



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE QUÍMICA

FELIPE DIAS LEAL



**QSAR 2D/3D E DOCAGEM DE INIBIDORES DA ENZIMA
DYRK1A COMO POTENCIAIS FÁRMACOS CONTRA A
DOENÇA DE ALZHEIMER**

RIO DE JANEIRO
2015

Felipe Dias Leal

QSAR 2D/3D E DOCAGEM DE INIBIDORES DA
ENZIMA DYRK1A COMO POTENCIAIS
FÁRMACOS CONTRA A DOENÇA DE
ALZHEIMER

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Química do Instituto de Química da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Ciências.

Orientadores:
Profa. Dra. Magaly Girão Albuquerque
Prof. Dr. Carlos Rangel Rodrigues

Rio de Janeiro
2015

L435

Leal, Felipe Dias.

QSAR 2D/3D e docagem de inibidores da enzima DYRK1A como potenciais fármacos contra a doença de Alzheimer / Felipe Dias Leal – Rio de Janeiro : UFRJ, 2015.

159 f.

Tese (Doutorado) – Universidade Federal do Rio de Janeiro, Instituto de Química, 2015.

Orientadores: Magaly Girão Albuquerque e Carlos Rangel Rodrigues

1. QSAR 2D/3D. 2. Docagem molecular. 3. Doença de Alzheimer. 4. 6- arilquinazolina-4-aminas. 5. Química medicinal. I. Albuquerque, Magaly Girão.II. Rodrigues, Carlos Rangel. III. Universidade Federal do Rio de Janeiro. Instituto de Química. Programa de Pós-Graduação em Química. IV. Título.

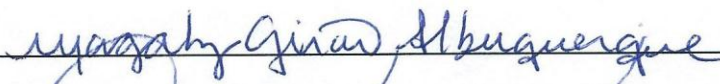
CDD 615

FELIPE DIAS LEAL


**QSAR 2D/3D E DOCAGEM MOLECULAR DE INIBIDORES DA ENZIMA
DYRK1A COMO POTENCIAIS FÁRMACOS CONTRA A DOENÇA DE
ALZHEIMER**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Química do Instituto de Química da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Ciências.

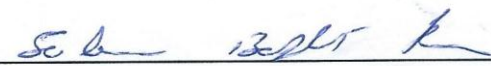
Rio de Janeiro, 28 de agosto de 2015.



Profa. Dra. Magaly Girão Albuquerque (IQ-UFRJ) (Orientador)



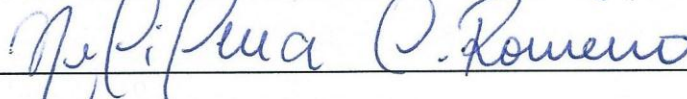
Prof. Dr. Carlos Rangel Rodrigues (FF-UFRJ) (Orientador)



Profa. Dra. Sabrina Baptista Ferreira (IQ-UFRJ)



Prof. Dr. Carlos Maurício Rabello de Sant'Anna (UFRRJ)



Profa. Dra. Nelilma Correia Romeiro (UFRJ)



Prof. Dr. Floriano Paes Silva Júnior (FIOCRUZ)

AGRADECIMENTOS

Definitivamente, não somos nada sozinhos. Quando me recordo dos anos de doutorado, percebo quanta gente contribuiu para que eu pudesse apresentar esta tese. São tantas pessoas que até me preocupa não lembrar todas.

Primeiramente, agradeço a Deus, meu Senhor e Salvador. Só a permissão Dele fez com que isto fosse possível. Espero ser capaz de retornar aos meus irmãos na sociedade os frutos desta titulação, pois sei que Ele conta com isso.

Em seguida, aos meus pais, Luis Fernando e Alda Maria, os primeiros incentivadores da minha carreira acadêmica e meus mestres, tanto da vida como da arte de ensinar. À minha esposa, Ana Carolina, que tanto me apoiou e que compreendeu minhas ausências em função da dedicação que um curso de doutorado exige. Aos meus avós, Luiz e Alda (*in memoriam*), que tanto zelo tiveram comigo durante este período e, também, à minha irmã, Fernanda. Realmente, nenhuma palavra poderá expressar minha gratidão. Meu amor por cada um de vocês será eterno.

À minha orientadora, Magaly, primeiramente, por aceitar o desafio de orientar um aluno oriundo de outro laboratório, já com quatro anos de doutorado e tendo que recomeçar uma tese do zero e também por me ensinar o que é ser um orientador de verdade. Este belíssimo exemplo levo comigo como um dos principais aprendizados deste curso. Estendo este agradecimento ao meu co-orientador, Prof. Carlos e à Profa. Helena, pelas contribuições dadas ao meu trabalho.

Às minhas chefes da Farmácia Universitária, Professoras Rita e Elisabete, e do INTO, Cláudia e Zilda, pela flexibilização de horários que me permitiram desempenhar essa difícil tarefa de trabalhar e concluir uma pós-graduação de excelência.

Ao Prof. Ricardo Bicca e aos colegas de LabMMol, Eugênio e Daniel, pela convivência sempre cordial e, em especial, ao Camilo, meu amigo e companheiro de trabalho no laboratório aos fins de semana, pelos ensinamentos e discussões científicas sempre construtivas.

Aos meus amigos de LabDop, Bruno, Vinícius, Fernanda, Gustavo e Marco, com os quais convivi boa parte do curso de doutorado e trouxeram grandes alegrias em tempos de tribulação.

Aos meus amigos Marcelo e Cristiano que, por também saberem como é cursar um doutorado com tantas outras obrigações, sempre me brindaram com palavras de incentivo durante todo o percurso.

Enfim, aos meus professores da pós-graduação, por todo o conhecimento adquirido e aos componentes da banca por todas as contribuições dadas a este trabalho.

“Comece fazendo o necessário, depois, o que é possível e, de repente, você estará fazendo o impossível.”

São Francisco de Assis

RESUMO

LEAL, Felipe Dias. QSAR 2D/3D e docagem de inibidores da enzima DYRK1A como potenciais fármacos contra a doença de Alzheimer. Rio de Janeiro, 2015. Tese (Doutorado em Química) – Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015

Este trabalho descreve estudos de modelagem molecular aplicados a uma série congênere de 46 derivados da classe das 6-arylquinazolina-4-aminas, sintetizados e avaliados farmacologicamente como inibidores da quinase-1A de especificidade dual regulada via fosforilação da tirosina (DYRK1A), uma enzima que participa ativamente de mecanismos de desenvolvimento da doença de Alzheimer, sendo um potencial alvo terapêutico. Primeiramente, foram realizados estudos de relação quantitativa estrutura-atividade em duas e três dimensões (QSAR 2D/3D), os quais permitiram a construção de modelos capazes de prever atividades biológicas de compostos ainda não sintetizados e indicar grupos funcionais na estrutura dos ligantes que sejam responsáveis pelo aumento ou redução da atividade biológica dos mesmos. No estudo de QSAR 2D, foi aplicado o método de holograma QSAR (HQSAR) e, no estudo de QSAR 3D, foram aplicados os métodos de análise comparativa de campos moleculares (CoMFA) e análise comparativa de índices de similaridade moleculares (CoMSIA). Os modelos foram analisados e todos apresentaram parâmetros estatísticos aceitáveis, resultando em modelos robustos e com boa capacidade preditiva (HQSAR: $q^2 = 0,757$; CoMFA: $q^2 = 0,608$; CoMSIA: $q^2 = 0,713$). Em seguida, foram realizados estudos de docagem molecular utilizando o programa Molegro Virtual Docker. Este método permitiu um melhor entendimento no modo de ligação dos derivados estudados à enzima alvo e, também, à corroboração de hipóteses desenvolvidas nos estudos de QSAR. Estes estudos permitiram a proposta de quatro novos análogos desta série a serem avaliados como inibidores da DYRK1A.

Palavras-chave: QSAR 2D/3D, docagem molecular, 6-arylquinazolina-4-aminas, doença de Alzheimer

ABSTRACT

LEAL, Felipe Dias. QSAR 2D/3D e docagem de inibidores da enzima DYRK1A como potenciais fármacos contra a doença de Alzheimer. Rio de Janeiro, 2015. Tese (Doutorado em Química) – Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015

Molecular modeling studies were applied to a series of 46 6-arylquinazolin-4-amine derivatives, which were synthesized and pharmacologically evaluated as dual specificity tyrosine-phosphorylation-regulated kinase-1A (DYRK1A) inhibitors. The DYRK1A enzyme is actively involved in mechanisms of the Alzheimer's disease development, thus being a potential therapeutic target for its treatment. Firstly, it were carried out two and three dimensional quantitative structure-activity relationship (2D/3D QSAR) studies, which allowed the development of models capable of predicting biological activities of non-synthesized compounds and indicating functional groups of the inhibitors responsible for increasing or decreasing their biological activities. In the 2D QSAR study, it was applied the hologram QSAR (HQSAR) method and in the 3D QSAR study, it were applied the Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Index Analysis (CoMSIA) methods. The models were statistically analysed and all of them presented acceptable statistic parameters, resulting in statistically robust models and with good predictive capability (HQSAR: $q^2 = 0.757$; CoMFA: $q^2 = 0.608$; CoMSIA: $q^2 = 0.713$). Subsequently, it were performed molecular docking studies using the Molegro Virtual Docker program. This approach allowed a better understanding about the binding mode between the inhibitors and the target enzyme, supporting the hypotheses developed in the 2D/3D QSAR approach. These studies enabled the proposal of four new analogs of this series to be evaluated as DYRK1A inhibitors.

Keywords: 2D/3D-QSAR, molecular docking, 6-arylquinazolin-4-amines, Alzheimer's disease

LISTA DE FIGURAS

- Figura 1.** Comparação entre cortes cerebrais de um cérebro de um paciente com DA e um cérebro normal (Adaptado de HOLTZMAN; MORRIS; GOATE, 2013). 29
- Figura 2.** Fotomicrografia de uma placa amiloide (PA) e de um emaranhado neurofibrilar (ENF) (Adaptado de HOLTZMAN; MORRIS; GOATE, 2011). 29
- Figura 3.** Estrutura tridimensional em modelo de fita da DYRK1A humana e detalhe do sítio de autofosforilação composto por um resíduo de tirosina fosforilado (O³POTyr321 = pTyr321) e dois resíduos de arginina (Arg325 e Arg328) (Adaptado de OGAWA et al., 2010). 34
- Figura 4.** Sítio de ligação do INDY (inibidor de DYRK) à DYRK1A humana (Adaptado de OGAWA et al., 2010). 35
- Figura 5.** Sítio de ligação dos derivados 3 (A) e 32 (B) da classe das pirido[2,3-d]pirimidinas à DYRK1A humana (Adaptado de ANDERSON et al., 2013). 36
- Figura 6.** Contribuição da superexpressão da DYRK1A para a β -amiloidose e degeneração neurofibrilar em doenças neurodegenerativas. ASF – Fator Alternativo de *Splicing*; PPA – Proteína Precursora Amiloide; ENFs – Emaranhados Neurofibrilares (Adaptado de WEGIEL; GONG; HWANG, 2011). 38
- Figura 7.** Estrutura química da harmina. 41
- Figura 8.** Estrutura química do galato de epigalocatequina (epigalocatequina-3-galato, EGCG). 42
- Figura 9.** Estruturas químicas dos compostos INDY e TG003. 43

Figura 10. Estrutura geral do composto mais ativo da série de derivados pirido[2,3- <i>d</i>]pirimidinas.	43
Figura 11. Estrutura geral da série de derivados 6-arilquinazolina-4-aminas.	44
Figura 12. Esquema geral do desenvolvimento de um modelo de QSAR (Adaptado de MYINT; XIE, 2011).	48
Figura 13. Esquema geral da geração de hologramas moleculares e modelos de HQSAR (Adaptado de MYINT; XIE, 2011).	51
Figura 14. Representação de um mapa de contribuição de um ligante da série de ácidos 2-(oxalilamino) benzoicos (Adaptado de CHENG et al., 2010).	52
Figura 15. Diagrama representando os potenciais eletrostático (Coulomb) e de van der Waals (Lennard-Jones) em função da distância interatômica (Adaptado de KUBINYI, 2003).	55
Figura 16. Esquema geral da análise por CoMFA (Adaptado de MYINT; XIE, 2011).	56
Figura 17. Gráfico da distribuição dos valores de atividade biológica (pIC_{50}) dos 36 compostos (Rosenthal et al. 2010; 2011) do grupo de treinamento.	70
Figura 18. Gráfico da distribuição dos valores de atividade biológica (pIC_{50}) dos 10 compostos (Rosenthal et al. 2010; 2011) do grupo de teste.	71
Figura 19. Modelo de ligação do inibidor DJM2005 à DYRK1A humana (Adaptado de SOUNDARARAJAN et al., 2013).	74

- Figura 20.** Estruturas químicas dos compostos 24 e DJM2005 usados na definição da hipótese farmacofórica. 74
- Figura 21.** Gráfico dos valores de pIC_{50} preditos *versus* os valores de pIC_{50} experimentais para os grupos de treinamento (azul) e de teste (vermelho), utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **A/L/C/Q/DA**. 90
- Figura 22.** Gráfico dos valores de pIC_{50} preditos *versus* os valores de pIC_{50} experimentais para os grupos de treinamento (azul) e de teste (vermelho), utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **A/L**. 91
- Figura 23.** Mapas de contribuição para os compostos mais ativo (**24**) e menos ativo (**6**), segundo os melhores modelos **A/L/C/Q/DA** (A) e **A/L** (B). As estruturas dos compostos estão representadas em modelo bastão e apenas os heteroátomos estão representados pela letra do elemento químico (Cl, N, O, S). O espectro do amarelo ao verde indica que o fragmento tem contribuição positiva para a atividade biológica; o espectro do laranja ao vermelho indica que o fragmento tem contribuição positiva para a atividade biológica; a cor branca indica que o fragmento tem contribuição neutra para a atividade biológica e a cor azul representa a estrutura comum a todos os compostos da série. 93
- Figura 24.** (A) Estrutura comum calculada pelo método *distill rigid* para a série das 6-arilquinazolina-4-aminas. (B) Sobreposição dos 46 derivados. 96
- Figura 25.** Gráfico com os valores de pIC_{50} preditos *versus* os valores de pIC_{50} experimentais para os grupos de treinamento (azul) e de teste (vermelho), utilizando o melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å). 103

Figura 26. Mapas de contorno do melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å) obtido para o composto **24** (modelo em bastão e colorido por elementos). (A) O mapa estérico indica as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a atividade biológica. (B) O mapa eletrostático indica áreas onde grupos com grande densidade eletrônica ou carga negativa aumentam (vermelho) ou diminuem (azul) a atividade biológica. 106

Figura 27. Mapa de contorno estérico do melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å) obtido para o composto **46** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a atividade biológica. 108

Figura 28. Mapa de contorno eletrostático do melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å) obtido para o composto **6** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos com alta densidade eletrônica aumentam (vermelho) ou diminuem (azul) a atividade biológica. 109

Figura 29. Gráfico com os valores de pIC₅₀ preditos *versus* os valores de pIC₅₀ experimentais para os grupos de treinamento (azul) e de teste (vermelho) utilizando o melhor modelo de CoMSIA. 116

Figura 30. Mapa de contorno eletrostático do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **24** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos com alta densidade eletrônica aumentam (vermelho) ou diminuem (azul) a atividade biológica. 118

Figura 31. Estrutura química do composto **24** com as cargas parciais calculadas pelo método AM1 destacadas para os átomos de oxigênio e carbono do anel 1,3-benzo-dioxola. 118

Figura 32. Mapa de contorno hidrofóbico do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **24** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos hidrofóbicos aumentam (amarelo) ou diminuem (roxo) a atividade biológica. 119

Figura 33. Mapa de contorno hidrofóbico do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **46** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos hidrofóbicos aumentam (amarelo) ou diminuem (roxo) a atividade biológica. 120

Figura 34. Mapa de contorno hidrofóbico do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **6** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos hidrofóbicos aumentam (amarelo) ou diminuem (roxo) a atividade biológica. 121

Figura 35. Mapa de contorno doador de ligação hidrogênio do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **24** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos doadores de ligação hidrogênio aumentam (azul claro) ou diminuem (preto) a atividade biológica. 122

Figura 36. Interações intermoleculares (linha pontilhada) entre o anel quinazolina da série de derivados 6-arylquinazolina-4-aminas, representada pelo enantiômero *R* do composto **34**, e os principais resíduos de aminoácidos da DYRK1A. A interação via ligação hidrogênio está representada pela cor verde e as interações hidrofóbicas pela cor rosa. 126

Figura 37. Mapa de superfície da DYRK1A construído no programa *Discovery Studio Visualizer* versão 4.1 (*Accelrys Software Inc*), indicando áreas da proteína contendo grupos doadores e aceptores de ligação hidrogênio, usando o enantiômero *R* do composto **34** como exemplo de ligante. 127

Figura 38. Mapa de superfície da DYRK1A construído no programa *Discovery Studio Visualizer* versão 4.1 (*Accelrys Software Inc*), indicando áreas hidrofóbicas e hidrofílicas com o composto **34** como exemplo de ligante. 128

Figura 39. Interações intermoleculares (linha pontilhada) entre o composto **24** (mais ativo) e os principais resíduos de aminoácidos da DYRK1A. A ligação hidrogênio está representada pela cor verde, as interações do tipo dipolo-dipolo estão representadas pela cor azul e as interações hidrofóbicas pela cor rosa.

Figura 40. Interações intermoleculares (linha pontilhada) entre o composto **6** (menos ativo) e os principais resíduos de aminoácidos da DYRK1A. As interações do tipo dipolo-dipolo estão representadas pela cor azul e as interações hidrofóbicas pela cor rosa. 130

Figura 41. Estruturas químicas das soluções obtidas para o composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) durante o processo de docagem molecular. Solução 1 é a melhor obtida segundo o *PLANTS Score* e a Solução 2 é a melhor obtida segundo o *Rerank Score*. 131

Figura 42. Interações intermoleculares (linha pontilhada) entre a solução **1** (*PLANTS Score*) do composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) e os principais resíduos de aminoácidos da DYRK1A. As interações do tipo dipolo-dipolo estão representadas pela cor azul, as interações hidrofóbicas pela cor rosa e a interação com o átomo de enxofre pela cor amarela. 133

Figura 43. Interações intermoleculares (linha pontilhada) entre a solução **2** (*Rerank Score*) do composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) e os principais resíduos de aminoácidos da DYRK1A. As interações hidrofóbicas pela cor rosa e a interação desfavorável entre grupos doadores de ligação hidrogênio pela cor vermelha. 134

Figura 44. Interações intermoleculares (linha pontilhada) entre o anel tetraidrofurano presente na região R₃ dos enantiômeros *R* (A) e *S* (B) do composto **32** e um resíduo de aminoácido da DYRK1A. As interações hidrofóbicas são representadas pela cor rosa. 135

Figura 45. Interações intermoleculares (linha pontilhada) entre o anel tetraidropirano presente na região R₃ dos enantiômeros *R* (A) e *S* (B) do composto **33** e resíduos de aminoácidos da DYRK1A. As interações hidrofóbicas são representadas pela cor rosa e as interações dipolo-dipolo pela cor azul. 136

Figura 46. Interações intermoleculares (linha pontilhada) entre o anel tetraidrofurano presente na região R₃ dos enantiômeros *R* (A) e *S* (B) do composto **34** e um resíduo de aminoácido da DYRK1A. As interações hidrofóbicas são representadas pela cor rosa. 136

Figura 47. Estruturas químicas dos compostos propostos **A1** e **A2** e dos seus respectivos análogos. (A) Composto **A1** é análogo ao composto **36**. (B) Composto **A2** é análogo ao composto **24**. (C) Composto **36**. (D) Composto **24**. 138

Figura 48. Estruturas químicas dos compostos propostos **A3** e **A4** e dos seus respectivos análogos. (A) Composto **A3** é análogo ao enantiômero (*R*) do composto **34**. (B) Composto **A4** é análogo ao enantiômero (*S*) do composto **34**. (C) Enantiômero (*R*) do composto **34**. (D) Enantiômero (*S*) do composto **34**. 139

LISTA DE TABELAS

- Tabela 1.** Parâmetros de distinção dos fragmentos moleculares empregados no método HQSAR. 50
- Tabela 2.** Estruturas químicas dos 46 compostos da classe das 6-arylquinazolina-4-aminas e seus respectivos valores de IC₅₀ (nM) e pIC₅₀. Os compostos do conjunto de teste estão marcados com um asterisco. 64
- Tabela 3.** Resumo dos índices estatísticos dos modelos de HQSAR para as diversas combinações de distinção de fragmentos (DF), utilizando o tamanho padrão de fragmentos (4-7 átomos) para os derivados da série das 6-arylquinazolina-4-aminas (N=36). Os quatro melhores modelos estão assinalados em negrito. 83
- Tabela 4.** Resumo dos índices estatísticos dos modelos de HQSAR para as diversas combinações de tamanho de fragmentos (TF), utilizando o parâmetro de distinção de fragmentos **A/L/C/Q/DA** para os derivados da série das 6-arylquinazolina-4-aminas (N=36). O melhor modelo está assinalado em negrito. 84
- Tabela 5.** Resumo dos índices estatísticos dos modelos de HQSAR para as diversas combinações de tamanho de fragmentos (TF), utilizando o parâmetro de distinção de fragmentos **A/L** para os derivados da série das 6-arylquinazolina-4-aminas (N=36). O melhor modelo está assinalado em negrito. 85
- Tabela 6.** Resumo dos índices estatísticos dos modelos de HQSAR do teste de randomização-Y utilizando o tamanho padrão de fragmentos (4-7 átomos) para os derivados da série das 6-arylquinazolina-4-aminas (N=36). 86

- Tabela 7.** Valores da atividade biológica experimental (Exp), predita (Pred) e resíduos (Res) dos compostos da classe das 6-irilquinazolina-4-aminas, utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **AL/C/Q/DA**. Os compostos do conjunto de teste estão marcados com um asterisco. 88
- Tabela 8.** Valores da atividade biológica experimental (Exp), predita (Pred) e resíduos (Res) dos compostos da classe das 6-irilquinazolina-4-aminas, utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **AL**. Os compostos do conjunto de teste estão marcados com um asterisco. 89
- Tabela 9.** Índices estatísticos dos modelos de CoMFA, obtidos para os derivados da série das 6-irilquinazolina-4-aminas em função do método de cálculo das cargas parciais (N=36). 97
- Tabela 10.** Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-irilquinazolina-4-aminas em função do átomo de prova (N=36). Método de cálculo de cargas parciais (AM1), valor de corte de energia (30 kcal.mol⁻¹), espaçamento da grade (2,0 Å). 98
- Tabela 11.** Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-irilquinazolina-4-aminas em função do nível de corte de energia (N=36). Método de cálculo de cargas parciais (AM1), átomo de prova (H +1), espaçamento da grade (2,0 Å). 99
- Tabela 12.** Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-irilquinazolina-4-aminas em função do espaçamento da grade (N=36). Método de cálculo de cargas parciais (AM1), átomo de prova (H +1), nível de corte de energia (30 kcal.mol⁻¹). 100

- Tabela 13.** Valores das atividades biológicas experimentais (Exp), preditas (Pred) e resíduos (Res) dos compostos da classe das 6-irilquinazolina-4-aminas utilizando o melhor modelo de CoMFA. Átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å. Os compostos do conjunto de teste estão marcados com um asterisco. 102
- Tabela 14.** Índices estatísticos obtidos no teste de randomização-Y para o melhor modelo de CoMFA: átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å. 104
- Tabela 15.** Índices estatísticos dos modelos de CoMSIA obtidos para os derivados da série das 6-irilquinazolina-4-aminas (N=36) em função das diversas combinações de campos moleculares (S, E, H, D, A). 111
- Tabela 16.** Índices estatísticos dos modelos de CoMSIA obtidos para os derivados da série das 6-irilquinazolina-4-aminas (N=36) em função do fator de atenuação (α) contendo os campos moleculares eletrostático (E), hidrofóbico (H) e doador de ligação hidrogênio (D). 112
- Tabela 17.** Valores de atividade biológica (pIC₅₀, M) experimental (Exp), predita (Pred) e resíduos (Exp – Pred) dos compostos da classe das 6-irilquinazolina-4-aminas utilizando o melhor modelo de CoMSIA (campos moleculares E, H, D e fator de atenuação $\alpha = 0,2$). Os compostos do conjunto de teste estão marcados com um asterisco. 114
- Tabela 18.** Índices estatísticos obtidos nas quatro randomizações da atividade biológica no teste de randomização-Y para o melhor modelo de CoMSIA. 117
- Tabela 19.** Resumo das estruturas tridimensionais da DYRK1A resolvidas por 123

difração de raios-X do cristal e seus respectivos ligantes disponíveis no PDB.

Tabela 20. Número de soluções com valor de RMSD menor do que 2,0 Å, considerando os complexos 2WO6 e 4MQ1 e as combinações de algoritmo de busca e função de pontuação. 124

Tabela 21. Energias internas e de complexação ligante-enzima para as duas melhores soluções encontradas para o composto **29** (*outlier* nos estudos de CoMFA/CoMSIA). 132

Tabela 22. Energias de complexação ligante-enzima para os pares de enantiômeros dos compostos **32**, **33** e **34**. 135

Tabela 23. Substituintes dos derivados da série das 6-arylquinazolina-4-aminas que geram melhor perfil de atividade inibitória contra a enzima DYRK1A, observados nos estudos de QSAR e docagem molecular. 137

Tabela 24. Valores de pIC_{50} preditos para os compostos propostos (**A1**, **A2**, **A3** e **A4**) e seus análogos correspondentes (**36**, **24**, **34 (R)** e **34 (S)**, respectivamente) segundo os melhores modelos de QSAR obtidos. 139

LISTA DE ABREVIATURAS E SIGLAS

AB	atividade biológica
ASF	<i>alternative splicing factor</i> (fator de <i>splicing</i> alternativo)
ATP	<i>adenosine triphosphate</i> (trifosfato de adenosina)
CDK	<i>cyclin-dependent kinase</i> (quinase dependente de ciclina)
CH	comprimento do holograma
CK	<i>creatine kinase</i> (creatina quinase)
CLK	<i>CDK-1 like kinase</i> (quinase semelhante à CDK-1)
CoMFA	<i>Comparative Molecular Field Analysis</i> (Análise Comparativa de Campos Moleculares)
CoMSIA	<i>Comparative Molecular Similarity Index Analysis</i> (Análise Comparativa de Índices de Similaridade Moleculares)
DA	doença de Alzheimer
DF	distinção de Fragmentos
DFT	<i>density functional theory</i> - teoria funcional de densidade
DNA	ácido desoxirribonucleico
DMAT	2-dimetilamino-4,5,6,7-tetrabromo-1 <i>H</i> -benzimidazol
DP	desvio padrão
DYRK1A	<i>Dual-Specificity Regulated-Tyrosine Kinase 1A</i> (quinase-1A de especificidade dual regulada via fosforilação da tirosina)
ECGC	epigallocatequina-3-galato
EP	erro padrão da estimativa
EP_{vc}	erro padrão da validação cruzada
ENF	emaranhados neurofibrilares
GSK	<i>glycogen synthase kinase</i> (glicogênio sintase quinase)
HQSAR	Holograma QSAR
LOO_{cv}	<i>leave-one-out cross validation</i> (validação cruzada por exclusão individual exaustiva)
MAO	monoamina oxidase
MAPK	proteína quinase ativada por mitógeno
MVD	<i>Molegro Virtual Docker</i>
NC	número de componentes principais

NFATc1	<i>Nuclear Factor of Activated T-Cells 1</i> (Fator nuclear ativador de células-T)
NINCDS-ADRDA	<i>National Institute for Communicative Disorders and Stroke – Alzheimer’s Disease and Related Disorders Association</i> (Instituto Nacional para Desordens Comunicativas e Derrame – Associação Doença de Alzheimer e Desordens Relacionadas)
NMDA	N-metil-D-aspartato
QSAR	<i>quantitative structure-activity relationship</i> (relação quantitativa estrutura-atividade)
PA	placas amiloides
PCA	<i>principal component analysis</i> (análise por componentes principais)
PLP	<i>piecewise linear potential</i> (potencial linear definido por partes)
PDB	<i>Protein Data Bank</i> (Banco de Dados de Proteínas)
PLS	<i>partial least squares</i> (mínimos quadrados parciais)
PPA	proteína precursora amiloide
RCAN1	regulador da Calcineurina 1
RESP	<i>restrained molecular electrostatic potential</i> (potencial eletrostático molecular restrito)
RMN	ressonância magnética nuclear
RMSD	<i>root-mean-square deviation</i> (raiz do desvio médio quadrático)
RNA	ácido ribonucleico
SE	<i>simplex evolution</i> (evolução <i>symplex</i>)
SLN	<i>Sybyl line notation</i> (Notação em linha Sybyl)
TBB	4,5,6,7-tetrabromo-1H-benzotriazol
TF	tamanho de fragmentos

SUMÁRIO

1	INTRODUÇÃO	27
1.1	DOENÇA DE ALZHEIMER	27
1.1.1	Aspectos Históricos	27
1.1.2	Aspectos Clínicos	27
1.1.3	Processo Patológico	28
1.1.4	Aspectos Genéticos	30
1.1.5	Diagnóstico	31
1.1.6	Aspectos Epidemiológicos	31
1.1.7	Tratamento	32
1.2	ENZIMA DYRK1A COMO UM NOVO ALVO CONTRA A DOENÇA DE ALZHEIMER	33
1.2.1	Quinase 1A de especificidade dual regulada via fosforilação de tirosina	33
1.2.2	Estrutura Tridimensional da DYRK1A	34
1.2.3	Papel da DYRK1A no Desenvolvimento da Doença de Alzheimer	37
1.2.4	Ativação da DYRK1A	39
1.2.5	Regulação da DYRK1A	40
1.2.6	Inibição da DYRK1A Humana	40
1.2.6.1	Inibidores da DYRK1A Humana	41
1.3	MODELAGEM MOLECULAR E PLANEJAMENTO DE NOVOS FÁRMACOS	45
1.3.1	QSAR	46
1.3.1.1	QSAR 2D: Holograma QSAR (HQSAR)	49
1.3.1.2	QSAR 3D: CoMFA e CoMSIA	53
1.3.2	Docagem Molecular	58

2	OBJETIVOS	63
3	METODOLOGIA	64
3.1	CONJUNTO DE DADOS ESTRUTURAL E BIOLÓGICO	64
3.1.1	Conjuntos de Treinamento e Teste	70
3.2	HQSAR	71
3.2.1	Construção das Estruturas dos Ligantes	71
3.2.2	Obtenção dos Modelos de HQSAR	71
3.2.3	Teste de Randomização da Variável Y	72
3.2.4	Validação Externa do Modelo Detecção de <i>Outliers</i>	72
3.3	CoMFA e CoMSIA	73
3.3.1	Construção e Otimização das Estruturas Tridimensionais dos Ligantes	73
3.3.2	Definição da Hipótese Farmacofórica	73
3.3.3	Cálculo das Cargas Atômicas Parciais	74
3.3.4	Alinhamento da Série de Compostos	75
3.3.5	Definição dos Átomos de Prova	75
3.3.6	Definição dos Valores de Corte dos Campos Moleculares	76
3.3.7	Definição do Espaçamento da Grade	76
3.3.8	Obtenção dos Modelos de CoMFA	76
3.3.9	Teste de Randomização da Variável Y do Melhor Modelo de CoMFA	77
3.3.10	Validação Externa do Modelo de CoMFA e Detecção de <i>Outliers</i>	77
3.3.11	Obtenção dos Modelos de CoMSIA	77
3.3.12	Teste de Randomização da Variável Y do Melhor Modelo de CoMSIA	78
3.3.13	Validação Externa do Modelo de CoMSIA e Detecção de <i>Outliers</i>	78
3.4	DOCAGEM MOLECULAR	78

3.4.1	Preparo dos Ligantes e da Proteína	78
3.4.2	Validação do Protocolo de Docagem (Re-docagem)	79
3.4.3	Docagem da Série de 6-irilquinazolina-4-aminas	80
4	RESULTADOS E DISCUSSÃO	82
4.1	HQSAR	82
4.1.1	Avaliação do Parâmetro Distinção de Fragmentos na Qualidade dos Modelos de HQSAR	82
4.1.2	Avaliação do Tamanho dos Fragmentos na Qualidade dos Modelos de HQSAR	83
4.1.3	Teste de Randomização-Y	85
4.1.4	Validação Externa do Modelo e Detecção de <i>Outliers</i>	87
4.1.5	Análise dos Mapas de Contribuição	92
4.2	CoMFA e CoMSIA	95
4.2.1	Análise da Sobreposição dos Compostos	95
4.2.2	Avaliação do Melhor Método de Cálculo das Cargas Parciais	96
4.2.3	Avaliação do Melhor Átomo de Prova	98
4.2.4	Avaliação do Melhor Nível de Corte de Energia	99
4.2.5	Avaliação do Melhor Espaçamento de Grade	100
4.2.6	Validação Externa do Melhor Modelo de CoMFA e Detecção de <i>Outliers</i>	101
4.2.7	Avaliação do Teste de Randomização-Y para o Melhor Modelo de CoMFA	104
4.2.8	Análise dos Mapas de Contorno	105
4.2.9	Avaliação do Melhor Modelo de CoMSIA	110
4.2.10	Validação Externa do Melhor Modelo de CoMSIA e Detecção de <i>Outliers</i>	113
4.2.11	Avaliação do Teste de Randomização-Y para o Melhor Modelo de CoMSIA	116

4.2.12	Análise dos Mapas de Contorno	117
4.3	DOCAGEM MOLECULAR	122
4.3.1	Validação do Protocolo de Docagem (Re-docagem)	122
4.3.2	Docagem da Série 6-arylquinazolina-4-aminas	124
4.4	PROPOSTA DE NOVOS COMPOSTOS	137
5	CONCLUSÕES	141
6	PERSPECTIVAS	142
	REFERÊNCIAS	
	ANEXO	

1 INTRODUÇÃO

1.1 DOENÇA DE ALZHEIMER

1.1.1 Aspectos Históricos

Em 1906, o psiquiatra alemão Alois Alzheimer apresentou, durante um congresso em Tübingen, na Alemanha, os resultados de seu estudo clínico em uma paciente que apresentava um caso peculiar de demência aos 51 anos de idade. Alzheimer correlacionou os aspectos cognitivos e comportamentais desta paciente, após sua morte aos 55 anos, com achados histopatológicos do seu córtex cerebral (ALZHEIMER et al., 1995). Nos anos seguintes, diversos pesquisadores estudaram pacientes em condições clínicas semelhantes, aumentando o entendimento sobre a patologia. O termo “Doença de Alzheimer” (DA) foi primeiro mencionado por Emil Kraepelin, em 1910, na oitava edição do seu livro *Handbook of Psychiatry*, onde ele relata as severas alterações celulares descobertas por Alzheimer (MAURER; VOLK; GERBALDO, 1997).

1.1.2 Aspectos Clínicos

A demência é uma síndrome caracterizada por perda ou diminuição da memória e outras habilidades cognitivas, e representa o declínio do nível intelectual de uma pessoa, interferindo negativamente na capacidade de execução de tarefas do seu cotidiano. A DA é uma doença neurodegenerativa, sendo a causa mais comum de demência, responsável por 60 a 70% dos casos (BARKER et al., 2002).

O aspecto essencial da DA é o declínio das habilidades cognitivas, que possui um início insidioso ao longo de vários meses, com subsequente progressão gradual e irreversível de etapas sucessivas de demência. O tempo médio desta patologia é de 7 a 10 anos e, inevitavelmente, culmina com a morte. O sintoma inicial da DA é a perda da memória recente, mas outros sintomas também são observados, como alterações no estado de atenção e na capacidade de resolver problemas. Com a progressão da

demência, sintomas como desorientação visuoespacial, disfunções de linguagem, perda de noção e alterações na personalidade aparecem com frequência (HOLTZMAN; MORRIS; GOATE, 2011).

No estágio inicial da DA, que compreende um período de 2 a 5 anos, os pacientes apresentam aparente normalidade, sendo ainda independentes e capazes de manter atividades cotidianas; embora pessoas próximas sejam capazes de identificar um declínio nos níveis das funções cognitivas do indivíduo. No estágio moderado da DA, que se estende de 2 a 4 anos, os pacientes apresentam um distúrbio de memória mais pronunciado, incluindo a perda de memória de longo prazo; maior dificuldade de realizar tarefas cotidianas, como banhar-se, vestir-se ou cuidar-se, as quais necessitam de supervisão; e dificuldade de socialização. No estágio avançado, os pacientes passam a ser totalmente dependentes para todas as atividades cotidianas e podem tornar-se mudos, incapazes de se locomover e de controlar a função de órgãos como bexiga e intestino (HOLTZMAN; MORRIS; GOATE, 2011).

1.1.3 Processo Patológico

Em nível macroscópico, analisando e comparando cortes cerebrais de pacientes com DA e normais, é possível observar em pacientes com DA uma grave atrofia cerebral (Figura 1); enquanto que, em nível microscópico, podem ser observados placas amiloides (PA) (também conhecidas como placas senis) e emaranhados neurofibrilares (ENFs) (Figura 2), além de extensa perda neuronal (HOLTZMAN; MORRIS; GOATE, 2011).

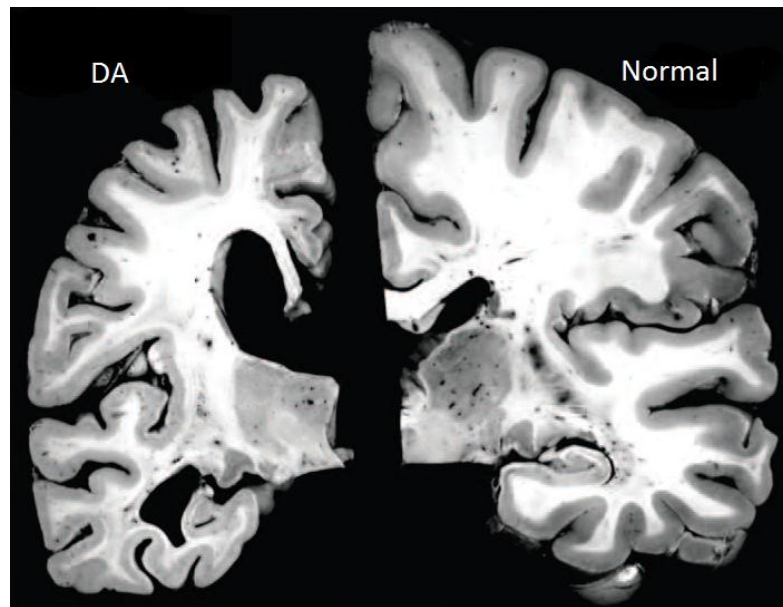


Figura 1. Comparação entre cortes cerebrais de um cérebro de um paciente com doença de Alzheimer (DA) e um cérebro normal (Adaptado de HOLTZMAN; MORRIS; GOATE, 2011).

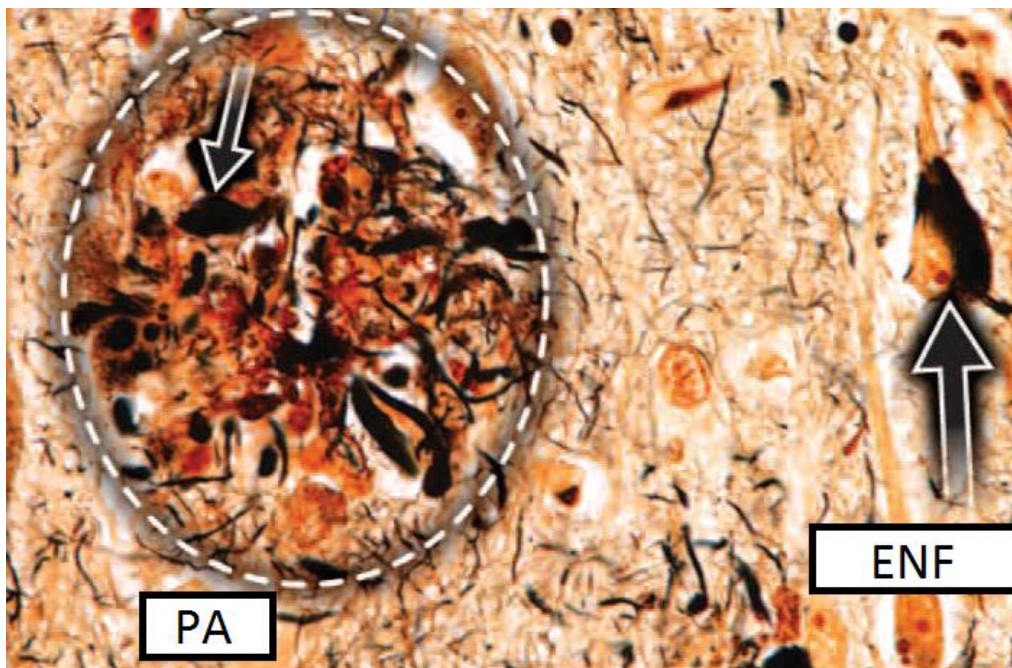


Figura 2. Fotomicrografia de uma placa amiloide (PA) e de um emaranhado neurofibrilar (ENF) (Adaptado de HOLTZMAN; MORRIS; GOATE, 2011).

Placas amiloides (PA) são depósitos moleculares no espaço extracelular do cérebro e seus principais componentes proteicos são peptídeos $A\beta$, que possuem de 38 a 43 aminoácidos e são derivados da proteína precursora amiloide (PPA). No interior

das placas, estes peptídeos formam agregados insolúveis, via um processo conhecido como β -amiloidose (GOLDE; ECKMAN; YOUNKIN, 2000). Quando as proteínas nos agregados apresentam uma conformação do tipo folha- β , os agregados são chamados de placas neuríticas e podem ser identificados por microscopia ótica (KAYED et al., 2003). Ao redor das placas neuríticas, ocorre também um processo inflamatório, envolvendo hipertrofia e alterações na morfologia das células gliais e multiplicação de astrócitos e micróglia, resultando em dano cerebral (LUCIN; WYSS-CORAY, 2009). Quando os agregados apresentam um formato diferente do tipo folha- β , eles são chamados de placas difusas e podem ser detectados por técnicas imuno-histoquímicas (KAYED et al., 2003).

Além da deposição do peptídeo A β na forma de placas amiloides, os corpos celulares neuronais também desenvolvem ENFs, que são estruturas formadas por uma forma agregada e hiperfosforilada da proteína Tau. Esta proteína é sintetizada em todos os neurônios e também está presente nas células da glia, tendo como função estabilizar os microtúbulos via ligação à tubulina. Entretanto, na DA, a proteína Tau hiperfosforilada sofre dissociação dos microtúbulos e ocorre uma autoagregação, gerando os ENFs (GRUNDKE-IQBAL et al., 1986).

Estudos genéticos, bioquímicos e neurobiológicos sugerem que a agregação, a modificação conformacional e o acúmulo de diferentes formas do peptídeo A β desempenham um papel central no desenvolvimento da patogênese da DA, embora uma variedade de modificações adicionais, dependentes ou não do peptídeo A β , parecem contribuir para o declínio cognitivo e a progressão da demência. Existem evidências que a disfunção da proteína Tau hiperfosforilada contribua para a progressão clínica da DA, provavelmente via processos posteriores à agregação do peptídeo A β , mas, também, em função de algum grau de dano neuronal independente da β -amiloidose (SMALL; DUFF, 2008).

1.1.4 Aspectos Genéticos

Existem evidências de que fatores genéticos podem ser muito relevantes para o desenvolvimento da DA, pois o histórico familiar positivo para esta patologia associado

à idade são os únicos fatores sistemáticos da DA (HEYMAN et al., 1984). A DA pode ser transmitida de forma autossômica dominante e as características relacionadas à idade de aparecimento e à evolução da doença são determinadas por subtipos genéticos distintos. Mutações genéticas presentes nos cromossomos 14 (gene pré-senilina-1) (LEVY-LAHAD et al., 1995) e 21 (gene PPA) (GOATE et al., 1991) estão diretamente envolvidas com o aparecimento precoce da doença (idade menor que 65 anos). Outros cromossomos envolvidos nos subtipos genéticos são o 1 (gene pré-senilina-2) e o 19 (apolipoproteína $\epsilon 4$) (VAN DUIJN, 1996).

1.1.5 Diagnóstico

Ainda não existem marcadores específicos para a investigação laboratorial e por imagem da DA, mas alguns exames convencionais na DA permitem a exclusão de causas reversíveis e, também, a observação de parâmetros de neuroimagem compatíveis com os variados estágios da DA; como atrofia cortical, alterações no hipocampo e lobo temporal mesial (POULIN; ZAKZANIS, 2002).

O diagnóstico definitivo da DA somente pode ser realizado pela observação de aspectos histopatológicos característicos da doença no tecido cerebral, como a perda neuronal nas camadas piramidais do córtex cerebral e degenerações sinápticas intensas nos níveis neocortical e hipocampal (BRAAK; BRAAK, 1991). Em função desta limitação, a única forma de diagnóstico é pelo exame neuropatológico segundo os critérios definidos pela NINCDS-ADRDA (*National Institute for Communicative Disorders and Stroke – Alzheimer’s Disease and Related Disorders Association*) em conjunto com a exclusão de outras possíveis causas de demência (MCKHANN et al., 1984).

