



IDENTIFICAÇÃO DE ALVOS NA TERAPIA  
ANTI-CANCERÍGENA UTILIZANDO  
TÉCNICAS DE MODELAÇÃO MOLECULAR

Dissertação para obtenção do Grau de Mestre em  
Bioquímica Aplicada

**Maria Inês Jardim de Sousa**

Orientador: Prof. Dr. Miguel Fernandes

Funchal, 2009



# IDENTIFICAÇÃO DE ALVOS NA TERAPIA ANTI-CANCERÍGENA UTILIZANDO TÉCNICAS DE MODELAÇÃO MOLECULAR

Dissertação para obtenção do Grau de Mestre em  
Bioquímica Aplicada

**Maria Inês Jardim de Sousa**

Orientador: Prof. Dr. Miguel Fernandes

Funchal, 2009



## Agradecimentos

Foram várias as pessoas que ao longo da realização do presente trabalho que colaboraram na experiência profissional e nos conhecimentos adquiridos. Sem esta colaboração a realização deste trabalho seria impossível, desta forma gostaria de agradecer ao:

Departamentos de Química e Biologia:

Ao *Prof. Dr. Miguel Fernandes*, na qualidade de orientador pela oportunidade, apoio, colaboração, incentivo, atenção e total disponibilidade que me concedeu ao longo de todo este trabalho.

Ao *Prof. Dr. Miguel Ângelo Carvalho*, Coordenador do Curso de Mestrado em Bioquímica Aplicada, pela organização e coordenação do mestrado.

À *Prof.<sup>a</sup> Dr.<sup>a</sup> Paula Castilho*, Presidente do Departamento de Química da Universidade da Madeira, pelas condições disponibilizadas para a realização deste trabalho.

Aos colegas do *Laboratório de Modelação Molecular da Madeira*, por toda a ajuda e apoio demonstrados durante a realização deste trabalho.

Institutos:

Ao *Instituto Canário de Investigación del Cáncer*, através do *Prof. Dr. J. M. Pádrón*, por ter disponibilizado as estruturas dos compostos estudados e os dados obtidos através dos testes celulares.

Ao *Centro de Química da Madeira*, através do coordenador *Prof. Dr. João Rodrigues*, pelas condições disponibilizadas para a realização deste trabalho.

Ao *Instituto de Emprego da Madeira*, pela oportunidade de realização do estágio profissional nesta área.

Ao *Centro de Ciência e Tecnologia da Madeira*, pela atribuição da bolsa de mestrado, permitindo assim a realização deste trabalho.

Família e amigas/os:

À minha mãe *Maria Sousa*, à minha irmã *Lígia Abreu*, ao meu cunhado *Alberto Abreu*, e aos meus sobrinhos *Daniela e Gonçalo Abreu*, por toda a compreensão, apoio, ajuda, incentivo e alegria que sempre me concederam.

A uma das pessoas mais importantes da minha vida, *Freddy Rodrigues*, por acreditar em mim, por todo o amor, carinho, felicidade, apoio, incentivo e compreensão demonstrada ao longo de todo este percurso.

Aos meus amigos, *Carolina Paulo, Michael Caldeira, Rosa Perestrelo*, entre outros, pela amizade, apoio e compreensão.

---

## Resumo

A realização deste trabalho teve como objectivos principais a identificação dos alvos moleculares, para uma série de compostos de origem natural que possuem actividade anti-cancerígena verificada através de ensaios celulares, a identificação das características estruturais relevantes para a actividade dos compostos e a obtenção de modelos, lineares e não lineares, para relacionar os descritores moleculares com a actividade biológica. Para efectuar estes estudos foram utilizadas diversas metodologias computacionais como o *docking*, a relação quantitativa entre a estrutura e a actividade (QSAR) e as redes neuronais artificiais (ANN).

Para este estudo foram seleccionadas várias enzimas que actuam nas diversas fases do ciclo celular e que são alvos vulgarmente utilizados na terapia anti-cancerígena, e também, enzimas que são alvos na terapia anti-inflamatória e anti-bacteriana.

As interacções entre as enzimas e os compostos com actividade anti-cancerígena demonstrada foram estudadas utilizando o método computacional de *docking* e o programa eHiTS. Os valores de energias de interacção obtidos foram posteriormente corrigidos, de modo a evitar a ocorrência de enviesamentos. Neste âmbito, foi ainda estudada a especificidade dos compostos, foram realizados testes de enriquecimento dos resultados e estabelecidas correlações entre os dados computacionais e dados obtidos através dos testes celulares.

A aplicação de métodos lineares e não lineares para relacionar os descritores moleculares com a actividade anti-cancerígena dos compostos em estudo foi efectuada utilizando os métodos QSAR e ANN, e os programas CODESSA e Statistica, respectivamente. Os descritores moleculares foram calculados recorrendo ao MOPAC (incluído no programa VEGA ZZ 2.2.0.), ao E-Dragon e ao CODESSA.

Através da realização deste trabalho, foi possível verificar que compostos em estudo apresentam uma maior especificidade em relação à cinase AURKA, que actua nas células cancerígenas do que em relação às cinases que actuam quer em células saudáveis, quer em células cancerígenas, tendo sido apresentada uma especificidade de 100 a 300 vezes superior. Os testes de enriquecimento realizados provocaram um enriquecimento de duas a cinquenta vezes. Foram ainda estabelecidas correlações entre os dados computacionais e os resultados experimentais para as diversas enzimas e linhas celulares estudadas.

Relativamente aos resultados referentes à aplicação dos métodos lineares e não lineares, identificou-se algumas das características estruturais relevantes para a actividade anti-cancerígena dos compostos, como o número de ligações triplas, descritores associados com as sombras projectadas nos planos XY, YZ e ZX (estando estes relacionados com o tamanho das moléculas), o momento de inércia A, entre outros.

Através deste trabalho, foi ainda possível determinar que os modelos lineares são mais adequados para relacionar os descritores moleculares com a actividade, em relação às linhas celulares A2780 e T-47D. Para relacionar a estrutura com a actividade para as linhas celulares HL60 e WiDr são mais adequados os modelos não lineares. Relativamente à linha celular SW1573, não foi possível determinar que tipo de modelos, linear ou não linear, é mais adequado para estabelecer este tipo de relações.

## Abstract

The main objectives of this work were the identification of molecular targets for a series of compounds from natural source, which in cellular assays showed anti-cancer activity, the identification of structural features relevant for the compounds' activity and the establishment of linear and nonlinear models to correlate molecular descriptors and biological activity. Several computational methodologies were used to perform these studies such as docking, quantitative structure-activity relationship (QSAR) and artificial neural networks (ANN).

In this study we selected several enzymes, which act in the different cell cycle phases and which are commonly used as targets in anti-cancer therapy, and some enzymes used as targets in anti-inflammatory and anti-bacterial therapy.

Docking methodology and the eHiTS program were used to study the interactions between enzymes and the compounds with anti-cancer activity. The obtained interaction energy values were corrected, in order to avoid some biases. In this work we also studied the specificity of compounds, performed some enrichment tests and established some correlations between computational data and the results obtained by cellular assays.

The application of linear and nonlinear methods to establish a relationship between molecular descriptors and anti-cancer activity of compounds under study was performed using QSAR and ANN methods and the CODESSA and Statistica programs, respectively. Molecular descriptors were calculated using MOPAC (included in VEGA ZZ 2.2.0.), E-Dragon and CODESSA.

In this work we verified that the compounds under study are 100 to 300 times more specific towards AURKA kinase, which act in cancer cells, than the kinases which act either in healthy and cancer cells. The enrichment tests caused an enrichment between two and fifty times, and were also obtained correlations between computational data and experimental results for several enzymes and the cancer cell lines studied.

The obtained results by application of linear and nonlinear methods, identified some structural features relevant to the anti-cancer activity of the compounds, such as the number of triple bonds, the descriptors associated with the shadow projected onto XY, YZ and ZX plane (related with molecule size), the moment of inertia A, among others. In this work was also possible to determine that the linear models are more

appropriate to establish a relationship between molecular descriptors and the activity to A2780 and T-47D cancer cell lines. To correlate the structure with the activity for HL60 and WiDr cancer cell lines are more appropriate nonlinear models. Relatively to the SW1573 cancer cell line no model, linear or nonlinear, stands out when establishing these relationships.

---

**Índice**

Agradecimentos	I
Resumo	III
Abstract	V
Índice	VII
Índice de Figuras	XII
Índice de Tabelas	XX
Lista de Abreviaturas	XXI
<b>Capítulo I – Introdução</b>	<b>1</b>
<b>Capítulo II – Retrospectiva</b>	<b>4</b>
2.1. Cancro	4
2.1.1. Proliferação celular e o seu controlo	6
2.1.2. Desenvolvimento do cancro e a organização dos tecidos	7
2.1.3. Oncogenes e Genes Supressores de Tumores	8
2.1.4. Alterações físicas e químicas no DNA	9
2.1.5. <i>Checkpoints</i> , mutações e cancro	10
2.1.6. Factores que predispõem o desenvolvimento de cancro	11
2.1.7. A apoptose	11
2.1.8. Alterações malignas nas células	13
2.1.8.1. Transformação	14
2.2. Terapias anti-cancerígenas	15
2.2.1. Quimioterapia	16
2.2.2. Alvos Terapêuticos	17
2.3. Mecanismos de acção de fármacos	18
2.3.1. Locais de acção dos fármacos	18
2.3.2. Modo de acção dos fármacos	19
2.3.3. Mecanismos de acção dos fármacos por inibição enzimática	19
2.3.4. Interações entre o fármaco e o receptor	20
2.3.5. Mecanismos dos fármacos anti-cancerígenos	21
2.4. Desenho de fármacos assistido por computador	23
2.4.1. Duas Classes de Problemas	25

---

<b>Capítulo III – Revisão Bibliográfica</b>	26
3.1. <i>Docking</i>	26
3.2. Relação Quantitativa entre a Estrutura e a Actividade (QSAR)	27
3.3. Redes Neurais Artificiais (ANN)	28
<b>Capítulo IV – Metodologia</b>	29
4.1. <i>Docking</i>	29
4.1.1. Requisitos Básicos para a Realização do <i>Docking</i>	29
4.1.2. Aplicações do <i>Docking</i>	30
4.1.3. Programa eHiTS	31
4.1.3.1. eHiTS: o método	33
4.1.3.2. Forma geométrica e características químicas	34
4.1.3.3. Reconstrução e optimização	34
4.1.3.4. Protonação	35
4.1.3.5. Função de <i>Scoring</i>	36
4.1.3.6. Vantagens do eHiTS	36
4.1.4. Realização do <i>Docking</i>	37
4.1.4.1. Enzimas	37
4.1.4.2. Modelação por homologia	39
4.1.4.3. Criação dos modelos por homologia	40
4.1.4.4. Ligandos	41
4.1.4.5. Optimização das condições de <i>docking</i>	41
4.1.4.6. Correção dos Resultados	44
4.1.4.7. Testes de Enriquecimento	45
4.2. Relação Quantitativa entre a Estrutura e a Actividade (QSAR)	46
4.2.1. Descritores Moleculares	48
4.2.2. Aplicações	50
4.2.3. Realização do QSAR	50
4.2.4. Método Heurístico	52
4.3. Redes Neurais Artificiais (ANN)	53
4.3.1. <i>Radial Basis Function</i> (RBF)	55
4.3.2. <i>Multilayer Perceptrons</i> (MLP)	56

