Tabela V). Em relação ao AUC, e EF a equação 56 demonstrou estar mais apta no reconhecimento de ligantes do que as equações 172 e 181.

Tabela V. Resultados e análises para AUC e EF. (Abreviações: *P*, *Plants*; *MD*, *MolDock*; *R*, *Re-Rank*; *I*, *Interaction*; *C*, *Cofactor*; *Pro*, *Protein*; *W*, *Water*; *Int*, *Internal*; *T*, *Torsions*; *SC*, *Soft Constraints*; *E*, *Electro*; *EL*, *ElectroLong*; *HB*, *HBond*; *D*, *Docking*; *Eq*, *Polyscore*).

Função Escore	AUC (%)	EF1	EF2	EF5	EF10	EF20
<i>P Score</i>	60,317	18,205	10,952	5,9883	3,5467	2,3111
<i>MD Score</i>	49,591	7,4603	7,3228	4,7058	2,8504	1,6896
<i>R Score</i>	50,842	13,913	7,7419	4,9051	2,5570	1,6896
<i>I Score</i>	52,039	15,581	8,6440	4,8648	2,8205	1,6896
<i>C Score</i>	18,273	0,0	0,0	0,0	0,1476	0,1804
<i>Pro Score</i>	54,532	17,5	11,359	5,0684	3,0177	1,8152
<i>W Score</i>	50,0	7,4603	7,3228	4,7058	2,8504	1,6896
<i>Int Score</i>	43,704	0,0	0,0	0,6589	0,9890	0,8641
<i>T Score</i>	69,031	0,0	0,0	1,6279	3,2370	1,8662
<i>SC Score</i>	50,0	7,4603	7,3228	4,7058	2,8504	1,6896
<i>E Score</i>	49,722	3,4146	2,4293	1,7773	1,7647	2,4787
<i>EL Score</i>	37,973	0,0	0,0	0,0732	2,1145	1,9760
<i>HB Score</i>	75,678	210,0	105,78	16,960	6,9753	3,6307
<i>LE1 Score</i>	52,390	16,829	8,6440	4,4356	2,5858	1,7395
<i>LE3 Score</i>	54,442	68,571	16,829	6,2721	3,1108	1,8470
<i>D Score</i>	60,317	18,205	10,952	5,9883	3,5467	2,3111
<i>Eq 56</i>	74,686	540,0	159,23	13,504	5,8959	3,3495
<i>Eq 416</i>	45,469	3,0952	1,9565	1,7021	1,4822	1,2647
<i>Eq 164</i>	54,532	17,5	11,359	5,0684	3,0177	1,8152
<i>Eq 181</i>	54,544	18,205	11,359	5,0684	3,0331	1,8152
<i>Eq 172</i>	54,532	17,5	11,359	5,0684	3,0177	1,8152

Em comparação a dados experimentais, *Polyscore 56* demonstrou ser mais eficiente na determinação do AUC em dados prévios de *docking* para Anidrase Carbônica II (EC 4.2.1.1) (Mysingert, *et al.* 2012), enzima similar à DHQD. Sendo assim, em estudos futuros recomenda-se a utilização da equação 56 para predição de afinidade da enzima DHQD contra possíveis ligantes antes da realização de teste *in vitro* e *in vivo*. A partir dos valores de AUC para *H-Bond* e *Polyscore56* foram gerados gráficos *ROC* para demonstrar a relação entre os positivos e falsos ligantes (Figura 15).

