Estevão Bombonato Pereira

Estudo da relação quantitativa estrutura-atividade de compostos β -carbolínicos, substituídos nas posições 1 e 3, utilizados no tratamento de câncer de pulmão e melanoma cutâneo

Dissertação apresentada ao Instituto de Química de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências (Físico-Química)

Área de Concentração: Físico-Química

Orientador: Prof^o Dr. Albérico Borges Ferreira da Silva

São Carlos

2014

“Na verdade, há um espírito no homem, e o sopro do Todo-Poderoso o faz sábio.” (Jó 32:8)

Agradecimentos

Primeiramente, a Deus por mais essa realização muito especial.

Aos meus pais, Antonio Carlos e Elaine, e minhas irmãs, Lídia e Júlia, que sempre me apoio e me encorajou a prosseguir nos estudos.

À minha namorada Ruhamah, que mesmo estando longe sempre me dá apoio e se faz presente.

Ao meu orientador Prof^º Dr. Albérico Borges Ferreira da Silva, pela oportunidade de fazer o mestrado sob a supervisão e orientação dele.

Aos meus caros colegas do laboratório, Aline Oliveira, Amanda Guimarães, Aldinéia Pereira, Emmanuela de Lima, Gláucio Nagurniak, Régis Santiago, Luiz Alberto Terrabuio, Tiago Teodoro, Ronaldo Fernandes, Rafael Vichiatti, Ana Cristina Tello e Milena Maringolo. A Ajuda de todos foi fundamental para a conclusão desse mestrado.

À Prof^a Dra. Káthia Maria Honório, por vir de São Paulo para ajudar não só a mim, mas também a outros colegas a desenvolverem melhor os trabalhos de mestrado ou doutorado.

Resumo

O câncer é uma doença que aflige pessoas no mundo todo, e mata muitas todos os anos. Muitos estudos são desenvolvidos para sintetizar novos compostos com atividade antitumoral cada vez melhor. O maior número de compostos possibilita um maior número de possibilidades de cura, já que alguns tumores podem adquirir resistência ao medicamento, sendo necessária a troca por outro.

Para que não seja feita uma síntese aleatória de compostos, desperdiçando tempo, reagentes e dinheiro sintetizando compostos inativos, um estudo teórico prévio para orientar a síntese é válido. Um estudo QSAR pode orientar a síntese indicando quais compostos são mais interessantes de serem sintetizados e quais são menos interessantes. Assim, há o desenvolvimento de mais compostos ativos em um menor tempo, do que se fosse realizada sínteses e testes para todos os compostos possíveis.

O presente estudo tem esse foco, avaliar a relação quantitativa entre a estrutura molecular e a atividade antitumoral de compostos β -carbólicos substituídos nas posições 1 e 3. Nesse estudo será possível elucidar quais características são relevantes para que os compostos sejam ativos no tratamento de câncer de pulmão e melanoma.

Abstract

Cancer is a disease that afflicts people worldwide and kills many every year. Many studies are developed to synthesize new compounds with better antitumor activity. The largest number of compounds allows a greater number of curability, because some tumors may acquire used drug resistance and are necessary different drugs.

To avoid a random synthesis of compounds, wasting time, money and reagents in synthesizing inactive compounds, a previous theoretical study as guide of synthesis is valid. A QSAR study can guide the synthesis indicating which compounds are most interesting to be synthesized and which are less interesting. Thus, more compounds active are developed in a shorter time.

The present study has this focus, assess the quantitative relationship between molecular structure and antitumor activity of β -carboline compounds substituted in positions 1 and 3. In this study will be possible to elucidate characteristics that are relevant to the compounds under study be active in the treatment of lung cancer and melanoma.

Lista de Figuras

Figura 1 – Molécula do tetrahidro- β -carbolina	11
Figura 2 – Estrutura principal dos compostos β -carbolínicos em estudo.	15
Figura 3. Esquema do método de regressão por PLS.	20
Figura 4. Esquema do procedimento autoconsistente de Kohn-Shan.	28
Figura 5 – Gráfico dos dados experimentais calculados pelo método <i>PLS</i> gerado com o conjunto de treinamento.	33
Figura 6 – Gráfico de Resíduos de <i>Student</i> vs Alavancagem para detecção de amostras atípicas.	34
Figura 7 – Gráfico dos valores preditos para o conjunto teste com os valores experimentais da atividade antitumoral contra câncer de pulmão.	35
Figura 8 – Valores de R^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral contra câncer de pulmão.	36
Figura 9 – Gráfico dos valores de Q^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral para câncer de pulmão.	37
Figura 10 – Gráfico dos valores de R^2 e Q^2 obtidos nas randomizações dos valores de atividade antitumoral para câncer de pulmão.	38
Figura 11 – Gráfico da variação de Q^2 com a retirada de compostos para geração do modelo.	39
Figura 12 – Gráfico das atividades experimentais calculadas com modelo <i>PLS</i> , gerado para o conjunto de treinamento contra melanoma.	43
Figura 13 – Gráfico de alavancagem para identificação de amostra atípica.	44
Figura 14 – Gráfico dos valores preditos com os valores experimentais da atividade antitumoral contra melanoma do conjunto teste.	45
Figura 15 – Gráfico dos valores de Q^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral para melanoma.	46
Figura 16 – Gráfico dos valores de R^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral para melanoma.	47
Figura 17 – Gráfica Q^2 e R^2 dos modelos aleatorizados de atividade antitumoral para melanoma.	48
Figura 18 – Gráfico de Q^2 para N amostras retiradas na validação cruzada.	49

Lista de Tabelas

Tabela 1 – Substituintes da estrutura principal dos compostos b-carbolínicos	15
Tabela 2 – Atividade antitumoral dos compostos contra câncer de pulmão e melanoma.	17
Tabela 3 – Valores de atividade antitumoral contra câncer de pulmão, obtidos experimentalmente e calculados pelo modelo.....	31
Tabela 4 – Resultados do método <i>PLS</i> gerados com conjunto de treinamento.	32
Tabela 5 – Valores experimentais e preditos pelo modelo, para o conjunto teste, de atividade antitumoral contra câncer de pulmão.	34
Tabela 6 – Valores de atividade antitumoral contra melanoma, obtidos experimentalmente e calculados pelo modelo.	41
Tabela 7 – Resultados extraídos do modelo gerado com o método <i>PLS</i>	42
Tabela 8 – Resultados da predição dos compostos do conjunto de teste.....	44

Sumário

1. Introdução	9
1.1. Câncer – Pulmão e Melanoma	9
1.2. Novos composto para o combate ao câncer.....	10
1.3. Estudos QSAR e métodos quimiométricos.....	13
2. Desenvolvimento	15
2.1. Objetivos	15
2.2. Materiais e Métodos	19
2.2.1. Método Quimiométrico – Quadrados Mínimos Parciais (<i>PLS – Partial Least Squares</i>).....	20
2.2.1.1. Métodos de Validação.....	21
2.2.1.2. Detecção de Amostras Atípicas.....	22
2.2.1.3. Teste de Robustez	23
2.2.1.4. Teste de Correlação ao Acaso	23
2.2.2. Método Mecânico Quântico – Teoria do Funcional de Densidade (DFT)	23
3. Resultados e Discussão	29
3.1. Estudo QSAR – Câncer de Pulmão.....	29
3.2. Estudo QSAR - Melanoma Cutâneo.....	39
4. Conclusão	50
5. Bibliografia	51

1. Introdução

1.1. Câncer – Pulmão e Melanoma

Uma das principais doenças que aflige a humanidade é o câncer. Seus sintomas podem ser confundidos com os de outras doenças e também pode ser assintomático, o que dificulta o diagnóstico da doença. Há alguns tipos de câncer que podem ser identificados mais facilmente, como o câncer de mama, que pode ser diagnosticado através do autoexame e também por mamografia, e o melanoma cutâneo, que inicia como pequenas manchas na pele, assim a pessoa deve estar atenta às manchas que surgirem.

Pela falta de atenção e informação e por achar que pequenas complicações de saúde não são nada, os tumores são detectados tardiamente e em estágio avançado. Há tratamento para os tumores e há índices de cura elevados para alguns tratamentos de tumores em estágio inicial, mas para tumores em estágio avançado os índices de cura não são tão bons. Por isso o diagnóstico de um tumor em estágio avançado é recebido pelos pacientes, muitas vezes, como uma sentença de sofrimento e morte.

As medidas mais comuns existentes para o combate ao câncer são a radioterapia, quimioterapia e intervenção cirúrgica de retirada do tumor e de alguns tecidos adjacentes a este. Estes quatro métodos podem ser utilizados isoladamente ou em conjunto. De acordo com o quadro do paciente e o avanço da doença os médicos escolhem qual é o método mais viável para o tratamento (Barros, Cecatti et al. 2012). Para alguns tipos de tumor, infelizmente, não há medicamentos eficazes a ponto de erradicá-los, mas que conseguem apenas amenizar as dores e outros efeitos causados pelo tumor à saúde do paciente e tentam prolongar o máximo possível sua vida.

Dentre os vários tipos de câncer, destaca-se o câncer de pulmão devido ao crescente aumento do número de casos. Este aumento é, muito provavelmente, um reflexo do índice de tabagismos no país (Silva, Noronha et al. 2011). Divulgado pela GLOBOCAN 2008, cerca de 7,8 milhões de óbitos no mundo foram causados pelo câncer, destacando-se o câncer de pulmão como o maior causador de mortes, 1,3 milhões (Saúde 2010). Cerca de 80% das causas do

câncer de pulmão são devidas ao tabagismo. Comparando o risco de um tabagista com o de um não fumante, o tabagista tem de 20 a 30 vezes mais chances de desenvolver câncer de pulmão (Silva, Noronha et al. 2011).

Além do tabagismo, fatores ambientais também interferem no organismo das pessoas propiciando o desenvolvimento do câncer de pulmão. Tais fatores podem ser a presença de amianto, arsênio, radônio e hidrocarbonetos aromáticos policíclicos, dentre outros compostos presentes na poluição atmosférica. Estima-se que em países industrializados, cerca de 5% a 10% dos casos de câncer de pulmão são causados pela exposição à poluição (Silva, Noronha et al. 2011).