1.1.6 Aspectos Epidemiológicos

A DA e a isquemia cerebrovascular são as causas mais frequentes da demência (NITRINI et al., 2005). Nos países desenvolvidos, a prevalência da DA é de, aproximadamente, 1,5% em pessoas com idade em torno de 65 anos e chega a 30%, em pessoas com cerca de 80 anos (RITCHIE; LOVESTONE, 2002). Nos Estados

Unidos, foi estimado que 3% dos indivíduos com idade entre 65 e 74 anos possuem DA. Este número aumenta para 17,6%, em pessoas com idade entre 75 e 85 anos, e para 32,3%, em pessoas com idade superior a 85 anos. Estima-se, ainda, que o número de pessoas com DA aumente de 4,7 milhões em 2010 para 13,8 milhões em 2050 (HEBERT et al., 2013). No Brasil, um estudo encontrou uma prevalência de 7,1% de casos de demência em uma população com idade igual ou maior que 65 anos. A taxa de incidência anual encontrada foi de 7,7 casos por 100.000 habitantes e a causa mais frequente da síndrome demencial foi a DA, representando um percentual de 55,1% (HERRERA et al., 2001).

1.1.7 Tratamento

O tratamento da DA envolve abordagens farmacológicas e não-farmacológicas, e o acompanhamento psicológico do paciente é extremamente importante neste processo. Dentre as principais opções farmacológicas atuais estão os inibidores reversíveis da enzima acetilcolinesterase (por exemplo, tacrina, galantamina e donepezil) e os inibidores pseudo-irreversíveis (por exemplo, rivastigmina), visto que se supõe uma deficiência colinérgica em pacientes com esta patologia; e os antagonistas competitivos do receptor glutamatérgico *N*-metil-*D*-aspartato (NMDA), por exemplo, memantina, visto que o glutamato pode atuar como uma excitotoxina e provocar morte neuronal, mesmo sendo também um dos neurotransmissores mais importantes do sistema nervoso central (FORLENZA, 2005).

Algumas outras abordagens farmacológicas menos comuns foram propostas, como o uso de antioxidantes (SANO et al., 1997), estrógenos (GREEN; GRIDLEY; SIMPKINS, 1996), estatinas (EVANS et al., 2004), anti-inflamatórios não esteroidais (LIM et al., 2001) e *Ginkgo biloba* (BIRKS; GRIMLEY EVANS, 2007); mas são menos utilizadas, pois estudos realizados com estas abordagens não foram capazes de comprovar os seus benefícios no tratamento da DA (BIRKS; FLICKER, 2003; HENDERSON et al., 2000; TABEL; FELDMAND, 2003; ZANDI et al., 2005). A importância da descoberta de novos fármacos para o tratamento de pacientes afetados

pela DA se dá em função dos fármacos que são utilizados atualmente não serem capazes de curar a doença, mas apenas retardar seu avanço (FORLENZA, 2005).

1.2 ENZIMA DYRK1A COMO UM NOVO ALVO CONTRA A DOENÇA DE ALZHEIMER

1.2.1 Quinase 1A de especificidade dual regulada via fosforilação de tirosina

As quinases de especificidade dual reguladas via fosforilação de tirosina – *Dual-specificity tyrosine-regulated kinases* (DYRKs) são uma família de quinases eucarióticas que pertencem a uma superfamília conhecida como CMGC quinases, que incluem, entre outras, proteínas quinases como: proteínas quinases ativadas por mitógeno – *mitogen-activated protein kinases* (MAPKs), quinases glicogênio sintase – *glycogen synthase kinases* (GSKs), quinases dependentes de ciclina - *cyclin dependente kinases* (CDKs) e as quinases do tipo CDK (BECKER; SIPPL, 2011). A família DYRK contém cinco subtipos: 1A, 1B, 2, 3 e 4. Entretanto, apenas o gene da DYRK1A está localizado no interior do cromossomo humano 21, precisamente, na região crítica da Síndrome de Down (SD) (BECKER; JOOST, 1999; GUIMERÁ et al., 1996). A família DYRK mantém alta similaridade entre as sequências de aminoácidos em diferentes espécies. A DYRK1A humana mantém 99% de conservação com as homólogas de ratos e camundongos, possuindo bons modelos experimentais para estudos de atividade e de participação em processos biológicos (MARTÍ et al., 2003).

A DYRK1A é uma proteína quinase capaz de catalisar a fosforilação de resíduos de serina e treonina em substratos proteicos exógenos, assim como a sua autofosforilação em resíduos de tirosina presentes em seu segmento de alça de ativação (HIMPEL et al., 2001; KENTRUP et al., 1996). Estudos indicam a participação desta enzima no desenvolvimento anormal do cérebro, levando a uma deficiência intelectual na SD (HÄMMERLE et al., 2003; PARK; SONG; CHUNG, 2009) e, também, no desenvolvimento de doenças neurodegenerativas como a DA (FERRER et al., 2005).

1.2.2 Estrutura Tridimensional da DYRK1A

A estrutura tridimensional da DYRK1A é semelhante à de outras quinases e apresenta o resíduo **Tyr321**, presente na alça de ativação, autofosforilado com interações com dois resíduos de arginina (**Arg325** e **Arg328**, Figura 3). Este é um domínio conservado em todas as enzimas da família DYRK (OGAWA et al., 2010). A região de homologia da família DYRK e o domínio catalítico da enzima estão localizados entre os resíduos **Val135** e **Lys480** (SOUNDARARAJAN et al., 2013).

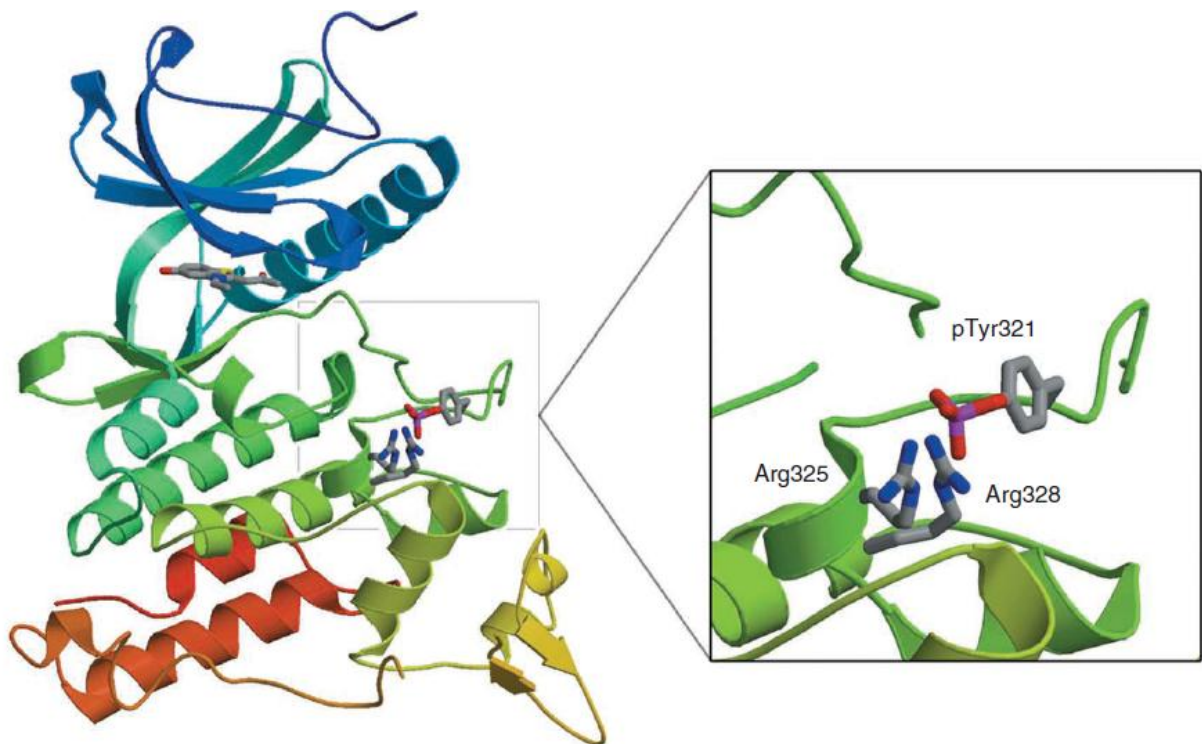


Figura 3. Estrutura tridimensional da DYRK1A humana e detalhe do sítio de autofosforilação composto por um resíduo de tirosina fosforilado (O3POTyr321 = pTyr321) e dois resíduos de arginina (Arg325 e Arg328) (Adaptado de OGAWA et al., 2010).

Existem algumas estruturas da DYRK1A co-cristalizadas com alguns ligantes disponíveis no banco de dados de proteínas – *Protein Data Bank* (PDB) (BERMAN et al., 2000). Todos estes compostos se ligam ao sítio de ligação do trifosfato de adenosina (ATP) necessário para as reações de fosforilação de substratos catalisada por esta enzima. O ligante benzotiazólico chamado INDY (inibidor de DYRK) interage com a enzima por ligações hidrogênio com o átomo de nitrogênio da cadeia principal do

resíduo **Leu241** e com o átomo de nitrogênio da cadeia lateral do resíduo **Lys188**, além de várias interações hidrofóbicas com os resíduos **Val173**, **Ala186**, **Phe238**, **Leu241**, **Leu294** e **Val306** (Figura 4) (Código no PDB – 2WO6) (OGAWA et al., 2010).

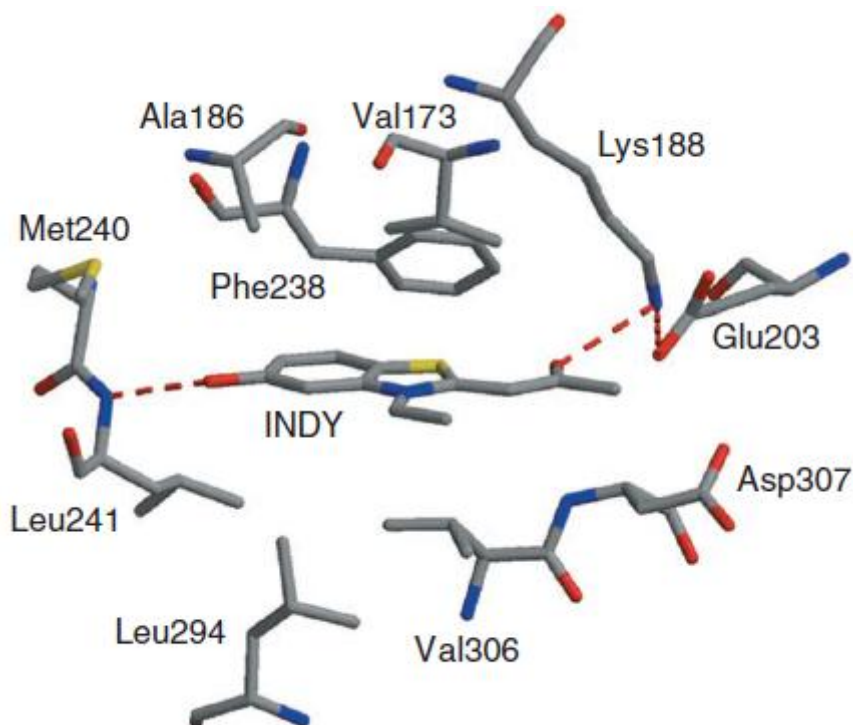


Figura 4. Sítio de ligação do INDY (inibidor de DYRK) à DYRK1A humana (Adaptado de OGAWA et al., 2010).

Dois inibidores da classe das pirido[2,3-*d*]pirimidinas, inicialmente desenvolvidas como inibidores da DYRK1B, também se apresentam co-cristalizadas à DYRK1A. O derivado 3 (Figura 5A) (Código no PDB – 4MQ1) interage com a enzima alvo por ligações hidrogênio com os resíduos **Leu241** e **Lys188** e por ligação halogênio (WILCKEN et al., 2013) com **Glu239**. O derivado 32 (Figura 5B) (Código no PDB – 4MQ2) mantém as interações por ligação hidrogênio com os resíduos **Leu241** e **Lys188** e faz uma ligação hidrogênio adicional com o resíduo **Asn292** (ANDERSON et al., 2013). Este mesmo padrão de interações por ligações hidrogênio é observado quando se analisa a estrutura da DYRK1A co-cristalizada com a leucettina L41, um derivado do alcaloide marinho leucettamina B (Código no PDB – 4MQ2). A leucettina L41 também faz interações hidrofóbicas com os resíduos **Val222**, **Leu241**, **Leu294** e **Val306** (TAHTOUH et al., 2012).

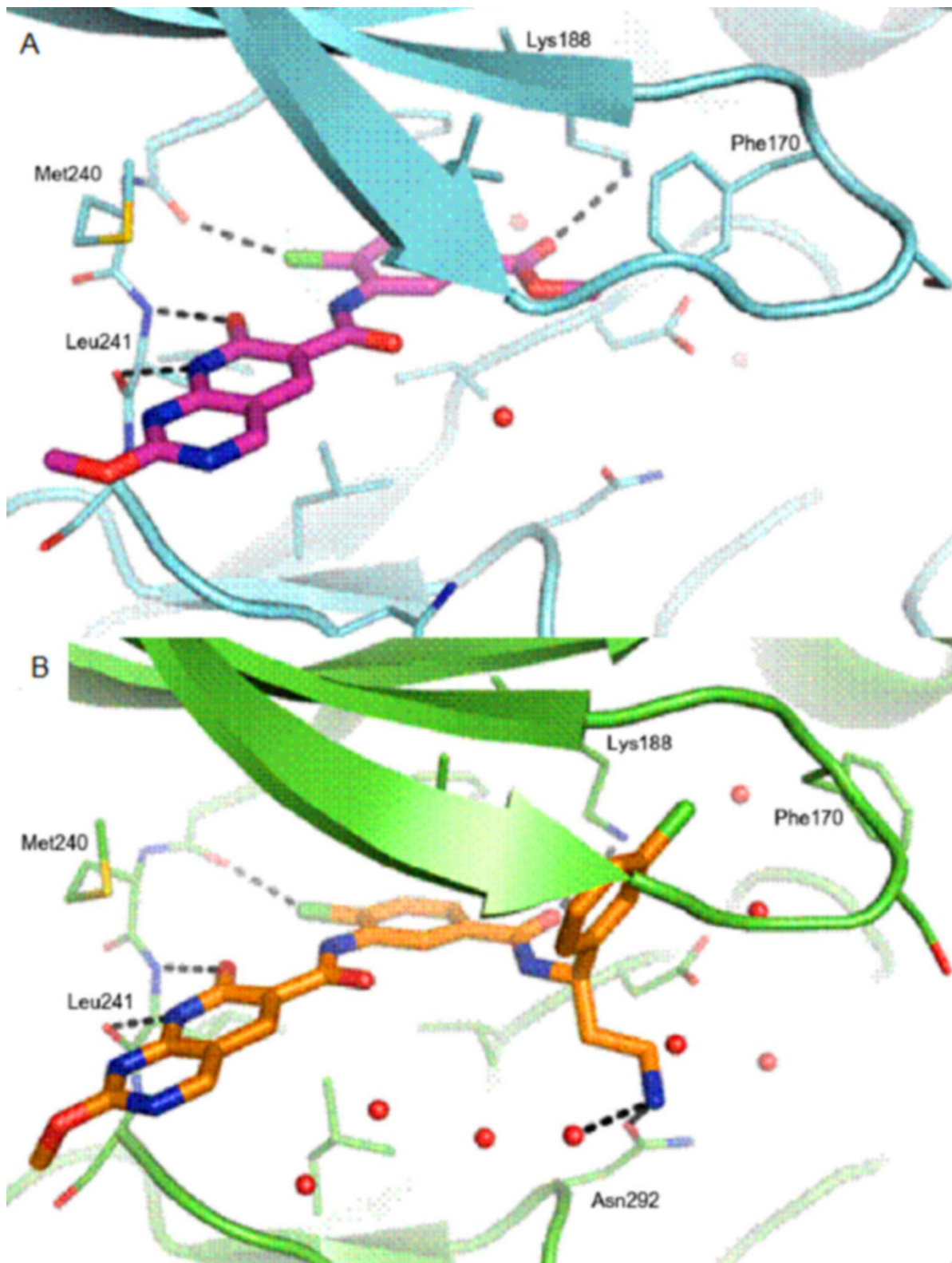


Figura 5. Sítio de ligação dos derivados **3** (A) e **32** (B) da classe das pirido[2,3-*d*]pirimidinas à DYRK1A humana (Adaptado de ANDERSON et al., 2013).

Outro inibidor co-cristalizado com a DYRK1A é o DJM2005, que faz diversas interações por ligações hidrogênio com os resíduos **Ser169**, **Tyr243**, **Lys289**, **Glu291**, **Asn292** e **Asp307**, além de várias interações hidrofóbicas (SOUNDARARAJAN et al., 2013).

1.2.3 Papel da DYRK1A no Desenvolvimento da Doença de Alzheimer

A DYRK1A possui múltiplas funções biológicas em função das suas interações com diversas proteínas do citoesqueleto celular, proteínas sinápticas e nucleares, incluindo fatores de transcrição e fatores envolvidos com a remoção pós-transcricional de íntrons e éxons (*splicing*) de ácidos ribonucleicos (RNAs) (GALCERAN et al., 2003). A expressão da DYRK1A em neurônios durante o período fetal e pós-natal, assim como em neurônios de indivíduos adultos e na velhice, sugere que a regulação desta expressão está envolvida com o desenvolvimento neuronal, sua maturação e envelhecimento posteriores (WEGIEL et al., 2004). Esta enzima é expressa em todo o organismo humano, mas é particularmente abundante no cerebelo, bulbo olfatório e hipocampo. Sua expressão é regulada positivamente nos primeiros estágios do desenvolvimento embrionário seguida de uma diminuição gradual a baixos níveis nos estágios posteriores (MARTÍ et al., 2003).

Foi observada imunorreatividade aumentada para DYRK1A no citoplasma e núcleo de neurônios dispersos no córtex entorrinal, hipocampo e neocórtex em doenças neurodegenerativas associadas à fosforilação de proteínas Tau; incluindo DA, SD e Doença de Pick (FERRER et al., 2005). A expressão aumentada da DYRK1A está envolvida tanto com a formação dos ENFs, via hiperfosforilação da proteína Tau; quanto com a formação das placas amiloides, via clivagem da PPA e consequente formação do peptídeo A β ; resultando em perda neuronal e demência (Figura 6) (WEGIEL; GONG; HWANG, 2011).

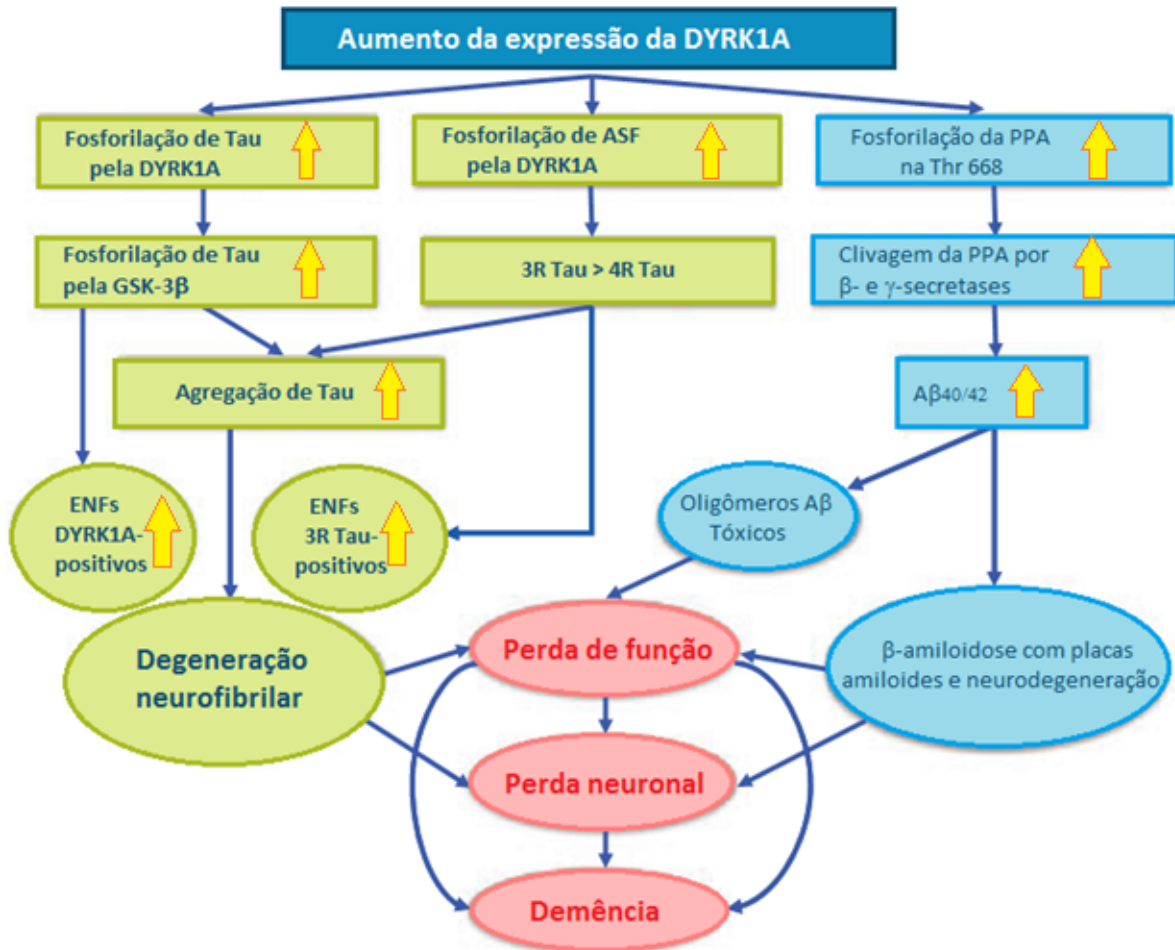


Figura 6. Contribuição da superexpressão da DYRK1A para a β -amiloidose e degeneração neurofibrilar em doenças neurodegenerativas. ASF (fator alternativo de *splicing*); PPA (proteína precursora amiloide); ENFs (emaranhados neurofibrilares) (Adaptado de WEGIEL; GONG; HWANG, 2011).

A fosforilação de proteínas Tau induzida pela DYRK1A com a atividade aumentada leva a subsequente fosforilação de proteínas Tau induzida pela GSK-3 β (LIU et al., 2008; WOODS et al., 2001). Paralelamente, a DYRK1A fosforila o fator de *splicing* alternativo (*alternative splicing factor*, ASF), que por sua vez leva ao aumento da expressão da proteína Tau-3R (SHI et al., 2008), causando a interrupção do equilíbrio nos níveis das proteínas Tau-3R e Tau-4R, fundamental para a função neuronal ótima (GOEDERT; JAKES, 2005). A formação de ENFs por estas duas vias de sinalização celular leva à degeneração neurofibrilar presente em doenças neurodegenerativas.

Estudos indicam que a fosforilação da PPA pela DYRK1A pode facilitar a clivagem da mesma por secretase- β 1 (LEE et al., 2003) e secretase- γ (VINGTDEUX et al., 2005); e aumentar a produção de peptídeos A β . Em camundongos transgênicos, que superexpressam o gene da DYRK1A humana, foram encontrados níveis aumentados de A β 40 e A β 42, em 160% e 17% respectivamente; indicando que a superexpressão da DYRK1A promove a clivagem da PPA e aumenta a produção de peptídeos A β (RYOO et al., 2008). O acúmulo destes oligômeros tóxicos inibe funções neuronais críticas e; portanto, contribui para a demência observada na DA (SAKONO; ZAKO, 2010).

A DYRK1A também fosforila o regulador da calcineurina-1 (RCAN1), o que diminui sua velocidade de degradação e aumenta a interação entre o RCAN1 e a calcineurina-1, contribuindo para o aumento da capacidade de inibição da atividade fosfatásica da mesma, o que resulta no aumento da quantidade de proteínas Tau fosforiladas e na redução da atividade transcricional da calcineurina/fator nuclear de células-T ativadas (NFAT) (JUNG et al., 2011).

1.2.4 Ativação da DYRK1A

A maioria das proteínas quinases pode adotar conformações ativa e inativa, que podem sofrer interconversão via fosforilações reversíveis em resíduos de serina, treonina ou tirosina presentes em uma região chamada de alça de ativação. A fosforilação destes resíduos presentes na alça estabiliza uma conformação da proteína que contém um sítio de ligação a um determinado substrato apropriadamente posicionado (NOLEN; TAYLOR; GHOSH, 2004). As quinases da família DYRK dependem da fosforilação de um resíduo conservado de tirosina presente na alça de ativação (**Tyr321**) para atingir a atividade catalítica máxima (HIMPEL et al., 2001). Esta fosforilação ocorre imediatamente após a tradução da mesma (LOCHHEAD et al., 2005).

A autofosforilação da DYRK1A de mamíferos é uma capacidade intrínseca do domínio catalítico e não depende de outros domínios ou quaisquer cofatores (GÖCKLER et al., 2009). Apesar da necessidade da autofosforilação para o alcance da

atividade catalítica máxima, ainda não existem evidências que a atividade catalítica das enzimas da família DYRK seja efetivamente regulada pela fosforilação e desfosforilação da tirosina presente na alça de ativação (BECKER; SIPPL, 2011).

Acerca do mecanismo de ativação da DYRK1A, sabe-se que a mutação de **Tyr321** para fenilalanina reduz a atividade catalítica da enzima em mais de 80% (WIECHMANN et al., 2003); entretanto, a desfosforilação deste resíduo, surpreendentemente, não provoca a inativação da mesma, o que sugere que a fosforilação deste resíduo é importante para mudar a conformação para a sua forma ativa, mas não para mantê-la (ADAYEV et al., 2007).

1.2.5 Regulação da DYRK1A

Como a fosforilação da **Tyr321** da alça de ativação é aparentemente constitutiva, a atividade da DYRK1A é regulada pelo seu nível de expressão gênica, o que pode ser explicado pela observação de pacientes com deficiência homozigota da DYRK1A (MØLLER et al., 2008). Dentre os fatores de transcrição descritos com propriedades inibidoras da expressão da DYRK1A estão o E2F1 (MAENZ et al., 2008) e o fator nuclear de células-T ativadas 1 – *Nuclear Factor of Activated T-Cells 1* (NFATc1) (LEE et al., 2009).

1.2.6 Inibição da DYRK1A Humana

Substâncias inibidoras de proteínas quinases capazes de permear membranas celulares são importantes no estudo de mecanismos de transdução de sinais intracelulares, pois possuem melhor controle temporal experimental e geram resultados fenotípicos diferentes quando comparadas com técnicas genéticas existentes, como a interferência por RNA (KNIGHT; SHOKAT, 2007).

A inibição da função da DYRK1A poderia, teoricamente, atenuar diversos processos relacionados à progressão da neurodegeneração associada à DA, visto que alvos da DYRKA, como as proteínas Tau, PPA e pré-senilina 1, estão claramente envolvidas em mecanismos pelos quais a atividade elevada da DYRK1A contribui para

o desenvolvimento da DA. Dada esta hipótese, há um grande interesse nesta enzima como potencial alvo farmacológico para novos inibidores (KIM et al., 2006; SAVAGE; GINGRICH, 2009).

1.2.6.1 Inibidores da DYRK1A Humana

A harmina é um alcaloide β -carbolínico conhecido por ser um potente inibidor da monoamina oxidase (MAO) (BAIN et al., 2007); e é produzida por diferentes espécies vegetais, incluindo o cipó sul-americano *Banisteriopsis caapi*. Este composto é um dos componentes da ayahuasca, uma mistura alucinógena utilizada em rituais religiosos e também na medicina tradicional dos povos amazônicos (KIM; SABLIN; RAMSAY, 1997). Como inibidor de quinases, a harmina apresenta grande especificidade para a DYRK1A ($IC_{50} = 80$ nM) entre 69 proteínas quinases (CALLAWAY et al., 1999). Há evidências que a harmina (Figura 7) é capaz de inibir a fosforilação direta de proteínas Tau via DYRK1A em células de neuroglioma H4 sem afetar a viabilidade celular (FROST et al., 2011). O principal problema relacionado à harmina é o seu potencial alucinogênico, mostrado como resultado de sua afinidade por receptores de serotonina e triptamina (AIRAKSINEN et al., 1987).

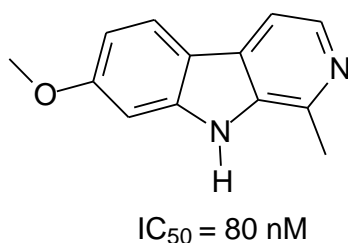


Figura 7. Estrutura química da harmina.

O galato de epigallocatequina (epigallocatequina-3-galato, EGCG) (Figura 8) é o maior componente polifenólico do chá verde e apresenta atividade inibidora sobre a DYRK1A humana, seletivamente, entre 29 quinases testadas ($IC_{50} = 300$ nM) (BAIN et al., 2003), sendo o mecanismo de inibição do tipo não-competitivo (ADAYEV et al., 2006). Estudos *in vivo* indicam que camundongos mutantes com DYRK1A

hiperexpressa apresentam melhor estrutura cerebral, plasticidade sináptica e função cognitiva; incluindo melhoria no aprendizado e nas memórias de curto e longo prazo, quando submetidos a uma dieta com chá verde contendo EGCG (GUEDJ et al., 2009). Apesar da atividade satisfatória, a EGCG possui uma farmacocinética complexa e baixa biodisponibilidade oral, o que limita sua utilização em modelos experimentais animais (LAMBERT; SANG; YANG, 2007).

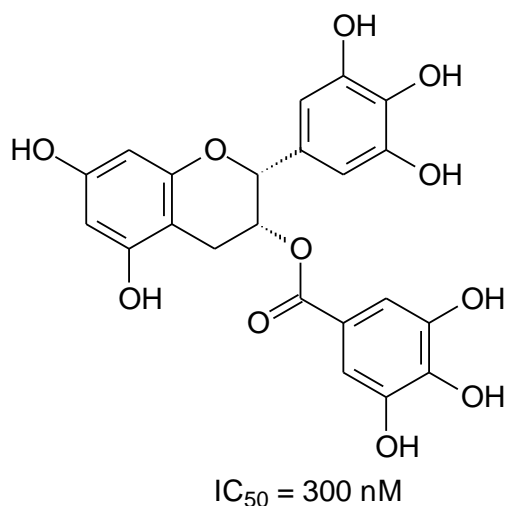


Figura 8. Estrutura química do galato de epigalocatequina (epigalocatequina-3-galato, EGCG).

Um composto benzotiazólico chamado INDY foi identificado como um inibidor dual de DYRK1A/CLK humanas. Este composto é um derivado do TG003, um inibidor de quinase tipo CDK-1 (*CDK-1 like kinase*, CLK), embora três vezes mais potente (Figura 9). Ambos os compostos também apresentaram atividade frente a outras enzimas da família DYRK humanas, como a DYRK1B. A atividade biológica do INDY também foi verificada em dois estudos *in vitro*: ele foi capaz de inibir a expressão da proteína Tau pThr212 em células de fibroblastos do tipo COS-7 e de bloquear a atividade reguladora negativa da DYRK1A na via calcineurina/NFAT de forma dependente da concentração (OGAWA et al., 2010).

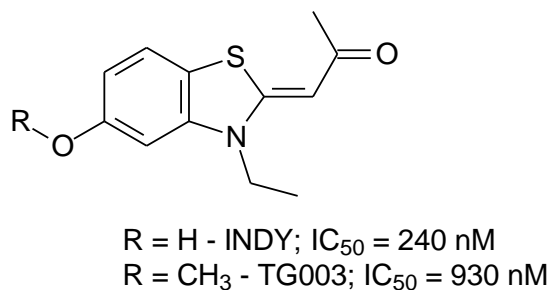


Figura 9. Estruturas químicas dos compostos INDY e TG003.

A busca por inibidores de outras enzimas da família DYRK também gerou compostos capazes de inibir a DYRK1A. A avaliação de uma série de derivados pirido[2,3-*d*]pirimidinas, objetivando o tratamento de cânceres via a inibição da DYRK1B, resultou em diversos compostos com atividade biológica pronunciada na DYRK1A humana, incluindo o derivado mostrado na Figura 10, o que levou ao estudo farmacocinético de alguns dos derivados obtidos (ANDERSON et al., 2013). Foi salientado, inclusive, que a diferença entre as DYRK1A e DYRK1B na região de ligação ao ATP é de apenas um aminoácido (leucina na DYRK1A e metionina na DYRK1B). Assim, é um desafio conseguir compostos seletivos para uma destas enzimas que se liguem a esta região (LEDER et al., 1999).

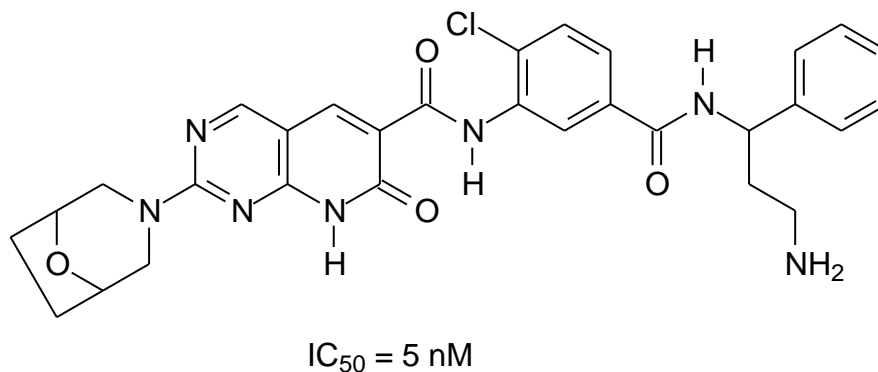


Figura 10. Estrutura geral do composto mais ativo da série de derivados pirido[2,3-*d*]pirimidinas.

Outros inibidores da DYRK1A humana foram avaliados e, dentre estes, podem ser citados produtos naturais como a quinalizarina que, apesar de inibir a DYRK1A, possui maior afinidade pela proteína quinase CK2 (COZZA et al., 2009); flavonoides como o acalinal A e B, que apresentam seletividade, mas atividades moderadas como

inibidores de DYRK1A (AHMADU et al., 2010); benzocumarinas (SARNO et al., 2012) e indolocarbazois, como a estaurosporina e a rebecamicina (SÁNCHEZ et al., 2009).

Dentre os compostos sintéticos, o universo de compostos é muito maior, e podem ser citados como exemplos as pirazolidino-3,5-dionas (KOO et al., 2009); as meriolinas (ECHALIER et al., 2008); as meridianinas (GIRAUD et al., 2011) e os cromeno[3,4-*b*]indóis (NEAGOIE et al., 2012). Todos os análogos ainda estão em fase de testes *in vitro*, e nenhum teste clínico foi realizado.

Um aspecto em comum em diversos trabalhos buscando inibidores da DYRK1A é a pequena quantidade de compostos com valores exatos de atividade biológica. Nestes trabalhos, a relação entre estrutura e atividade biológica é discutida de forma apenas qualitativa, pois não há um número de compostos suficientes para a geração de modelos quantitativos de relação estrutura-atividade. Uma série que atende a esta característica é a de 6-arylquinazolina-4-aminas (Figura 11) testadas para diversas CLKs e para DYRK1A e 1B.

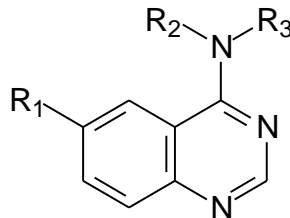


Figura 11. Estrutura geral da série de derivados 6-arylquinazolina-4-aminas.

Os derivados 6-arylquinazolina-4-aminas foram inicialmente idealizados como inibidores de CLKs, onde um destes derivados, o NCGC00010037 foi reportado como inibidor das enzimas CLK1, CLK4 e DYRK1A (MOTT et al., 2009), o que despertou o interesse sobre esse grupo de derivados, visto que quinazolininas substituídas são farmacóforos comuns para inibidores competitivos da ATP (ROSENTHAL et al., 2010, 2011).

Em função deste interesse, dois trabalhos mais aprofundados sobre estes derivados foram realizados. Nestes trabalhos, foram empregadas diversas substituições nas posições R₁, R₂ e R₃ (Figura 11), e os derivados obtidos foram testados contra uma série de proteínas quinases em ensaios enzimáticos utilizando ATP marcado com ³³P.

Com os resultados dos ensaios *in vitro*, foi proposta a relação qualitativa entre a estrutura química e atividade biológica destes compostos (ROSENTHAL et al., 2010, 2011). A partir destes trabalhos, foram selecionados 46 compostos com atividade sobre a DYRK1A humana para estudos quantitativos de relação estrutura atividade no presente trabalho.

1.3 MODELAGEM MOLECULAR E PLANEJAMENTO DE NOVOS FÁRMACOS

Inicialmente, os métodos de obtenção de novos fármacos eram baseados em tentativas e erros, e qualquer substância com potencial valor terapêutico poderia ser testada diretamente em indivíduos doentes para tratar empiricamente uma determinada patologia e, conseqüentemente, atestar sua eficácia. Com o avanço das técnicas químicas, os pesquisadores passaram a purificar substâncias de preparações de origem vegetal com propriedades medicinais conhecidas e elucidar suas estruturas químicas (SNEADER, 2005). A ciência da descoberta de novos fármacos avançou ainda mais em função dos progressos nas áreas de bioquímica e biologia molecular, utilizando conceitos de expressão gênica, síntese de proteínas e a relação ligante-receptor (DREWS, 2000).

Com a ampliação do conhecimento em genômica e proteômica, ocorrido nos anos 90 e com os avanços nas áreas de bioinformática e quimioinformática, foram desenvolvidas e introduzidas novas técnicas de análise capazes de identificar genes e proteínas relacionadas com diversas patologias. Esta introdução foi feita, primeiramente, em genômica e proteômica (BUTCHER; BERG; KUNKEL, 2004) e, posteriormente, em bioinformática (CHEN; CHEN, 2008) e quimioinformática (ENGEL, 2006), possibilitando o processamento de dados em larga escala e a experimentação *in silico* (ou seja, por simulação computacional) (LEE; HUANG; JUAN, 2011).

Um dos objetivos do planejamento de fármacos é selecionar um alvo terapêutico na etiologia de uma determinada patologia e encontrar uma ou mais substâncias capazes de interagir com este alvo, ativando-o ou desativando-o (MOON; HOWE, 1991). As modificações resultantes desta interação desencadearão uma série de reações, gerando um determinado efeito terapêutico (SLENO; EMILI, 2008).

A modelagem molecular compreende um conjunto diverso de métodos que permite construir, visualizar, analisar e armazenar sistemas moleculares complexos, contribuindo no planejamento de novos fármacos. Existem duas abordagens principais na concepção de novos fármacos utilizando técnicas de modelagem molecular: a abordagem indireta (ou dependente do ligante) e a abordagem direta (ou dependente da estrutura do receptor) (COHEN et al., 1990).

Na abordagem indireta, o planejamento de fármacos é baseado em processos que utilizam, como dados de partida, ligantes conhecidos de um determinado receptor alvo. Modelos farmacofóricos, modelos de relação quantitativa estrutura-atividade de uma, duas, três ou quatro dimensões, independente do receptor, são exemplos de métodos deste tipo de abordagem (TINTORI; MANETTI; BOTTA, 2010).

Na abordagem direta, informações sobre a estrutura tridimensional do receptor alvo devem estar disponíveis. Estas informações podem ser obtidas por técnicas de ressonância magnética nuclear (RMN) ou cristalografia e difração de raios-X. Estudos de docagem molecular e simulações de dinâmica molecular são exemplos de técnicas desta abordagem (IVANOV et al., 2006).

1.3.1 QSAR

O método de relação quantitativa estrutura-atividade (*quantitative structure-activity relationship*, QSAR) tenta correlacionar a variação da resposta da atividade biológica com a variação da estrutura química (ou propriedades derivadas da estrutura química) em uma série de compostos. Se a atividade de um grupo de ligantes puder ser determinada, é possível construir um modelo matemático que descreva esta relação (ESPOSITO; HOPFINGER; MADURA, 2004). A descrição da estrutura de um fármaco é um problema complexo, visto que não pode ser representada por um único parâmetro. Desta forma, um grupo de propriedades químicas ou estruturais conhecidas como descritores são calculados a partir da estrutura e usados nesta quantificação. Utilizando os descritores como variáveis independentes e a atividade biológica como a variável dependente, pode-se construir um modelo que as relacione (PERKINS et al., 2003).

Uma grande variedade de métodos de QSAR foi desenvolvida desde a introdução deste conceito em 1964 (FREE; WILSON, 1964; HANSCH; FUJITA, 1964). Os métodos tradicionais de QSAR 1D e 2D, como Free-Wilson e Hansch-Fujita, usam substituintes moleculares ou fragmentos bidimensionais e propriedades físico-químicas para gerar modelos preditivos quantitativos. O primeiro método de QSAR 3D chamado de análise comparativa de campos moleculares - *Comparative Molecular Field Analysis* (CoMFA) foi introduzido em 1988 (CRAMER; PATTERSON; BUNCE, 1988) e deu origem a outros métodos de QSAR 3D, como a análise comparativa de índices de similaridade moleculares - *Comparative Molecular Similarity Index Analysis* (CoMSIA), assim como métodos de QSAR multidimensionais como o QSAR 4D, QSAR 5D, entre outros, com o objetivo de solucionar o problema crucial dos métodos de QSAR 3D, a limitação de uma única conformação para a estrutura tridimensional de cada ligante (VEDANI; DOBLER; LILL, 2006).

Em geral, a modelagem por QSAR envolve um processo sistemático com múltiplas etapas (Figura 12), incluindo o preparo do conjunto de dados, seleção e geração dos descritores moleculares, desenvolvimento do modelo matemático, validação interna do modelo, utilizando um conjunto de treinamento e a validação externa utilizando um conjunto de teste (MYINT; XIE, 2010).

Durante a primeira etapa (Figura 12), é importante estar atento à qualidade dos dados para desenvolver um modelo de QSAR confiável. Os dados devem advir do mesmo tipo de protocolo de ensaio biológico e é preferível utilizar dados de um único laboratório, a fim de evitar inconsistência e variabilidade interlaboratorial dos dados (MYINT; XIE, 2010). Além disso, o conjunto de dados deve conter um número mínimo de compostos para assegurar a consistência estatística; suas atividades biológicas devem cobrir uma larga faixa de valores e possuir uma boa distribuição dentro desta faixa e o mesmo deve ser dividido em grupos de treinamento e de teste (PERKINS et al., 2003).

A segunda etapa (Figura 12) envolve a geração e a seleção dos descritores moleculares. Vários descritores podem ser gerados, mas apenas alguns deles podem estar correlacionados de forma significativa com a atividade biológica. Portanto, a seleção apropriada destes descritores é fundamental para gerar um modelo de QSAR

robusto. Na terceira etapa, é necessário escolher o método estatístico adequado para relacionar os descritores com as atividades biológicas. Alguns modelos lineares, como a regressão linear múltipla ou o método dos mínimos quadrados parciais – do inglês *partial least squares* (PLS) ou não lineares, como as redes neurais, podem ser utilizados como funções de correlação (LEACH, 2011).

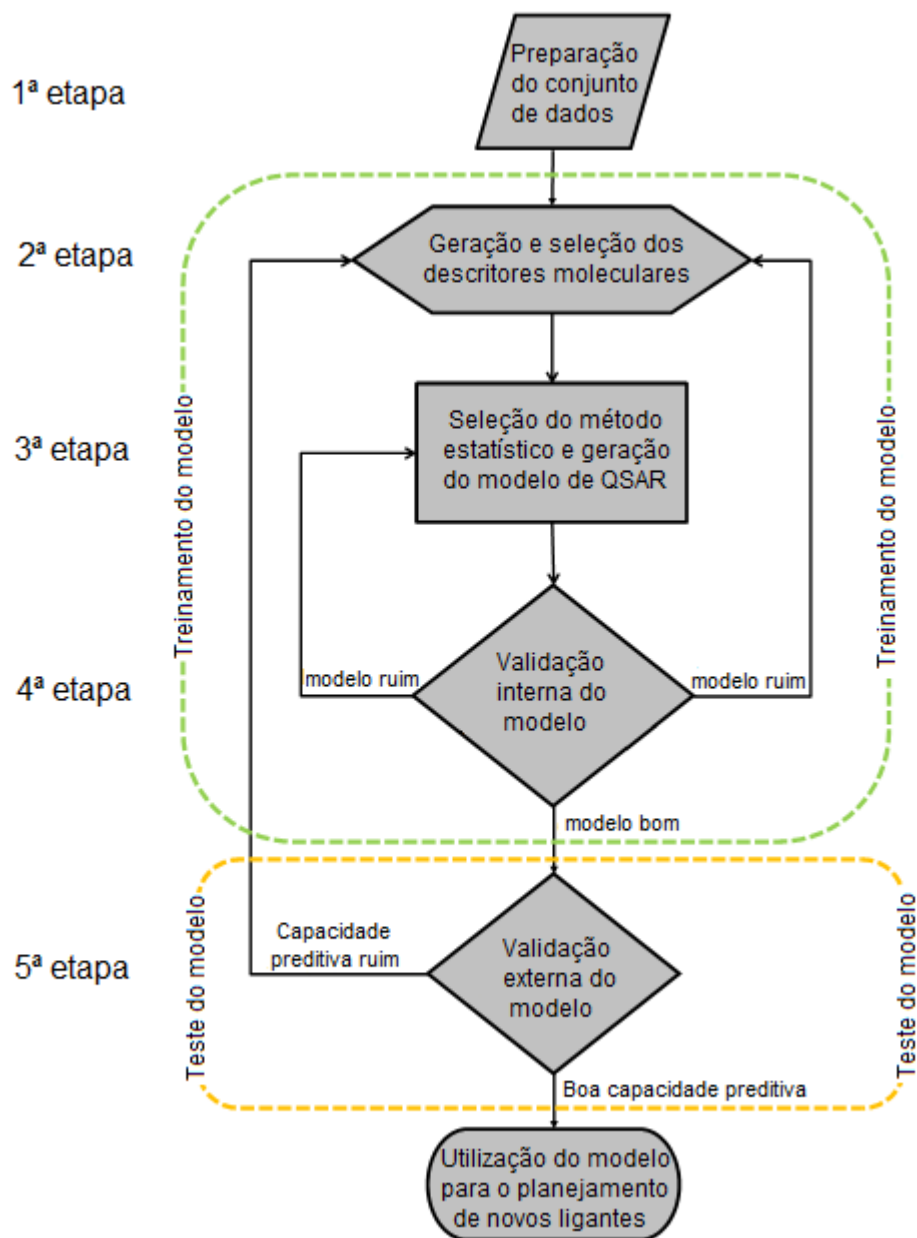


Figura 12. Esquema geral do desenvolvimento de um modelo de QSAR (Adaptado de MYINT; XIE, 2011).

