



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós Graduação em Química

Uso de descritores químico-quânticos como etapa de pós-processamento do protocolo de redesenho de enzimas

Wenison Marrone Souza Farias

João Pessoa

2022



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós Graduação em Química

Wenison Marrone Souza Farias

**Uso de descritores químico-quânticos como etapa de
pós-processamento do protocolo de redesenho de
enzimas**

Dissertação apresentada à Universidade Federal da Paraíba como parte dos requisitos para a obtenção do título de Mestre em Química.

Orientador: Gerd Bruno da Rocha

João Pessoa

2022

Catálogo na publicação
Seção de Catalogação e Classificação

F224u Farias, Wenison Marrone Souza.

 Uso de descritores químico-quânticos como etapa de pós-processamento do protocolo de redesenho de enzimas / Wenison Marrone Souza Farias. - João Pessoa, 2022. 85 f. : il.

 Orientação: Gerd Bruno da Rocha.
 Dissertação (Mestrado) - UFPB/CCEN.

 1. Química quântica. 2. Enzimas. 3. Catálise enzimática. 4. Descritores de reatividade. 5. Descritores moleculares. I. Rocha, Gerd Bruno da. II. Título.

UFPB/BC

CDU 544.18(043)



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE QUÍMICA EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Ata da Sessão Pública para avaliar o Trabalho de Dissertação de Mestrado do aluno **Wenison Marrone Souza Farias**, submetido à obtenção do grau de Mestre em Química – **Área de Concentração em FÍSICO-QUÍMICA**, deste Programa de Pós-Graduação em Química, do CCEN/UFPB.

Aos vinte e oito dias do mês de julho do ano de dois mil e vinte e dois, às catorze horas, reuniram-se em cerimônia pública em modelo webconferência no endereço eletrônico <https://meet.google.com/kzi-kvix-aox>, divulgado amplamente por e-mail e no SIGAA, os membros da comissão constituída na forma e termos do Artigo 82º do Regulamento Geral dos Programas de Pós-Graduação *Stricto Sensu* da Universidade Federal da Paraíba, composta pelos professores: **Dr. Gerd Bruno da Rocha**, do Departamento de Química, do Centro de Ciências Exatas e da Natureza, da Universidade Federal da Paraíba, João Pessoa-PB, Orientador e Presidente da Banca Examinadora; **Dr. Luciano Tavares da Costa**, do Instituto de Química, da Universidade Federal Fluminense, Niterói-RJ (Examinador) e **Dra. Karen Caclda Weber**, do Departamento de Química, do Centro de Ciências Exatas e da Natureza, da Universidade Federal da Paraíba, João Pessoa-PB (Examinador) a fim de examinarem o mestrando **Wenison Marrone Souza Farias** ao grau de mestre em química, **área de concentração em FÍSICO-QUÍMICA**. Além dos examinadores e do examinado, compareceram remotamente na webconferência também representantes do corpo docente e do corpo discente. Iniciando a sessão, o professor **Gerd Bruno da Rocha**, na qualidade de Presidente da referida banca, comunicou aos presentes qual o fim da reunião e os procedimentos de encaminhamentos da mesma. A seguir, concedeu à palavra ao mestrando para que fizesse oralmente a exposição do seu trabalho de dissertação de mestrado, sob o título **“Uso de descritores químico-quânticos como etapa de pós-processamento do protocolo de redesenho de enzimas”**. Concluída a exposição, o senhor presidente solicitou que fosse feita a argüição por cada um dos examinadores. Após o que, foi concedida a palavra ao mestrando para que respondesse e esclarecesse as questões levantadas. Terminadas as argüições, a Banca Examinadora passou a proceder à avaliação e julgamento do mestrando. Em seguida, o senhor presidente declarou que a Banca Examinadora considerou, segundo o Artigo 83º do Regulamento Geral dos Programas de Pós-Graduação *Stricto Sensu*, da Universidade Federal da Paraíba, **Aprovada**, a dissertação apresentada e defendida pelo mestrando **Wenison Marrone Souza Farias**, concedendo-lhe, assim, o grau de **Mestre em Química**, área de concentração em **FÍSICO-QUÍMICA**. Logo após a avaliação da comissão julgadora, o Orientador preencherá um formulário juntamente com o Relatório Final de Dissertação de Mestrado, a fim de que o mesmo seja homologado pelo Colegiado do Programa de Pós-Graduação em Química. E, para constar, eu, **Dr. Marcos Antonio Gomes Pequeno**, secretário da reunião, lavrei a presente Ata, que assino junto com os membros da Banca Examinadora. João Pessoa, 28 de julho de 2022. E, segue com as assinaturas de:

Dr. Marcos Antonio Gomes Pequeno
Secretário

Prof. Dr. Gerd Bruno da Rocha
Orientador/Presidente

Prof. Dr. Luciano Tavares da Costa
Examinador

Prof. Dra. Karen Caclda Weber
Examinadora

Assinaturas da Banca realizada em modo Webconferência, digitalizadas e certificadas pelo Prof. Dr. Gerd Bruno da

Rocha (SIAPE 1520134) em 28/07/2022 _____



UNIVERSIDADE FEDERAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO GERAL DE PÓS-GRADUAÇÃO

FICHA DE AVALIAÇÃO

EXAME DE: DISSERTAÇÃO DE MESTRADO

PROGRAMA

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ÁREA DE CONCENTRAÇÃO

FÍSICO-QUÍMICA

TÍTULO DO TRABALHO

Uso de descritores químico-quânticos como etapa de pós-processamento do protocolo de redesenho de enzimas

ORIENTADOR

Prof. Dr. Gerd Bruno da Rocha

OBSERVAÇÕES DA BANCA EXAMINADORA

A defesa seguiu normalmente, onde o discente expôs seu trabalho no tempo estipulado e respondeu todas as perguntas feitas pela banca de forma apropriada. As correções no texto foram apresentadas pelos membros da banca e o discente concordou em executá-las para a versão final da dissertação. A banca foi unânime em considerá-lo aprovado.

A BANCA EXAMINADORA, TENDO EM VISTA A EXPOSIÇÃO ORAL APRESENTADA PELO ALUNO E PROCEDIDA A ARGÜIÇÃO PERTINENTE AO TRABALHO FINAL, CONSIDEROU O MESTRANDO Wenison Marrone Souza Farias [x] APROVADO [] REPROVADO

MEMBROS DA BANCA	INSTITUIÇÃO DE ORIGEM
Dr. Gerd Bruno da Rocha	UFPB/JPA-PB
Dr. Luciano Tavares da Costa	UFF/Niterói-RJ
Dra. Karen Cacilda Weber	UFPB/JPA-PB

LOCAL
JOÃO PESSOA-PB

HORA/DATA
14:00 – 28/07/2022

INSTRUÇÕES À BANCA EXAMINADORA

O conceito da avaliação deverá ser expresso como: **Aprovado e Reprovado**.
A avaliação é feita logo após o encerramento da exposição oral e argüição do candidato.
Caso seja sugerida reformulação do trabalho final, a **Banca Examinadora** deverá estabelecer um prazo disponível para o aluno procedê-la.
Após o preenchimento desta **Ficha de Avaliação**, a mesma deverá ser entregue a **Secretaria de Pós-Graduação em Química, pelo Presidente da Banca Examinadora**.

Assinaturas da Banca realizadas em modo Webconferência, digitalizadas e certificadas pelo Prof. Dr. Gerd Bruno da

Rocha (SIAPE 1520134) em 28/07/2022

Gerd Bruno da Rocha

Uso de descritores químico-quânticos como etapa de pós-processamento do protocolo de redesenho de enzimas.

Dissertação de Mestrado apresentada pelo aluno Wenison Marrone Souza Farias e aprovada pela banca examinadora em 28 de julho de 2022.

Gerd Bruno da Rocha

Prof. Dr. Gerd Bruno da Rocha
Departamento de Química – CCEN/UEPB
Orientador/Presidente

Luciano Tavares da Costa

Prof. Dr. Luciano Tavares da Costa
IQ/UEFF/Niterói-RJ
Examinador Externo

Profa. Dra. Karen Caclda Weber
Departamento de Química – CCEN/UEPB
Examinadora Interno

Assinaturas da Banca realizadas em modo Webconferência, digitalizadas e certificadas pelo Prof. Dr. Gerd Bruno da Rocha (SIAPE 1520134) em 28/07/2022

Gerd Bruno da Rocha

“Com grandes poderes, vêm grandes responsabilidades.”

– Tio Ben

Agradecimentos

Agradeço a Deus por tudo e meus amigos pelos bons momentos. A minha família pelo apoio que sempre me deram durante toda a minha vida.

Agradeço ao meu orientador professor Dr. Gerd Bruno da Rocha pelo suporte e apoio na elaboração dessa dissertação.

Agradeço a bolsa do CNPq que possibilitou o auxílio financeiro.

Agradeço a UFPB por fornecer todo o ambiente possibilitando minha formação.

Resumo

Na busca constante pela redução da energia consumida em processos na indústria, o uso de catalisadores tem se desenvolvido muito, dentre os quais, as enzimas são os que possuem maior performance nesse sentido em meios homogêneos. Em busca desses catalisadores para reações industriais, uma ampla gama de estudos teóricos focado no desenvolvimento de enzimas veem sendo conduzidos, como o protocolo de *design* enzimático Rosetta, que promovendo mutações em estruturas proteicas buscam criar ou melhorar as reações por elas catalisadas. Nessa dissertação, buscamos contribuir para o melhoramento desse protocolo, pois, o mesmo ainda carece de aperfeiçoamentos na etapa de seleção das prováveis estruturas com capacidade catalítica. Com isso em mente, propomos uma etapa adicional ao protocolo de *design* baseado no pós-processamento do complexo enzima-substrato pelo uso de entalpia de ligação enzima-ligante ($\Delta H_{binding}$) e descritores químico-quânticos de reatividade. Desse modo, desenvolvemos uma metodologia que consiste em, após obter as estruturas redesenhadas pelo Rosetta, realizar cálculos quânticos *single-point* utilizando o Hamiltoniano semiempírico, PM7, com efeito de solvatação (modelo COSMO) e o algoritmo de escalonamento linear MOZYME, obtendo, assim, a entalpia de ligação enzima-ligante. Em seguida, após obter a função de onda da enzima, são calculados os descritores de reatividade pela aproximação dos orbitais congelados além de considerar o *gap* HOMO-LUMO como uma banda, BD, calculando também a sua forma ponderada, EW, através do programa PRIMoRDiA. Para validar essa metodologia buscou-se na literatura, estruturas redesenhadas pelo Rosetta e que possuíam dados de ensaios enzimáticos disponíveis. Como resultado, ao utilizar a entalpia de ligação calculada de uma série de estruturas obtidas pelo protocolo Rosetta, conseguiu-se melhorar a classificação, antes ranqueada pela pontuação do Rosetta. Desse modo, melhorando para reação de eliminação de *Kemp* 17 colocações e para a reação de *Diels-Alder* deixando entre as dez melhores colocações. Com relação aos descritores de reatividade, para a reação de *Kemp* foi possível observar que a mudança de apenas um resíduo, na estrutura supleKE59, em relação à estrutura que apresentou maior capacidade catalítica, nativeKE59, fez com que a propriedade NAS (*Nucleophilic Attack Susceptibility*) se afaste do ligante, desse modo, indicando a inatividade catalítica da estrutura supleKE59. Portanto, a combinação dessas duas abordagens pode auxiliar no ranqueamento de estruturas melhores e, dentre elas, ajudar a selecionar as que possuem suas propriedades NAS e EAS localizados próximas ao sítio-ativo.

Palavras-chave: Reclassificação de poses, Desenho Enzimático, Descritores Químico-Quânticos, Métodos Semiempíricos.

Abstract

In a constant search for the reduction of energy consumed, the use of catalysts has developed a lot, among which enzymes are the ones that have the highest performance in a homogeneous environment. In search of these catalysts for industrial reactions, a wide range of theoretical studies focused on enzyme development have been developed, such as the Rosetta enzymatic *design* protocol, which promotes mutations in protein structures and seeks to create or improve the reactions that it catalyzes. In this study, we try to improve this protocol, presenting improvements in the selection of probable structures with catalytic capacity and thus propose an additional step to the *design* protocol based on the post-processing of the enzyme-substrate complex by the use of enzyme-binding enthalpy ($\Delta H_{binding}$) and reactivity descriptors. Thus, we developed a method that comprises using the structures designed by Rosetta and re-ranking them by performing quantum chemical calculations of $\Delta H_{binding}$ through the semiempirical Hamiltonian PM7 considering solvation effect (COSMO model) and the MOZYME algorithm. Additionally, from the wave function of the enzymes, reactivity descriptors were calculated by using the frozen orbitals approximation considering molecular orbitals near gap HOMO-LUMO as a band, BD, or calculating its weighted formulation, EW. To validate this method, we sought the literature, structures designed by Rosetta with their kinetic characterization performed. As a result, by using the calculation of $\Delta H_{binding}$, it was possible to improve the ranking, previously ranked by Rosetta's score. Thus improving for elimination reaction of *Kemp* 17 positions and for the reaction of *Diels-Alder* leaving in the top 10 places. Regarding the reactivity descriptors, for the reaction of *Kemp*, it was possible to observe that the change of only one residue in the supleKEstructure59 in relation to the structure that presented the highest catalytic capacity, nativeKE59, caused the NAS property to move away from the ligand, thus indicating the catalytic inactivity of the supleKE structure59. Therefore, the combination of these two approaches contributed to the ranking of better structures and, among them, aided to select those that have their NAS and EAS quantities located at the enzymes' active site.

Keywords: Reclassification of poses, Enzyme Design, Quantum-chemical Descriptors, Semiempirical methods.

Sumário

Agradecimentos	iii
Resumo	v
Abstract	vii
Lista de Abreviações	xv
1 Introdução	1
2 Objetivo	5
2.1 Objetivo Geral	5
2.2 Objetivos Específicos	5
3 Revisão Bibliográfica	7
3.1 Enzimas e catálise enzimática	7
3.2 <i>Design</i> racional de enzimas	8
3.3 Método evolutivo de enzimas	17
3.4 Métodos de Química Quântica	20
3.5 Descritores de reatividade	22
4 Metodologia	27
5 Resultados e Discussão	33
5.1 Usando o $\Delta H_{binding}$ para <i>rescoring</i>	33
5.2 Descritores moleculares	39
6 Conclusões	49
7 Perspectivas	51
Referências	52
Anexo	59

Lista de Figuras

3.1	Grupos funcionais de um aminoácido genérico: grupo amina, grupo carboxil, carbono- α e R da cadeia lateral. Fonte: o autor.	8
3.2	Representação esquemática do protocolo de <i>design</i> Rosetta. Fonte: (<i>Richter et al.</i> , 2011).	10
3.3	Representação do arquivo de restrição .cst. Fonte: o autor.	12
3.4	Esquema do PCR. Fonte: (<i>Biolabs</i> , 2022).	18
3.5	Ideia que envolve a mudança de rota catalítica na evolução dirigida (<i>Heckmann and Paradisi</i> , 2020)	20
4.1	Fluxograma do protocolo usado nesse estudo.	28
4.2	As linhas RT e o PR são responsáveis por especificar o tipo de cálculo e os <i>outputs</i> gerados pelo PRIMorDiA configurado para leitura no pymol. A terceira linha especifica o cálculo de banda e o número máximo de orbitais moleculares a ser utilizados pelo método BD, onde foi utilizado 10.	29
4.3	O RT e o PR são as linhas responsáveis por especificar o tipo de cálculo e os <i>outputs</i> gerados com configuração de pymol. A terceira coluna especifica o método de cálculo de descritores para macromoléculas ponderado por energia e a banda foi definida como eband 5, indicando que foi considerados os orbitais moleculares próximos ao HOMO-LUMO dentro de 5eV.	29
5.1	Exemplo de pontuações do Rosetta.	34
5.2	Relação <i>Score</i> Rosetta e $\Delta H_{binding}$ para reação de eliminação de Kemp.	36
5.3	Relação <i>Score</i> Rosetta e $\Delta H_{binding}$ para reação de <i>Diels-Alder</i>	36
5.4	Alinhamento da estrutura KE59-native, sendo a forma catalítica, com a estrutura não catalítica nomeada KE59.pdb.	40
5.5	As 4 estruturas desenhadas segundo o artigo referente à reação de eliminação de <i>Kemp</i> pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. A KE70 apresentou NAS no ligante e EAS no resíduo Met-112. KE07 EAS no resíduo Met-186. NativeKE59 EAS no resíduo Met-153. SupleKE59 NAS no resíduo Pro-1 e EAS no Met-153.	41

5.6	<i>Design</i> da reação de eliminação de <i>Kemp</i> segundo a pontuação do Rosetta pelos descritores que expressa a propriedade Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. A estrutura rep_193.39_UM_101_E211W89 apresentou propriedade EAS no resíduo Met-130 e o NAS localizado no ligante.	42
5.7	<i>Design</i> da reação de eliminação de <i>Kemp</i> segundo a pontuação $\Delta H_{binding}$ pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. A estrutura rep_198.07_UM_46_E159W51 apresentou propriedade EAS no resíduo Met-183 e NAS localizado no ligante.	43
5.8	As 4 estruturas BestHit segundo o <i>Design</i> da reação de <i>Diels-Alder</i> utilizando a propriedade do descritor de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. BestHit_rep_222.7_UM_2_Y121Q195 EAS no resíduo Met-90 e NAS no resíduo Cys-157. BestHit_rep_249.41_UM_2_Y121Q195 EAS no resíduo Met-231 e NAS no resíduo Cys-119. BestHit_UM_2_Y121Q195 EAS no resíduo Met-231 e NAS no resíduo Cys-119. BestHit_rep_225.74_UM_2_Y121Q195 EAS no resíduo Met-90 e NAS no resíduo Cys-157.	44
5.9	<i>Design</i> da reação de <i>Diels-Alder</i> segundo a pontuação do $\Delta H_{binding}$ utilizando a propriedade do descritor de Netfilicidade na forma EW. A estrutura UM_90_T195Q120 EAS no resíduo Cys-265 e NAS no resíduo Ala-173-174.	45
5.10	Essa estrutura é obtida usando como base a estrutura BestHit-6M da reação de <i>Diels-Alder</i> pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. O EAS no resíduo Cys-173 e NAS no resíduo Arg-91. Foram alterados seis mutações, sendo A21T, A74I, Q149R, A173C, S271A, e A272N.	46
5.11	As 5 estruturas desenhadas segundo o artigo referente a reação de <i>Retro-Aldol</i> pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. RA22 EAS no resíduo Met-236 e NAS no resíduo Pro-1. RA34 EAS no resíduo Met-199 e NAS localizados no resíduo Asp-127-Ser-31-Ile-30-Pro105. RA46 EAS no resíduo Met-153 e NAS no resíduo Pro-1. RA60 EAS no resíduo Met-163 e NAS no resíduo Pro-125. RA61 EAS no resíduo Met-163 e NAS no resíduo Pro-125.	47
5.12	A RA34 estrutura desenhada segundo o artigo referente a reação de <i>Retro-Aldol</i> pelos descritores de NAS na forma BD. O NAS nos resíduos Pro-56 e Trp-57 e EAS localizado nos resíduos Met-52-236. . . .	48

Lista de Tabelas

3.1	Código dos resíduos de aminoácidos	11
4.1	Constantes de velocidade catalítica para a reação de eliminação <i>Kemp</i>	30
4.2	Constantes de velocidade catalítica para a reação de <i>Diels-Alder</i>	31
4.3	Constantes de velocidade catalítica para a reação de <i>Retro-Aldol</i>	31
5.1	Top 10 reação de eliminação de Kemp	35
5.2	Top 10 Reação de Diels-Alder	37
5.3	Top 10 eliminação de Kemp e suas mutações	37
5.4	Top 10 Reação de Diels-Alder e suas mutações	37
5.5	Top 10 eliminação de Kemp e as contabilização das interações	38
5.6	Top 10 Reação de Diels-Alder e as contabilização das interações	38
1	Reação de eliminação de Kemp	60
2	Reação de Diels-Alder	61

Lista de Abreviações

Abreviação	Descrição
$\Delta H_{binding}$	Entalpia de ligação enzima-ligante
Arquivo cst	Arquivo de restrição para formulação da <i>Theozima</i>
PCR	Reação em cadeia da polimerase
epPCR	Reação em cadeia da polimerase propensa a erros
CASTing	teste de site ativo combinatório
DFT	Teoria do funcional da densidade
ProSAR	relações de atividade de sequência de proteína
PRIMorDIA	Acesso aos Descritores de Reatividade Macromolecular PRIMorDiA
EAS	Suscetibilidade de ataque eletrofílico
NAS	Suscetibilidade de ataque nucleofílico
Netfilicidade	Agrupa as propriedades NAS E EAS
BD	Calculo dos descritores de banda
EW	Calculo dos descritores de banda na forma ponderada
k_{cat}	Constante de velocidade catalítica
PDB	Banco de dados de proteínas
QSAR	Relações Quantitativas Estrutura-Atividade
QSPR	Relações Quantitativas Estrutura-Propriedade
Arb.U	Unidade Arbitrária

Capítulo 1

Introdução

As enzimas são catalisadores biológicos eficazes capazes de acelerar uma reação química específica mais de 14 ordens de grandeza (*Radzicka and Wolfenden, 1995*). Apresentam melhor controle sobre a estereoquímica, regioseletividade e especificidade, sendo estas características atrativas para aplicações industriais e comerciais. No entanto, poucas enzimas são viáveis para aplicação industrial de forma útil e economicamente vantajosa (*Bornscheuer and Pohl, 2001*). Assim, muitos esforços estão atualmente engajados no *design* enzimático para atender às necessidades comerciais atuais (*Chica et al.*).