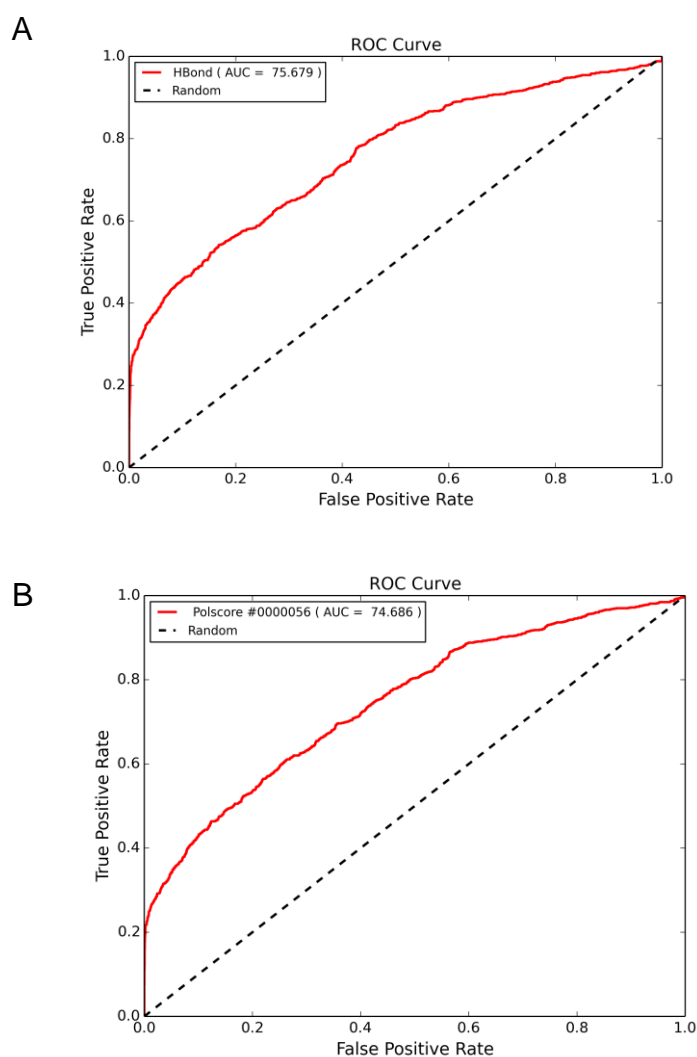


Figura 15. Curvas ROC para os dois melhores resultados de AUC. A) H-Bond Score, AUC = 75,678%; B) Polyscore0000056 = 74,686%.

6. CONSIDERAÇÕES FINAIS

A identificação de novos compostos antibióticos capazes de controlar, ou prevenir cepas bacterianas é uma necessidade atual para a saúde pública. A descoberta de novas rotas metabólicas e enzimas alvo para o desenho de drogas vem ganhando força, uma vez que enzimas utilizadas a anos como foco para drogas já não respondem aos diversos compostos presentes no mercado. Nesse paradigma, a via metabólica do chiquimato apresenta-se como um forte alvo para o desenho de drogas por dois motivos: nunca foi utilizada como composto antibiótico, somente como herbicida (glifosato); e não está presente em organismos animais, o que diminui as chances de os compostos provocarem efeitos colaterais. O foco desse trabalho, a enzima DHQD, apresenta-se como possível alvo de novos compostos, pois está presente em todos os organismos bacterianos, além de ser essencial para o crescimento e desenvolvimento desses indivíduos.

As simulações de *docking* do estudo realizado focaram em buscar FEE's que fossem capazes de prever a afinidade entre a enzima e os ligantes naturais, facilitando, assim, testes *in vitro* de novos compostos. Além das simulações de *docking*, também foram realizados métodos SML na confecção de novas funções escores polinomiais a partir de FE's presentes no *software* MVD. O critério de seleção utilizado em todas as fases de testes foi o Coeficiente de Spearman (ρ) e p -value < 0,05 (significativamente estatístico para sistemas biológicos).

Ao final de todas as simulações e testes realizados ao longo do projeto, chegamos à conclusão de que a equação *Polyscore 56* apresentou-se como a mais hábil para prever a afinidade entre o sítio ativo de DHQD com compostos testados. Para esse polinômio os resultados de *test set* ($\rho = 0,900$; p -value = 0,037), AUC (74,686%), EF1 (540) e EF2 (159,23) foram, na maioria das categorias avaliadas, os melhores, confirmando as hipóteses formuladas sobre a equação. Assim, ao término do trabalho, sugerimos, aos futuros estudos com a enzima DHQD de qualquer organismo bacteriano, a utilização do polinômio *Polyscore56* em análises prévias aos testes *in vitro* e *in vivo* para uma otimização e, possível, economia de tempo e recursos.

7. REFERÊNCIAS

1. <MINISTÉRIO DA SAÚDE BRASILEIRO>. Disponível em: <http://dados.saude.gov.br>. Acesso em: 26/12/2016.
2. <ORGANIZAÇÃO MUNDIAL DA SAÚDE>. Disponível em: <http://www.who.int>. Acesso em: 26/12/2016
3. VENTER H, HENNINGSEN ML, BEGG SL. Antimicrobial resistance in healthcare, agriculture and the environment: the biochemistry behind the headlines. *Essays Biochemistry*. 2017; 61:1 – 10.
4. <ANVISA>. Disponível em: <http://www.anvisa.gov.br>. Acesso em: 17/11/2016.
5. ENZLER MJ, BERBARI E, OSMON DR. Antimicrobial prophylaxis in adults. *Mayo Clinic Proceedings*. 2011; 86(7):686-701.
6. LEEKHA S, TERRELI CL, EDSON RS. General Principles of Antimicrobial Therapy. *Mayo Clinic Proceedings*. 2011; 86(2):156-167.
7. WRIGHT GD. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews: Microbiology*. 2007; 5:175 – 186.
8. WRIGHT GD. Bacterial resistance to antibiotics: Enzymatic degradation and modification. *Advanced Drug Delivery Reviews*. 2005; 5:1451-1470.
9. <EUROPEAN MEDICINES AGENCY>. Disponível em: <http://www.ema.europa.eu>. Acesso em: 26/12/2016.
10. BENTLEY R & HASLAM E. The Shikimate Pathway: a Methabolic Tree with Many Branches. *Critical Reviews in Biochemistry and Molecular Biology*. 1990; 25(5):307-384.