Na quarta etapa, o modelo de QSAR deve ser submetido à validação interna. Um método de validação bastante utilizado é a validação cruzada por exclusão individual exaustiva, mais conhecida pelo seu termo em inglês – *leave-one-out cross validation* (LOO_{CV}) – realizada para avaliar a estabilidade estatística do modelo, onde um composto por vez é excluído do conjunto de treinamento. Na quinta etapa, realiza-se a validação externa do modelo; utilizando o conjunto de teste que é constituído por compostos relacionados estruturalmente aos compostos do conjunto de treinamento, mas que não deram origem ao modelo, o que é importante para avaliar a capacidade preditiva externa do mesmo (LIVINGSTONE, 1995).

1.3.1.1 QSAR 2D: Holograma QSAR (HQSAR)

HQSAR é um método de QSAR 2D baseado em fragmentos moleculares e que, portanto, não requer informações que são necessárias em técnicas tridimensionais; como a determinação da estrutura tridimensional, conformação de ligação ao receptor e alinhamento molecular (HURST; HERITAGE, 2001; LOWIS, 1997). Nesta técnica, cada molécula do grupo de treinamento é representada na forma de uma notação linear específica, neste caso, a *Sybyl line notation* (SLN) e desmembrada em fragmentos estruturais únicos (lineares, ramificados ou cíclicos) os quais irão formar um holograma molecular, uma espécie de impressão digital que codifica todos os possíveis fragmentos formados. Em seguida, é atribuído um número inteiro positivo a cada fragmento por um algoritmo de checagem de redundância e estes fragmentos são ordenados em hologramas (matrizes contendo posições chamadas *bins*) de comprimentos variados. As frequências de ocupação destes *bins* são os descritores estruturais desta técnica que contém informações topológicas e constitucionais das moléculas, incluindo algumas informações tridimensionais, como hibridização e quiralidade (SALUM; ANDRICOPULO, 2009).

Existem alguns parâmetros que podem afetar a qualidade dos modelos de HQSAR; como tamanho dos fragmentos, comprimento do holograma e distinção dos fragmentos. O tamanho dos fragmentos é um parâmetro que define o número máximo e mínimo de átomos em que as moléculas do conjunto de dados são fragmentadas; o

comprimento do holograma expressa o número de *bins* disponíveis para a alocação dos fragmentos e o parâmetro de distinção de fragmentos define características diferenciais entre os fragmentos, descritas na Tabela 1.

Tabela 1. Parâmetros de distinção dos fragmentos moleculares empregados no método HQSAR.

Distinção de fragmentos	Definição
Átomos (A)	Tipos de átomos
Ligações (L)	Tipos de ligação: simples, dupla ou tripla
Conectividade (C)	Tipos de hibridização dos átomos
Hidrogênio (H)	Número de átomos de hidrogênio
Quiralidade (Q)	Centros de quiralidade
Doador e aceptor de ligação hidrogênio (DA)	Átomos doadores ou aceptores de ligação hidrogênio

As frequências são, então, relacionadas aos valores de atividade biológica na forma de uma equação matemática pelo método de regressão por PLS. Este método estatístico é semelhante à análise por componentes principais, do inglês – *principal component analysis* (PCA), a qual busca correlacionar a variável dependente (atividade biológica) com as variáveis independentes (descritores moleculares) de uma única vez, sendo extremamente útil em estudos de QSAR nos quais é gerada uma quantidade muito grande de variáveis. O método de PLS reduz as variáveis iniciais pela construção de variáveis latentes, que são combinações lineares das variáveis iniciais (SALUM; ANDRICOPULO, 2009).

A análise por PLS gera uma equação de regressão que correlaciona os valores dos *bins* dos hologramas com as respectivas atividades biológicas (ABs) (Equação 1).

$$AB_i = const + \sum_{j=1}^L x_{ij} C_j \quad (\text{Eq. 1})$$

Na Equação 1, AB_i é a atividade biológica do i -ésimo composto, x_{ij} é o valor de ocupação do holograma molecular do i -ésimo composto no bin j , C_j é o coeficiente para o bin j derivado da análise por PLS, e L é o comprimento do holograma (Figura 13).

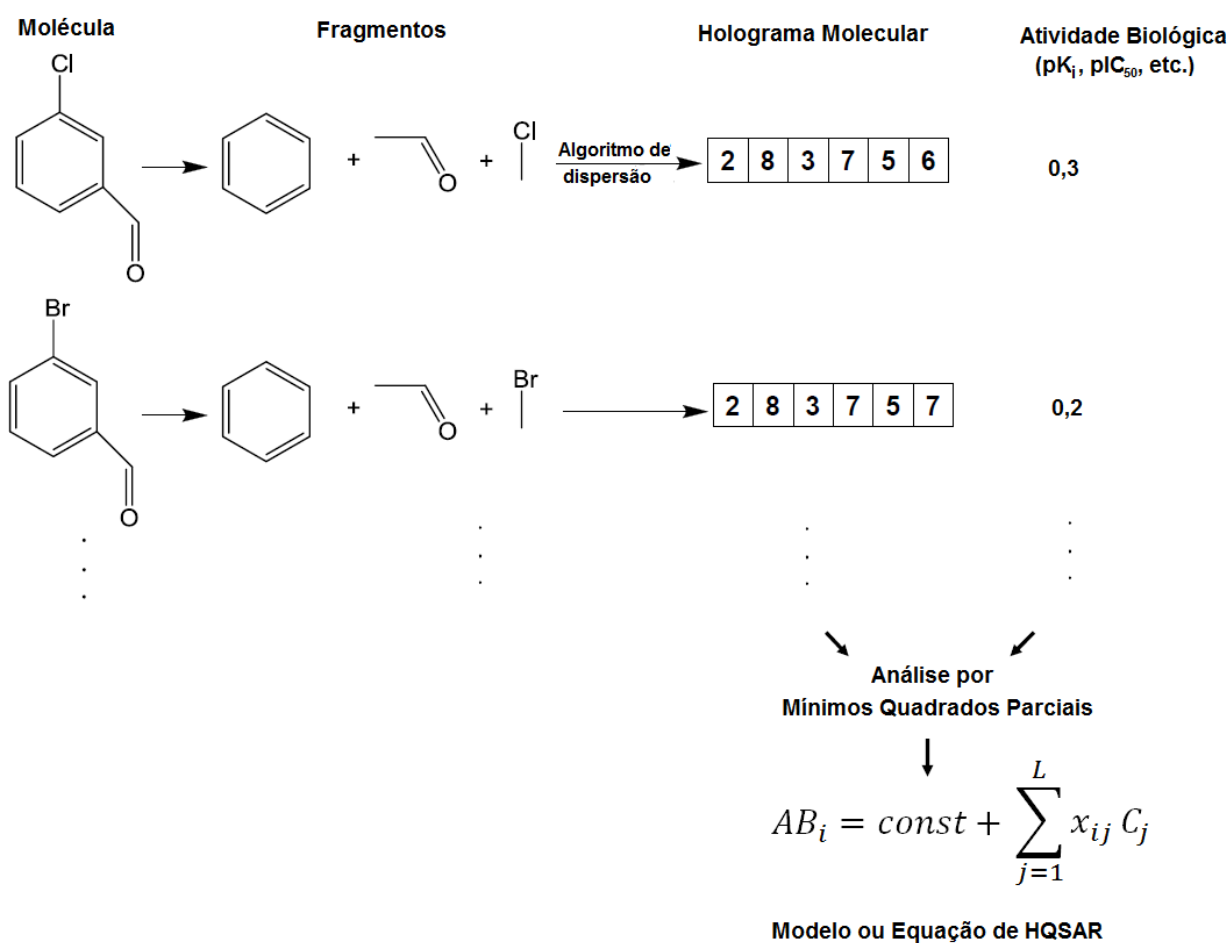


Figura 13. Esquema geral da geração de hologramas moleculares e modelos de QSAR (Adaptado de MYINT; XIE, 2010).

Para a identificação do número adequado de variáveis ou componentes que resulta em um modelo ótimo, é realizado o processo de validação interna do modelo por LOO_{CV} , uma técnica que consiste na exclusão sistemática de cada composto do

conjunto de treinamento, seguida pela predição da sua respectiva atividade biológica pelo novo modelo gerado a partir dos compostos restantes até que todos os compostos tenham sido excluídos uma vez.

Um problema do método de HQSAR é o fenômeno chamado de colisão de fragmentos que pode ocorrer durante o processo de ordenação dos fragmentos, o que pode fazer com que existam hologramas com diferentes fragmentos no mesmo *bin*. A alteração dos comprimentos dos hologramas, os quais controlam o número de *bins* em cada holograma, modifica os padrões das ocupações dos *bins*. Os comprimentos dos hologramas podem ser definidos e existem 12 comprimentos padrões previamente testados com bons resultados em diferentes conjuntos de dados que geram grupos únicos de colisões de fragmentos (LOWIS, 1997).

A técnica de HQSAR também fornece informações úteis sobre a relação entre cada fragmento de um ligante e a atividade biológica do mesmo. Estes fragmentos podem ser facilmente visualizados nos denominados “mapas de contribuição” (Figura 14), que refletem a contribuição individual de cada átomo ou fragmento estrutural de uma dada molécula bioativa do conjunto de dados (HURST; HERITAGE, 2001). Estes mapas de contribuição podem fornecer informações diretas para os químicos sintéticos medicinais e podem auxiliar no processo de planejamento de novos fármacos (ALBUQUERQUE et al., 2007).

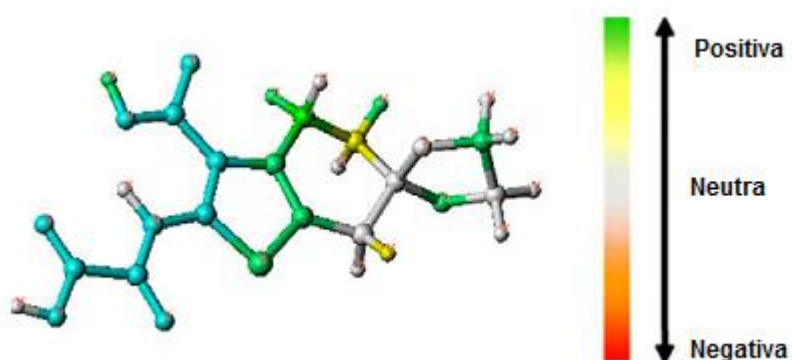


Figura 14. Representação de um mapa de contribuição de um ligante da série de ácidos 2-(oxalilamino) benzoicos (Adaptado de CHENG et al., 2010).

A técnica de HQSAR tem sido frequentemente aplicada a diversos conjuntos de dados com resultados excelentes (BHANSALI; KULKARNI, 2014; CHENG et al., 2010; PALANGSUNTIKUL et al., 2013; TROSSINI; MALTAROLLO; SCHMIDT, 2014).

1.3.1.2 QSAR 3D: CoMFA e CoMSIA

O CoMFA é um método de QSAR que descreve as relações entre estrutura tridimensional e atividade, pois considera as estruturas tridimensionais dos ligantes, para as quais são gerados campos moleculares estéricos e eletrostáticos que são sobrepostos e, portanto, é considerado um método que depende do alinhamento molecular e da determinação das conformações bioativas dos ligantes. Estes campos moleculares são os descritores que são correlacionados com a atividade biológica por meio de uma equação matemática (CRAMER; PATTERSON; BUNCE, 1988).

O conjunto de dados a ser modelado pelo método CoMFA deve possuir uma característica: todos os ligantes devem interagir com a mesma proteína alvo, seja ela um receptor, enzima ou transportador, segundo um mesmo modo de ligação com a mesma geometria relativa (KUBINYI, 2003). Este conjunto deve ser dividido em conjuntos de treinamento e teste de forma semelhante ao executado na análise por HQSAR.

Em seguida, é necessário o cálculo das cargas atômicas parciais e a seleção das prováveis conformações bioativas pela definição da hipótese farmacofórica, a partir de estudos de docagem previamente realizados ou de dados cristalográficos obtidos da literatura (GOLBRAIKH; BERNARD; CHRÉTIEN, 2000). Quando não há informações sobre qual é a conformação bioativa, o consenso é a utilização da conformação de menor energia (RODRIGUES, 1999). A etapa de alinhamento (e sobreposição) do conjunto de dados é complexa, principalmente quando as estruturas 3D dos ligantes não são semelhantes (IVANCUIC, 2000), o que significa que um alinhamento ideal é algo muito difícil, mesmo para séries congêneres.

Após o alinhamento, os compostos sobrepostos são inseridos numa caixa, mantendo-se uma distância mínima entre os compostos e os limites da caixa. Define-

se, então, a distância da grade que gerará pequenos cubos no interior da caixa, cujo valor padrão é de 2 Å, podendo ser variado, embora a diminuição desta distância aumente exponencialmente o tempo de cálculo dos campos moleculares (BRITO, 2008).

Os campos moleculares são calculados, separadamente, para cada composto do conjunto de dados, utilizando um átomo de prova, que simula átomos da proteína receptora, podendo ser de diversos tipos, os mais comuns são: carbono sp^3 com carga +1, hidrogênio com carga +1 e oxigênio sp^3 com carga -1 (CHO; TROPSHA, 1995). Os campos estéricos são calculados de acordo com a função potencial de Lennard-Jones (Equação 2), enquanto os campos eletrostáticos são calculados segundo a função potencial de Coulomb (Equação 3).

$$E_C = \sum_{i=1}^n \frac{q_i q_j}{D r_{ij}} \quad (\text{Eq. 2})$$

Na Equação 2, E_C é a energia de interação de Coulomb, q_i é a carga parcial do átomo i na molécula, q_j é a carga do átomo de prova, r_{ij} é a distância entre o átomo i e o ponto de interseção da grade onde o átomo de prova j é inserido, e D é a constante dielétrica do meio (KUBINYI, 2003).

$$E_{vdW} = \sum_{i=1}^n (A_{ij} r_{ij}^{-12} - C_{ij} r_{ij}^{-6}) \quad (\text{Eq. 3})$$

Na Equação 3, E_{vdW} é a energia de interação de van der Waals, r_{ij} é a distância entre o átomo i e o ponto de interseção da grade onde o átomo de prova j é inserido, A_{ij} e C_{ij} são constantes que dependem do raio de van der Waals de cada átomo na molécula. É importante ressaltar que o termo $A_{ij} r_{ij}^{-12}$ descreve a repulsão, enquanto o termo $C_{ij} r_{ij}^{-6}$ descreve a atração. Conforme a diminuição da distância entre os átomos que interagem diminui, as energias de interação estérica e eletrostática tendem ao infinito, o que ocorreria com a sobreposição dos mesmos, portanto, valores de corte

(*cut-off*) são definidos para evitar que este fenômeno ocorra (Figura 15) (KUBINYI, 2003).

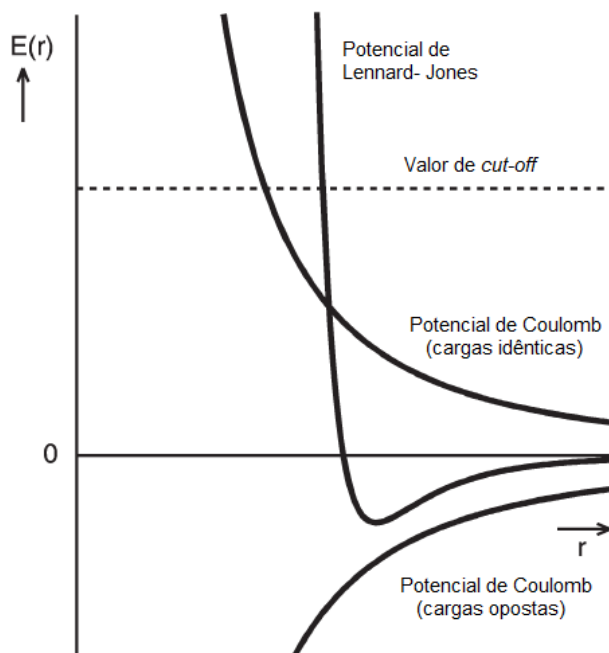


Figura 15. Diagrama representando os potenciais eletrostático (Coulomb) e de van der Waals (Lennard-Jones) em função da distância interatômica (Adaptado de KUBINYI, 2003).

A obtenção do modelo de CoMFA (incluindo a validação interna) e a validação externa do mesmo são realizadas pelas mesmas técnicas utilizadas na análise por HQSAR: PLS e LOO_{CV}, respectivamente. A Equação 4 mostra a relação entre a atividade biológica (AB) e os campos moleculares obtidos. Também é necessária a validação externa do modelo para estabelecer a sua capacidade de prever a atividade biológica de outros compostos relacionados (i.e., dos compostos do conjunto de teste).

$$AB_i = const + \sum_{j=i}^N a_{ij} F_{ij} \quad (\text{Eq. 4})$$

Na Equação 4, AB_i é a atividade biológica do i-ésimo composto, F_{ij} é a j-ésima energia de campo (eletrostática/estérica) calculada para o composto i, e a_{ij} é o coeficiente associado a esta energia.

A visualização das interações eletrostáticas e estéricas favoráveis e desfavoráveis para a atividade biológica do conjunto de dados pode ser realizada pelos mapas de contorno tridimensionais, sendo possível a identificação de sítios de interação com o receptor. Um esquema geral da análise por CoMFA é mostrado na Figura 16.

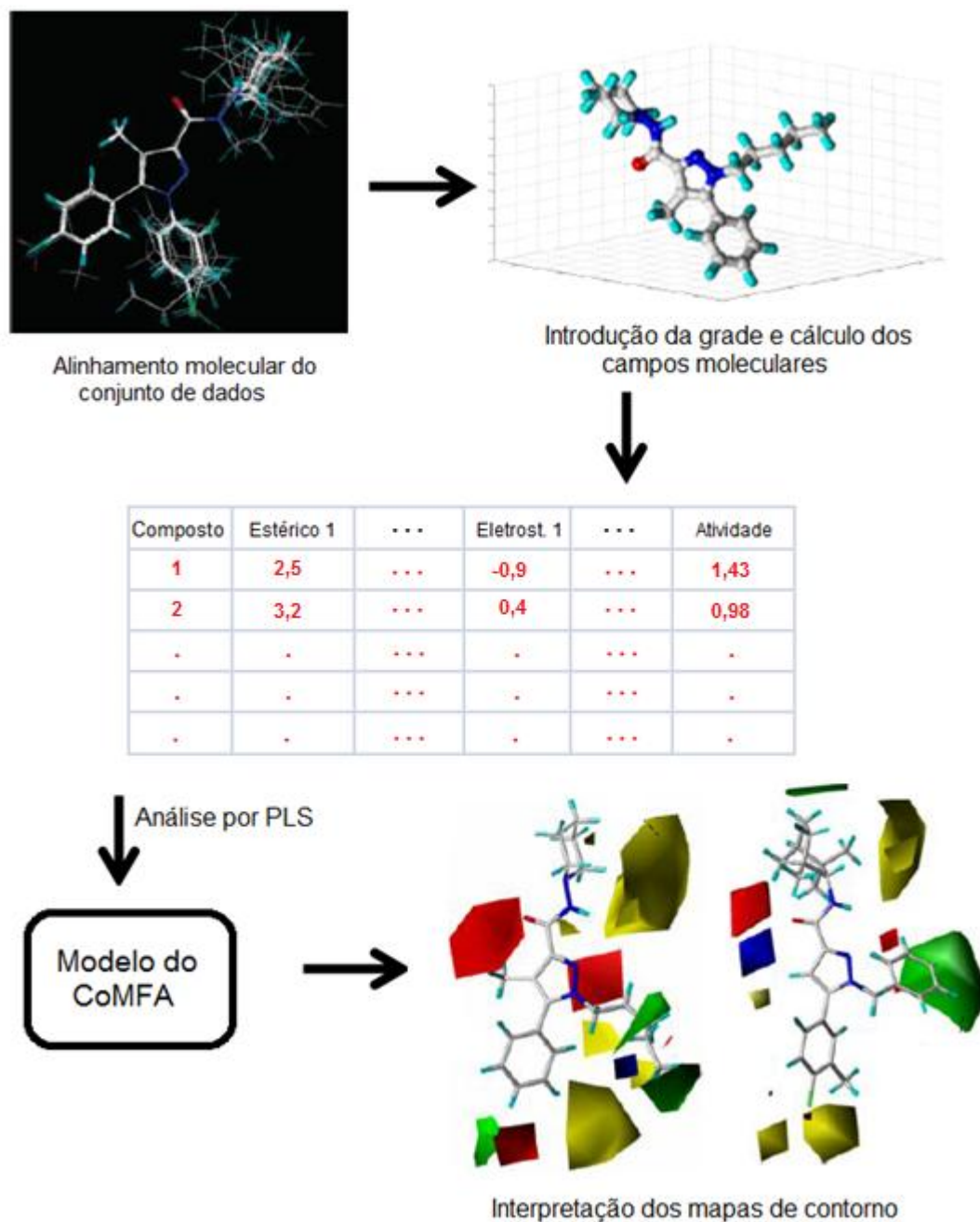


Figura 16. Esquema geral da análise por CoMFA (Adaptado de MYINT; XIE, 2011).

O CoMSIA é um método semelhante ao CoMFA, pois também é dependente de alinhamento molecular, onde os compostos sobrepostos também são inserido numa caixa gradeada, com diferenças no método de cálculo dos descritores. No CoMSIA, os descritores são calculados em termos de índices de similaridade, ao invés de funções de potenciais de interação, como as funções de Coulomb e de Lennard-Jones. O índice de semelhança A_F para um determinado composto j com átomos i em um ponto da caixa q é uma função do tipo Gaussiana, dependente da distância e é determinado de acordo com a Equação 5.

$$A_{Fk}(j) = - \sum_{i=1}^n w_{prova,k} w_{ik} e^{-\alpha r_{iq}^2} \quad (\text{Eq. 5})$$

Na Equação 5, $w_{prova,k}$ é o átomo de prova, $w_{i,k}$ é o valor da propriedade química k do átomo i , r_{iq} é a distância entre o átomo de prova e o átomo i do composto, e α é o fator de atenuação (KLEBE; ABRAHAM; MIETZNER, 1994).

Neste método, podem ser avaliadas cinco propriedades físico-químicas: estérica, eletrostática, hidrofóbica, doadora e aceptora de ligação hidrogênio. Para esta avaliação, usa-se um único átomo de prova com raio de 1 Å, carga igual a +1 e propriedades de ligação hidrogênio e hidrofóbica. Os índices estéricos estão relacionados à terceira potência da distância, os índices eletrostáticos derivam das cargas atômicas parciais previamente calculadas, os índices hidrofóbicos são baseados no logaritmo do coeficiente de partição (n-octanol/água) (VISWANADHAN et al., 1989) e os índices relacionados às propriedades doadoras e aceptoras de ligação hidrogênio baseiam-se em um conjunto de regras determinadas a partir de valores determinados experimentalmente (KLEBE; MIETZNER; WEBER, 1999).

Os valores do fator de atenuação podem variar entre 0,2 e 0,4. Valores maiores do que 0,4 resultam em funções Gaussianas mais íngremes e que atenuam mais fortemente os efeitos dependentes da distância no cálculo dos índices de similaridade. Desta forma, valores mais altos de α significam avaliações mais localizadas da similaridade, enquanto valores mais baixos de α significam avaliações mais globais (DA CUNHA, 2006).

1.3.2 Docagem Molecular

A docagem molecular (do inglês, *molecular docking*) é uma técnica de modelagem molecular que busca prever modos de ligação (encaixe) entre duas (ou mais) moléculas, dadas as suas respectivas coordenadas atômicas (HALPERIN et al., 2002). Um método de docagem bem sucedido deve conseguir prever a conformação nativa do ligante no interior do sítio de ligação da proteína, assim como as interações moleculares relacionadas ao complexo ligante-proteína (JAIN; NICHOLLS, 2008). Além disso, o método deve permitir a distinção entre ligantes e não ligantes e ordenar corretamente os ligantes do conjunto de dados (KOLB; IRWIN, 2009).

Os algoritmos básicos de um método de docagem capaz de gerar e avaliar as conformações de um determinado ligante são os algoritmos de busca (pesquisa) e de função de pontuação. A capacidade de tratar a flexibilidade molecular intrínseca do sistema e de descrever corretamente a energia das interações ligante-proteína é um fator crítico no desenvolvimento de métodos de docagem úteis no planejamento de fármacos (GUEDES; DE MAGALHÃES; DARDENNE, 2013).

Os algoritmos de busca podem ser classificados em três grupos principais de acordo com o método empregado na exploração da flexibilidade do ligante: algoritmos de busca sistemática, estocástica e determinística; sendo que alguns algoritmos são considerados híbridos, pois associam duas ou todas as estratégias (BROOIJMANS; KUNTZ, 2003).

Os algoritmos de busca sistemática exploram todos os graus de liberdade do ligante e, por sua vez, podem ser classificados em três grupos: exaustivo, construção incremental e conjunto de conformações. Os algoritmos por busca exaustiva exploram, sistematicamente, os valores de cada grau de liberdade de modo combinatório (HUANG; GRINTER; ZOU, 2010). Os algoritmos de busca por construção incremental fragmentam o ligante em subunidades e selecionam e encaixam um fragmento-base no sítio de ligação do receptor, reconstruindo o ligante via ligações covalentes sucessivas dos demais fragmentos ao fragmento-base (SMELLIE; CRIPPEN; RICHARDS, 1991). Os algoritmos de busca por conjunto de conformações encaixam de forma rígida um determinado grupo de conformações ao sítio de ligação (EWING et al., 2001).

Os algoritmos de busca estocástica ou randômica modificam, aleatoriamente, todos os graus de liberdade do ligante (translacional, rotacional e conformacional) a cada etapa, gerando uma grande diversidade de soluções (poses). Estas poses são avaliadas de acordo com um determinado critério probabilístico para ser decidido quais serão rejeitadas ou não. Estes métodos não garantem a convergência para uma solução ótima e, portanto, devem ser realizadas várias corridas independentes para encontrar o mínimo global de energia. Como exemplos destes algoritmos, podem ser citados os métodos de Monte Carlo, algoritmos evolucionários, busca tabu e otimização por enxame (HUANG; GRINTER; ZOU, 2010).

Nos algoritmos determinísticos, o estado atual do sistema determina as modificações que serão realizadas para alcançar o próximo estado; ou seja, uma solução depende fortemente da estrutura antecessora, pois uma determinada configuração estrutural inicial associada a um determinado conjunto de parâmetros resultará sempre no mesmo estado final. Como exemplos destes algoritmos, podem ser citados os métodos de minimização de energia e de dinâmica molecular (GUEDES; DE MAGALHÃES; DARDENNE, 2013).

As funções de pontuação avaliam a qualidade das soluções geradas pelos algoritmos de busca, indicando quais poses são as mais relevantes durante o processo de docagem. Estas funções devem ser capazes de distinguir as soluções em função da sua energia de ligação, classificar ligantes como ativos ou inativos e prever constantes de afinidade e classificar as diversas soluções de acordo com a potência (CHENG et al., 2009). Estas funções podem ser divididas em três tipos: baseadas em campos de força, empíricas e baseadas em conhecimento (WANG et al., 2005).

As funções baseadas em campos de força consistem em uma soma de termos de energia e, usualmente, consideram as energias de interação entre ligante e proteína, além da energia interna do ligante (MORRIS et al., 1998; VERDONK et al., 2003). As funções empíricas são baseadas na ideia de que é possível correlacionar a energia livre de ligação a uma soma ponderada de variáveis não relacionadas, cujos coeficientes são obtidos por análise de regressão, utilizando dados de afinidade experimentais (DE AZEVEDO; DIAS, 2008). As funções baseadas em conhecimento empregam análise estatística das interações entre de pares de átomos de complexos sistema ligante-

proteína experimentais, e o resultado desta análise é convertido em um pseudopotencial que descreve a geometria dos pares de átomos do sistema ligante-proteína problema (WALLQVIST; COVELL, 1996).

A inclusão da flexibilidade proteica nos protocolos de docagem é um grande desafio neste campo. Apesar de gerar resultados mais próximos da realidade em função da grande liberdade conformacional de proteínas em geral, esta inclusão gera um custo de tempo de cálculo e de capacidade computacional muito maior do que em análises que consideram proteínas rígidas (GUEDES; DE MAGALHÃES; DARDENNE, 2013).

O Molegro Virtual Docker (MVD) é um programa para estudos de docagem que utiliza algoritmos de busca combinados com um algoritmo de predição de cavidades que identifica de forma rápida e precisa os potenciais modos de ligação entre ligante e proteína. Este programa contém três tipos de algoritmos de busca e dois tipos de funções de pontuação, que podem ser combinados para obter a associação que melhor se aplica ao sistema problema. Os algoritmos de busca inclusos no MVD são: *MolDock Optimizer*, *MolDock Simplex Evolution* (SE) e *Iterated Simplex*; e as funções de pontuação são: *MolDock Score* e *PLANTS Score*.

O algoritmo de busca padrão do MVD é o *MolDock Optimizer* que é baseado em uma classe de algoritmos evolucionários conhecida como evolução diferencial, que gera soluções-filhas a partir da diferença ponderada das soluções-pais selecionadas aleatoriamente da população de soluções (STORN; PRICE, 1997). O *MolDock SE* é um algoritmo estocástico alternativo aplicável a ligantes com muitos graus de liberdade, funcionando em sistemas onde o algoritmo padrão falha, pois combina a construção incremental das poses com a escolha aleatória de indivíduos para o início da busca pelas soluções de menor energia. O *Iterated Simplex* gera soluções que são sucessivamente refinadas, utilizando níveis de tolerância de diferenças de energia de ligação entre as melhores e piores soluções cada vez menores até um número máximo de repetições previamente determinado (THOMSEN; CHRISTENSEN, 2006).

A função de pontuação *MolDock Score* é derivada de funções do tipo potencial linear composto - *piecewise linear potential* (PLP), originalmente propostas por Gehlhaar e colaboradores (GEHLHAAR; BOUZIDA; REJTO, 1998; GEHLHAAR et al.,

1995), e posteriormente modificada por Yang e Chen (YANG & CHEN, 2004), definida pela Equação 6.

$$E_{score} = E_{intra} + E_{inter} \quad (\text{Eq. 6})$$

Na Equação 6, E_{inter} é a energia de interação entre ligante e proteína e E_{intra} é a energia interna do ligante. A energia de interação ligante-proteína (E_{inter}), por sua vez, é definida pela equação 7.

$$E_{inter} = \sum_{i \in lig.} \sum_{j \in ptn.} \left[E_{PLP}(r_{ij}) + 332 \frac{q_i q_j}{4r_{ij}^2} \right] \quad (\text{Eq. 7})$$

Na Equação 7, E_{PLP} é o potencial linear composto (PLP); enquanto o segundo termo descreve as interações eletrostáticas do sistema. Nesta equação, a soma engloba todos os átomos pesados (ou seja, todos os átomos com exceção do átomo de hidrogênio), incluindo água e cofatores presentes no sistema, e o valor de 332 é utilizado para que a unidade de energia seja Kcal/mol (THOMSEN; CHRISTENSEN, 2006).

A energia interna do ligante (E_{intra}) é descrita pela Equação 8.

$$E_{intra} = \sum_{i \in lig.} \sum_{j \in ptn.} E_{PLP}(r_{ij}) + \sum_{lig.simp} [1 - \cos(m \cdot \theta - \theta_0)] + E_{col} \quad (\text{Eq. 8})$$

Na Equação 8, a soma dupla no primeiro termo considera todos os pares de átomos no ligante, excluindo os pares de átomos que estão conectados por duas ligações ou menos. O segundo termo representa a energia torsional do ligante e é parametrizado de acordo com a hibridização dos átomos que definem o ângulo de torção (θ). O terceiro termo, E_{col} , acrescenta penalidade à energia interna do ligante, quando dois átomos não ligados estão a menos de 2,0 Å entre si, evitando conformações improváveis do ligante (THOMSEN; CHRISTENSEN, 2006).

A função de pontuação PLANTS Score é definida pela Equação 9.

$$E_{plantsscore} = f_{PLP} + f_{col} + f_{tors} + c_{esfera} - 20 \quad (\text{Eq. 9})$$

Na Equação 9, o potencial f_{PLP} é semelhante ao potencial PLP da função *MolDock Score*, porém contendo mais tipos de interação. Os potenciais de colisão (f_{col}) e torsional (f_{tors}) levam em consideração colisões internas e contribuições torsionais para ligações simples (KORB; STÜTZLE; EXNER, 2009), o termo c_{esfera} aplica uma penalidade caso o ligante encontre-se fora do sítio de ligação definido por uma esfera no espaço e o último termo, -20, foi necessário para que os valores desta nova função sejam comparáveis aos da função PLANTS inicial.

2 OBJETIVOS

Em função da grande importância que a enzima DYRK1A possui no processo de desenvolvimento da doença de Alzheimer e do interesse no desenvolvimento de inibidores desta enzima, este estudo tem por objetivo compreender a relação entre a estrutura química e a atividade inibitória sobre a DYRK1A de uma série de 6-arilquinazolina-4-aminas (46 compostos) e o modo de interação com a enzima alvo para orientar o planejamento de novos potenciais candidatos a fármacos para o tratamento de indivíduos acometidos por esta patologia. Para atingir estes objetivos, serão construídos modelos de QSAR 2D (HQSAR) e 3D (CoMFA e CoMSIA) e serão obtidos complexos ligante-enzima por método de docagem molecular.

3 METODOLOGIA

3.1 CONJUNTO DE DADOS ESTRUTURAL E BIOLÓGICO

Uma série de 46 compostos da classe das 6-arylquinazolina-4-aminas foi obtida de trabalhos publicados na literatura por Rosenthal e colaboradores, onde os ensaios para a determinação da atividade biológica *in vitro* de cada composto da série seguiu o mesmo protocolo experimental (ROSENTHAL et al., 2010, 2011). Alguns dos compostos presentes nestes trabalhos não possuíam o valor exato da sua atividade biológica, pois ultrapassaram o valor limite de quantificação da atividade do método utilizado (10000 nM) e, portanto, não foram utilizados na construção dos modelos.

Os valores da atividade biológica dos compostos são expressos como a concentração mínima necessária para inibir 50% da atividade catalítica da DYRK1A humana (IC_{50} , em nM). Estes valores foram convertidos para a unidade molar (M) e, posteriormente, transformados em seus respectivos logaritmos inversos (pIC_{50}). A Tabela 2 apresenta as estruturas químicas dos compostos da série e seus respectivos valores de pIC_{50} , que estão regularmente distribuídos entre 5,05 e 7,85.

Tabela 2. Estruturas químicas dos 46 compostos da classe das 6-arylquinazolina-4-aminas e seus respectivos valores de IC_{50} (nM) e pIC_{50} . Os compostos do conjunto de teste estão marcados com um asterisco.

Composto	R ₁	R ₂	R ₃	IC ₅₀ (nM)	pIC ₅₀
1		H		62	7,21
2		H		1262	5,90

Tabela 2. (cont.)

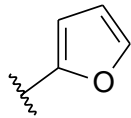
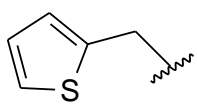
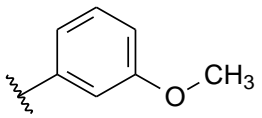
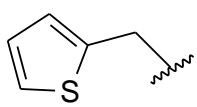
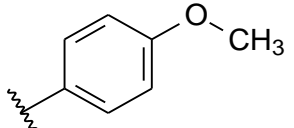
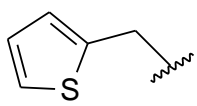
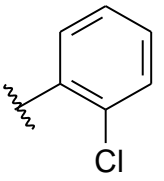
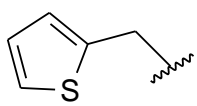
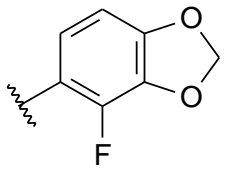
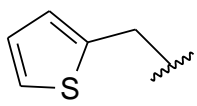
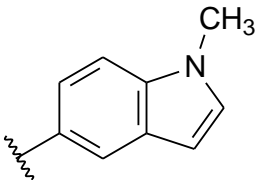
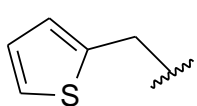
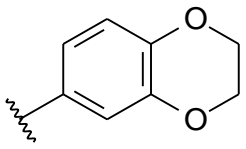
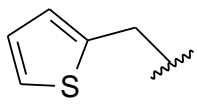
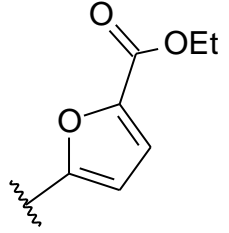
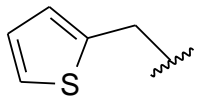
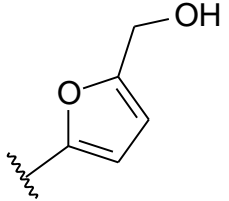
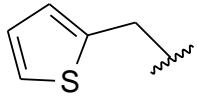
# ^{a,b,c}	R ₁	R ₂	R ₃	IC ₅₀ (nM)	pIC ₅₀
3*		H		3480	5,46
4		H		5697	5,24
5		H		3152	5,50
6		H		9012	5,05
7		H		164	6,79
8		H		4517	5,35
9		H		180	6,74
10		H		1437	5,84
11*		H		4657	5,33

Tabela 2. (cont.)

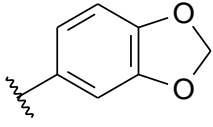
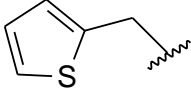
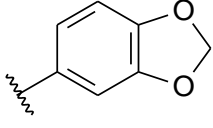
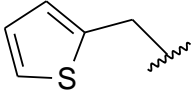
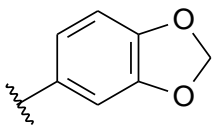
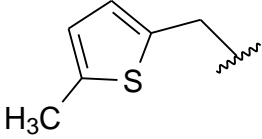
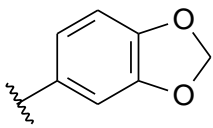
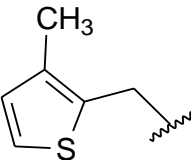
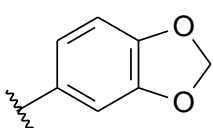
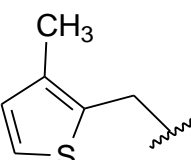
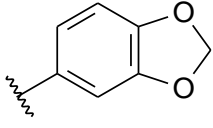
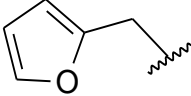
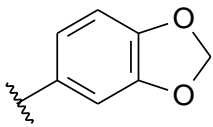
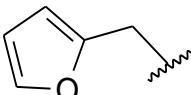
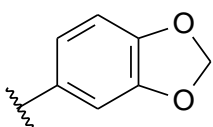
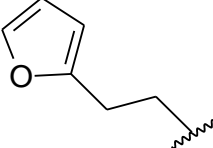
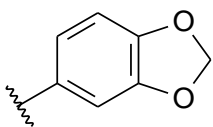
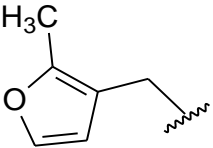
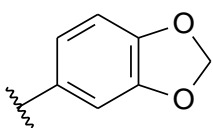
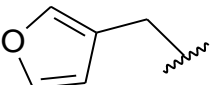
Composto	R ₁	R ₂	R ₃	IC ₅₀ (nM)	pIC ₅₀
12		CH ₃		31	7,51
13		CH ₂ CH ₃		38	7,42
14		H		1158	5,94
15		CH ₃		260	6,59
16		H		35	7,46
17		H		84	7,08
18		CH ₃		98	7,01
19		H		74	7,13
20*		H		57	7,24
21		H		126	6,90

Tabela 2. (cont.)

Composto	R ₁	R ₂	R ₃	IC ₅₀ (nM)	pIC ₅₀
22		H		93	7,03
23*		H		206	6,69
24		CH ₃		14	7,85
25		H		70	7,15
26		CH ₃		51	7,29
27		CH ₃		26	7,59
28		H		155	6,81
29*		CH ₃		922	6,04
30		H		541	6,27
31		CH ₃		120	6,92

Tabela 2. (cont.)

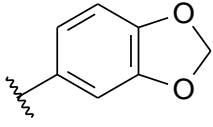
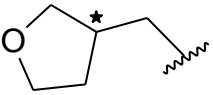
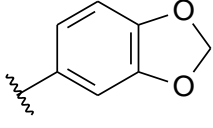
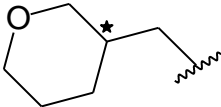
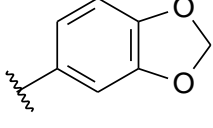
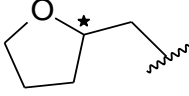
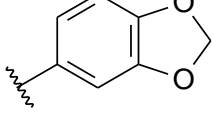
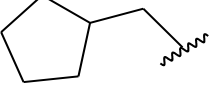
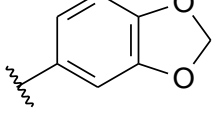
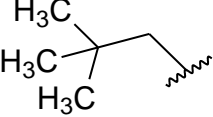
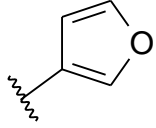
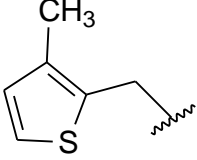
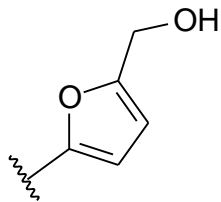
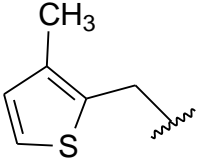
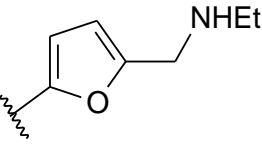
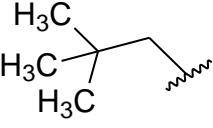
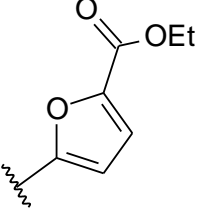
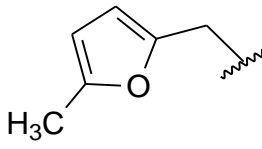
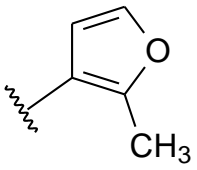
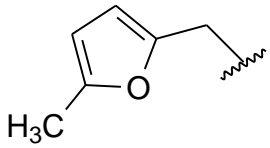
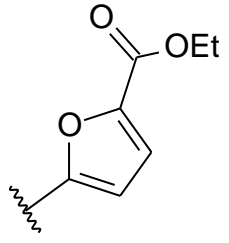
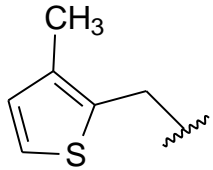
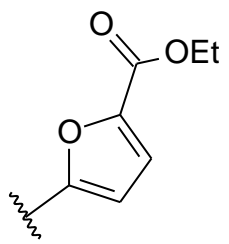
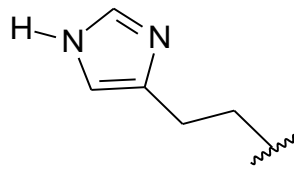
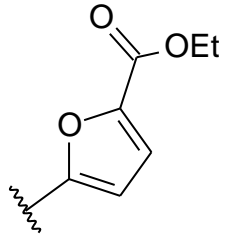
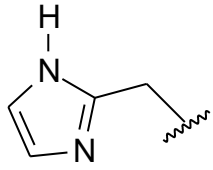
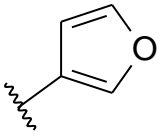
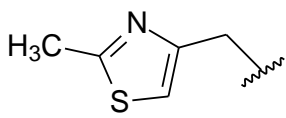
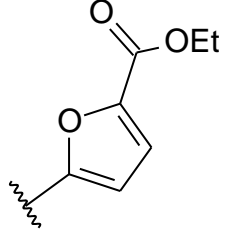
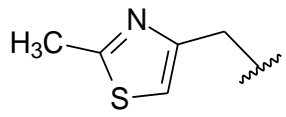
Composto	R ₁	R ₂	R ₃	IC ₅₀ (nM)	pIC ₅₀
32*		H		93	7,03
33*		H		135	6,87
34*		H		30	7,52
35		H		76	7,12
36*		H		17	7,77
37		CH ₃		1136	5,94
38		CH ₃		557	6,25
39		H		594	6,23
40		H		3629	5,44

Tabela 2. (cont.)