As enzimas normalmente se apresentam em solução na forma globular, forma esta devido à predileção dessas estruturas para se configurarem como uma esfera (*Engelking and Engelking, 2015*). Seus pesos moleculares variam de 10^4 a 10^6 Da (*Bhatia and Bhatia, 2018*), sendo categorizadas de acordo com suas funções, que variam em seis classes principais: oxirredutase (enzimas que catalisam reações de transferência de elétrons), transferases (enzimas que realizam transferência de grupos), hidrolases (quebra ligação de compostos através da utilização de água), liases (adição de grupos às ligações duplas ou remoção desse grupo formando ligação dupla), isomerases (catalisam mudanças estruturais da molécula), ligases (enzimas que catalisam a formação de ligação entre moléculas) (*Nelson and Cox, 2018*).

O uso de biocatalisadores enzimáticos na indústria podem ser benéficos para várias plantas produtivas, em vários aspectos, tais como: promovendo a diminuição do uso energético, aumentando o rendimento e qualidade do produto, além de reduzir o impacto ambiental, o que acaba por promover setores que se apoiam na química verde (*Ramos, 2009*).

As enzimas, em muitos casos, são capazes de substituir estratégias de produção caras e poluentes por condições reacionais notavelmente mais brandas (*Frazzetto, 2003*). No estudo de 2016 de Rajendra Singh et. al. (*Singh et al., 2016*), os autores estimaram em 2014 que a indústria enzimática gerou cerca de 4,2 bilhões de dólares e que no período de 2015 a 2020 chegaria a 6,2 bilhões de dólares, sendo amplamente

usadas em processos industriais, como panificação, fabricação de cerveja, detergentes, produtos fermentados, produtos farmacêuticos, têxteis como processamento de couro (*Singh et al.*, 2016).

O mecanismo de ação enzimática atua principalmente na estabilização do estado de transição da reação catalisada (*Warshel et al.*, 2006) e redução de sua barreira de ativação (*Cox and Nelson*, 2013). Essa redução juntamente com o fato de biocatalizadores funcionarem em condições amenas de temperatura, pressão e pH são os motivos pelos quais a indústria diminuiu consideravelmente os gastos energéticos e conseguiu, em termos cinéticos, um robusto melhoramento. Sabendo disso, desenvolveram-se projetos de redesenho de enzimas, buscando utilizar modelos empíricos e teóricos para modificar sequências de aminoácidos, com o objetivo de melhorar ambos os atributos (*Heckmann and Paradisi*, 2020).

Devido a esses ganhos, muitas propostas de melhorias de enzimas surgiram e ainda ocorrem, por exemplo, o projeto de Justin B. Siegel que desenvolveu uma reação inexistente em ocorrência natural, a reação de Diels-Alder, que infelizmente não atingiu um robusto poder catalítico, porém, mostrou a possibilidade de desenhar reações enzimáticas inéditas (*Siegel et al.*, 2010). Outro estudo foi o de Röthlisberger *et al.* (*Röthlisberger et al.*, 2008) que desenvolveu enzimas para a reação de eliminação de *Kemp* (formação de um grupo nitrila), apresentando capacidade catalítica de 10^6 em relação à reação não catalisada combinando duas técnicas distintas. Os autores usaram conjuntamente para isso a técnica de *design* racional, através de *software*, e a evolução dirigida (*Röthlisberger et al.*, 2008).

Graças aos avanços tecnológicos, o *design* racional hoje pode ser realizado por meio de *softwares*, trazendo todo um arcabouço derivado de conhecimentos experimentais e cálculos quânticos, se tornando cada vez mais eficiente. Entre os programas mais populares está o protocolo Rosetta de *design* de enzima (*Heckmann and Paradisi*, 2020). De forma resumida, esse protocolo consiste na descrição minimalista do estado de transição do ligante rodeado pelos aminoácidos catalíticos posicionados de forma ideal para a catálise (*Jiang et al.*, 2008). O protocolo ainda está em desenvolvimento e, sob certos aspectos, cabe ainda melhoramentos e nesse atual estágio de desenvolvimento é bastante comum o emprego da estratégia de evolução dirigida como complemento em etapas posteriores ao Rosetta (*Röthlisberger et al.*, 2008).

A combinação da técnica de *design* de enzimas e evolução dirigida apresenta resultados robustos (*Packer and Liu*, 2015). A abordagem de projeto de enzimas conhecida por evolução dirigida consiste na inspiração do método adotado pela natureza, evolução por variabilidade mutagênica e seleção adaptativa, ou seja, através de mudanças aleatórias nos genes correspondentes a aminoácidos específicos são gerados um grande conjunto de enzimas com diversas substituições. Em seguida, todas as enzimas mutantes são testadas e aquelas que apresentarem uma melhor capacidade catalítica são selecionadas, como ponto de partida, para outros ciclos mutagênicos. Com isso, as mutações que forem benéficas são mantidas enquanto se busca outras para melhorar a eficiência (*Strohmeier et al.*, 2011).

Inspirado na observação expressa no artigo *Computational Design of Synthetic Enzymes* (Vaissier Welborn and Head-Gordon, 2018) à respeito do protocolo de *design* enzimático do Rosetta, este trabalho de mestrado teve o *insight* necessário para ser realizado. O que se observou ao ler os trabalhos de revisão dessa área é que o desempenho geral desses *designs* não foram muito bons, recorrendo muitas vezes a técnica de mutação dirigida para alcançar melhores resultados. Com isso buscamos desenvolver uma metodologia, utilizando métodos quânticos, capazes de fornecer informações para melhor selecionar as possíveis enzimas com potencial catalítico. Desse modo, estamos propondo um melhoramento no *rescoring* para o protocolo Rosetta, utilizando de forma combinada descritores de reatividade e o cálculo da entalpia de ligação enzima-substrato ($\Delta H_{binding}$).

Capítulo 2

Objetivo

2.1 Objetivo Geral

Propor uma etapa adicional ao protocolo de *design* de enzimas Rosetta baseada no pós-processamento de complexos enzima-substrato pelo uso de descritores de reatividade e entalpia de ligação enzima-ligante ($\Delta H_{binding}$).

2.2 Objetivos Específicos

- Reproduzir dados publicados de *design* de enzimas de outros estudos.
- Melhorar a classificação de enzimas modificadas pelo protocolo do Rosetta utilizando entalpia de ligação enzima-ligante.
- Identificar, através de descritores quânticos, as estruturas com maior potencial para catálise enzimática através da localização dos orbitais HOMO-LUMO.

Capítulo 3

Revisão Bibliográfica

3.1 Enzimas e catálise enzimática

Enzimas na natureza

Ao longo dos anos, os humanos cultivaram microrganismos e todo seu aparato enzimático, especialmente aqueles relacionados à produção de açúcares fermentados com etanol, hidromel, cerveja, vinho e pão. Hoje, temos os recursos tecnológicos para superar a fermentação e usar esses catalisadores para fabricar medicamentos, biocombustíveis, produtos químicos e terapia de reposição enzimática como remédio para problemas metabólicos (Neufeld, 2006).

Na indústria de biotecnologia, uma alta estabilidade estrutural referente às enzimas é uma característica valiosa, pois, tem um grande impacto na quantidade de enzimas e no tempo de processamento necessário, afetando diretamente os custos. Com o aumento da estabilidade (Cao *et al.*, 2015), há a possibilidade de elevação da temperatura de atuação enzimática, o que por sua vez irá acelerar a conversão do substrato, pois permite que mais substrato atinja o estado de transição, que é a etapa que requer maior energia empenhada para que o reagente seja convertido em produto (Klibanov, 1983). Sendo assim, se faz necessário conhecer todos os fenômenos químicos responsáveis pela estabilidade e funcionalidade enzimática.

Estrutura Enzimática

Em 1961, Anfinsen e colaboradores mostraram que os peptídeos podiam se dobrar espontaneamente *in vitro* e hoje se sabe que tanto a estrutura quanto estabilidade dependem da sua sequência de aminoácidos (Godoi Contessoto *et al.*, 2018). Os aminoácidos são formados por 4 grupos: α -amino, α -carboxil, α -carbono e a cadeia lateral, conforme mostrado na Figura 3.1.

A união desses aminoácidos, considerando suas diversidades estruturais e químicas na tabela 3.1, unido por ligações peptídicas gera um polipeptídeo (polímero

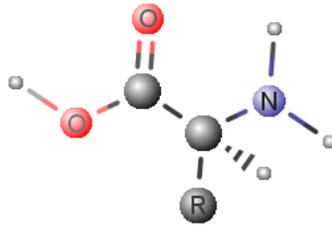


Figura 3.1: Grupos funcionais de um aminoácido genérico: grupo amina, grupo carboxil, carbono- α e R da cadeia lateral. Fonte: o autor.

de aminoácidos) conhecida como estrutura primária, que se dobra em uma estrutura tridimensional (3D) específica, denominada estrutura terciária ou conformação natural. Algumas proteínas possuem a capacidade de acelerar reações sendo nomeadas de enzimas. Se essa estrutura tridimensional mudar por qualquer motivo, a enzima perderá sua atividade e função, sendo assim, para o bom funcionamento é necessário que a enzima mantenha sua forma nativa (*Anfinsen et al.*, 1961).

3.2 *Design* racional de enzimas

O *design* racional consiste na alteração de resíduos de aminoácidos utilizando conhecimento empírico ou teórico a respeito da estrutura e função, ou seja, a partir de um conhecimento previamente estabelecido. Essa abordagem vem se tornando mais eficiente graças ao desenvolvimento computacional com que diversas técnicas baseadas em mecânica molecular ou mecânica quântica permitem entender melhor a estrutura enzimática, sua função e dinâmica. Outro ponto relevante é a técnica experimental de cristalografia de raio-x que, pela estrutura elucidada, permite entender a configuração nativa da enzima. Essa abordagem apresentou bons resultados e no estudo de James A. Wells et. al. foi possível projetar uma protease subtilisina BPN pela otimização racional das interações eletrostáticas do substrato aos aminoácidos carregados, após 16 mutações a enzima aumentou a capacidade catalítica em 1900 vezes em relação à reação não catalisada (*Wells et al.*, 1987).

Outro uso do *design* racional foi realizado pelos pesquisadores N S Scrutton et. al., que pela substituição sistemática dos resíduos presentes na glutatona reductase, uma enzima responsável pela especificidade da coenzima NADP⁺, foi possível alterar a preferência da coenzima para NAD⁺, sem que com isso fosse alterada a especificidade da enzima para com o substrato (*Scrutton et al.*, 1990). Outras formas de *design* racional muito difundida é através de *softwares*. Os trabalhos de Daniela Röthlisberger et. al (*Röthlisberger et al.*, 2008) e Justin B. Siegel et. al (*Siegel et al.*, 2010) conseguiram, utilizando o protocolo de *design* Rosetta, demonstrar a capacidade dessa abordagem no desenho enzimático.

O primeiro trabalho teve como foco o projeto de uma enzima capaz de catalisar

a reação de eliminação de *Kemp*, que envolve essencialmente a remoção de prótons do carbono e a formação de um radical nitrila. Desse modo, usando o protocolo de *design* Rosetta, foi desenvolvida uma enzima chamada KE59 que apresentou uma capacidade catalítica de 10^5 em comparação com a reação não catalisada.

O segundo trabalho teve como objetivo desenvolver, utilizando o Rosetta, uma enzima capaz de catalisar a reação bimolecular de Diels-Alder, reação que consiste na adição do dieno conjugado ao dienófilo. Essa reação catalisada por enzimas não existe na natureza, mas após o redesenho apresentou atividade de $0.10 \pm 0.02 \text{ hr}^{-1}$ em relação à reação não catalisada, sendo confirmada por cristalografia de raios-X e caracterização cinética. Essa estrutura recebeu a identificação de DA_20_00. Nesse mesmo trabalho, utilizando a intuição química, foram alterados seis aminoácidos, alcançando uma catálise de $2.13 \pm 0.24 \text{ hr}^{-1}$, essa denominada de DA_20_10.

O sucesso do *design* Rosetta, mesmo ainda carecendo de melhorias, se divide em dois aspectos, o seu protocolo e sua função score. Ambas aplicadas visando selecionar as possíveis estruturas que apresentaram as melhores atividades catalíticas. Como o objetivo desse trabalho é propor melhoria no procedimento de ranqueamento e seleção desse *software*, seus aspectos fundamentais serão melhor apresentados nos tópicos a seguir.

Protocolo de *Design* Enzimático no Rosetta

Um dos mecanismos pelo qual as enzimas aceleram reações é a estabilização do estado de transição. Buscando aprimorar essa característica, o algoritmo de *design* enzimático Rosetta foi desenvolvido. O protocolo começa com uma descrição da geometria otimizada em um modelo de estado de transição cercado pelos resíduos catalíticos do sítio ativo de forma adequada para catálise, recebendo o nome de *theozima*. Em seguida, se utiliza o algoritmo *match* implementado no Rosetta para criar um conjunto de estruturas enzimáticas com o ligante acoplado em possíveis posições onde o sítio ativo será desenhado. Por fim, a etapa de *design* se inicia, variando-se os resíduos não catalíticos, visando aumentar a afinidade do sítio ativo ao ligante no estado de transição (*Richter et al.*, 2011). A figura 3.2 reflete o fluxograma referente ao funcionamento do protocolo de *design* Rosetta.

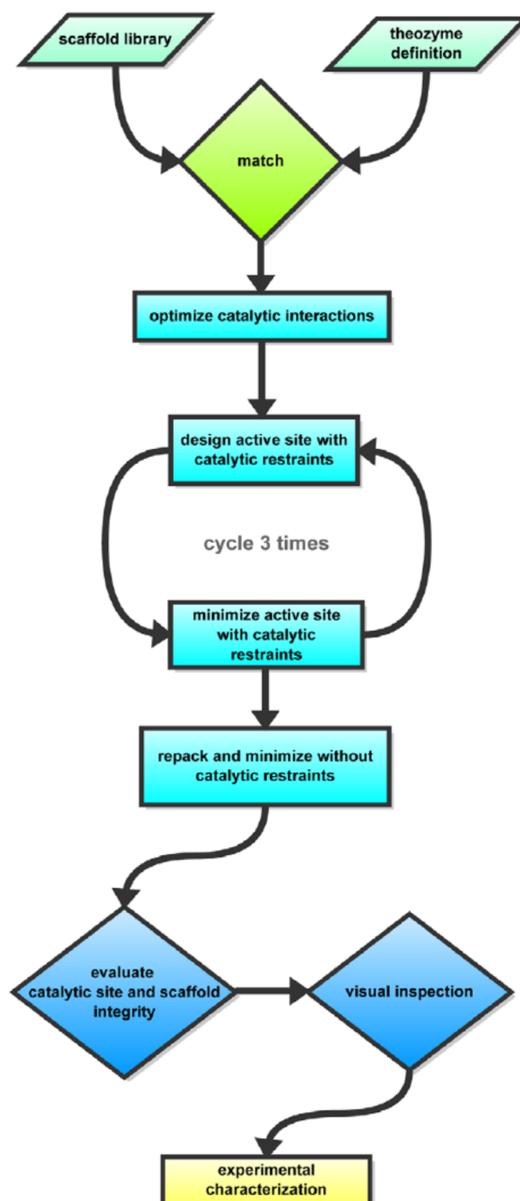


Figura 3.2: Representação esquemática do protocolo de *design* Rosetta. Fonte: (Richter *et al.*, 2011).

Theozima

A primeira etapa, citada anteriormente, consiste em modelar e avaliar o estado de transição do ligante posicionado junto aos resíduos catalíticos encontrados na literatura e otimizá-los, especialmente, através de cálculos quânticos. Definida a *theozima*, um arquivo de restrição geométrica chamada “arquivo cst” é gerado, consistindo em blocos - para cada interação entre o resíduo catalítico e o ligante. Esse arquivo é fundamental para o uso do algoritmo de pesquisa do Rosetta, *Matching*, pois, utiliza os dados desse arquivo para transplantar a *theozima* na estrutura

Tabela 3.1: Código dos resíduos de aminoácidos

Código de 1 letra	Código de 3 letra	Nome do resíduo
A	Ala	Alanina
C	Cys	Cistina
D	Asp	Aspartato (Ácido aspártico)
E	Glu	Glutamato (Ácido glutâmico)
F	Phe	Fenilalanina
G	Gly	Glicina
H	His	Histidina
I	Ile	Isoleucina
K	Lys	Lisina
L	Leu	Leucina
M	Met	Metionina
N	Asn	Aspargina
P	Pro	Prolina
Q	Gln	Glutamina (Glutamida)
R	Arg	Arginina
S	Ser	Serina
T	Thr	Treonina
V	Val	Valina
W	Trp	Triptofano (Triptofana)
Y	Tyr	Tirosina

cristalográfica de diversas enzimas/proteínas. Outro uso ocorre durante a etapa de *design*, sendo usado para restringir os resíduos catalíticos à geometria desejada durante a otimização e minimização baseada em gradiente do sítio ativo a ser projetado (*Richter et al.*, 2011).

A descrição resumida de um bloco no arquivo de restrição é expressa na figura 3.3 A primeira coluna é a distância do resíduo ao ligante. O valor de 3,06 Å é a distância; Próxima coluna é uma faixa de valores tolerável para a distância; A terceira coluna é a constante de força, usada na penalidade de pontuação; Para a quarta coluna, insere-se valor 1 ou 0, sendo 1 para ligação covalente e 0 para interação iônica; A última coluna é opcional e aumenta pontos de amostragem em $2n + 1$ (*Richter et al.*, 2011).

Já para os ângulos e torções as variáveis têm as seguintes utilidades: A primeira coluna é o ângulo do resíduo em relação ao ligante, que corresponde a 73,6°; A próxima coluna é a faixa aceitável para esse ângulo; A terceira coluna é a constante de força, usada na penalidade de pontuação; Já a quarta coluna é responsável por especificar a periodicidade da restrição. Por exemplo, se x_0 é 73,6° e a periodicidade, definida na imagem 3.3 como per , sendo 180°, a função de restrição terá 2 mínimos, um em 73,6° e outro em 253,6° graus; A última coluna é opcional semelhante a já explicada anteriormente (*Richter et al.*, 2011).

Por fim, a nomenclatura de *residue3* e *residue1* são, respectivamente, o código

```

#block 1 for glutamate proton abstracting residue

CST::BEGIN
  TEMPLATE::  ATOM_MAP: 1 atom_name: C3 C1 O6
  TEMPLATE::  ATOM_MAP: 1 residue3: 1n1

  TEMPLATE::  ATOM_MAP: 2 atom_type: OOC ,
  TEMPLATE::  ATOM_MAP: 2 residue1: ED

#Identificação das variáveis: x0  xtol  k  per  n
CONSTRAINT:: distanceAB:  3.06  0.2  100.  0  0
CONSTRAINT::  angle_A:   73.6  10.0  80.0  360.  1
CONSTRAINT::  angle_B:  120.0  15.0  80.0  360.  1
CONSTRAINT::  torsion_A: -101.2  15.0  60.0  360.  1
CONSTRAINT::  torsion_AB: 180.0  90.0  0.00  360.  3
CONSTRAINT::  torsion_B: 180.0  15.0  0.00  360.  1
CST::END

```

Figura 3.3: Representação do arquivo de restrição .cst. Fonte: o autor.

de 3 e 1 letra dos aminoácidos, visto na tabela 3.1, sendo, no caso do 1n1, a representação do ligante. Já para o `atom_type` e `atom_name` são responsáveis por identificar, respectivamente, o tipo atômico, como funções orgânicas, e os nomes dos átomos e suas posições numéricas, como carbono 3 e oxigênio 6 (*Richter et al.*, 2011).

A etapa de *Matching*

Nesta etapa, o módulo Rosetta *match* é usado para inserir a *theozima*, usando as informações do arquivo .cst em cada estrutura proteica cristalográfica. O acoplamento do ligante à cavidade da estrutura juntamente com os resíduos catalíticos não afeta a estrutura proteica a suas cadeias laterais.

Normalmente, apenas um conjunto de cavidade das proteínas são considerados durante o *match*, devido à dificuldade de gerar locais de ligação em regiões enterradas ou hidrofóbicas. Pois, se sabe que tentar criar um sítio de ligação muito enterrado à macromolécula é problemático devido à instabilidade estrutural gerada. Em segundo lugar, muitas cavidades exigirão maior poder computacional, especialmente recursos como memória e tempo de execução, muito provavelmente porque mais átomos serão necessários para o modelo estrutural. Por fim, o usuário deve preparar cada estrutura e selecionar qual dos dois algoritmos existentes no Rosetta será usado (*Richter et al.*, 2011).

O protocolo do Rosetta *match* se divide em dois algoritmos sendo nomeados por correspondência clássica e secundária, podendo ambas ser usadas na mesma rodada do protocolo. No caso do algoritmo de correspondência clássico, é necessário fornecer valores para os 6 parâmetros, presentes no arquivo .cst expressa na figura 3.3, que conectam a cadeia lateral e o ligante no estado de transição. Em suma, para

cada bloco presente no arquivo, o algoritmo constrói essa interação da *theozima* em cada posição possível na cavidade alvo. Para cada teste construído, é verificado se há colisões com o *backbone* da proteína, registrando, caso não haja alguma, a sua configuração em três coordenadas cartesianas e três euclidianas, recebendo o nome de *hit* (Richter et al., 2011).