---

4.3.4. Aplicações	58
4.3.5. Realização das Redes Neurais Artificiais	58
<b>Capítulo V – Resultados e Discussão</b>	<b>60</b>
5.1. <i>Docking</i>	60
5.1.1. Distribuição de Energia	60
5.1.2. Constante de Dissociação	61
5.1.3. Correção dos Resultados de <i>Docking</i>	63
5.1.4. Cálculo do Erro (MRE)	64
5.1.5. Preferências das Enzimas	65
5.1.6. Estudo da Especificidade	68
5.1.7. Testes de Enriquecimento	71
5.1.8. Correlação entre os Dados Computacionais e Experimentais	72
5.2. Relação Quantitativa entre a Estrutura e a Actividade (QSAR)	83
5.2.1. Série 1	83
5.2.1.1. Linha Celular do Cancro do Ovário (A2780)	83
5.2.1.2. Linha Celular do Cancro do Pulmão (SW1573)	88
5.2.1.3. Linha Celular do Carcinoma Mamário (T-47D)	92
5.2.2. Série 7	96
5.2.2.1. Linha Celular da Leucemia (HL60)	96
5.2.2.2. Linha Celular do Cancro do Ovário (A2780)	99
5.2.3. Série 9	101
5.2.3.1. Linha Celular do Cancro do Ovário (A2780)	101
5.2.3.2. Linha Celular do Cancro do Pulmão (SW1573)	103
5.2.3.3. Linha Celular do Cancro do Cólon (WiDr)	107
5.2.4. Série 12	109
5.2.4.1. Linha Celular do Cancro do Ovário (A2780)	109
5.2.4.2. Linha Celular do Carcinoma Mamário (T-47D)	111
5.3. Redes Neurais Artificiais (ANN)	113
5.3.1. Série 1	113
5.3.1.1. Linha Celular do Cancro do Ovário (A2780)	113
5.3.1.2. Linha Celular do Cancro do Pulmão (SW1573)	118
5.3.1.3. Linha Celular do Carcinoma Mamário (T-47D)	121

---

5.3.2. Série 7	123
5.3.2.1. Linha Celular da Leucemia (HL60)	123
5.3.3. Série 9	126
5.3.3.1. Linha Celular do Cancro do Ovário (A2780)	126
5.3.3.2. Linha Celular do Cancro do Pulmão (SW1573)	128
5.3.3.3. Linha Celular do Cancro do Cólon (WiDr)	131
5.3.4. Série 12	133
5.3.4.2. Linha Celular do Carcinoma Mamário (T-47D)	133
<b>Capítulo VI – Conclusões e Perspectivas Futuras</b>	
Conclusões	135
Perspectivas Futuras	138
<b>Bibliografia</b>	139
<b>Anexos</b> (encontram-se em formato electrónico)	
Anexo A – Optimização das Condições de <i>Docking</i>	145
Anexo B – Distribuição de Energias	160
Anexo C – Cálculo do Erro (MRE)	185
Anexo D – Testes de Enriquecimento	190
Anexo E – Correlações entre os Dados Computacionais e Experimentais	193
E.1. Linha celular do cancro do ovário (A2780)	193
E.2. Linha celular da leucemia (HL60)	201
E.3. Linha celular do cancro do pulmão (SW1573)	208
E.4. Linha celular do carcinoma mamário (T-47D)	213
E.5. Linha celular do cancro do cólon (WiDr)	219
Anexo F – Relação Quantitativa entre a Estrutura e a Actividade (QSAR) e Redes Neurais Artificiais (ANN)	224
F.1. Série 1	224
F.1.1. QSAR	225
F.1.2. ANN	227
F.2. Série 7	229
F.2.1. QSAR	231

---

F.2.2. ANN	232
F.3. Série 9	234
F.3.1. QSAR	235
F.3.2. ANN	237
F.4. Série 12	239
F.4.1. QSAR	240
F.4.2. ANN	241
F.5. Tabelas das Equações de QSAR	242
F.5.1. Série 1	242
F.5.2. Série 7	245
F.5.3. Série 9	245
F.5.4. Série 12	247
F.6. Tabelas de Estatística dos modelos de QSAR	248
F.6.1. Série 1	248
F.6.2. Série 7	248
F.6.3. Série 9	249
F.6.4. Série 12	249

---

## Índice de Figuras

Figura 1: Esquematização das diversas fases que ocorrem no desenvolvimento do cancro.	5
Figura 2: Esquema representativo dos processos de proliferação celular e da apoptose.	13
Figura 3: Influência da variação do número de conformações no tempo necessário para a realização do <i>docking</i> para a proteína Tubulina.	42
Figura 4: Influência da variação da margem no tempo necessário para a realização do <i>docking</i> para a proteína Tubulina.	43
Figura 5: Influência da utilização ou não da opção <i>fast</i> no tempo necessário para a realização do <i>docking</i> para a proteína Tubulina.	43
Figura 6: Influência da utilização ou não da opção <i>fast</i> nos valores obtidos pelo <i>docking</i> para a proteína Tubulina.	43
Figura 7: Exemplos de estruturas de redes neuronais artificiais, <i>Multilayer Perceptrons</i> (à esquerda) e <i>Kohonen self-organising maps</i> (à direita).	54
Figura 8: Estrutura típica de uma rede neuronal do tipo RBF.	55
Figura 9: Estrutura de uma rede neuronal do tipo MLP com uma camada “escondida”.	57
Figura 10: Distribuição das energias obtidas para a enzima CDC25C.	60
Figura 11: Distribuição das energias obtidas para a enzima ERK1.	60
Figura 12: Representação gráfica do número de compostos que apresentam valores de constante de dissociação de diferentes ordens de grandeza para a enzima GSK-3B.	62
Figura 13: Representação gráfica do número de compostos que apresentam valores de constante de dissociação de diferentes ordens de grandeza para a enzima ERK2.	62
Figura 14: Distribuição da energia antes e após a correcção para a enzima AKT1.	63
Figura 15: Distribuição da energia antes e após a correcção para a enzima CHK1.	63
Figura 16: Distribuição da energia antes e após a correcção para a enzima CDC2.	64
Figura 17: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correcção.	65
Figura 18: Estrutura base dos compostos com os quais a enzima AMD1 interage preferencialmente.	66
Figura 19: Estrutura base dos compostos com os quais a enzima WEE1 interage	66

preferencialmente.	
Figura 20: Estrutura base dos compostos com os quais a enzima CDC25B interage preferencialmente.	67
Figura 21: Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para VJH-6.	69
Figura 22: Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para VJH-8.	69
Figura 23: Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para NHR-38.	69
Figura 24: Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para NHR-47.	70
Figura 25: Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para VJH-21.	70
Figura 26: Teste de Enriquecimento realizado para a enzima CDC7.	71
Figura 27: Teste de Enriquecimento realizado para a enzima CDK4.	72
Figura 28: Correlação entre os dados computacionais e experimentais para a enzima CDC25C para a linha celular A2780.	74
Figura 29: Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular A2780.	74
Figura 30: Correlação entre os dados computacionais e experimentais para a enzima PAK1 para a linha celular HL60.	75
Figura 31: Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular HL60.	75
Figura 32: Correlação entre os dados computacionais e experimentais para a enzima 5-LOX para a linha celular SW1573.	76
Figura 33: Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular SW1573.	77
Figura 34: Correlação entre os dados computacionais e experimentais para a enzima NA para a linha celular T-47D.	77
Figura 35: Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular T-47D.	78
Figura 36: Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular WiDr.	79

Figura 37: Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular WiDr.	79
Figura 38: Correlação entre os dados computacionais e experimentais para a enzima HO-1 para a linha celular A2780.	80
Figura 39: Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular HL60.	81
Figura 40: Correlação entre os dados computacionais e experimentais para a proteína CAP-G para a linha celular SW1573.	81
Figura 41: Correlação entre os dados computacionais e experimentais para a enzima NEK2 para a linha celular T-47D.	81
Figura 42: Correlação entre os dados computacionais e experimentais para a proteína MAX para a linha celular WiDr.	82
Figura 43: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $IC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	83
Figura 44: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $GI_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	84
Figura 45: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $LC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $LC_{50}$ , em função dos descritores mais e menos significativos.	85
Figura 46: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	87
Figura 47: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $IC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	89

Figura 48: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $GI_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	90
Figura 49: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros da actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	91
Figura 50: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $IC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	93
Figura 51: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro TGI para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro TGI, em função dos descritores mais e menos significativos.	94
Figura 52: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	95
Figura 53: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $GI_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	96
Figura 54: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	98
Figura 55: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média	99

dos parâmetros de actividade, em função dos descritores mais e menos significativos.	
Figura 56: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	102
Figura 57: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $IC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	104
Figura 58: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $LC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $LC_{50}$ , em função dos descritores mais e menos significativos.	105
Figura 59: a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	106
Figura 60: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $GI_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	108
Figura 61: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $GI_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	110
Figura 62: a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $IC_{50}$ para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	111
Figura 63: a) Representação esquemática da rede neuronal (3-8-1), b)	114

Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $IC_{50}$ para os conjuntos de treino e de teste.	
Figura 64: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	114
Figura 65: a) Representação esquemática da rede neuronal (3-7-1), b)	
Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro $LC_{50}$ para os conjuntos de treino e de teste.	115
Figura 66: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $LC_{50}$ , em função dos descritores mais e menos significativos.	115
Figura 67: a) Representação esquemática da rede neuronal (2-6-1), b)	
Representação gráfica das correlações entre os valores experimentais e previstos da média dos vários parâmetros de actividade para os conjuntos de treino e de teste.	117
Figura 68: Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	117
Figura 69: a) Representação esquemática da rede neuronal (4-3-1), b)	
Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro $GI_{50}$ para os conjuntos de treino e de teste.	119
Figura 70: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	119
Figura 71: a) Representação esquemática da rede neuronal (3-8-1), b)	
Representação gráfica das correlações entre os valores experimentais e previstos para a média dos diversos parâmetros de actividade para os conjuntos de treino e de teste.	120
Figura 72: Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	120
Figura 73: a) Representação esquemática da rede neuronal (2-3-1), b)	
Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro de actividade $IC_{50}$ para os conjuntos de treino e de teste.	122
Figura 74: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	122

Figura 75: a) Representação esquemática da rede neuronal (3-5-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro $GI_{50}$ para os conjuntos de treino e de teste.	124
Figura 76: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	124
Figura 77: a) Representação esquemática da rede neuronal (3-6-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para a média dos parâmetros de actividade para os conjuntos de treino e de teste.	125
Figura 78: Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	125
Figura 79: a) Representação esquemática da rede neuronal (3-5-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para a média dos parâmetros de actividade para os conjuntos de treino e de teste.	127
Figura 80: Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.	127
Figura 81: a) Representação esquemática da rede neuronal (2-2-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro $IC_{50}$ para os conjuntos de treino e de teste.	128
Figura 82: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $IC_{50}$ , em função dos descritores mais e menos significativos.	129
Figura 83: a) Representação esquemática da rede neuronal (2-4-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro $LC_{50}$ para os conjuntos de treino e de teste.	130
Figura 84: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $LC_{50}$ , em função dos descritores mais e menos significativos.	130
Figura 85: a) Representação esquemática da rede neuronal (3-8-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro $GI_{50}$ para os conjuntos de treino e de teste.	132
Figura 86: Gráfico de superfície dos valores experimentais e previstos para o parâmetro $GI_{50}$ , em função dos descritores mais e menos significativos.	132
Figura 87: a) Representação esquemática da rede neuronal (2-5-1), b)	133

Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro  $IC_{50}$  para os conjuntos de treino e de teste.