O melanoma cutâneo também se destaca por ser o mais frequente no Brasil, sendo responsável por 25% dos casos de câncer malignos. O melanoma cutâneo tem origem em células chamadas melanócitos, responsáveis pela pigmentação da pele, e é o principal tipo de câncer de pele. O melanoma, mesmo ocorrendo na parte mais externa do nosso corpo, é extremamente grave devido à alta possibilidade de metástase, mas se for detectado no estágio inicial o prognóstico é bom. O aumento da sobrevivência dos pacientes com esse câncer é devido principalmente ao diagnóstico precoce que possibilita o tratamento mais eficaz e evita a ocorrência de metástase (INCA 2013).

1.2. Novos composto para o combate ao câncer

É importante haver uma pesquisa constante para o desenvolvimento de novos medicamentos, pois os agentes causadores das doenças adquirem resistência aos medicamentos já existentes, ou o medicamento gera algum efeito adverso em algumas pessoas ou efeito nenhum. Assim, gerar novos medicamentos propicia um aumento nas possibilidades de tratamento e maior eficiência para obtenção da cura do paciente, além de ser possível conseguir medicamentos novos que reduzam os efeitos colaterais.

Visando melhorar o tratamento dos vários tipos de câncer e também a eficiência dos fármacos utilizados atualmente, vários estudos de desenvolvimento, síntese e avaliação biológica, tem sido desenvolvidos por grupos de pesquisa (Savariz 2009; Kumar, Gupta et al. 2010; Barbosa, Formagio et al. 2011; Cao, Fan et al. 2013). Um destes grupos é do

Departamento de Química da Universidade Estadual de Maringá, que sintetizou e testou vários derivados de β -carbolinas quanto à atividade antitumoral contra vários tipos de tumores. Alguns trabalhos desenvolvidos nesse grupo de pesquisa desenvolveram novos derivados de β -carbolinas variando o substituinte nas posições 1 e 3 da estrutura padrão (Junior 2009; Tonin, Barbosa et al. 2009; Savariz, Formagio et al. 2010; Valdez, Tonin et al. 2012).

As β -carbolinas são compostos da família dos alcaloides e apresentam atividade citotóxica podendo ser utilizadas para remissão de células tumorais, e também como composto de ação antimicrobiana e antiprotozoaria. As β -carbolinas naturais, como a tetrahydro- β -carbolina (Figura 1), são geradas a partir da reação de condensação entre uma indolamina e um acetaldeído ou formaldeído (Cao, Peng et al. 2007).

Figura 1 – Molécula do tetrahydro- β -carbolina

É possível encontrar β -carbolinas naturais em vegetais, frutas, sucos, tecidos de mamíferos e em alguns alimentos industrializados. Em alimento industrializados as reações naturais para surgimento as β -carbolinas são favorecidas devido a fatores como pH, temperatura, tempo de armazenamento, oxidantes, antioxidantes, leveduras, conservantes, e também por processos de fermentação, defumação, cozimento, aquecimento e amadurecimento (Herraiz 2000; Herraiz and GALISTEO 2003; Papavergou and Herraiz 2003).

A ação antitumoral das β -carbolinas pode dar-se pela intercalação com o DNA das células cancerosas impedindo que ele seja replicado pela topoisomerase I e II. Não se sabe ainda onde se localiza o sitio ativo, se na molécula de DNA ou nas enzimas topoisomerase I ou II, mas há estudos de Doking trabalhando moléculas de β -carbolinas interagindo com a molécula de DNA.

Com a finalidade de desenvolver novos derivados de β -carbolicas eficientes na remissão de células cancerosas, vários estudos vêm sendo desenvolvidos com foco na síntese e teste biológicos destes novos derivados de β -carbolicas (Ishida, Wang et al. 1999; Cao, Peng et al. 2005; Guan, Chen et al. 2006; Junior 2009; Chen, Cao et al. 2010). Buscando uma correlação entre a intercalação com o DNA e a atividade antitumoral, pesquisadores têm realizado estudos para determinar quais interações ocorrem entre as β -carbolicas e o DNA, e quais são os pontos de maior interação (Xiao, Lin et al. 2001; Guan, Chen et al. 2006; Chen, Cao et al. 2010). Para orientar estas sínteses e testes biológicos, estudos teóricos para determinação de parâmetros que influenciam na atividade antitumoral têm sido desenvolvidos juntamente com estudos experimentais, correlacionando a atividade biológica com a estrutura dos compostos, como feito por Weber, Wu, Cao e Rocha (Weber 2008; Wu, Zhao et al. 2009; Cao, Guan et al. 2010; Rocha 2010).

Outros grupos de pesquisa, também desenvolvem estudos de síntese e avaliação antitumoral usando outras classes de moléculas, mantendo uma estrutura padrão e variando apenas os substituintes em determinadas posições. O grupo canadense da *Unité des Biotechnologies et de Bioingénierie*, trabalha com derivados de benzossulfonamidas substituídas, que além da síntese e avaliação antitumoral, também desenvolveram um estudo quantitativo de relação entre a estrutura e a atividade de seus compostos. Este estudo buscou desenvolver compostos com atividade antitumoral mais potente contra células cancerosas e associar quais propriedades estruturais que estão envolvidas na melhora desta atividade (Cao, Peng et al. 2005; Formagio, Tonin et al. 2008; Chen, Cao et al. 2010; Chen, Cao et al. 2010; Kumar, Gupta et al. 2010; Savariz, Formagio et al. 2010; Barbosa, Formagio et al. 2011).

Há grupos de pesquisa que desenvolvem estudos teóricos relacionando a estrutura molecular de uma classe de compostos com a suas respectivas atividades biológicas, as quais foram determinadas experimentalmente. Estes estudos teóricos visam orientar a sínteses de novos compostos, de maneira a economizar reagentes e otimizar o tempo, evitando síntese de derivados não interessantes (Honorio, Richard C. Garratt et al. 2005; Honório, Garratt et al. 2007; Weber, Salum et al. 2010).

1.3. Estudos QSAR e métodos quimiométricos

Estes estudos podem ser divididos em dois tipos, os de relação quantitativa estrutura/atividade (QSAR) e os de relação estrutura/atividade (SAR). Os estudos QSAR são geralmente estudos quantitativos, relacionando as propriedades estruturais e eletrônicas das moléculas ao valor de sua atividade biológica. Já os estudos SAR, são estudos qualitativos que relacionam as características estruturais e eletrônicas com a classificação da molécula, geralmente ativa ou inativa.

As propriedades analisadas para obter-se um conjunto que se relacione com a atividade biológica são chamadas de descritores, e estes podem ser classificados em algumas classes devido à sua natureza (Montanari 2011). Assim, é possível separar os descritores nas seguintes classes:

- Constitucionais, são descritores obtidos da simples observação dos átomos das moléculas, como número de elétrons de valência, massa molar e número de átomos de carbono.
- Físico-químicos, seriam descritores referentes às propriedades físico-químicas das moléculas, como pontos de fusão e ebulição, solubilidade, refratividade molar, coeficiente de partição, solubilidade, pKa e constantes dielétricas.
- Topológicos, que são as variáveis indicadoras tabeladas, como índice de Wiener, índice de Randic, índice de Kier e de Kier-Hall, e índice de Balaban e Hosoya.
- Geométricos ou estruturais, são propriedades intrínsecas à estrutura espacial da molécula, como volume de Van der Waals e de Connolly, áreas superficiais de Van der Waals e de Connolly, comprimentos e ângulos de ligação, ângulo de torção e distância não covalente.
- Mecânico quânticos ou eletrônicos, são descritores obtidos por meio de cálculos mecânico quânticos referentes à estrutura eletrônica das moléculas, como hiperpolarizabilidade, momento de dipolo, densidade eletrônica, cargas atômicas parciais, calor de formação, propriedades termodinâmicas, energia e densidade dos orbitais de fronteira, potencial de ionização e descritores de ligação de hidrogênio.

- Combinados, são descritores derivados da combinação de outros descritores, como os índices eletro-topológicos e a densidade de carga superficial.

Os descritores depois de definidos são selecionados para gerar modelos que os relacionem com a atividade biológica. Esses modelos podem ser gerados por métodos de regressão, combinação linear, dentre outros. Um dos principais métodos de regressão é de quadrados mínimos parciais (*PLS – Partial Least Square*).

1.4. Fundamentos de Química Teórica

Para calcular as propriedades utilizadas na análise quimiométrica, é necessário fazer uma modelagem da estrutura das moléculas sob estudo. Para esta modelagem são utilizados métodos da mecânica quântica, que podem ser *ab initio*, semi-empíricos ou *DFT (Density Functional Theory)*. Estes métodos podem ser utilizados para calcular propriedades eletrônicas das moléculas, além de fornecer a estrutura molecular espacial para que as propriedades físicas referentes a ela sejam calculadas.

A principal diferença entre estes métodos são as próprias características que os definem e não estão presentes nos outros métodos. Os métodos *ab initio* tem por característica a resolução da equação de *Schrödinger* utilizando apenas aproximações matemáticas e geralmente são cálculos com um custo computacional mais elevado. Os métodos semi empíricos, resolvem a equação de *Schrödinger* utilizando parâmetros obtidos de dados experimentais, tem um custo computacional geralmente menor que o dos outros métodos. Os métodos da *DFT* resolvem a equação de *Schrödinger* descrevendo os elétrons através de um funcional de densidade eletrônica, em vez de descreve-los individualmente como nos outros métodos. Os métodos *DFT* apresentam um custo computacional moderado e geram bons resultados quanto à caracterização da estrutura eletrônica das moléculas (Morgon and Coutinho 2007).

2. Desenvolvimento

2.1. Objetivos

O presente estudo tem por objetivo determinar um modelo quantitativo que relacione a estrutura dos compostos β -carbolínicos presentes Tabela 1 com suas respectivas atividades antitumoral (Tabela 2). O modelo deve relacionar algumas de propriedades advindas da estrutura molecular dos compostos com a atividade antitumoral destes contra os cânceres de pulmão e um segundo modelo contra melanoma.