Já para o algoritmo de correspondência secundária, esse consiste em gerar conformações dos resíduos de aminoácidos para cada bloco presente no arquivo .cst em toda cavidade alvo. Então, para cada conformação *r*, é avaliada sua geometria com *hit* gerado no algoritmo de correspondência clássica. Se a geometria da conformação *r* e do *hit* *h*, corresponder à geometria do bloco existente no arquivo .cst suas seis coordenadas são consideradas como uma interação sem colisão sendo registrada como um *hit* (Zanghellini et al., 2006).

Depois de gerar *hits* para todas as interações presentes em cada bloco no arquivo .cst, todas as coordenadas são colocadas em uma tabela nomeada de *bin hash*, responsável por reunir todas as possíveis coordenadas capazes de gerar uma *theozima*. Se verificar que nos *bin hashes* houver no mínimo um *hit* para cada um dos blocos no arquivo .cst então ocorre o *match*, desse modo, o ligante e resíduos catalíticos são transplantadas com sucesso na estrutura proteica (Richter et al., 2011).

O algoritmo de combinação primária e secundária têm seus ônus e bônus. A secundária tem por vantagem o fato de ser derivado do *hit* gerado anteriormente, sendo assim, não há necessidade de especificar os seis parâmetros no arquivo .cst. Nos problemas em que os pontos importantes são a distância e os 2 ângulos de ligação, enquanto os 3 diedros não sejam importantes, o algoritmo secundário pode esquecer os diedros, reduzindo o tempo de execução. Esse problema na correspondência primária necessitaria em varrer esses diedros em toda a faixa dos 360 graus. Em segundo lugar, a correspondência secundária também pode ser usada para identificar interações da *theozyme* entre duas cadeias laterais, uma vez que cada posição do ligante, bem como os rotâmeros, dos quais ela se originou, são armazenados no mesmo *hit*. Por fim, outra grande vantagem da correspondência secundária está na sua capacidade de avaliar a geometria entre rotâmeros projetados para interações na *theozyme* e rotâmeros anteriormente gerado no *hit* (Richter et al., 2011).

O benefício da correspondência primária em detrimento a secundária é que a sua execução significativamente mais rápida, caso haja uma grande variedade de colocações geradas pelas interações dos resíduos-ligante na *theozyme*. Pois, a eficiência de colocação das interações referente à *theozyme* utilizando a correspondência secundária diminui quanto mais colocações anteriores aos ligantes existirem, pois, cada um deles terá de ser testado novamente, como citado anteriormente. O algoritmo de correspondência primário, porém, gera as colocações de ligante a partir das coordenadas dos rotâmeros da cadeia lateral e os registros dentro do arquivo de restrição geométrica no arquivo .cst, portanto, são imparciais as colocações do ligante anteriormente gerado. Outra desvantagem da correspondência secundária se dá de-

vido a não conformação geométrica das interações do ligante-resíduo de maneira muito apropriada, forçando o usuário a tolerar grandes valores para uma interação específica ou amostrar densamente uma certa interação na *theozyme* (Richter et al., 2011).

Etapa do *design*

Depois que se coloca o ligante na estrutura proteica começa a etapa de *design*. Essa etapa no Rosetta consiste em quatro passos:

- A Indicação de quais resíduos entram no *design* e quais precisam ser minimizados;
- B Otimização da interação catalítica;
- C Ciclos de desenho e minimização com limitação geométrica presente no arquivo .cst;
- D Minimização da sequência projetada sem restrições.

No projeto da estrutura de entrada é possível a determinação automática dos resíduos a serem redesenhados. O Rosetta divide os resíduos em 5 grupos com base na sua distância, em Å, do ligante:

1. Resíduos cujo C_α estão dentro de uma distância definida em cut1 de qualquer átomo pesado pertencente ao ligante são selecionados para *design*;
2. Para os C_α dos resíduos que estão dentro de uma distância cut2, de qualquer átomo pesado do ligante e o C_β mais próximo a esse átomo do que C_α são selecionados para *design*, é necessário que cut2 seja maior que cut1;
3. Os resíduos que têm seu C_α dentro de uma distância de corte cut3, de qualquer átomo pesado do ligante são selecionados para minimização. É necessário que cut2 seja menor que cut3;
4. Já os resíduos que possuem o C_α dentro de uma distância, cut4, de qualquer átomo pesado do ligante e o C_β mais próximo desse átomo do ligante serão ajustados para minimização. A quantidade cut4 precisa ser maior que cut3;
5. Por fim, para os resíduos que não estão em nenhum dos quatro grupos citados anteriormente, estes são mantidos fixos.

Para a minimização do sítio, apenas os ângulos ϕ e ψ dos resíduos podem se mover dentro da área de projeto citada acima, movendo seus átomos de C_α apenas em 0,5 Å a partir de sua posição inicial, evitando assim, com que o *backbone* da estrutura proteica se mova permitindo otimizar o seu enovelamento, melhorando seu *score*. Uma alternativa para a otimização baseada em gradiente é a realização da amostragem de Monte Carlo (Richter et al., 2011).

Nas posições para redesenho, se aplica o algoritmo de Monte Carlo do Rosetta (Moretti et al., 2016), sendo usado para alterar a sequência de aminoácidos não

catalíticos, na região do projeto, buscando reduzir a energia do estado de transição. As restrições geométricas são mantidas ao longo deste estágio e uma vez criada a nova sequência, o sítio ativo resultante é minimizado. Essas 2 etapas são iterativas e se repetem por volta de 2 a 4 vezes.

No final do processo iterativo, de *design* e minimização dos resíduos não catalíticos, ocorre uma última minimização sem aplicação das restrições presentes no arquivo .cst. Com isso, até mesmo os resíduos catalíticos são minimizados, etapa essa necessária para verificar se o novo sítio ativo desenvolvido é capaz de manter os resíduos catalíticos em sua conformação definida no arquivo de restrição (*Richter et al.*, 2011).

Avaliação e classificação das sequências projetadas

Em um projeto enzimático típico, um grande número de até milhares de estruturas são normalmente projetadas. Por isso, se faz necessário para cada modelo criado pelo protocolo de *design* Rosetta uma pontuação que reflita as sequências com maior propensão catalítica.

É possível simplesmente escolher o modelo com a classificação geral, porém, essa pontuação não irá refletir todas as características necessárias para o bom funcionamento enzimático. Visto isso, se faz necessário ponderar outras propriedades e sua pontuação, sendo as principais o número de ligações de hidrogênio, número de átomos polares enterrados insatisfatoriamente, contatos não locais como entre resíduos que estão distantes na sequência, pontuação geral da enzima utilizando uma função ponderada com diversos fatores e a estimativa da energia de ligação.

Por fim, para cada modelo gerado no Rosetta, tanto suas pontuações quanto sua localização são depositadas, a cada linha, em um documento de saída. Na sintaxe geral, os rótulos das colunas terminando em -pm” são determinados usando métrica de estrutura - pose. Já para os resíduos catalíticos, as suas pontuações são representados pela sigla SRN, onde N se refere aos resíduos, e a estimativa da energia de ligação é reconhecido pelo termo SR(N+1) (*Richter et al.*, 2011).

Funções Score

A proteína tem uma variedade de formas tridimensionais para suportar e acomodar uma ampla gama de mecanismos, visando catalisar inúmeras reações. Devido ao potencial referente a sequência de aminoácidos, sua estrutura e o delicado equilíbrio entre as contribuições de entalpia e entropia das interações não covalentes, se faz necessário o desenvolvimento de funções de pontuação/energia, mesmo de forma aproximada, para contabilizar as interações, afim de que auxiliem na modelagem molecular. A função de energia se apoia na hipótese de Anfinsen, em que a conformação nativa - de baixa energia - é termodinamicamente estável. Esse estado se encontra no fundo do poço de energia, no qual a energia de Gibbs é favorável (*Godoi Contessoto et al.*, 2018).

Para distinguir entre estruturas enoveladas, desenoveladas e nativas, os programas de modelagem molecular requerem relações matemáticas. Essas funções são uma mistura linear de termos que calculam a energia como uma função de diferentes graus de liberdade. Os primeiros modelos integram a capacidade do potencial de Lennard-Jones, que representa as interações de van der Waals, sendo esta parametrizada através de um conjunto de moléculas pequenas. Hoje em dia, há uma vasta família de funções de energia, entre elas estão o AMBER (*Cornell et al.*, 1995), OPLS (*Jorgensen and Tirado-Rives*, 1988), e CHARMM (*Brooks et al.*, 1983).

Com maior poder computacional permitindo melhorias nos parâmetros moleculares usando cálculo quântico *ab initio*, melhorias nos métodos usados para determinar propriedades referente à estrutura proteica possibilitaram funções de energia com parcelas e parametrizações mais eficientes. Para regular a modelagem de alta performance, Kuhlman, Baker e colaboradores apresentaram a função de energia Rosetta, que enfatizava o empacotamento de átomos, ligação hidrogênio, solvatação e ângulos de torção de macromoléculas enoveladas (*Alford et al.*, 2017). A função inclui uma versão de solvatação implícita de pares complementares e termo eletrostático derivado estatisticamente, potencial de Lennard-Jones e um termo para preferências de rotâmero, conformação da cadeia lateral do aminoácido, estruturado no *backbone* (*Alford et al.*, 2017). As principais equações são descritas a seguir:

$$\Delta E_{total} = \sum_i w_i E_i(\Theta_i, aa_i) \quad (3.1)$$

A equação 3.1 descreve a ΔE_{total} , é calculado como uma combinação linear de termos de energia E_i que são calculados como funções dos graus de liberdade geométricos (Θ_i) e sua identidades químicas (aa_i) além do dimensionamento por um peso em cada termo (w_i).

$$E_{vdw} = \epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{d_{i,j}} \right)^{12} - 2 \left(\frac{\sigma_{i,j}}{d_{i,j}} \right)^6 \right] \quad (3.2)$$

Já a equação 3.2 descreve a forma calculada para a interações de van der Waals no Rosetta, esse potencial de Lennard-Jones é calculado com a energia de interação dos átomos i e j em diferentes resíduos dada a soma de seus raios atômicos, $\sigma_{i,j}$, a distância de pares dos átomos, $d_{i,j}$, além da sua distância média referente ao mínimo no poço de energia potencial da interação, $\epsilon_{i,j}$.

$$E_{hbond} = \sum_{H,A} w_H \cdot w_A \cdot f \left(E_{hbond}^{HA} (d_{HA}) + E_{hbond}^{AHD} (\Theta_{AHD}) + E_{hbond}^{BAH} (\Theta_{BAH}) + E_{hbond}^{B_2BAH} (\rho, \Phi_{B_2BAH}, \Theta_{BAH}) \right) \quad (3.3)$$

$$f(x) = \begin{cases} x & x < -0.1 \\ -0.0025 + \frac{x}{2} - 2.5x^2 - 0.1 & -0.1 \leq x < 0.1 \\ 0 & 0.1 \leq x \end{cases} \quad (3.4)$$

A equação 3.3 refere-se a energia da ligação de hidrogênio sendo descrita pela multiplicação de dois pesos específicos a entidade atômica, w_H e w_A , que explicam a força variável das ligações de hidrogênio. O termo $E_{hbond}^{B_2BAH}$ depende da hibridização do orbital acceptor de prótons, ρ , e a função $f(x)$, equação 3.4 e usada na equação 3.3, é usada para suavização, evitando a descontinuidade da derivada, garantindo as ligações de hidrogênio nos casos mais extremos. A distância entre o receptor e o doador de prótons é definida pelo termo d_{HA} , seus ângulos pelos Θ_{AHD} e Θ_{BAH} além de seu ângulo de torção Φ_{B_2BAH} .

A combinação da mecânica molecular e potenciais de torção derivados de dados estatísticos permitiu uma boa previsão dos sistemas. Em 2004, avanços consideráveis nas ligações de hidrogênio, modelos de solvatação, análises conformacionais do *backbone* e rotâmeros foram realizadas (Alford *et al.*, 2017). No processo, esses desenvolvimentos permitiram o Rosetta lidar com novos desafios na modelagem biomolecular como o refinamento estrutural obtido de cristalografia de raios-X. Esses desenvolvimentos permitiram o *design* de imunogênicos, tipos de peptídeos de biomineralização e enzimas que realizam novas funções. Foi adicionado, no *Rosetta Energy Function 2015* (que possui o codinome REF15), funções de energia adicionais para estimar a energia em kcal/mol ao invés de funções com unidades arbitrárias. Desse modo, a energia calculada e usada no Rosetta, se mostra bem robusta, em relação a sua versão anterior para modelar e avaliar estruturas geradas pelo mesmo (Alford *et al.*, 2017).

3.3 Método evolutivo de enzimas

Métodos de evolução enzimática utilizam a teoria de Darwin, pois, aplicam a variabilidade mutagênica e seleção adaptativa, buscando aprimorar certas características. Essas técnicas usam como base, aplicando certas alterações, o PCR, *polymerase chain reaction*. Essa técnica foi desenvolvido na década de 1980 por Kary Mullis, ganhador do prêmio Nobel de Química em 1993 junto com Michael Smith, que permite a multiplicação de uma fita de DNA (Heckmann and Paradisi, 2020).

Como mostrado na figura 3.4, os princípios da PCR são a desnaturação do DNA em alta temperatura, em seguida, os *primers* fornecidos na mistura reacional são acoplados na fita e o DNA é copiado. Desse modo, ciclos repetidos amplificam exponencialmente a sequência alvo. Já os *primers* são uma sequência curta de nucleotídeos que fornecem um ponto de partida na síntese do DNA, definindo a região que será copiada.

No início da década de 1990, Frances Arnold, utilizou essa técnica incorporando mutações aleatórias nessas fitas, gerando uma ampla gama de mutantes. Desenvol-

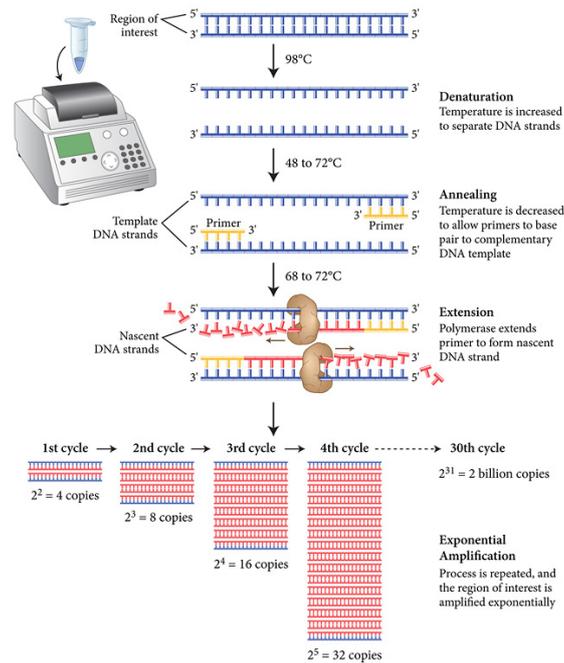


Figura 3.4: Esquema do PCR. Fonte: (Biolabs, 2022).

vendo, assim, a evolução dirigida, recebendo em 2018 o prêmio Nobel de Química (Heckmann and Paradisi, 2020). Ela modulou a polimerase para que ocasionalmente incorporasse nucleotídeos incorretos, recebendo o nome de epPCR, que promove a mutação aleatória na fita. Com essa técnica, foi possível desenvolver uma subtilisina E, após introduzir 10 mutações capaz de ser tolerante às altas concentrações de dimetilformamida (DMF) (Chen and Arnold, 1993).

Em anos posteriores foi adicionado o DNA *shuffling* ao método evolutivo, esse, implementado por Pim Stemmer em 1994 (Stemmer, 1994), permite ampliar a utilização de genes, imitando a recombinação de DNA que ocorre em organismos como forma de aumentar a diversidade genética, permitindo a mistura de diversas proteínas, e assim, gerando diversas sequências distantes das naturais (Heckmann and Paradisi, 2020).

Posteriormente, devido ao fato de que o maior problema da evolução dirigida ser a etapa de triagem, foram desenvolvidos softwares para melhorar a capacidade de identificar as estruturas com maior potencial catalítica. Para isso, foi necessária uma maior compreensão referente a estrutura proteica e graças aos avanços experimentais e teóricos-computacionais essa lacuna foi preenchida, utilizando técnicas como cristalografia de raio-X, *docking*, dinâmica molecular e modelagem proteica, é possível analisar a contribuição de certas mutações ao sistema. Destaca-se entre os *softwares*, o protocolo de *design* enzimática do Rosetta e sua implementação de-novo *design* de proteínas (Heckmann and Paradisi, 2020).

Com esse conhecimento desenvolvido, processos como o SCHEMA (Voigt *et al.*, 2002), que permite identificar e avaliar as mutações segundo a sua capacidade em afetar o processo de enovelamento da enzima, e CASTing (teste de sítio ativo combinatório), em que vários resíduos que revestem o sítio ativo são alterados ao mesmo tempo, permitindo que surjam efeitos sinérgicos entre as mutações, conseguem reduzir o tamanho da biblioteca de mutantes a serem testadas (Heckmann and Paradisi, 2020). A segunda técnica, citada acima, teve sucesso na criação de amino desidrogenases enantiosseletiva, a partir de uma leucina desidrogenase (Abrahamson *et al.*, 2012), além de permitir mudanças de enantiosseletividade (Reetz, 2013) e aumento do escopo dos ligantes à enzima (Reetz *et al.*, 2005).

Na evolução dirigida, recursos como dados estatísticos e teóricos referente às mutações veem sendo aplicados nas técnicas de *machine learning*. Dentre as técnicas destacam-se as relações de atividade de sequência de proteína (ProSAR), que veem fornecendo bons resultados ao serem aplicados (Heckmann and Paradisi, 2020). O ProSAR consiste em classificar se uma certa mutação é benéfica, neutra ou prejudicial, desse modo, informando qual desses devem ser incorporadas em bibliotecas posteriores, para sempre ser testadas, ao invés de eleger uma melhor variante e a partir dela gerar uma nova biblioteca, ignorando as mutações anteriores. Essa estratégia foi aplicada com sucesso na criação de uma halodrina dehalogenase (HHDH), ao longo de 18 rodadas de evolução implementando 35 mutações. Essa enzima sintetiza o (R)-4-ciano-3-hidroxi-butirato, sendo esse composto um intermediário para síntese de atorvastatina, medicamento capaz de reduzir o colesterol. Nesse trabalho, os autores observam que metade das mutações presentes na estrutura final não estavam presentes no início nas melhores mutações estruturais selecionadas, desse modo, teria sido perdida se não fosse aplicada a técnica ProSar (Fox *et al.*, 2007).

Através da combinação das técnicas, citadas anteriormente, outros casos de sucessos surgiram. O pesquisador Savile utilizou essas ideias para gerar uma enzima transaminase (Savile *et al.*, 2010). Sendo utilizado a partir de uma enzima que apresentava baixa atividade para um ligante e que era análoga ao composto sitagliptina. Com base nessa baixa catálise para um ligante parecido ao que se desejava projetar, foi possível desviar sua rota catalítica, permitindo catalisar esse novo ligante, ideia essa ilustrada na figura 3.5. Foram utilizadas 11 rodadas de evolução incorporando 27 mutações, utilizando as técnicas de CASTing, proSAR e epPCR, o que gerou 36.480 variantes. Esse processo conseguiu desenvolver uma estratégia mais eficiente que o processo alternativo de aminação reductiva catalisada por ródio (Heckmann and Paradisi, 2020).

Outro caso de sucesso foi a produção de uma enzima imina redutase (IRED) para sintetizar o inibidor de LSD1-GSK2879552, um composto que inibe o crescimento celular em células tumorais. Devido à escassez de informações estruturais, foi selecionada uma estrutura que apresentava potencial para evolução dirigida e posteriormente foi utilizado a técnica de CASTing em quase toda a estrutura, sendo alterados os 256 das 296 posições residuais. Em seguida foram analisadas as mutações benéficas à luz da análise estatística e o *software* CodeEvolver para, assim,

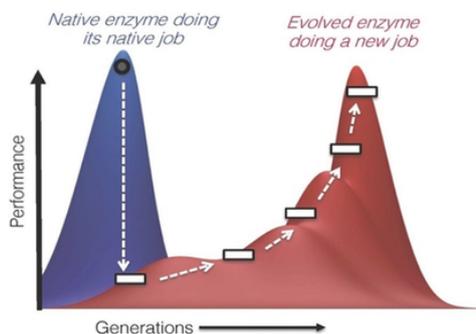


Figura 3.5: Ideia que envolve a mudança de rota catalítica na evolução dirigida (Heckmann and Paradisi, 2020)

gerar uma biblioteca de mutantes. Por fim, após 3 rodadas de evolução, foi desenvolvida uma enzima com 13 mutações que conseguiu sintetizar 1,4 kg de GSK2879552 (Schober *et al.*, 2019).

Os avanços dessa técnica permitiu o grupo de Frances Arnold desenvolver, utilizando a evolução dirigida, uma reação em uma proteína que não possui nenhuma atividade catalítica (Heckmann and Paradisi, 2020). Eles utilizaram a citocromo *c* (*cytc*) como base para catalisar uma ligação carbono-silício, mais uma reação inexistente na natureza. Após a triagem, identificaram que um *cytc* de *Rhodothermus marinus* apresentava baixa atividade e utilizando o CASTing identificaram 3 mutações em resíduos, gerando uma melhoria de 15 vezes em relação aos catalisadores sintéticos de última geração, apresentando alta enantiosseletividade (Kan *et al.*, 2016).