11. MILLAR G & COGGINS, J.R. The complete amino acid sequence of dehydroquinase synthase of *Escherichia coli* K12. FEBS Lett. 1986; 200:11-17.
12. HERRMANN KM. The Shikimate Pathway as Entry to Aromatic Secondary Metabolism. Plant Physiology. 1995; 107:7-12.
13. HAWKINS AR & SMITH M. Domain structure and interaction within the pentafunctional AROM polypeptide. European Journal of Biochemistry. 1991; 196(7): 17-724.
14. ROBERTS CW, *et al.* The Shikimate Pathway and Its Branches in Apicomplexan Parasites. The Journal of Infectious Diseases. 2002; 185(Supplement 1):25-36.
15. MAEDA H & DUDAREVA N. The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. The Annual Review of Plant Biology. 2013; 53:73-105.
16. ROBERTS F, *et al.* Evidence for the Shikimate Pathway in Apicomplexan Parasites. Letters to Nature. 1998; 393:801-805.
17. MIR R, JALLU S, SINGH TP. The Shikimate pathway: Review of amino acid sequence, function and three-dimensional structures of the enzymes. Critical reviews in Microbiology. 2013; Early online:1-18.
18. SHUMILIN IA, KRETSINGER RH, BAUER RH. Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosanate-7-phosphate synthase from *Escherichia coli*. Structure. 1999; 7:865-875.
19. WALLACE BJ & PITTARD J. Genetic and Biochemical Analysis of the Isoenzymes Concerned in the First Reaction of Aromatic Biosynthesis in *Escherichia coli*. Journal of Bacteriology. 1967; 39(1):237-244.

20. SHUMILIN IA, ZHAO C, BAUERLE R, KRETSINGER RH. Allosteric Inhibition of 3-deoxy-D-arabino-heptulosanato-7-phosphate Synthase Alters the Coordination of Both substrates. *Journal of Molecular Biology*. 2002; 320:1147-1156.
21. CHANDRAN SS & FROST JW. Aromatic Inhibitors of Dehydroquinase Synthase: Synthesis, Evaluation and Implications for Gallic Acid Biosynthesis. *Bioorganic & Medicinal Chemistry Letters*. 2001; 11:1493-1496.
22. HERRMANN KM & WEAVER LM. The shikimate pathway. *Annu. Rev. Plant Physiology Plant Molecular Biology*. 1999; 50:473-503.
23. MINASOV G, LIGHT SH, SHUVALOVA L, PAPAZISI L, ANDERSON WF. 2.5 Angstrom Resolution Crystal Structure of 3-Dehydroquinase (aroB) from *Vibrio cholerae*. Artigo aceito para futura publicação. Disponível em: <http://www.rcsb.org/pdb/explore/explore.do?structureId=3okf>. Acesso em: 25/06/2016.
24. ROSZAK AW, *et al.* The Structure and Mechanism of the Type II Dehydroquinase from *Streptomyces coelicolor*. *Structure*. 2002; 10(4): 493-503.
25. GOURLEY DG, *et al.* The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nature Structural Biology*. 1999; 6:521-525.
26. TIZON L, *et al.* A Prodrug Approach for Improving Antituberculosis Activity of Potent Mycobacterium Tuberculosis Type II Dehydroquinase Inhibitors. 2011; 54(17):6063-6084.
27. HARRIS JM, GONZALEZ-BELLO C, KLEANTHOUS C, HAWKINS AR, COGGINS JR, ABELL C. Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases. *Biochemical Journal*. 1996; 319(2):333-336.