Figura 2 – Estrutura principal dos compostos β -carbolínicos em estudo.

Fonte: (Savariz, Formagio et al. 2010; Barbosa, Formagio et al. 2011)

Tabela 1 – Substituintes da estrutura principal dos compostos β -carbolínicos

Composto	R1	R2	Composto	R1	R2
BC1	H		BC20	2-Cl	
BC2	4-OCH ₃		BC21	2-Cl	

BC3	3-NO ₂		BC22	H
BC4	3-OCH ₃ , 4-OH		BC23	3-NO ₂
BC5	4-N(CH ₃) ₂		BC24	4-N(CH ₃) ₂
BC6	2-Cl		BC25	H
BC7	H		BC26	4-N(CH ₃) ₂
BC8	4-OCH ₃		BC27	H
BC9	4-OH		BC28	3-NO ₂
BC10	3-NO ₂		BC29	4-N(CH ₃) ₂
BC11	3-OCH ₃ , 4-OH		BC30	H

BC12	2-Cl		BC31	3-NO ₂
BC13	H		BC32	4-N(CH ₃) ₂
BC14	4-OCH ₃		BC33	H
BC15	4-OH		BC34	3-NO ₂
BC16	4-N(CH ₃) ₂		BC35	4-N(CH ₃) ₂
BC17	H		BC36	H
BC18	4-OH		37	3-NO ₂
BC19	4-N(CH ₃) ₂		38	4-N(CH ₃) ₂

Fonte: (Savariz, Formagio et al. 2010; Barbosa, Formagio et al. 2011)

Tabela 2 – Atividade antitumoral dos compostos contra câncer de pulmão e melanoma.

Composto	Pulmão (IC50/(μmol/L))	Melanoma (IC50/(μmol/L))
BC1	20,6	>100

BC2	9,2	16,76
BC3	73,41	24,1
BC4	89,29	17,33
BC5	19,09	23,82
BC6	20,75	18,08
BC7	3,42	99,48
BC8	1,28	29,08
BC9	17,82	19,07
BC10	24,47	16,38
BC11	4,82	17,8
BC12	1,64	15,19
BC13	6,36	6,36
BC14	0,1	66,8
BC15	6,69	10,6
BC16	0,06	1,74
BC17	1,59	24,3
BC18	0,15	22,5
BC19	0,2	50,4
BC20	15,4	20,2
BC21	13,9	18,6
BC22	5,72	6
BC23	4,26	7,08
BC24	1,01	0,88
BC25	2,74	5,29
BC26	3,99	10,34
BC27	7,02	8,89
BC28	4,33	7,58
BC29	15,19	8,74
BC30	4,63	>100

BC31	5,25	11,3
BC32	1,28	2,98
BC33	7,58	7,61
BC34	17,98	6,97
BC35	11,04	6,63
BC36	36,77	6,77
BC37	40,43	6,81
BC38	38,55	24,19

Fonte: (Savariz, Formagio et al. 2010; Barbosa, Formagio et al. 2011)

2.2. Materiais e Métodos

Para a modelagem molecular foram utilizados o programa de interface gráfica GaussView05 e o programa de cálculo Gaussian09 (FRISCH M. J. 2009), em ambiente Windows e Linux, respectivamente. Foi empregado o método de cálculo utilizando DFT, com função de base B3LYP e funcional de densidade 6-31G(d,p), para os cálculos de estrutura eletrônica das moléculas no estado fundamental. Para os cálculos da estrutura eletrônica das moléculas nos estados oxidado e reduzido, foi acrescida uma função difusa ao cálculo, para descrever melhor a nuvem eletrônica.

Para geração de alguns descritores foi utilizado o programa e-Dragon. Outros descritores foram obtidos com cálculos com os resultados gerados após os cálculos mecânico-quânticos de otimização e frequência feitos com o Gaussian 09.

A seleção destes descritores foi feita mediante o uso de análise de correlação dos descritores com a atividade antitumoral e o uso do algoritmo genético do programa BuildQSAR (Oliveira and Gaudio 2001; Ferreira, Montanari et al. 2002).

A geração do modelo QSAR foi feita pelo método de regressão por quadrados mínimos parciais (PLS – *Partia Least Squares*) com o programa Pirouett.

Os teste de robustez por *Leave-N-out* e o teste de correlação acaso por aleatorização de y , sendo y a atividade antitumoral dos compostos, foram realizados pelo programa QSARModeling (Ferreira 2013).

2.2.1. Método Quimiométrico – Quadrados Mínimos Parciais (PLS – *Partial Least Squares*)

Neste método de regressão os descritores (matriz X), ou variáveis independentes, são correlacionados com a atividade biológica (vetor y) através da alteração das coordenadas dos compostos sob estudo, gerando uma nova matriz T . Estas novas coordenadas são otimizadas para que seja possível chegar ao valor do vetor y . Logo, estão inclusos no processo de regressão os valores de atividade biológica (Montanari 2011).

Maximizando a covariância de cada componente, entre os escores de t de X e o vetor y , chegando a $Xw = t$ e a $w = \frac{X^T \cdot y}{\|X^T \cdot y\|}$.

No cálculo de escores há a participação da atividade biológica, como mostrado na Figura 3.

Figura 3. Esquema do método de regressão por PLS.

$$\left\{ \begin{array}{l} \text{Bloco } X \left\{ \begin{array}{ll} X = yw^T + E_x & (X^T = wy^T) \\ \hat{w} = X^T(y^T y)^{-1}y & (\text{pesos}) \\ \hat{w} = \frac{\hat{w}}{\sqrt{\hat{w}^T \hat{w}}} & (\text{normalização}) \\ t = X\hat{w} & (\text{escores}) \\ l = X^T t (t^T t)^{-1} & (\text{loadings}) \end{array} \right. \\ \\ \text{Bloco } y \left\{ \begin{array}{l} \text{Relacionando os vetores } t \text{ e } y \\ q = (t^T t)^{-1} t^T y \end{array} \right. \\ \\ \text{Atualização de } X \text{ e } y \left\{ \begin{array}{l} E = X - tl^T \\ f = y - qt \\ X = E \\ y = f \end{array} \right. \end{array} \right.$$

Esse algoritmo é feito para a primeira variável latente e depois repetido para as demais. Com os resultados gerados é calculado o vetor de regressão $\hat{b}$, dado pela equação (1).

$$\hat{b} = W(L^T W)^{-1} q^T \quad (1)$$

2.2.1.1. Métodos de Validação

Os métodos de validação são utilizados para verificar a eficiência dos modelos de regressão. Para a validação do método, a separação em dois grupos, sendo um para geração do modelo e outro para a validação externa, ou aplicação do modelo. Para que este conjunto de validação, ou conjunto de testes, tenha uma aceitabilidade maior, é recomendado o uso de 30% do número de compostos utilizado para a geração do modelo (Montanari 2011).

Para validar o modelo, é necessário testá-lo utilizando apenas os compostos empregados na confecção do modelo, também chamado de conjunto de treinamento. Um método é a validação cruzada com *leave-one-out*, onde uma das amostras é deixada de fora na geração do modelo e é estimada sua atividade biológica. Cada um dos compostos do conjunto de treinamento é excluído uma única vez, então são calculados os respectivos erros de previsão e a soma do quadrado dos erros de previsão, PRESS. A equação (2) representa o cálculo de PRESS, sendo y_{iexp} a atividade biológica experimental e $\hat{y}_{ival}$ a atividade gerada pelo modelo sem o composto i .

$$PRESS = \sum (y_{iexp} - \hat{y}_{ival})^2 \quad (2)$$

Quando menor a diferença entre o valor experimental e o gerado pelo modelo, maior é a capacidade preditiva do modelo em relacionar os descritores com a atividade biológica. Logo, quanto menor o valor de PRESS, melhor será o modelo.

Outro parâmetro que pode ser utilizado para avaliar a validade do modelo é o erro padrão de validação, ou SEP_{val} . Ele consiste na avaliação da capacidade de predição do modelo, por meio do valor de PRESS e o número de amostras do conjunto treino (I), como mostra a equação (3).

$$SEP_{val} = \sqrt{\frac{PRESS}{I}} \quad (3)$$

Há também o erro padrão de calibração, SEP_{cal} , que é calculado quando todas os compostos do conjunto treino são utilizados na geração do modelo. O SEP_{cal} leva em consideração o $\hat{y}_{ical}$, (valor estimado da atividade biológica, obtida utilizando todos os compostos no modelo) e os

graus de liberdades do modelo, dado pelo número de variáveis latentes (A), ou fatores, e o número de compostos usado no modelo, como mostra a equação (4).

$$SEP_{cal} = \sqrt{\frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{I-A-1}} \quad (4)$$

Utilizando os valores de Q^2 e R^2 é possível avaliar a qualidade da predição e do ajuste do modelo. Sendo Q^2 a variância explicada na validação cruzada ou na previsão do modelo, e R^2 a variância explicada no ajuste do modelo. Podem Q^2 e R^2 ser calculados pelas equações (5) e (6):

$$Q^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ival})^2}{\sum (y_{iexp} - y_{med})^2} \quad (5)$$

$$R^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - y_{med})^2} \quad (6)$$

sendo y_{med} o valor médio da atividade biológica experimental dos compostos. Os valores de Q^2 e R^2 remetem-se a um bom modelo quando $Q^2 > 0,5$ e $R^2 > 0,6$ (Montanari 2011).

2.2.1.2. Detecção de Amostras Atípicas

A determinação de amostras atípicas pode ser feita utilizando o resíduo de *Student* e a alavancagem das amostras. O resíduo de *Student* consiste no resíduo padronizado de cada uma das amostras do conjunto de treinamento, calculado durante a validação cruzada do modelo. A alavancagem é a medida da influência de cada uma das amostras no modelo. A melhor maneira de utilizar essas duas informações é em um gráfico com resíduo de *Student* pela alavancagem, com cada amostra sendo um ponto.