Por fim, como foi possível observar, o sucesso dos estudos citados estava atrelado além da geração de bibliotecas eficientes a escolha de estruturas. Desse modo algoritmos de busca como o BLAST têm sido desenvolvidos há muito tempo, onde esse consegue a partir de modelos de comparação estrutural definir a provável função de uma sequência de DNA e conseqüentemente estruturas candidatas a evolução. Com o avanços do conhecimento e dos bancos de dados em relação às estruturas proteicas e enzimáticas, a evolução dirigida vem ganhando espaço como uma ferramenta poderosa para auxiliar a descoberta de novos catalisadores (Heckmann and Paradisi, 2020).

3.4 Métodos de Química Quântica

Um dos principais objetivos da química quântica é buscar resolver a equação de Schrödinger para verificar com exatidão as propriedades dos sistemas atômicos e moleculares. Devido à complexidade das equações, surgiram os métodos de química quântica que se dividem em *ab initio*, DFT e semiempírico.

Os métodos *ab initio* buscam resolver a equação de Schrödinger de forma mais

próxima da solução exata, evitando ao máximo as aproximações. Esse fato implica a esses métodos a necessidade de um maior poder computacional devido à complexidade das equações em sistemas moleculares com muitos átomos, conseqüentemente, levando muito mais tempo para encontrar a solução. Dentre os mais aplicados hoje estão: Teoria de perturbação de muitos corpos de Møller-Plesset (MPn) e Campo multi-configuracional autoconsistente (MCSCF, incluindo CASSCF e RASSCF) e *Coupled Cluster* (Pessoa et al., 2018)

Os métodos DFT surgiram como uma alternativa para os métodos *ab initio* que substitui a função de onda de muitos elétrons pela função de densidade eletrônica. Essa alteração, juntamente com o aprimoramento dos funcionais de troca e correlação, fornece resultados exatos em relação aos dados experimentais e com custo computacional menor quando comparado com os métodos *ab initio*. Dois funcionais que se destacam são: B3LYP e M06-2X (Pessoa et al., 2018)

Os métodos semiempíricos se destacam por permitir o tratamento de grandes sistemas moleculares a baixo custo computacional. Para isso certas integrais em seus elementos de cálculo são zeradas ou substituída por parâmetros experimentais/ajustáveis, como, por exemplo, a energia de ionização, entalpias de formação atômicas, entre outras. Além disso, é realizado o tratamento explícito, que consiste em resolvendo as equações, apenas dos elétrons de valência utilizando a base mínima, enquanto os elétrons internos são abordados de maneira implícita (Laschuk, 2005). Com tais aproximações, esses métodos necessitam de parametrização e validação estatística como uma etapa posterior ao seu desenvolvimento. Os métodos atuais possuem uma boa relação custo computacional e poder preditivo, destacando-se hoje o método PM7 no tratamento de biomoléculas (Pessoa et al., 2018).

Como visto, o custo computacional é fundamental para a viabilidade de um método a depender do tamanho do sistema. Assim, buscando tratar grandes estruturas de maneira ainda mais rápida foi implementado no *software* MOPAC (Stewart, 1990), que comporta todos os métodos semi-empíricos até aqui citados, o algoritmo de escalonamento MOZYME (Stewart, 1996). A ideia por trás dessa estratégia está em trabalhar com orbitais moleculares localizados baseados em estruturas de Lewis. Durante a convergência da energia E_SCF e da densidade eletrônica, esses orbitais moleculares localizados não são permitidos a perderem a condição de localização, mas mantendo a ortogonalidade entre eles, como é requerido no método de Hartree-Fock-Roothaan. Esse método ainda usa um formato de matrizes esparsas próprio, o que confere um menor custo em alocar dados na memória, bem como de operações de ponto flutuante. Isso só é possível porque interações eletrônicas de longo alcance, na maior parte das vezes, tende a um valor muito baixo e podem ser ignoradas, dentro do contexto de uso de orbitais moleculares localizados e da aplicação de raios de corte no início do cálculo. Diante desses aspectos, ao analisar a relação do tempo gasto para encontrar a solução do sistema, é possível reduzir o grau de complexidade do algoritmo Hartree-Fock-Roothaan-SCF de N^3 para N^2 , onde N é o número de átomos da molécula. Assim, a complexidade do algoritmo SCF passa de uma relação cúbica para quadrática, entre o tempo gasto e o número de átomos do

sistema molecular, permitindo encontrar essa solução em um menor tempo.

3.5 Descritores de reatividade

PRIMorDIA

O PRIMorDIA (*PRIMorDIA Macromolecular Reactivity Descriptors Access*), é um *software* desenvolvido pelo Grupo de Química Quântica Aplicada a Sistemas Biológicos da Universidade Federal da Paraíba - UFPB (*Grillo et al.*, 2020d). Sendo escrito em C++ com recursos de programação paralela, possui o objetivo de calcular de forma rápida e eficiente descritores de reatividade derivado do formalismo referente ao *Conceptual Density Functional Theory*. Seu funcionamento se baseia no uso de *outputs* de softwares de mecânica quântica, como Gaussian, GAMESS, ORCA e MOPAC, e calculando os descritores com base nos orbitais presentes na função de onda retirada desses programas gera seu próprio *outputs*. Esse arquivo de saída se divide em 2 tipos, sendo o primeiro arquivo formatado para trabalhar com dados numéricos, enquanto o segundo para ser aberto em programas gráficos, onde seus dados podem ser visualizados e analisados de forma ágil. É possível extrair desses *outputs* dados volumétricos, que auxiliam em análises semelhantemente ao volume de van Der Waals, ou pontuais presentes nos átomos ou resíduos de aminoácidos.

Até esse momento, o PRIMorDIA foi utilizado em 4 estudos, sendo os artigos indicados nas referências em seguida: (*Grillo et al.*, 2020c), (*Grillo et al.*, 2020b), (*Grillo et al.*, 2020a) e (*Grillo et al.*, 2020d).

O primeiro realizou cálculos *single-point* semiempíricos para obter os descritores utilizando a aproximação de orbitais congelados com o formalismo dos orbitais moleculares da banda de fronteira. Através do cálculo das sucessibilidades à ataques nucleofílico e eletrofílico foi possível demonstrar o grande potencial em prever os ligantes com capacidade inibitória ativa referente a protease do HIV-1 (*Grillo et al.*, 2020c).

O segundo estudo foi produzido pela mesma equipe e buscou elucidar a catálise enzimática de três enzimas utilizando os descritores moleculares de reatividade. Para isso foi utilizado os métodos quânticos semiempíricos para descrever três caminhos de reação, calculando os descritores locais de dureza e sucessibilidades à ataques nucleofílico e eletrofílico. Para a dureza do sitio ativo não houve mudança substancial, principalmente para os resíduos que estabilizam a estrutura de reação. Já a sucessibilidades, por outro lado, apresentou mudanças abruptas durante a coordenada de reação, revelando sua influência durante a catálise. Por fim, os autores do estudo concluíram que o PRIMorDIA pode fornecer muitas informações e melhorar as análises proteicas (*Grillo et al.*, 2020b).

Já no terceiro estudo, os autores estudaram a enzima chiquimato 5-desidrogenase, que causa tuberculose (*Grillo et al.*, 2020a). Nesse estudo, os autores aplicaram dinâmica molecular, scan relaxado, descritores moleculares e simulação

de energia livre. Dentre as análises obtidas foi observado uma possível relevância para a Thr65 (resíduo do sítio ativo) ainda não relatada na literatura (*Grillo et al.*, 2020a). Por último, o PRIMoRDIA foi utilizado para monitorar as características presentes no sítio ativo da protease do SARS-CoV-2 que evoluíram por mutação a partir do SARS-CoV-1, e, com isso, identificar um potencial inibidor, fornecendo, desse modo, *insights* para auxiliar outros pesquisadores no desenvolvimento de medicamentos (*Grillo et al.*, 2020d).

Os *outputs* gráficos gerados pelo PRIMoRDIA são formatados para que possam ser abertos e manipulados no Pymol, um programa de visualização de estruturas moleculares, facilitando a análise visual desses descritores. Por fim, o formalismo que é utilizado para obter esses descritores será descrito em detalhes na seção seguinte.

Formalismo teórico do cálculo dos Descritores de Reatividade

Os avanços no desenvolvimento em hardware e algoritmos eficientes, impulsionaram os cálculos de mecânica quântica em sistemas cada vez maiores (*Mati Karelson and Katritzky*, 1996). Os novos métodos semiempíricos fornecem, para grandes sistemas moleculares, boas resoluções em um tempo de processamento consideravelmente curto, o que torna os cálculos quânticos uma fonte atraente para o cálculo de novos descritores moleculares, que, em princípio, podem ser usados para descrever todas as propriedades moleculares. Esse aparato matemático fornece uma descrição mais detalhada e precisa dos efeitos eletrônicos, além disso, os descritores eletrônicos gerados são capazes de ser usados sozinhos ou adicionados, em combinação, a outros descritores, podendo também ser separados com base nos átomos ou grupos, descrevendo, assim, muitas regiões moleculares diferentes na mesma estrutura. Por fim, dados experimentais corroboram com a eficiência desses descritores, visto a sua relação com os dados teóricos a respeito das reatividades dos compostos orgânicos, coeficientes de partição octanol/água entre outras (*Mati Karelson and Katritzky*, 1996).

Posto isso, buscando uma forma sistemática de obtenção desses descritores, após a obtenção da função de onda, foi desenvolvido o *software* PRIMoRDIA (*Grillo et al.*, 2020d) que através da teoria da CDFT (*Conceptual DFT*) fornece conceitos químicos abrangentes. Os conceitos são derivados da equação diferencial fundamental do método DFT (*Geerlings et al.*, 2003), conforme mostrado na eq. 3.5, devido à manipulação da série de Taylor ou suas derivadas, a fim de fornecer dados numéricos relativos à estrutura eletrônica no que diz respeito à variação eletrônica.

$$dE = \mu dN + \int \rho(r) d\nu(r) dr \quad (3.5)$$

Descritores globais como potencial químico (μ), dureza (η) e suavidade (S) são derivados da energia em relação ao número de elétrons (*Parr et al.*, 1978). A primeira é obtida a partir da primeira derivada mostrada na equação 3.6, enquanto

a segunda e a terceira, essas definidas como o inverso da dureza, são obtidas a partir da segunda derivada da energia, expressa pela equação 3.7.

$$\mu = \left(\frac{\partial E}{\partial N} \right)_{\nu} \quad (3.6)$$

$$\eta = \frac{1}{S} = \left(\frac{\partial^2 E}{\partial^2 N} \right)_{\nu} \quad (3.7)$$

A derivada da densidade eletrônica em relação ao número de elétrons gera descritores de reatividade local, que fornecem dados numéricos para átomos distribuídos no espaço tridimensional. Sua primeira derivada é definida como função de Fukui (*Parr and Yang, 1984*), conforme expressa na equação 3.8.

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_{\nu} \quad (3.8)$$

Através da aproximação do orbital molecular fronteira, que permite através do cálculo quântico single point, obter dois descritores locais, sendo suscetibilidade de ataque eletrofílico (EAS), eq. 3.9, e suscetibilidade de ataque nucleofílico (NAS), eq. 3.10.

$$EAS(r) = |\alpha_{HOMO(r)}|^2 \quad (3.9)$$

$$NAS(r) = |\alpha_{LUMO(r)}|^2 \quad (3.10)$$

Para estruturas grandes, como proteínas, Fukushima et al. (*Fukushima et al., 2008*), demonstraram que os orbitais nas proximidades do *gap* HOMO-LUMO estão, na catálise enzimática, dentro de uma banda na região de 1-5 eV. Outra forma de se calcular é a utilização das funções fukui ponderadas, expressas nas equações 3.11 e 3.12, apresenta melhores resultados no cálculo do EAS e NAS. Tanto o cálculo de banda quanto a forma ponderada foram utilizadas nesse trabalho.

$$f_{EW}^- = \sum_{i=b}^{HOMO} e^{-|E_i - E_{HOMO}|} \cdot |\psi_i|^2 \quad (3.11)$$

$$f_{EW}^+ = \sum_{i=LUMO}^b e^{-|E_i - E_{LUMO}|} \cdot |\psi_i|^2 \quad (3.12)$$

Nas duas equações citadas acima, descrevem a soma dos orbitais com uma diferença de energia dentro da banda de energia definida em b, do HOMO ou LUMO.

Para auxiliar nos cálculos de NAS e EAS foi utilizada o formalismo matemático da netfilicidade, expressa na equação 3.13, que compila às duas informações em uma mesma equação.

$$\text{Netfilicidade} = NAS(r) - EAS(r) \quad (3.13)$$

Capítulo 4

Metodologia

Nesse capítulo, apresentamos detalhes sobre a metodologia que usamos para realizar o pós-processamento ou *rescoring* da saída do programa Rosetta e, desse modo, buscando apresentar uma alternativa na tentativa de solucionar o problema citado no estudo *Computational Design of Synthetic Enzymes* (Vaissier Welborn and Head-Gordon, 2018). Para isso, será descrito o protocolo de *design* do Rosetta, como foi realizado os cálculos de energia de ligação enima-substrato ($\Delta H_{binding}$) e descritores de reatividade, a fim de obter a estrutura redesenhada que melhor apresenta potencial catalítico. Um resumo do protocolo é mostrado na Figura 4.1.

A fim de melhorar a forma de classificação do Rosetta, o $\Delta H_{binding}$ foi usado em uma primeira estratégia e para análise visual foram utilizados os descritores quânticos como uma segunda estratégia. Essas duas quantidades são obtidas a partir da função de onda das estruturas biomoleculares a partir de cálculos quânticos semiempíricos. Dentre os muitos Hamiltonianos existentes, os semiempíricos apresentam hoje a capacidade de lidar com grandes estruturas com baixo custo computacional. Portanto, as funções de onda e entalpias foram obtidas usando um cálculo *single point* com o Hamiltoniano PM7 (Stewart, 2013), considerando efeitos de solvatação através do modelo implícito COSMO, e utilizando o algoritmo MOYZME, que reduz o tempo de cálculo pelo uso de orbitais localizados e pseudo diagonalização (Stewart, 1996). Todos os cálculos foram realizados na versão 2016 do programa MOPAC (Stewart, 1990).

Uma vez obtida a entalpia de formação do complexo (ligante-enzima), foram realizados cálculos do ligante e da enzima isolados para cada estrutura desenhada. Feito isso, é possível, por meio da Equação 4.1, obter a entalpia de ligação, $\Delta H_{binding}$, que servirá para a reclassificação.

$$\Delta H_{binding} = \Delta H_f^{complexo} - (\Delta H_f^{enzima} + \Delta H_f^{ligante}) \quad (4.1)$$

Na equação 4.1, o $\Delta H_f^{complexo}$, ΔH_f^{enzima} e $\Delta H_f^{ligante}$ consistem nas entalpias de formação da enzima com o ligante, enzima e ligante isolados, respectivamente.

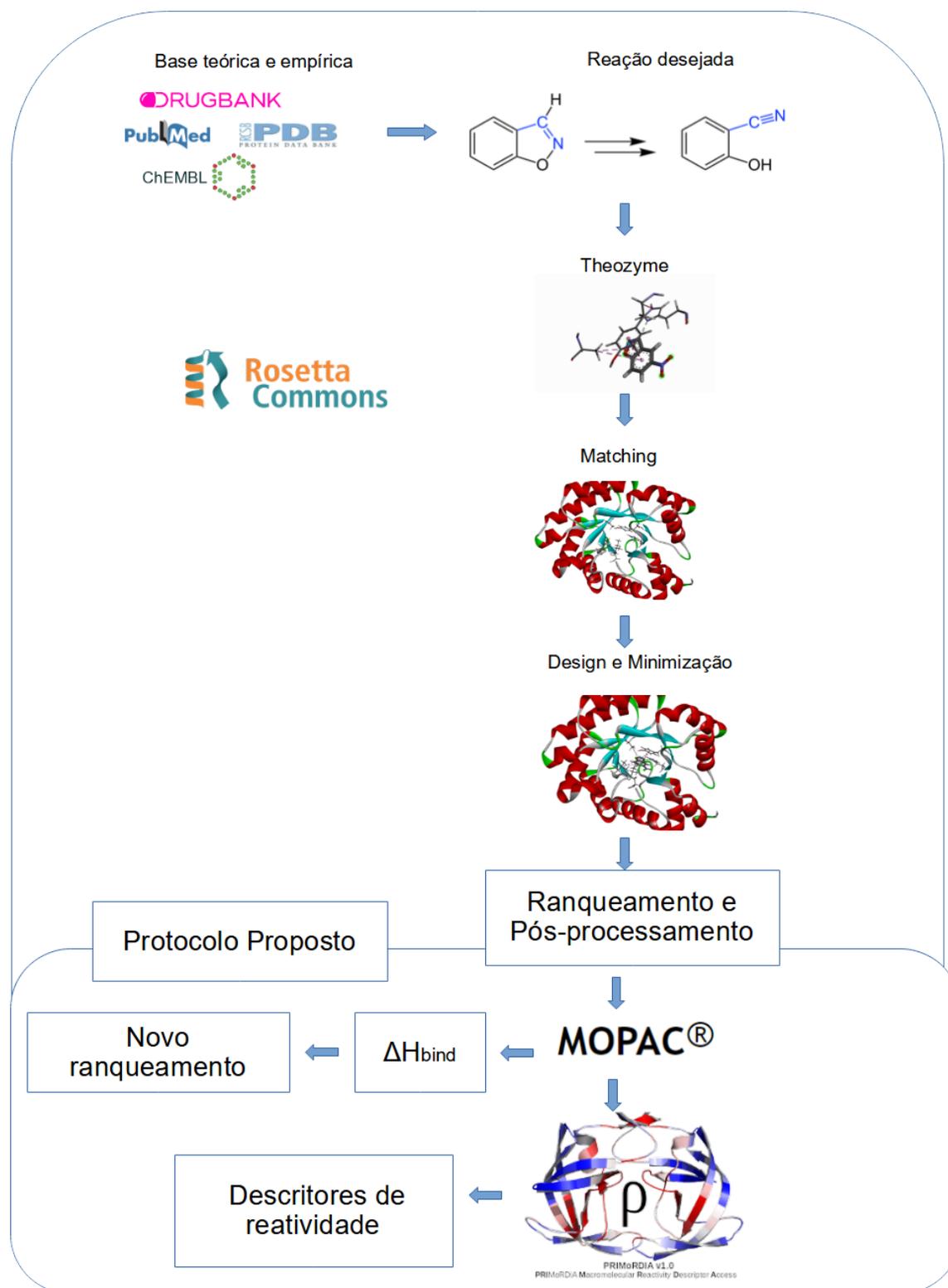


Figura 4.1: Fluxograma do protocolo usado nesse estudo.

Obtida a densidade eletrônica do complexo enzima-ligante, o software PRIMorDiA foi utilizado e configurado para calcular os descritores com a banda HOMO-LUMO dentro de 5eV usando a forma ponderada EW. Todos os descritores foram obtidos utilizando em seu *input* as seguintes especificações resumidas nas figuras 4.2 e 4.3

```
#RT normal
#PR eband 5 pymols dos
3 KE07_ok.aux LCP 0 10 KE07_ok.pdb mopac 0 0 0 0 BD
```

Figura 4.2: As linhas RT e o PR são responsáveis por especificar o tipo de cálculo e os *outputs* gerados pelo PRIMorDiA configurado para leitura no pymol. A terceira linha especifica o cálculo de banda e o número máximo de orbitais moleculares a ser utilizados pelo método BD, onde foi utilizado 10.

```
#RT normal
#PR eband 5 pymols dos
3 KE07_ok.aux LCP 0 0 KE07_ok.pdb mopac 0 0 0 0 EW
```

Figura 4.3: O RT e o PR são as linhas responsáveis por especificar o tipo de cálculo e os *outputs* gerados com configuração de pymol. A terceira coluna especifica o método de cálculo de descritores para macromoléculas ponderado por energia e a banda foi definida como eband 5, indicando que foi considerados os orbitais moleculares próximos ao HOMO-LUMO dentro de 5eV.

Visando validar o protocolo proposto, buscamos estudos que utilizaram o protocolo de *design* Rosetta e realizaram a parte experimental, obtendo a constante catalítica da enzima. Os artigos que atenderam a esses critérios foram os trabalhos de Daniela Röthlisberger et. al (*Röthlisberger et al.*, 2008) e Justin B. Siegel et. al (*Siegel et al.*, 2010).

Ambos os artigos forneceram o arquivo fasta, o que inclui a estrutura primária das enzimas desenhadas, os dados do arquivo .cst, as informações referente ao ligante no estado de transição, bem como algumas estruturas no formato PDB das estruturas projetadas. Com esses arquivos foi possível reproduzir os dados desses artigos e usá-los como base para validar o nosso trabalho.

No artigo da Daniela Röthlisberger et. al (*Röthlisberger et al.*, 2008), foi necessário adicionar a estrutura catalítica do KE59, pois, o que disponibilizaram eram a versão com mutações que não possuía atividade catalítica. Para isso foi necessário utilizar a estrutura com código pdb de 1a53 e alterar os resíduos para que correspondesse ao arquivo FASTA, sequencias de código de uma letra referente a sequencia primária da enzima, disponibilizado no material suplementar do artigo. Ao alterar os resíduos a estrutura foi submetida a etapa de *match* e *design* do Rosetta de modo que se permanece a estrutura fasta, alterando apenas a conformação seguindo os

critérios do *software*, desse modo foi possível obter a estrutura catalítica informada no artigo. As informações catalíticas estão expressas na tabela 4.1.