Composto	R ₁	R ₂	R ₃	IC ₅₀ (nM)	pIC ₅₀
41		H		3388	5,47
42		CH ₃		1501	5,82
43*		H		1406	5,85
44		H		2706	5,57
45		H		4820	5,31
46		H		8307	5,08

^a Os compostos do conjunto de teste estão marcados com asterisco. ^b Os compostos **42** até o **46** são de (Rosenthal et al., 2010), os outros compostos são de (Rosenthal et al., 2011). ^c Os compostos **32**, **33**, e **34** possuem centros de quiralidade e suas atividades biológicas são das suas respectivas misturas racêmicas.

3.1.1 Conjuntos de Treinamento e Teste

Para os estudos de HQSAR, CoMFA e CoMSIA, o conjunto de dados contendo os 46 compostos foi dividido em um conjunto de treinamento com 36 compostos, incluindo o mais e o menos ativo; e um conjunto de teste com 10 compostos, que estão assinalados com um asterisco na Tabela 2, representando, aproximadamente, 20% do número total de compostos. A divisão não foi completamente aleatória para garantir a diversidade química e de atividade biológica para os dois grupos. Os compostos **32**, **33** e **34**, contendo um único centro de quiralidade cada, foram incluídos no grupo teste, pois seus valores de atividade biológica são das correspondentes misturas racêmicas, entretanto, as suas formas enantioméricas (*R* e *S*) foram modeladas separadamente.

As distribuições dos valores de atividade biológica dos compostos do grupo de treinamento e de teste podem ser observadas nas Figuras 17 e 18, respectivamente.

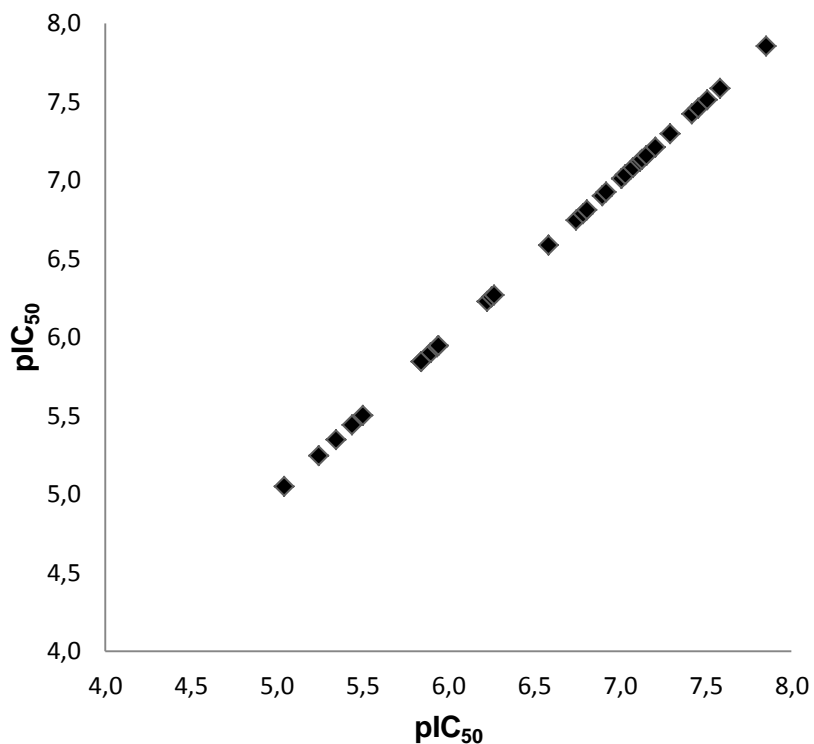


Figura 17. Gráfico da distribuição dos valores de atividade biológica (pIC₅₀) dos 36 compostos (Rosenthal et al. 2010; 2011) do grupo de treinamento.

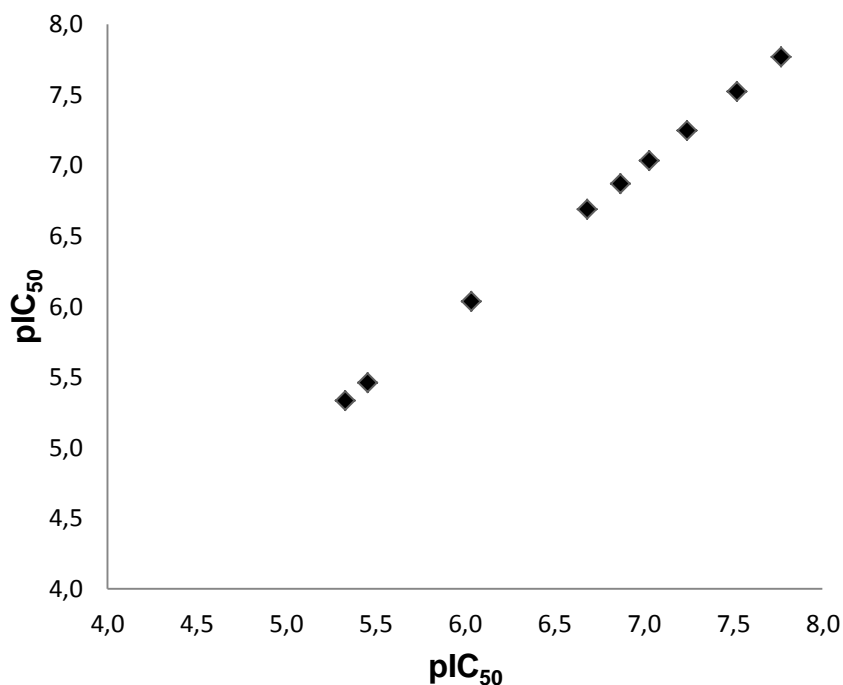


Figura 18. Gráfico da distribuição dos valores de atividade biológica (pIC_{50}) dos 10 compostos (Rosenthal et al. 2010; 2011) do grupo de teste.

3.2 HQSAR

3.2.1 Construção das Estruturas dos Ligantes

As estruturas dos 46 compostos, incluindo os pares de enantiômeros dos compostos **32**, **33** e **34**, foram construídas utilizando a ferramenta *Build* disponível no programa Spartan versão 10 (Wavefunction, Inc.) (DEWAR et al., 1985).

3.2.2 Obtenção dos Modelos de HQSAR

A modelagem por HQSAR foi realizada utilizando o programa SYBYL versão 8.0 (Tripos, Inc.). Durante o desenvolvimento dos modelos, foram utilizados os comprimentos padrões de hologramas (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401 *bins*), mantendo o tamanho de fragmento padrão (4-7 átomos). Em seguida, foi

verificada a influência dos tamanhos dos fragmentos na qualidade dos modelos, avaliando-se o tamanho de 2 a 12 átomos por fragmento. Finalmente, seis tipos de distinções de fragmentos foram combinadas, utilizando átomos (A), ligações (L), conectividade (C), átomos de hidrogênio (H), quiralidade (Q) e átomos doadores/aceptores de ligação hidrogênio (DA) (DE SOUZA et al., 2012).

Os modelos de HQSAR foram gerados por análise de PLS, enquanto a validação interna foi realizada pelo método LOO_{CV} . Subsequentemente, os melhores modelos de HQSAR foram selecionados baseados na análise de diversos parâmetros estatísticos, como o coeficiente de correlação quadrático (r^2), o coeficiente de correlação da validação cruzada (q^2), o erro padrão da estimativa (EP) e o erro padrão da validação cruzada (EP_{VC}). Este procedimento também informa o número ótimo de componentes principais (NC) e o comprimento do holograma utilizado (CH).

3.2.3 Teste de Randomização da Variável Y

Com o objetivo de avaliar uma possível correlação fortuita, foi realizado teste de randomização (ou aleatorização) da variável Y (i.e., variável dependente), também conhecido como randomização (ou aleatorização) da resposta; que consiste em um procedimento de validação adicional onde os valores da atividade biológica são randomizados e a análise por HQSAR é realizada novamente para o mesmo conjunto de treinamento. Caso não haja correlação fortuita, os parâmetros estatísticos obtidos devem ser desfavoráveis (RÜCKER; RÜCKER; MERINGER, 2007).

3.2.4 Validação Externa do Modelo e Detecção de *Outliers*

A validação externa do modelo é realizada com os compostos do conjunto de teste, que não foram utilizados na construção do modelo, com o objetivo de avaliar a capacidade preditiva do mesmo. Esta avaliação consiste no cálculo do coeficiente de correlação quadrático preditivo (r^2_{pred}) definido pela Equação 10.

$$r_{pred}^2 = \frac{DP - PRESS}{DP} \quad (\text{Eq. 10})$$

Na Equação 10: DP é a soma dos desvios quadráticos entre os valores de atividade biológica dos compostos do conjunto de teste e a atividade biológica média dos compostos do conjunto de treinamento e PRESS é a soma dos desvios quadráticos entre os valores de atividade biológica experimentais e os valores preditos pelo modelo (CRAMER; PATTERSON; BUNCE, 1988). O requisito para um composto ser considerado um *outlier* neste trabalho é apresentar um valor residual, ou seja, valor da diferença entre a atividade biológica experimental e a predita, superior a uma unidade logarítmica (SOBHIA; BHARATAM, 2005).

3.3 CoMFA e CoMSIA

3.3.1 Construção e Análise Conformacional das Estruturas dos Ligantes

A construção dos ligantes foi realizada como descrito no item 3.2.1. As geometrias dos compostos foram otimizadas por mecânica molecular, utilizando o campo de força MMFF e, em seguida, pelo método semi-empírico AM1, ambos disponíveis no programa Spartan. As estruturas foram ainda submetidas à análise conformacional sistemática padrão utilizando o método mecânico quântico semi-empírico AM1 também disponível no programa Spartan.

3.3.2 Definição da Hipótese Farmacofórica

A hipótese farmacofórica foi derivada do modo de ligação do composto DJM2005 na DYRK1A (código no PDB: 2WO6 – resolução 2,5 Å) (Figura 19) (SOUNDARARAJAN et al., 2013), um inibidor competitivo com o ATP, que possui alguma semelhança estrutural com o composto **24**, o mais ativo da série. Ambos os compostos possuem um anel heteroaromático nitrogenado central ligado a dois grupos, cada um contendo um anel aromático (Figura 20).

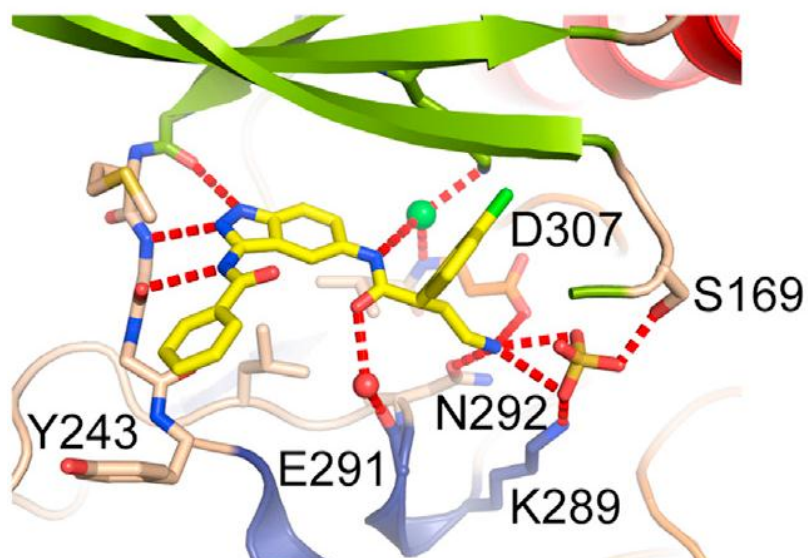
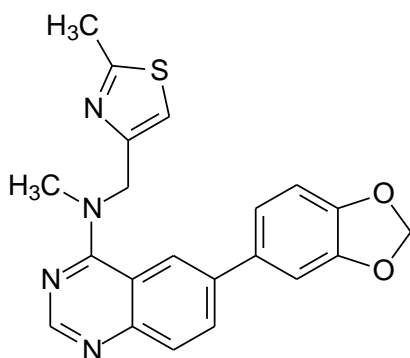
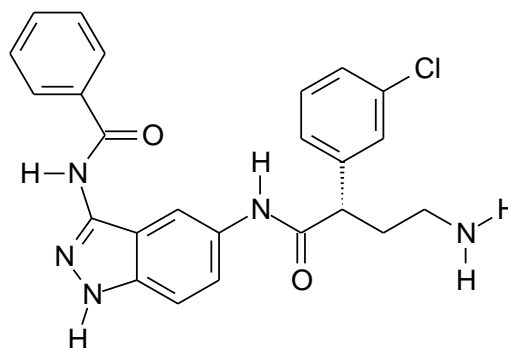


Figura 19. Modelo de ligação do inibidor DJM2005 à DYRK1A humana (Adaptado de SOUNDARARAJAN et al., 2013).



Composto 24



DJM2005

Figura 20. Estruturas químicas dos compostos 24 e DJM2005 usados na definição da hipótese farmacofórica.

3.3.3 Cálculo das Cargas Atômicas Parciais

Em função da contribuição eletrostática, tanto na análise por CoMFA quanto na análise por CoMSIA, é importante verificar o efeito de diferentes abordagens de cálculos de cargas atômicas parciais na qualidade dos modelos obtidos. Para isso, o assinalamento das cargas atômicas parciais da série de compostos foi realizado pelo ajuste das cargas a um potencial eletrostático molecular restrito (*Restrained Molecular*

Electrostatic Potential, RESP), empregando quatro métodos de cálculo de mecânica quântica; os métodos semi-empíricos AM1 e PM3; o método da teoria funcional de densidade (*Density Funcional Theory*, DFT), empregando o funcional B3LYP e o conjunto de bases 6-31G*; e o método *ab initio* Hartree-Fock (HF), usando o conjunto de bases 6-31G*; todos disponíveis no programa Spartan versão 10 (Wavefunction, Inc.) (BRITO, 2008).

3.3.4 Alinhamento da Série de Compostos

Em estudos de CoMFA e CoMSIA, a etapa do alinhamento da série de compostos é considerada crítica, pois as energias de interação entre ligante e proteína dependem da posição espacial relativa entre diferentes grupos funcionais presentes nas moléculas (PETERSON; SCHAAL; KARLÉN, 2006).

Um alinhamento pode ser considerado ótimo quando ocorre um máximo de superposição dos campos eletrostáticos e estéricos de um determinado conjunto de compostos. Assim, somente quando modelos forem gerados com alinhamentos válidos, os seus resultados podem ser considerados pertinentes e significantes (PERKINS et al., 2003).

Os compostos foram exportados para o programa SYBYL versão 8.0 (Tripos, Inc.) e foram alinhados segundo o método *distill rigid*. Este método identifica uma estrutura comum a toda série congênere e baseia-se nesta estrutura comum para alinhar os compostos. A eficiência deste método de alinhamento e sua aplicação de estudos de QSAR-3D já foram descrita em estudos anteriores (PATEL; GHATE, 2014; VYAS et al., 2013).

3.3.5 Definição dos Átomos de Prova

Após o alinhamento, os compostos do conjunto de treinamento foram inseridos em uma caixa tridimensional virtual, regularmente gradeada com espaçamento de 2,0 Å. Assim, em cada vértice das grades da caixa foram calculados os campos moleculares eletrostáticos e estéricos, utilizando três tipos diferentes de átomos de

prova: átomo de carbono com hibridização sp^3 e carga +1; átomo de hidrogênio com carga +1; e átomo de oxigênio com hibridização sp^3 e carga -1, mantendo o valor de corte padrão do programa de 30 kcal.mol⁻¹.

3.3.6 Definição dos Valores de Corte dos Campos Moleculares

Para avaliar a influência de diferentes valores de corte (*cut-off*) dos campos moleculares na qualidade dos modelos de CoMFA, estes valores foram variados em três níveis idênticos para ambos os campos eletrostático e estérico: 30 kcal.mol⁻¹, 20 kcal.mol⁻¹ e 10 kcal.mol⁻¹, enquanto mantiveram-se fixos o método de assinalamento das cargas parciais e o átomo de prova que gerou o melhor resultado.

3.3.7 Definição do Espaçamento da Grade

O melhor espaçamento da grade foi avaliado mantendo-se fixos os parâmetros avaliados anteriormente que geraram as melhores equações durante o processo de otimização dos modelos de CoMFA, e variando-se o espaçamento da grade em três níveis: 1,0 Å, 1,5 Å e 2,0 Å.

3.3.8 Obtenção dos Modelos de CoMFA

Como na análise por HQSAR, os modelos de CoMFA foram obtidos por análise de PLS, enquanto a validação interna foi realizada pelo método LOO_{CV}, com a diferença que, neste método, a análise é realizada duas vezes. A primeira análise é feita juntamente com a validação cruzada, definindo-se o número máximo de componentes principais extraídos como sendo $N/5$, onde N é o número de compostos do grupo de treinamento, com o objetivo de evitar o superajuste do modelo (MELVILLE; HIRST, 2004). Esta análise gera os seguintes parâmetros estatísticos: coeficiente de correlação da validação cruzada (q^2); o erro padrão da validação cruzada (EP_{VC}) e número ótimo de componentes principais (NC); que corresponde ao primeiro mínimo do EP_{VC} . A segunda é realizada sem a validação cruzada e com o NC obtido na primeira análise e

gera os seguintes parâmetros: coeficiente de correlação quadrático (r^2), o erro padrão da estimativa (EP), o valor de Fisher (F) e as porcentagens de contribuição das interações eletrostática (E) e estérica (S). As colunas com variação de energia menor do que $2,0 \text{ kcal.mol}^{-1}$ são automaticamente excluídas do cálculo, o que reduz, significativamente, o tempo da análise sem prejudicar a qualidade do modelo gerado pela mesma.

3.3.9 Teste de Randomização da Variável Y do Melhor Modelo de CoMFA

Nos estudos de QSAR 3D, também é necessário avaliar uma possível correlação fortuita pelo teste de randomização da variável dependente Y. Este teste foi realizado pela randomização dos valores da atividade biológica dos derivados da série e construindo novos modelos de CoMFA para cada conjunto randomizado. Assim como na análise por HQSAR, os índices estatísticos obtidos neste teste devem ser desfavoráveis.

3.3.10 Validação Externa do Modelo de CoMFA e Detecção de *Outliers*

Os procedimentos de validação externa do modelo e de identificação de *outliers* foram realizados de forma análoga a da análise de HQSAR descrita no item 3.2.4.

3.3.11 Obtenção dos Modelos de CoMSIA

A metodologia de CoMSIA foi aplicada ao mesmo conjunto de dados do estudo de CoMFA, utilizando alinhamento, nível de teoria do assinalamento de cargas parciais, átomo de prova e espaçamento da grade idênticos aos que geraram o melhor modelo de CoMFA. Contudo, como a análise de CoMSIA permite a representação de campos moleculares adicionais, além do eletrostático e do estérico, a mesma foi executada utilizando cinco campos moleculares: eletrostático (E), estérico (S), hidrofóbico (H), acceptor de ligação hidrogênio (A) e doador de ligação hidrogênio (D).

Além da utilização dos cinco campos moleculares em conjunto para gerar modelos de QSAR, combinações entre 2, 3 e 4 índices foram testadas para gerar outros modelos. A razão para isso é a possibilidade da substituição de um grupo funcional de um ligante por outro que possa afetar os índices de similaridade de forma semelhante; o que geraria modelos com informações redundantes (BÖHM; ST RZEBECHER; KLEBE, 1999); por exemplo, a substituição de um grupo funcional carregado por um não carregado afeta os índices eletrostático e hidrofóbico. Após a geração do modelo de CoMSIA com os melhores parâmetros, investigou-se a influência do fator de atenuação (α) na qualidade dos modelos, para isso, foram avaliados três valores de α : 0,2, 0,3 e 0,4.

3.3.12 Teste de Randomização da Variável Y do Melhor Modelo de CoMSIA

O teste de randomização-Y realizado para o melhor modelo de CoMSIA foi idêntico ao realizado para o melhor modelo de CoMFA descrito no item 3.3.9.

3.3.13 Validação Externa do Modelo de CoMSIA e Detecção de *Outliers*

Os procedimentos de validação externa do modelo e de identificação de *outliers* foram realizados de forma análoga às das análises de HQSAR e CoMFA descritas nos itens 3.2.4 e 3.3.9, respectivamente.

3.4 DOCAGEM MOLECULAR

3.4.1 Preparo dos Ligantes e da Proteína

Todas as simulações de docagem molecular foram realizadas no programa *Molegro Virtual Docker* (MVD) versão 2008.2.4 (Molegro ApS), devido ao melhor desempenho global de simulações de re-docagem e à maior velocidade de obtenção de resultados utilizando este programa quando comparado com outros programas de docagem molecular como o Autodock (ARAÚJO et al., 2011).

Para as simulações realizadas, todos os tipos de átomos e as ordens de ligação foram corrigidos para as estruturas da enzima e dos ligantes, empregando o método de preparação automático do programa (THOMSEN; CHRISTENSEN, 2006). Também foram adicionados os átomos de hidrogênio em todas as estruturas. As cargas parciais dos átomos dos ligantes utilizados na validação do protocolo de docagem e na docagem propriamente dita foram assinaladas pelo método semi-empírico AM1 do programa Spartan. Já as cargas parciais dos átomos da enzima foram assinaladas pelo MVD. É importante ressaltar que, em todas as simulações, as ligações rotacionáveis dos ligantes foram mantidas livres, a proteína foi considerada fixa, e todas as moléculas de água, cofatores e íons foram excluídos dos complexos. As cavidades da enzima foram identificadas pelo emprego de um algoritmo baseado em *grid* disponível no programa (ARAÚJO et al., 2011).

3.4.2 Validação do Protocolo de Docagem (Re-docagem)

Nesta etapa, é importante identificar qual das combinações entre algoritmos de busca e funções de pontuação se adequa melhor ao sistema. Como não há um ligante da série congênere cristalizado com a DYRK1A, foram obtidas no *Protein Data Bank* (PDB) estruturas tridimensionais da enzima co-cristalizadas com ligantes que se adequavam aos seguintes critérios: resolução menor que 2,50 Å e número de ligações rotacionáveis entre 4 e 8, mesma faixa de liberdade conformacional dos ligantes da série das 6-arilquinazolina-4-aminas.

Nesta avaliação, foram combinados os algoritmos de busca *MolDock Optimizer*, *MolDock SE* e *Iterated Simplex* com as funções de pontuação *MolDock Score* e *PLANTS Score*. A única exceção foi a combinação entre o algoritmo *MolDock SE* e a função *PLANTS Score* que não é possível de ser realizada no programa.

Os parâmetros das simulações utilizados nesta simulação foram os seguintes: a população de partida utilizada foi de 50 para os algoritmos *MolDock Optimizer* e *MolDock SE* e de 20 para o algoritmo *Iterated Simplex*; o número máximo de iterações foi 2000 para o algoritmo *MolDock Optimizer*, 1500 para o *MolDock SE* e 100 para o *Iterated Simplex*; todos valores padrão do programa. O número de corridas

independentes foi de 40 para todas as combinações, cada uma retornando uma única solução.

No processo de avaliação do melhor protocolo, busca-se uma combinação que gere soluções mais próximas possíveis da estrutura experimental. Esta similaridade é uma medida da distância entre a solução gerada pelo programa e a estrutura do complexo ligante-enzima experimental. Esta distância é calculada pela raiz do desvio médio quadrático (*root-mean-square deviation*, RMSD), entre dois conjuntos de coordenadas atômicas, neste caso, um do complexo ligante-proteína cristalizado e outra das coordenadas das soluções obtidas pelo método de docagem, segundo a Equação 11.

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{c,i} - x_{p,i})^2 + (y_{c,i} - y_{p,i})^2 + (z_{c,i} - z_{p,i})^2} \quad (\text{Eq. 11})$$

Na Equação 11, $x_{c,i}$, $y_{c,i}$ e $z_{c,i}$ são as coordenadas espaciais do átomo i no ligante cristalizado, $x_{p,i}$, $y_{p,i}$ e $z_{p,i}$ são as coordenadas espaciais do átomo i na solução resultante do processo de docagem, e n é o número de átomos do ligante. Um resultado pode ser considerado aceitável quando o valor de RMSD é menor do que 2,0 Å (FRIESNER et al., 2004) e a combinação a ser escolhida é a que retorne o maior número de soluções possíveis com valores de RMSD aceitáveis.

3.4.3 Docagem da Série de 6-arilquinazolina-4-aminas

Com a definição da melhor combinação de algoritmo de busca e função de pontuação, os 46 ligantes foram submetidos ao processo de docagem e a enzima utilizada foi a de código do PDB 4MQ1, pois é a estrutura com melhor resolução (2,35 Å). Os parâmetros foram idênticos aos utilizados no processo de re-docagem, entretanto o número de poses retornadas foi de cinco.

A análise pormenorizada das soluções com as melhores pontuações foi realizada para os compostos mais e menos ativos da série, os *outliers* dos estudos de QSAR e os compostos com centros de quiralidade. Foram analisadas interações eletrostáticas,

hidrofóbicas (e.g., van der Waals e empilhamento- π) e ligações hidrogênio. A apresentação dos resultados da docagem e a construção de superfícies hidrofóbica e de ligação hidrogênio da proteína foram realizadas no programa *Discovery Studio Visualizer* versão 4.1 (*Accelrys Software Inc*), devido à sua qualidade gráfica na apresentação dos dados.

4 RESULTADOS E DISCUSSÃO

4.1 HQSAR

4.1.1 Avaliação do Parâmetro Distinção de Fragmentos na Qualidade dos Modelos de HQSAR

Nesta primeira avaliação, foram investigadas diferentes combinações de distinção de fragmentos, mantendo constante o tamanho dos fragmentos (4 – 7 átomos) e selecionando todos os valores padrões de comprimento de hologramas disponíveis. O número máximo de componentes principais foi definido em 15, um valor menor do que a metade do número de compostos do grupo de treinamento e a análise gerou um total de 16 modelos. A influência deste parâmetro na qualidade dos modelos de HQSAR está indicada na Tabela 3.

De acordo com a Tabela 3, todos os modelos de HQSAR gerados são aceitáveis, pois o menor valor de q^2 é 0,640. Para diminuir o número de modelos analisados, foram considerados somente os modelos com q^2 maiores do que 0,730. Desta forma, foram selecionados quatro modelos: **A/L/C/Q/DA** ($q^2=0,743$), **A/C/Q/DA** ($q^2=0,742$), **C/H** ($q^2=0,740$), e **A/L** ($q^2=0,732$); os quais foram avaliados para investigar a influência do tamanho dos fragmentos na qualidade dos modelos.

Tabela 3. Resumo dos índices estatísticos dos modelos de HQSAR para as diversas combinações de distinção de fragmentos (DF), utilizando o tamanho padrão de fragmentos (4-7 átomos) para os derivados da série das 6-arylquinazolina-4-aminas (N=36). Os quatros melhores modelos estão assinalados em negrito.

DF ^b	Índices Estatísticos ^a					
	q ²	r ²	EP	EP _{vc}	NC	CH
A/L	0,732	0,847	0,373	0,493	3	61
A/C	0,728	0,799	0,421	0,489	2	353
A/H	0,640	0,782	0,444	0,571	3	199
A/DA	0,697	0,896	0,323	0,551	6	59
L/C	0,711	0,841	0,380	0,512	3	53
L/H	0,727	0,824	0,400	0,498	3	59
C/H	0,740	0,801	0,419	0,478	2	353
C/DA	0,720	0,834	0,394	0,512	4	61
A/L/C	0,724	0,855	0,323	0,500	3	53
A/L/H	0,670	0,781	0,446	0,547	3	401
A/C/H	0,656	0,818	0,413	0,567	4	401
A/C/DA	0,721	0,842	0,394	0,511	4	61
L/C/Q	0,711	0,841	0,380	0,512	3	53
A/L/C/H	0,691	0,777	0,443	0,521	2	353
A/C/Q/DA	0,742	0,876	0,341	0,491	4	257
A/L/C/Q/DA	0,743	0,917	0,284	0,498	5	53

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; CH, comprimento do holograma; ^b DF, Parâmetros de distinção de fragmentos: átomos (A); ligações (L); conexões (C); quiralidade (Q); átomos de hidrogênio (H) e átomos doadores/aceptores de ligação hidrogênio (DA).

4.1.2 Avaliação do Tamanho dos Fragmentos na Qualidade dos Modelos de HQSAR

Com o objetivo de melhorar os modelos construídos, oito novos modelos foram gerados a partir de cada um dos quatro melhores modelos selecionados anteriormente, considerando diferentes faixas de tamanhos de fragmentos, contendo entre 2 e 12

átomos, com variação de quatro unidades cada (2-5, 3-6, 4-7, 5-8, 6-9, 7-10, 8-11 e 9-12 átomos). A variação do parâmetro de tamanho dos fragmentos não foi capaz de melhorar significativamente os modelos com as seguintes distinções de fragmentos: **A/C/Q/DA** e **C/H**. Entretanto, os modelos com distinções de fragmentos **A/L/C/Q/DA** e **A/L** obtiveram uma melhora significativa e seus índices estatísticos estão mostrados nas Tabelas 4 e 5, respectivamente.

Tabela 4. Resumo dos índices estatísticos dos modelos de HQSAR para as diversas combinações de tamanho de fragmentos (TF), utilizando o parâmetro de distinção de fragmentos **A/L/C/Q/DA** para os derivados da série das 6-arylquinazolina-4-aminas (N=36). O melhor modelo está assinalado em negrito.

TF ^b	Índices Estatísticos ^a					
	q ²	r ²	EP	EP _{Vc}	NC	CH
2-5	0,734	0,855	0,362	0,491	3	401
3-6	0,757	0,937	0,251	0,493	6	53
4-7	0,743	0,917	0,284	0,498	5	53
5-8	0,751	0,883	0,331	0,483	4	53
6-9	0,738	0,871	0,347	0,496	4	61
7-10	0,732	0,920	0,282	0,518	6	53
8-11	0,681	0,906	0,302	0,556	5	151
9-12	0,642	0,804	0,421	0,570	3	151

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{Vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; CH, comprimento do holograma; ^b TF, Tamanho dos fragmentos, de 2 a 12 átomos.

Tabela 5. Resumo dos índices estatísticos dos modelos de HQSAR para as diversas combinações de tamanho de fragmentos (TF), utilizando o parâmetro de distinção de fragmentos **A/L** para os derivados da série das 6-arilquinazolina-4-aminas (N=36). O melhor modelo está assinalado em negrito.

TF ^b	Índices Estatísticos ^a					
	q ²	r ²	EP	EP _{VC}	NC	HL
2-5	0,737	0,848	0,372	0,488	3	61
3-6	0,717	0,858	0,359	0,507	3	83
4-7	0,732	0,847	0,373	0,493	3	61
5-8	0,713	0,839	0,382	0,510	3	61
6-9	0,719	0,848	0,377	0,513	4	61
7-10	0,748	0,847	0,372	0,478	3	199
8-11	0,724	0,848	0,371	0,500	3	401
9-12	0,705	0,829	0,394	0,517	3	83

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{VC}, erro padrão da validação cruzada; NC, número ótimo de componentes; CH, comprimento do holograma; ^b TF, Tamanho dos fragmentos, de 2 a 12 átomos.

O melhor modelo com o parâmetro de distinção de fragmentos **A/L/C/Q/DA** contém de 3 a 6 átomos por fragmento (Tabela 3), enquanto o melhor modelo com o parâmetro de distinção de fragmentos **A/L** contém de 7 a 10 átomos por fragmento (Tabela 4). É importante ressaltar que o modelo com maior valor de q² avaliado contém cinco parâmetros de distinção de fragmentos (**A/L/C/Q/DA**) e um tamanho de fragmento de 3 a 6 átomos, o que significa que a atividade biológica desta série de compostos parece ser mais bem explicada por um modelo mais diversificado e com fragmentos de tamanho reduzido, logo, a remoção de algum destes parâmetros de distinção resulta em perda significativa de informação.

4.1.3 Teste de Randomização-Y

O teste de randomização-Y foi realizado de forma a avaliar a robustez dos melhores modelos obtidos. Neste teste, os valores das atividades biológicas foram

randomizados e novas análises de HQSAR foram feitas. De acordo com a Tabela 6, todos os modelos obtidos por este teste são estatisticamente ruins (o maior valor de q^2 foi 0,211), o que reforça a robustez dos modelos originais, visto que a probabilidade de que a correlação original tenha sido obtida ao acaso é baixa.

Tabela 6. Resumo dos índices estatísticos dos modelos de HQSAR do teste de randomização-Y, utilizando o tamanho padrão de fragmentos (4-7 átomos) para os derivados da série das 6-irilquinazolina-4-aminas (N=36).

DF ^b	Índices Estatísticos ^a					
	q^2	r^2	EP	EP _{vc}	NC	CH
A/L	0,143	0,396	0,694	0,827	2	353
A/C	0,117	0,722	0,502	0,895	6	59
A/H	0,058	0,381	0,703	0,867	2	199
A/DA	0,113	0,586	0,593	0,868	4	59
L/C	0,062	0,183	0,795	0,852	1	53
L/H	0,041	0,824	0,400	0,498	3	59
C/H	0,055	0,264	0,756	0,868	2	401
C/DA	0,089	0,202	0,785	0,840	1	53
A/L/C	0,211	0,713	0,510	0,846	6	61
A/L/H	0,044	0,351	0,719	0,873	2	401
A/C/H	0,045	0,359	0,715	0,872	2	353
A/C/DA	0,098	0,215	0,779	0,835	1	71
B/C/Q	0,062	0,183	0,794	0,852	1	53
A/B/C/H	0,051	0,314	0,739	0,870	2	257
A/C/Q/DA	0,106	0,222	0,776	0,832	1	71
A/B/C/Q/DA	0,099	0,235	0,770	0,835	1	151

^a q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; CH, comprimento do holograma; ^b DF, Parâmetros de distinção de fragmentos: átomos (A); ligações (L); conexões (C); quiralidade (Q); átomos de hidrogênio (H) e átomos doadores/aceptores de ligação hidrogênio (DA).

4.1.4 Validação Externa do Modelo e Detecção de *Outliers*

A capacidade preditiva dos dois modelos selecionados foi avaliada utilizando os compostos do conjunto de teste pelo cálculo do parâmetro r^2_{pred} . Os valores das atividades biológicas experimentais (Exp), preditas (Pred) e os resíduos (Exp – Pred) dos derivados da classe das 6-arilquinazolina-4-aminas obtidos pelos melhores modelos de HQSAR com os parâmetros de distinção **AL/C/Q/DA** e **AL** são mostrados nas Tabelas 7 e 8, respectivamente. Os gráficos comparativos entre os valores de Exp e Pred dos grupos de treinamento e de teste com os parâmetros de distinção **AL/C/Q/DA** e **AL** são mostrados nas Figuras 21 e 22, respectivamente

Tabela 7. Valores da atividade biológica experimental (Exp), predita (Pred) e resíduos (Res) dos compostos da classe das 6-arylquinazolina-4-aminas, utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **A/L/C/Q/DA**. Os compostos do conjunto de teste estão marcados com um asterisco.

#	Exp	Pred	Res	#	Exp	Pred	Res
1	7,21	6,86	0,35	26	7,29	7,36	-0,07
2	5,90	5,52	0,38	27	7,59	7,38	0,21
3*	5,46	5,26	0,20	28	6,81	6,79	0,02
4	5,24	5,53	-0,29	29*	6,04	6,78	-0,74
5	5,50	5,45	0,05	30	6,27	6,25	0,02
6	5,05	4,98	0,07	31	6,92	6,76	0,16
7	6,79	6,82	-0,03	32 (R)*	7,03	7,13	-0,10
8	5,35	5,35	0,00	32 (S)*	7,03	7,25	-0,22
9	6,74	6,81	-0,07	33 (R)*	6,87	6,99	-0,12
10	5,84	5,89	-0,05	33 (S)*	6,87	6,98	-0,11
11*	5,33	6,14	-0,81	34 (R)*	7,52	7,20	0,32
12	7,51	7,26	0,25	34 (S)*	7,52	6,97	0,55
13	7,42	7,30	0,12	35	7,12	7,09	0,03
14	5,94	6,39	0,45	36*	7,77	6,93	0,84
15	6,59	7,33	-0,74	37	5,94	5,98	0,04
16	7,46	6,97	0,49	38	6,25	6,00	0,25
17	7,08	7,16	-0,08	39	6,23	6,20	0,03
18	7,01	7,08	-0,07	40	5,44	5,43	0,01
19	7,13	7,03	0,10	41	5,47	5,25	0,22
20*	7,24	6,78	0,46	42	5,82	5,19	0,63
21	6,90	6,83	0,07	43*	5,85	5,66	0,19
22	7,03	7,15	-0,12	44	5,57	5,06	0,51
23*	6,69	6,56	0,13	45	5,31	5,19	0,13
24	7,85	7,92	-0,07	46	5,08	4,73	0,35
25	7,15	7,28	-0,13				

Tabela 8. Valores da atividade biológica experimental (Exp), predita (Pred) e resíduos (Res) dos compostos da classe das 6-arylquinazolina-4-aminas, utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **A/L**. Os compostos do conjunto de teste estão marcados com um asterisco.

#	Exp	Pred	Res	#	Exp	Pred	Res
1	7,21	6,89	0,32	26	7,29	7,25	0,04
2	5,90	5,35	0,55	27	7,59	7,30	0,29
3*	5,46	5,53	-0,07	28	6,81	6,57	0,24
4	5,24	5,63	-0,39	29*	6,04	6,98	-0,94
5	5,50	5,89	-0,39	30	6,27	6,88	-0,61
6	5,05	5,19	-0,14	31	6,92	7,08	-0,16
7	6,79	7,07	-0,28	32 (R)*	7,03	7,10	-0,07
8	5,35	5,00	0,35	32 (S)*	7,03	7,10	-0,07
9	6,74	6,68	0,06	33 (R)*	6,87	7,13	-0,26
10	5,84	5,39	0,45	33 (S)*	6,87	7,13	-0,26
11*	5,33	5,60	-0,27	34 (R)*	7,52	6,93	0,59
12	7,51	7,32	0,19	34 (S)*	7,52	6,93	0,59
13	7,42	7,45	-0,03	35	7,12	7,13	-0,01
14	5,94	6,84	-0,90	36*	7,77	7,25	0,52
15	6,59	7,35	-0,76	37	5,94	5,81	0,13
16	7,46	6,91	0,55	38	6,25	6,05	0,20
17	7,08	6,82	0,26	39	6,23	6,20	0,03
18	7,01	7,26	-0,25	40	5,44	5,36	0,08
19	7,13	6,99	0,14	41	5,47	5,19	0,28
20*	7,24	6,97	0,27	42	5,82	5,41	0,41
21	6,90	6,89	0,01	43*	5,85	5,39	0,46
22	7,03	6,85	0,18	44	5,57	5,07	0,50
23*	6,69	6,70	-0,01	45	5,31	5,16	0,16
24	7,85	7,24	0,61	46	5,08	5,21	-0,13
25	7,15	6,83	0,32				

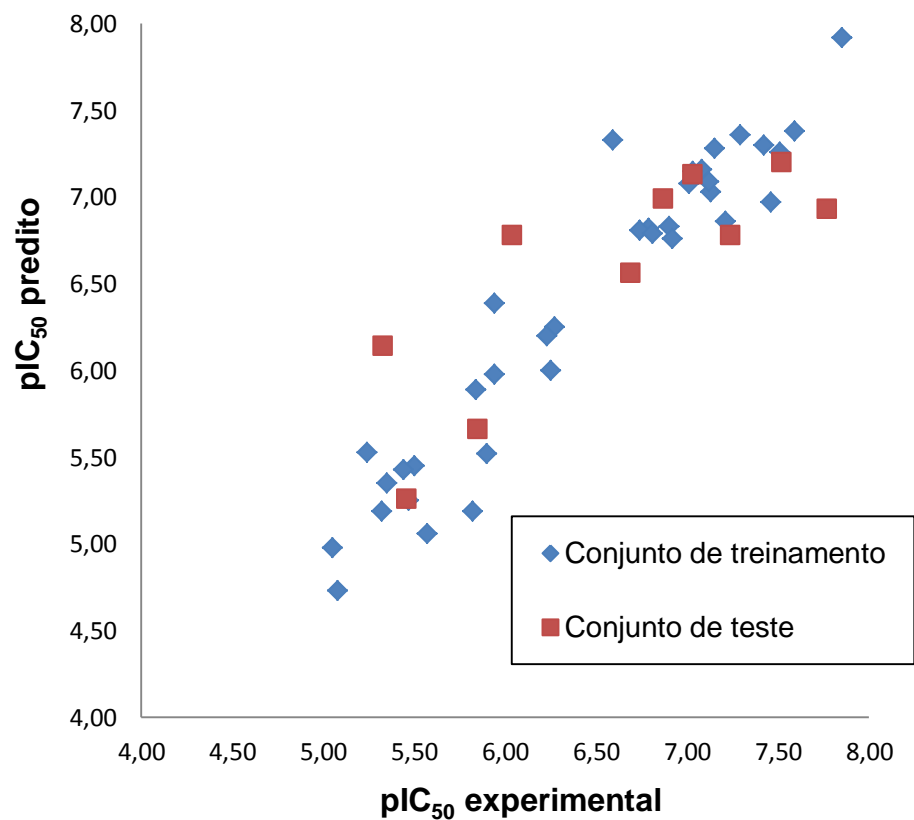


Figura 21. Gráfico dos valores de pIC_{50} preditos *versus* os valores de pIC_{50} experimentais para os grupos de treinamento (azul) e de teste (vermelho), utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **A/L/C/Q/DA**.

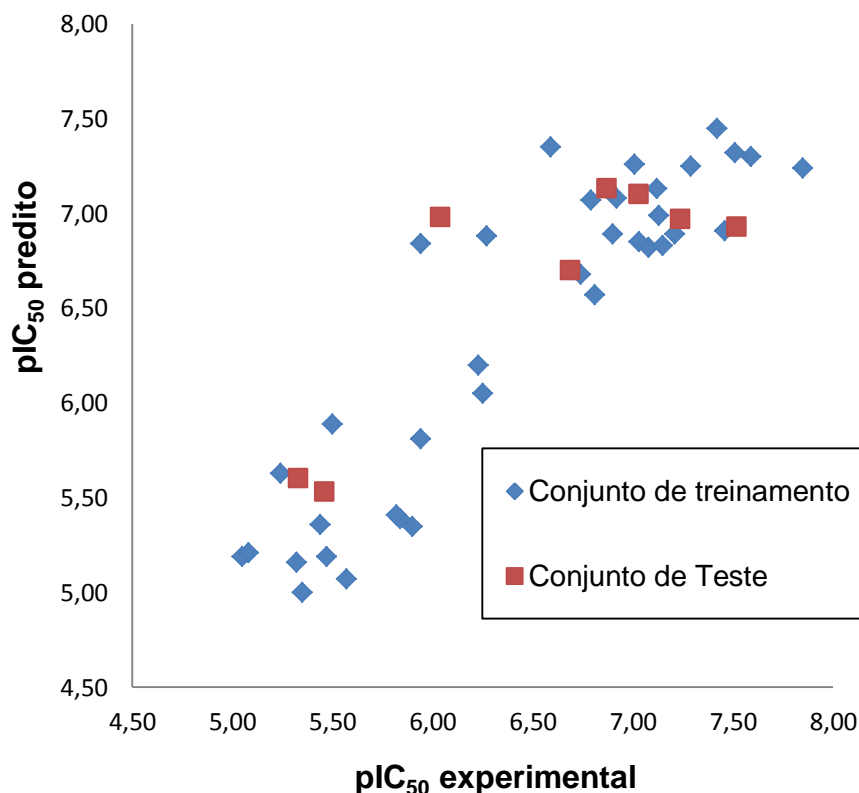


Figura 22. Gráfico dos valores de pIC₅₀ preditos *versus* os valores de pIC₅₀ experimentais para os grupos de treinamento (azul) e de teste (vermelho), utilizando o melhor modelo de HQSAR com o parâmetro de distinção de fragmentos **A/L**.