Para que uma amostra seja considerada típica, ela deve ter resíduo de *Student* abaixo de 2,0 e alavancagem menor que $3A/I$, sendo A o número de fatores e I o número de amostras no conjunto de treinamento. Caso uma amostra apresente baixa alavancagem e elevado resíduo de *Student*, é provável que tenha ocorrido algum problema na determinação experimental de sua atividade biológica. As demais amostras que forem identificadas com elevada alavancagem e elevado resíduo de *Student*, devem ser avaliadas. As amostras identificadas e avaliadas como atípicas devem ser retiradas do modelo (Montanari 2011).

2.2.1.3. Teste de Robustez

Esse teste consiste em avaliar a robustez do modelo, se ele sofre poucas variações com a retirada de mais compostos do conjunto treino. Para isso são gerados vários modelos com validação cruzada *leave-N-out*, variando o N de 1 até 20 ou 30% do número de compostos presentes no conjunto de treino.

A avaliação é mediante o valor de Q^2 de cada um dos modelos gerados. Se os modelos gerados tiverem valores de $Q^2 > 0,5$ e cada modelo tiver um desvio de Q^2 variando entre 0,05 e -0,05, o modelo pode ser considerado robusto (Montanari 2011).

2.2.1.4. Teste de Correlação ao Acaso

Para verificação de correlação ao acaso no modelo, é utilizada a aleatorização do vetor y , pois é possível que descritores artificiais estejam bem correlacionados com a atividade biológica que gerem um bom modelo, mas sem nenhum significado químico. Esse teste faz uma permutação aleatória da atividade biológica dos compostos do conjunto de treino e mantendo a disposição dos descritores inalterada. Assim, é esperado que o resultado gerado pelo novo modelo seja o pior possível, mas se forem gerados bons modelos, é sinal de que há correlação ao acaso e o modelo inicial não tem significado químico relevante.

Pela literatura, o uso de 10 aleatorizações do vetor y seriam o suficiente para demonstrar se no modelo há ou não correlação ao acaso. Os valores de Q^2 e R^2 obtidos nos modelos randomizados são comparados com o modelo inicialmente proposto. Para um modelo sem correlação ao acaso o esperado é que os valores aleatorizados de Q^2 e R^2 sejam inferiores a 0,3 e 0,4, respectivamente (Montanari 2011).

2.2.2. Método Mecânico Quântico – Teoria do Funcional de Densidade (DFT)

Com relação aos métodos mecânico-quânticos, o uso de cálculos que seguem o modelo da Teoria do Funcional de Densidade (*DFT – Density Functional Theory*) para otimização de estrutura geométrica e cálculos de energia das moléculas é interessante, pois estes cálculos

consideram a correlação eletrônica, a qual é desprezada por outros métodos. A DFT procura resolver a equação de Schrödinger utilizando uma função de probabilidade ($\rho(r)$) como variável para descrever um sistema eletrônico ao em vez de descrever os elétrons individualmente (Morgon and Coutinho 2007).

Desde quando foi proposta a DFT, ela passou por vários aprimoramentos, como o funcional de energia de Thomas-Fermi-Dirac, o teorema de Horemberg e Kohn, as equações de Kohn-Sham, dentre outros aprimoramentos que estão sendo realizados até hoje. O grande diferencial da DFT é ter baixo custo computacional, quando comparado com alguns métodos *ab initio*, que buscam resolver a equação de Schrodinger com o mínimo de aproximações (Morgon and Coutinho 2007).

Na DFT, o ponto de partida foi o uso da densidade eletrônica ($\rho(r)$) para geração de modelos matemáticos, que surgiam e eram aperfeiçoados com o passar dos anos. Quando Dirac aperfeiçoou o modelo de Thomas-Fermi, incluindo a energia de troca obteve-se o funcional de energia Thomas-Fermi-Dirac:

$$E_{TFD} = C_F \int \rho(r)^{\frac{5}{3}} dr + \int \rho(r)v(r) dr + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1-r_2|} dr_1 dr_2 - C_X \int \rho(r)^{\frac{4}{3}} dr, \quad (7)$$

sendo as incógnitas $C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}}$ e $C_X = \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}$.

Os termos da equação do funcional de energia de Thomas-Fermi-Dirac correspondem respectivamente, da esquerda para a direita, à energia de troca, ao potencial de Coulomb, ao potencial externo e à energia cinética. Na equação r , corresponde às coordenadas, e ρ à densidade eletrônica.

Considerando a aproximação de Born-Oppenheimer, desprezando os efeitos relativísticos, tem-se um hamiltoniano de um sistema eletrônico de M núcleos e N elétrons, como mostra a equação (8).

$$\hat{H}_{BO} = -\sum_i^N \frac{1}{2} \nabla_i^2 - \sum_A^M \sum_i^N \frac{Z_A}{|R_A - r_i|} + \sum_{i<j}^N \sum_j^N \frac{1}{|r_i - r_j|} + \sum_{B<A}^M \sum_A^M \frac{Z_A Z_B}{|R_A - R_B|} \quad (8)$$

Nesta equação (8), i e j representam os elétrons, A e B representam os núcleos, r_i são as coordenadas do elétron i , R_A são coordenadas núcleo A , e Z_A é o número atômico do átomo A .

Da esquerda para a direita nesta equação os termos representam, respectivamente, $\hat{T}$ (operador de energia cinética), $\hat{U}$ (potencial externo relacionado aos elétrons), $\hat{V}_e$ (operador de repulsão elétron-elétron, com repulsão Coulombiana os termos de troca e correlação), e $\hat{E}_{rep}$ (operador de repulsão entre os núcleos).

O potencial externo relacionado aos elétrons pode ser substituído por outro termo:

$$\hat{U} = \sum_A^M \sum_i^N \frac{Z_A}{|R_A - r_i|} = \sum_i^N v(r_i) \quad (9)$$

A densidade eletrônica é definida pela seguinte equação:

$$\rho(r) = \int \dots \int \Psi(r_1, r_2, r_3, \dots, r_N)^* \Psi(r_1, r_2, r_3, \dots, r_N) dr_2 dr_3 \dots dr_N \quad (10)$$

O termo $\Psi(r_1, r_2, r_3, \dots, r_N)$ da equação (10) é a solução para o estado fundamental do Hamiltoniano. Logo, pode-se chegar à energia total do sistema, da seguinte maneira,

$$E_0 = \int \Psi(r_1, r_2, r_3, \dots, r_N)^* \hat{H}_{BO} \Psi(r_1, r_2, r_3, \dots, r_N) dr_1 dr_2 \dots dr_N = \left\langle \Psi \left| \hat{H}_{BO} \right| \Psi \right\rangle \quad (11)$$

Se separarmos o potencial externo em um funcional de densidade eletrônica, a energia total pode ser reescrita como na equação a seguir,

$$E_0 = \left\langle \Psi \left| \hat{T} + \hat{V}_e \right| \Psi \right\rangle + \int \rho(r) v(r) dr \quad (12)$$

Com esta equação (12) é possível perceber a relevância do número de elétrons, N, e do potencial externo relacionado aos elétrons, já que estes definem o hamiltoniano do sistema.

Hohemberg-Kohn estabeleceram dois teoremas, sendo que o primeiro diz que o potencial externo é um funcional único de $\rho(r)$, além de uma constante aditiva. Como a energia do sistema é obtida pela resolução da equação de Schrödinger, assim a energia do sistema eletrônico é determinada pela densidade eletrônica $\rho(r)$,

$$E = E_v[\rho] \quad (13)$$

Para elucidar a dependência do potencial externo $u(r)$, coloca-se o índice u.

Em seu segundo teorema, Hohenberg-Kohn estabelecem que, havendo qualquer aproximação da densidade eletrônica, $\tilde{\rho}(r)$, de maneira que $\tilde{\rho}(r) \geq 0$ e $\int \tilde{\rho}(r)dr = N$, a energia total será sempre maior ou igual a energia exata do sistema, logo, $E[\tilde{\rho}] \geq E[\rho] = E_0$. Assim pode-se definir o seguinte funcional como universal, levando em conta que T e V_e se aplicam universalmente a todos os sistemas eletrônicos:

$$F[\rho] = \left\langle \Psi \left| \hat{T} + \hat{V}_e \right| \Psi \right\rangle, \quad (14)$$

Sabendo que $\tilde{\rho}(r)$ define seu próprio $\tilde{v}(r)$, ela acaba definindo também o Hamiltoniano $\tilde{H}$ e a função $\tilde{\Psi}(r_1, r_2, r_3, \dots, r_N)$. É possível utilizar a função $\tilde{\Psi}(r_1, r_2, r_3, \dots, r_N)$ como função tentativa para um sistema de potencial externo $v(r)$. Logo, pelo princípio variacional, obtém-se a equação (15).

$$E_0 = E_v[\rho] = F[\rho] + \int \rho(r)v(r)dr \leq E_v[\tilde{\rho}] = F[\tilde{\rho}] + \int \tilde{\rho}(r)v(r)dr \quad (15)$$

Em 1965 Kohn-Shan reescreveram a equação (12), explicitando a repulsão elétron-elétron de Coulomb e estabeleceram uma nova função universal, $G[\rho]$:

$$E_v[\rho] = G[\rho] + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + \int \rho(r)v(r)dr, \quad (16)$$

onde,

$$G[\rho] = T_S[\rho] + E_{XS}[\rho], \quad (17)$$

sendo $T_S[\rho]$ o funcional de energia cinética de um sistema de elétrons que não interagem e tem a mesma densidade eletrônica do sistema de elétrons que interagem, e $E_{XS}[\rho]$ inclui o termo de interação elétron-elétron de troca e correlação e a parte residual da energia cinética, que é a diferença entre a energia cinética exata e a do sistema de elétrons que interagem ($T[\rho] - T_S[\rho]$).