No artigo de Justin B. Siegel et. al (*Siegel et al.*, 2010), foram realizadas 2 estruturas, uma que foi obtida da saída do protocolo Rosetta e que apresentou baixa atividade catalítica, $0.10 \pm 0.02 \text{ h}^{-1}$, e a outra obtida ao aplicar nela a intuição química para alterar certos resíduos alcançando $2.13 \pm 0.24 \text{ h}^{-1}$. Desse modo, a primeira estrutura foi gerada passando pelo protocolo de *design* Rosetta, correspondendo aquela do mesmo código fasta do artigo, obtendo-se assim a estrutura no formato pdb para análise dos descritores. Já a segunda foi obtida alterando os resíduos de aminoácidos seguindo a intuição química do autor. As informações catalíticas estão expressas na tabela 4.2.

Ao produzir as estruturas redesenhadas pelo protocolo de *design* Rosetta o *script* DesignSelect.pl, disponibilizado pelo próprio *software*, foi aplicado para filtrar o número de estruturas. Esse *script* utiliza uma série de valores, escolhidos pelo usuário para os parâmetros que julgar mais relevantes a partir das ligações de hidrogênio, interações hidrofóbicas, iônicas, orientação do ligante no sítio ativo entre outras, busca-se avaliar a estrutura redesenhada. Esse processo não busca achar a melhor estrutura, porém busca reduzir a quantidade de estruturas para análise visual por meio da intuição química, ou seja, tornando humanamente tratável. Desse modo, as estruturas que estiverem dentro do filtro aplicado serão selecionadas para análises posteriores.

Não foi possível reproduzir o estudo *Retro-Aldol* (*Jiang et al.*, 2008), devido à impossibilidade de replicar o arquivo de restrição. Porém, como disponibilizaram, em seu material suplementar, as estruturas em formato pdb, foi possível realizar os cálculos de descritores visando entender seus descritores nessas estruturas. As informações catalíticas estão expressas na tabela 4.3.

Tabela 4.1: Constantes de velocidade catalítica para a reação de eliminação *Kemp*.

catalisador	$k_{cat} \times \text{s}^{-1}$
KE07	1.80×10^{-2}
KE59-native	2.90×10^{-1}
KE70	1.60×10^{-1}

Tabela 4.2: Constantes de velocidade catalítica para a reação de *Diels-Alder*.

catalisador	$k_{cat} \times s^{-1}$
DA_20_00	2.70×10^{-5}
DA_20_10	5.91×10^{-4}

Tabela 4.3: Constantes de velocidade catalítica para a reação de *Retro-Aldol*.

catalisador	$k_{cat} \times s^{-1}$
RA22	5.16×10^{-5}
RA34	7.00×10^{-5}
RA46	1.00×10^{-5}
RA60	1.55×10^{-4}
RA61	1.15×10^{-4}

Capítulo 5

Resultados e Discussão

Nesse capítulo iremos apresentar os resultados obtidos ao calcular o $\Delta H_{binding}$ na seção 5.1 e discutiremos o quanto o *rescoring*, através dos cálculos quânticos, melhoram a colocação de cada estrutura em relação ao *score* do Rosetta. Na seção apresentaremos os resultados obtidos pelos descritores de reatividade retirados do PRIMoRDIA e discutiremos como eles podem auxiliar na escolha das estruturas com maior potencial catalítico.

5.1 Usando o $\Delta H_{binding}$ para *rescoring*

Após a reprodução dos artigos de Daniela Röthlisberger et. al (*Röthlisberger et al.*, 2008) e Justin B. Siegel et. al (*Siegel et al.*, 2010), um total de 154 *matches* para a reação de *kemp*, e 105 *matches* para a reação de *Diels-alder* foram encontrados, resultando em 1540 e 1050 enzimas redesenhadas, respectivamente. A estrutura projetada KE59, que no artigo de Daniela et. al apresentou melhor atividade catalítica após a análise experimental das muitas testadas, foi renomeada para Best_25_E231W110 e adicionado ao conjunto de estruturas projetadas. O *script* DesignSelect.pl, foi utilizado para filtrar estruturas com base em pontuações de Rosetta que sai no final do processo de *design*, expressa na figura 5.1, reduzindo a quantidade de estruturas para análise.

O resultado da filtragem está resumida na seção de anexa na tabela 1. Apenas 51 estruturas projetadas passaram pelo filtro usado, incluindo a estrutura Best_25_E231W110. Nessa tabela estão contidos os valores calculados para $\Delta H_{binding}$ e a reclassificação, além da pontuação Rosetta. A partir dela, foi possível observar que a estrutura Best_25_E231W110, deixou a última posição para a posição 34, subindo 17 posições, porém, ainda longe das 10 melhores posições.

No artigo de Justin B. Siegel, a estrutura projetada DA_20_00, que foi realizada a caracterização catalítica derivada diretamente do protocolo de *design* Rosetta, foi renomeado para Best_UM_2_Y121Q195 e adicionado ao conjunto de estruturas

File	Edit	Search	View	Document	Help	total_score	fa_rep	hbond_sc	all_cst	tot_pstat_pm	tot_nlpstat_pm	tot_nlpstat_tm	tot_burunsat_pm	tot_hbond_pm	tot_MLconts_pm	tot_nlsurfaceE_pm	tot_total_neg_charges	tot_total_pos_charges	tot_total_charge	tot_total_neg_charges	tot_total_pos_charges	tot_sec
-859.58	260.70	-45.17	1.64	0.65	0.66	0.66	52.00	251.00	161.00	5.52	35.00	36.00										
-873.12	268.24	-41.48	1.61	0.66	0.66	0.66	52.00	253.00	159.00	5.53	35.00	36.00										
-866.68	194.47	-37.66	1.10	0.63	0.63	0.63	54.00	248.00	161.00	6.07	34.00	35.00										
-864.74	195.70	-38.95	3.74	0.63	0.64	0.64	48.00	242.00	162.00	6.21	34.00	35.00										
-844.04	265.72	-38.05	3.13	0.63	0.64	0.64	52.00	242.00	162.00	12.86	34.00	36.00										
-859.36	185.83	-42.23	1.08	0.65	0.62	0.62	53.00	251.00	163.00	8.47	36.00	36.00										
-864.20	192.95	-37.68	3.26	0.64	0.64	0.64	50.00	245.00	152.00	5.89	34.00	35.00										
-862.01	264.86	-37.80	2.19	0.64	0.63	0.63	51.00	247.00	157.00	6.15	36.00	36.00										
-869.04	193.39	-39.21	1.85	0.63	0.63	0.63	57.00	248.00	168.00	9.37	35.00	37.00										
-861.77	191.06	-41.37	1.95	0.61	0.63	0.63	54.00	246.00	166.00	8.46	35.00	36.00										
-859.36	206.23	-43.50	3.82	0.66	0.64	0.64	63.00	246.00	160.00	2.67	36.00	36.00										
-854.39	206.77	-38.59	1.21	0.66	0.65	0.65	54.00	248.00	161.00	4.77	36.00	36.00										
-854.81	202.74	-38.50	1.03	0.66	0.66	0.66	53.00	246.00	161.00	4.73	36.00	36.00										
-865.56	192.80	-38.55	3.03	0.76	0.66	0.66	56.00	246.00	170.00	6.01	35.00	37.00										
-867.54	203.07	-39.55	0.82	0.67	0.67	0.67	56.00	249.00	160.00	8.77	35.00	37.00										
-866.25	205.09	-42.37	1.52	0.66	0.67	0.67	57.00	249.00	161.00	5.74	36.00	36.00										
-855.52	184.58	-43.68	3.24	0.66	0.63	0.63	44.00	252.00	160.00	6.05	35.00	39.00										
-873.92	180.84	-41.40	1.86	0.66	0.61	0.61	46.00	250.00	156.00	5.97	35.00	35.00										
-860.83	205.66	-37.78	3.33	0.68	0.66	0.66	51.00	245.00	159.00	5.81	35.00	35.00										
-852.38	268.72	-39.94	1.73	0.65	0.65	0.65	46.00	244.00	159.00	9.17	35.00	36.00										
-854.59	199.50	-39.06	1.13	0.65	0.65	0.65	57.00	245.00	159.00	7.91	35.00	36.00										
-846.74	203.61	-37.77	2.63	0.63	0.63	0.63	51.00	248.00	153.00	5.38	35.00	35.00										
-857.30	205.59	-44.83	0.57	0.68	0.65	0.65	46.00	252.00	152.00	5.44	36.00	37.00										
-871.18	204.54	-41.56	3.58	0.66	0.66	0.66	51.00	245.00	162.00	4.75	35.00	36.00										
-872.04	208.78	-42.36	0.67	0.67	0.67	0.67	47.00	251.00	153.00	6.56	35.00	37.00										
-848.58	203.02	-38.15	2.33	0.65	0.63	0.63	48.00	247.00	154.00	5.45	34.00	36.00										
-861.85	202.57	-36.64	2.16	0.68	0.67	0.67	53.00	242.00	158.00	5.88	34.00	36.00										
-841.29	208.60	-37.77	0.88	0.67	0.66	0.66	54.00	242.00	161.00	9.27	34.00	36.00										
-855.09	194.74	-41.69	0.97	0.66	0.64	0.64	46.00	249.00	152.00	4.63	35.00	36.00										
-862.42	196.62	-43.34	0.86	0.66	0.66	0.66	46.00	253.00	157.00	6.86	34.00	35.00										
-851.90	201.03	-39.09	1.49	0.63	0.61	0.61	55.00	248.00	156.00	5.81	34.00	36.00										
-862.22	202.63	-41.35	0.41	0.67	0.65	0.65	50.00	250.00	154.00	6.12	35.00	36.00										
-871.52	206.85	-42.73	1.26	0.65	0.65	0.65	49.00	250.00	159.00	7.05	34.00	38.00										
-877.47	201.35	-41.37	2.57	0.65	0.64	0.64	53.00	247.00	154.00	6.08	36.00	36.00										
-864.89	194.40	-40.05	3.66	0.63	0.63	0.63	50.00	244.00	160.00	6.00	36.00	36.00										
-859.09	207.00	-36.87	2.40	0.65	0.65	0.65	57.00	241.00	161.00	7.13	34.00	34.00										
-871.82	191.94	-36.16	1.98	0.66	0.67	0.67	53.00	240.00	162.00	4.74	34.00	34.00										
-857.08	199.71	-38.50	2.95	0.66	0.66	0.66	56.00	246.00	152.00	6.43	36.00	36.00										
-852.98	202.90	-40.43	0.93	0.67	0.66	0.66	55.00	245.00	159.00	6.06	35.00	37.00										
-860.18	208.94	-44.19	3.51	0.65	0.64	0.64	59.00	245.00	165.00	5.82	37.00	37.00										
-865.87	197.32	-37.63	2.85	0.66	0.66	0.66	45.00	248.00	154.00	5.89	36.00	36.00										
-846.62	201.56	-37.95	3.93	0.61	0.61	0.61	50.00	245.00	154.00	8.73	36.00	36.00										
-863.99	201.88	-40.28	0.41	0.66	0.66	0.66	52.00	246.00	155.00	5.48	35.00	37.00										
-860.96	199.88	-41.86	2.04	0.68	0.66	0.66	61.00	249.00	159.00	6.31	36.00	36.00										
-862.90	207.81	-43.73	1.98	0.64	0.65	0.65	50.00	253.00	154.00	4.98	35.00	35.00										
-853.27	199.78	-36.99	1.23	0.67	0.65	0.65	50.00	246.00	160.00	6.34	36.00	38.00										
-848.05	205.64	-35.85	3.03	0.65	0.64	0.64	52.00	245.00	165.00	8.13	35.00	35.00										
-861.19	194.51	-40.44	2.50	0.67	0.66	0.66	51.00	244.00	167.00	4.69	35.00	37.00										
-856.88	184.15	-42.18	1.62	0.64	0.64	0.64	54.00	253.00	155.00	8.31	34.00	37.00										

Figura 5.1: Exemplo de pontuações do Rosetta.

Tabela 5.1: Top 10 reação de eliminação de Kemp

Reação de eliminação de Kemp	Score Rosetta (Arb.U)	Ranqueamento	Entalpia	Reação de eliminação de Kemp	Entalpia (kcal/mol)
UML66_E180W83	-12.35	1	1	UML66_E180W83	-40.06
UML55_E159W110	-11.34	2	3	rep_201.35_UML126_E233W110	-34.29
rep_194.47_UML40_E133W110	-10.37	3	29	UML55_E159W110	-31.35
UML86_E184W89	-10.33	4	5	UML101_E211W89	-30.9
rep_205.72_UML101_E211W89	-10.3	5	7	UML86_E184W89	-29.87
UML102_E211Y110	-10.22	6	25	UML118_E231W110	-29.21
UML10_E51W89	-10.06	7	36	rep_205.72_UML101_E211W89	-28.8
UML6_E51W110	-10	8	10	rep_198.07_UML46_E159W51	-27.93
rep_193.39_UML101_E211W89	-9.81	9	46	UML43_E133W110	-27.73
UML79_E180W233	-9.81	10	22	UML6_E51W110	-27.3

projetadas. No processo de replicação do artigo, foi encontrado para essa estrutura 4 conformações na etapa de *match*, gerando, portanto, 40 estruturas desenhadas, e devido a esse fato, todas essas estruturas passaram pelo *script* DesignSelect.pl, juntos as outras estruturas de saída de protocolo Rosetta, reduzindo assim a quantidade de estruturas ruins para a etapa de pós-processamento.

A filtragem resultante está resumida na tabela 2, na seção de anexos, e apresenta as 4 estruturas que representa a estrutura proteica da DA_20_00. Apenas 52 estruturas projetadas passaram pelo filtro e com base nelas foram calculadas os $\Delta H_{binding}$ e sua reclassificação está presente na mesma tabela. Desse modo, com base na reclassificação, foi possível observar que as conformações da estrutura Best_UM_2_Y121Q195, deixaram as posições 10, 12, 13 e 16, indicada na tabela 2, para as posições 4, 5, 6 e 9, desse modo, ficando entre as 10 melhores estruturas.

A partir das entalpias de ligação enzima-ligante calculadas foram analisadas a correlação dessas com a pontuação proposta pelo Rosetta, sendo expressas nas figuras 5.2 e 5.3. Foi observado os R^2 de 0.189 e 0.006 para reação de eliminação de *kemp* e a reação de *Diels-Alder*, respectivamente. Com isso é possível inferir que a relação entre o $\Delta H_{binding}$ calculado e *Score* Rosetta não são lineares, portanto, ao calcular o $\Delta H_{binding}$ não se trata de informação redundante, o que pode acrescentar mais informação ao protocolo de *design* enzimático Rosetta. A reclassificação pelo $\Delta H_{binding}$, ao ser aplicado nessas reações, mostraram melhoria no ranqueamento, destacando a reação de *Diels-Alder* que foi reclassificada deixando a mutante que foi testada experimentalmente entre as 10 melhores posições referente as estruturas saídas do protocolo de *design* enzimático do Rosetta.

Em seguida, foram selecionadas as dez melhores estruturas classificadas pelo Rosetta e $\Delta H_{binding}$, tanto para reação de eliminação de *Kemp*, na tabela 5.1, quanto para a reação de *Diels-Alder*, na tabela 5.2. Para essas estruturas foram analisadas as mutações para cada reação, sendo definida pelo código de uma letra do novo resíduo de aminoácido com a numeração referente sua posição na sequência da estrutura primária, expressa nas tabelas 5.3 e 5.4 respectivamente.

Devido à presença de inúmeras mutações torna-se difícil e complexa a análise visual de cada uma delas, desse modo, foi necessário recorrer a ferramentas de bioinformática. Um *software* que atendeu a essa necessidade foi o Hbind (*Raschka et al.*, 2018), que contabiliza as interações no complexo proteína-ligante fornecendo

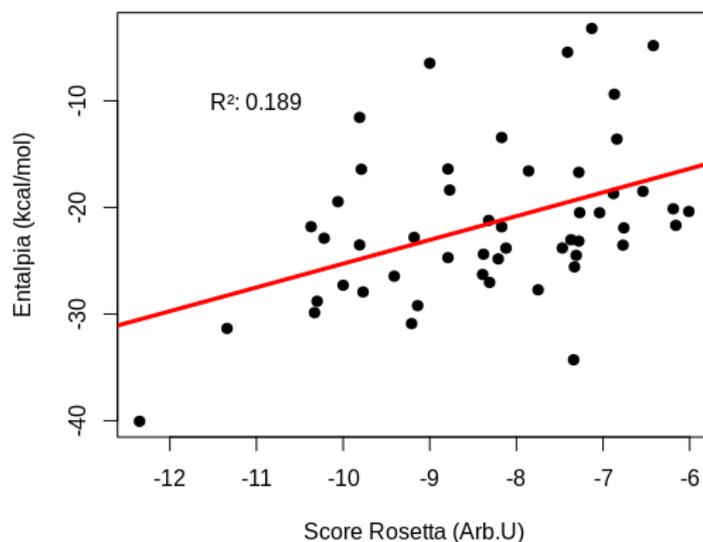


Figura 5.2: Relação *Score Rosetta* e $\Delta H_{binding}$ para reação de eliminação de Kemp.

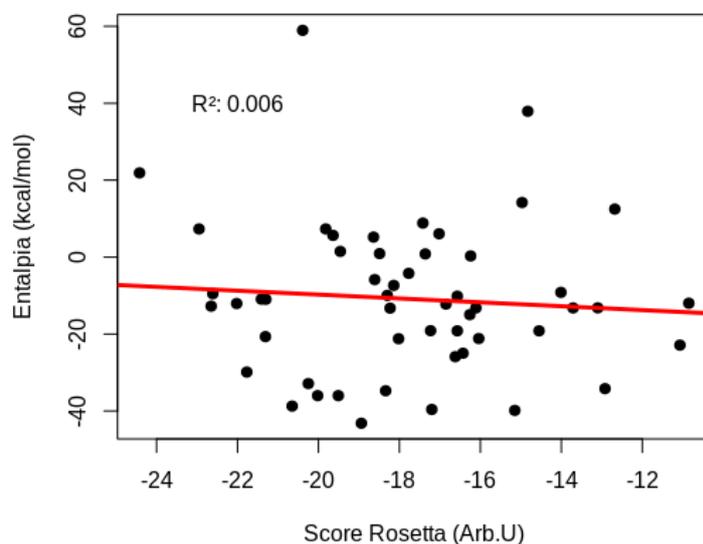


Figura 5.3: Relação *Score Rosetta* e $\Delta H_{binding}$ para reação de *Diels-Alder*.

o número de ligações de hidrogênio, Protein-Ligand H-bond, contatos hidrofóbicos, Protein-Ligand Hydrophobic Contacts, orientação do ligante no sítio projetado, SLIDE OrientScore, e uma pontuação referente ao número de ligações favoráveis e

Tabela 5.2: Top 10 Reação de Diels-Alder

Reação de Diels-Alder	Score Rosetta (Arb.U)	Ranqueamento	Entalpia	Reação de Diels-Alder	Entalpia (kcal/mol)
UM_81_T179Q77	-24.42	1	50	UM_52_T121Q74	-43.17376
rep_249.26_UM_64_Y121Q74	-22.95	2	45	rep_243.22_UM_72_Y121Q195	-39.83902
UM_89_T195Q74	-22.65	3	25	UM_90_T195Q120	-39.58869
rep_228.14_UM_64_Y121Q74	-22.61	4	33	BestHit_rep_222.73_UM_2_Y121Q195	-38.71903
rep_231.74_UM_68_Y121N74	-22.02	5	27	BestHit_rep_249.41_UM_2_Y121Q195	-36.00375
rep_245.99_UM_28_S120Q74	-21.77	6	10	BestHit_UM_2_Y121Q195	-36.00375
rep_243.32_UM_71_Y121Q195	-21.41	7	30	UM_35_S120Q74	-34.73546
UM_77_Y121Q74	-21.31	8	16	UM_37_S120Q74	-34.1821
UM_73_Y121Q195	-21.3	9	29	BestHit_rep_225.74_UM_2_Y121Q195	-32.86865
BestHit_rep_222.7_UM_2_Y121Q195	-20.65	10	4	rep_245.99_UM_28_S120Q74	-29.86748