Figura 88: Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

134

---

## Índice de Tabelas

Tabela 1: Enzimas utilizadas para o <i>docking</i> , sendo alvos utilizados na terapia anti-cancerígena.	38
Tabela 2: Enzimas utilizadas para o <i>docking</i> , sendo alvos da terapia anti-inflamatória.	38
Tabela 3: Enzimas utilizadas para o <i>docking</i> , sendo alvos da terapia anti-bacteriana.	39
Tabela 4: Substituintes para a estrutura base representada na figura 18.	66
Tabela 5: Substituintes para a estrutura base representada na figura 19.	67
Tabela 6: Substituintes para a estrutura base representada na figura 20.	67
Tabela 7: Cinases utilizadas no estudo da especificidade e respectiva família.	68
Tabela 8: Substituintes para a estrutura base representada na figura 29.	74
Tabela 9: Substituintes para a estrutura base representada na figura 31.	76
Tabela 10: Substituintes para a estrutura base representada na figura 33.	77
Tabela 11: Substituintes para a estrutura base representada na figura 35.	78
Tabela 12: Substituintes para a estrutura base representada na figura 37.	80

## Lista de Abreviaturas

5-LOX: 5-Lipoxigenase.

A2780: Linha celular do cancro do ovário.

AKT1: Cinase proteica de serina/treonina alfa-RAC.

AMD1: Descarboxilase da adenosilmetionina 1.

AMMP: *Molecular Mechanics Program*.

ANN: Redes neuronais artificiais.

Asp: Complemento do Componente C3.

AURKA: Cinase da Aurora A.

AURKB: Cinase da Aurora B.

AZT: Zidovudina.

CADD: Desenho de fármacos assistido por computador.

CAMD: Desenho molecular assistido por computador.

CAP-G: Proteína de *capping*.

CCNH: Ciclina H.

CDC2: Proteína de controlo da divisão celular 2.

CDC25B: Ciclo de divisão celular 25 homólogo B.

CDC25C: Ciclo de divisão celular 25 homólogo C.

CDC42: Ciclo de divisão celular 42.

CDC6: Ciclo de divisão celular 6.

CDC7: Ciclo de divisão celular 7.

CDH1: Caderina 1.

CDK2: Cinase dependente da ciclina 2.

CDK4: Cinase dependente da ciclina 4.

CDK7: Cinase dependente da ciclina 7.

CENP-A: Proteína do centrómero A.

CHK1: Cinase proteica de serina/treonina Chk1.

CHK2: Cinase proteica de serina/treonina Chk2.

COX: Ciclooxygenase.

DHFR: Redutase do Dihidrofolato.

DHPS: Sintase do Dihidropteroato.

DNA: Ácido desoxirribonucleico.

DP: Desmoplaquina.

DUD: *Directory of Useful Decoys*.

E2F: Factor de transcrição E2F1.

EF-G: Factor de alongamento G.

Eg5: Membro da família da cinesina 11.

eHiTS: *electronic High Throughput Screening*.

ERK1: Cinase proteica activada por mitogénio 3.

ERK2: Cinase proteica activada por mitogénio 1.

FADD: *FAS-associated death domain protein*.

FNTB: Farnesiltransferase.

GFRs: Receptores de factores de crescimento.

GI<sub>50</sub>: Concentração necessária para a inibição de 50% do crescimento da amostra.

GOLD: *Genetic Optimization for Ligand Docking*.

GSK-3B: Cinase da sintase do glicogénio 3 beta.

GTP: Proteína de ligação ao GTP 7.

HER: Receptor de cinase proteica de tirosina erb.

HL60: Linha celular da leucemia.

HO-1: Oxigenase do heme 1.

IC<sub>50</sub>: Concentração necessária para a inibição de 50% da amostra.

IMPDH: Desidrogenase do inositol-5-monofosfato.

iNOS: Sintase do óxido nítrico induzível.

K<sub>D</sub>: Constante de dissociação.

KITLG: Ligando KIT.

LC<sub>50</sub>: Concentração necessária para induzir a morte de 50% da amostra.

M2TM: Proteína M2.

MAD2: Proteína de interacção MAX 1.

MAP2K1: Cinase da cinase proteica activada por mitogénio 1.

MASC: *Multiple Active Site Corrections for Docking and Virtual Screening*.

MAX: Proteína Max.

MCM2: Factor para a replicação do DNA MCM2.

MEK: Cinase proteica activada por mitogénio.

mPGES-1: Sintase da prostaglandina E.

MRE: *Mean Ranking Error*.

MYC: Proteína proto-oncogene Myc.

NA: Neuraminidase.

NEK2: Cinase proteica da serina/treonina Nek2.

NMR: Ressonância Magnética Nuclear.

p107: Proteína associada ao retinoblastoma 107 kDa.

p21CIP1: Inibidor da cinase dependente da ciclina 1.

p27 KIP1: Inibidor da cinase dependente da ciclina p27.

p53: Supressor de tumor p53.

PAK1: Cinase proteica da serina/treonina PAK1.

PDB: Base de dados das proteínas.

PIK3CG: Fosfoinositida-3-cinase polipéptido gama.

PIN1: Rotamase Pin1.

PKA: Cinase proteica alfa.

Plk1: Cinase proteica de serina/treonina Plk1.

QSAR: Relação quantitativa entre a estrutura e a actividade.

RMSE: *Root Mean Squared Error*.

RNA: Ácido ribonucleico.

RNAP: Polimerase do RNA.

ROCK: Cinase proteica associada a Rho.

SEE: *Standard Error of Estimate*.

SKP1: Proteína associada à cinase da fase S1.

SV40: Simian Vírus 40.

SW1573: Linha celular do cancro do pulmão.

T-47D: Linha celular do carcinoma mamário.

TGI: Concentração necessária para a inibição total do crescimento.

TNF: Factores para a necrose de tumores.

TOP1: Topoisomerase I.

TOP2A: Topoisomerase IIA.

TOPOII: Topoisomerase II.

TOPOIV: Topoisomerase IV.

TRAIL: *TNF-related apoptosis inducing ligand*.

VIF: *Variance Inflation Factor*.

WEE1: Cinase proteica Wee1.

WiDr: Linha celular do cancro do cólon.

$\Delta G^0$ : Energia livre de Gibbs padrão.

# CAPÍTULO I



# INTRODUÇÃO

## Capítulo I - Introdução

A obtenção de fármacos utilizados no tratamento das várias doenças passa por duas fases principais, a fase da descoberta e a fase do desenvolvimento. Estas fases asseguram que apenas os produtos que são seguros e eficazes sejam comercializados. Neste contexto, o estudo realizado neste trabalho insere-se na fase de desenvolvimento de potenciais fármacos. <sup>[1]</sup>

Este trabalho teve como ponto de partida o estudo, a nível molecular, da actividade anti-cancerígena de diversos compostos de origem natural que já tinham sido sintetizados anteriormente e que, aquando da realização de testes celulares, apresentaram actividade anti-cancerígena. Estes testes celulares foram realizados para diversas linhas celulares, sendo elas, HL60 (leucemia), A2780 (cancro no ovário), SW1573 (cancro no pulmão), T-47D (carcinoma mamário), WiDr (cancro no cólon).

Os principais objectivos deste estudo são a identificação dos alvos moleculares para estes compostos, a determinação dos seus mecanismos de acção e a obtenção de modelos lineares e não lineares para estabelecer relações quantitativas entre a estrutura e a actividade biológica dos mesmos.

O primeiro objectivo prende-se com o facto dos compostos apresentarem actividade anti-cancerígena, mas ainda serem desconhecidos os alvos moleculares dos mesmos. Os métodos de QSAR lineares e não lineares aplicados têm como objectivo identificar que características estruturais dos compostos são relevantes para as suas actividades biológicas.

Para a identificação dos alvos moleculares dos compostos em estudo, foram seleccionadas diversas enzimas que são alvos utilizados no estudo da actividade anti-cancerígena e também algumas enzimas que são alvos terapêuticos no estudo de anti-inflamatórios e de anti-bacterianos.

Neste trabalho foram utilizados diversos métodos computacionais, o *docking*, o QSAR e as ANN. O *docking* permite encontrar o melhor ajuste possível entre as enzimas e os compostos em estudo, prever a posição e a orientação do ligando quando este estabelece uma ligação com a enzima e também prever a energia do complexo proteína-ligando. Quanto aos métodos de QSAR e ANN, estes correlacionam os descritores moleculares com a actividade biológica dos compostos utilizando modelos

lineares e não lineares. Com base nos modelos obtidos, é possível prever a actividade biológica de outros compostos que pertençam à mesma série congénica. <sup>[2][3][4]</sup>

A identificação dos alvos moleculares e das características estruturais responsáveis pela actividade biológica permite a passagem para a fase seguinte do processo de produção de fármacos. A fase de desenvolvimento é constituída por diversos passos, entre eles a modificação de compostos de forma a melhorar as suas propriedades, o estudo das propriedades físico-químicas, da toxicidade, da solubilidade e da estabilidade, tudo de modo a que os fármacos que chegam à fase final de todo este processo, possam ser comercializados apresentando-se seguros em relação à saúde das pessoas e eficazes no tratamento das doenças. <sup>[1]</sup>

Existem diversos programas que executam o *docking*, como o AutoDock, o AMMP, o GOLD, entre outros. No entanto, o programa escolhido para a realização deste estudo foi o programa eHiTS, uma vez que este estabelece boas previsões, é de fácil utilização, permite um bom controlo das condições em que é realizado o *docking* e também porque apresenta uma elevada velocidade de execução. Este programa fornece as coordenadas tridimensionais das conformações quando as moléculas de ligando interagem com o centro activo do receptor e a energia de interacção de cada conformação. <sup>[5]</sup>

A abordagem efectuada para a identificação de alvos moleculares constitui uma inovação neste tipo de estudos, uma vez que vulgarmente o ponto de partida é um alvo molecular específico, sendo depois desenvolvidos compostos para esse alvo. Este estudo foi desenvolvido da forma inversa, procedendo-se à identificação dos alvos moleculares para compostos que demonstraram em testes celulares, actividade anti-cancerígena.

Com base nos resultados obtidos por *docking*, é possível determinar com que compostos as enzimas interagem preferencialmente e identificar as matrizes (*scaffolds*) dos mesmos. Também é possível estudar a especificidade dos compostos em relação às enzimas que actuam especificamente em células cancerígenas e aquelas que estão presentes quer em células saudáveis, quer em células cancerígenas. Os dados obtidos computacionalmente podem ainda ser correlacionados com os dados provenientes dos testes celulares, sendo possível através dessas correlações estabelecer quais os compostos que poderão ser utilizados no tratamento de cada tipo de cancro estudado.