Além de estabelecer nova função universal, Kohn-Shan também estipulam que é possível utilizar um sistema de referência de elétrons que não interagem com o Hamiltoniano que tenha um potencial local efetivo, $v_{ef}(r)$. Logo,

$$H^{KS} = -\frac{1}{2}\nabla^2 + v_{ef}(r) \quad (18)$$

Este operado Hamiltoniano de Kohn-Shan, H^{KS} , descreve um sistema em que os elétrons não interagem, e não contém interação elétron-elétron.

Como o objetivo é a resolução da equação de Schrödinger, há a necessidade de se estabelecer uma função de onda para o Hamiltoniano H^{KS} . Neste caso a função Ψ^{KS} do estado fundamental do sistema de referência de elétrons que não interagem, é obtida utilizando a mesma aproximação do método de Hartree-Fock. Então, Ψ^{KS} resultará da aproximação do produto anti-simetrizado de N funções de onda de um elétron, $\Psi_i(r_i)$, que podem ser representadas pelo determinante de Slater:

$$\Psi^{KS} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Psi_1^{KS}(r_1) & \Psi_2^{KS}(r_1) & \dots & \Psi_N^{KS}(r_1) \\ \Psi_1^{KS}(r_2) & \Psi_2^{KS}(r_2) & \dots & \Psi_N^{KS}(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1^{KS}(r_N) & \Psi_2^{KS}(r_N) & \dots & \Psi_N^{KS}(r_N) \end{vmatrix} \quad (19)$$

Logo, define-se que os orbitais Kohn-Shan (KS), Ψ^{KS} , são obtidos através da equação de Schrödinger com um elétron:

$$\left(-\frac{1}{2}\nabla^2 + v_{ef}\right)\Psi_i^{KS} = \varepsilon_i \Psi_i^{KS} \quad (20)$$

Para que possamos conectar esse sistema hipotético a um sistema real, é necessário que seja estabelecido um potencial efetivo, de maneira que a densidade eletrônica obtida seja igual à densidade eletrônica fundamental. Assim, obtém-se a seguinte equação:

$$\rho_S(r) = \sum_i^N 2|\Psi_i^{KS}(r)|^2 = \rho_0(r) \quad (21)$$

Utilizando um procedimento autoconsistente, pode-se calcular a energia cinética:

$$T_S[\rho] = \sum_i^N \left\langle \Psi_i^{KS} \left| -\frac{1}{2}\nabla_i^2 \right| \Psi_i^{KS} \right\rangle \quad (22)$$

Se minimizarmos energia cinética na equação (14), restringindo as funções de um elétron a serem ortonormais, $\langle \Psi_i^{KS} | \Psi_j^{KS} \rangle = \delta_{ij}$, o resultado é a equação do potencial efetivo,

$$v_{ef}(r) = v(r) + \int \frac{\rho(r')}{|r-r'|} + v_{XC}(r), \quad (23)$$

sendo que,

$$v_{XC}(r) = \frac{\delta E_{XC}[\rho]}{\delta \rho(r)} \quad (24)$$

Se analisarmos conjuntamente as equações (20), (23) e (24), obteremos a representação do esquema de campo autoconsistente de Kohn-Shan (KS-SCF, *Kohn-Shan – Self Consistent Field*).

Assim, as equações de Kohn-Shan são resolvidas mediante um processo autoconsistente, devido à dependência que o potencial efetivo ($v_{ef}(r)$) tem da densidade eletrônica ($\rho(r)$). Isso é mostrado na Figura 4.

Figura 4. Esquema do procedimento autoconsistente de Kohn-Shan.

As equações de Kohn-Shan geram equações para um elétron, que são aplicáveis para sistemas de muitos elétrons. A princípio, estas equações são exatas, uma vez que elas contemplam totalmente os efeitos da correlação eletrônica de troca e correlação. As soluções destas equações de Kohn-Shan equivalem à resolução do problema variacional da DFT.

Sendo a densidade eletrônica, $\rho(r)$, e o potencial efetivo, $v(r)$, são obtidas mediante as coordenadas, constata-se que os resultados gerados são locais. Assim, o potencial efetivo ($v_{ef}(r)$) responsável pelos efeitos de troca-correlação e a diferença de energia cinética ($T[\rho] - T_S[\rho]$), são também locais.

Utilizando-se a equação (16), é possível determinar a fórmula da energia total:

$$E = \sum \varepsilon_i - \frac{1}{2} \int \frac{\rho(r_1)\rho(r_2)}{|r_1-r_2|} dr_1 dr_2 + E_{XC}[\rho] - \int \rho(r)v_{XC}(r)dr \quad (25)$$

3. Resultados e Discussão

Com o programa e-Dragon e o cálculo dos descritores eletrônicos, utilizando os resultados da otimização da estrutura eletrônica com o programa Gaussian 09, foram obtidos entre 1200 e 1300 descritores. Estes descritores apresentavam baixa correlação com as atividades biológicas dos compostos, e não havia diferença suficiente entre a atividade biológica dos compostos para separá-los em classes e utilizar métodos não supervisionados. Assim, foi utilizado apenas o método supervisionado de regressão por quadrados mínimos parciais.

Como este método é uma análise multivariada, a relevância maior é o que o conjunto de determinadas variáveis pode exprimir, não sendo tão relevante a correlação delas, individualmente, com a atividade biológica.

Para cada tipo de tumor foi gerado um modelo diferente, com diferentes variáveis. Isto se deve ao fato das atividades biológicas serem diferentes no tratamento de cada tipo de tumor, e, matematicamente, uma função real não pode expressar diferentes valores de y para um mesmo valor de x . Sendo necessário gera um modelo diferente para cada descrição dos compostos no tratamento de cada tipo de câncer.

3.1. Estudo QSAR – Câncer de Pulmão

Usando o algoritmo genético do programa BuildQSAR e análise da correlação entre a atividade biológica e os descritores gerados, para selecionar as variáveis, foi possível chegar a um modelo com 6 descritores dos, aproximadamente, 1300 descritores gerados. Destes descritores, quase todos, são valores tabelados ou que podem ser obtidos por meio de equações embutidas no programa e-Dragon e alguns são calculados utilizando dados dos cálculos quânticos. Com o programa Pirouett foi gerado o modelo PLS, utilizando 6 descritores que representavam bem o conjunto de treinamento. Assim, pela regressão, foi obtida a seguinte equação:

$$y = 0,5045\mu + 0,8879Mor31m + 0,7552Mor27u - 0,5506GATS2p - 0,6211HATS3p - 0,5663H8m$$

Como se trata de uma análise multivariada, os descritores não tem significado relevante sozinhos, sendo possível propor algo analisando o conjunto de descritores selecionados. Os descritores μ , HATS3p e GATS2p são a polarizabilidade da molécula e valores ponderados nela, respectivamente. Isso indica que a polarizabilidade e outros efeitos relacionados a ela são importantes para a atividade dos compostos. Como os descritores HATS3p e GATS2p referem-se à região das posições 2 e 3, podemos supor que o efeito de polarizabilidade próximos a essa região tem maior influência na atividade dos compostos contra o câncer de pulmão.

Os descritores Mor31m e Mor27u, são dados de difração eletrônica ponderados não massa e não ponderados, respectivamente. Esses descritores são o sinal da difração eletrônica gerado pela estrutura estática da molécula. Como se trata de uma avaliação de sinal eletrônico, podemos presumir que a polarização da molécula afeta o cálculo dessa propriedade.

O descritor H8m mostra que há alguma importância na posição 8 da molécula. Não houver variação na posição 8, mantendo apenas o H, mas em outros artigos é relatado que a substituição na posição 8 influencia a atividade da antitumoral dos compostos b-carbolínicos. Pode ser que a polarizabilidade na molécula gere algum efeito atrativo tornando o H da posição 8 mais, ou menos, disponível ocasionando a variação que gera o descritor H8m.

Os valores de atividade antitumoral gerados pela equação, tanto os valores calculados quanto os valores obtidos por validação interna, são valores próximos do determinado experimentalmente, já que os resíduos são, na maioria dos casos, valores baixos de acordo com a Tabela 3. As médias da atividade antitumoral calculada e a de validação geradas pelo modelo são, respectivamente, idêntica e muito próxima à média da determinada experimentalmente. Isso pode gerar uma falsa sensação de eficiência do modelo.

Os valores muito baixos de média simples dos resíduos corroboram que a média dos valores calculados pelo modelo seja próxima ou igual à média dos valores experimentais. O que acaba ocorrendo, entretanto, é que há um autocancelamento dos erros. Como alguns são positivos e outros são negativos, na somatória para calcular a média, ocorre o cancelamento e isso dá a falsa impressão de que os erros são baixos e falseia a ideia sobre a eficiência do modelo.

Para melhor elucidar a variação dos valores determinados pelo modelo com relação aos valores experimentais, foi feita a média dos resíduos em módulo, melhorando a percepção da

intensidade desvio dos resultados. Essa média com o valor dos resíduos em módulo gera um valor muito maior do que o remetido pela média simples, dando uma percepção real sobre o desvio dos valores obtidos pelo modelo em relação aos valores experimentais.

Tabela 3 – Valores de atividade antitumoral contra câncer de pulmão, obtidos experimentalmente e calculados pelo modelo.

	pIC_{50exp}	pIC_{50Val}	Resíduo_{Val}	pIC_{50Cal}	Resíduo_{Cal}
BC21	4,857	5,145	-0,288	4,971	-0,114
BC1	4,686	4,673	0,013	4,635	0,051
BC2	5,036	5,038	-0,002	4,984	0,052
BC3	4,134	4,070	0,064	4,118	0,016
BC6	4,683	5,175	-0,492	4,880	-0,197
BC7	5,466	5,205	0,260	5,276	0,190
BC8	5,893	5,651	0,242	5,645	0,248
BC9	4,749	5,130	-0,381	4,980	-0,231
BC10	4,611	4,597	0,014	4,680	-0,069
BC11	5,317	5,647	-0,331	5,514	-0,197
BC12	5,785	5,740	0,045	5,720	0,065
BC13	5,197	5,736	-0,539	5,658	-0,462
BC15	5,175	5,867	-0,692	5,723	-0,549
BC16	7,222	6,344	0,878	6,622	0,600
BC18	6,824	6,007	0,817	6,160	0,664
BC19	6,699	6,827	-0,128	6,855	-0,156
BC20	4,812	4,797	0,016	4,803	0,009
BC23	5,371	5,864	-0,494	5,730	-0,359
BC25	5,562	6,037	-0,475	5,978	-0,416
BC26	5,399	5,573	-0,174	5,533	-0,134
BC27	5,154	5,175	-0,021	5,180	-0,027
BC28	5,364	4,409	0,954	4,777	0,587

BC29	4,818	4,858	-0,039	4,800	0,019
BC30	5,334	4,560	0,775	4,812	0,523
BC32	5,893	5,701	0,192	5,708	0,185
BC33	5,120	5,422	-0,301	5,384	-0,263
BC37	4,393	4,411	-0,017	4,429	-0,035
Média dos valores	5,317	5,321	0,320 ^a /-0.00384 ^b	5,317	0,238 ^a /-1.5E-07 ^b

a: A média foi feita utilizando os resíduos em módulo.

b: Média simples.