Tabela 5.3: Top 10 eliminação de Kemp e suas mutações

Reação de eliminação de Kemp	Mutações Score Rosetta
UM_66_E180W83	E50L_K52Q_S55D_S80A_L81V_L82W_E84N_G90M_L107A_K109S_L130V_I132V_K134H_I135L_E158V_N160T_G177Q_N179E_S180R_R181L_I183R_Q193L_E209L_L230Q
UM_55_E159W110	E50L_K52A_S55D_S80A_L82S_K109W_D110L_L130T_K134W_L156M_I157L_N160L_G177V_I178V_N179V_R181T_L183R_E209A_L230G_G232T
rep_194.47_UM_40_E133W110	E50V_K52A_S80Q_L81V_L82T_L107A_K109W_D110A_F111M_L130V_I132E_V133T_K134M_E158V_N160L_G177A_N179V_R181L_I183Q_E209L_S210G_L230M_G232E
UM_86_E184W89	E50L_K52V_K54T_P56L_S80A_L82V_E84L_F88W_K109E_F111H_I132L_K134Q_E158V_N160T_N179L_R181L_D182L_I183E_E209R_S210G_G232T
rep_205.72_UM_101_E211W89	E50V_K52V_K54T_P56L_S57T_S80T_L82D_E84L_Y87L_F88W_K109M_D110A_F111W_L130A_E158L_N179L_R181L_I183W_E209L_S210E_G232L_S233E
UM_102_E211Y110	E50A_K52L_S55H_P56L_S80V_L81V_L82A_Y87L_F88K_L107M_K109Y_D110Q_F111H_L130V_S180A_R181L_I188Q_E209I_S210E_L230T_G232L
UM_10_E51W89	V11L_K52F_K54Q_P56L_S80A_L81V_L82S_T83V_E84Q_Y87L_F88W_L107M_K109V_D110A_E158Q_N179L_R181M_L183E_E209Q_L231H_G232T
UM_6_E51W110	A49V_Y51V_S55D_S80T_K109W_D110A_L130S_L156A_E158V_G177V_N179L_R181L_E209A_L230I
rep_193.39_UM_101_E211W89	E50L_K52T_K54A_P56L_S57M_S80T_L82D_E84L_F88W_N89K_K109T_L130M_N179L_S180V_R181L_E209L_S210E_L230M_G232I_S233L
UM_79_E180W233	E50V_K52T_S55H_P56L_S80A_L82M_E84V_F88Y_K109L_L130A_E158L_N160T_I178V_N179E_R181L_Q193L_E209L_S210G_L230L_G232W_S233E

Reação de eliminação de Kemp	Mutações Entalpia de ligação
UM_66_E180W83	E50L_K52Q_S55D_S80A_L81V_L82W_E84N_G90M_L107A_K109S_L130V_I132V_K134H_I135L_E158V_N160T_G177Q_N179E_S180R_R181L_I183R_Q193L_E209L_L230Q
rep_201.35_UM_126_E233W110	H48V_E50V_K52V_S55L_S80T_L82T_F88L_L107V_K109W_L130V_E158T_G177Q_N179L_E209A_S210G_L230V_G232E_S233L
UM_55_E159W110	E50L_K52A_S55D_S80A_L82S_K109W_D110L_L130T_K134W_L156M_I157L_N160L_G177V_I178V_N179V_R181T_L183R_E209A_L230L_G232T
UM_101_E211W89	E50L_K52F_K54V_P56L_L159V_S80A_L82T_F88W_K109L_D110A_F111W_E158A_N179L_R181L_I188L_E209M_S210E_L230A_G232L_S233E
UM_86_E184W89	E50L_K52V_K54T_P56L_S80A_L82V_E84L_F88W_K109E_F111H_I132L_K134Q_E158V_N160T_N179L_R181L_D182L_I183E_E209R_S210G_G232T
rep_118_E231W110	H48L_E50V_K52M_S80L_L81V_L82T_E84N_L107V_K109W_L130T_E158V_N160H_G177V_N179V_E209T_L230E_L231F_G232T
UM_205.72_UM_101_E211W89	E50V_K52V_K54T_P56L_S57T_S80T_L82D_E84L_Y87L_F88W_K109M_D110A_F111W_L130A_E158L_N179L_R181L_I183W_E209L_S210E_G232L_S233E
rep_198.07_UM_46_E159W51	H48V_E50V_Y51V_K52M_P56L_S80A_L82V_K109L_L130V_L156L_I157L_G177A_R181N_L183M_E209L_L230S_G232S
UM_43_E133W110	E50V_K52L_S80A_L82A_F88W_L107A_K109W_D110L_F111Y_I112N_L130H_I132E_K134S_E158A_N160T_G177L_N179L_R181Q_D182L_E209A_L230Q_G232T
UM_6_E51W110	A49V_Y51V_S55D_S80T_L82A_K109W_D110A_L130S_L156A_E158V_G177V_N179L_R181L_E209A_L230I

desfavoráveis no complexo, SLIDE AffiScore. Os resultados estão, para cada reação, expressa nas tabelas 5.5 e 5.6.

Analisando as tabelas apresentadas nota-se que a classificação tanto do $\Delta H_{binding}$ quanto do *score* Rosetta fornece estruturas com contabilização de interações semelhantes, dentro de uma estreita faixa, com relação à orientação do ligante, contatos hidrofóbicos e ligações de hidrogênio para a reação de eliminação de *Kemp* e reação de *Diels-Alder*. Desse modo, em termos de interações as duas classificações são satisfatórias.

Tabela 5.4: Top 10 Reação de Diels-Alder e suas mutações

Reação de Diels-Alder	Mutações Score Rosetta
UM_81_T179Q77	E21S_P39A_V34A_C70V_C78L_F86V_D121A_C122L_F124L_Y126N_V179T_D229S_G230S_F233L_N272G_H274A_K302D_K303V_Q304S_Y305A_T308L
rep_249.26_UM_64_Y121Q74	E21S_G22A_P36T_P44L_F27P_P73T_A74Q_G75T_D89V_R91S_C119A_N120A_D121Y_T133V_P135V_A136M_Y144R_R146M_L148L_F173A_P174G_N175T_G176A_I177V_D229G_G230Q_W244E_K269Y_S271N_N272G_H287E
UM_80_T195Q74	E21M_G22A_V34T_P36Y_P44L_I72A_P78L_A74Q_G75D_M90L_R91Q_N120V_D121A_P135T_A136S_Y144R_R146M_M148L_F173A_P174G_N175S_P196Y_D229S_W244E_K269N_S271G_N272D
rep_228.14_UM_64_Y121Q74	E21T_G22A_V34A_P36V_E37V_F72S_P73V_A74Q_G75S_R91Q_C119A_N120A_D121Y_C122L_T133V_P135T_Y144T_L145K_R146L_F173A_N175S_G176A_I177V_D229G_W244E_K269H_H287E
rep_231.74_UM_68_Y121N74	E21T_G22A_V34A_P36V_E37A_P44R_G71A_P72T_P73L_A74S_G75A_F88S_M90L_R91A_C119T_N120A_D121Y_C122M_T133A_P135Y_Y144V_R146M_F173Q_N175E_P177V_D229V_G230Q_W244E_K269Y_S271N_N272G_H287E
rep_245.99_UM_28_S120Q74	E21S_P36T_E39V_P44T_G70A_I72A_P73M_A74Q_G75D_D89A_M90L_C119A_N120S_D121Y_T133V_Y144L_R146M_L148L_F173A_N175T_T195A_D229S_W244E_S271L_N272A
rep_243.32_UM_71_Y121Q195	E21A_F72L_A74D_G75K_M90G_N120D_D121Y_T133V_P135T_R146L_M148V_F173M_P174A_E194T_T195Q_P196L_E225S_D229V_G230D_W244E_K269H_S271L_H287K
UM_77_Y121Q74	E21N_G22A_V34A_P36A_E37F_I72T_P73V_A74Q_G75A_M90L_R91K_N120A_D121Y_C122M_T133V_P135L_Y144Q_R146L_M148E_F173A_N175T_T195A_D229T_W244E_K269H_H287D
UM_73_Y121Q195	E21T_G22A_P36E_I72L_A74V_G75S_A88T_M90G_C119L_N120D_D121Y_T133V_P135L_R146L_M148V_F173L_P174A_E194Q_E195Q_P196A_D225S_D229G_A242V_W244H_S271L_N272A_H287E
BestHit_rep_222.7_UM_2_Y121Q195	E21A_E37V_N120A_D121Y_Y144F_R146L_M148L_F173A_N175A_T195Q_E225K_D229A_N272A

Reação de Diels-Alder	Mutações Entalpia de ligação
UM_52_T121Q74	E21A_G22A_V34A_P36A_E37Q_P73V_A74Q_M90L_N120S_D121T_T133A_P135A_N175A_I177V_D229S_W244M_K269Y_S271A_N272D_H287I
rep_243.22_UM_72_Y121Q195	E21N_A74V_G75A_M90G_C119L_N120D_D121Y_T133V_P135T_R146Q_F173A_N175G_I177V_E194S_T195Q_P196E_E225S_D229A_G230V_W244L_G245N_K269Q_S271T_N272A_H287W
UM_90_T195Q120	E21F_P36A_P72M_A74L_M90A_R91S_N120Q_D121A_P135A_A136T_G137T_Y144L_T145V_R146E_M148A_F173A_P174A_N175G_E194S_P196L_D229S_W244F_K269V_S271T_N272D_H287L
BestHit_rep_222.73_UM_2_Y121Q195	E21A_E37V_N120A_D121Y_Y144F_R146L_M148L_F173A_N175A_T195Q_E225K_D229A_N272A
BestHit_rep_249.41_UM_2_Y121Q195	E21A_E37V_N120A_D121Y_Y144F_R146L_M148L_F173A_N175A_T195Q_E225K_D229A_N272A
BestHit_UM_2_Y121Q195	E21A_E37V_N120A_D121Y_Y144F_R146L_M148L_F173A_N175A_T195Q_E225K_D229A_N272A
UM_35_S120Q74	E21V_G22A_V34A_P36V_E37L_I72L_P73V_A74Q_G75A_M90G_N120S_P135V_F173Q_N175A_G176A_T195A_D229S_W244M_K269L_S271A_N272M_H287Q
UM_37_S120Q74	E21A_G22A_V34A_P36S_E37S_P73V_A74Q_G75S_M90L_G183A_N198S_D191S_F178V_N179V_T195A_D229G_W244F_K269V_S271T_N272D_H287V
BestHit_rep_225.74_UM_2_Y121Q195	E21A_E37V_N120A_D121Y_Y144F_R146L_M148L_F173A_N175A_T195Q_E225K_D229A_N272A
rep_245.99_UM_28_S120Q74	E21S_P36T_E39V_P44T_G70A_I72A_P73M_A74Q_G75D_D89A_M90L_C119A_N120S_D121Y_T133V_Y144L_R146M_L148L_F173A_N175T_T195A_D229S_W244E_S271L_N272A

Tabela 5.5: Top 10 eliminação de Kemp e as contabilização das interações

Reação de eliminação de Kemp (Score Rosetta)	Protein-Ligand Hydrophobic Contacts	Protein-Ligand H-bond	SLIDE OrientScore	SLIDE AffiScore
UML66_E180W83	40	3	-6.451	-6.32
UML55_E159W110	43	2	-6.399	-6.227
rep_194.47_UML40_E133W110	43	2	-6.394	-6.357
UML86_E184W89	39	3	-6.428	-6.836
rep_205.72_UML101_E211W89	52	1	-6.301	-6.148
UML102_E211Y110	49	1	-6.162	-6.271
UML10_E51W89	34	2	-6.25	-5.953
UML6_E51W110	40	2	-6.033	-6.602
rep_193.39_UML101_E211W89	39	1	-6.128	-5.996
UML79_E180W233	35	2	-5.922	-6.507

Reação de eliminação de Kemp (Entalpia de ligação)	Protein-Ligand Hydrophobic Contacts	Protein-Ligand H-bond	SLIDE OrientScore	SLIDE AffiScore
UML66_E180W83	40	3	-6.451	-6.32
rep_201.35_UML126_E233W110	46	0	-6.141	-5.911
UML55_E159W110	43	2	-6.399	-6.227
UML101_E211W89	56	1	-6.44	-6.151
UML86_E184W89	39	3	-6.428	-6.836
UML118_E231W110	35	2	-5.978	-6.177
rep_205.72_UML101_E211W89	52	1	-6.301	-6.148
rep_198.07_UML46_E159W51	42	2	-6.192	-6.34
UML43_E133W110	37	1	-5.852	-6.111
UML6_E51W110	40	2	-6.033	-6.602

Tabela 5.6: Top 10 Reação de Diels-Alder e as contabilização das interações

Reação de Diels-Alder (Score Rosetta)	Protein-Ligand Hydrophobic Contacts	Protein-Ligand H-bond	SLIDE OrientScore	SLIDE AffiScore
UML81_T179Q77	56	4	-7.573	-6.605
rep_249.26_UML64_Y121Q74	67	1	-6.988	-7.355
UML89_T195Q74	58	4	-7.397	-8.472
rep_228.14_UML64_Y121Q74	37	6	-6.988	-7.562
rep_231.74_UML68_Y121N74	61	3	-7.282	-7.237
rep_245.99_UML28_S120Q74	39	6	-7.087	-8.08
rep_243.32_UML71_Y121Q195	34	4	-6.384	-7.351
UML77_Y121Q74	57	4	-7.221	-7.926
UML73_Y121Q195	42	4	-6.683	-8.427
BestHit_rep_222.7_UML2_Y121Q195	40	4	-6.622	-7.87

Reação de Diels-Alder (Entalpia de ligação)	Protein-Ligand Hydrophobic Contacts	Protein-Ligand H-bond	SLIDE OrientScore	SLIDE AffiScore
UML52_T121Q74	56	5	-7.467	-8.241
rep_243.22_UML72_Y121Q195	34	3	-6.23	-7.559
UML90_T195Q120	50	1	-6.726	-7.164
BestHit_rep_222.73_UML2_Y121Q195	40	4	-6.622	-7.87
BestHit_rep_249.41_UML2_Y121Q195	38	4	-6.562	-7.881
BestHit_UML2_Y121Q195	38	4	-6.562	-7.881
UML35_S120Q74	71	3	-7.618	-7.916
UML37_S120Q74	53	3	-6.841	-8.059
BestHit_rep_225.74_UML2_Y121Q195	40	4	-6.597	-7.905
rep_245.99_UML28_S120Q74	39	6	-7.087	-8.08

5.2 Descritores moleculares

Utilizando o programa PRIMoRDiA buscou-se estudar o sítio de ligação de 3 enzimas desenhadas sendo a reação de *Retro-Aldol* (Jiang et al., 2008), *Diels-Alder* (Siegel et al., 2010) e eliminação *Kemp* (Röthlisberger et al., 2008). Para o artigo que apresentou a reação de *Kemp* foram retiradas do seu material suplementar as 3 estruturas denominadas KE59, KE07 e KE70. A primeira estrutura citada não apresentou atividade catalítica, sendo renomeada para SupleKE59. Devido a esse fato, para incluir a versão catalítica da KE59 foi feito o processo de redesenho enzimático do Rosetta obtendo a enzima com mesmo código FASTA, sendo nomeada KE59-nativa. Para a reação de Diels-Alder, devido à ausência de estruturas disponíveis, foi necessário reproduzir o estudo, e após o uso do Rosetta foi possível obter suas estruturas para posterior cálculo dos descritores. Por fim, não foi possível reproduzir o estudo da reação *Retro-Aldol*, mas como ele apresentou em seu material suplementar as estruturas desenhadas em formato pdb, o estudo dos descritores foi realizado buscando entendê-las. Todos os descritores foram calculados com ligante acoplado, buscando, assim como no artigo (Grillo et al., 2020d), entender o modo de ligação enzima-ligante.

Visando encontrar os melhores descritores a serem utilizados, as estruturas foram calculadas com os descritores de banda (BD) e energia ponderada (EW). Ao analisar as estruturas referentes a eliminação *Kemp* foi possível observar para as estruturas KE70, KE07 e KE59-nativa, a localização da propriedade NAS no ligante. Já a estrutura SupleKE59, estrutura que não apresentou atividade catalítica e apresenta apenas um resíduo diferente em relação ao KE59-nativa, mostrada na figura 5.4, não apresentou seus descritores no ligante ou proximidades. Essa estrutura junto com as KE70, KE07 e KE59-nativa estão expressada na figura 5.5. Desse modo, essas estruturas apresentaram concordância com o que o estudo de Fukushima e colaboradores observou, sobre a relação da atividade enzimática e a região em que seus orbitais HOMO-LUMO se encontram (Fukushima et al., 2008). Nesse estudo, os autores observaram que os orbitais HOMO-LUMO da enzima se localizam próximos do seu sítio ativo, quando maior for a especificidade de uma enzima por um ligante.

Usando como base a análise anterior, buscou-se encontrar estruturas no grupo das TOP 10, ranqueada tanto pela pontuação Rosetta quanto pelo $\Delta H_{binding}$, padrões similares ou até melhores em relação à figura 5.5, utilizando a observação de Fukushima.

Analisando essas estruturas podemos observar que a propriedade NAS para todas elas se apresentam localizados no ligante, enquanto o EAS varia bastante. Comparando esses descritores com os descritores das estruturas que foram testadas experimentalmente, tanto as ranqueadas pela pontuação do Rosetta quanto pelo valor de $\Delta H_{binding}$, apresentaram boa probabilidade de possuírem atividade catalítica. Dentre elas, devido à proximidade do NAS e EAS, ou seja, do HOMO e LUMO, se destacam as estruturas rep_193.39_UM_101_E211W89 e

```

CLUSTAL O(1.2.4) multiple sequence alignment

KE59.pdb      PRYKLGWLKDVVQLSRRPSFRASRQRPDIISLNERILEFNKRNITAIIVYKRKSPSGLD 60
KE59-native  PRYKLGWLKDVVQLSRRPSFRASRQRPDIISLNERILEFNKRNITAIIVYKRKSPSGLD 60
*****

KE59.pdb      VERDPIEYSKFMERYAVGLVILTEEKYFNGSYETLRKIASSVSIPILMDFIVKESQIDD 120
KE59-native  VERDPIEYSKFMERYAVGLVILTEEKYFNGSYETLRKIASSVSIPILMDFIVKESQIDD 120
*****

KE59.pdb      AYNLGADTVSLIVKILTERELESLLLEYARSYGMPEAIVINDENDLDIALRIGARFIISS 180
KE59-native  AYNLGADTVGLIVKILTERELESLLLEYARSYGMPEAIVINDENDLDIALRIGARFIISS 180
*****

KE59.pdb      RDLETLEINKENQRKLISMIPSNVVKVAASGISERNEIEELRKLGVNAFEIGSSLMRNP 240
KE59-native  RDLETLEINKENQRKLISMIPSNVVKVAASGISERNEIEELRKLGVNAFEIGSSLMRNP 240
*****

KE59.pdb      KIKEFIL      247
KE59-native  KIKEFIL      247
*****

```

Figura 5.4: Alinhamento da estrutura KE59-native, sendo a forma catalítica, com a estrutura não catalítica nomeada KE59.pdb.

rep_198.07_UM_46_E159W51, expressas nas figuras 5.6 e 5.7 respectivamente. Contudo, para ter certeza de sua melhor atividade catalítica será necessário realizar as etapas experimentais buscando a caracterização cinética dessas enzimas. Outro resultado interessante foi que a reclassificação pelo $\Delta H_{binding}$ fez com que a primeira, rep_193.39_UM_101_E211W89, estrutura saísse da posição 9 para 46 enquanto a segunda, rep_198.07_UM_46_E159W51, da 12 para a 8.

Para a reação de *Diels-Alder* foi necessário refazer os dados do artigo original, pois os autores só disponibilizaram os arquivos FASTA e as informações sobre o arquivo de restrição e proteína usada como base no protocolo enzimático de *design* Rosetta. A enzima obtida através do protocolo foi nomeada como BestHit e devido sua baixa atividade o pesquisador utilizou a intuição química nessa estrutura, alterando 6 resíduos de aminoácido. Desse modo, foi possível aumentar sua capacidade catalítica, e essa estrutura foi replicada sendo nomeada por BestHit-6M.

Como o BestHit encontrou 4 *matches* calculamos seus descritores, expressos na figura 5.8. Duas dessas estruturas, a BestHit_rep_222.7_UM_2_Y121Q195 e BestHit_rep_225.74_UM_2_Y121Q195 apresentou a propriedade EAS localizado próximo ao ligante e seu NAS variou bastante, enquanto para os outros dois BestHit, seus descritores se dispersaram para longe do ligante. No artigo publicado, os autores chegam a um resultado de que essa catálise de $0.10 \pm 0.02 \text{ hour}^{-1}$. Analisando a BestHit-6M, na figura 5.10, que apresentou atividade catalítica de $2.13 \pm 0.24 \text{ hour}^{-1}$, podemos perceber que o mesmo padrão se repete, EAS localizado próximo ao ligante e NAS distante do sítio-ativo, indicando que as mutações não promoveram tanta mudança nos descritores, o que reflete ao realizar a caracterização cinética a baixa catálise. Dentro das estruturas desenhadas pelo protocolo do Rosetta, gerados para reproduzir o artigo, se destacou a UM_90_T195Q120, expressa na figura 5.9, que pela reclassificação utilizando os valores de $\Delta H_{binding}$ saiu da posição 31 para a 3, e apresentou NAS localizado em parte no ligante e nos resíduos Ala-173-174 e EAS no resíduo Cys-265.