Quanto aos estudos de QSAR lineares e não lineares efectuados, estes permitiram a identificação de características estruturais relevantes para a actividade

anti-cancerígena dos compostos em estudo com base nos descritores moleculares envolvidos nos modelos obtidos. Também foi possível, determinar que modelos, lineares ou não lineares, são mais adequados para relacionar os descritores moleculares com a actividade biológica para cada tipo de cancro estudado neste trabalho.

Mediante a realização deste estudo foi possível estabelecer com que matrizes dos compostos estudados, as enzimas testadas interagem preferencialmente. Também foi possível, através do estudo da especificidade, determinar que compostos são mais específicos em relação a enzimas que actuam apenas em células cancerígenas do que em relação a enzimas que actuam também em células normais.

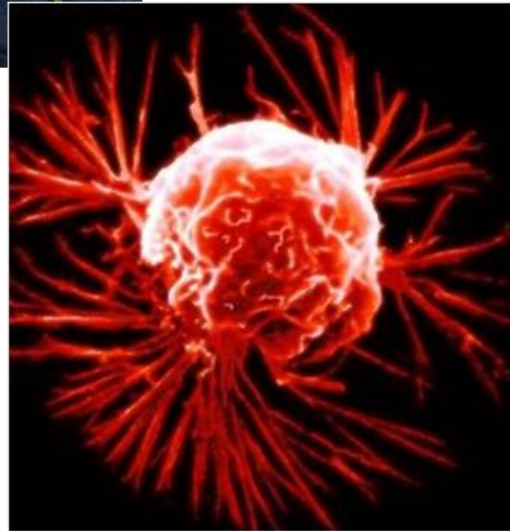
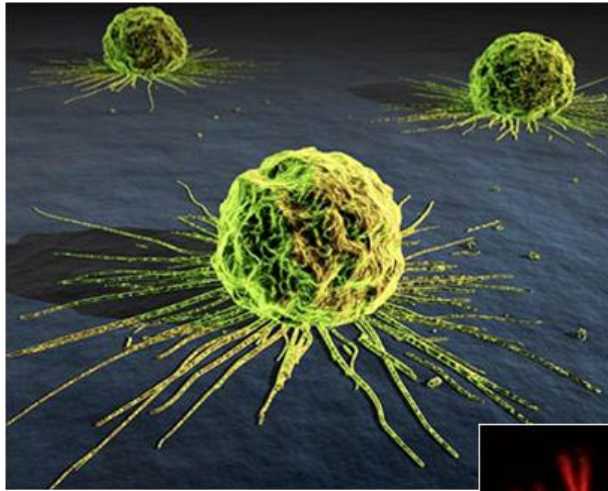
Os resultados obtidos com a realização deste trabalho demonstram que as estratégias utilizadas permitem identificar potenciais alvos moleculares para os compostos estudados e as características estruturais relevantes para a actividade anti-cancerígena apresentada pelos mesmos. No entanto, para determinar a qualidade das previsões realizadas será sempre necessária a confirmação experimental. \*

\* Divulgação dos resultados obtidos neste trabalho nos seguintes encontros:

- 4<sup>th</sup> Materials Group Meeting/ CQM, 30<sup>th</sup> January 2009, Funchal, Madeira.

-3<sup>rd</sup> European Conference on Chemistry for Life Sciences: *Linking Chemistry with Biological Activity*, 2<sup>th</sup> – 5<sup>th</sup> September 2009, Frankfurt am Main, Germany.

# CAPÍTULO II



**RETROSPECTIVA**

## Capítulo II - Retrospectiva

### 2.1. Cancro

O termo *cancro* designa um grupo de doenças nas quais as células crescem anormalmente e formam um tumor maligno. As células malignas podem invadir os tecidos vizinhos e alastrar-se para outras partes do corpo onde estabelecem áreas secundárias de crescimento. Este padrão de crescimento anormal resulta de mutações que ocorrem nos genes que regulam a proliferação, diferenciação e sobrevivência das células nos organismos multicelulares. Devido a estas alterações genéticas, as células cancerígenas não respondem aos sinais que controlam o crescimento das células normais. <sup>[6]</sup>

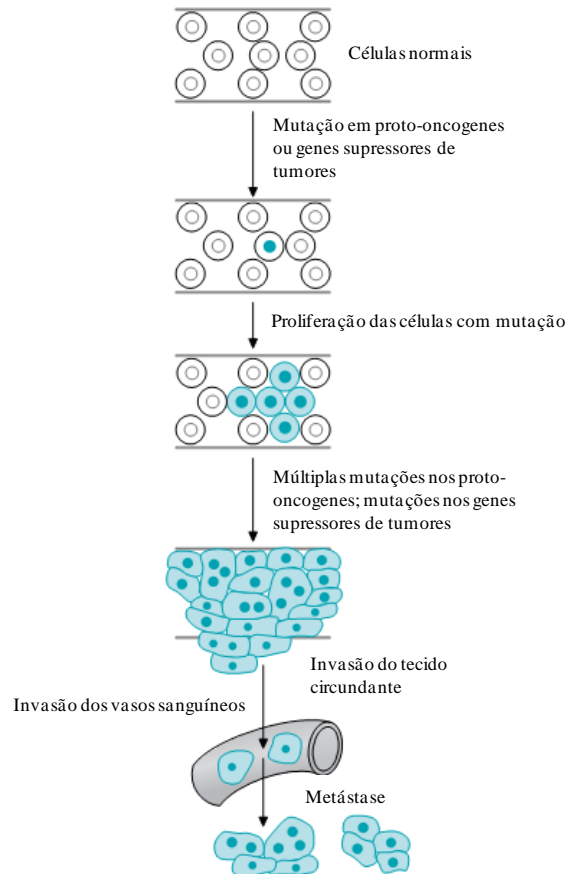
A carcinogénese é o processo que origina mutações genéticas induzidas por agentes físicos ou químicos, podendo este ser dividido em três fases, a iniciação, a promoção e a progressão. A iniciação envolve uma alteração genética irreversível, que é geralmente uma mutação que ocorre num único gene. A promoção está geralmente associada com o aumento da proliferação das células iniciais, aumentando assim a sua população. A progressão é a acumulação de mais mutações genéticas que levam à aquisição de fenótipo maligno ou invasivo. <sup>[7]</sup>

As células normais presentes no organismo respondem a determinados sinais, como a inibição por contacto, que faz com que as células parem de se proliferar. Por sua vez, as células infectadas com cancro não necessitam de sinais estimuladores de crescimento e são resistentes aos sinais de inibição do crescimento. No entanto, existem diversos tipos de células que constituem uma exceção a esta regra, como as células da medula óssea, do epitélio intestinal e as células tumorais. <sup>[6][8]</sup>

As células cancerígenas são também resistentes à apoptose, que é a morte programada das células e que é realizada para que as células com danos indesejáveis e irreparáveis sejam autodestruídas. Estas células cancerígenas possuem uma capacidade de proliferação infinita e não morrem, podendo também crescer independentemente do suporte estrutural. <sup>[6]</sup>

Uma única célula que se divida anormalmente poderá, eventualmente, formar uma massa designada de tumor. Os tumores desenvolvem a capacidade de formar novos vasos sanguíneos. Assim, os tumores podem possuir o seu próprio abastecimento de

sangue para conduzir oxigénio e nutrientes. As células cancerígenas podem também alastrar-se a outras partes do organismo, separando-se da massa de crescimento do tumor e movimentando-se através do sangue ou linfa para outros órgãos, onde as células cancerígenas também irão crescer. Na figura seguinte estão representadas as várias fases que ocorrem no desenvolvimento de cancro. [6]



**Figura 1:** Esquemática das diversas fases que ocorrem no desenvolvimento do cancro. [6]

Na medicina, a distinção entre tumores malignos e benignos é efectuada com base no facto dos tumores benignos serem constituídos por células que crescem lentamente e que são muito diferenciadas. Os tumores malignos apresentam um crescimento rápido e invasivo e tendem a formar metástases. [8]

Algumas das células existentes no organismo humano são específicas de determinados tecidos (como as células epiteliais da bexiga, pulmões, mama, pele, entre outros), contribuindo para cerca de 70% dos cancros. Qualquer célula possui o potencial para se transformar em célula cancerígena e pode originar o desenvolvimento de carcinoma. [9]

A transformação de uma célula normal numa célula cancerígena inicia-se com lesões no DNA (alterações na base ou quebra da cadeia) causada por compostos químicos cancerígenos, radiação UV, vírus ou erros de replicação. As mutações resultam do DNA danificado se este não for reparado adequadamente, ou se não for reparado antes de ocorrer a replicação. A mutação que pode levar à transformação também pode ser herdada. Quando uma célula com uma mutação prolifera, origina um elevado número de células contendo esta mutação, podendo as células adquirir uma segunda mutação importante para o controlo do crescimento ou da morte celular. Em cada expansão, a probabilidade de ocorrer outra mutação aumenta. Como as mutações se acumulam nos genes que controlam a proliferação, as mutações subsequentes ocorrem mais rapidamente até as células adquirirem múltiplas mutações, sendo estas necessárias para a transformação total. <sup>[6]</sup>

As mutações ocorrem nos genes que regulam a proliferação e diferenciação celular (proto-oncogenes) e também nos genes que suprimem o crescimento (genes supressores de tumores), fazendo com que as células danificadas sejam considerados alvos irreparáveis, não podendo o seu DNA ser reparado e sendo estas células mortas através da apoptose. O cancro é causado pela acumulação de mutações nos genes que estão envolvidos no crescimento e na diferenciação das células normais. Estas mutações originam células cancerígenas capazes de se proliferar irregular, autónoma e infinitamente. <sup>[6]</sup>

Através da análise dos genes envolvidos no desenvolvimento do cancro, foi verificado que um determinado tipo de cancro pode surgir de diversas maneiras. Deste modo, os tratamentos que são eficazes para um paciente com um determinado tipo de cancro podem não ser eficazes noutro paciente com o mesmo tipo de cancro, esta situação deve-se às diferenças na base molecular da doença de cada indivíduo. Futuramente, é necessário que sejam identificadas as lesões moleculares envolvidas na doença e desenvolver os tratamentos apropriados. <sup>[6]</sup>

### **2.1.1. Proliferação celular e o seu controlo**

O controlo da divisão celular é um processo importante em todos os tecidos do organismo, sendo particularmente relevante em tecidos que apresentem uma rápida

auto-renovação, uma vez que nestes casos tem de existir um equilíbrio entre a proliferação e a perda celular. A proliferação é regulada por uma complexa rede de sinais e mensagens que envolvem factores de crescimento, citocinas e hormonas. Estas mensagens podem ser produzidas pelas próprias células (regulação autócrina), pelas células vizinhas de tipos celulares semelhantes ou não relacionados (regulação parácrina) e por hormonas de circulação (regulação do sistema endócrino).<sup>[10]</sup>

Algumas das redes de sinais que controlam a homeostase do tecido podem prevenir a superprodução de células ou podem parar o ciclo celular se a célula estiver danificada. Nos *checkpoints* do ciclo celular, o DNA danificado pode ser reparado ou as células podem cometer suicídio através do processo da apoptose, caso a danificação seja significativa.<sup>[10]</sup>