O valor baixo de PRESS demonstra que os erros tem baixa intensidade, já que por avaliar a somatória do quadrado dos erros, o sinal positivo ou negativo não é relevante. SEP_{Val} e SEP_{Cal} , que esboçam os erros de validação e calibração, também apresentam valores baixos (Tabela 4). Como estes três parâmetros, PRESS, SEP_{Val} e SEP_{Cal} , avaliam o erro entre os valores gerados pelo modelo e os dados experimentais, baixos valores indicam que o modelo é bom.

O modelo apresenta um bom ajuste e uma capacidade de predição boa, como indicam os valores de R^2 e Q^2 presentes na Tabela 4. O valor de R^2 está bem acima do que a literatura diz ser um bom valor para o modelo ($r^2 > 0,6$), mostrando que ao descrever os compostos os valores teóricos foram muito próximos dos experimentais e o os compostos estão muito próximos da reta descrita pela equação (Figura 5). O valor de Q^2 também está acima do que a literatura diz ser um bom modelo ($q^2 > 0,5$), mostrando que o modelo tem boa capacidade em prever a atividade biológica dos compostos do modelo, ou seja, tem uma boa capacidade preditiva.

Tabela 4 – Resultados do método *PLS* gerados com conjunto de treinamento.

Variância	37,01908
Descrição dos compostos pelo modelo (%)	100
SEP_{Val}	0,43369
SEP_{Cal}	0,363267
PRESS	5,078352
Q^2	0,625268

R^2	0,805249
-------	----------

Figura 5 – Gráfico dos dados experimentais calculados pelo método *PLS* gerado com o conjunto de treinamento.

Analisando o modelo quanto a amostras atípicas, na Figura 6 estão expostos os compostos no gráfico de resíduos de Student vs Alavancagem. Pelo que diz a literatura, para um modelo com 6 fatores e 27 compostos, as amostras típicas devem estar abaixo de 0,67 de alavancagem. No gráfico não há compostos com valor de alavancagem maior que 0,67, ao contrário, os compostos apresentam baixos valores de alavancagem, sendo 15 compostos abaixo de 0,1, 9 compostos entre 0,1 e 0,2 e 3 compostos entre 0,2 e 0,3.

Nos resíduos de *Student* há apenas dois compostos um pouco acima de 2, que são os compostos BC28 e BC16. Como eles apresentam valores baixos de alavancagem, não se tratam de amostras atípicas, mas é provável que apresentem algum problema quanto à determinação experimental de suas atividades biológicas.

Figura 6 – Gráfico de Resíduos de *Student* vs Alavancagem para detecção de amostras atípicas.

Para avaliar se o modelo é eficaz quanto à predição, ele foi aplicado aos compostos do conjunto de teste, predizendo os possíveis valores de atividade antitumoral destes compostos. Como expresso na Tabela 5, a diferença entre os valores preditos e os valores experimentais é pequena, mostrando que o modelo apresenta boa capacidade de predição. Isto também fica claro quando observamos a Figura 7, onde é vemos que os compostos se localizam próximo à reta descrita pela equação do modelo. Assim, há boa relação entre os valores de atividade antitumoral preditos e os determinados experimentalmente.

Tabela 5 – Valores experimentais e preditos pelo modelo, para o conjunto teste, de atividade antitumoral contra câncer de pulmão.

	pIC50 (predito)	pIC50 (experimental)	Resíduos
BC4	3,902516	4,049197	0,146681
BC5	5,074039	4,719194	-0,354845
BC14	6,657093	7,000	0,342907
BC17	6,039574	5,798603	-0,240971
BC22	5,448682	5,242604	-0,206078

BC24	5,589952	5,995679	0,405727
BC31	4,497111	5,279841	0,78273
BC34	4,840135	4,74521	-0,094925
BC35	5,593274	4,957031	-0,636243
BC36	4,605764	4,434506	-0,171258

Figura 7 – Gráfico dos valores preditos para o conjunto teste com os valores experimentais da atividade antitumoral contra câncer de pulmão.

Os baixos valores de R^2 dos modelos randomizados, como mostra a Figura 8, mostram que os compostos estão longe da reta descrita pela equação gerada por cada modelo. Estes modelos não descrevem os compostos do conjunto de treinamento de maneira adequada, tendo um ajuste baixo dos compostos com a reta. Os valores de atividade antitumoral teórica estão longe da reta que descreve equação gerada por regressão. Isso demonstra que o conjunto de descritores escolhidos gera um modelo bem ajustado, que descreve bem a atividade dos compostos, apenas para os compostos com os valores originais de atividade.

Figura 8 – Valores de R^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral contra câncer de pulmão.

A capacidade preditiva dos modelos com y aleatorizado é reduzida, como mostra a Figura 9. Os valores baixos de Q^2 mostram que os modelos randomizados não conseguem fazer uma boa predição dos valores de atividade antitumoral dos compostos do conjunto de treinamento. Este resultado reforça o que mostrou a avaliação dos valores de R^2 , que os descritores selecionados geram um bom modelo apenas para os valores originais de atividade.

Figura 9 – Gráfico dos valores de Q^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral para câncer de pulmão.

Analisando os resultados que constam na Figura 10, fica claro que não há correlação ao acaso e que o modelo descreve bem apenas os dados originais de atividade. Os valores de R^2 e Q^2 para os modelos aleatorizado estão abaixo de 4,0 e 0,0, respectivamente, confirmando que o modelo original não foi gerado ao acaso.

Figura 10 – Gráfico dos valores de R^2 e Q^2 obtidos nas randomizações dos valores de atividade antitumoral para câncer de pulmão.

O teste de robustez do modelo pelo processo de *leave-N-out*, com a retirada de até 12 compostos do conjunto treino para a confecção do modelo, como mostra a Figura 11, apresentou uma flutuação considerável. Provavelmente essa flutuação ocorre devido ao número de compostos pequeno, 27 compostos. A retirada de 10 compostos é uma retirada de aproximadamente 40% dos compostos do conjunto de treinamento. Isso mostra que o modelo é sensível à retirada de compostos.

Figura 11 – Gráfico da variação de Q^2 com a retirada de compostos para geração do modelo.

3.2. Estudo QSAR - Melanoma Cutâneo

Para confecção de um modelo que descrevesse a atividade antitumoral dos compostos para o tratamento de melanoma, foram selecionadas as variáveis G2m, G1u, G3u, RDF095e, BELp4 e BEHv8.

Com o método de regressão por quadrados mínimos parciais foi possível obter um modelo para relacionar as propriedades advindas da estrutura dos compostos do conjunto de treinamento com a atividade antitumoral destes. Esse modelo gerou a seguinte equação:

$$y = -0,6761G2m + 1,1365RDF095e + 0,5072BELp4 + 1,0076BEHv8 + 2,2900G1u + 0,5079G3u$$

Os descritores G2m, G1u e G3u são descritores globais do tipo WHIM, o primeiro ponderado pela massa da molécula e os outros dois sem ponderação. Estes descritores são construídos de forma que capturem informações relevantes quanto a tamanho molecular,

forma, simetria e átomos na molécula. São descritores que levam em consideração apenas parâmetros físicos e não eletrônicos.

O Descritor RDF095e baseia-se numa função de distribuição radial sobre a distribuição de distâncias dos átomos na molécula. Essa função de distribuição pode ser interpretada como a probabilidade localizar estes átomos em um determinado volume esférico. Este descritor, como os três anteriores, também faz uso apenas de parâmetros físicos da molécula.

BELp4 e BEHv8 são descritores topológicos, relacionados respectivamente à polarizabilidade e ao volume de Van der Waals da molécula. Eles são gerados por autovalores da matriz de Burden, gerada com as distâncias topológicas dos átomos da molécula.

Podemos perceber que dos 6 descritores mencionados, 5 são referentes à estrutura física da molécula e um à uma propriedade eletrônica. Isso demonstra que o arranjo físico da estrutura da molécula é o principal fator de influência em sua atividade antitumoral, sendo os parâmetros eletrônicos de importância menor.

O modelo gerado pela regressão pode calcular os valores de atividade antitumoral dos compostos, fazendo uma descrição dos compostos, e tentar, apenas pela equação gerada, prever valores de atividade antitumoral para estes mesmos compostos. Como mostra a Tabela 6, podemos ver estes resultados e os resíduos destes com a relação à atividade antitumoral experimental.

Os valores calculados pelo modelo próximos dos valores experimentais, avaliando-os individualmente. A média dos valores de atividade antitumoral é praticamente a mesma, tendo uma diferença mínima entre a calculada para atividade experimental e a de validação, mostrando que o modelo descreve bem o conjunto de treinamento. Analisando os resíduos e a média dos resíduos é perceptível que a média simples calculada embute o efeito de autocancelamento de erros, também embutido na média das atividades calculadas pelo modelo. Calculando a média do módulo dos resíduos, determinamos a média da intensidade dos resíduos, desconsiderando se o desvio do valor é real de atividade antitumoral é para mais ou para menos. Assim não há efeito de cancelamento de erros para os resíduos e fica claro que há um desvio dos valores calculados com relação aos experimentais.