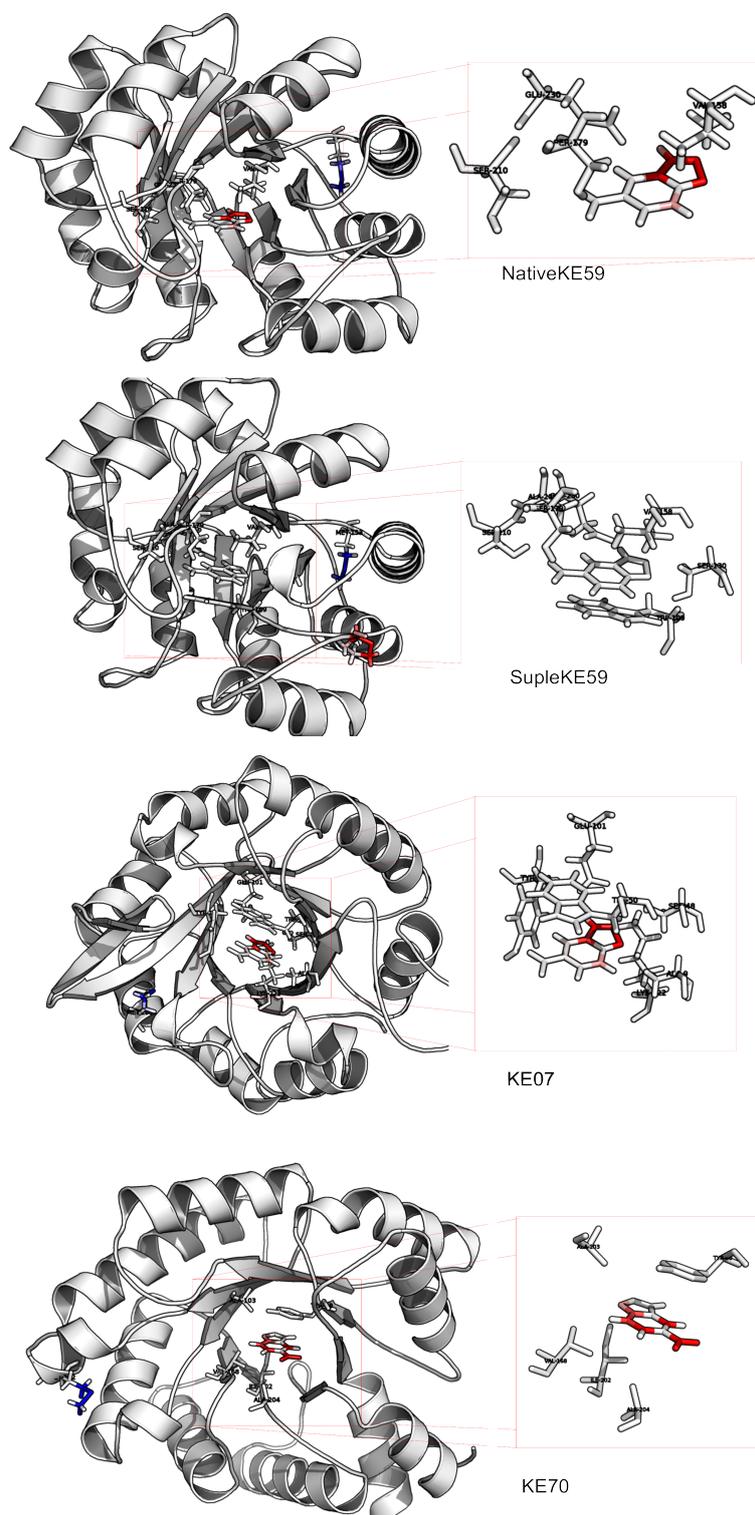


Figura 5.5: As 4 estruturas desenhadas segundo o artigo referente à reação de eliminação de *Kemp* pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. A KE70 apresentou NAS no ligante e EAS no resíduo Met-112. KE07 EAS no resíduo Met-186. NativeKE59 EAS no resíduo Met-153. SupleKE59 NAS no resíduo Pro-1 e EAS no Met-153.

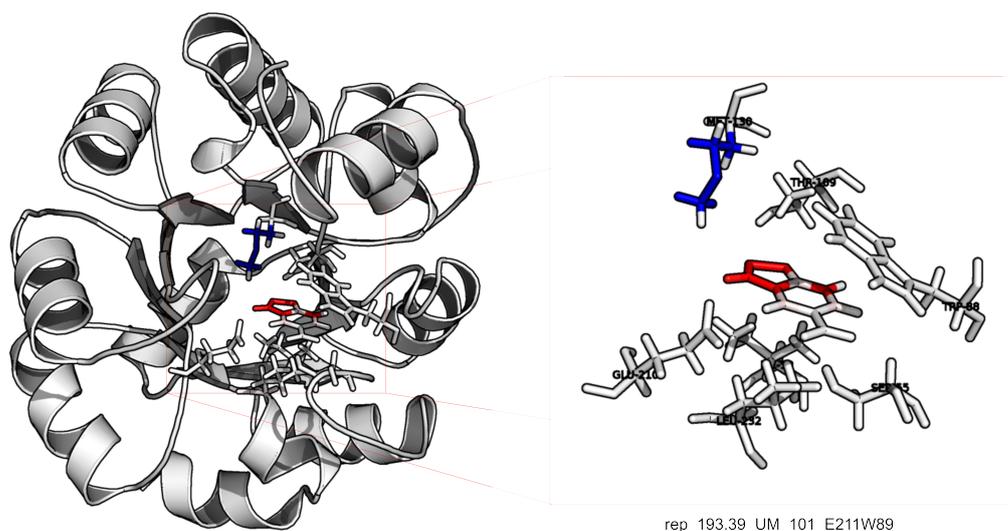


Figura 5.6: *Design* da reação de eliminação de *Kemp* segundo a pontuação do Rosetta pelos descritores que expressa a propriedade Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. A estrutura rep_193.39_UM_101_E211W89 apresentou propriedade EAS no resíduo Met-130 e o NAS localizado no ligante.

No estudo da reação de *Retro-Aldol*, que disponibilizaram 5 estruturas em formato PDB, foi analisado os descritores calculados a partir de das estratégias EW e BD. Enquanto descritores calculados com EW é mais focado em poucos resíduos, aqueles calculados com BD se espalham por uma maior superfície. A análise do EAS e NAS, no formalismo EW, demonstraram uma maior presença de um deles em resíduos próximos ao ligante, figura 5.11, exceto a estrutura nomeada de RA34, que mostrou NAS próximo ao ligante apenas quando utilizamos o formalismo, BD, figura 5.12.

Portanto, ao analisar as estruturas utilizando os descritores de reatividade foram observadas uma conformidade com os dados experimentais. Pois, utilizando as descobertas de Fukushima e colaboradores, ao comparar a reação de *Retro-Aldol* e *Diels-Alder* com a reação de *Kemp* foi observado a propriedade NAS ou EAS próximo ao ligante com k_{cat} muito abaixo quando comparado a reação de *Kemp*, que por possuir a propriedade NAS localizado no ligante apresentou k_{cat} superior.

A partir dessa observação é possível propor estruturas que poderiam apresentar melhor atividade, contudo, a estrutura RA46, pertencente a reação de *Retro-Aldol*, não apresentou o mesmo padrão observado nas outras reações, mesmo após o cálculo dos seus descritores na forma ponderada, EW, e de banda, BD. Apesar da divergência, sua atividade catalítica foi baixa, indicando que a falta de propriedades NAS e EAS localizadas no ligante representa baixo k_{cat} .

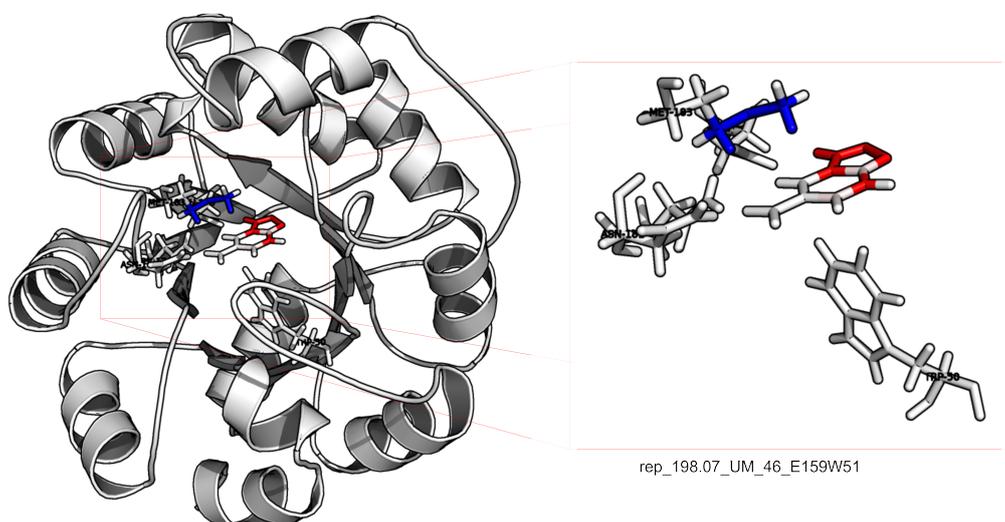


Figura 5.7: *Design* da reação de eliminação de *Kemp* segundo a pontuação $\Delta H_{binding}$ pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. A estrutura rep_198.07_UM_46_E159W51 apresentou propriedade EAS no resíduo Met-183 e NAS localizado no ligante.

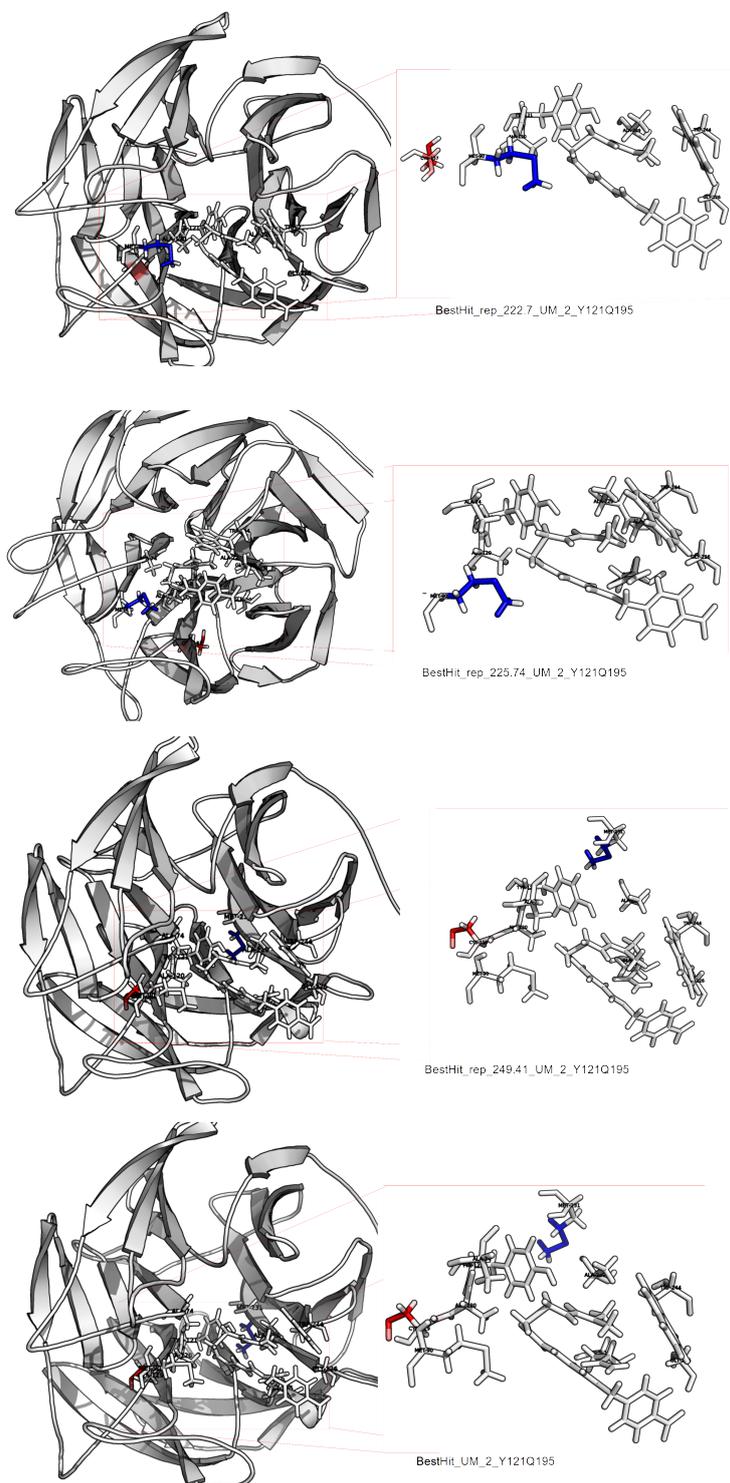


Figura 5.8: As 4 estruturas BestHit segundo o *Design* da reação de *Diels-Alder* utilizando a propriedade do descritor de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. BestHit_rep_222.7_UM_2_Y121Q195 EAS no resíduo Met-90 e NAS no resíduo Cys-157. BestHit_rep_249.41_UM_2_Y121Q195 EAS no resíduo Met-231 e NAS no resíduo Cys-119. BestHit_UM_2_Y121Q195 EAS no resíduo Met-231 e NAS no resíduo Cys-119. BestHit_rep_225.74_UM_2_Y121Q195 EAS no resíduo Met-90 e NAS no resíduo Cys-157.

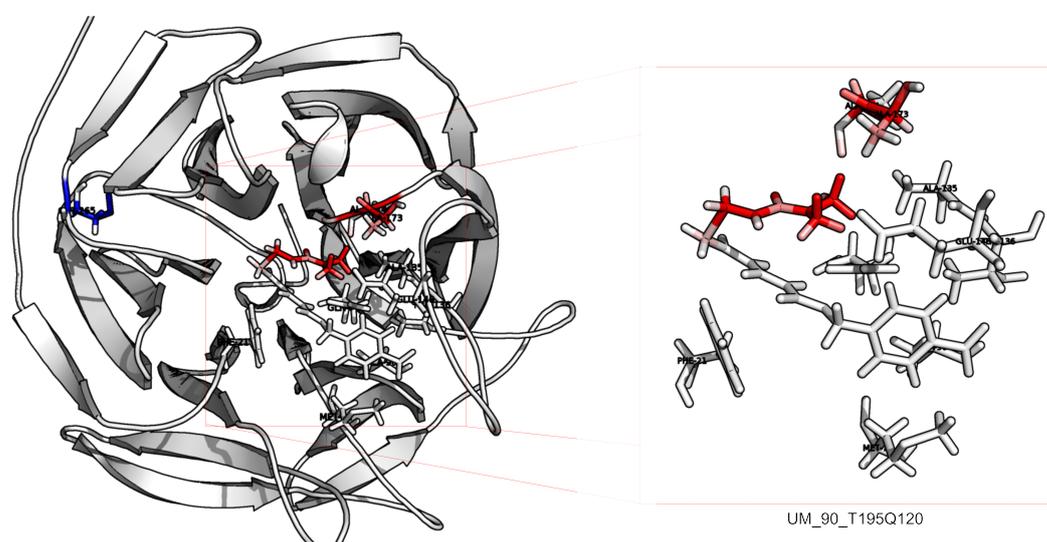


Figura 5.9: *Design* da reação de *Diels-Alder* segundo a pontuação do $\Delta H_{binding}$ utilizando a propriedade do descritor de Netfilicidade na forma EW. A estrutura UM_90_T195Q120 EAS no resíduo Cys-265 e NAS no resíduo Ala-173-174.

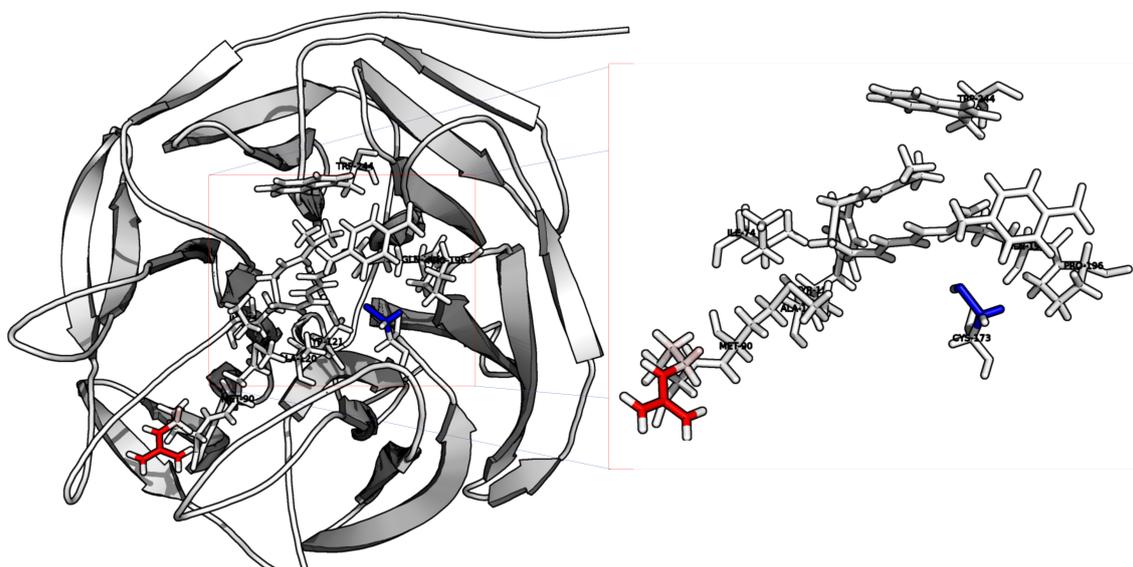


Figura 5.10: Essa estrutura é obtida usando como base a estrutura BestHit-6M da reação de *Diels-Alder* pelos descritores de Netfilicidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. O EAS no resíduo Cys-173 e NAS no resíduo Arg-91. Foram alterados seis mutações, sendo A21T, A74I, Q149R, A173C, S271A, e A272N.

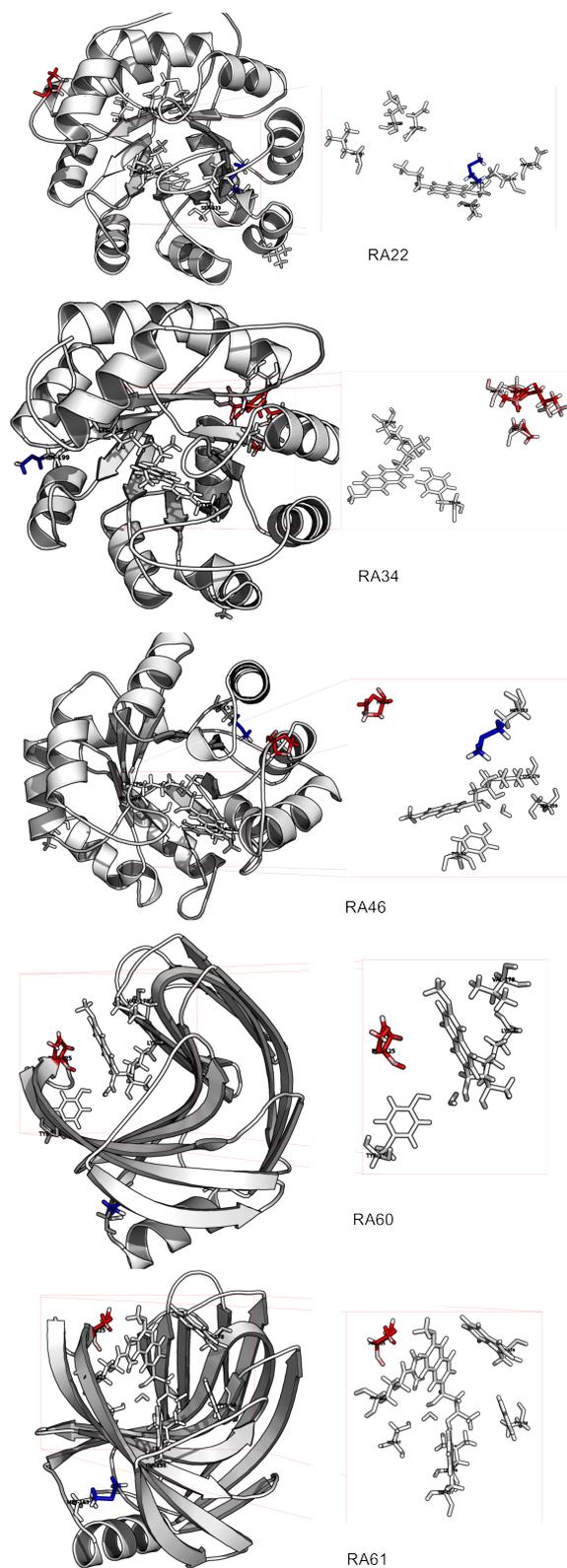


Figura 5.11: As 5 estruturas desenhadas segundo o artigo referente a reação de *Retro-Aldol* pelos descritores de Netfilterabilidade na forma EW. Foi definida em vermelho o NAS e em azul o EAS. RA22 EAS no resíduo Met-236 e NAS no resíduo Pro-1. RA34 EAS no resíduo Met-199 e NAS localizados no resíduo Asp-127-Ser-31-Ile-30-Pro105. RA46 EAS no resíduo Met-153 e NAS no resíduo Pro-1. RA60 EAS no resíduo Met-163 e NAS no resíduo Pro-125. RA61 EAS no resíduo Met-163 e NAS no resíduo Pro-125.