As primeiras fases da formação do cancro estão associadas com o mau funcionamento dos mecanismos de controlo da divisão celular, estando o equilíbrio entre a proliferação e a perda celular perturbado ou desregulado. No desenvolvimento do cancro, o equilíbrio sofre um desvio que favorece a proliferação, fazendo com que o tecido se expanda de uma forma progressiva e descontrolada, alterando a sua estrutura e a sua função.<sup>[10]</sup>

### **2.1.2. Desenvolvimento do cancro e a organização dos tecidos**

O processo de desenvolvimento do cancro é constituído por diversas fases. Numa primeira fase, a célula sofre uma mutação que pode ou não influenciar o seu comportamento, podendo ficar mais susceptível a subseqüentes mutações e, acumular gradualmente lesões suficientes, perturbando o controlo normal ou os mecanismos de “paragem”. Como esta doença se desenvolve devido à acumulação gradual de mutações, esta aparece normalmente em pessoas de idade avançada, com excepção para os casos em que a mutação primária é uma desordem hereditária.<sup>[10]</sup>

Quando um tecido sofre este processo, é possível observar histologicamente a hiperplasia ou o crescimento celular excessivo. No entanto, a hiperplasia pode ser benigna, uma vez que a regeneração de tecidos como resposta a ferimentos constitui uma forma de hiperplasia. Nos adultos, a hiperplasia apenas pode ocorrer nos tecidos que proliferam, sendo que a maioria dos cancros aparece em tecidos que se renovam

rapidamente ou que estão continuamente a renovar-se. Em cada caso, existe um mau funcionamento na homeostase celular e a produção celular excede a perda das células.<sup>[10]</sup>

O cancro pode ser também uma das doenças das células estaminais. Ao ocorrer uma expansão do número de células estaminais pode ser originada uma hiperplasia. Geralmente, quando ocorre este tipo de expansão, o tecido detecta-a e o excesso de células estaminais é removido através de apoptose, se esta remoção não ocorrer, a produção de células irá aumentar drasticamente. As mutações que ocorrem nas células estaminais podem ser originadas através de três passos vitais, a regulação da divisão celular na população renovada (retenção), a reparação do DNA (o DNA normal não é mantido) e as interacções com o ambiente extracelular (células ou matriz). Em conjunto, estes subvertem o processo normal de diferenciação, permitindo o crescimento ilimitado dos tecidos sem o acompanhamento dos níveis da morte celular, seguida da invasão e da metástase noutras zonas dos tecidos.<sup>[10]</sup>

### **2.1.3. Oncogenes e Genes Supressores de Tumores**

As mutações podem causar transformações em duas classes de genes, os oncogenes e os genes supressores de tumores. Os oncogenes são formas de genes que codificam as proteínas que regulam o ciclo celular mas que sofreram mutações. Inicialmente estes genes foram identificados como genes originados por vírus que causam transformação nas células alvo. A principal classe de oncogenes virais possui uma parte celular que está envolvida nas funções das células normais. Este tipo de genes deriva dos proto-oncogenes que codificam as proteínas reguladoras do crescimento, apresentando características muito semelhantes às destes genes. Os genes celulares são designados de proto-oncogenes, nalguns casos a sua mutação ou activação anormal na célula transformando-o num oncogene está associada com a formação de tumores.<sup>[6][11][12]</sup>

Durante as infecções produzidas por vírus, a sequência de DNA do proto-oncogene é por vezes copiada pelo vírus e incorporada no seu genoma, podendo o gene tornar-se defeituoso. Quando o oncogene viral é expresso na célula hospedeira durante a

subsequente infecção, a proteína anormal produzida interfere com a regulação do crescimento normal da célula, resultando por vezes em tumores. <sup>[6][11][12]</sup>

Os proto-oncogenes podem transformar-se em oncogenes sem um intermediário viral. Os rearranjos nos cromossomas, os agentes químicos e a radiação são alguns dos factores que podem provocar mutações oncogénicas. Este tipo de mutação é geneticamente dominante, necessitando apenas que um dos cromossomas do par contenha um gene deficiente para transmitir o sinal para a divisão celular anormal, o que levará posteriormente ao desenvolvimento de tumores. A produção de um oncogene representa um “ganho de função”, este processo pode envolver uma mutação na proteína, ou uma activação constitutiva, a sobre-expressão, ou falha para parar a expressão na altura apropriada. <sup>[6][11][12]</sup>

Os oncogenes codificam proteínas, factores de crescimento, proteínas transmembranares (receptores), proteínas citoplasmáticas (proteínas G e cinases das proteínas) e factores de transcrição nuclear que controlam a expressão dos genes essenciais para a divisão celular. <sup>[6][11]</sup>

Os genes supressores de tumores codificam proteínas que restringem a divisão celular. A ocorrência de uma mutação em um ou mais destes genes pode provocar o desenvolvimento de tumores. Os tumores formam-se apenas se ambos os cromossomas do par contêm o gene deficiente. Os genes supressores de tumores são detectados através de deleções, ou outras mutações inactivantes, que são tumorigénicas. As mutações representam uma “perda de função” nos genes que restringem o ciclo ou o crescimento celular, a ausência de restrições é tumorigénica. <sup>[6][11][12]</sup>

As mutações nos oncogenes e nos genes supressores de tumores não são totalmente responsáveis pela formação de tumores, uma vez que em determinados tipos de cancro, para que uma célula normal progrida para uma célula cancerígena é necessário que sejam acumuladas várias mutações. <sup>[6][11]</sup>

#### **2.1.4. Alterações físicas e químicas no DNA**

Para que o cancro se desenvolva é fundamental que ocorram alterações na estrutura química do DNA ou na sequência das bases presentes nos genes. A função do DNA depende da presença de diversos grupos químicos polares nas bases do DNA que

formam ligações de hidrogénio entre cadeias de DNA ou participam noutras reacções químicas. Os átomos de oxigénio e de azoto nas bases do DNA são alvos para uma variedade de ataques electrófilos. <sup>[6]</sup>

Os compostos químicos cancerígenos (compostos que podem causar mutações) que existem no meio ambiente e que são ingeridos através da alimentação são geralmente compostos lipófilos estáveis, que têm de ser activados pelo metabolismo do organismo para reagir com o DNA. Muitos agentes de quimioterapia, que são concebidos para parar a proliferação das células através da interacção com o DNA, podem também actuar como cancerígenos e causar novas mutações e tumores, enquanto erradicam os tumores antigos. As alterações estruturais no DNA também ocorrem por exposição a radiação UV. <sup>[6]</sup>

### **2.1.5. Checkpoints, mutações e cancro**

O cancro demora um longo período de tempo a se desenvolver nos seres humanos devido às múltiplas alterações genéticas que é necessário ocorrer para que as células normais se possam transformar em células malignas. Uma única mudança num oncogene ou num gene supressor de tumor numa célula individual, não é suficiente para a transformação. Quando os danos no DNA ocorrem numa célula que se prolifera normalmente é produzido um conjunto de células possuindo uma mutação. A expansão da população que possui essa mutação aumenta substancialmente a probabilidade de ocorrência de uma segunda mutação nas células que já contenham a primeira mutação. <sup>[6]</sup>

Após uma ou mais mutações nos proto-oncogenes ou genes supressores de tumores, a célula pode proliferar mais rapidamente na presença de um estímulo de crescimento e com a promoção de mutações, a célula irá crescer autonomamente, isto é, independentemente dos controlos do crescimento normal. Desta forma, o crescimento acentuado aumenta a probabilidade de promoção de mutações. <sup>[6]</sup>

A organização e a duração do ciclo celular das células cancerígenas não sofre nenhuma alteração em relação ao das células normais, no entanto, os *checkpoints* são afectados. Os *checkpoints* asseguram a ocorrência de poucas mutações nas células normais, mas quando estes estão alterados aumentam a proporção das mutações nas

células cancerígenas, resultando na perda progressiva do controlo e, conseqüentemente, na doença neoplásica. <sup>[10]</sup>

Uma minoria das mutações propensas ao aparecimento do cancro é hereditária. Nestes casos, um gene mutado num cromossoma ou num par de cromossomas é herdado. Se, ao longo da vida, ocorrer uma mutação neste gene nesta parte do cromossoma, podem ser originadas células cancerígenas. Por outro lado, a maioria das mutações relacionadas com o desenvolvimento de cancro aparece ao longo da vida, podendo também serem causadas por vírus como o SV40 e o papilomavírus. <sup>[10]</sup>

Os cancros também estão muito associados com os meios ambientes mutagénicos, como os produzidos pelo tabaco. Exposições repetidas podem produzir diversas mutações que são necessárias para causar o cancro. As mutações dominantes podem activar o crescimento promovido pelos oncogenes, ou a perda ou inactivação dos genes supressores de tumores que limitam o crescimento. <sup>[10]</sup>

### **2.1.6. Factores que predispõem o desenvolvimento de cancro**

O número de casos de cancro tem vindo a aumentar ao longo dos anos, podendo esta situação estar relacionada com diversos factores, entre os quais o aumento da riqueza e da longevidade das sociedades. Como as pessoas vivem durante mais tempo, a probabilidade de desenvolverem cancro aumenta. Por outro lado, as sociedades mais ricas consomem mais quantidades de “comida rápida”, álcool e tabaco, o que aumenta a probabilidade do aparecimento de cancro.

A predisposição para desenvolver cancro está então relacionada com diversos factores como dieta alimentar, tabaco, vírus, bactérias, radiação, amianto, químicos, poluição, factores genéticos, entre outros. <sup>[9]</sup>

### **2.1.7. A apoptose**

Nos organismos complexos ocorre um processo de regulação das células, que inclui a embriogénese, a manutenção do número de células adequadas para os tecidos, a remoção das células infectadas ou danificadas, a manutenção do sistema imunitário, o

envelhecimento e a apoptose. Estes processos são, por sua vez, regulados por factores estimuladores e inibitórios. <sup>[6][8]</sup>

Geneticamente, a apoptose é a morte programada da célula, que origina uma desagregação e disposição organizada das células. Morfologicamente, a apoptose caracteriza-se por alterações que ocorrem na membrana celular, encolhimento do núcleo, condensação da cromatina e fragmentação do DNA. Os macrófagos e outras células fagocíticas reconhecem as células apoptóticas e removem-nas através de fagocitose sem desenvolver um fenómeno inflamatório. <sup>[8]</sup>

O crescimento do tecido, ou do número de células, é regulado pela apoptose. Este processo permite a eliminação de células supérfluas ou indesejáveis, e também de células doentes (células tumorais, células infectadas com vírus, etc). <sup>[8]</sup>