Tabela 6 – Valores de atividade antitumoral contra melanoma, obtidos experimentalmente e calculados pelo modelo.

Compostos	pIC_{50EXP}	pIC_{50Val}	Resíduo_{Val}	pIC_{50Cal}	Resíduo_{Cal}
BC21	4,730	4,622	0,109	4,636	0,095
BC4	4,761	4,684	0,077	4,691	0,071
BC6	4,743	4,686	0,057	4,711	0,032
BC8	4,536	4,637	-0,100	4,605	-0,069
BC9	4,720	4,756	-0,036	4,747	-0,027
BC10	4,786	4,786	0,000	4,793	-0,007
BC12	4,818	4,816	0,003	4,815	0,003
BC13	5,197	5,036	0,160	5,107	0,089
BC14	4,175	4,242	-0,066	4,201	-0,026
BC15	4,975	4,902	0,073	4,911	0,064
BC16	5,759	5,836	-0,077	5,810	-0,051
BC17	4,614	4,572	0,043	4,580	0,034
BC18	4,648	4,656	-0,008	4,656	-0,008
BC19	4,298	4,489	-0,191	4,397	-0,100
BC20	4,695	5,045	-0,350	4,936	-0,242
BC22	5,222	5,137	0,085	5,153	0,069
BC23	5,150	5,199	-0,049	5,193	-0,043
BC25	5,277	5,062	0,214	5,091	0,185
BC26	4,985	4,978	0,008	4,981	0,004
BC27	5,051	5,125	-0,074	5,115	-0,064
BC28	5,120	5,229	-0,108	5,205	-0,085
BC29	5,058	5,115	-0,056	5,093	-0,035
BC31	4,947	4,973	-0,026	4,963	-0,016
BC33	5,119	4,943	0,176	4,972	0,147
BC34	5,157	5,183	-0,026	5,178	-0,021
Média dos valores	4,902	4,908	0,087 ^a /-0,007 ^b	4,902	0,063 ^a /0,0 ^b

a: A média foi feita utilizando os resíduos em módulo.

b: Média simples.

Outras propriedades que permitem avaliar a qualidade do modelo estão na Tabela 7. Os valores baixos de PRESS, SEP_{Val} e SEP_{Cal} indicam que os erros são baixos, logo os resultados gerados pelo modelo estão próximos do experimental. A baixa variância também denota que os valores calculados não apresentam grandes variações com relação aos valores experimentais.

O ajuste dos compostos à reta descrita pelo modelo e a capacidade preditiva são elucidados pelos valores de R^2 e Q^2 , respectivamente. Por estarem próximos de 1,0, esses valores indicam que o modelo descreve bem os compostos quanto a sua atividade antitumoral e apresenta boa capacidade de predizer esses valores.

Tabela 7 – Resultados extraídos do modelo gerado com o método PLS.

Variância	6,264347
Descrição dos compostos pelo modelo (%)	100
PRESS	0,342972
SEP_{Val}	0,117128
SEP_{Cal}	0,100314
Q²	0,872615
R²	0,932725

Observando o gráfico da Figura 12, fica evidente o bom ajuste do modelo, sendo quase linear a relação entre o valor experimental e o calculado da atividade antitumoral. A disposição dos compostos em relação à reta descrita pelo modelo corrobora os dados da Tabela 7, denotando um erro baixo entre o valor experimental e o calculado de atividade antitumoral dos compostos.

Figura 12 – Gráfico das atividades experimentais calculadas com modelo *PLS*, gerado para o conjunto de treinamento contra melanoma.

A determinação de amostras atípicas pode ser feita pelo gráfico da Figura 13. Neste gráfico podemos perceber a dispersão dos compostos de acordo com o valor de alavancagem e resíduos de *Student*. Pela literatura o limite de alavancagem para os compostos seria de 0,7, já que são usados 6 fatores para 25 compostos. Os compostos que mais se destacam neste gráfico são BC20 e BC16.

O composto BC20 apresente baixo valor de alavancagem, mas um valor de resíduo de *Student* discrepante em relação aos outros compostos, isto mostra que, provavelmente, algum erro ocorreu durante a determinação de sua atividade antitumoral experimental.

O composto BC16 apresenta baixo valor de resíduo de *Student*, mas também um alto resíduo. Mesmo ainda dentro do limite de alavancagem, esse composto provavelmente se distancia dos demais, pois é o único com atividade antitumoral elevada o suficiente para ser considerado ativo.

Figura 13 – Gráfico de alavancagem para identificação de amostra atípica.

Avaliando a capacidade preditiva do modelo, e se este realmente é eficiente, foi feita a predição da atividade antitumoral para os compostos do conjunto de teste. Os resultados constam na Tabela 8, onde podemos observar que o valor predito para os compostos é próximo do valor experimental de sua atividade antitumoral. Analisando o valor médio da atividade antitumoral experimental e calculada, há uma diferença pequena, mostrando que o modelo descreve melhor o conjunto de treinamento que o conjunto teste. O valor da média dos resíduos, tanto a simples quanto a dos resíduos em módulo, demonstra a diferença do valor predito com o experimental.

Tabela 8 – Resultados da predição dos compostos do conjunto de teste.

	pIC_{50EXP}	pIC_{50PREDITO}	Resíduos
BC2	4,776	4,816	-0,040
BC3	4,618	4,743	-0,125
BC5	4,623	4,489	0,134
BC7	4,002	4,455	-0,453
BC11	4,750	4,904	-0,155

BC24	6,056	6,140	-0,084
BC32	5,526	5,603	-0,077
BC35	5,178	5,021	0,157
BC36	5,169	5,404	-0,235
BC37	5,167	5,050	0,117
BC38	4,616	5,020	-0,404
Média dos valores	4.953	5.059	0.180 ^a / -0.106 ^b

a: A média foi feita utilizando os resíduos em módulo.

b: Média simples.

O gráfico da Figura 14 elucida os resultados mostrados na Tabela 8. Podemos observar também que a predição foi mais eficiente para compostos com valores mais altos de atividade antitumoral. Os compostos estão localizados próximos à reta descrita pelo modelo, mostrando uma boa aproximação dos resultados preditos com relação à atividade antitumoral.

Figura 14 – Gráfico dos valores preditos com os valores experimentais da atividade antitumoral contra melanoma do conjunto teste.

Para verificar se não houve a uma correlação ao acaso na geração do modelo, foi feito o teste de aleatorização dos valores de atividade antitumoral. Avaliando os parâmetros Q^2 e R^2 por gráficos destes em relação ao coeficiente de correlação Pearson, Figuras 15 e 16, podemos perceber que os modelos aleatorizados não foram bons para descrever os compostos do conjunto de treinamento. Esses modelos são avaliados como ruins, pois apresentam baixa correlação de Pearson, ou seja, o valor de atividade antitumoral é diferente do original, os valores de Q^2 estão abaixo de 0,0 e os valores de R^2 encontram-se abaixo de 0,4.

Na Figura 16 fica claro a concentração dos modelos aleatorizados na região onde os valores de Q^2 são menores que zero e os valores de R^2 são menores que 0,4. Isso mostra que o modelo gerado com os dados originais não ocorreu ao acaso.

Figura 15 – Gráfico dos valores de Q^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral para melanoma.

Figura 16 – Gráfico dos valores de R^2 pela correlação de Pearson, obtidos na aleatorização dos valores de atividade antitumoral para melanoma.

Figura 17 – Gráfica Q^2 e R^2 dos modelos aleatorizados de atividade antitumoral para melanoma.

A Figura 18 mostra a variação do valor de Q^2 com a retirada dos compostos, teste conhecido como *leave-N-out*, para determinação da robustez do modelo. Para a retirada de até 5 compostos o modelo apresentou uma baixa flutuação, mas para a retirada de mais compostos essa flutuação aumentou, demonstrando que o modelo é sensível à retirada de compostos. A sensibilidade do modelo em relação à retirada de compostos provavelmente se deve ao número de compostos que estão presentes no modelo. Como o conjunto de treinamento tem 25 compostos, a retirada de 10 compostos representa praticamente a retirada de 40% do conjunto de treinamento. Caso o conjunto de treinamento tivesse mais compostos, seria esperado uma robustez maior, ou seja, uma menor flutuação dos valores de Q^2 com retirada de compostos. Assim, este modelo é razoavelmente robusto, suportando a retirada de até 5 compostos.

Figura 18 – Gráfico de Q^2 para N amostras retiradas na validação cruzada.

Comparando os resultados para geração dos dois modelos, podemos perceber alguma relação entre suas variáveis e o local onde ocorrem os tumores. No caso dos descritores para o modelo contra o câncer de pulmão, percebermos que efeitos relacionados à polarização têm maior influencia na atividade antitumoral dos compostos. Provavelmente isto está relacionado ao fato de o pulmão ser um órgão altamente vascularizado, devido às trocas gasosas ocorridas durante a respiração, e o sangue tem caráter polar. Assim o caráter polar das moléculas propicia sua ação nas células daquele meio.

No caso do melanoma, trata-se de uma região mais externa do corpo, a pele. Essa região do corpo é um pouco mais oleosa que as demais. Assim características relacionadas à polaridade da molécula não seriam tão influentes quanto características relacionadas às estrutura física da molécula. Efeitos de polaridade ainda são relevantes, como evidencia a presença do descritor BELp4, mas muito maior é a contribuição de características físicas da molécula.

4. Conclusão

Os modelos gerados para relacionar quantitativamente as propriedades estruturais dos compostos b-carbolínicos com sua atividade antitumoral contra câncer de pulmão e contra melanoma cutâneo, foram satisfatórios. Ambos apresentaram bons valores de ajustes para o conjunto de treinamento e também bons valores na predição da atividade antitumoral dos compostos do conjunto de teste. Os valores de erro foram baixos concordando com os bons valores de ajuste e predição.