Ambos os modelos não contêm *outliers*, pois em nenhum deles surgiu algum composto com valor de resíduo maior do que uma unidade logarítmica. Entretanto, o desvio padrão (DP) dos valores residuais do modelo com o parâmetro de distinção de fragmentos **A/L/C/Q/DA** (DP = 0,322) é menor do que o do modelo com o parâmetro de distinção de fragmentos **A/L** (DP = 0,379), indicando que os valores de pIC₅₀ preditos pelo primeiro modelo estão mais próximos dos valores experimentais. Os valores calculados de r^2_{pred} para os modelos **A/L/C/Q/DA** ($r^2_{\text{pred}} = 0,659$) e **A/L** ($r^2_{\text{pred}} = 0,743$) são maiores do que 0,5, indicando que ambos os modelos possuem capacidade preditiva aceitável.

Os compostos contendo um centro de quiralidade (**32**, **33** e **34**), modelados nas suas duas formas enantioméricas (*R* e *S*), apresentaram valores de pIC₅₀ idênticos ou muito próximos, independente do modelo ou do enantiômero, indicando que estes

centros de quiralidade não possuem relevância no estudo da relação entre estrutura e atividade destes compostos, segundo o método HQSAR.

4.1.5 Análise dos Mapas de Contribuição

A análise abrangente de HQSAR envolve a interpretação dos mapas de contribuição, que consistem em diagramas coloridos onde as contribuições positivas para a atividade são representadas em amarelo e verde, as contribuições neutras, em branco e as contribuições negativas, em laranja e vermelho. A Figura 23 mostra os mapas de contribuição para o composto mais ativo (**24**) e para o menos ativo (**6**), segundo os melhores modelos **A/L/C/Q/DA** (Figura 23A) e **A/L** (Figura 23B). Neste diagrama, a cor azul representa a parte da estrutura química comum a todos os compostos da série.

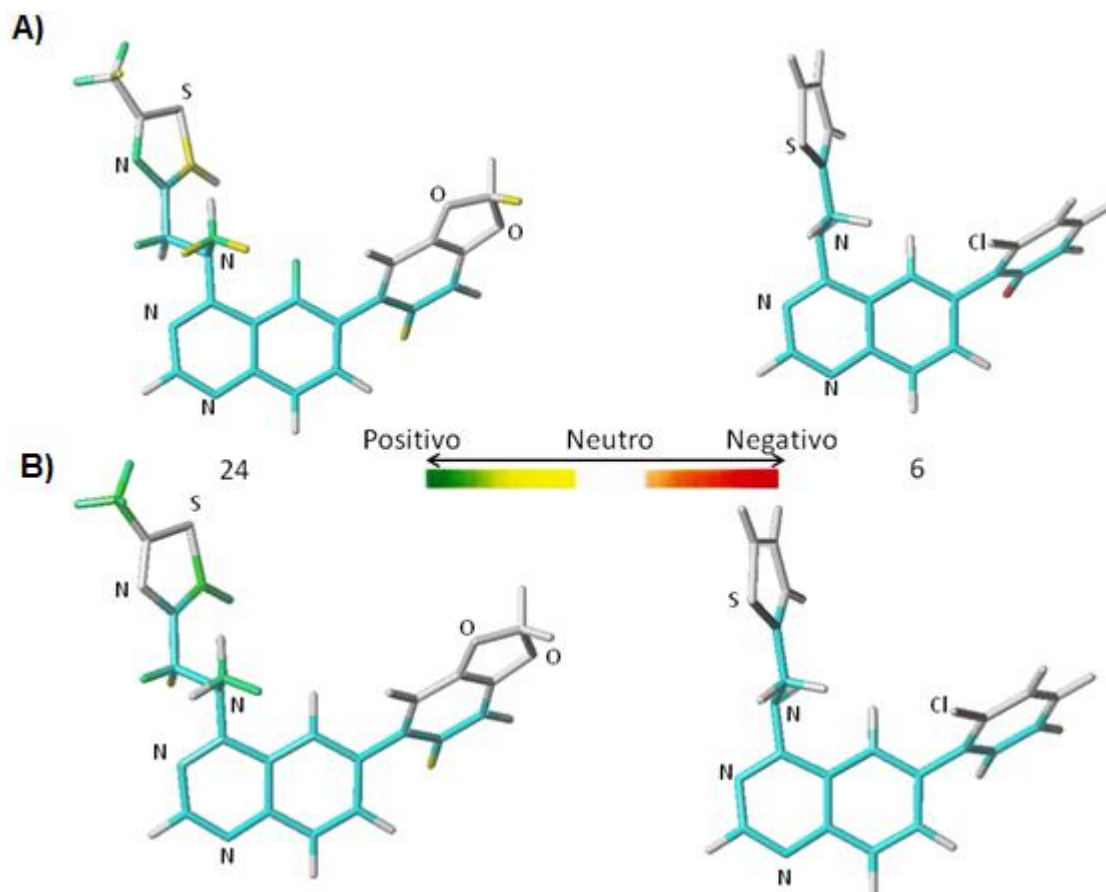


Figura 23. Mapas de contribuição para os compostos mais ativo (**24**) e menos ativo (**6**) segundo os melhores modelos **A/L/C/Q/DA** (A) e **A/L** (B). As estruturas dos compostos estão representadas em modelo bastão e apenas os heteroátomos estão representados pela letra do elemento químico (Cl, N, O, S). O espectro do amarelo ao verde indica que o fragmento tem contribuição positiva para a atividade biológica; o espectro do laranja ao vermelho indica que o fragmento tem contribuição positiva para a atividade biológica; a cor branca indica que o fragmento tem contribuição neutra para a atividade biológica e a cor azul representa a estrutura comum a todos os compostos da série.

Considerando apenas os mapas de contribuição do composto **24** (mais ativo, Figura 23), ambos os modelos foram capazes de identificar fragmentos que contribuem positivamente para a atividade biológica, já que em ambos existem fragmentos coloridos em amarelo e verde. Entretanto, o mesmo não ocorreu para o composto **6** (menos ativo, Figura 23), onde apenas o modelo **A/L/C/Q/DA** foi capaz de identificar fragmentos que diminuem a atividade biológica, visto que apenas no mapa de contribuição deste modelo existe um fragmento colorido em vermelho. No mapa obtido pelo modelo **A/L**, somente aparecem fragmentos em branco (contribuição neutra) e azul

(estrutura comum), que não se correlacionam com a atividade biológica. Conseqüentemente, o modelo **A/L/C/Q/DA** parece ser o mais capaz de fazer distinções entre os compostos mais e menos ativos e, portanto, seria o modelo mais útil do ponto de vista da química medicinal.

Uma característica adicional, observada apenas no mapa de contribuição do composto **24** do modelo **A/L/C/Q/DA** (Figura 23A) é a presença de um fragmento colorido em verde, que corresponde ao átomo de nitrogênio do anel tiazolila (substituinte R₃, Tabela 2). Como apenas este modelo possui o parâmetro de distinção de fragmentos relacionado à presença de átomos doadores ou aceptores de ligação hidrogênio, essa característica salienta a importância deste átomo como aceptor de ligação hidrogênio em uma possível interação entre este composto e a DYRK1A, e reforça o modelo **A/L/C/Q/DA** como sendo, de fato, o melhor; e, portanto, o único a ser discutido a partir deste ponto.

O mapa de contribuição do composto **24**, de acordo com o melhor modelo de HQSAR, identifica três substituintes chamados R₁, R₂ e R₃ (Tabela 2), que possuem influência significativa na atividade biológica desta série de compostos. O anel 1,3-benzo-dioxola (R₁), o grupo metila (R₂) e o grupo tiazolila (R₃) estão presentes na maioria dos compostos mais ativos, como os compostos **24**, **26** e **27**. De fato, todos estes grupos possuem fragmentos (pelo menos um átomo) coloridos em verde ou amarelo, salientando suas contribuições positivas para a atividade biológica.

O mapa de contribuição do composto **6**, de acordo com o melhor modelo de HQSAR, identifica um átomo colorido em vermelho localizado no grupo *orto*-cloro-fenila (R₁), que é prejudicial à atividade biológica, provavelmente, pelo fato do átomo de cloro na posição *orto* impedir uma maior co-planaridade entre os anéis aromáticos, uma característica possivelmente importante para a interação ligante-proteína. Além disso, a presença de um fragmento colorido em vermelho e a ausência de fragmentos coloridos em verde ou amarelo explicam a baixa atividade biológica do composto **6**, assim como, a substituição do grupo metila (R₂) por hidrogênio e a substituição do grupo tiazolila (R₃) pelo grupo tiofenila.

Alguns destes resultados estão de acordo com os apresentados em um estudo de modelagem por QSAR 3D, utilizando esta mesma série de derivados 6-

arilquinazolina-4-aminas (PAN; WANG; BRYANT, 2013). Neste estudo, foi observado que a atividade inibitória da DYRK1A aumenta quando R_1 é um grupo fenila com substituintes hidrofílicos e retiradores de elétrons, R_3 é um anel heterocíclico com um substituinte hidrofóbico e R_2 um substituinte volumoso e hidrofóbico. Por outro lado, a atividade inibitória da DYRK1A decresce quando R_2 é um átomo de hidrogênio, e R_1 e R_2 são grupos aromáticos hidrofóbicos (PAN; WANG; BRYANT, 2013).

4.2 CoMFA e CoMSIA

4.2.1 Análise do Alinhamento dos Compostos

A estrutura comum calculada pelo método *distill rigid* (Figura 24A) corresponde ao anel quinazolina com o grupo amino na posição 4, seguido por dois átomos de carbono (R_3) e o primeiro átomo de carbono do grupo R_1 na posição 6. Analisando a sobreposição da série congênere (Figura 24B), é possível observar uma excelente sobreposição do anel quinazolina de todos os 46 derivados e, também, uma boa sobreposição dos anéis aromáticos em R_1 e dos grupos metila e hidrogênio em R_2 . Os substituintes em R_3 apresentam uma sobreposição mais distante da ideal em função da grande diversidade estrutural de grupos nessa posição. Estes resultados indicam que o alinhamento pelo método *distill rigid* foi capaz de sobrepor adequadamente os derivados da série de 6-arilquinazolina-4-aminas.

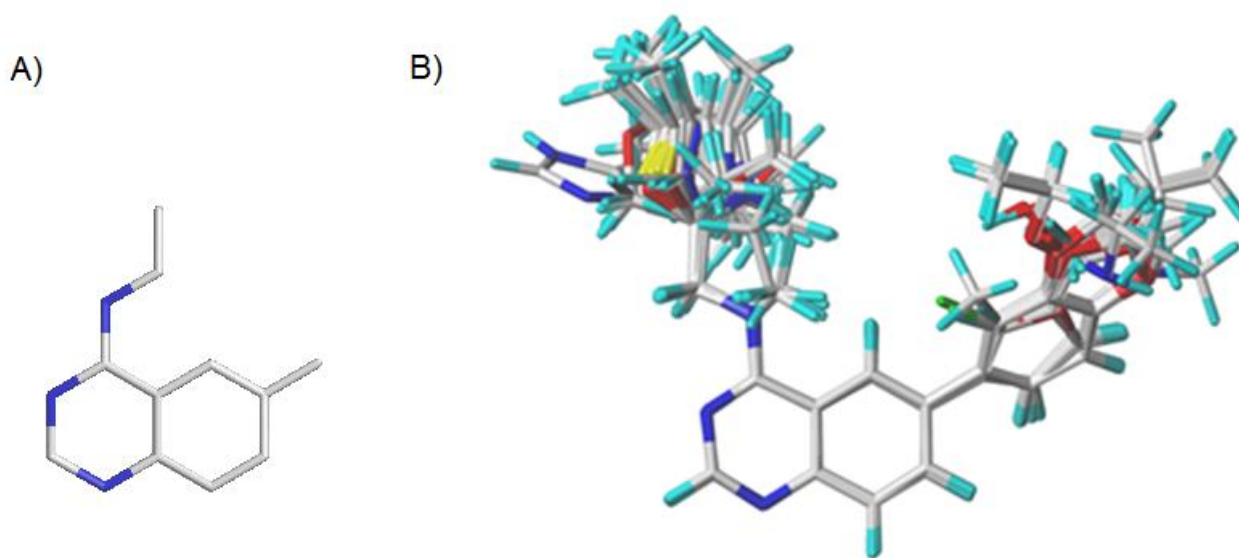


Figura 24. (A) Estrutura comum calculada pelo método *distill rigid* para a série das 6-arylquinazolina-4-aminas. (B) Sobreposição dos 46 derivados.

4.2.2 Avaliação do Melhor Método de Cálculo das Cargas Parciais

Posteriormente ao alinhamento, investigou-se qual método empregado para o cálculo das cargas atômicas parciais geraria um melhor método de CoMFA, mantendo-se constantes o átomo de prova (carbono sp^3 e carga +1), o nível de corte de energia (30 kcal.mol^{-1}) e o espaçamento da grade ($2,0 \text{ \AA}$). A Tabela 9 mostra os parâmetros estatísticos obtidos nesta primeira análise.

Tabela 9. Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-arilquinazolina-4-aminas em função do método de cálculo das cargas parciais (N=36). Átomo de prova ($Csp^3 +1$), valor de corte de energia (30 kcal.mol⁻¹), espaçamento da grade (2,0 Å).

Método ^b	Índices Estatísticos ^a							
	q ²	r ²	EP	EP _{vc}	NC	F	S(%)	E(%)
AM1	0,576	0,816	0,366	0,556	2	73280	40,3	59,7
PM3	0,568	0,812	0,370	0,561	2	71193	38,1	61,9
DFT	0,555	0,808	0,373	0,569	2	69601	40,2	59,8
HF	0,550	0,815	0,367	0,572	2	72765	43,3	56,7

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher; S, porcentagem da contribuição estérica; E, porcentagem da contribuição eletrostática. ^b AM1, método semi-empírico AM1; PM3, método semi-empírico PM3; DFT, método da teoria funcional de densidade empregando o funcional B3LYP e o conjunto de bases 6-31G*; HF, método *ab initio* Hartree-Fock usando o conjunto de bases 6-31G*.

É possível observar na Tabela 9 que os maiores valores de q² e r² foram obtidos quando as cargas parciais foram calculadas pelo método semi-empírico AM1. Este mesmo modelo apresentou os menores valores de erro padrão, com e sem a validação cruzada, resultando no melhor desempenho preditivo entre os quatro modelos avaliados. Segundo os valores de r² encontrados, pode-se afirmar que todos os modelos são capazes de explicar mais de 80% da variabilidade da atividade biológica com dois componentes principais.

Em relação às porcentagens das contribuições estéricas e eletrostáticas e, observa-se pouca diferença entre os modelos, visto que todos apresentam cerca de 40% de contribuição estérica e cerca de 60% de contribuição eletrostática. A maior contribuição eletrostática pode ser devida à presença de heteroátomos mais eletronegativos, como oxigênio e nitrogênio, na estrutura dos derivados que podem se ligar à enzima via este tipo de interação intermolecular.

Ao final desta análise, o modelo utilizando o método AM1 para o cálculo das cargas parciais foi escolhido como aquele a ser utilizado na continuidade das análises dos impactos da variação de parâmetros de cálculo dos modelos de CoMFA.

4.2.3 Avaliação do Melhor Átomo de Prova

A segunda avaliação envolveu a análise do impacto do átomo de prova na qualidade dos modelos de CoMFA. Para isso, foram mantidos o espaçamento da grade (2,0 Å) e o nível de corte de energia (30 kcal.mol⁻¹) e utilizou-se o método de cálculos de cargas parciais AM1, identificado como o melhor método na etapa anterior, variando apenas os átomos de prova: carbono *sp*³ com carga +1, hidrogênio com carga +1 e oxigênio *sp*³ com carga -1. Os índices estatísticos obtidos nesta etapa da análise são mostrados na Tabela 10.

Tabela 10. Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-irilquinazolina-4-aminas em função do átomo de prova (N=36). Método de cálculo de cargas parciais (AM1), valor de corte de energia (30 kcal.mol⁻¹), espaçamento da grade (2,0 Å).

Átomo de prova	Índices Estatísticos ^a							
	q ²	r ²	EP	EP _{vc}	NC	F	S(%)	E(%)
Csp³ (+1)	0,576	0,816	0,366	0,556	2	73280	40,3	59,7
H (+1)	0,603	0,827	0,355	0,546	3	78669	47,4	52,6
Osp³ (-1)	0,580	0,823	0,359	0,553	2	76724	43,3	56,7

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher; S, porcentagem da contribuição estérica; E, porcentagem da contribuição eletrostática.

É possível observar na Tabela 10 que a mudança do átomo de prova promoveu uma melhora na capacidade preditiva do modelo, visto que ambos os modelos com hidrogênio ou oxigênio como átomos de prova possuem maiores valores de q² do que o modelo com carbono. Entre os dois modelos novos, o modelo que utilizou hidrogênio

como átomo de prova apresentou os melhores parâmetros estatísticos e, portanto, foi utilizado na etapa seguinte. O uso do hidrogênio como átomo de prova aumentou a resolução do cálculo dos campos moleculares devido ao seu menor raio atômico, o que permitiu uma maior proximidade com os ligantes no momento do cálculo.

4.2.4 Avaliação do Melhor Nível de Corte de Energia

Esta etapa consistiu na obtenção de modelos de CoMFA mantendo o método de cálculo de cargas parciais (AM1), o átomo de prova (hidrogênio com carga +1), e o espaçamento da grade (2,0 Å) e variando apenas o nível de corte de energia. A Tabela 11 mostra os modelos gerados em três níveis de corte de energia distintos: 10, 20 e 30 kcal.mol⁻¹.

Tabela 11. Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-arilquinazolina-4-aminas em função do nível de corte de energia (N=36). Método de cálculo de cargas parciais (AM1), átomo de prova (H +1), espaçamento da grade (2,0 Å).

Nível de corte (kcal.mol ⁻¹)	Índices Estatísticos ^a							
	q ²	r ²	EP	EP _{VC}	NC	F	S(%)	E(%)
10	0,588	0,815	0,367	0,548	2	72548	42,7	57,3
20	0,584	0,821	0,361	0,550	2	75715	46,4	53,6
30	0,603	0,827	0,355	0,546	3	78669	47,4	52,6

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{VC}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher; S, porcentagem da contribuição estérica; E, porcentagem da contribuição eletrostática.

Ao contrário do ocorrido na etapa anterior, o modelo obtido anteriormente apresentou melhor desempenho do que os novos modelos gerados, ou seja, maiores valores de q² e r² e menores valores de EP e EP_{VC} e, portanto, foi mantido o valor de corte em 30 kcal.mol⁻¹ a ser utilizado na próxima etapa.

4.2.5 Avaliação do Melhor Espaçamento da Grade

A última etapa de refinamento dos modelos de CoMFA envolveu a manutenção dos melhores parâmetros obtidos até agora (método de cálculo das cargas parciais: AM1, átomo de prova: hidrogênio +1, nível de corte de energia: 30 kcal.mol⁻¹) e variando somente o espaçamento da grade em três níveis: 1, 1,5 e 2,0 Å. Os parâmetros estatísticos obtidos nesta última etapa podem ser observados na Tabela 12.

Tabela 12. Índices estatísticos dos modelos de CoMFA obtidos para os derivados da série das 6-arylquinazolina-4-aminas em função do espaçamento da grade (N=36). Método de cálculo de cargas parciais (AM1), átomo de prova (H +1), nível de corte de energia (30 kcal.mol⁻¹).

Espaçamento da grade (Å)	Índices Estatísticos ^a							
	q ²	r ²	EP	EP _{VC}	NC	F	S(%)	E(%)
1,0	0,608	0,891	0,286	0,542	3	87459	38,8	61,2
1,5	0,601	0,819	0,363	0,548	2	74750	43,9	56,1
2,0	0,603	0,827	0,355	0,546	3	78669	47,4	52,6

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{VC}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher; S, porcentagem da contribuição estérica; E, porcentagem da contribuição eletrostática.

Segundo os dados da Tabela 12, é possível observar que a diminuição do espaçamento da grade de 2,0 para 1,5 Å gerou um modelo com quase o mesmo desempenho, mas a diminuição 2,0 para 1,0 Å gerou o modelo com os melhores parâmetros estatísticos desta análise. Este modelo possui um valor de q² ligeiramente maior que o anterior e um valor de r² bem maior do que o anterior, passando a explicar quase 90% da variabilidade da atividade biológica dos compostos da série, além dos menores valores de EP e EP_{VC} e o maior valor de Fisher (único acima de 80000) entre todos os modelos gerados até então. Pode ser observado, também, que em todas as etapas da análise, modelos com NC iguais a 3 eram de qualidade superior aos modelos com NC iguais a 2. Em função do melhor modelo de CoMFA obtido possuir

espaçamento da grade igual a 1,0 Å e o hidrogênio como átomo de prova, o qual possui o menor raio atômico entre os átomos de prova testados, pode-se inferir que o modelo responde de forma estatisticamente favorável a uma exploração mais minuciosa da superfície da molécula, que não seria possível com átomos de prova com raios atômicos maiores e com vértices da grade mais afastados entre si.

4.2.6 Validação Externa do Melhor Modelo de CoMFA e Detecção de *Outliers*

A capacidade preditiva do melhor modelo de CoMFA obtido foi verificada de forma análoga ao do método de HQSAR, onde foi calculado o valor da diferença entre as atividades biológicas experimentais e preditas pelo modelo e também o valor de r^2_{pred} para este modelo. A Tabela 13 mostra os valores de Exp, Pred e dos resíduos (Exp – Pred) dos derivados da classe das 6-irilquinazolina-4-aminas obtidos pelo melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å).

Tabela 13. Valores das atividades biológicas experimentais (Exp), preditas (Pred) e resíduos (Res) dos compostos da classe das 6-arylquinazolina-4-aminas utilizando o melhor modelo de CoMFA. Átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å. Os compostos do conjunto de teste estão marcados com um asterisco.

#	Exp	Pred	Res	#	Exp	Pred	Res
1	7,21	6,72	0,49	26	7,29	7,44	-0,15
2	5,90	5,28	0,62	27	7,59	7,51	0,08
3*	5,46	5,13	0,33	28	6,81	7,06	-0,25
4	5,24	5,54	-0,30	29*	6,04	7,44	-1,40
5	5,50	5,70	-0,20	30	6,27	6,27	0,00
6	5,05	4,96	0,09	31	6,92	6,96	-0,04
7	6,79	6,57	0,22	32 (R)*	7,03	7,19	-0,16
8	5,35	5,31	0,04	32 (S)*	7,03	7,01	0,02
9	6,74	6,30	0,44	33 (R)*	6,87	7,01	-0,14
10	5,84	5,66	0,18	33 (S)*	6,87	7,11	-0,24
11*	5,33	5,52	-0,19	34 (R)*	7,52	6,98	0,54
12	7,51	7,26	0,25	34 (S)*	7,52	7,08	0,44
13	7,42	7,36	0,06	35	7,12	7,14	-0,02
14	5,94	6,60	-0,66	36*	7,77	7,13	0,64
15	6,59	7,18	-0,59	37	5,94	5,85	0,09
16	7,46	7,14	0,32	38	6,25	6,03	0,22
17	7,08	6,99	0,09	39	6,23	6,28	-0,05
18	7,01	7,33	-0,32	40	5,44	5,79	-0,35
19	7,13	7,15	-0,02	41	5,47	5,46	0,01
20*	7,24	7,07	0,17	42	5,82	5,95	-0,13
21	6,90	6,75	0,15	43*	5,85	5,90	-0,05
22	7,03	6,89	0,14	44	5,57	5,75	-0,18
23*	6,69	6,93	-0,24	45	5,31	5,51	-0,19
24	7,85	7,68	0,17	46	5,08	5,33	-0,25
25	7,15	7,10	0,05				

A análise da Tabela 13 informa que apenas 13% dos compostos apresentaram valor residual superior a 0,5 em valor modular. Este fato, somado ao valor calculado de r^2_{pred} igual a 0,572 indicam que o modelo de CoMFA gerado possui capacidade preditiva aceitável para novos compostos desta série congênere. É possível também observar a presença de um composto *outlier*, composto **29**, que pertence ao grupo de teste. Como o composto **29** não foi considerado um *outlier* no estudo de HQSAR, o comportamento do composto **29** provavelmente se deve a fatores tridimensionais. A Figura 25 mostra o gráfico comparativo entre os valores de Exp e Pred dos grupos de treinamento e de teste do melhor modelo de CoMFA, onde é possível observar o ponto relativo ao composto **29** afastado dos demais, os quais seguem uma distribuição semelhante. Foi possível observar também que os centros de quiralidade nos substituintes localizados na posição R₃ não parecem impactar na atividade biológica dos compostos, visto que as diferenças nos valores de pIC₅₀ preditos não diferem de forma significativa para os enantiômeros dos compostos **32**, **33** e **34**.

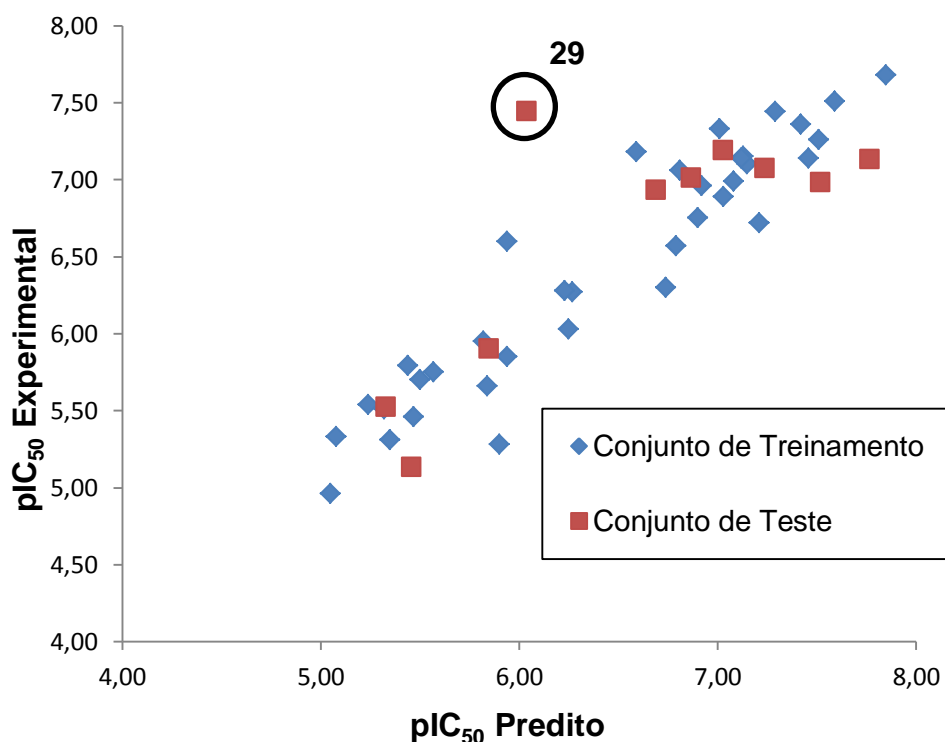


Figura 25. Gráfico com os valores de pIC₅₀ preditos *versus* os valores de pIC₅₀ experimentais para os grupos de treinamento (azul) e de teste (vermelho) utilizando o melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å).

4.2.7 Avaliação do Teste de Randomização-Y para o Melhor Modelo de CoMFA

O teste de randomização-Y foi realizado de forma ligeiramente diferente da análise por HQSAR. Nesta etapa, quatro randomizações dos valores de atividade biológica foram realizadas, seguidas pela elaboração de um modelo de CoMFA para cada randomização, mantendo os mesmos parâmetros do melhor modelo de CoMFA obtido no processo de refinamento. Os parâmetros estatísticos resultantes deste teste encontram-se na Tabela 14.

Tabela 14. Índices estatísticos obtidos no teste de randomização-Y para o melhor modelo de CoMFA: átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å.

#	Índices Estatísticos ^a							
	q ²	r ²	EP	EP _{vc}	NC	F	S(%)	E(%)
1	-0,333	0,315	0,696	0,972	1	15671	36	64
2	-0,320	0,326	0,691	0,967	1	16408	36,1	63,9
3	-0,666	0,291	0,709	1,086	1	13948	40	60
4	-0,489	0,217	0,745	1,027	1	9398	41,4	58,6

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher; S, porcentagem da contribuição estérica; E, porcentagem da contribuição eletrostática.

Como esperado, os parâmetros estatísticos obtidos para as quatro randomizações realizadas foram ruins. Valores baixos de r², chegando a valores negativos para q² e valores muito altos para os desvios padrões com e sem validação cruzada, o que descarta a correlação ao acaso entre a atividade biológica e os campos moleculares calculados dos compostos pelo melhor método de CoMFA.

4.2.8 Análise dos Mapas de Contorno

Os mapas de contorno são representações dos vértices da grade onde as diferenças nos valores dos campos estão associadas com diferenças na afinidade entre ligante e proteína e, portanto, na atividade biológica dos mesmos, contribuindo com informações úteis acerca do modo de ligação à enzima alvo.

No mapa de contorno estérico, áreas representadas por poliedros verdes são regiões onde grupos volumosos são favoráveis à atividade biológica, enquanto áreas representadas por poliedros amarelos são regiões onde grupos volumosos são desfavoráveis à atividade biológica. No mapa de contorno eletrostático, áreas representadas por poliedros vermelhos são regiões onde grupos com grande densidade eletrônica ou carga negativa são favoráveis à atividade biológica, enquanto áreas representadas por poliedros azuis grupos são regiões onde grupos com grande densidade eletrônica ou carga negativa diminuem a atividade biológica.

Os mapas de contorno estérico e eletrostático para o composto mais ativo da série (**24**) são mostrados nas Figuras 26A e 26B, respectivamente.

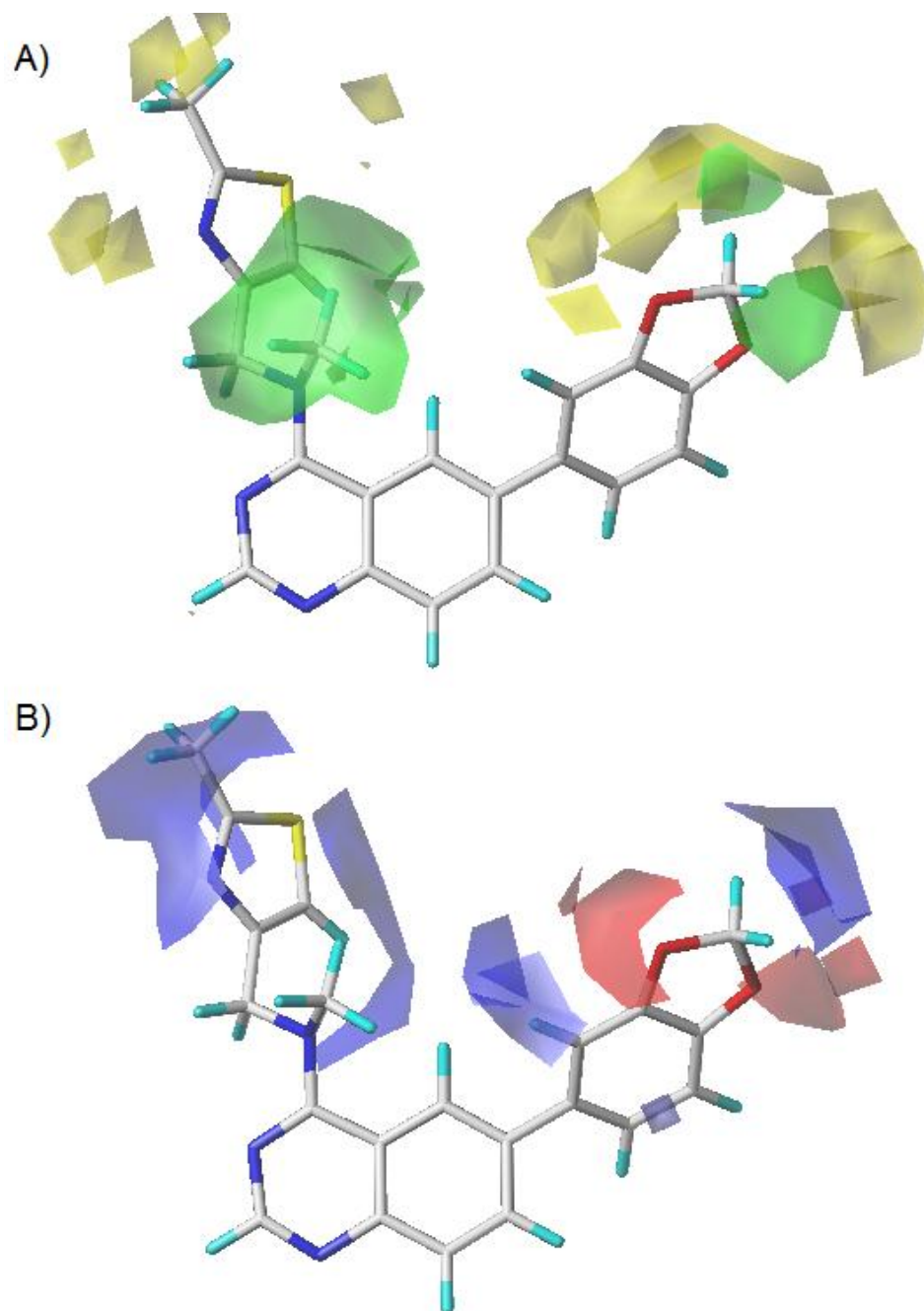


Figura 26. Mapas de contorno do melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å) obtido para o composto **24** (modelo em bastão e colorido por elementos). (A) O mapa estérico indica as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a atividade biológica. (B) O mapa eletrostático indica áreas onde grupos com grande densidade eletrônica ou carga negativa aumentam (vermelho) ou diminuem (azul) a atividade biológica.

No mapa de contorno estérico do composto **24** (Figura 26A), podem ser observadas duas regiões onde grupos volumosos aumentam a atividade biológica: a primeira é na região onde se situa o grupo metila (R_2) ligado ao átomo de nitrogênio do grupo amino e a segunda é na região onde se localizam os átomos de hidrogênio do grupo metilenodioxo do anel 1,3-benzo-dioxola (R_1). A primeira região ilustra o fato dos compostos que contém o grupo metila em R_2 serem, em geral, mais ativos do que os compostos que contém hidrogênio nesta posição, por exemplo, os cinco compostos menos ativos da série possuem hidrogênio na posição R_2 e, dos cinco compostos mais ativos da série, os compostos **36** e **34** não possuem o grupo metila em R_2 . Compostos análogos de **36** e **34** com grupos metila em R_2 não fazem parte do conjunto de dados e poderiam ser compostos com atividade maior do que os apresentados nesta série.

Regiões onde a presença de grupos volumosos diminui a atividade biológica podem ser observadas ao redor no anel 1,3-benzo-dioxola em R_1 e em alguns pontos próximos do anel heterocíclico em R_3 . Isso reforça que anéis diferentes do 1,3-benzo-dioxola em R_1 como os ésteres etílicos do furano presentes nos compostos **40**, **42**, **43**, **44** e **46** os quais possuem este grupo, possuem baixa atividade biológica. De fato, todos estes compostos possuem pIC_{50} menores do que 6,0. A Figura 27 exemplifica a presença do grupo éster etílico do furano do composto **46** na região onde grupos volumosos são prejudiciais à atividade biológica dos compostos.

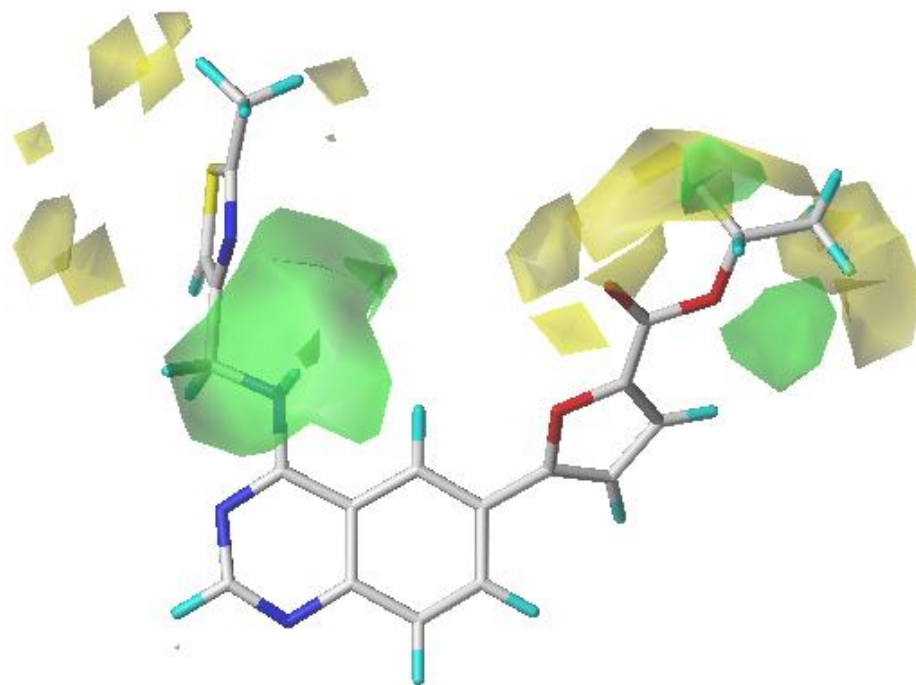


Figura 27. Mapa de contorno estérico do melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å) obtido para o composto **46** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a atividade biológica.

Observando o mapa de contorno eletrostático do composto **24** (Figura 26B), é possível observar duas regiões onde átomos com densidade eletrônica elevada são favoráveis à atividade biológica. Estas regiões estão localizadas ao redor dos átomos de oxigênio do grupo metilenodioxo do anel 1,3-benzo-dioxola, que são eletronegativos e, portanto, conferem um aumento na atividade biológica dos compostos contendo esse grupo.

Também é possível observar na Figura 26B que existem regiões onde átomos com densidade eletrônica elevada são prejudiciais à atividade biológica, como por exemplo, a região posicionada diante do carbono sp³ do grupo metilenodioxo do anel 1,3-benzo-dioxola, que possui carga parcial positiva em função da presença dos átomos de oxigênio na vizinhança, o que reforça a importância deste anel para a atividade biológica. A presença de grupos com alta densidade eletrônica também diminui a atividade biológica em outra região próxima ao anel fenila na posição 6 do

anel quinazolina. Esta região se localiza precisamente onde o átomo de cloro do composto **6** (o menos ativo da série) está presente (Figura 28).

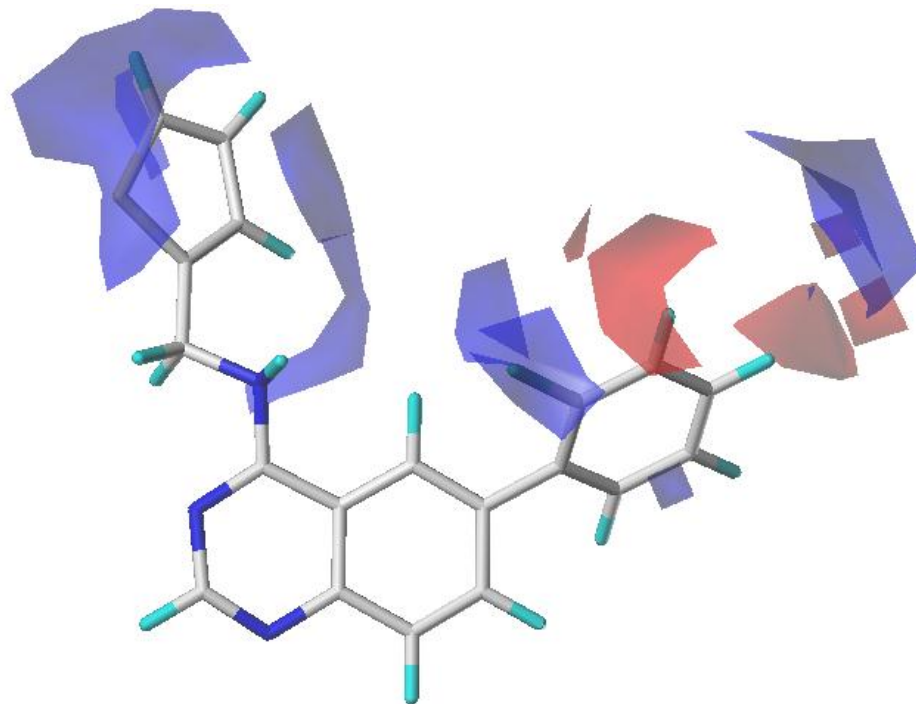


Figura 28. Mapa de contorno eletrostático do melhor modelo de CoMFA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; nível de corte de energia: 30 kcal.mol⁻¹ e espaçamento da grade: 1,0 Å) obtido para o composto **6** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos com alta densidade eletrônica aumentam (vermelho) ou diminuem (azul) a atividade biológica.

Na figura 28, a presença de regiões de cor azul em volta do anel heterocíclico em R3 deve-se ao fato de alguns derivados contendo anéis heterocíclicos com substituintes metila apresentarem baixa atividade biológica. Entretanto, existem compostos com baixa atividade biológica que não possuem grupos metila em anéis heterocíclicos e, portanto este campo molecular não é uma região capaz de diferenciar os compostos da série.

Os resultados de CoMFA apresentados estão em conformidade com um estudo de QSAR 3D realizado por Pan e colaboradores com esta mesma série de compostos. Este estudo utiliza características tridimensionais assim como o método de CoMFA, mas a apresentação gráfica dos resultados é semelhante à do HQSAR, pois aponta regiões onde a presença de determinados grupos favorece ou não a atividade biológica,

sem especificar a razão do favorecimento. Como resultados deste estudo, os autores estabelecem que o anel 1,3-benzo-dioxola em R_1 e o grupo metila em R_2 são grupos favoráveis para a atividade biológica, como no composto **24**, assim como o átomo de cloro ligado ao anel benzênico em R_1 e o átomo de hidrogênio em R_2 no composto **6** como desfavoráveis à atividade biológica (PAN; WANG; BRYANT, 2013).

4.2.9 Avaliação do Melhor Modelo de CoMSIA

O mesmo conjunto de treinamento utilizado para obter os modelos de CoMFA foram utilizados para obter os modelos de CoMSIA, que foram gerados com o mesmo conjunto de parâmetros utilizado para o melhor modelo de CoMFA, ou seja, método de cálculo das cargas atômicas parciais (AM1), átomo de prova (hidrogênio com carga +1) e espaçamento da grade (1,0 Å). As etapas de validação interna e externa, determinação de *outliers* e visualização de mapas de contorno foram executadas da mesma forma que no método CoMFA. Como o método CoMSIA utiliza funções gaussianas para gerar os índices de similaridade, não há a necessidade de avaliação de valores de corte de energia (*cut-off*), uma limitação do método CoMFA.

Desta forma, a primeira etapa da análise foi a obtenção de modelos de CoMSIA, considerando diferentes combinações dos índices estérico (S), eletrostático (E), hidrofóbico (H), doador (D) e aceptor (A) de ligação hidrogênio, mantendo o valor do fator de atenuação (α) padrão igual a 0,3. Os resultados desta análise são mostrados na Tabela 15.

Tabela 15. Índices estatísticos dos modelos de CoMSIA obtidos para os derivados da série das 6-arylquinazolina-4-aminas (N=36) em função das diversas combinações de campos moleculares (S, E, H, D, A).

Campos Moleculares ^a	Índices Estatísticos ^b						Contribuição (%)				
	q ²	r ²	EP	EP _{vc}	NC	F	S	E	H	D	A
S + E	0,529	0,766	0,412	0,586	2	54044	14,4	85,6	-	-	-
	0,355	0,653	0,511	0,697	4	16024	53,0	-	-	-	47,0
S + H	0,686	0,792	0,395	0,470	2	90627	23,2	-	76,8	-	-
E + D	0,550	0,710	0,459	0,572	2	40382	-	80,5	-	19,5	-
E + A	0,514	0,733	0,441	0,594	2	45312	-	74,9	-	-	25,1
E + H	0,700	0,762	0,410	0,460	1	108576	-	45,0	55,0	-	-
H + D	0,675	0,794	0,395	0,486	2	63445	-	-	79,5	20,5	-
H + A	0,663	0,745	0,393	0,497	3	60451	-	-	66,2	-	33,8
S + E + H	0,690	0,758	0,413	0,468	1	106443	9,1	40,9	50,0	-	-
S + E + A	0,518	0,751	0,426	0,592	2	49731	11,5	69,9	-	-	18,6
S + E + D	0,526	0,703	0,465	0,587	2	39090	10,9	69,7	-	19,3	-
S + D + A	0,326	0,668	0,500	0,712	4	17106	36,5	-	-	18,4	45,2
S + H + D	0,683	0,790	0,397	0,469	2	64247	14,4	-	67,8	17,8	-
S + H + A	0,665	0,785	0,401	0,495	3	61076	18,1	-	55,7	-	26,2
E + H + D	0,712	0,795	0,386	0,457	2	64168	-	33,2	49,5	17,3	-
E + D + A	0,499	0,774	0,412	0,613	3	36514	-	61,4	-	16,3	22,3
E + H + A	0,672	0,800	0,382	0,489	2	65937	-	35,8	49,5	-	14,7
H + D + A	0,668	0,784	0,414	0,490	2	53276	-	-	57,3	26,2	16,5
S + E + H + A	0,672	0,806	0,376	0,489	2	68606	8,6	32,3	45,0	-	14,1
S + E + H + D	0,700	0,793	0,388	0,467	2	63138	7,2	30,6	45,8	16,5	-
S + E + D + A	0,499	0,766	0,419	0,613	3	34936	12,6	51,1	-	15,8	20,6
E + H + D + A	0,679	0,791	0,390	0,483	2	32451	-	26,4	42,5	14,6	16,5
S + H + D + A	0,701	0,819	0,368	0,474	3	48422	8,9	-	56,9	18,1	16
S + E + H + D + A	0,673	0,792	0,389	0,488	2	63005	5,9	24,7	39,9	14,1	15,4

^a Campos moleculares: estérico (S), eletrostático (E), hidrofóbico (H), doador (D) e aceptor (A) de ligação hidrogênio. ^b q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher; o símbolo "-" foi utilizado quando o campo molecular correspondente não foi considerado na construção do modelo. Foram excluídos os modelos com valor de q² menor ou igual a 0,3.