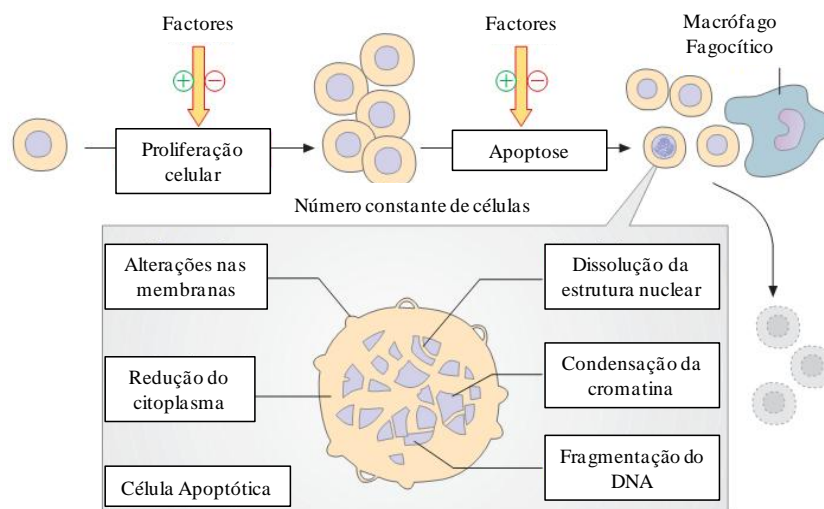
Na apoptose, os produtos da degradação das proteínas e do DNA (aminoácidos e nucleótidos) são libertados num processo controlado, podendo ser reutilizados pelas células vizinhas. A apoptose permite assim que o organismo elimine a célula sem desperdiçar os seus componentes. <sup>[11]</sup>

Muitas células podem controlar com precisão o tempo da sua própria morte por apoptose. A apoptose também ocorre noutros processos para além do processo de desenvolvimento. Algumas vezes, o suicídio celular não é programado, mas ocorre como uma resposta às circunstâncias biológicas que ameaçam o resto do organismo. Os mecanismos reguladores que accionam a apoptose envolvem algumas das proteínas que regulam o ciclo celular. O sinal para a ocorrência da apoptose provém frequentemente do exterior, passando para o interior através do receptor de superfície. Desta forma, a apoptose é activada por diversos estímulos, como a eliminação dos factores de crescimento, o aumento da proteína p53 como resposta aos danos existentes no DNA, a monitorização do DNA danificado pelas enzimas de reparação ou pela libertação de factores para a necrose de tumores (TNF) ou de outros factores imunitários. <sup>[6][11]</sup>

A apoptose pode ser iniciada em resposta a vários factores, entre eles a danificação das células, a radiação, os radicais livres ou outras toxinas. Este processo pode proteger os organismos do efeito negativo das mutações e provoca a destruição das células com danos irreparáveis no DNA antes da sua proliferação. O excesso do sinal de crescimento faz com que ocorra uma produção de um número excessivo de células indesejáveis, se a apoptose falhar na remoção destas células ou dos danos celulares irá, deste modo, contribuir para o desenvolvimento do cancro. <sup>[6]</sup>

A apoptose pode ser desencadeada através de vários sinais que utilizam diversas vias de transmissão. Existem, no entanto, outras vias de sinalização que previnem a apoptose. No processo apoptótico estão envolvidas as caspases que são enzimas com resíduos de cisteína, estas enzimas activam-se mutuamente provocando uma cascata enzimática. Neste grupo estão ainda incluídas outras enzimas, as caspases efectoras, que depois de serem activadas clivam os componentes celulares ou activam DNases especiais que fragmentam o DNA do núcleo. [8]

Na figura que se segue está representado um esquema acerca da proliferação celular e apoptose.



**Figura 2:** Esquema representativo dos processos de proliferação celular e da apoptose. [8]

### 2.1.8. Alterações malignas nas células

O processo de desenvolvimento de cancro é constituído por diversas fases, em cada uma destas fases ocorre a alteração genética que transforma a célula normal em maligna. As alterações que ocorrem na fisiologia da célula que influenciam o crescimento maligno são: [9]

- Auto-suficiência nos sinais de crescimento: ao contrário das células normais que necessitam de sinais de crescimento para proliferar, a maioria das células cancerígenas produzem os seus próprios sinais de crescimento, aos quais respondem, funcionando de um modo independente e não como fazendo parte de um organismo.

- Insensibilidade para sinais inibidores: as células monitorizam o seu meio ambiente externo e decidem se proliferam ou não. Muitos sinais anti-proliferativos

funcionam via a proteína do Retinoblastoma, e quando esta via é perturbada o ciclo celular não é controlado e as células proliferam.

- Evasão da apoptose: a apoptose programada existe em quase todas as células do organismo, no entanto, a maioria das células cancerígenas resiste a este processo.

- Potencial de replicação ilimitado: ao contrário do que se verifica para as células normais, as células cancerígenas apresentam a capacidade de replicar-se infinitamente, devendo-se este facto, na maioria dos casos, à enzima telomerase.

- Angiogénese sustentada: a angiogénese corresponde ao processo de formação de novos vasos sanguíneos, sendo este um processo essencial quando as células da massa do tumor possuem o seu próprio fornecimento de oxigénio e nutrientes.

- Invasão de tecidos e metástase: a maioria das mortes causadas pelo cancro deve-se à metástase do tumor primário para outras zonas do organismo. Numa primeira fase, os citoesqueletos das células rearranjam-se permitindo que as células tumorais adiram a outras células e se movimentem. Quando as células cancerígenas encontram um bloqueio, estas segregam enzimas que destroem essa barreira, possibilitando a sua entrada na corrente sanguínea e, desta forma, circulam ao longo do organismo até encontrarem um local adequado para se fixarem e crescerem. <sup>[9]</sup>

Quando as células normais tornam-se tumorigénicas, ocorrem três tipos de alterações, a imortalização, a transformação e a metástase. A imortalização deve-se ao facto das células crescerem indefinidamente, sem que tenham ocorrido outras alterações no fenótipo. A transformação refere-se ao facto das células cancerígenas não possuírem restrições no crescimento, tornando-se independentes dos factores que são necessários para o crescimento e sobrevivência das células. Por sua vez, a metástase corresponde à fase em que as células cancerígenas conseguem invadir o tecido normal, podendo deslocar-se ao longo do organismo e estabelecer uma nova “colónia” noutra zona. <sup>[12]</sup>

### **2.1.8.1. Transformação**

A transição das células normais para células tumorais é designada de transformação. As células normais apresentam características de células diferenciadas especializadas para uma determinada função, sendo inibidas na fase G0 do ciclo celular. A sua forma exterior é variável e é determinada por um citoesqueleto fortemente estruturado. <sup>[8]</sup>

Quanto às células tumorais, estas dividem-se sem estarem sujeitas à inibição e, em geral, não são diferenciadas. A superfície destas células é alterada, sendo este facto particularmente evidente no distúrbio da inibição por contacto pelas células vizinhas. O citoesqueleto das células tumorais é também reestruturado e frequentemente reduzido, dando-lhes uma forma arredondada. O núcleo das células tumorais pode ser atípico em termos de forma, número e tamanho.

A transição de um estado normal para um estado transformado é um processo que envolve diversos passos, que se encontram seguidamente citados. <sup>[8]</sup>

- Iniciação tumoral: a maioria dos tumores inicia-se com a lesão do DNA das células individuais. O defeito genético é quase sempre causado por factores do meio ambiente. É principalmente os defeitos nos proto-oncogenes que são relevantes para a iniciação do tumor, sendo estes a causa decisiva da transformação. A perda de um gene supressor de tumor também pode contribuir para a iniciação do tumor.

- Promoção do tumor: esta fase corresponde à proliferação preferencial da célula danificada através da transformação. Este é um processo muito lento que pode demorar vários anos, no entanto, determinadas substâncias são capazes de acelerar fortemente este processo.

- Progressão do tumor: nesta fase é originado um tumor macroscopicamente visível, em resultado do crescimento. Quando os tumores sólidos excedem um determinado tamanho, formam a sua própria rede vascular, possuindo o seu próprio fornecimento de sangue. <sup>[8]</sup>

## 2.2. Terapias anti-cancerígenas

Existem diversas formas de terapia anti-cancerígena, entre as quais podemos salientar: a terapia biológica, por vezes designada de imunoterapia, bioterapia ou terapia de modificação da resposta biológica, e a quimioterapia. <sup>[13][14]</sup>

A quimioterapia, que corresponde à utilização de fármacos anti-cancerígenos para destruir as células cancerígenas, efectuando a disrupção do crescimento destas células; a terapia de radiação (também chamado de radioterapia, terapia de raios-X, ou irradiação) que consiste na utilização de um determinado tipo de energia (chamada de radiação ionizante) para matar as células cancerígenas e diminuir os tumores; a terapia

fotodinâmica, que utiliza o laser ou outras fontes de luz, combinadas com um fármaco sensível à luz (agente fotosensibilizante) para destruir as células cancerígenas; a cirurgia; as vacinas; e a terapia genética. Neste trabalho estamos interessados na quimioterapia. <sup>[15][16]</sup>

### 2.2.1. Quimioterapia

Os fármacos utilizados na quimioterapia são geralmente administrados através de injeção intravenosa, circulando através da corrente sanguínea, de modo a chegar às células cancerígenas que se encontrem em qualquer parte do organismo.

A quimioterapia é o tratamento utilizado para tratar diversos tipos de cancro, podendo ser utilizado antes ou após a cirurgia, ou a radioterapia para os tornar mais eficazes, ou administrada conjuntamente com a radioterapia. Este tratamento também pode ser administrado a pessoas em que ocorreu o alastramento do cancro para outras partes do organismo, ou em casos em que o cancro reapareceu após a radioterapia. Também é utilizado para tentar diminuir e controlar o cancro, para aliviar os sintomas provocados por esta doença e para tentar prolongar a boa qualidade de vida dos pacientes. <sup>[16]</sup>

Existem três objectivos associados ao uso dos agentes anti-cancerígenos mais comuns, sendo estes a reparação das lesões no DNA das células cancerígenas afectadas; a inibição da síntese de novas cadeias de DNA de modo a parar a replicação da célula, uma vez que esta replicação permite que o tumor cresça; parar a mitose, porque desta forma a divisão celular (replicação) do cancro é parada, podendo parar a progressão do cancro. <sup>[17]</sup>

A maioria dos fármacos utilizada não é específica, provocando diversos efeitos secundários, que estão vulgarmente associados com a quimioterapia cancerígena. A quimioterapia tem como objectivo diminuir a divisão celular das células cancerígenas, tentando diminuir e, se possível, parar o crescimento e o alastramento do cancro. No entanto, os efeitos secundários são visíveis a nível físico, uma vez que afecta as células da pele, do cabelo, do sistema gastrointestinal e da medula óssea, provocando anemia, queda de cabelo, náuseas, vómitos, entre outros efeitos. <sup>[16][17]</sup>

Em geral, os agentes de quimioterapia podem ser divididos em três categorias principais com base no seu mecanismo de acção. Uma destas categorias é constituída pelos fármacos que param a produção de unidades para a síntese da molécula de pré-DNA. Os precursores do DNA são o ácido fólico, bases heterocíclicas e nucleótidos, que são produzidos naturalmente no interior das células. Todos estes agentes trabalham nos passos de formação de nucleótidos ou deoxiribonucleótidos (necessários para a produção do DNA). Quando estes passos são bloqueados, os nucleótidos, que são partes constituintes do DNA e RNA, não podem ser sintetizados. Deste modo, as células não podem replicar-se já que não é possível produzir o DNA sem nucleótidos. Nesta categoria estão incluídos fármacos como o metotrexato, o fluorouracil, o hidroxurea e o mercaptopurina. <sup>[17]</sup>