28. LASKOWSKI RA, SWINDELLS MB. LigPlot⁺: multiple ligand-protein interaction diagrams for drug discovery. *J Chem. Inf. Model.* 2011, 51: 2778-2786.
29. PEEK J, SHI T, CHRISTENDAT D. Identifications of Novel Polyphenolic Inhibitors of Shikimate Dehydrogenase (*AroE*). *Journal of Biomolecular Screening.* 2014; 19: 1090-1098.
30. AVITIA-DOMÍNGUEZ C, *et al.* *Molecules.* 2014; 19:4491-4509.
31. MICHEL G, *et al.* Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities. *J Biol Chem.* 2003; 278(21): 19463-19472.
32. WHIPP MJ & PITTARD AJ. A Reassessment of the Relationship between *aroK*- and *aroL*- Encoded Shikimate Kinase Enzymes of *Escherichia coli*. *Journal of Bacteriology.* 1995; 177:1627-1629.
33. DIAS MV, *et al.* Effects of the magnesium and chloride ions and shikimate on the structure of shikimate kinase from *Mycobacterium tuberculosis*. *Acta Crystallografica.* 2007; 63:1 – 6.
34. GU Y, *et al.* Crystal Structure of Shikimate Kinase from *Mycobacterium tuberculosis* Reveals the Dynamic Role of the LID Domain in Catalysis. *Journal of Molecular Biology.* 2002; 319(3):779-789.
35. PEREIRA JH, *et al.* Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. *Acta Crystallographica.* 2004; 60:2310-2319.
36. FUNKE T, *et al.* Structural Basis of Glyphosate Resistance Resulting from the Double Mutation Thr97-Ile and Pro101-Ser in 5-Enolpyruvylshikimate-3-phosphate Synthase from *Escherichia coli*. *The Journal of Biological Chemistry.* 2009; 284; 9854-9860.

37. STALLINGS WC, *et al.* Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase: a distinctive protein fold. Proceedings of the National Academy of Sciences. 1991; 88(11):5046-5050.
38. SCHÖNBRUNN E, *et al.* Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proc. Natl. Acad. Sci. USA. 2001; 98(4):1376 – 1380.
39. WEAVER LM & HERRMANN KM. Dynamics of the shikimate pathway in plants. Trends in Plant Science. 1997; 2(9):346-351.
40. THOMAS MG, *et al.* A series of 2(Z)-2-Benzylidene-6,7-dihydroxybenzofuran-3[2H]-ones as inhibitors of chorismate synthase. Biorganic & medicinal chemistry letters. 2003; 13(3):423-426.
41. AHN HJ, YOON HJ, LEE B, SUH SW. Crystal Structure of Chorismate Synthase: A novel FMN-binding Protein Fold and Functional Insights. Journal of Molecular Biology. 2003; 336:903-915.
42. QUEVILLON-CHERUEL S, *et al.* Crystal Structure of the Bifunctional Chorismate Synthase from *Saccharomyces cerevisiae*. The Journal of Biological Chemistry. 2004; 279(1):619-625.
43. VIOLA CM, SARIDAKIS V, CHRISTENDAT D. Crystal structure of chorismate synthase from *Aquifex aeolicus* reveals a novel beta alpha beta sandwich topology. Proteins. 2004; 54(1):166-169.
44. BENSON DA, *et al.* GenBank. Nucleic Acids Research. 1999; 27(1):12-17
45. BERMAN HM, *et al.* The Protein Data Bank. Nucleic Acids Research. 2000; 28:235-242.
46. KITANO H. Computational Systems Biology. Nature. 2002; 420:206-210.

47. DE ÁVILA MB & DE AZVEDO WF Jr. Data Mining of Docking Results. Application to 3-Dehydroquinase Dehydratase. *Current Bioinformatics*. 2014; 9:361-379.
48. LIU T, LIN Y, WEN C, JORISSEN RN, GILSON MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research*. 2007; 1:198-201.
49. THOMSEM R & CHRISTENSEN MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem*. 2006; 49:3315-3321.
50. FLOREANO D., MATTIUSI C. *Bio-inspired Artificial Intelligence: Theories, Methods, and Technologies*. Cambridge: The MIT Press 2008.
51. MENG EC, SHOICHET BK, KUNTZ ID. Automated docking with grid-based energy evaluation. *J Comput Chem*. 1992; 13:505-524.
52. WEINER SJ, *et al.* A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J Am Chem Soc*. 1984; 106:765-784.
53. BÖHM HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*. 1994; 8(3):243-256.
54. BÖHM HJ. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des*. 1998; 12(4):309-323.
55. TANAKA S & SCHERAGA HA. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*. 1976 9(6):945-950.