Considerando apenas os campos moleculares estérico e eletrostático, podem ser comparados os valores de q^2 do melhor modelo de CoMFA ($q^2 = 0,608$) com o modelo de CoMSIA ($q^2 = 0,529$), onde é possível observar que o modelo de CoMFA é mais preditivo do que o de CoMSIA correspondente. Entretanto, a inclusão dos três demais campos moleculares gera um modelo de CoMSIA mais preditivo ($q^2 = 0,673$) do que o modelo de CoMFA ($q^2 = 0,608$), indicando que a variação da atividade biológica desta série congênere não depende apenas de parâmetros estéricos e eletrostáticos.

Ao analisar-se o modelo de CoMSIA contendo todos os campos moleculares, é possível observar que o índice de similaridade hidrofóbico é o que mais contribui para a explicação da variação da atividade biológica desta série (aproximadamente 40%), este fato é corroborado pelo fato de que todos os modelos que contêm este índice possuem maiores valores de q^2 do que os que não o contêm.

Dentre os modelos de CoMSIA obtidos, o que possui maior valor de q^2 e, portanto, é o modelo com melhor validação interna, é o que possui três dos cinco campos moleculares estudados: eletrostático, hidrofóbico e doador de ligação hidrogênio e, portanto, foi utilizado para investigar o impacto do fator de atenuação (α) nos modelos de CoMSIA. Este fator foi variado em três níveis: 0,2, 0,3 e 0,4 e os resultados encontram-se na Tabela 16.

Tabela 16. Índices estatísticos dos modelos de CoMSIA obtidos para os derivados da série das 6-arylquinazolina-4-aminas (N=36) em função do fator de atenuação (α) contendo os campos moleculares eletrostático (E), hidrofóbico (H) e doador de ligação hidrogênio (D).

α	Índices Estatísticos ^a					Contribuição (%) ^b					
	q^2	r^2	EP	EP _{vc}	NC	F	S	E	H	D	A
0,2	0,713	0,796	0,385	0,457	2	64327	-	29,9	51,2	18,9	-
0,3	0,712	0,795	0,386	0,457	2	64168	-	33,2	49,5	17,3	-
0,4	0,712	0,794	0,387	0,458	2	63702	-	33,4	49,8	16,7	-

^a q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher. ^b Campos moleculares: estérico (S), eletrostático (E), hidrofóbico (H), doador (D) e acceptor (A) de ligação hidrogênio.

De acordo com os resultados obtidos nesta etapa, é possível concluir que o fator de atenuação não impacta nos índices estatísticos, visto que os parâmetros estatísticos obtidos para todos os modelos praticamente não diferem entre si, sendo equivalentes estatisticamente. Logo, os três modelos foram submetidos à validação externa.

4.2.10 Validação Externa do Melhor Modelo de CoMSIA e Verificação da Presença de *Outliers*

A capacidade preditiva do melhor modelo de CoMSIA obtido foi verificada de forma análoga ao procedimento realizado para o melhor modelo de CoMFA. Os valores calculados de r^2_{pred} para os modelos de CoMSIA com fator de atenuação igual a 0,2, 0,3 e 0,4 foram 0,506, 0,487 e 0,490, respectivamente. Em função deste resultado, foi possível concluir que apesar do fator de atenuação não impactar nos parâmetros estatísticos da etapa de validação interna, o mesmo não ocorreu na validação externa dos modelos, visto que para ser considerado um modelo preditivo o valor de r^2_{pred} deve ser maior do que 0,5, algo que somente foi observado para o modelo com fator de atenuação igual a 0,2 que obviamente foi considerado o melhor modelo.

A Tabela 17 apresenta os valores experimentais e calculados de atividade biológica (pIC_{50} , M) e os valores residuais considerando o melhor modelo de CoMSIA dos derivados da classe das 6-irilquinazolina-4-aminas.

Tabela 17. Valores de atividade biológica (pIC_{50} , M) experimental (Exp), predita (Pred) e resíduos (Exp – Pred) dos compostos da classe das 6-arylquinazolina-4-aminas utilizando o melhor modelo de CoMSIA (campos moleculares E, H, D e fator de atenuação $\alpha = 0,2$). Os compostos do conjunto de teste estão marcados com um asterisco.

#	Exp	Pred	Resíduo	#	Exp	Pred	Resíduo
1	7,21	6,77	0,44	26	7,29	7,33	-0,04
2	5,90	5,39	0,51	27	7,59	7,31	0,28
3*	5,46	5,22	0,24	28	6,81	6,88	-0,07
4	5,24	5,84	-0,60	29*	6,04	7,30	-1,26
5	5,50	6,06	-0,56	30	6,27	6,88	-0,61
6	5,05	5,26	-0,21	31	6,92	7,29	-0,37
7	6,79	7,14	-0,35	32 (R)*	7,03	6,72	0,31
8	5,35	5,28	0,07	32 (S)*	7,03	6,66	0,37
9	6,74	7,01	-0,27	33 (R)*	6,87	6,85	0,02
10	5,84	5,53	0,31	33 (S)*	6,87	6,84	0,03
11*	5,33	5,65	-0,32	34 (R)*	7,52	6,81	0,71
12	7,51	7,26	0,25	34 (S)*	7,52	6,74	0,78
13	7,42	7,28	0,14	35	7,12	6,91	0,21
14	5,94	6,77	-0,83	36*	7,77	6,87	0,90
15	6,59	7,23	-0,64	37	5,94	5,84	0,10
16	7,46	6,81	0,65	38	6,25	6,10	0,15
17	7,08	6,78	0,30	39	6,23	5,93	0,30
18	7,01	7,25	-0,24	40	5,44	5,51	-0,07
19	7,13	6,89	0,24	41	5,47	5,49	-0,02
20*	7,24	6,81	0,43	42	5,82	5,54	0,28
21	6,90	6,80	0,10	43*	5,85	5,64	0,21
22	7,03	6,77	0,26	44	5,57	5,60	-0,03
23*	6,69	6,78	-0,09	45	5,31	5,43	-0,11
24	7,85	7,32	0,53	46	5,08	5,59	-0,51
25	7,15	6,76	0,39				

Para o modelo de CoMSIA, aproximadamente 26% dos compostos apresentaram valor residual superior a 0,5 (em módulo). O composto **29**, considerado um *outlier* na análise do modelo de CoMFA, também foi considerado um *outlier* nesta análise. Este

comportamento se deve, provavelmente, devido à similaridade estrutural com compostos bastante ativos, como o composto **24**, que também possui um grupo metila em R₂, um anel 1,3-benzo-dioxola em R₁ e um anel heterocíclico em R₃. A atividade biológica deste composto foi superestimada em ambos os estudos de QSAR 3D.

Como estes estudos são independentes da estrutura tridimensional da enzima e partem da premissa que todos os compostos interagem com a proteína da mesma forma, a explicação da atividade biológica deste composto ser menor que a estimada é, provavelmente, em função da conformação bioativa deste composto ser distinta dos demais, o que pode ser avaliado em estudos de docagem molecular.

A Figura 29 mostra o gráfico comparativo entre os valores de Exp e Pred dos compostos dos grupos de treinamento e de teste do melhor modelo de CoMSIA, onde também é possível observar o ponto relativo ao composto **29** afastado dos demais, bastante semelhante ao observado na Figura 25. As figuras 29 e 25 também apontam uma perda de sensibilidade dos métodos de QSAR-3D, visto que compostos com valores de Exp acima de 6,50 possuem valores de Pred subestimados, resultando em um platô nessa região dos gráficos Exp *versus* Pred.

A exemplo do ocorrido para o modelo de CoMFA, o modelo de CoMSIA também não é capaz de discriminar compostos com centros de quiralidade posicionados na posição R₃, visto que para os enantiômeros dos compostos **32**, **33** e **34** não foram preditas atividades biológicas distintas.

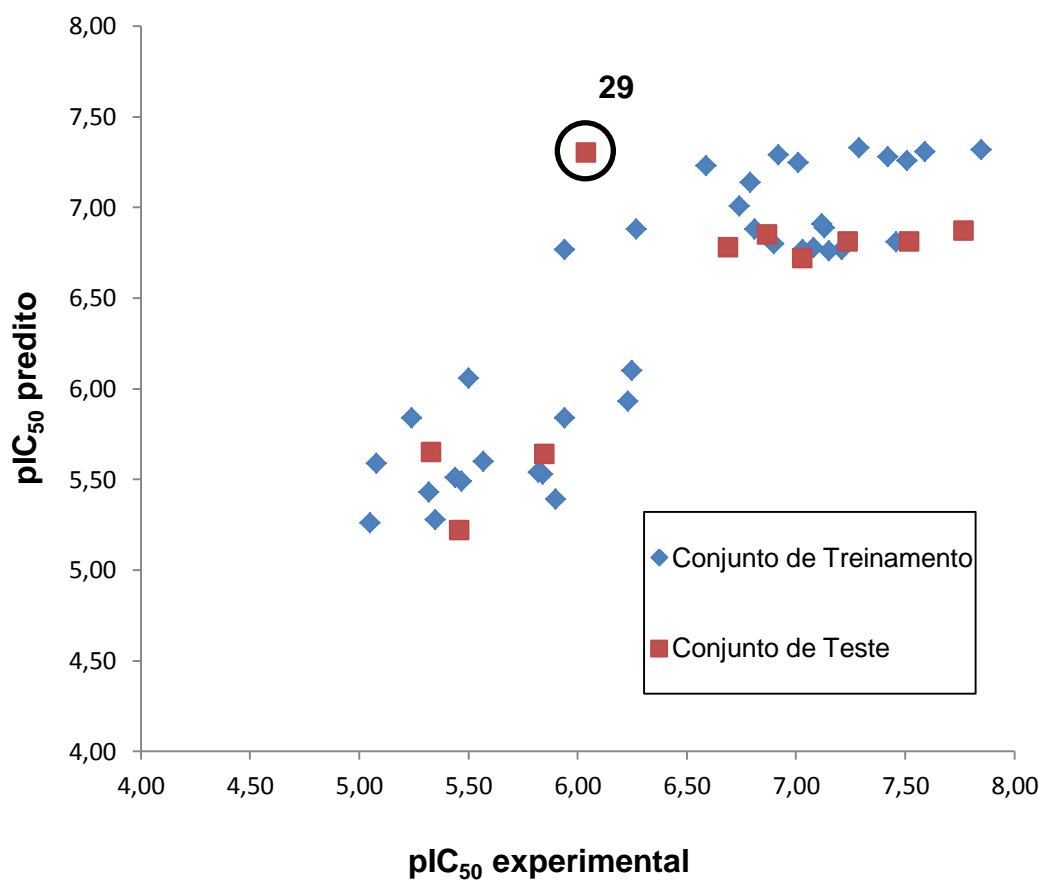


Figura 29. Gráfico com os valores de pIC_{50} preditos *versus* os valores de pIC_{50} experimentais para os grupos de treinamento (azul) e de teste (vermelho) utilizando o melhor modelo de CoMSIA.

4.2.11 Avaliação do Teste de Randomização-Y para o Melhor Modelo de CoMSIA

O teste de randomização-Y foi realizado de forma idêntica a da análise por CoMFA, onde quatro randomizações dos valores de atividade biológica foram realizadas, seguidas pela elaboração de um modelo de CoMSIA para cada randomização. Os parâmetros estatísticos resultantes deste teste encontram-se na Tabela 18.

Tabela 18. Índices estatísticos obtidos nas quatro randomizações da atividade biológica no teste de randomização-Y para o melhor modelo de CoMSIA.

#	Índices Estatísticos ^a					Contribuição (%) ^b					
	q ²	r ²	EP	EP _{vc}	NC	F	S	E	H	D	A
1	-0,046	0,163	0,770	0,861	1	6622	-	36,8	22,1	41,2	-
2	-0,040	0,166	0,769	0,858	1	6752	-	35,9	21,7	42,3	-
3	-0,306	0,237	0,746	0,976	2	5120	-	65,1	24,2	10,6	-
4	-0,435	0,118	0,790	1,008	1	4555	-	56,5	21,9	21,7	-

^a q², coeficiente de correlação da validação cruzada; r², coeficiente de determinação; EP, erro padrão sem a validação cruzada; EP_{vc}, erro padrão da validação cruzada; NC, número ótimo de componentes; F, valor de Fisher. ^b Campos moleculares: estérico (S), eletrostático (E), hidrofóbico (H), doador (D) e acceptor (A) de ligação hidrogênio.

Assim como no teste para o modelo de CoMFA, foram encontrados valores de r² muito baixos, valores de q² negativos para e valores altos para os desvios padrões com e sem validação cruzada, o que também exclui a possibilidade de correlação ao acaso entre a atividade biológica e os campos moleculares calculados para os compostos no melhor modelo de CoMSIA.

4.2.12 Análise dos Mapas de Contorno

Em função do melhor modelo de CoMSIA conter apenas campos moleculares eletrostático, hidrofóbico e doador de ligação hidrogênio, somente estes três tipos de mapas foram obtidos. O mapa eletrostático para o composto **24** (o mais ativo da série) pode ser observado na Figura 30, onde é possível perceber a semelhança com o mapa eletrostático obtido pelo método de CoMFA (Figura 26B). Ambos os mapas indicam a importância do anel 1,3-benzo-dioxola para a atividade biológica da série congênere, tanto pela presença de regiões nas quais átomos com elevada densidade eletrônica aumentam a atividade biológica correspondendo aos átomos de oxigênio deste grupo, quanto pela presença de regiões nas quais átomos com elevada densidade eletrônica diminuem a atividade biológica correspondendo aos átomos de carbono vizinhos aos

átomos de oxigênio os quais possuem cargas parciais positivas, como exemplificado na Figura 31.

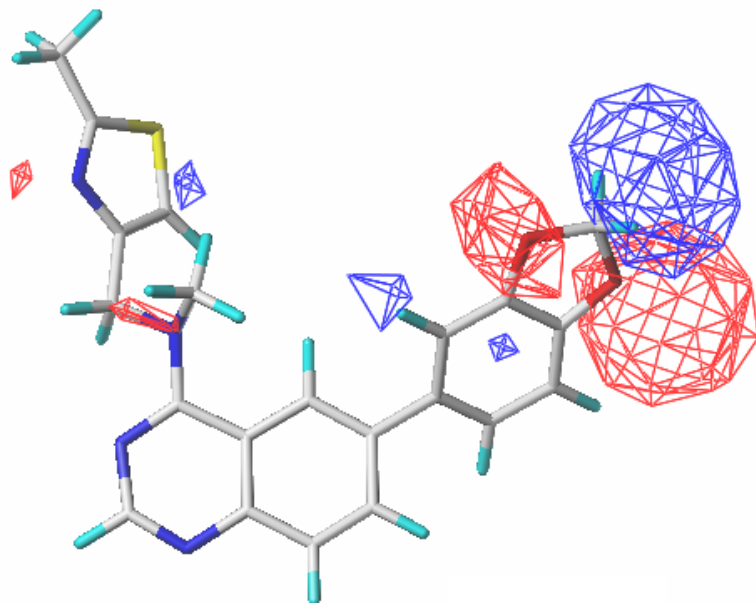


Figura 30. Mapa de contorno eletrostático do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **24** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos com alta densidade eletrônica aumentam (vermelho) ou diminuem (azul) a atividade biológica.

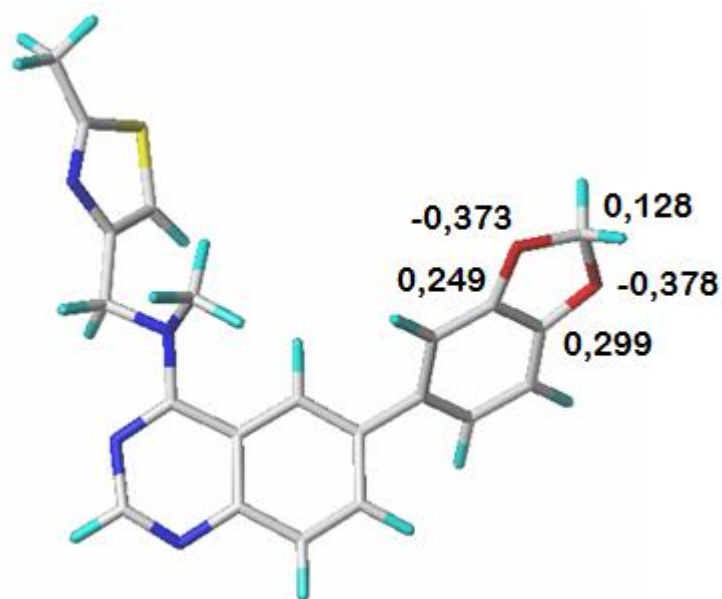


Figura 31. Estrutura química do composto **24** com as cargas parciais calculadas pelo método AM1 destacadas para os átomos de oxigênio e carbono do anel 1,3-benzo-dioxola.

Este mapa eletrostático de CoMSIA também posiciona uma região azul próxima ao átomo de cloro do composto **6**, assim como no mapa eletrostático do modelo de CoMFA deste composto, entretanto este mapa posiciona uma região vermelha próxima ao átomo de nitrogênio do grupo amino da série, o que também indica que quanto mais grupos doadores de elétrons estiverem ligados a ele, maior é a atividade biológica dos compostos, reforçando a importância de grupos metila na posição R₂ para o aumento da atividade biológica dos derivados.

O mapa de contorno do índice de similaridade hidrofóbico para o composto **24** (Figura 32) também está posicionado no anel 1,3-benzo-dioxola. Uma região onde grupos hidrofóbicos são desfavoráveis para a atividade biológica dos compostos está posicionada no centro deste anel, onde existem ligações covalentes polares entre carbono e oxigênio. Esta diferença de polaridade é mais bem evidenciada pelas cargas parciais calculadas anteriormente (Figura 31).

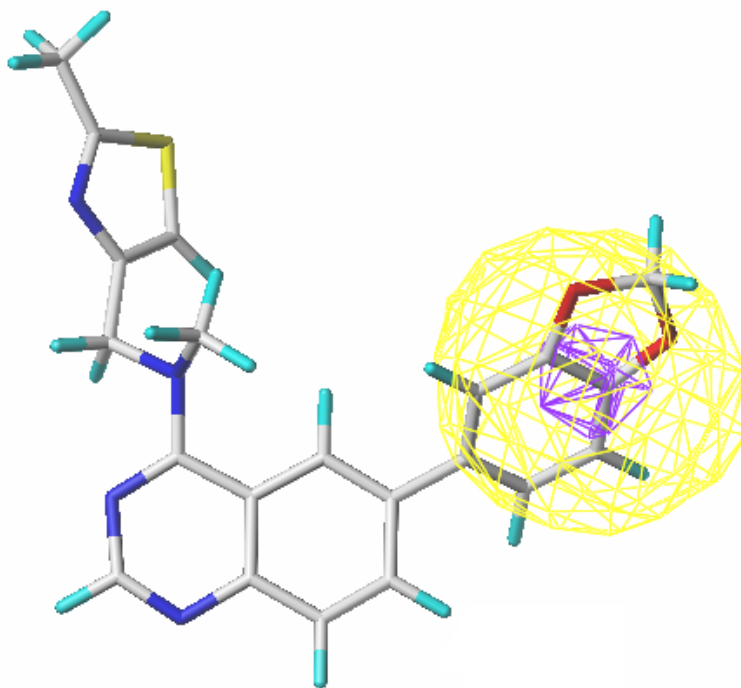


Figura 32. Mapa de contorno hidrofóbico do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **24** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos hidrofóbicos aumentam (amarelo) ou diminuem (roxo) a atividade biológica.

Ao analisarmos o mesmo mapa de contorno hidrofóbico para o composto **46** (Figura 33), é possível perceber que diversos átomos de oxigênio do substituinte em R₁ estão posicionados em uma região onde grupos hidrofóbicos aumentam a atividade biológica de compostos desta série, o que explica a baixa atividade biológica deste composto e de outros com o mesmo substituinte em R₁ (**40**, **42**, **43** e **44**). O composto **6**, possui um átomo de cloro hidrofóbico, mas que está posicionado fora da região onde grupos hidrofóbicos aumentam a atividade biológica dos compostos e, portanto, não contribui para a redução da atividade biológica (Figura 34).

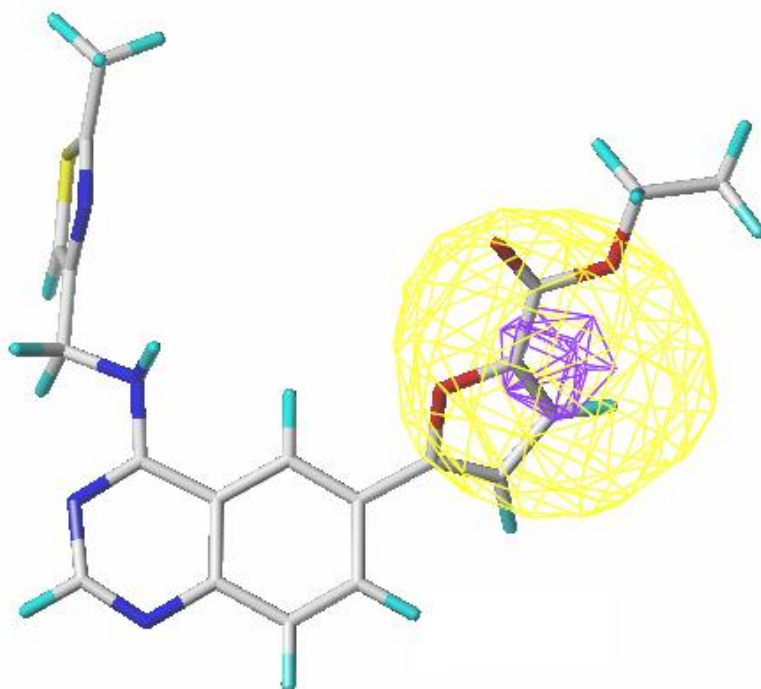


Figura 33. Mapa de contorno hidrofóbico do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **46** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos hidrofóbicos aumentam (amarelo) ou diminuem (roxo) a atividade biológica.

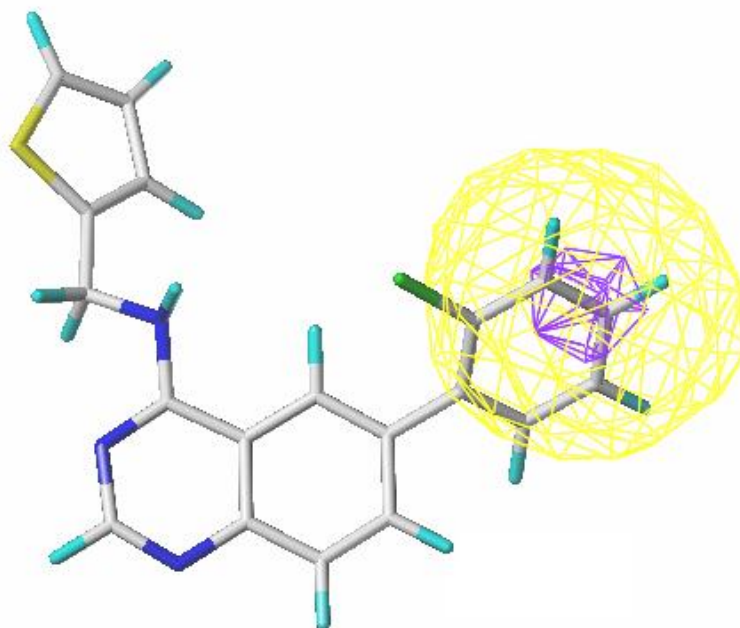


Figura 34. Mapa de contorno hidrofóbico do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **6** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos hidrofóbicos aumentam (amarelo) ou diminuem (roxo) a atividade biológica.

O mapa de contorno do índice de similiaridade doador de ligação hidrogênio para o composto **24** (Figura 35) mostra uma região na qual grupos doadores de ligação hidrogênio são desfavoráveis para a atividade biológica. Esta região é próxima ao grupo metila da posição R₂, indicando que compostos contendo átomos de hidrogênio polares (N-H) nesta posição possuem menor atividade biológica do que compostos semelhantes contendo grupos metila nesta mesma posição, já que o grupo metila não possui propriedade de doação de ligação hidrogênio.

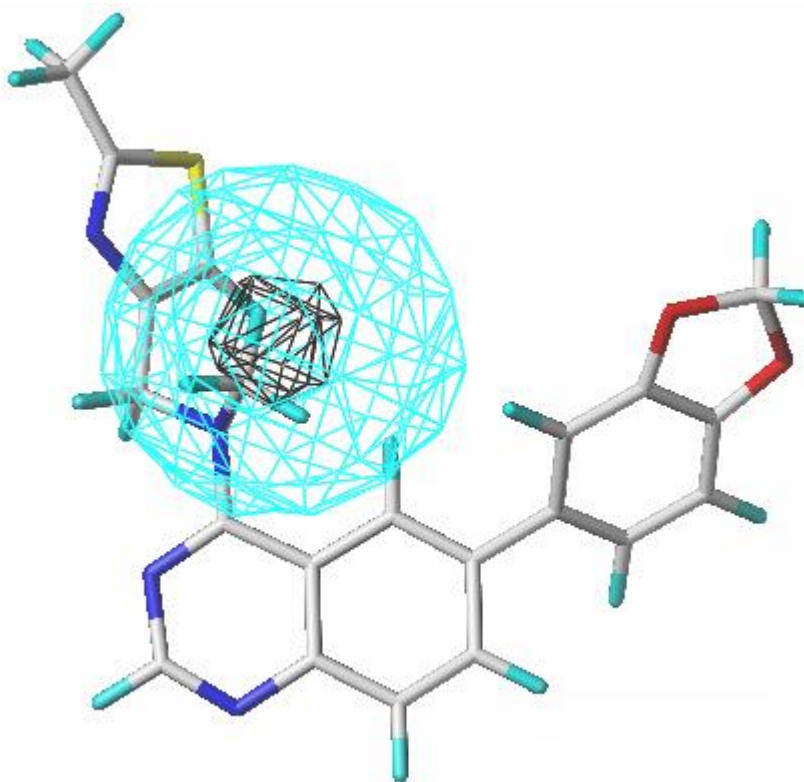


Figura 35. Mapa de contorno doador de ligação hidrogênio do melhor modelo de CoMSIA (átomo de prova: H +1; método de cálculo das cargas parciais: AM1; espaçamento da grade: 1,0 Å; campos moleculares: estérico, hidrofóbico e doador de ligação hidrogênio; fator de atenuação 0,2) obtido para o composto **24** (modelo em bastão e colorido por elementos), mostrando as áreas onde grupos doadores de ligação hidrogênio aumentam (azul claro) ou diminuem (preto) a atividade biológica.

4.3 DOCAGEM MOLECULAR

4.3.1 Validação do Protocolo de Docagem (Re-docagem)

Para a avaliação do melhor protocolo de docagem foram obtidas cinco estruturas tridimensionais da DYRK1A disponíveis no PDB, cujas informações mais relevantes estão resumidas na Tabela 19.

Tabela 19. Resumo das estruturas tridimensionais da DYRK1A resolvidas por difração de raios-X do cristal e seus respectivos ligantes disponíveis no PDB.

PDB ^a	Ligante ^b	Rot. ^c	R (Å) ^d
2WO6	D15	7	2,50
3ANQ	EHB	2	2,60
3ANR	HRM	1	2,60
4MQ1	2C3	7	2,35
4MQ2	2C4	5	2,80

^a Código do complexo da proteína no PDB. ^b Código do ligante no PDB. ^c Número de ligações rotacionáveis do ligante. ^d Resolução da estrutura do complexo ligante-proteína.

Em função da necessidade de complexos ligante-enzima com alta resolução (R igual ou menor do que 2,5 Å) e um número de ligações rotacionáveis dos ligantes entre quatro e oito (números de ligações rotacionáveis da série das 6-amilquinazolina-4-aminas), apenas dois complexos satisfizeram essas duas condições e foram selecionados para a validação. Os complexos selecionados contêm os seguintes ligantes: DJM2005 (código do complexo do PDB: 2WO6; código do ligante: D15; página 73) com resolução igual a 2,50 Å e sete ligações rotacionáveis (SOUNDARARAJAN et al., 2013) e 2C3, um dos ligantes da série das pirido[2,3-*d*]pirimidinas (código do complexo no PDB: 4MQ1; página 42), com resolução igual a 2,35 Å e sete ligações rotacionáveis (ANDERSON et al., 2013).

Utilizando os dois complexos selecionados (2WO6 e 4MQ1), a busca do melhor protocolo de docagem foi realizada pela análise dos valores de RMSD obtidos para cada combinação entre algoritmo de busca e função de pontuação. Para o complexo 2WO6, todas as combinações retornaram, ao menos, uma solução com valor de RMSD menor do que 2,0 Å e a faixa de valores de RMSD menor do que 2,0 Å encontrada foi de 0,88 a 1,95 Å. Para o complexo 4MQ1, todas as combinações também retornaram, ao menos, uma solução com valor de RMSD menor do que 2,0 Å e faixa entre 0,78 e 2,00 Å. A Tabela 20 resume as informações obtidas no processo de validação do protocolo de docagem.

Tabela 20. Número de soluções com valor de RMSD menor do que 2,0 Å, considerando os complexos 2WO6 e 4MQ1 e as combinações de algoritmo de busca e função de pontuação.

PDB ^a	Combinação ^b	Soluções ^c
2WO6	<i>MolDock Optimizer/MolDock Score</i>	1
	<i>MolDock SE/MolDock Score</i>	1
	<i>Iterated Simplex/ MolDock Score</i>	2
	<i>MolDock Optimizer/PLANTS Score</i>	1
	<i>Iterated Simplex/PLANTS Score</i>	3
4MQ1	<i>MolDock Optimizer/MolDock Score</i>	2
	<i>MolDock SE/MolDock Score</i>	3
	<i>Iterated Simplex/ MolDock Score</i>	4
	<i>MolDock Optimizer/PLANTS Score</i>	1
	<i>Iterated Simplex/PLANTS Score</i>	5

^a Código do complexo da proteína no PDB. ^b Combinação entre algoritmo de busca e função de pontuação. ^c Número de soluções com valor de RMSD menor ou igual a 2,0 Å.

Na análise da Tabela 20, é possível observar que, independente da função de pontuação e do complexo analisado, o algoritmo de busca *Iterated Simplex* retornou um maior número de soluções com valor de RMSD menor do que 2,0 Å e que a melhor função de pontuação para este algoritmo de busca foi a *PLANTS Score*, as quais obtiveram os maiores números de soluções com valor de RMSD menor do que 2,0 Å, para ambos os complexos ligante-enzima avaliados. Desta forma, a combinação *Iterated Simplex/PLANTS Score* foi utilizada para a docagem da série de derivados 6-arilquinazolina-4-aminas.

4.3.2 Docagem da Série 6-arilquinazolina-4-aminas

Para a avaliação das interações intermoleculares presentes nos complexos entre a série de derivados 6-arilquinazolina-4-aminas e a DYRK1A, foram considerados três compostos: o composto **24** (mais ativo da série); o composto **6** (menos ativo da série); o composto **29** (*outlier* segundo os modelos de CoMFA e CoMSIA). Além disso, também

foram analisados ambos os enantiômeros dos compostos **32**, **33** e **34**, para avaliar a importância do centro de quiralidade para a atividade destes compostos.

As diferentes soluções obtidas para os ligantes complexados com a DYRK1A foram avaliadas por dois tipos de função de pontuação: *PLANTS Score*, pois foi a função de pontuação escolhida para a docagem dos derivados, e *Rerank Score*, uma função que reclassifica as corridas de docagem molecular, considerando um termo de torção sp^2-sp^2 e um potencial 12-6 de Lennard-Jones. Os procedimentos de reclassificação, após o uso de funções de pontuações simples, são mais adequados para a identificação de modos de ligação de alta qualidade ao invés do uso de funções de pontuação mais avançadas (MORRIS et al., 1998).

As melhores soluções retornadas pelo MVD, classificadas tanto em função do *Rerank Score* como em função do *PLANTS Score*, eram bastante semelhantes ou até mesmo iguais entre si, com exceção das soluções obtidas para o composto **29**. Assim, arbitrariamente, foram escolhidas as melhores soluções classificadas em função do *Rerank Score* para a avaliação das interações intermoleculares entre os ligantes e a DYRK1A.

Em função de todos os derivados possuírem uma subestrutura em comum, era esperado que algumas interações com a proteína fossem comuns a todos os derivados. Estas interações em comum foram identificadas e estão representadas na Figura 36, usando o composto **34 R** como exemplo. É possível observar na Figura 36 uma ligação hidrogênio entre o átomo de nitrogênio na posição 3 do anel quinazolina e um dos átomos de hidrogênio do grupo amino da cadeia lateral do aminoácido **Lys188**. Como este grupo amino está protonado, esta ligação hidrogênio é reforçada por interação eletrostática com a carga formal positiva. Este resíduo já havia sido identificado como importante na ligação de outros inibidores da DYRK1A (NEAGOIE et al., 2012; SARNO et al., 2012).

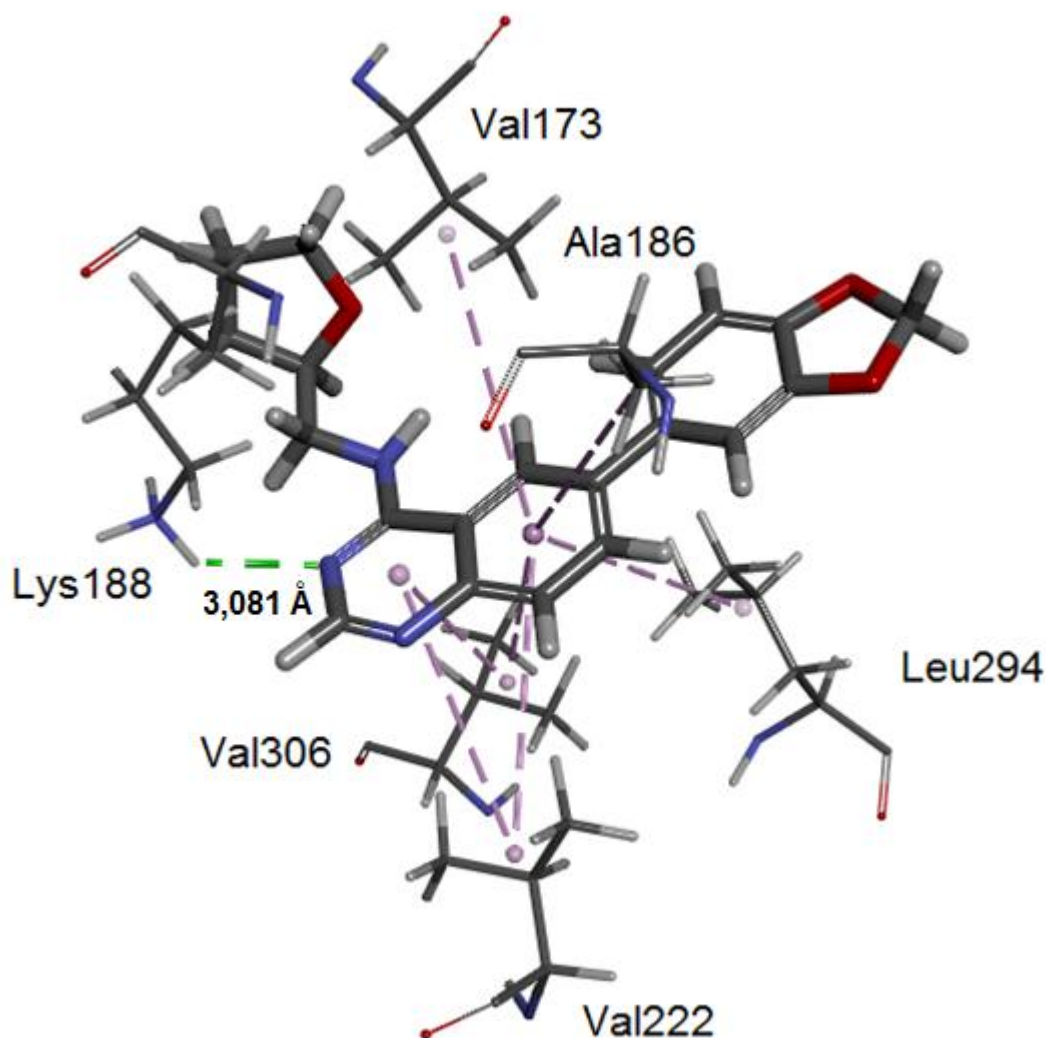


Figura 36. Interações intermoleculares (linha pontilhada) entre o anel quinazolina da série de derivados 6-arylquinazolina-4-aminas, representada pelo enantiômero *R* do composto **34**, e os principais resíduos de aminoácidos da DYRK1A. A interação via ligação hidrogênio está representada pela cor verde e as interações hidrofóbicas pela cor rosa.

Além da interação via ligação hidrogênio, podem ser observadas ainda na Figura 36 diversas interações hidrofóbicas entre o sistema de elétrons π do anel quinazolina aromático e grupos alquila dos aminoácidos **Val173**, **Ala 186**, **Val222**, **Val306** e **Leu294**. Esta quantidade de interações hidrofóbicas também é capaz de explicar o fato de diversos inibidores desta enzima possuírem anéis aromáticos em suas regiões mais centrais.

Na análise do mapa de superfície de grupos doadores e aceptores de ligação hidrogênio da DYRK1A (Figura 37), percebemos uma região de grupos doadores de ligação hidrogênio na proteína direcionados à região onde está localizado o átomo de nitrogênio na posição 3 do anel quinazolina do ligante (enantiômero *R* do composto **34**), o que reforça a importância desta estrutura comum para a atividade biológica desta série congênere. Da mesma forma, a importância da presença de grupos hidrofóbicos também é reforçada na análise do mapa de superfície hidrofóbica da DYRK1A (Figura 38), onde se percebe uma grande área hidrofóbica na proteína localizada no sítio de ligação dos inibidores desta enzima.

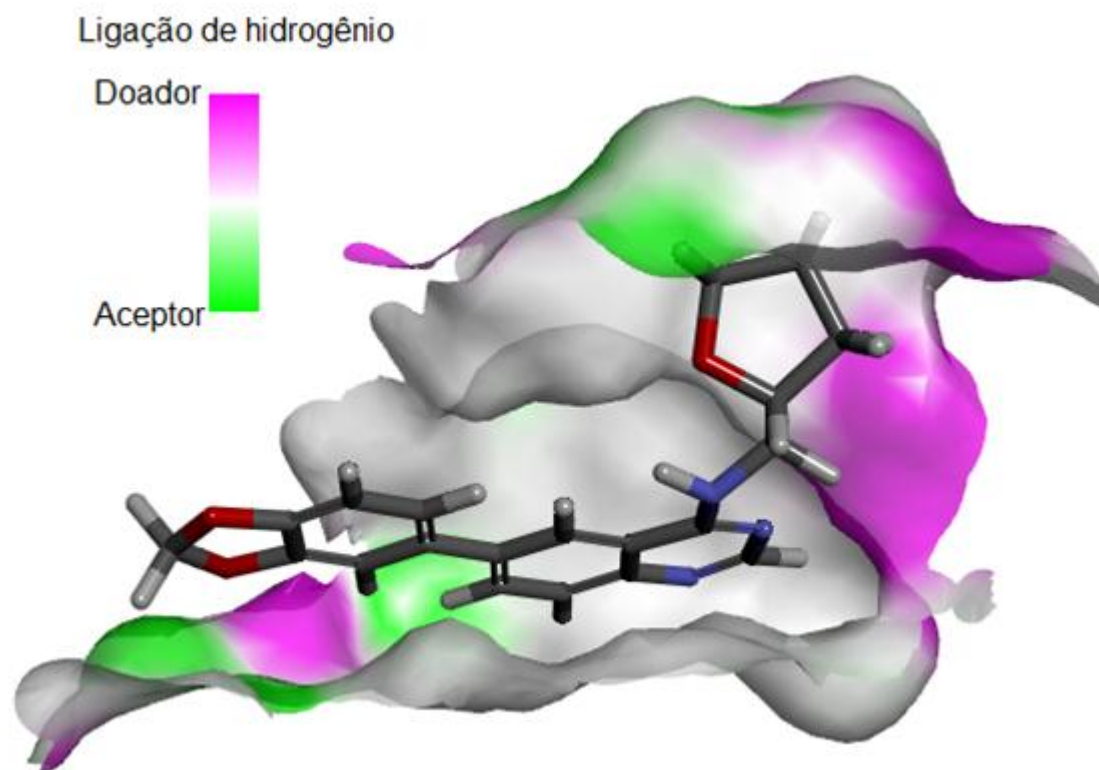


Figura 37. Mapa de superfície da DYRK1A construído no programa *Discovery Studio Visualizer* versão 4.1 (*Accelrys Software Inc*), indicando áreas da proteína contendo grupos doadores e aceptores de ligação hidrogênio, usando o enantiômero *R* do composto **34** como exemplo de ligante.

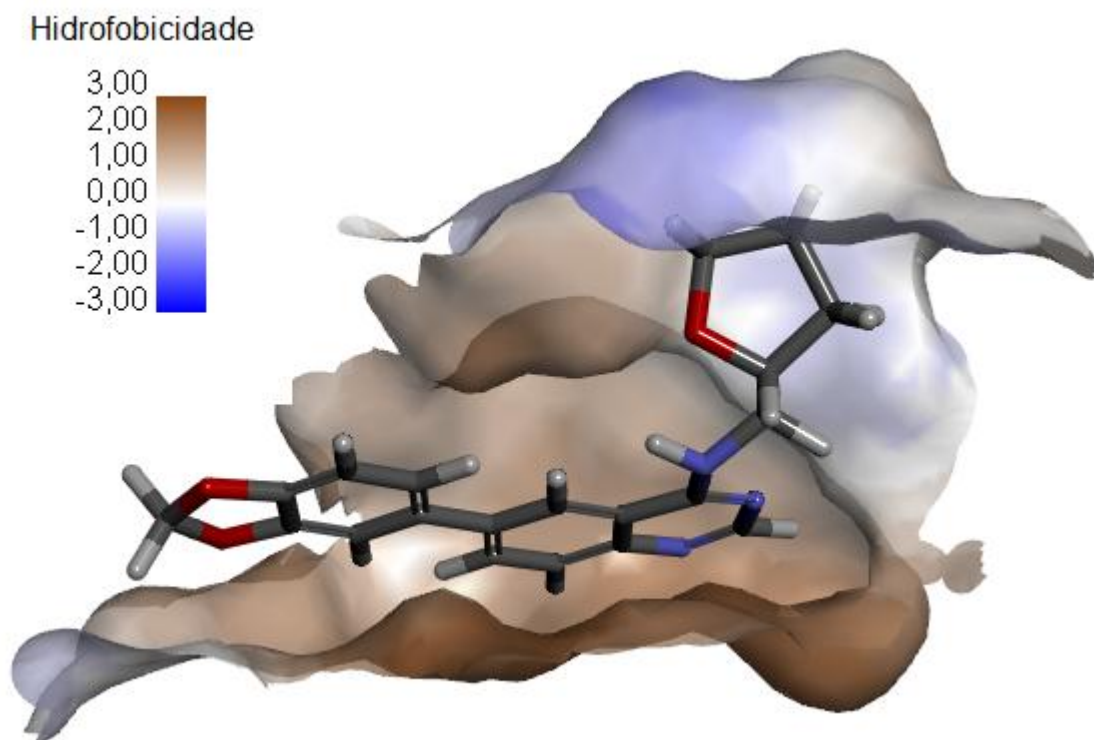


Figura 38. Mapa de superfície da DYRK1A construído no programa *Discovery Studio Visualizer*, versão 4.1 (*Accelrys Software Inc*), indicando áreas hidrofóbicas e hidrofílicas da proteína, usando o enantiômero *R* do composto **34** como exemplo de ligante.

Em relação ao composto **24** (mais ativo), é possível observar diversas interações entre o mesmo e a DYRK1A, além das interações do anel quinazolina comum a todas as estruturas (Figura 39). A exemplo do que ocorre com o anel quinazolina, o sistema π do anel 1,3-benzo-dioxola faz diversas interações hidrofóbicas com os aminoácidos **Ile165**, **Ala186**, **Leu294** e **Met240**. Interações hidrofóbicas também são observadas entre o anel tiazola e os aminoácidos **Lys188** e **Val173**. O átomo de nitrogênio do anel tiazola é capaz de interagir com a **Lys188** via ligação hidrogênio e interação dipolo-dipolo. Além das interações explicitadas anteriormente, o composto **24** ainda pode interagir via interação dipolo-dipolo com o oxigênio do grupo carboxilato do aminoácido **Asp307**.