Outra das categorias dos fármacos utilizados na quimioterapia é a que engloba os fármacos que reparam os danos no DNA do núcleo da célula. Os agentes de quimioterapia danificam quimicamente o DNA e o RNA, afectam a replicação do DNA e/ou a paragem total da replicação ou provocam a produção de DNA ou RNA sem sentido (isto é, o novo DNA ou RNA não codifica nada útil). Alguns dos fármacos que pertencem a esta categoria são o cisplatina, os antibióticos daunorubicina, o doxorubicina e o etoposido. <sup>[17]</sup>

Existem também fármacos que evitam os efeitos da síntese ou falhas nos fusos mitóticos. Os fusos mitóticos são muito importantes devido à sua ajuda na divisão para a obtenção de uma nova cópia de DNA, que estará em cada uma das duas novas células durante a divisão celular. Estes fármacos interrompem a formação destes fusos e, portanto, interrompem a divisão celular. A esta categoria pertencem fármacos como a vimblastina, a vincristina e o paclitaxel. <sup>[17]</sup>

### **2.2.2. Alvos Terapêuticos**

Os alvos terapêuticos são moléculas chave, geralmente de natureza proteica, que participam no crescimento e na divisão das células cancerígenas de diferentes formas durante o desenvolvimento, crescimento e dispersão do cancro. Ao interferir com os alvos terapêuticos os sinais que estes transmitem às células cancerígenas para crescerem

e se dividirem descontroladamente são bloqueados, sendo possível, desta forma, ajudar a parar o crescimento e a divisão das células cancerígenas. <sup>[18]</sup>

Focando-nos nas alterações específicas que ocorrem a nível molecular e celular para o desenvolvimento do cancro, o tratamento pode tornar-se mais eficaz e ser menos nocivo para as células normais. Esta intervenção poderá fornecer assim uma forma mais adequada de tratamento do cancro, uma vez que os tratamentos podem ser individualizados com base num único conjunto de alvos moleculares produzidos pelos tumores dos pacientes. Estas terapias fazem com que o tratamento seja mais selectivo, afectando menos células normais, reduzindo os efeitos secundários e aumentando assim a qualidade de vida dos pacientes.

Os fármacos utilizados para interactuar com os alvos terapêuticos são “moléculas pequenas” que bloqueiam as enzimas específicas e os receptores dos factores de crescimento (GFRs) envolvidos no crescimento das células cancerígenas, podendo também serem designados de inibidores de transdução de sinais. Existem também fármacos que induzem a apoptose, fazendo com que as células cancerígenas sofram a morte celular, pois interferem com as proteínas envolvidas no processo da apoptose. <sup>[18]</sup>

## **2.3. Mecanismos de acção de fármacos**

### **2.3.1. Locais de acção dos fármacos**

Os fármacos actuam em diversos locais da célula, dependendo do tipo de acção que estes apresentam. Quando ocorre uma inibição enzimática, os fármacos actuam no interior da célula modificando as reacções bioquímicas. Este tipo de inibição pode ser reversível ou irreversível, competitiva ou não competitiva. <sup>[19]</sup>

Na interacção entre o fármaco e o receptor, os fármacos actuam na membrana celular através das interacções físicas e/ou químicas que vulgarmente ocorrem através dos sítios específicos do receptor do fármaco que estão localizados na membrana.

Quando ocorrem interacções não específicas, os fármacos actuam apenas fisicamente no exterior das células, podendo interagir quimicamente no exterior das membranas celulares. <sup>[19]</sup>

### 2.3.2. Modo de acção dos fármacos

É importante distinguir entre a actuação dos fármacos e os seus efeitos. Os fármacos actuam através de mecanismos fisiológicos utilizando produtos químicos.

Um dos maiores problemas da farmacologia é que nenhum fármaco produz apenas um único efeito. O efeito primário é o efeito terapêutico desejado, enquanto que os efeitos secundários são todos os outros efeitos próximos do efeito desejado, que podem ser benéficos ou nocivos. <sup>[19]</sup>

Os efeitos biológicos observados após o fármaco ser administrado são o resultado de uma interacção entre o composto químico e uma parte do organismo. Os mecanismos da acção dos fármacos podem ser vistos de perspectivas diferentes, como por exemplo, focando-se no sítio onde os fármacos actuam e a natureza geral da interacção entre o fármaco e a célula. <sup>[19]</sup>

Os fármacos podem actuar matando organismos estranhos, ou estimulando ou inibindo, as funções fisiológicas normais. No primeiro caso, há o exemplo dos agentes quimioterapêuticos que actuam matando ou enfraquecendo os organismos estranhos como bactérias, vermes e vírus. O principal princípio da acção é a toxicidade selectiva, isto é, o fármaco tem de ser mais tóxico para o parasita do que para o hospedeiro. No segundo caso, os fármacos actuam estimulando ou inibindo as funções fisiológicas normais. A estimulação aumenta a proporção da actividade enquanto a inibição reduz a proporção da actividade. <sup>[19]</sup>

### 2.3.3. Mecanismos de acção dos fármacos por inibição enzimática

Os mecanismos da acção dos fármacos através da inibição enzimática podem incluir a inibição enzimática directa, eliminação da expressão dos genes e dos antimetabolitos.

A maioria dos efeitos dos fármacos é produzida através de inibição enzimática. A inibição causada pelos fármacos pode ser reversível ou irreversível. Uma situação reversível ocorre quando o equilíbrio se estabelece entre a enzima e o fármaco inibidor. <sup>[20]</sup>

A inibição também poderá ser competitiva ou não competitiva. A inibição competitiva ocorre quando o fármaco, como é semelhante ao substrato normal, compete com este pelo centro activo da enzima. Os efeitos da concentração são importantes para a inibição competitiva. Na inibição não competitiva, o fármaco combina-se com a enzima num sítio diferente do centro activo. O substrato normal não pode substituir o fármaco a partir deste sítio e não pode interagir com qualquer centro activo, desde que a forma da enzima tenha sido alterada.

Os fármacos também poderão actuar suprimindo a expressão dos genes, sendo desta forma que actuam alguns antibióticos, fungicidas, antimaláricos, antivirais, entre outros fármacos. A expressão dos genes pode ser suprimida em diversos passos da síntese das proteínas ou da inibição da biosíntese dos ácidos nucleicos. Muitas substâncias que inibem a biosíntese dos ácidos nucleicos são muito tóxicas, caso o fármaco não seja muito selectivo na sua acção entre o parasita e o hospedeiro. <sup>[20]</sup>

Os metabolitos são as substâncias usadas ou produzidas nas reacções bioquímicas. O fármaco que possui uma similaridade química muito próxima ao metabolito normal é chamado de antimetabolito. O antimetabolito participa na reacção sintética normal através da interacção com uma enzima e produzindo um metabolito falso. O metabolito falso inibe outra enzima, podendo o produto final da interacção ser inútil e não poder ser utilizado pela célula para o crescimento ou para a reprodução. Alguns metabolitos têm sido usados como agentes anti-bacterianos e anti-cancerígenos. <sup>[20]</sup>

#### **2.3.4. Interações entre o fármaco e o receptor**

A maioria dos fármacos possui uma elevada correlação entre a sua estrutura e a especificidade que exhibe em relação ao alvo em questão, actuando de modo eficaz contra a doença a tratar sem produzir efeitos tóxicos. Na maioria dos casos, é necessária uma estrutura química específica para o centro activo do receptor e para a estrutura do fármaco complementar. Se a estrutura molecular do fármaco sofrer alterações, mesmo que estas sejam insignificantes podem alterar drasticamente a especificidade do mesmo. <sup>[21]</sup>

Diversas forças químicas podem participar numa ligação temporária do fármaco ao receptor, podendo qualquer destas ligações estar depois envolvida na interacção entre estes. Por definição, a interacção entre o fármaco e o receptor é reversível, sendo muito raro a formação da ligação covalente. Se os fármacos contiverem grupos funcionais ácido e amina, que são ionizados a pH fisiológico, poderão ser formadas ligações iónicas através da atracção das cargas opostas no centro activo do receptor.

As interacções dipolo-dipolo, como as pontes de hidrogénio, são uma extensão da atracção de cargas opostas. A reacção entre o fármaco e o receptor poderá basear-se essencialmente na formação da ligação de hidrogénio entre a molécula do fármaco, a água circundante e o centro activo do receptor. <sup>[21]</sup>

As ligações hidrófobas poderão ser formadas entre os grupos não polares existentes no fármaco e os que estão presentes no centro activo do receptor. Estas ligações não são muito específicas mas as interacções ocorrem de modo a excluir as moléculas de água. As forças repulsivas que poderão diminuir a estabilidade a interacção entre o fármaco e o receptor incluem a repulsão de cargas semelhantes e o impedimento estérico. O impedimento estérico refere-se a determinadas características tridimensionais onde a repulsão ocorre entre as nuvens electrónicas, ligações químicas inflexíveis ou grupos alquilo grandes. <sup>[21]</sup>

### **2.3.5. Mecanismos dos fármacos anti-cancerígenos**

Os fármacos anti-cancerígenos têm mecanismos de acção distintos, podendo variar nos efeitos que provocam nos diferentes tipos de células normais e cancerígenas. Muitos dos fármacos anti-cancerígenos têm como alvos o DNA, topoisomerasas, microtúbulos, deacetilases da histona e cinases proteicas essenciais (como a CDK9). <sup>[22]</sup>

Alguns dos fármacos anti-cancerígenos exploram os defeitos na regulação do ciclo celular, que são comuns nas células cancerígenas, outros envolvem moléculas que actuam preferencialmente nas células cancerígenas, matando-as, não afectando as células normais. Outros fármacos podem ser constituídos por proteínas virais que necessitam de células que se dividam rapidamente para completar o seu ciclo de vida, também existem fármacos que induzem a apoptose utilizando os componentes que

regulam este processo, como caspases, FADD, receptores de TRAIL, para matar as células cancerígenas. <sup>[17][23]</sup>

Existem fármacos que actuam ao nível da replicação do DNA, mais precisamente nas polimerases do DNA, sendo estes designados de agentes danificadores do DNA. Estes agentes modificam o ácido nucleico, não funcionando durante muito tempo como substrato eficaz para a polimerase do DNA. Alguns agentes de quimioterapia apresentam este modo de funcionamento, como é o caso da temozolomida e da cisplatina, que modificam a composição e a estrutura do DNA para inibir a síntese do mesmo e para prevenir a proliferação celular. <sup>[24]</sup>

Outra estratégia adoptada por alguns fármacos anti-cancerígenos é a inibição de enzimas envolvidas na replicação do DNA, como é o caso dos etoposídeos que inibem a actividade da topoisomerase. Esta inibição provoca a apoptose criando quebras na cadeia dupla ou simples do DNA interrompendo, deste modo, a continuidade da síntese do mesmo.