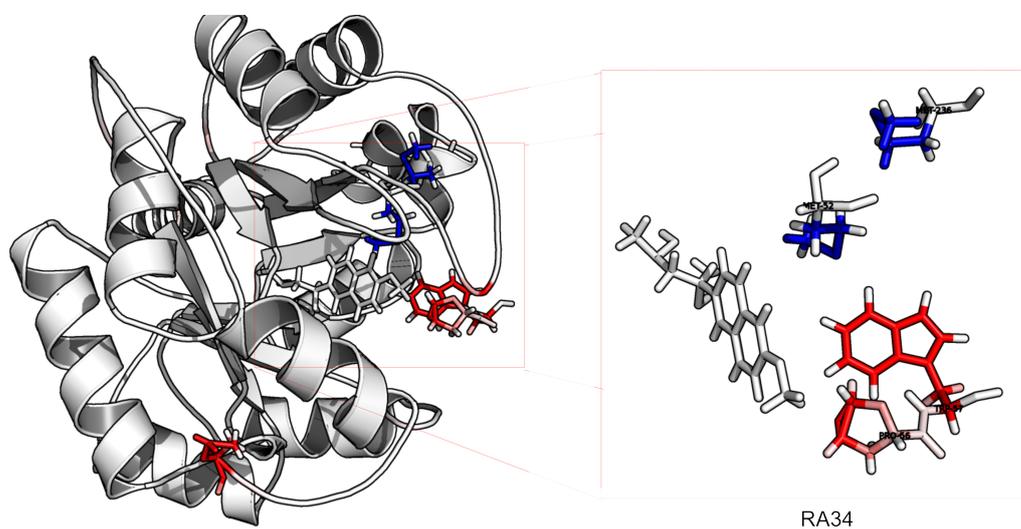


Figura 5.12: A RA34 estrutura desenhada segundo o artigo referente a reação de *Retro-Aldol* pelos descritores de NAS na forma BD. O NAS nos resíduos Pro-56 e Trp-57 e EAS localizado nos resíduos Met-52-236.

Capítulo 6

Conclusões

Nesse trabalho, procuramos apresentar um melhoramento ao procedimento de seleção e classificação usado no desenho computacional de enzimas usado no programa Rosetta. Sendo apresentado o conceito e o uso dos valores calculados a partir de métodos quânticos semiempíricos da entalpia de ligação enzima-ligante, $\Delta H_{binding}$, para reclassificação de estruturas mutantes, além do uso de descritores de reatividade obtidos com o programa PRIMoRDIA para análise visual. Com essas ferramentas em mãos, o cálculo de $\Delta H_{binding}$ auxiliou no re-ranqueamento, permitindo que estruturas que possuem dados experimentais, para a reação de eliminação de *Kemp* e *Diels-Alder*, assumisse melhores classificações, desse modo, surge a possibilidade de que ao realizar a etapa experimental, as enzimas promissoras que possam surgir aumente. A partir do cálculo dos descritores de reatividade foi possível observar que as estruturas projetadas nos estudos da reação de *Retro-Aldol* e *Diels-Alder*, em que ambas possuem k_{cat} entre 10^{-5} e 10^{-4} s $^{-1}$, mostraram NAS ou EAS localizado próximo ao ligante, no sítio ativo, enquanto que as estruturas calculadas para a reação de eliminação *Kemp*, com k_{cat} entre 10^{-1} e 10^{-2} s $^{-1}$, mostraram as propriedades NAS localizadas no ligante. Para a reação de eliminação de *Kemp* foi possível distinguir a estrutura supleKE59, que não possui atividade catalítica, com propriedades NAS e EAS localizados fora do sítio ativo, e a estrutura nativeKE59, com a propriedade NAS no ligante, que apresentou k_{cat} igual a 2.90×10^{-1} . Portanto, o uso do protocolo de *design* enzimático Rosetta combinado com os descritores permitiram distinguir estruturas com potencial catalítico observando as propriedades referentes a suscetibilidade de ataque nucleofílico e eletrofílico, ou seja, a localização dos seus orbitais próximos em energia ao HOMO e LUMO nas proximidades do sítio ativo das enzimas.

Capítulo 7

Perspectivas

Para trabalhos futuros, planeja-se buscar um banco de dados com dados experimentais que apresente mutantes de diversas enzimas para reações diferentes, confirmando seus efeitos através parâmetros catalíticos como k_{cat} , que indica a constante de velocidade para reação enzimática. Com esses dados em mão, a intenção seria de projetar o estado de transição dessas catálises para os ligantes para que, em seguida, seja possível aplicar o protocolo de *design* enzimático do Rosetta. Desse modo, será aplicado o protocolo desenvolvido nesse estudo, propondo diversas funções *escore*, que serão montadas utilizando técnica de aprendizado de máquina, testando automaticamente os vários modelos e escolhendo o melhor.

Referências

- Abrahamson, M. J., E. Vázquez-Figueroa, N. B. Woodall, J. C. Moore, and A. S. Bommarius, Development of an amine dehydrogenase for synthesis of chiral amines, *Angewandte Chemie International Edition*, 51(16), 3969–3972, 2012.
- Alford, R. F., et al., The rosetta all-atom energy function for macromolecular modeling and design, *Journal of chemical theory and computation*, 13(6), 3031–3048, 2017.
- Anfinsen, C. B., E. Haber, M. Sela, and F. White Jr, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proceedings of the National Academy of Sciences of the United States of America*, 47(9), 1309, 1961.
- Bhatia, S., and S. Bhatia, Introduction to enzymes and their applications, *Introduction to pharmaceutical biotechnology*, 2, 1–29, 2018.
- Biolabs, N. E., Master mixes, url <https://international.neb.com/products/pcr-qpcr-and-amplification-technologies/master-mixes/master-mixes>, 2022.
- Bornscheuer, U. T., and M. Pohl, Improved biocatalysts by directed evolution and rational protein design, *Current opinion in chemical biology*, 5(2), 137–143, 2001.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. a. Swaminathan, and M. Karplus, Charmm: a program for macromolecular energy, minimization, and dynamics calculations, *Journal of computational chemistry*, 4(2), 187–217, 1983.
- Cao, L.-c., R. Chen, W. Xie, and Y.-h. Liu, Enhancing the thermostability of feruloyl esterase estf27 by directed evolution and the underlying structural basis, *Journal of agricultural and food chemistry*, 63(37), 8225–8233, 2015.
- Chen, K., and F. H. Arnold, Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin e for catalysis in dimethylformamide, *Proceedings of the National Academy of Sciences*, 90(12), 5618–5622, 1993.
- Chica, R. A., N. Doucet, and J. N. Pelletier, Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, *Current opinion in biotechnology*, 16.

- Cornell, W. D., et al., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *Journal of the American Chemical Society*, *117*(19), 5179–5197, 1995.
- Cox, M., and D. Nelson, Chapter 6.2: How enzymes work, *Lehninger Principles of Biochemistry*, p. 195, 2013.
- Engelking, L. R., and L. Engelking, Chapter 6—enzyme kinetics, *Textbook of veterinary physiological chemistry*, *3*, 2015.
- Fox, R. J., et al., Improving catalytic function by prosar-driven enzyme evolution, *Nature biotechnology*, *25*(3), 338–344, 2007.
- Frazzetto, G., White biotechnology, *EMBO reports*, *4*(9), 835–837, doi:<https://doi.org/10.1038/sj.embor.embor928>, 2003.
- Fukushima, K., M. Wada, and M. Sakurai, An insight into the general relationship between the three dimensional structures of enzymes and their electronic wave functions: Implication for the prediction of functional sites of enzymes, *Proteins: Structure, Function, and Bioinformatics*, *71*(4), 1940–1954, 2008.
- Geerlings, P., F. De Proft, and W. Langenaeker, Conceptual density functional theory, *Chemical reviews*, *103*(5), 1793–1874, 2003.
- Godoi Contessoto, V. d., A. B. d. Oliveira Junior, J. Chahine, R. J. d. Oliveira, and V. B. Pereira Leite, Introdução ao problema de enovelamento de proteínas: uma abordagem utilizando modelos computacionais simplificados, *Revista Brasileira de Ensino de Física*, *40*, 2018.
- Grillo, I. B., J. F. R. Bachega, L. F. S. Timmers, R. A. Caceres, O. N. de Souza, M. J. Field, and G. B. Rocha, Theoretical characterization of the shikimate 5-dehydrogenase reaction from mycobacterium tuberculosis by hybrid qc/mm simulations and quantum chemical descriptors, *Journal of Molecular Modeling*, *26*(11), 1–12, 2020a.
- Grillo, I. B., G. A. Urquiza-Carvalho, J. F. R. Bachega, and G. B. Rocha, Elucidating enzymatic catalysis using fast quantum chemical descriptors, *Journal of Chemical Information and Modeling*, *60*(2), 578–591, 2020b.
- Grillo, I. B., G. A. Urquiza-Carvalho, E. J. F. Chaves, and G. B. Rocha, Semiempirical methods do fukui functions: Unlocking a modeling framework for biosystems, *Journal of Computational Chemistry*, *41*(9), 862–873, 2020c.
- Grillo, I. B., G. A. Urquiza-Carvalho, and G. B. Rocha, Primordia: A software to explore reactivity and electronic structure in large biomolecules, *Journal of Chemical Information and Modeling*, *60*(12), 5885–5890, 2020d.

- Heckmann, C. M., and F. Paradisi, Looking back: A short history of the discovery of enzymes and how they became powerful chemical tools, *ChemCatChem*, *12*(24), 6082–6102, 2020.
- Jiang, L., et al., De novo computational design of retro-aldol enzymes, *science*, *319*(5868), 1387–1391, 2008.
- Jorgensen, W. L., and J. Tirado-Rives, The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *Journal of the American Chemical Society*, *110*(6), 1657–1666, 1988.
- Kan, S. J., R. D. Lewis, K. Chen, and F. H. Arnold, Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life, *Science*, *354*(6315), 1048–1051, 2016.
- Klibanov, A. M., Stabilization of enzymes against thermal inactivation, *Advances in applied microbiology*, *29*, 1–28, 1983.
- Laschuk, E., Novo formalismo semi-empírico para cálculos químico-quânticos. 143p, Ph.D. thesis, Tese (Doutorado em Química)–Instituto de Química, UFRGS, Porto Alegre, 2005.
- Mati Karelson, V. S. L., and A. R. Katritzky, Quantum-chemical descriptors in qsar/qspr studies, *Chemical reviews*, *96*(3), 1027–1044, 1996.
- Moretti, R., B. J. Bender, B. Allison, and J. Meiler, Rosetta and the design of ligand binding sites, in *Computational Design of Ligand Binding Proteins*, pp. 47–62, Springer, 2016.
- Nelson, D. L., and M. M. Cox, *Princípios de Bioquímica de Lehninger-7*, Artmed Editora, 2018.
- Neufeld, E. F., Enzyme replacement therapy—a brief history, *Fabry disease: perspectives from 5 years of FOS*, 2006.
- Packer, M. S., and D. R. Liu, Methods for the directed evolution of proteins, *Nature Reviews Genetics*, *16*(7), 379–394, 2015.
- Parr, R. G., and W. Yang, Density functional approach to the frontier-electron theory of chemical reactivity, *Journal of the American Chemical Society*, *106*(14), 4049–4050, 1984.
- Parr, R. G., R. A. Donnelly, M. Levy, and W. E. Palke, Electronegativity: the density functional viewpoint, *The Journal of Chemical Physics*, *68*(8), 3801–3807, 1978.

- Pessôa, K. F., J. C. G. Correia, A. N. M. Carauta, and F. B. d. Silva, Revisão de alguns principais métodos utilizados em modelagem molecular. parte ii-métodos quânticos, 2018.
- Radzicka, A., and R. Wolfenden, A proficient enzyme, *Science*, *267*(5194), 90–93, 1995.
- Ramos, M. A. F. d. C., Química verde: potencialidades e dificuldades da sua introdução no ensino básico e secundário, Ph.D. thesis, 2009.
- Raschka, S., A. J. Wolf, J. Bemister-Buffington, and L. A. Kuhn, Protein–ligand interfaces are polarized: discovery of a strong trend for intermolecular hydrogen bonds to favor donors on the protein side with implications for predicting and designing ligand complexes, *Journal of computer-aided molecular design*, *32*(4), 511–528, 2018.
- Reetz, M. T., Biocatalysis in organic chemistry and biotechnology: past, present, and future, *Journal of the American Chemical Society*, *135*(34), 12,480–12,496, 2013.
- Reetz, M. T., M. Bocola, J. D. Carballeira, D. Zha, and A. Vogel, Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test, *Angewandte Chemie International Edition*, *44*(27), 4192–4196, 2005.
- Richter, F., A. Leaver-Fay, S. D. Khare, S. Bjelic, and D. Baker, De novo enzyme design using rosetta3, *PloS one*, *6*(5), e19,230, 2011.
- Röthlisberger, D., et al., Kemp elimination catalysts by computational enzyme design, *Nature*, *453*(7192), 190–195, 2008.
- Savile, C. K., et al., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture, *Science*, *329*(5989), 305–309, 2010.
- Schober, M., et al., Chiral synthesis of lsd1 inhibitor gsk2879552 enabled by directed evolution of an imine reductase, *Nature Catalysis*, *2*(10), 909–915, 2019.
- Scrutton, N. S., A. Berry, and R. N. Perham, Redesign of the coenzyme specificity of a dehydrogenase by protein engineering, *Nature*, *343*(6253), 38–43, 1990.
- Siegel, J. B., et al., Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction, *Science*, *329*(5989), 309–313, 2010.
- Singh, R., M. Kumar, A. Mittal, and P. K. Mehta, Microbial enzymes: industrial progress in 21st century, *3 Biotech*, *6*(2), 1–15, 2016.
- Stemmer, W. P., Rapid evolution of a protein in vitro by dna shuffling, *Nature*, *370*(6488), 389–391, 1994.

- Stewart, J. J., Mopac: a semiempirical molecular orbital program, *Journal of computer-aided molecular design*, 4(1), 1–103, 1990.
- Stewart, J. J., Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations, *International Journal of Quantum Chemistry*, 58(2), 133–146, 1996.
- Stewart, J. J., Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters, *Journal of molecular modeling*, 19(1), 1–32, 2013.
- Strohmeier, G. A., H. Pichler, O. May, and M. Gruber-Khadjawi, Application of designed enzymes in organic synthesis, *Chemical reviews*, 111(7), 4141–4164, 2011.
- Vaissier Welborn, V., and T. Head-Gordon, Computational design of synthetic enzymes, *Chemical reviews*, 119(11), 6613–6630, 2018.
- Voigt, C. A., C. Martinez, Z.-G. Wang, S. L. Mayo, and F. H. Arnold, Protein building blocks preserved by recombination, *Nature structural biology*, 9(7), 553–558, 2002.
- Warshel, A., P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. Olsson, Electrostatic basis for enzyme catalysis, *Chemical reviews*, 106(8), 3210–3235, 2006.
- Wells, J. A., D. B. Powers, R. R. Bott, T. P. Graycar, and D. A. Estell, Designing substrate specificity by protein engineering of electrostatic interactions, *Proceedings of the National Academy of Sciences*, 84(5), 1219–1223, 1987.
- Zanghellini, A., L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, and D. Baker, New algorithms and an in silico benchmark for computational enzyme design, *Protein Science*, 15(12), 2785–2794, 2006.

Anexo

Tabela 1: Reação de eliminação de Kemp

Reação de eliminação de Kemp	Score Rosetta (Arb.U)	Classificação Rosetta	Entalpia (kcal/mol)	Classificação Entalpia
UM_66_E180W83	-12.35	1	-40.06	1
UM_55_E159W110	-11.34	2	-31.35	3
rep_194.47_UM_40_E133W110	-10.37	3	-21.81	29
UM_86_E184W89	-10.33	4	-29.87	5
rep_205.72_UM_101_E211W89	-10.3	5	-28.80	7
UM_102_E211Y110	-10.22	6	-22.89	25
UM_10_E51W89	-10.06	7	-19.46	36
UM_6_E51W110	-10.00	8	-27.30	10
rep_193.39_UM_101_E211W89	-9.81	9	-11.56	46
UM_79_E180W233	-9.81	10	-23.52	22
rep_206.23_UM_58_E180F210	-9.79	11	-16.42	42
rep_198.07_UM_46_E159W51	-9.77	12	-27.93	8
UM_58_E180F210	-9.41	13	-26.45	12
UM_101_E211W89	-9.21	14	-30.90	4
rep_203.07_UM_43_E133W110	-9.18	15	-22.81	26
UM_118_E231W110	-9.14	16	-29.21	6
rep_184.58_UM_25_E110W211	-9.00	17	-6.47	48
UM_18_E83Y210	-8.79	18	-24.71	16
UM_83_E184W51	-8.79	19	-16.40	43
UM_34_E133Y159	-8.77	20	-18.37	39
UM_25_E110W211	-8.39	21	-26.29	13
UM_75_E180W110	-8.38	22	-24.39	18
UM_115_E211Y159	-8.32	23	-21.22	31
UM_105_E211Y110	-8.31	24	-27.04	11
UM_128_E233Y110	-8.21	25	-24.83	15
UM_22_E89Y51	-8.17	26	-13.44	45
UM_63_E180W110	-8.17	27	-21.82	28
UM_106_E211W89	-8.12	28	-23.82	19
rep_194.74_UM_24_E110Y210	-7.86	29	-16.58	41
UM_43_E133W110	-7.75	30	-27.73	9
UM_23_E110Y210	-7.47	31	-23.81	20
rep_202.63_UM_125_E233F51	-7.41	32	-5.44	49
UM_46_E159W51	-7.37	33	-23.03	24
rep_201.35_UM_126_E233W110	-7.34	34	-34.29	2
UM_17_E83Y210	-7.33	35	-25.58	14
UM_64_E180W51	-7.31	36	-24.50	17
UM_40_E133W110	-7.28	37	-16.71	40
3cst_UM_3_E83Y210R184	-7.28	38	-23.16	23
UM_126_E233W110	-7.27	39	-20.49	33
UM_11_E51Y210	-7.13	40	-3.21	51
UM_85_E184W159	-7.04	41	-20.5	32
rep_197.32_UM_27_E110F210	-6.88	42	-18.7	37
UM_81_E184W51	-6.87	43	-9.38	47
UM_24_E110Y210	-6.84	44	-13.59	44
UM_125_E233F51	-6.77	45	-23.54	21
UM_27_E110F210	-6.76	46	-21.92	27
UM_108_E211W180	-6.54	47	-18.49	38
UM_88_E184F180	-6.42	48	-4.82	50
UM_53_E159Y110	-6.19	49	-20.13	35
UM_35_E133W110	-6.16	50	-21.68	30
Best_UM_25_E231W110	-6.01	51	-20.38	34

Tabela 2: Reação de Diels-Alder

Reação de Diels-Alder	Score Rosetta (Arb.U)	Classificação Rosetta	Entalpia (kcal/mol)	Classificação Entalpia
UM_81_T179Q77	-24.42	1	21.90	50
rep_249.26_UM_64_Y121Q74	-22.95	2	7.32	45
UM_89_T195Q74	-22.65	3	-12.74	25
rep_228.14_UM_64_Y121Q74	-22.61	4	-9.51	33
rep_231.74_UM_68_Y121N74	-22.02	5	-12.06	27
rep_245.99_UM_28_S120Q74	-21.77	6	-29.87	10
rep_243.32_UM_71_Y121Q195	-21.41	7	-10.92	30
UM_77_Y121Q74	-21.31	8	-20.66	16
UM_73_Y121Q195	-21.30	9	-10.96	29
BestHit_rep_222.7_UM_2_Y121Q195	-20.65	10	-38.72	4
rep_235.92_UM_69_Y121N74	-20.39	11	58.96	52
BestHit_rep_225.74_UM_2_Y121Q195	-20.25	12	-32.87	9
BestHit_rep_249.41_UM_2_Y121Q195	-20.02	13	-36.00	5
UM_64_Y121Q74	-19.82	14	7.32	46
UM_88_S195Q74	-19.64	15	5.65	43
BestHit_UM_2_Y121Q195	-19.51	16	-36.00	6
UM_71_Y121Q195	-19.46	17	1.49	41
UM_52_T121Q74	-18.94	18	-43.17	1
UM_69_Y121N74	-18.64	19	5.21	42
UM_94_T195N120	-18.61	20	-5.84	36
UM_68_Y121N74	-18.49	21	0.92	40
UM_35_S120Q74	-18.34	22	-34.74	7
UM_30_S120N74	-18.30	23	-9.95	32
UM_51_S121Q74	-18.23	24	-13.27	21
UM_40_T120Q74	-18.14	25	-7.35	35
UM_93_T195Q120	-18.02	26	-21.21	14
rep_234.27_UM_43_S121Q74	-17.77	27	-4.22	37
rep_240.81_UM_41_T120N74	-17.42	28	8.85	47
rep_234.36_UM_72_Y121Q195	-17.36	29	0.83	39
rep_237.82_UM_72_Y121Q195	-17.23	30	-19.12	19
UM_90_T195Q120	-17.20	31	-39.59	3
UM_38_S120Q74	-17.02	32	6.07	44
rep_246.73_UM_57_S121Q74	-16.85	33	-12.21	26
rep_238.29_UM_72_Y121Q195	-16.62	34	-25.88	11
rep_245.40_UM_37_S120Q74	-16.57	35	-10.15	31
rep_244.99_UM_56_S121N74	-16.57	36	-19.16	17
UM_43_S121Q74	-16.43	37	-24.98	12
UM_32_S120Q74	-16.26	38	-14.95	20
UM_41_T120N74	-16.24	39	0.30	38
rep_233.44_UM_57_S121Q74	-16.11	40	-13.21	23
UM_45_S121Q74	-16.04	41	-21.16	15
rep_243.22_UM_72_Y121Q195	-15.15	42	-39.84	2
UM_15_Y75Q120	-14.97	43	14.20	49
UM_25_S120Q74	-14.83	44	37.91	51
UM_72_Y121Q195	-14.55	45	-19.16	18
UM_34_S120N74	-14.01	46	-9.16	34
UM_57_S121Q74	-13.71	47	-13.21	22
UM_12_Y74N195	-13.10	48	-13.19	24
UM_37_S120Q74	-12.92	49	-34.18	8
UM_28_S120Q74	-12.68	50	12.49	48
UM_56_S121N74	-11.07	51	-22.87	13
UM_42_T120N74	-10.85	52	-11.98	28