Possivelmente, este grande número de interações com caráter polar e, portanto, mais fortes que as interações hidrofóbicas explicam a melhor atividade do composto **24** em relação aos demais derivados da série. É possível observar, inclusive, uma proximidade entre o grupo metila em R_2 e a região hidrofóbica do resíduo **Val173**, o que

provavelmente implica em uma interação de van der Waals que não existe nos compostos da série que não possuem este grupo metila, os quais são, por sua vez, em geral, menos ativos.

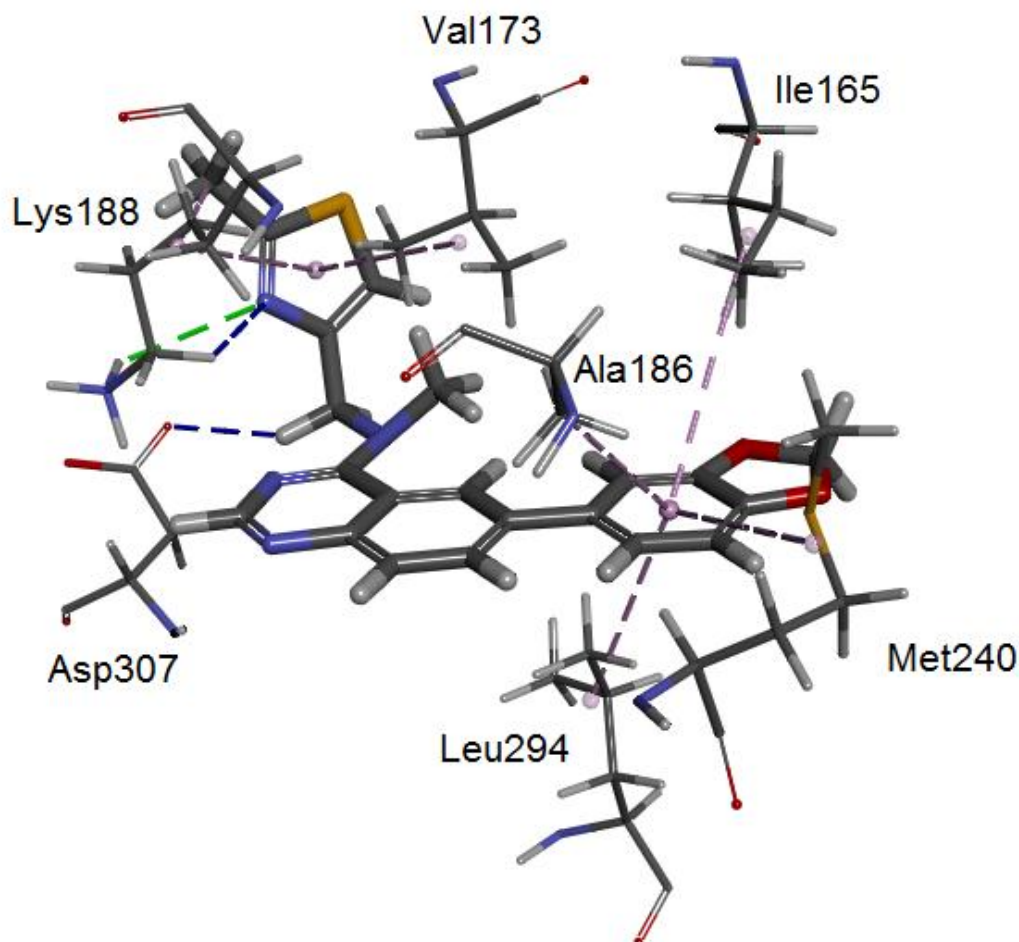


Figura 39. Interações intermoleculares (linha pontilhada) entre o composto **24** (mais ativo) e os principais resíduos de aminoácidos da DYRK1A. A ligação hidrogênio está representada pela cor verde, as interações do tipo dipolo-dipolo estão representadas pela cor azul e as interações hidrofóbicas pela cor rosa.

A análise das interações intermoleculares entre o composto **6** (menos ativo) e a DYRK1A (Figura 40) revela um padrão de interações hidrofóbicas semelhante ao que ocorre no composto **24**, pois o grupo fenila também interage com os mesmos quatro aminoácidos apolares (**Ile165**, **Ala186**, **Leu294** e **Met240**) e as mesmas interações hidrofóbicas também são observadas entre o anel heterocíclico em R₃ e os aminoácidos

Lys188 e **Val173**. A única interação com caráter polar é a ligação dipolo-dipolo com o resíduo **Asp307**, que também é observada no composto **24**. É importante ressaltar que não há ligação hidrogênio entre o anel heterocíclico em R₃ e o resíduo **Lys188**, característica que pode ser fundamental para explicar a menor atividade deste derivado. Em função da presença do átomo de cloro, duas novas interações hidrofóbicas entre este átomo e os resíduos **Ile165** e **Val173** surgem, mas estas interações não parecem contribuir para uma melhor atividade biológica deste composto.

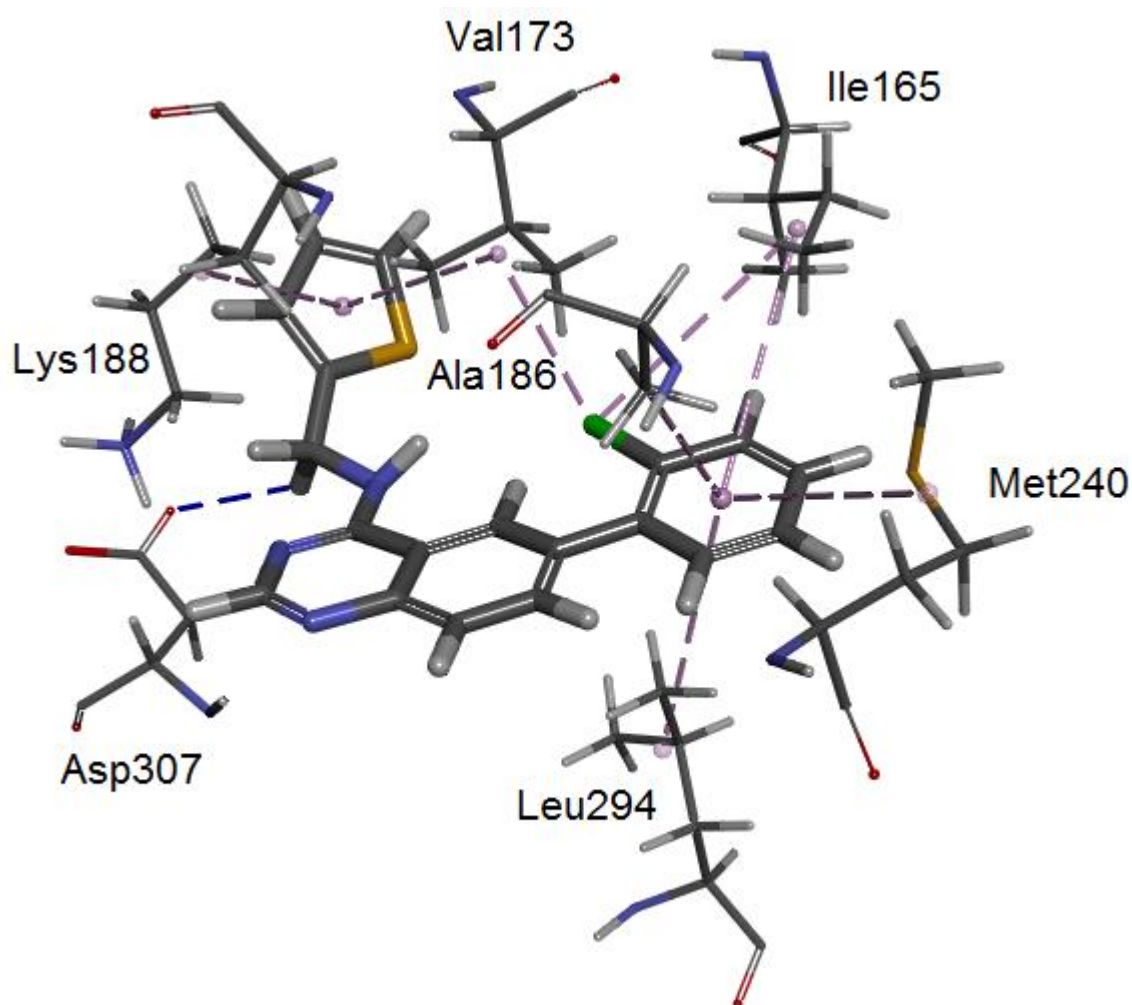


Figura 40. Interações intermoleculares (linha pontilhada) entre o composto **6** (menos ativo) e os principais resíduos de aminoácidos da DYRK1A. As interações do tipo dipolo-dipolo estão representadas pela cor azul e as interações hidrofóbicas pela cor rosa.

A análise do composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) é mais complexa, pois a solução com melhor valor de *Rerank Score* apresenta uma conformação diferente da solução com melhor valor de *PLANTS Score*. Durante as análises de CoMFA e CoMSIA, o composto **29** foi considerado um *outlier* em função da sua atividade biológica predita ser maior do que a atividade biológica experimental em mais de uma unidade logarítmica. Naquela análise, a hipótese proposta era que, provavelmente, a conformação bioativa deste composto deveria ser diferente dos demais compostos da série.

Ao visualizarmos as duas soluções lado a lado (Figura 41), verificamos que a única diferença está na conformação da molécula. A solução 1 (*PLANTS Score*) contém o anel imidazola orientado para o mesmo lado do anel 1,3-benzo-dioxola e a solução 2 (*Rerank Score*) contém o anel imidazola orientado para o lado oposto ao anel 1,3-benzo-dioxola e é a mesma conformação observada para os demais compostos da série das 6-arylquinazolina-4-aminas.

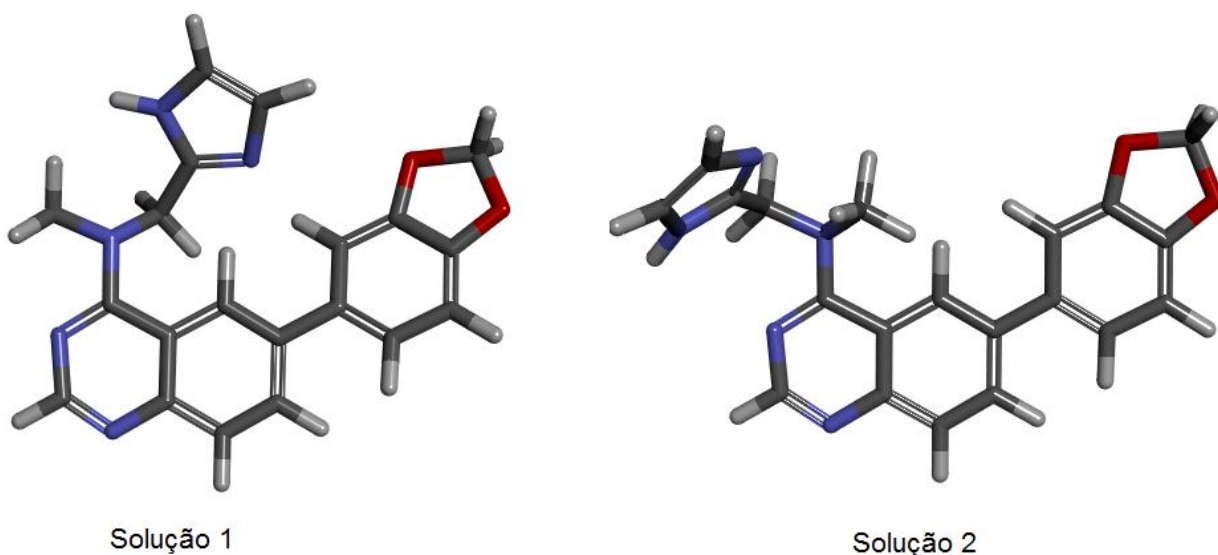


Figura 41. Estruturas químicas das soluções obtidas para o composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) durante o processo de docagem molecular. Solução 1 é a melhor obtida segundo o *PLANTS Score* e a Solução 2 é a melhor obtida segundo o *Rerank Score*.

A Tabela 21 apresenta os valores das energias internas das soluções 1 e 2 assim como os valores das energias dos complexos ligante-enzima. Quando se compara os valores de energia do sistema através dos valores de *Rerank Score*,

observa-se que solução **2** possui um valor de energia menor do que a solução **1** (-134,852 kJ/mol e -133,781 kJ/mol, respectivamente), entretanto, a diferença entre as energias das soluções é de apenas 1,071 kJ/mol, o que é um valor baixo de energia. Entretanto, quando se compara apenas as energias internas das soluções, verifica-se uma diferença pronunciada nos valores, visto que a energia interna da solução **2** é de 21,790 kJ/mol e a da solução **1** é de 11,109 kJ/mol, ou seja, pode-se afirmar que a energia interna da solução **2** é, aproximadamente, o dobro da energia interna da solução **1** e, portanto, a solução **2** representa uma conformação menos estável do que a solução **1**, mesmo possuindo energias do sistema semelhantes entre si. É importante ressaltar que essa diferença não foi encontrada para nenhum dos demais compostos da série analisados neste trabalho.

Tabela 21. Energias internas e de complexação ligante-enzima para as duas melhores soluções encontradas para o composto **29** (*outlier* nos estudos de CoMFA/CoMSIA).

Solução ^a	EC (kJ/mol) ^b	EC (kcal/mol)	EI (kJ/mol) ^c	EI (kcal/mol)
1	-133,781	-31,974	11,109	2,655
2	-134,852	-32,230	21,790	5,208

^a Número da solução obtida pela docagem. ^b EC: energia de complexação ligante-enzima obtida pela função *Rerank Score*. ^c EI: energia interna da solução obtida pela função *Rerank Score*.

Ao avaliarmos as interações intermoleculares entre a solução **1** e a DYRK1A (Figura 42), é possível observar interações hidrofóbicas entre o anel 1,3-benzo-dioxola e os resíduos **Ile165**, **Ala186** e **Leu294** e também entre o anel imidazola e o aminoácido **Val173**. Uma interação do tipo dipolo-dipolo entre o grupo metila, diretamente ligado a um átomo de nitrogênio, na posição R₂ e um átomo de oxigênio do grupo carboxilato do resíduo **Asp307** também pode ser observada. Uma interação observada para esta solução, em especial, é entre o átomo de enxofre do resíduo **Met240** e o anel 1,3-benzo-dioxola que foi anteriormente descrita em outros sistemas (ZAUHAR et al., 2000).

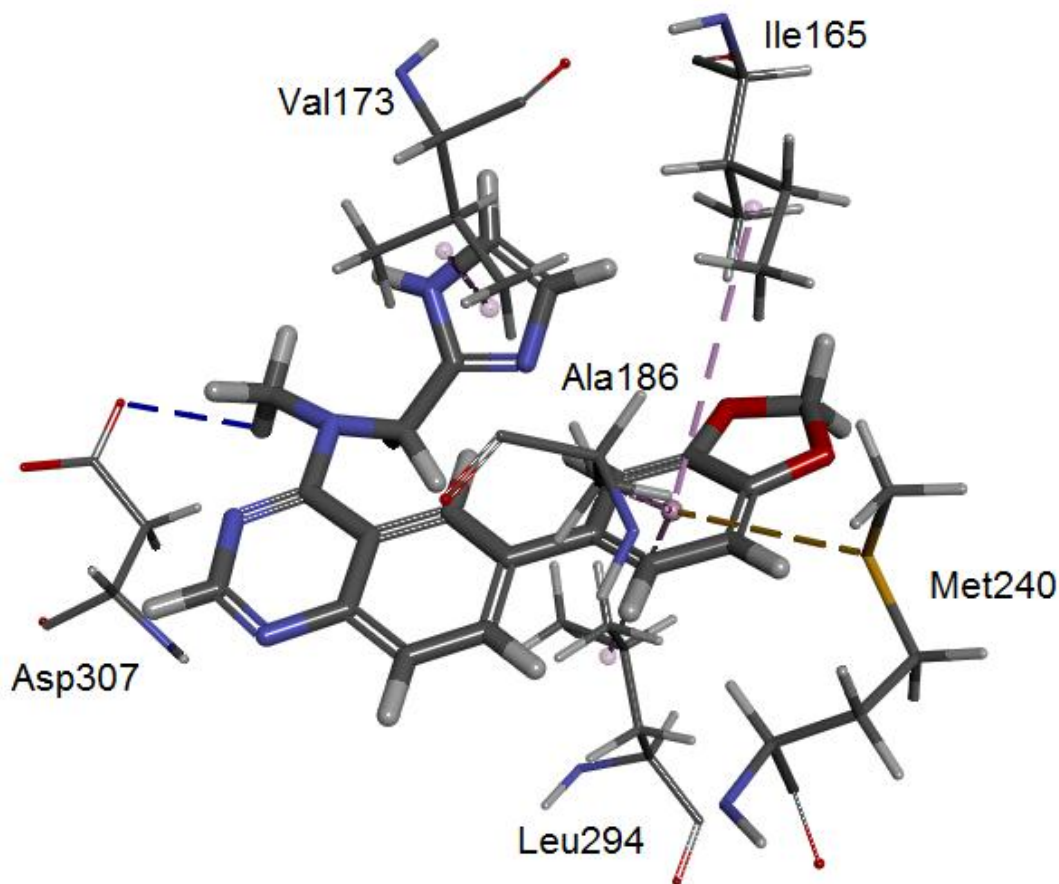


Figura 42. Interações intermoleculares (linha pontilhada) entre a solução 1 (*PLANTS Score*) do composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) e os principais resíduos de aminoácidos da DYRK1A. As interações do tipo dipolo-dipolo estão representadas pela cor azul, as interações hidrofóbicas pela cor rosa e a interação com o átomo de enxofre pela cor amarela.

A solução **2** possui as mesmas interações hidrofóbicas com os resíduos **Ile165**, **Ala186**, **Leu294** e **Val173**, porém não possui a interação dipolo-dipolo com o resíduo **Asp307**. Ademais, esta solução possui uma interação desfavorável entre dois grupos doadores de ligação hidrogênio, o hidrogênio ligado ao nitrogênio do anel imidazola e o grupo amino do resíduo **Lys188**. Em função da maior energia interna e da ausência de uma interação intermolecular desta solução com a DYRK1A, é possível sugerir que o composto **29** possui uma conformação bioativa diferente das conformações observadas para os demais derivados da série e, por isso, foi assinalado como *outlier* nos estudos de QSAR 3D.

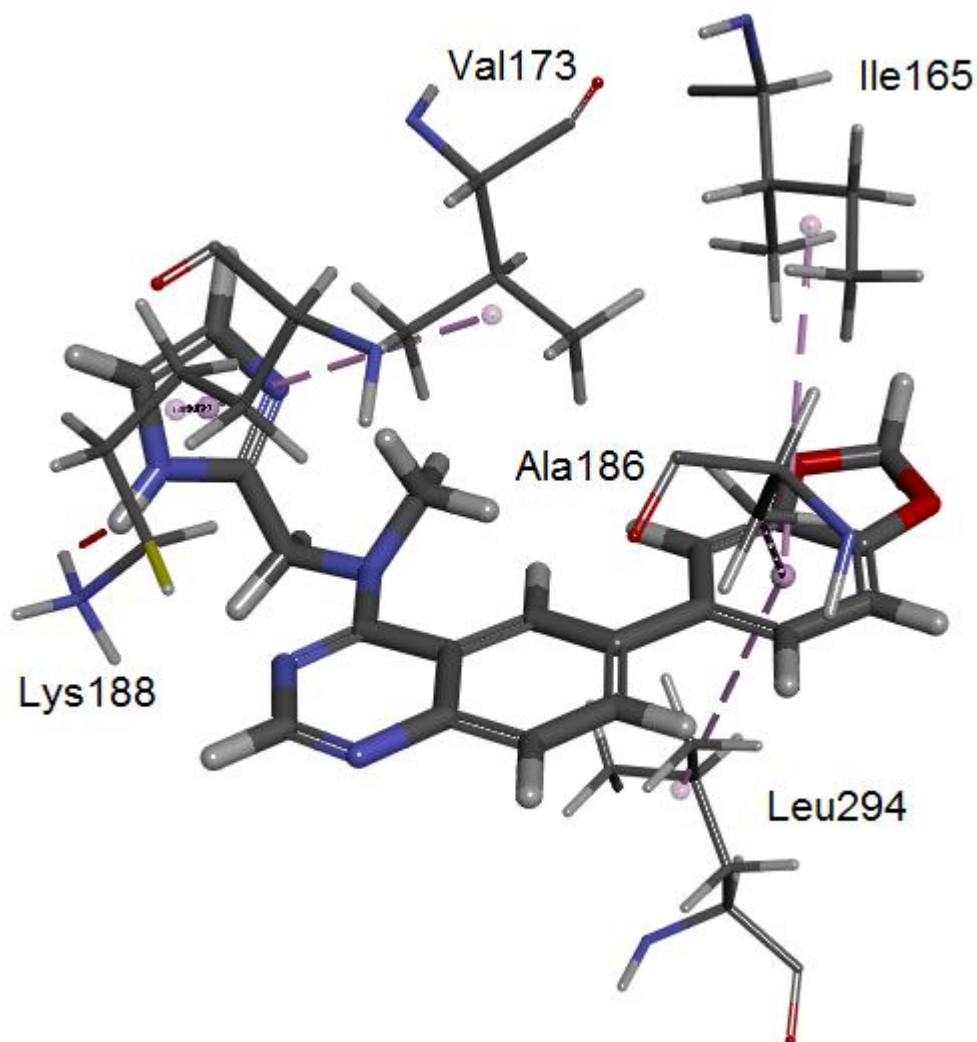


Figura 43. Interações intermoleculares (linha pontilhada) entre a solução **2** (*Rerank Score*) do composto **29** (*outlier* nos estudos de CoMFA/CoMSIA) e os principais resíduos de aminoácidos da DYRK1A. As interações hidrofóbicas pela cor rosa e a interação desfavorável entre grupos doadores de ligação hidrogênio pela cor vermelha.

Para verificar a possível influência do centro de quiralidade presente no substituinte R_3 dos compostos **32**, **33** e **34**, os pares de enantiômeros de cada composto foram docados separadamente. As pontuações dos complexos ligante-enzima são mostradas na Tabela 22. É importante ressaltar que as diferenças entre as pontuações entre cada par de enantiômeros não ultrapassou 3,0 KJ/mol (0,717 Kcal/mol) para nenhum dos três compostos. Em relação ao composto **32**, tanto o enantiômero *R* (Figura 44A) quanto o enantiômero *S* (Figura 44B), fazem apenas uma

interação hidrofóbica com um resíduo de **Val173** na região onde se localiza o centro de quiralidade e essa interação independe da orientação do átomo de oxigênio do anel tetraidrofurano em relação à enzima.

Tabela 22. Energias de complexação ligante-enzima para os pares de enantiômeros dos compostos encontradas para os compostos **32**, **33** e **34**.

#	EC (kJ/mol) ^a	EC (kcal/mol)
32(R)	-100,759	-24,066
32(S)	-98,063	-23,422
33(R)	-108,035	-25,804
33(S)	-106,001	-25,318
34(R)	-100,391	-23,978
34(S)	-98,974	-23,640

^a EC: energia de complexação ligante-enzima obtida pela função *Rerank Score*.

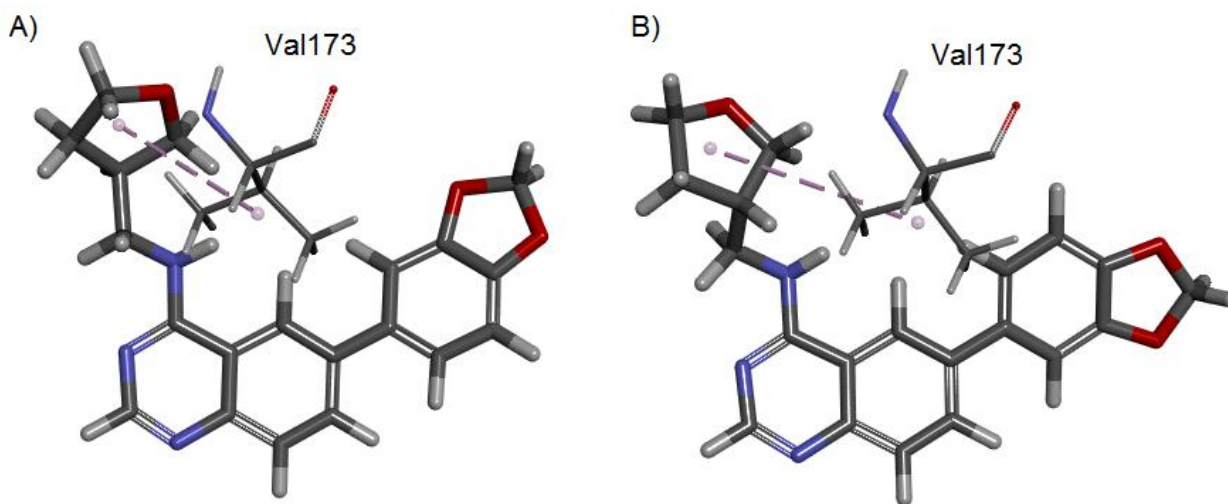


Figura 44. Interações intermoleculares (linha pontilhada) entre o anel tetraidrofurano presente na região R₃ dos enantiômeros R (A) e S (B) do composto **32** e um resíduo de aminoácido da DYRK1A. As interações hidrofóbicas são representadas pela cor rosa.

Em relação aos enantiômeros dos compostos **33** e **34** são observadas as mesmas interações entre cada par de enantiômeros. No caso composto **33** (Figura 45),

o anel tetraidropirano de cada enantiômero participa de uma interação hidrofóbica com o resíduo **Val173** e de uma interação dipolo-dipolo com o átomo de oxigênio da carbonila do resíduo **Gly171**. Os enantiômeros do composto **34** (Figura 46) fazem unicamente a interação hidrofóbica com o resíduo **Val173**.

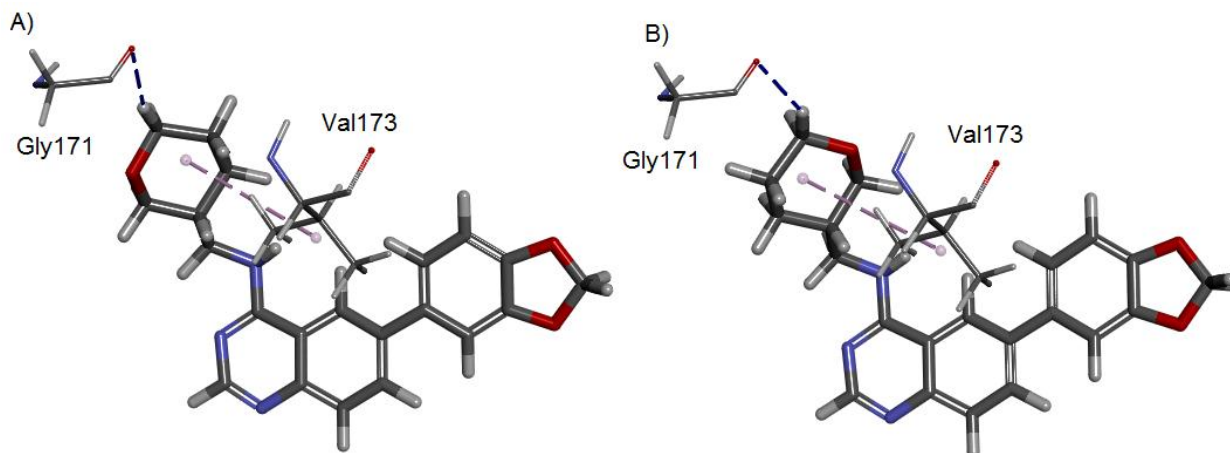


Figura 45. Interações intermoleculares (linha pontilhada) entre o anel tetraidropirano presente na região R₃ dos enantiômeros R (A) e S (B) do composto **33** e resíduos de aminoácidos da DYRK1A. As interações hidrofóbicas são representadas pela cor rosa e as interações dipolo-dipolo pela cor azul.

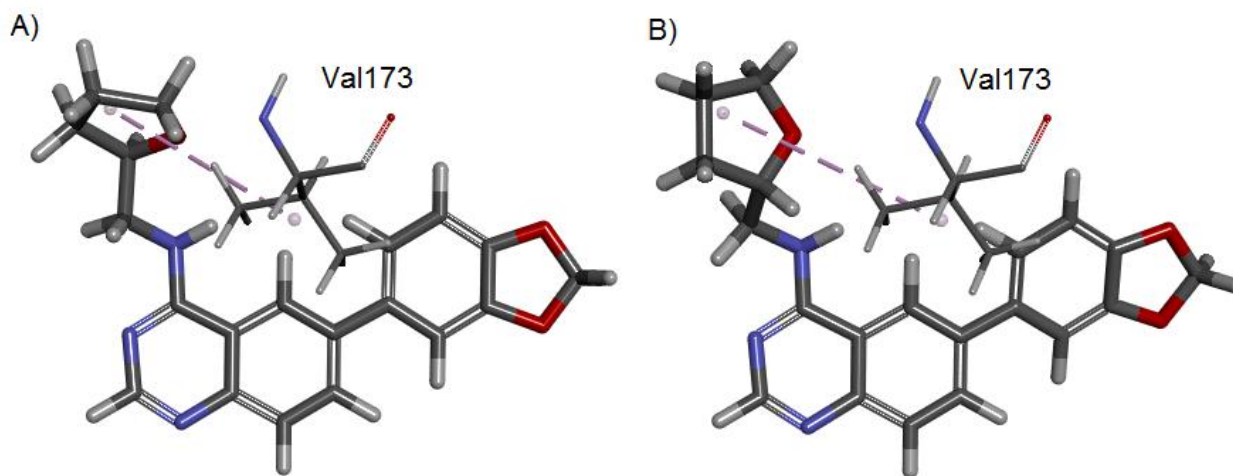


Figura 46. Interações intermoleculares (linha pontilhada) entre o anel tetraidrofurano presente na região R₃ dos enantiômeros R (A) e S (B) do composto **34** e um resíduo de aminoácido da DYRK1A. As interações hidrofóbicas são representadas pela cor rosa.

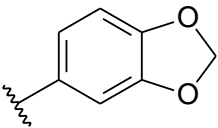
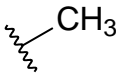
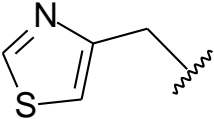
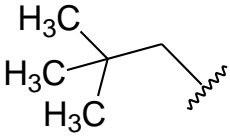
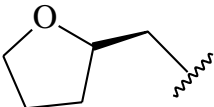
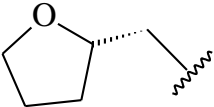
Em suma, a docagem molecular foi capaz de confirmar resultados obtidos nos estudos de QSAR, já que ela foi capaz de identificar interações importantes entre os

ligantes e a DYRK1A, identificando mais interações relevantes no composto mais ativo (**24**) do que no composto menos ativo da série (**6**), além de ter sido capaz de dar suporte a uma hipótese para o comportamento *outlier* do composto **29** e confirmado a pouca importância do centro de quiralidade presente no substituinte R₃ dos compostos **32**, **33** e **34** para a atividade biológica.

4.4 PROPOSTA DE NOVOS COMPOSTOS

A compilação dos resultados obtidos nos estudos de QSAR e de docagem molecular pode ser observada na Tabela 23. Estes resultados evidenciaram a importância do grupo metila em R₂, do anel 1,3-benzo-dioxola em R₁ e quatro estruturas diferentes em R₃. Baseados nesses resultados, quatro novos compostos foram propostos com o objetivo de gerar novos derivados da classe das 6-arylquinazolina-4-aminas os quais podem possuir melhor atividade biológica do que os derivados apresentados.

Tabela 23. Substituintes dos derivados da série das 6-arylquinazolina-4-aminas que geram melhor perfil de atividade inibitória contra a enzima DYRK1A, observados nos estudos de QSAR e docagem molecular.

Substituinte	Estrutura
R ₁	
R ₂	
R ₃	
	
R ₃	
	

O composto **A1** (Figura 47A) é um análogo do **36** ($pIC_{50} = 7,77$), cuja modificação estrutural, refletida na troca do átomo de hidrogênio por um grupo metila, poderia aumentar ainda mais a atividade biológica, visto que o composto **36** é um dos mais potentes da série ($pIC_{50} = 7,77$). O composto **A2** (Figura 47B) é um análogo do **24** ($pIC_{50} = 7,85$) sem o grupo metila ligado ao anel tiazolila, pois a importância de grupos metila ligados a anéis aromáticos localizados na posição R_3 não foi totalmente confirmada.

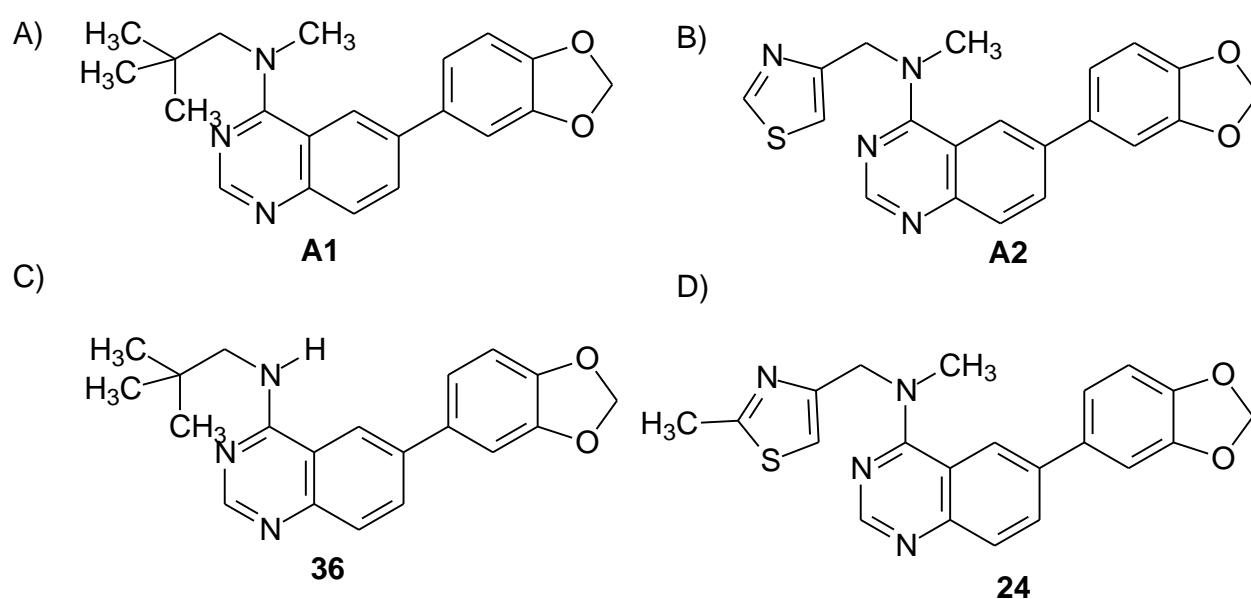


Figura 47. Estruturas químicas dos compostos propostos **A1** e **A2** e dos seus respectivos análogos. (A) Composto **A1** é análogo ao composto **36**. (B) Composto **A2** é análogo ao composto **24**. (C) Composto **36**. (D) Composto **24**.

Os compostos **A3** e **A4** (Figura 48A e B) são resultantes da mesma modificação estrutural feita para o composto **A1**, mas em relação aos enantiômeros do composto **34** ($pIC_{50} = 7,52$).

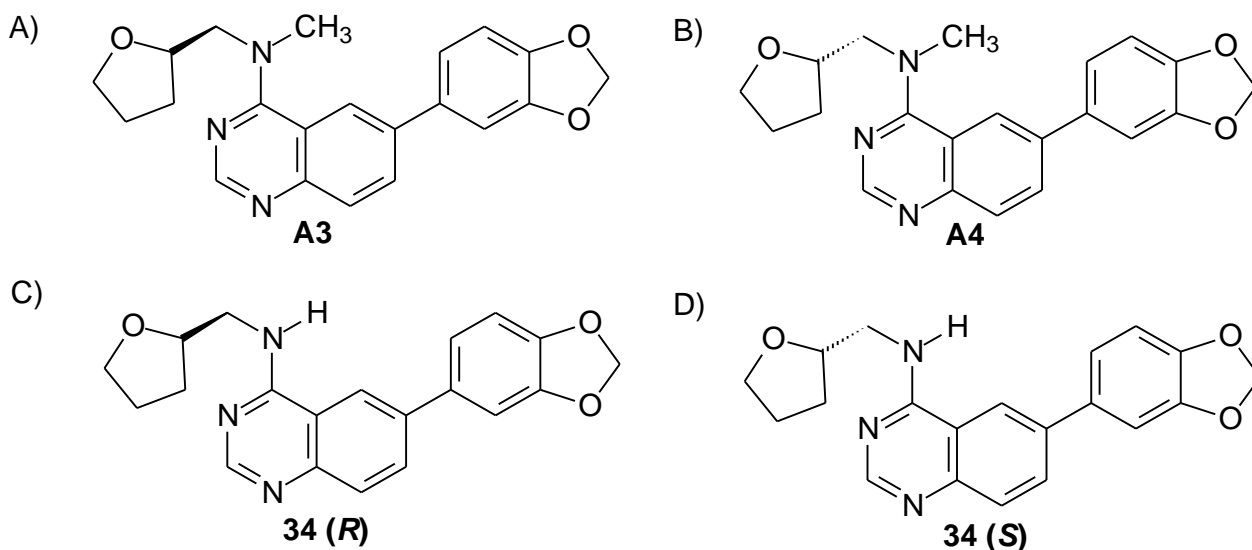


Figura 48. Estruturas químicas dos compostos propostos **A3** e **A4** e dos seus respectivos análogos. (A) Composto **A3** é análogo ao enantiômero (*R*) do composto **34**. (B) Composto **A4** é análogo ao enantiômero (*S*) do composto **34**. (C) Enantiômero (*R*) do composto **34**. (D) Enantiômero (*S*) do composto **34**.

Estes quatro compostos possuem grupos considerados importantes para a atividade biológica da série de derivados 6-arylquinazolina-4-aminas e podem se tornar candidatos a novos fármacos para o tratamento da DA por inibição da DYRK1A. A Tabela 24 contém os valores preditos por todos os modelos de QSAR estudados para cada um dos compostos propostos.

Tabela 24. Valores de pIC₅₀ (M) preditos para os compostos propostos (**A1**, **A2**, **A3** e **A4**) e seus análogos correspondentes (**36**, **24**, **34 (R)** e **34 (S)**, respectivamente) segundo os melhores modelos de QSAR obtidos.

Composto	HQSAR	CoMFA	CoMSIA
A1	6,95	7,32	7,35
36	6,93	7,13	6,87
A2	8,16	7,40	7,30
24	7,92	7,68	7,32
A3	7,50	7,46	7,34
34 (R)	7,20	6,98	6,81
A4	7,50	7,12	7,31
34 (S)	6,97	7,08	6,74

De acordo com os dados da Tabela 24, os compostos **A1**, **A3** e **A4** apresentam todos os valores de atividade biológica preditos pelos modelos de HQSAR, CoMFA e CoMSIA superiores aos valores dos seus respectivos análogos desmetilados (composto **36**, **34 R** e **34 S**), o que reforça o impacto positivo do grupo metila na posição R₂, segundo os três modelos de QSAR estudados.

Para o composto **A2**, observa-se que apenas o modelo de HQSAR prevê a sua atividade biológica como sendo maior do que a do seu análogo (composto **24**), apresentando, inclusive, um valor de pIC₅₀ acima de 8,0 M, algo não obtido por nenhum dos compostos da série das 6-arilquinazolina-4-aminas. Os modelos de CoMFA e CoMSIA preveem valores de pIC₅₀ ligeiramente mais baixos para este composto em comparação ao análogo **24**, mas ainda, assim, valores altos de atividade biológica quando comparado com diversos outros compostos desta série. Este composto **A2**, além de possuir um alto valor de pIC₅₀, não possui um grupo metila ligado a um anel aromático, conhecida como posição benzílica, a qual é extremamente suscetível a reações de oxidação catalisadas por enzimas do citocromo P450 e, portanto, provavelmente, possui um tempo de meia-vida mais alto do que o composto **24**, uma característica que pode ser interessante do ponto de vista terapêutico.

Em função das características apresentadas por estes novos análogos propostos, todos os compostos podem ser sintetizados e testados para a confirmação de suas atividades biológicas, pois se tratam de potenciais fármacos desta série congênere para o tratamento da DA.

5 CONCLUSÕES

Neste trabalho, foram construídos e avaliados modelos de QSAR 2D (HQSAR) e 3D (CoMFA e CoMSIA) de uma série congênere de 46 derivados 6-arylquinazolina-4-aminas inibidores da enzima DYRK1A, onde o conjunto de treinamento possui 36 compostos e o conjunto de teste possui 10 compostos. Além disso, os derivados desta série foram submetidos à docagem molecular no sítio ativo da enzima para a predição do modo de ligação.

Todos os modelos de QSAR obtidos apresentaram boa capacidade preditiva (HQSAR: $q^2 = 0,757$; CoMFA: $q^2 = 0,608$; CoMSIA: $q^2 = 0,713$) e, portanto, podem ser utilizados para a predição da atividade biológica de novos compostos. Em geral, estes modelos apontaram a importância de grupos metila na posição R_2 e do anel 1,3-benzodioxola na posição R_1 . Como a posição R_3 apresenta uma maior possibilidade de substituintes, observou-se que a presença de anéis aromáticos heterocíclicos (metilados ou não) e o grupo *tert*-butila nesta posição geram derivados com alta atividade biológica.

A docagem molecular permitiu a identificação de interações comuns a todos os compostos da série congênere, como uma ligação hidrogênio reforçada por carga entre o átomo de nitrogênio na posição 3 do anel quinazolina e o grupo amino protonado da **Lys188** e, também, interações diferentes entre os compostos e a enzima, que podem explicar a variação nos valores de atividade biológica. Além disso, a docagem molecular conseguiu dar suporte a hipóteses geradas durante os estudos de QSAR 3D, como a conformação bioativa sugerida para o composto **29** (um *outlier* do CoMFA/ CoMSIA) e a irrelevância de um centro de quiralidade presente no substituinte R_3 de três compostos da série.

Todas estas informações obtidas neste trabalho permitiram a proposta de quatro novos compostos desta série, os quais podem ser importantes moléculas para o tratamento contra a DA.

6 PERSPECTIVAS

A DA é uma patologia neurodegenerativa caracterizada por um estado de demência progressiva e que, usualmente, leva ao falecimento em aproximadamente nove anos (CITRON, 2010). A neurodegeneração ocorre, principalmente, devido ao acúmulo de ENFs e à β -amiloidose, mecanismos patogênicos que contam com a participação da DYRK1A, o que desperta interesse nesta enzima como um novo alvo terapêutico no tratamento desta patologia (WEGIEL; GONG; HWANG, 2011).

A síntese de inibidores da DYRK1A tem sido o esforço de diversos grupos de pesquisa no mundo, mas poucos conseguem gerar uma quantidade de derivados suficientes para a construção de modelos relacionando a estrutura e a atividade biológica de forma quantitativa (SMITH et al., 2012).

Em função da capacidade da modelagem molecular de identificar modos de ligação entre ligante e enzima e, também, de gerar modelos com alta capacidade preditiva, novos estudos devem ser realizados com o objetivo de aumentar o conhecimento sobre estes sistemas. Entre os estudos que podem ser realizados, pode ser citada a simulação por dinâmica molecular, que pode gerar informações não acessíveis por estudos *in silico* estáticos e *in vitro*.

Além de avançar nos estudos de modelagem molecular, a perspectiva é que os derivados propostos possam ser sintetizados e avaliados como inibidores da DYRK1A, já que o processo sintético é bastante viável, a caracterização química dos produtos é simples e o teste *in vitro* também é factível. Dessa forma os derivados propostos serão validados e, enfim, disponibilizados para ensaios pré-clínicos e clínicos futuros.