56. SIPPL MJ. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol.* 1990; 213(4):859-883.
57. HU L, BENSON ML, SMITH RD, LERNER MG, CARLSON HA. Binding MOAD (Mother Of All Databases). *Proteins.* 2005; 60(3):333-340.
58. WANG R, FANG, X, LU Y, WANG S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. *J Med Chem.* 2004; 47(12):2977-2980.
59. DE AZEVEDO WF JR & DIAS R. Evaluation of ligand-binding affinity using polynomial empirical scoring functions. *Bioorg Med Chem.* 2008; 16(20):9378-9382.
60. ZAR JH. Significance Testing of the Spearman Rank Correlation Coefficient. *J Amer Statist Assoc.* 1972; 67(339):578-580.
61. MITRASINOVIC PM. Towards an Experimental and Systems Biology Framework for Cancer Cell Therapeutics. *Curr Bioinformatics.* 2012; 7:490-504.
62. MITRASINOVIC PM. Progress in structure-based design of EGFR inhibitors. *Curr Drug Targets.* 2013; 14:817-829.
63. HUANG SY, GRINTER SZ, ZOU X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 2010; 12:899-908.
64. IRWIN JJ & SHOBICHET. ZINC – A free database of commercially available compounds for Virtual Screening. *Journal of Chemical Information and Modeling.* 2005; 45(1): 177-182.
65. CERQUEIRA NMFS, *et al.* Receptor-based virtual screening protocol for drug discovery. *Archives of Biochemistry and Biophysics.* 2015; 582:1-12.

66. XAVIER MM, *et al.* SAnDReS a Computational Tool for Statistical Analysis of Docking Results and Development of Scoring Functions. *Combinatorial Chemistry & High Throughput Screening*. 2016; 19:1-12.
67. KUMAR A & SIDDIQI MI. New molecular scaffolds for the design of *Mycobacterium tuberculosis* type II dehydroquinase inhibitors identified using ligand and receptor based virtual screening. *J Mol Model*. 2010; 16:693-712.
68. MYSINGER MM, CARCHIA M, IRWIN JJ, SHOICHET BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry*. 2012; 55(14):6582-6594.

ANEXO A

Lista dos trinta e dois protocolos utilizados durante o processo de *re-docking* do projeto realizado. Cada um dos protocolos consiste na integração entre as FE's (*Scoring Functions*) e o algoritmo de busca (*Search algorithm*). Devido a presença de água em todos os sistemas biológicos, foram testados, principalmente, protocolos que usam a água como fator determinante da equação.

Protocol	Scoring functions	Search algorithm	Displaceable water?	RMSD (Å)	RMSD (Å)	RMSD (Å)	RMSD (Å)
Sorting criteria				MolDock Score	Rerank Score	HBond	RMSD
1	MolDock Score	MolDock Optimizer	Yes				
2	MolDock Score	MolDock Optimizer	No				
3	MolDock Score	MolDock (SE)	Yes				
4	MolDock Score	MolDock (SE)	No				
5	MolDock Score	Iterated Simplex	Yes				
6	MolDock Score	Iterated Simplex	No				
7	MolDock Score	Iterated Simplex (ANT)	Yes				
8	MolDock Score	Iterated Simplex (ANT)	No				
9	MolDock Score [GRID]	MolDock Optimizer	Yes				
10	MolDock Score [GRID]	MolDock Optimizer	No				
11	MolDock Score [GRID]	MolDock (SE)	Yes				
12	MolDock Score [GRID]	MolDock (SE)	No				
13	MolDock Score [GRID]	Iterated Simplex	Yes				
14	MolDock Score [GRID]	Iterated Simplex	No				
15	MolDock Score [GRID]	Iterated Simplex (ANT)	Yes				
16	MolDock Score [GRID]	Iterated Simplex (ANT)	No				
17	PLANTS Score	MolDock Optimizer	Yes				
18	PLANTS Score	MolDock Optimizer	No				

19	PLANTS Score	MolDock (SE)	Yes				
20	PLANTS Score	MolDock (SE)	No				
21	PLANTS Score	Iterated Simplex	Yes				
22	PLANTS Score	Iterated Simplex	No				
23	PLANTS Score	Iterated Simplex (ANT)	Yes				
24	PLANTS Score	Iterated Simplex (ANT)	No				
25	PLANTS Score [GRID]	MolDock Optimizer	Yes				
26	PLANTS Score [GRID]	MolDock Optimizer	No				
27	PLANTS Score [GRID]	MolDock (SE)	Yes				
28	PLANTS Score [GRID]	MolDock (SE)	No				
29	PLANTS Score [GRID]	Iterated Simplex	Yes				
30	PLANTS Score {GRID}	Iterated Simplex	No				
31	PLANTS Score [GRID]	Iterated Simplex (ANT)	Yes				
32	PLANTS Score [GRID]	Iterated Simplex (ANT)	No				