Quanto à validação do modelo, a robustez para os dois modelos gerados é razoável, devido ao número de compostos ser baixo. Uma quantidade maior de compostos para o estudo geraria modelos mais robustos. Ficou comprovado pelos resultados de aleatorização da atividade antitumoral, que os modelos não foram gerados por uma correlação ao acaso, mas que realmente o conjunto de descritores correlaciona-se bem com a atividade antitumoral original.

Os descritores utilizados em cada modelo estão bem relacionados com as características químicas dos órgãos onde ocorrem os tumores, no pulmão e na pele.

Embasado nestes resultados é possível afirmar que os modelos relacionam adequadamente as propriedades derivadas da estrutura molecular com os valores quantitativos de atividade antitumoral dos compostos.

5. Bibliografia

Barbosa, V. A., A. S. N. Formagio, et al. (2011). "Synthesis and antitumor activity of b-carboline 3-(substituted-carbohydrazide) derivatives." Bioorganic & Medicinal Chemistry **19**: 6400–6408.

Barros, L. H. C., S. G. P. Cecatti, et al. (2012). "Tratamentos." Retrieved 03/10, 2012, from <http://www.oncoguia.org.br/conteudo/tratamentos/77/50/>.

Cao, R., W. Fan, et al. (2013). "Synthesis and structure-activity relationships of harmine derivatives as potential antitumor agents." European Journal of Medicinal Chemistry **60**: 135 - 143.

Cao, R., X. Guan, et al. (2010). "Design, synthesis and 3D-QSAR of b-carboline derivatives as potent antitumor agents." European Journal of Medicinal Chemistry **45**: 2503-2515.

Cao, R., W. Peng, et al. (2005). "Synthesis and in vitro cytotoxic evaluation of 1,3-bisubstituted and 1,3,9-trisubstituted b-carboline derivatives." European Journal of Medicinal Chemistry **40**: 249–257.

Cao, R., W. Peng, et al. (2007). "b-Carboline Alkaloids: Biochemical and Pharmacological Functions." Current Medicinal Chemistry **14**: 479-500.

Chen, Z., R. Cao, et al. (2010). "Synthesis of novel b-carbolines with efficient DNA-binding capacity and potent cytotoxicity." Bioorganic & Medicinal Chemistry Letters **20**: 3876–3879.

Chen, Z., R. Cao, et al. (2010). "Synthesis, cytotoxic activities and DNA binding properties of b-carboline derivatives." European Journal of Medicinal Chemistry **45**: 4740-4745.

Ferreira, J. P. A. M. M. C. (2013). "QSAR MODELING: UM NOVO PACOTE COMPUTACIONAL OPEN SOURCE PARA GERAR E VALIDAR MODELOS QSAR." Química Nova **36**(4): 554-560.

Ferreira, M. M. C., C. A. Montanari, et al. (2002). "Seleção de Variáveis em QSAR." Química Nova **25**(3): 439-448.

Formagio, A. S. N., L. T. D. Tonin, et al. (2008). "Synthesis and antitumoral activity of novel 3-(2-substituted-1,3,4-oxadiazol-5-yl) and 3-(5-substituted-1,2,4-triazol-3-yl) b-carboline derivatives." Bioorganic & Medicinal Chemistry **16**: 9660–9667.

FRISCH M. J., T. G. W., SCHLEGEL H. B., SCUSERIA G. E., ROBB M. A. e CHEESEMAN J. R.; SCALMANI, G.; BARONE, V.; MENNUCCI, B.; PETERSSON, G. A.; NAKATSUJI, H.; CARICATO, M.; LI, X.; HRATCHIAN, H. P.; IZMAYLOV, A. F.; BLOINO, J.; ZHENG, G.; SONNENGERG, J. L.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; VREVEN, T.; MONTGOMERY, J. A; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M.; HEYD, J. J.; BROTHERS, E.; KUDIN, K. N.; STAROVEROV, V. N.; KOBAYASHI, R.; NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A.; BURANT, J. C.; IYENGAR, S. S.; TOMASI, J.; COSSI, M.; REGA, N.; MILLAM, J. M.; KLENE, M.; KNOX, J. E.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERTS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; ZAKRZEWSKI, V. G.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; DAPPRICH, S.; DANIELS, A. D.; FARKAS, O.; FORESMAN, J. B.; ORTIZ, J. V.; CIOŚŁOWSKI, J.; FOX, D. J. (2009). Gaussian 09.

Guan, H., H. Chen, et al. (2006). "Design of β -carboline derivatives as DNA-targeting antitumor agents." European Journal of Medicinal Chemistry **41**: 1167–1179.

Herraiz, T. (2000). "Analysis of the bioactive alkaloids tetrahydro-b-carboline and b-carboline in food." Journal of Chromatography A **881**: 483–499.

Herraiz, T. and J. GALISTEO (2003). "Tetrahydro- β -carboline Alkaloids Occur in Fruits and Fruit Juices. Activity as Antioxidants and Radical Scavengers." Journal of Agricultural and Food Chemistry **51**: 7156-7161.

Honório, K. M., R. C. Garratt, et al. (2007). "3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators " Journal of Molecular Graphics and Modelling **25**(6).

Honorio, K. M., Richard C. Garratt, et al. (2005). "Hologram quantitative structure–activity relationships for a series of farnesoid X receptor activators." Bioorganic & Medicinal Chemistry Letters **15**(12): 3119-3125.

INCA, I. N. d. C.-. (2013). "Câncer de pele." from http://www.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pele_melanoma.

Ishida, J., H. K. Wang, et al. (1999). "Antitumor agents 201.1 Cytotoxicity of harmine and β -carboline analogs." Bioorganic & Medicinal Chemistry Letters: 3319-3324.

Junior, M. T. R. (2009). Síntese e Atividade Citotóxica, Leishmanicida e Sobre o Sistema Nervoso Central de Compostos b-Carbólicos. Instituto de Química UNICAMP. Campinas, UNICAMP.

Kumar, R., L. Gupta, et al. (2010). "Synthesis and cytotoxicity evaluation of (tetrahydro-b-carboline)-1,3,5-triazine hybrids as anticancer agents." European Journal of Medicinal Chemistry **45**: 2265–2276.

Montanari, C. A. (2011). QUÍMICA MEDICINAL: Métodos e Fundamentos em Planejamento de Fármacos São Paulo, Editora da Universidade de São Paulo.

Morgon, N. H. and K. Coutinho (2007). Métodos de Química Teórica e Modelagem Molecular. São Paulo, Livraria da Física.

Oliveira, D. B. d. and A. C. Gaudio (2001). "BuildQSAR: A new computer program for QSAR analysis." Quantitative Structure-Activity Relationships **19**(6): 599–601.

Papavergou, E. and T. Herraiz (2003). "Identification and occurrence of 1,2,3,4-tetrahydro-b-carboline-3-carboxylic acid: the main b-carboline alkaloid in smoked foods." Food Research International **36**: 843–848.

Rocha, J. R. d. (2010). Planejamento de Inibidores das Enzimas Gliceraldeído-3-fosfato Desidrogenase e Diidrorotato Desidrogenase de Trypanosoma cruzi. Instituto de Química de São Carlos. São Carlos, Universidade de São Paulo.

Saúde, M. d. (2010). "Doenças associadas ao uso dos derivados do tabaco." from <http://www.inca.gov.br/tabagismo/frameset.asp?item=atento&link=doencas.htm>.

Savariz, F. C. (2009). SÍNTESE, ATIVIDADE ANTITUMORAL E ANTIMICROBIANA DE NOVOS DERIVADOS 3-[2-TIOXO-3-(ALQUILAMINOMETIL)-1,3,4-OXADIAZOL-5-IL] - β -CARBOLÍNICOS-1-

FENILSSUBSTITUÍDOS. Departamento de Química do Centro de Ciências Exatas. Maringá, Universidade Estadual de Maringá.

Savariz, F. C., A. S. N. Formagio, et al. (2010). "Synthesis, Antitumor and Antimicrobial Activity of Novel 1-Substituted Phenyl-3-[3-alkylamino(methyl)-2-thioxo-1,3,4-oxadiazol-5-yl] b-Carboline Derivatives." Journal of Brazilian Chemical Society **21**(2): 288-298.

Silva, A. P. R., C. P. Noronha, et al. (2011). "Estimativa | 2012 Incidência de Câncer no Brasil." Retrieved 30/09, 2012, from <http://www.inca.gov.br/estimativa/2012/estimativa20122111.pdf>.

Tonin, L. T. D., V. A. Barbosa, et al. (2009). "Comparative study of the trypanocidal activity of the methyl 1-nitrophenyl-1,2,3,4-9H-tetrahydro-b-carboline-3-carboxylate derivatives and benznidazole using theoretical calculations and cyclic voltammetry." European Journal of Medicinal Chemistry **44**: 1745 - 1750.

Valdez, R. H., L. T. D. Tonin, et al. (2012). "In Vitro and In Vivo Trypanocidal Synergistic Activity of N-Butyl-1-(4-Dimethylamino)Phenyl-1,2,3,4-Tetrahydro--Carboline-3-Carboxamide Associated with Benznidazole." Antimicrobial Agents and Chemotherapy **56**(1): 507-512.

Weber, K. C. (2008). Modelagem Molecular de Compostos Arilpiperazínicos e suas Interações com o Receptor 5-HT_{1A}. Instituto de Química de São Carlos. São Carlos, Universidade de São Paulo.

Weber, K. C., L. B. Salum, et al. (2010). "Pharmacophore-based 3D QSAR studies on a series of high affinity 5-HT_{1A} receptor ligands." European Journal of Medicinal Chemistry **45**(4): 1508-1514.

Wu, J., M. Zhao, et al. (2009). "Novel N-(3-carboxyl-9-benzyl-b-carboline-1-yl)ethylamino acids: Synthesis, anti-tumor evaluation, intercalating determination, 3D QSAR analysis and docking investigation." European Journal of Medicinal Chemistry **44**: 4153–4161.

Xiao, S., W. Lin, et al. (2001). "Synthesis and biological evaluation of DNA targeting flexible side-chainsubstituted b-carboline derivatives." Bioorganic & Medicinal Chemistry Letters **11**: 437-441.