REFERÊNCIAS

- ADAYEV, T.; CHEN-HWANG, M.-C.; MURAKAMI, N.; LEE, E.; BOLTON, D. C.; HWANG, Y.-W. Dual-specificity tyrosine phosphorylation-regulated kinase 1A does not require tyrosine phosphorylation for activity in vitro. **Biochemistry**, v. 46, n. 25, p. 7614–7624, 2007.
- ADAYEV, T.; CHEN-HWANG, M.-C.; MURAKAMI, N.; WEGIEL, J.; HWANG, Y.-W. Kinetic properties of a MNB/DYRK1A mutant suitable for the elucidation of biochemical pathways. **Biochemistry**, v. 45, n. 39, p. 12011–12019, 2006.
- AHMADU, A.; ABDULKARIM, A.; GROUGNET, R.; MYRIANTHOPOULOS, V.; TILLEQUIN, F.; MAGIATIS, P.; SKALTSOUNIS, A. L. Two new peltogynoids from acacia nilotica delile with kinase inhibitory activity. **Planta Medica**, v. 76, n. 5, p. 458–460, 2010.
- AIRAKSINEN, M. M.; LECKLIN, A.; SAANO, V.; TUOMISTO, L.; GYNTER, J. Tremorigenic effect and inhibition of tryptamine and serotonin receptor binding by beta-carbolines. **Pharmacology & toxicology**, v. 60, n. 1, p. 5–8, 1987.
- ALBUQUERQUE, M. A.; BRITO, M. A.; CUNHA, E. F. F.; ALENCASTRO, R. B.; ANTUNES, O. A. C.; CASTRO, H. C.; RODRIGUES, C. R. Multidimensional-QSAR: beyond the third-dimension in drug design. In: TAFT, C. A.; SILVA, C. H. T. P. (Eds.). **Current Methods In Medicinal Chemistry and Biological Physics**. 1st. ed. [s.l.] Research Signpost, 2007. p. 91–101.
- ALZHEIMER, A.; STELZMANN, R. A.; SCHNITZLEIN, H. N.; MURTAGH, F. R. An English translation of Alzheimer's 1907 paper, "Über eine eigenartige Erkrankung der Hirnrinde". **Clinical anatomy (New York, N.Y.)**, v. 8, n. 6, p. 429–431, 1995.
- ANDERSON, K.; CHEN, Y.; CHEN, Z.; DOMINIQUE, R.; GLENN, K.; HE, Y.; JANSON, C.; LUK, K.-C.; LUKACS, C.; POLONSKAIA, A.; QIAO, Q.; RAILKAR, A.; ROSSMAN, P.; SUN, H.; XIANG, Q.; VILENCHIK, M.; WOVKULICH, P.; ZHANG, X. Pyrido[2,3-d]pyrimidines: discovery and preliminary SAR of a novel series of DYRK1B and DYRK1A inhibitors. **Bioorganic & medicinal chemistry letters**, v. 23, n. 24, p. 6610–6615, 2013.
- ARAÚJO, J. Q.; LIMA, J. A.; PINTO, A. DA C.; DE ALENCASTRO, R. B.; ALBUQUERQUE, M. G. Docking of the alkaloid geissospermine into acetylcholinesterase: a natural scaffold targeting the treatment of Alzheimer's disease. **Journal of molecular modeling**, v. 17, n. 6, p. 1401–1412, 2011.
- BAIN, J.; MCLAUCHLAN, H.; ELLIOTT, M.; COHEN, P. The specificities of protein kinase inhibitors: an update. **The Biochemical journal**, v. 371, n. Pt 1, p. 199–204, 2003.

BAIN, J.; PLATER, L.; ELLIOTT, M.; SHPIRO, N.; HASTIE, C. J.; MCLAUCHLAN, H.; KLEVERNIC, I.; ARTHUR, J. S. C.; ALESSI, D. R.; COHEN, P. The selectivity of protein kinase inhibitors: a further update. **The Biochemical journal**, v. 408, n. 3, p. 297–315, 2007.

BARKER, W. W.; LUIS, C. A.; KASHUBA, A.; LUIS, M.; HARWOOD, D. G.; LOEWENSTEIN, D.; WATERS, C.; JIMISON, P.; SHEPHERD, E.; SEVUSH, S.; GRAFF-RADFORD, N.; NEWLAND, D.; TODD, M.; MILLER, B.; GOLD, M.; HEILMAN, K.; DOTY, L.; GOODMAN, I.; ROBINSON, B.; PEARL, G.; DICKSON, D.; DUARA, R. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. **Alzheimer disease and associated disorders**, v. 16, n. 4, p. 203–212, 2002.

BECKER, W.; JOOST, H. G. Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. **Progress in nucleic acid research and molecular biology**, v. 62, p. 1–17, 1999.

BECKER, W.; SIPPL, W. Activation, regulation, and inhibition of DYRK1A. **The FEBS journal**, v. 278, n. 2, p. 246–256, 2011.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The Protein Data Bank. **Nucleic acids research**, v. 28, n. 1, p. 235–242, 2000.

BHANSALI, S. G.; KULKARNI, V. M. Combined 2D and 3D-QSAR, molecular modelling and docking studies of pyrazolodiazepinones as novel phosphodiesterase 2 inhibitors. **SAR and QSAR in environmental research**, v. 25, n. 11, p. 905–937, 2014.

BIRKS, J.; FLICKER, L. Selegiline for Alzheimer's disease. **The Cochrane database of systematic reviews**, n. 1, p. CD000442, 2003.

BIRKS, J.; GRIMLEY EVANS, J. Ginkgo biloba for cognitive impairment and dementia. **The Cochrane database of systematic reviews**, n. 2, p. CD003120, 2007.

BÖHM, M.; ST RZEBECHER, J.; KLEBE, G. Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and f. **Journal of medicinal chemistry**, v. 42, n. 3, p. 458–477, 1999.

BRAAK, H.; BRAAK, E. Neuropathological staging of Alzheimer-related changes. **Acta neuropathologica**, v. 82, n. 4, p. 239–259, 1991.

BRITO, M. A. **QSAR-3D de Inibidores Não Nucleosídeos da Transcriptase Reversa do HIV-1: Estudos Independente e Dependente da Enzima**. 2008. 231 f. Tese de

Doutorado em Química Orgânica) - Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.

BROOIJMANS, N.; KUNTZ, I. D. Molecular recognition and docking algorithms. **Annual review of biophysics and biomolecular structure**, v. 32, p. 335–373, 2003.

BUTCHER, E. C.; BERG, E. L.; KUNKEL, E. J. Systems biology in drug discovery. **Nature biotechnology**, v. 22, n. 10, p. 1253–1259, 2004.

CALLAWAY, J. C.; MCKENNA, D. J.; GROB, C. S.; BRITO, G. S.; RAYMON, L. P.; POLAND, R. E.; ANDRADE, E. N.; ANDRADE, E. O.; MASH, D. C. Pharmacokinetics of Hoasca alkaloids in healthy humans. **Journal of ethnopharmacology**, v. 65, n. 3, p. 243–256, 1999.

CHEN, Y.-P. P.; CHEN, F. Identifying targets for drug discovery using bioinformatics. **Expert opinion on therapeutic targets**, v. 12, n. 4, p. 383–389, 2008.

CHENG, T.; LI, X.; LI, Y.; LIU, Z.; WANG, R. Comparative assessment of scoring functions on a diverse test set. **Journal of chemical information and modeling**, v. 49, n. 4, p. 1079–1093, 2009.

CHENG, Y.; ZHOU, M.; TUNG, C.-H.; JI, M.; ZHANG, F. Studies on two types of PTP1B inhibitors for the treatment of type 2 diabetes: Hologram QSAR for OBA and BBB analogues. **Bioorganic & medicinal chemistry letters**, v. 20, n. 11, p. 3329–3337, 2010.

CHO, S. J.; TROPSHA, A. Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. **Journal of medicinal chemistry**, v. 38, n. 7, p. 1060–1066, 1995.

CITRON, M. Alzheimer's disease: strategies for disease modification. **Nature reviews. Drug discovery**, v. 9, n. 5, p. 387–398, 2010.

COHEN, N. C.; BLANEY, J. M.; HUMBLET, C.; GUND, P.; BARRY, D. C. Molecular modeling software and methods for medicinal chemistry. **Journal of medicinal chemistry**, v. 33, n. 3, p. 883–894, 1990.

COZZA, G.; MAZZORANA, M.; PAPINUTTO, E.; BAIN, J.; ELLIOTT, M.; DI MAIRA, G.; GIANONCELLI, A.; PAGANO, M. A.; SARNO, S.; RUZZENE, M.; BATTISTUTTA, R.; MEGGIO, F.; MORO, S.; ZAGOTTO, G.; PINNA, L. A. Quinalizarin as a potent, selective and cell-permeable inhibitor of protein kinase CK2. **The Biochemical journal**, v. 421, n. 3, p. 387–395, 2009.

CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, v. 110, n. 18, p. 5959–5967, 1988.

DA CUNHA, E. F. F. **Estudos Quantitativos de Correlação Estrutura-Atividade em 3D e 4D de inibidores peptídicos da HIV-1 Protease**. 2006. 165 f. Tese (Doutorado em Química Orgânica) - Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

DE AZEVEDO, W. F.; DIAS, R. Computational methods for calculation of ligand-binding affinity. **Current drug targets**, v. 9, n. 12, p. 1031–1039, 2008.

DE SOUZA, S. D.; DE SOUZA, A. M. T.; DE SOUSA, A. C. C.; SODERO, A. C. R.; CABRAL, L. M.; ALBUQUERQUE, M. G.; CASTRO, H. C.; RODRIGUES, C. R. Hologram QSAR models of 4-[(diethylamino)methyl]-phenol inhibitors of acetyl/butrylcholinesterase enzymes as potential anti-Alzheimer agents. **Molecules (Basel, Switzerland)**, v. 17, n. 8, p. 9529–9539, 2012.

DEWAR, M. J. S.; ZOEBISCH, E. G.; HEALY, E. F.; STEWART, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. **Journal of the American Chemical Society**, v. 107, n. 13, p. 3902–3909, 1985.

DREWS, J. Drug discovery: a historical perspective. **Science (New York, N.Y.)**, v. 287, n. 5460, p. 1960–1964, 2000.

ECHALIER, A.; BETTAYEB, K.; FERANDIN, Y.; LOZACH, O.; CLÉMENT, M.; VALETTE, A.; LIGER, F.; MARQUET, B.; MORRIS, J. C.; ENDICOTT, J. A.; JOSEPH, B.; MEIJER, L. Meriolins (3-(pyrimidin-4-yl)-7-azaindoles): synthesis, kinase inhibitory activity, cellular effects, and structure of a CDK2/cyclin A/meriolin complex. **Journal of medicinal chemistry**, v. 51, n. 4, p. 737–751, 2008.

ENGEL, T. Basic overview of chemoinformatics. **Journal of chemical information and modeling**, v. 46, n. 6, p. 2267–2277, 2006.

ESPOSITO, E. X.; HOPFINGER, A. J.; MADURA, J. D. Methods for applying the quantitative structure-activity relationship paradigm. **Methods in molecular biology (Clifton, N.J.)**, v. 275, p. 131–214, 2004.

EVANS, R. M.; HUI, S.; PERKINS, A.; LAHIRI, D. K.; POIRIER, J.; FARLOW, M. R. Cholesterol and APOE genotype interact to influence Alzheimer disease progression. **Neurology**, v. 62, n. 10, p. 1869–1871, 2004.

EWING, T. J.; MAKINO, S.; SKILLMAN, A. G.; KUNTZ, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. **Journal of computer-aided molecular design**, v. 15, n. 5, p. 411–428, 2001.

FERRER, I.; BARRACHINA, M.; PUIG, B.; MARTÍNEZ DE LAGRÁN, M.; MARTÍ, E.; AVILA, J.; DIERSSEN, M. Constitutive Dyrk1A is abnormally expressed in Alzheimer

disease, Down syndrome, Pick disease, and related transgenic models. **Neurobiology of disease**, v. 20, n. 2, p. 392–400, 2005.

FORLENZA, O. V. Tratamento farmacológico da doença de Alzheimer. **Revista de Psiquiatria Clínica**, v. 32, n. 3, p. 137–148, 2005.

FREE, S. M.; WILSON, J. W. A Mathematical Contribution to Structure-Activity Studies. **Journal of Medicinal Chemistry**, v. 7, n. 4, p. 395–399, 1964.

FRIESNER, R. A.; BANKS, J. L.; MURPHY, R. B.; HALGREN, T. A.; KLICIC, J. J.; MAINZ, D. T.; REPASKY, M. P.; KNOLL, E. H.; SHELLEY, M.; PERRY, J. K.; SHAW, D. E.; FRANCIS, P.; SHENKIN, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. **Journal of medicinal chemistry**, v. 47, n. 7, p. 1739–1749, 2004.

FROST, D.; MEECHOOVET, B.; WANG, T.; GATELY, S.; GIORGETTI, M.; SHCHERBAKOVA, I.; DUNCKLEY, T. β -carboline compounds, including harmine, inhibit DYRK1A and tau phosphorylation at multiple Alzheimer's disease-related sites. **PloS one**, v. 6, n. 5, p. e19264, 2011.

GALCERAN, J.; DE GRAAF, K.; TEJEDOR, F. J.; BECKER, W. The MNB/DYRK1A protein kinase: genetic and biochemical properties. **Journal of neural transmission. Supplementum**, n. 67, p. 139–148, 2003.

GEHLHAAR, D. K.; BOUZIDA, D.; REJTO, P. A. **Evolutionary Programming VII**. Berlin, Heidelberg: Springer Berlin Heidelberg, 10 Apr. 1998. v. 1447p. 449–461

GEHLHAAR, D. K.; VERKHIVKER, G. M.; REJTO, P. A.; SHERMAN, C. J.; FOGEL, D. B.; FOGEL, L. J.; FREER, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. **Chemistry & biology**, v. 2, n. 5, p. 317–324, 1995.

GIRAUD, F.; ALVES, G.; DEBITON, E.; NAUTON, L.; THÉRY, V.; DURIEU, E.; FERANDIN, Y.; LOZACH, O.; MEIJER, L.; ANIZON, F.; PEREIRA, E.; MOREAU, P. Synthesis, protein kinase inhibitory potencies, and in vitro antiproliferative activities of meridianin derivatives. **Journal of medicinal chemistry**, v. 54, n. 13, p. 4474–4489, 2011.

GOATE, A.; CHARTIER-HARLIN, M. C.; MULLAN, M.; BROWN, J.; CRAWFORD, F.; FIDANI, L.; GIUFFRA, L.; HAYNES, A.; IRVING, N.; JAMES, L. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. **Nature**, v. 349, n. 6311, p. 704–706, 1991.

GÖCKLER, N.; JOFRE, G.; PAPADOPOULOS, C.; SOPPA, U.; TEJEDOR, F. J.; BECKER, W. Harmine specifically inhibits protein kinase DYRK1A and interferes with neurite formation. **The FEBS journal**, v. 276, n. 21, p. 6324–6337, 2009.

- GOEDERT, M.; JAKES, R. Mutations causing neurodegenerative tauopathies. **Biochimica et biophysica acta**, v. 1739, n. 2-3, p. 240–250, 2005.
- GOLBRAIKH, A.; BERNARD, P.; CHRÉTIEN, J. R. Validation of protein-based alignment in 3D quantitative structure-activity relationships with CoMFA models. **European journal of medicinal chemistry**, v. 35, n. 1, p. 123–136, 2000.
- GOLDE, T. E.; ECKMAN, C. B.; YOUNKIN, S. G. Biochemical detection of A β isoforms: implications for pathogenesis, diagnosis, and treatment of Alzheimer's disease. **Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease**, v. 1502, n. 1, p. 172–187, 2000.
- GREEN, P. S.; GRIDLEY, K. E.; SIMPKINS, J. W. Estradiol protects against beta-amyloid (25-35)-induced toxicity in SK-N-SH human neuroblastoma cells. **Neuroscience letters**, v. 218, n. 3, p. 165–168, 1996.
- GRUNDKE-IQBAL, I.; IQBAL, K.; TUNG, Y. C.; QUINLAN, M.; WISNIEWSKI, H. M.; BINDER, L. I. Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 13, p. 4913–4917, 1986.
- GUEDES, I. A.; DE MAGALHÃES, C. S.; DARDENNE, L. E. Receptor–ligand molecular docking. **Biophysical Reviews**, v. 6, n. 1, p. 75–87, 2013.
- GUEDJ, F.; SÉBRIÉ, C.; RIVALS, I.; LEDRU, A.; PALY, E.; BIZOT, J. C.; SMITH, D.; RUBIN, E.; GILLET, B.; ARBONES, M.; DELABAR, J. M. Green tea polyphenols rescue of brain defects induced by overexpression of DYRK1A. **PloS one**, v. 4, n. 2, p. e4606, 2009.
- GUIMERÁ, J.; CASAS, C.; PUCHARCÒS, C.; SOLANS, A.; DOMÈNECH, A.; PLANAS, A. M.; ASHLEY, J.; LOVETT, M.; ESTIVILL, X.; PRITCHARD, M. A. A human homologue of Drosophila minibrain (MNB) is expressed in the neuronal regions affected in Down syndrome and maps to the critical region. **Human molecular genetics**, v. 5, n. 9, p. 1305–1310, 1996.
- HALPERIN, I.; MA, B.; WOLFSON, H.; NUSSINOV, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. **Proteins**, v. 47, n. 4, p. 409–443, 2002.
- HÄMMERLE, B.; ELIZALDE, C.; GALCERAN, J.; BECKER, W.; TEJEDOR, F. J. The MNB/DYRK1A protein kinase: neurobiological functions and Down syndrome implications. **Journal of neural transmission. Supplementum**, n. 67, p. 129–137, 2003.

- HANSCH, C.; FUJITA, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. **Journal of the American Chemical Society**, v. 86, n. 8, p. 1616–1626, 1964.
- HEBERT, L. E.; WEUVE, J.; SCHERR, P. A.; EVANS, D. A. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. **Neurology**, v. 80, n. 19, p. 1778–1783, 2013.
- HENDERSON, V. W.; PAGANINI-HILL, A.; MILLER, B. L.; ELBLE, R. J.; REYES, P. F.; SHOUBE, D.; MCCLEARY, C. A.; KLEIN, R. A.; HAKE, A. M.; FARLOW, M. R. Estrogen for Alzheimer's disease in women: randomized, double-blind, placebo-controlled trial. **Neurology**, v. 54, n. 2, p. 295–301, 2000.
- HERRERA, E.; CARAMELLI, P.; SILVEIRA, A. S. B.; NITRINI, R. Epidemiologic survey of dementia in a community-dwelling Brazilian population. **Alzheimer disease and associated disorders**, v. 16, n. 2, p. 103–108, 2001.
- HEYMAN, A.; WILKINSON, W. E.; STAFFORD, J. A.; HELMS, M. J.; SIGMON, A. H.; WEINBERG, T. Alzheimer's disease: a study of epidemiological aspects. **Annals of neurology**, v. 15, n. 4, p. 335–341, 1984.
- HIMPEL, S.; PANZER, P.; EIRMBTER, K.; CZAJKOWSKA, H.; SAYED, M.; PACKMAN, L. C.; BLUNDELL, T.; KENTRUP, H.; GRÖTZINGER, J.; JOOST, H. G.; BECKER, W. Identification of the autophosphorylation sites and characterization of their effects in the protein kinase DYRK1A. **The Biochemical journal**, v. 359, n. Pt 3, p. 497–505, 2001.
- HOLTZMAN, D. M.; MORRIS, J. C.; GOATE, A. M. Alzheimer's disease: the challenge of the second century. **Science translational medicine**, v. 3, n. 77, p. 77sr1, 2011.
- HUANG, S.-Y.; GRINTER, S. Z.; ZOU, X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. **Physical chemistry chemical physics : PCCP**, v. 12, n. 40, p. 12899–12908, 2010.
- HURST, J. R.; HERITAGE, T. W. **Molecular hologram QSARUSA**, 2001. Disponível em: <<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=/netahtml/PTO/search-adv.htm&r=497&f=G&l=50&d=PTXT&s1=hurst.INNM.&p=10&OS=IN/hurst&RS=IN/hurst>>. Acesso em: 1 jan. 2015
- IVANCUIC, O. QSPR/QSAR Studies by Molecular Descriptors. In: DIUDEA, M. V. **3D QSAR Models**. New Science Publishers, 2000. p. 251–253.
- IVANOV, A. S.; VESELOVSKY, A. V.; DUBANOV, A. V.; SKVORTSOV, V. S. Bioinformatics platform development: from gene to lead compound. **Methods in molecular biology (Clifton, N.J.)**, v. 316, p. 389–431, 2006.

JAIN, A. N.; NICHOLLS, A. Recommendations for evaluation of computational methods. **Journal of computer-aided molecular design**, v. 22, n. 3-4, p. 133–139, 2008.

JUNG, M.-S.; PARK, J.-H.; RYU, Y. S.; CHOI, S.-H.; YOON, S.-H.; KWEN, M.-Y.; OH, J. Y.; SONG, W.-J.; CHUNG, S.-H. Regulation of RCAN1 protein activity by Dyrk1A protein-mediated phosphorylation. **The Journal of biological chemistry**, v. 286, n. 46, p. 40401–40412, 2011.

KAYED, R.; HEAD, E.; THOMPSON, J. L.; MCINTIRE, T. M.; MILTON, S. C.; COTMAN, C. W.; GLABE, C. G. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. **Science (New York, N.Y.)**, v. 300, n. 5618, p. 486–489, 2003.

KENTRUP, H.; BECKER, W.; HEUKELBACH, J.; WILMES, A.; SCHÜRMAN, A.; HUPPERTZ, C.; KAINULAINEN, H.; JOOST, H. G. Dyrk, a dual specificity protein kinase with unique structural features whose activity is dependent on tyrosine residues between subdomains VII and VIII. **The Journal of biological chemistry**, v. 271, n. 7, p. 3488–3495, 1996.

KIM, H.; SABLIN, S. O.; RAMSAY, R. R. Inhibition of monoamine oxidase A by beta-carboline derivatives. **Archives of biochemistry and biophysics**, v. 337, n. 1, p. 137–142, 1997.

KIM, N. D.; YOON, J.; KIM, J. H.; LEE, J. T.; CHON, Y. S.; HWANG, M.-K.; HA, I.; SONG, W.-J. Putative therapeutic agents for the learning and memory deficits of people with Down syndrome. **Bioorganic & medicinal chemistry letters**, v. 16, n. 14, p. 3772–3776, 2006.

KLEBE, G.; ABRAHAM, U.; MIETZNER, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. **Journal of Medicinal Chemistry**, v. 37, n. 24, p. 4130–4146, 1994.

KLEBE, G.; MIETZNER, T.; WEBER, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. **Journal of computer-aided molecular design**, v. 13, n. 1, p. 35–49, 1999.

KNIGHT, Z. A.; SHOKAT, K. M. Chemical genetics: where genetics and pharmacology meet. **Cell**, v. 128, n. 3, p. 425–430, 2007.

KOLB, P.; IRWIN, J. J. Docking screens: right for the right reasons? **Current topics in medicinal chemistry**, v. 9, n. 9, p. 755–770, 2009.

KOO, K. A.; KIM, N. D.; CHON, Y. S.; JUNG, M.-S.; LEE, B.-J.; KIM, J. H.; SONG, W.-J. QSAR analysis of pyrazolidine-3,5-diones derivatives as Dyrk1A inhibitors. **Bioorganic & medicinal chemistry letters**, v. 19, n. 8, p. 2324–2328, 2009.

KORB, O.; STÜTZLE, T.; EXNER, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. **Journal of chemical information and modeling**, v. 49, n. 1, p. 84–96, 2009.

KUBINYI, H. Comparative Molecular Field Analysis (CoMFA). In: GASTEIGER, J. (Ed.). **Handbook of Chemoinformatics: From Data to Knowledge in 4 Volumes**. 1st. ed. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2003. p. 1–12.

LAMBERT, J. D.; SANG, S.; YANG, C. S. Biotransformation of green tea polyphenols and the biological activities of those metabolites. **Molecular pharmaceuticals**, v. 4, n. 6, p. 819–825, 2007.

LEACH, A. R. **Molecular Modelling: Principles and Applications**. 5th. ed. New York: Longman, 2011. p. 595

LEDER, S.; WEBER, Y.; ALTAF AJ, X.; ESTIVILL, X.; JOOST, H. G.; BECKER, W. Cloning and characterization of DYRK1B, a novel member of the DYRK family of protein kinases. **Biochemical and biophysical research communications**, v. 254, n. 2, p. 474–479, 1999.

LEE, C.-H.; HUANG, H.-C.; JUAN, H.-F. Reviewing ligand-based rational drug design: the search for an ATP synthase inhibitor. **International journal of molecular sciences**, v. 12, n. 8, p. 5304–5318, 2011.

LEE, M.-S.; KAO, S.-C.; LEMERE, C. A.; XIA, W.; TSENG, H.-C.; ZHOU, Y.; NEVE, R.; AHLIJANIAN, M. K.; TSAI, L.-H. APP processing is regulated by cytoplasmic phosphorylation. **The Journal of cell biology**, v. 163, n. 1, p. 83–95, 2003.

LEE, Y.; HA, J.; KIM, H. J.; KIM, Y.-S.; CHANG, E.-J.; SONG, W.-J.; KIM, H.-H. Negative feedback Inhibition of NFATc1 by DYRK1A regulates bone homeostasis. **The Journal of biological chemistry**, v. 284, n. 48, p. 33343–33351, 2009.

LEVY-LAHAD, E.; WIJSMAN, E. M.; NEMENS, E.; ANDERSON, L.; GODDARD, K. A.; WEBER, J. L.; BIRD, T. D.; SCHELLENBERG, G. D. A familial Alzheimer's disease locus on chromosome 1. **Science (New York, N.Y.)**, v. 269, n. 5226, p. 970–973, 1995.

LIM, G. P.; YANG, F.; CHU, T.; GAHTAN, E.; UBEDA, O.; BEECH, W.; OVERMIER, J. B.; HSIAO-ASHEC, K.; FRAUTSCHY, S. A.; COLE, G. M. Ibuprofen effects on Alzheimer pathology and open field activity in APP^{sw} transgenic mice. **Neurobiology of aging**, v. 22, n. 6, p. 983–991, 2001.

LIU, F.; LIANG, Z.; WEGIEL, J.; HWANG, Y.-W.; IQBAL, K.; GRUNDKE-IQBAL, I.; RAMAKRISHNA, N.; GONG, C.-X. Overexpression of Dyrk1A contributes to neurofibrillary degeneration in Down syndrome. **FASEB journal : official publication of the Federation of American Societies for Experimental Biology**, v. 22, n. 9, p. 3224–3233, 2008.

LIVINGSTONE, D. **Data analysis for chemists : applications to QSAR and chemical product design**. 1st. ed. New York: Oxford University Press, 1995. p. 239

LOCHHEAD, P. A.; SIBBET, G.; MORRICE, N.; CLEGHON, V. Activation-loop autophosphorylation is mediated by a novel transitional intermediate form of DYRKs. **Cell**, v. 121, n. 6, p. 925–936, 2005.

LOWIS, D. R. **HQSAR: A New, Highly Predictive QSAR Technique, Tripos Technique Notes**Tripos. St. Louis, MO, USA: [s.n.]. Disponível em: <http://tripos.com/data/SYBYL/HQSAR_Application_Note_072605.pdf>. Acesso em: 1 jan. 2015.

LUCIN, K. M.; WYSS-CORAY, T. Immune activation in brain aging and neurodegeneration: too much or too little? **Neuron**, v. 64, n. 1, p. 110–122, 2009.

MAENZ, B.; HEKERMAN, P.; VELA, E. M.; GALCERAN, J.; BECKER, W. Characterization of the human DYRK1A promoter and its regulation by the transcription factor E2F1. **BMC molecular biology**, v. 9, n. 1, p. 30, 2008.

MARTÍ, E.; ALTAF AJ, X.; DIERSSEN, M.; DE LA LUNA, S.; FOTAKI, V.; ALVAREZ, M.; PÉREZ-RIBA, M.; FERRER, I.; ESTIVILL, X. Dyrk1A expression pattern supports specific roles of this kinase in the adult central nervous system. **Brain research**, v. 964, n. 2, p. 250–263, 2003.

MAURER, K.; VOLK, S.; GERBALDO, H. Auguste D and Alzheimer's disease. **Lancet**, v. 349, n. 9064, p. 1546–1549, 1997.

MCKHANN, G.; DRACHMAN, D.; FOLSTEIN, M.; KATZMAN, R.; PRICE, D.; STADLAN, E. M. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. **Neurology**, v. 34, n. 7, p. 939–944, 1984.

MELVILLE, J. L.; HIRST, J. D. On the stability of CoMFA models. **Journal of chemical information and computer sciences**, v. 44, n. 4, p. 1294–1300, 2004.

MØLLER, R. S.; KÜBART, S.; HOELTZENBEIN, M.; HEYE, B.; VOGEL, I.; HANSEN, C. P.; MENZEL, C.; ULLMANN, R.; TOMMERUP, N.; ROPERS, H.-H.; TÜMER, Z.; KALSCHUEER, V. M. Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. **American journal of human genetics**, v. 82, n. 5, p. 1165–1170, 2008.

MOON, J. B.; HOWE, W. J. Computer design of bioactive molecules: a method for receptor-based de novo ligand design. **Proteins**, v. 11, n. 4, p. 314–328, 1991.

MORRIS, G. M.; GOODSSELL, D. S.; HALLIDAY, R. S.; HUEY, R.; HART, W. E.; BELEW, R. K.; OLSON, A. J. Automated docking using a Lamarckian genetic algorithm

and an empirical binding free energy function. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1639–1662, 1998.

MOTT, B. T.; TANEGA, C.; SHEN, M.; MALONEY, D. J.; SHINN, P.; LEISTER, W.; MARUGAN, J. J.; INGLESE, J.; AUSTIN, C. P.; MISTELI, T.; AULD, D. S.; THOMAS, C. J. Evaluation of substituted 6-arylquinazolin-4-amines as potent and selective inhibitors of cdc2-like kinases (Clk). **Bioorganic & medicinal chemistry letters**, v. 19, n. 23, p. 6700–6705, 2009.

MYINT, K. Z.; XIE, X.-Q. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. **International journal of molecular sciences**, v. 11, n. 10, p. 3846–3866, 2010.

NEAGOIE, C.; VEDRENNE, E.; BURON, F.; MÉROUR, J.-Y.; ROSCA, S.; BOURG, S.; LOZACH, O.; MEIJER, L.; BALDEYROU, B.; LANSIAUX, A.; ROUTIER, S. Synthesis of chromeno[3,4-b]indoles as Lamellarin D analogues: a novel DYRK1A inhibitor class. **European journal of medicinal chemistry**, v. 49, p. 379–396, 2012.

NITRINI, R.; CARAMELLI, P.; BOTTINO, C. M. DE C.; DAMASCENO, B. P.; BRUCKI, S. M. D.; ANGHINAH, R. Diagnóstico de doença de Alzheimer no Brasil: avaliação cognitiva e funcional. Recomendações do Departamento Científico de Neurologia Cognitiva e do Envelhecimento da Academia Brasileira de Neurologia. **Arquivos de Neuro-Psiquiatria**, v. 63, n. 3a, p. 720–727, 2005.

NOLEN, B.; TAYLOR, S.; GHOSH, G. Regulation of protein kinases; controlling activity through activation segment conformation. **Molecular cell**, v. 15, n. 5, p. 661–675, 2004.

OGAWA, Y.; NONAKA, Y.; GOTO, T.; OHNISHI, E.; HIRAMATSU, T.; KII, I.; YOSHIDA, M.; IKURA, T.; ONOGI, H.; SHIBUYA, H.; HOSOYA, T.; ITO, N.; HAGIWARA, M. Development of a novel selective inhibitor of the Down syndrome-related kinase Dyrk1A. **Nature communications**, v. 1, p. 86, 2010.

PALANGSUNTIKUL, R.; BERNER, H.; BERGER, M. L.; WOLSCHANN, P. Holographic quantitative structure-activity relationships of tryptamine derivatives at NMDA, 5HT(1A) and 5HT(2A) receptors. **Molecules (Basel, Switzerland)**, v. 18, n. 8, p. 8799–8811, 2013.

PAN, Y.; WANG, Y.; BRYANT, S. H. Pharmacophore and 3D-QSAR characterization of 6-arylquinazolin-4-amines as Cdc2-like kinase 4 (Clk4) and dual specificity tyrosine-phosphorylation-regulated kinase 1A (Dyrk1A) inhibitors. **Journal of chemical information and modeling**, v. 53, n. 4, p. 938–947, 2013.

PARK, J.; SONG, W.-J.; CHUNG, K. C. Function and regulation of Dyrk1A: towards understanding Down syndrome. **Cellular and molecular life sciences : CMLS**, v. 66, n. 20, p. 3235–3240, 2009.

PATEL, B. D.; GHATE, M. D. 3D-QSAR studies of dipeptidyl peptidase-4 inhibitors using various alignment methods. **Medicinal Chemistry Research**, v. 24, n. 3, p. 1060–1069, 2014.

PERKINS, R.; FANG, H.; TONG, W.; WELSH, W. J. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. **Environmental toxicology and chemistry / SETAC**, v. 22, n. 8, p. 1666–1679, 2003.

PETERSON, S. D.; SCHAAL, W.; KARLÉN, A. Improved CoMFA modeling by optimization of settings. **Journal of chemical information and modeling**, v. 46, n. 1, p. 355–364, 2006.

POULIN, P.; ZAKZANIS, K. K. In vivo neuroanatomy of Alzheimer's disease: evidence from structural and functional brain imaging. **Brain and cognition**, v. 49, n. 2, p. 220–225, 2002.

RITCHIE, K.; LOVESTONE, S. The dementias. **Lancet**, v. 360, n. 9347, p. 1759–1766, 2002.

RODRIGUES, C. R. **Modelagem molecular de agentes antiprotozoários utilizando cálculos semi-empíricos AM1, mecânica molecular e métodos de QSAR (CoMFA/HQSAR)**. 1999. 169 f. Tese (Doutorado em Química Orgânica) - Instituto de Química. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1999.

ROSENTHAL, A. S.; TANEGA, C.; SHEN, M.; MOTT, B. T.; BOUGIE, J. M.; NGUYEN, D.-T.; MISTELI, T.; AULD, D. S.; MALONEY, D. J.; THOMAS, C. J. **An inhibitor of the Cdc2-like kinase 4 (Clk4)**. In: Probe Reports from the NIH Molecular Libraries Program, 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/books/NBK56236/>>. Acesso em: 4 jan. 2015.

ROSENTHAL, A. S.; TANEGA, C.; SHEN, M.; MOTT, B. T.; BOUGIE, J. M.; NGUYEN, D.-T.; MISTELI, T.; AULD, D. S.; MALONEY, D. J.; THOMAS, C. J. Potent and selective small molecule inhibitors of specific isoforms of Cdc2-like kinases (Clk) and dual specificity tyrosine-phosphorylation-regulated kinases (Dyrk). **Bioorganic & medicinal chemistry letters**, v. 21, n. 10, p. 3152–3158, 2011.

RÜCKER, C.; RÜCKER, G.; MERINGER, M. γ -Randomization and its variants in QSPR/QSAR. **Journal of chemical information and modeling**, v. 47, n. 6, p. 2345–2357, 2007.

RYOO, S.-R.; CHO, H.-J.; LEE, H.-W.; JEONG, H. K.; RADNAABAZAR, C.; KIM, Y.-S.; KIM, M.-J.; SON, M.-Y.; SEO, H.; CHUNG, S.-H.; SONG, W.-J. Dual-specificity tyrosine(Y)-phosphorylation regulated kinase 1A-mediated phosphorylation of amyloid precursor protein: evidence for a functional link between Down syndrome and Alzheimer's disease. **Journal of neurochemistry**, v. 104, n. 5, p. 1333–1344, 2008.

SAKONO, M.; ZAKO, T. Amyloid oligomers: formation and toxicity of Abeta oligomers. **The FEBS journal**, v. 277, n. 6, p. 1348–1358, 2010.

SALUM, L. B.; ANDRICOPULO, A. D. Fragment-based QSAR: perspectives in drug design. **Molecular diversity**, v. 13, n. 3, p. 277–285, 2009.

SÁNCHEZ, C.; SALAS, A. P.; BRAÑA, A. F.; PALOMINO, M.; PINEDA-LUCENA, A.; CARBAJO, R. J.; MÉNDEZ, C.; MORIS, F.; SALAS, J. A. Generation of potent and selective kinase inhibitors by combinatorial biosynthesis of glycosylated indolocarbazoles. **Chemical communications (Cambridge, England)**, n. 27, p. 4118–4120, 2009.

SANO, M.; ERNESTO, C.; THOMAS, R. G.; KLAUBER, M. R.; SCHAFER, K.; GRUNDMAN, M.; WOODBURY, P.; GROWDON, J.; COTMAN, C. W.; PFEIFFER, E.; SCHNEIDER, L. S.; THAL, L. J. A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. The Alzheimer's Disease Cooperative Study. **The New England journal of medicine**, v. 336, n. 17, p. 1216–1222, 1997.

SARNO, S.; MAZZORANA, M.; TRAYNOR, R.; RUZZENE, M.; COZZA, G.; PAGANO, M. A.; MEGGIO, F.; ZAGOTTO, G.; BATTISTUTTA, R.; PINNA, L. A. Structural features underlying the selectivity of the kinase inhibitors NBC and dNBC: role of a nitro group that discriminates between CK2 and DYRK1A. **Cellular and molecular life sciences : CMLS**, v. 69, n. 3, p. 449–460, 2012.

SAVAGE, M. J.; GINGRICH, D. E. Advances in the development of kinase inhibitor therapeutics for Alzheimer's disease. **Drug Development Research**, v. 70, n. 2, p. 125–144, 2009.

SHI, J.; ZHANG, T.; ZHOU, C.; CHOHAN, M. O.; GU, X.; WEGIEL, J.; ZHOU, J.; HWANG, Y.-W.; IQBAL, K.; GRUNDKE-IQBAL, I.; GONG, C.-X.; LIU, F. Increased dosage of Dyrk1A alters alternative splicing factor (ASF)-regulated alternative splicing of tau in Down syndrome. **The Journal of biological chemistry**, v. 283, n. 42, p. 28660–28669, 2008.

SLENO, L.; EMILI, A. Proteomic methods for drug target discovery. **Current opinion in chemical biology**, v. 12, n. 1, p. 46–54, 2008.

SMALL, S. A.; DUFF, K. Linking Abeta and tau in late-onset Alzheimer's disease: a dual pathway hypothesis. **Neuron**, v. 60, n. 4, p. 534–542, 2008.

SMELLIE, A. S.; CRIPPEN, G. M.; RICHARDS, W. G. Fast drug-receptor mapping by site-directed distances: a novel method of predicting new pharmacological leads. **Journal of chemical information and computer sciences**, v. 31, n. 3, p. 386–392, 1991.

SMITH, B.; MEDDA, F.; GOKHALE, V.; DUNCKLEY, T.; HULME, C. Recent advances in the design, synthesis, and biological evaluation of selective DYRK1A inhibitors: a new avenue for a disease modifying treatment of Alzheimer's? **ACS chemical neuroscience**, v. 3, n. 11, p. 857–872, 2012.

SNEADER, W. **Drug Discovery: A History**. Chichester, UK: John Wiley & Sons, Ltd, 15 Apr. 2005. p. 1–468

SOBHIA, M. E.; BHARATAM, P. V. Comparative molecular similarity indices analysis (CoMSIA) studies of 1,2-naphthoquinone derivatives as PTP1B inhibitors. **Bioorganic & medicinal chemistry**, v. 13, n. 6, p. 2331–2338, 2005.

SOUNDARARAJAN, M.; ROOS, A. K.; SAVITSKY, P.; FILIPPAKOPOULOS, P.; KETTENBACH, A. N.; OLSEN, J. V.; GERBER, S. A.; ESWARAN, J.; KNAPP, S.; ELKINS, J. M. Structures of Down syndrome kinases, DYRKs, reveal mechanisms of kinase activation and substrate recognition. **Structure (London, England : 1993)**, v. 21, n. 6, p. 986–996, 2013.

STORN, R.; PRICE, K. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. **Journal of Global Optimization**, v. 11, n. 4, p. 341–359, 1997.

TABET, N.; FELDMAND, H. Ibuprofen for Alzheimer's disease. **The Cochrane database of systematic reviews**, n. 2, p. CD004031, 2003.

TAHTOUH, T.; ELKINS, J. M.; FILIPPAKOPOULOS, P.; SOUNDARARAJAN, M.; BURG, G.; DURIEU, E.; COCHET, C.; SCHMID, R. S.; LO, D. C.; DELHOMMEL, F.; OBERHOLZER, A. E.; PEARL, L. H.; CARREAUX, F.; BAZUREAU, J.-P.; KNAPP, S.; MEIJER, L. Selectivity, cocrystal structures, and neuroprotective properties of leucettines, a family of protein kinase inhibitors derived from the marine sponge alkaloid leucettamine B. **Journal of medicinal chemistry**, v. 55, n. 21, p. 9312–9330, 2012.

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: a new technique for high-accuracy molecular docking. **Journal of medicinal chemistry**, v. 49, n. 11, p. 3315–3321, 2006.

TINTORI, C.; MANETTI, F.; BOTTA, M. Pharmacophoric models and 3D QSAR studies of the adenosine receptor ligands. **Current topics in medicinal chemistry**, v. 10, n. 10, p. 1019–1035, 2010.

TROSSINI, G. H. G.; MALTAROLLO, V. G.; SCHMIDT, T. J. Hologram QSAR studies of antiprotozoal activities of sesquiterpene lactones. **Molecules (Basel, Switzerland)**, v. 19, n. 7, p. 10546–10562, 2014.

VAN DUIJN, C. M. Epidemiology of the dementias: recent developments and new approaches. **Journal of neurology, neurosurgery, and psychiatry**, v. 60, n. 5, p. 478–488, 1996.

VEDANI, A.; DOBLER, M.; LILL, M. A. The challenge of predicting drug toxicity in silico. **Basic & clinical pharmacology & toxicology**, v. 99, n. 3, p. 195–208, 2006.

VERDONK, M. L.; BERDINI, V.; HARTSHORN, M. J.; MOOIJ, W. T. M.; MURRAY, C. W.; TAYLOR, R. D.; WATSON, P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. **Journal of chemical information and computer sciences**, v. 44, n. 3, p. 793–806, 2003.

VINGTDEUX, V.; HAMDANE, M.; GOMPEL, M.; BÉCARD, S.; DROBECQ, H.; GHESTEM, A.; GROSJEAN, M.-E.; KOSTANJEVECKI, V.; GROGNET, P.; VANMECHELEN, E.; BUÉE, L.; DELACOURTE, A.; SERGEANT, N. Phosphorylation of amyloid precursor carboxy-terminal fragments enhances their processing by a gamma-secretase-dependent mechanism. **Neurobiology of disease**, v. 20, n. 2, p. 625–637, 2005.

VISWANADHAN, V. N.; GHOSE, A. K.; REVANKAR, G. R.; ROBINS, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain . **Journal of Chemical Information and Modeling**, v. 29, n. 3, p. 163–172, 1989.

VYAS, V. K.; BHATT, H. G.; PATEL, P. K.; JALU, J.; CHINTHA, C.; GUPTA, N.; GHATE, M. CoMFA and CoMSIA studies on C-aryl glucoside SGLT2 inhibitors as potential anti-diabetic agents. **SAR and QSAR in environmental research**, v. 24, n. 7, p. 519–551, 2013.

WALLQVIST, A.; COVELL, D. G. Docking enzyme-inhibitor complexes using a preference-based free-energy surface. **Proteins**, v. 25, n. 4, p. 403–419, 1996.

WANG, J.; KANG, X.; KUNTZ, I. D.; KOLLMAN, P. A. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. **Journal of medicinal chemistry**, v. 48, n. 7, p. 2432–2444, 2005.

WEGIEL, J.; GONG, C.-X.; HWANG, Y.-W. The role of DYRK1A in neurodegenerative diseases. **The FEBS journal**, v. 278, n. 2, p. 236–245, 2011.

WEGIEL, J.; KUCHNA, I.; NOWICKI, K.; FRACKOWIAK, J.; DOWJAT, K.; SILVERMAN, W. P.; REISBERG, B.; DELEON, M.; WISNIEWSKI, T.; ADAYEV, T.; CHEN-HWANG, M.-C.; HWANG, Y.-W. Cell type- and brain structure-specific patterns of distribution of minibrain kinase in human brain. **Brain research**, v. 1010, n. 1-2, p. 69–80, 2004.

WIECHMANN, S.; CZAJKOWSKA, H.; DE GRAAF, K.; GRÖTZINGER, J.; JOOST, H.-G.; BECKER, W. Unusual function of the activation loop in the protein kinase DYRK1A. **Biochemical and biophysical research communications**, v. 302, n. 2, p. 403–408, 2003.

WILCKEN, R.; ZIMMERMANN, M. O.; LANGE, A.; JOERGER, A. C.; BOECKLER, F. M. Principles and applications of halogen bonding in medicinal chemistry and chemical biology. **Journal of medicinal chemistry**, v. 56, n. 4, p. 1363–1388, 2013.

WOODS, Y.; COHEN, P.; BECKER, W.; JAKES, R.; GOEDERT, M.; WANG, X.; PROUD, C. The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2B ϵ at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. **The Biochemical Journal**, v. 355, n. Pt3, p. 609–615, 2001.

ZANDI, P. P.; SPARKS, D. L.; KHACHATURIAN, A. S.; TSCHANZ, J.; NORTON, M.; STEINBERG, M.; WELSH-BOHMER, K. A.; BREITNER, J. C. S. Do statins reduce risk of incident dementia and Alzheimer disease? The Cache County Study. **Archives of general psychiatry**, v. 62, n. 2, p. 217–224, 2005.

ZAUHAR, R. J.; COLBERT, C. L.; MORGAN, R. S.; WELSH, W. J. Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. **Biopolymers**, v. 53, n. 3, p. 233–248, 2000.

ANEXO