Outra aproximação inclui a redução da disponibilidade de dNTPs que estão envolvidos na síntese do DNA. Inibindo as enzimas envolvidas no metabolismo dos nucleótidos, os antimetabolitos como metotrexato e hidroxiureia reduzem o número de nucleótidos para impedir indirectamente a síntese de DNA. A utilização de análogos de nucleósidos é a aproximação mais directa, estes podem ser AZT ou fludarabina que têm como alvo a actividade enzimática da polimerase do DNA. <sup>[24]</sup>

Uma das formas de actuação de alguns fármacos utilizados na terapia do cancro envolve o aumento da imunidade endógena, mediada pelas células T através da interrupção das vias de regulação destas células. A vantagem desta aproximação é o facto de ter como alvo o sistema imunitário do hospedeiro, eliminando a necessidade de identificar os antígenos tumorais específicos e fornecendo, deste modo, um alvo em vez de vários alvos. <sup>[25]</sup>

Como os agentes anti-cancerígenos que têm como alvo o DNA por vezes não curam tumores sólidos, têm sido desenvolvidos estudos para encontrar fármacos que actuem nos reguladores da transdução de sinal (família do receptor HER, Ras, Raf e cinases MEK); reguladores da sobrevivência das células (Bcl-2 e os seus homólogos); proteínas oncogénicas como Bcr/Abl; proteínas reguladoras do ciclo celular como as ciclinas, as cinases dependentes de ciclinas e os inibidores das cinases dependentes das ciclinas; e as proteínas envolvidas na angiogénese tumoral, como as metaloproteínas de matriz e receptores do factor de crescimento endoteliais. <sup>[26]</sup>

Apesar dos diversos tratamentos existentes para o cancro, esta continua a ser uma área de investigação difícil, uma vez que existem vários tipos de cancro e são uma minoria os tratamentos que conseguem distinguir as diferenças bioquímicas entre as células cancerígenas e as células normais. Por esta razão, a eficácia de muitos fármacos anti-cancerígenos é limitada pela sua toxicidade em relação ao crescimento das células normais.

Outro problema dos tratamentos para o cancro está relacionado com o facto das células cancerígenas, que são inicialmente eliminadas por um fármaco específico, poderem tornar-se resistentes a esse mesmo fármaco. Desta forma, a quimioterapia utilizada no tratamento do cancro pode consistir na utilização de diversos fármacos nos diversos períodos de tempo do tratamento. <sup>[17]</sup>

## 2.4. Desenho de fármacos assistido por computador

O desenho de fármacos assistido por computador (CADD), também designado de desenho molecular assistido por computador (CAMD), representa uma das mais recentes aplicações dos computadores como ferramentas no processo de desenho de medicamentos. <sup>[27]</sup>

Ao utilizar o CADD, são realizadas tentativas para encontrar um ligando que irá interagir favoravelmente com o centro activo do receptor. As ligações entre o ligando e o receptor podem ser estabelecidas através de interações hidrófobas, electrostáticas e pontes de hidrogénio. As energias de solvatação do ligando e do sítio do receptor também são importantes devido à parcial ou completa solvatação que pode ocorrer previamente ao estabelecimento da ligação.

A aproximação usada em CADD depende da quantidade de informação que está disponível acerca do ligando e do receptor. Idealmente, deve-se possuir a informação estrutural a três dimensões para o receptor e para o complexo ligando-receptor a partir da difracção de raios-X ou NMR. Por outro lado, podem não existir dados experimentais para a construção de modelos do ligando e do receptor e assim os métodos computacionais podem ser aplicados sem as restrições fornecidas pelos dados experimentais. <sup>[27]</sup>

Com base na informação disponível, esta aproximação pode ser aplicada nos métodos de desenho molecular baseados no ligando ou no receptor. A aproximação baseada no ligando é aplicada quando a estrutura do centro activo do receptor é desconhecida e temos uma série de compostos que exercem uma determinada actividade de interesse. Para esta aproximação ser usada com uma maior eficácia, os compostos em estudo deverão possuir uma estrutura semelhante e uma variedade de actividades, ou seja, alguns compostos deverão apresentar uma elevada actividade, outros deverão ter uma gama de actividades intermédia e deverá também existir compostos sem actividade.<sup>[27]</sup>

Quando existe um bom modelo do centro activo do receptor (obtido através da difracção de raios-X ou NMR, por exemplo) é utilizada outro tipo de aproximação que consiste no desenho de ligandos que poderão interagir favoravelmente com o centro activo através do *docking*.<sup>[27]</sup>

Os métodos de desenho de fármacos baseado na estrutura envolvem a modelação da estrutura tridimensional da proteína, que é um potencial alvo, ao interagir com diversos compostos orientadores (*lead compound*). Estes compostos orientadores são pequenas moléculas que constituem o ponto de partida para uma optimização envolvendo diversas moléculas pequenas, que estão muito relacionadas com a estrutura do composto inicial.<sup>[28]</sup>

Os compostos são modelados computacionalmente para calcular a sua ligação ao alvo através de uma função de energia. A maioria dos algoritmos tem em consideração as interacções estruturais e funcionais, como as ligações de hidrogénio e as interacções hidrófobas.

Após a fase inicial de desenho, segue-se a fase de síntese do composto promotor, os testes de ligação à proteína alvo, a co-cristalização do composto e do alvo para os estudos estruturais com raios-X. Após a obtenção dos dados experimentais que informam sobre a forma como os compostos orientadores se ligam ao alvo, é possível alterar os compostos orientadores de forma a melhorar a ligação entre estes compostos e o alvo. Os compostos aperfeiçoados são sintetizados e formam um complexo com o alvo, sendo depois novamente melhorados através de um novo processo de optimização.<sup>[28]</sup>

O desenho de fármacos baseado na estrutura é frequentemente utilizado conjuntamente com aproximações da química combinatória, envolvendo a gestão de bases de dados de moléculas pequenas ou de bibliotecas combinatórias.<sup>[28]</sup>

### 2.4.1. Duas Classes de Problemas

Os problemas de desenho de fármacos estão divididos em duas categorias dependendo da estrutura química e da geometria do receptor ser conhecida ou não. Se o receptor é conhecido, é necessário encontrar um ligando que possa ser colocado no centro activo do receptor numa conformação que resulte numa menor energia para o complexo, este é o problema do *docking*. Este problema possui diversas variações, pois poderá ser pretendida uma descrição precisa da interacção ou poderá ser estimada uma aproximação para saber que ligandos, presentes na base de dados, se ligam preferencialmente ao receptor. <sup>[29]</sup>

Geralmente, o local de ligação é desconhecido tendo de ser utilizadas aproximações indirectas que utilizam diversos ligandos que interagem com um receptor específico. Estes ligandos podem ser obtidos experimentalmente e, usando a estrutura geométrica e as características químicas destas moléculas, pode ser obtida informação acerca do receptor, sendo importante identificar o farmacóforo presente nestes ligandos. O farmacóforo é um conjunto de características num arranjo tridimensional específico contido em todas as conformações activas das moléculas consideradas, é a parte (ou partes) da molécula que é responsável pela actividade do fármaco, enquanto que o resto da molécula é a matriz para as características farmacológicas. Se o farmacóforo for determinado ao analisar as diferentes actividades, as formas relativas e as estruturas químicas das moléculas iniciais, este pode ser utilizado para desenhar um fármaco mais eficaz. <sup>[29]</sup>

# CAPÍTULO III



# REVISÃO BIBLIOGRÁFICA

## Capítulo III – Revisão Bibliográfica

### 3.1. Docking

O *docking* é um método computacional muito utilizado no *design* de fármacos, no estudo da interacção entre fármacos, ou potenciais fármacos, e as enzimas e também na identificação de alvos moleculares para alguns compostos. Desta forma, têm sido diversos os trabalhos que reportam estudos utilizando o *docking*.

Alguns dos trabalhos desenvolvidos, utilizando esta metodologia focam-se na identificação de alvos moleculares para compostos que tenham uma actividade biológica conhecida, efectuando o *docking* inverso de modo a identificar potenciais proteínas com as quais os compostos interajam favoravelmente. Utilizando também o método de *docking* inverso, é possível identificar o alvo preferencial para um determinado ligando, utilizando para este fim uma base de dados de proteínas (PDB, por exemplo), realizar o *docking* entre os ligandos e as proteínas, de modo a recuperar o receptor com o qual cada ligando interage preferencialmente. <sup>[30][31]</sup>

O *docking* é uma metodologia que também é aplicada no estudo do local da ligação entre fármacos e o receptor. Neste tipo de estudos é utilizado um inibidor conhecido (um fármaco, por exemplo) e o alvo que este inibe, efectuando o *docking* entre estes, é possível identificar o local onde ocorre a ligação entre ambos, sendo este tipo de aplicação importante para identificar potenciais centros activos desconhecidos nos alvos terapêuticos. <sup>[32]</sup>

O referido método computacional pode ainda ser utilizado no desenho de fármacos baseado na estrutura, de modo a identificar o farmacóforo e estudar a interacção entre ligandos e o alvo que se pretende inibir. Os resultados obtidos podem auxiliar no desenho de novos fármacos e na modificação dos compostos de modo a melhorar a sua interacção com o alvo em questão. <sup>[33]</sup>

### 3.2. Relação Quantitativa entre a Estrutura e a Actividade (QSAR)

No âmbito das relações quantitativas entre a estrutura e a actividade dos compostos utilizando métodos lineares, têm sido desenvolvidos diversos estudos demonstrando as diversas áreas em que esta metodologia pode ser aplicada.

Uma das áreas em que os estudos de QSAR têm relevado uma elevada aplicabilidade é na toxicidade aquática. Neste tipo de trabalhos, é efectuada a determinação da toxicidade dos pesticidas nos peixes, sendo depois correlacionados os descritores moleculares, calculados para os pesticidas, com as actividades biológicas em estudo. A aplicação do QSAR à toxicidade aquática é importante pois possibilita a identificação de subestruturas biologicamente activas responsáveis pela actividade dos pesticidas, podendo orientar na síntese de novos compostos químicos que não sejam tão prejudiciais para o meio ambiente aquático e, conseqüentemente, para os peixes. [34][35][36]

A maioria dos estudos de QSAR focam-se na influência da estrutura dos compostos na sua actividade biológica, como é caso do trabalho desenvolvido por Chang, *et al.* que incide sobre os compostos fenólicos que apresentam a capacidade de induzir a apoptose das células cancerígenas. Este trabalho é muito importante, já que percebendo o mecanismo da actividade destes compostos e que características estruturais são determinantes para a actividade indutora da apoptose, é possível alterar estruturalmente os compostos fenólicos de forma a melhorar a sua actividade biológica. [37]

O QSAR tem ainda apresentado uma elevada aplicabilidade no auxílio da síntese de novos compostos com propriedades desejadas. Neste tipo de estudos, o QSAR é utilizado geralmente, como método complementar à síntese, ao *design*, ao *docking* e à criação de farmacóforos dos compostos em estudo. A maioria destes trabalhos é desenvolvida com o intuito de melhorar a actividade de compostos (alguns utilizados como fármacos) e, ao mesmo tempo, torná-los mais selectivos e menos tóxicos. Desta forma, os resultados obtidos são relevantes do ponto de vista da compreensão das ligações que se estabelecem entre os compostos e as enzimas, da identificação das características responsáveis pela actividade biológica dos compostos e também pela melhoria da eficácia de alguns fármacos, diminuindo os efeitos secundários dos mesmos. [38][39][40]

### 3.3. Redes Neurais Artificiais (ANN)

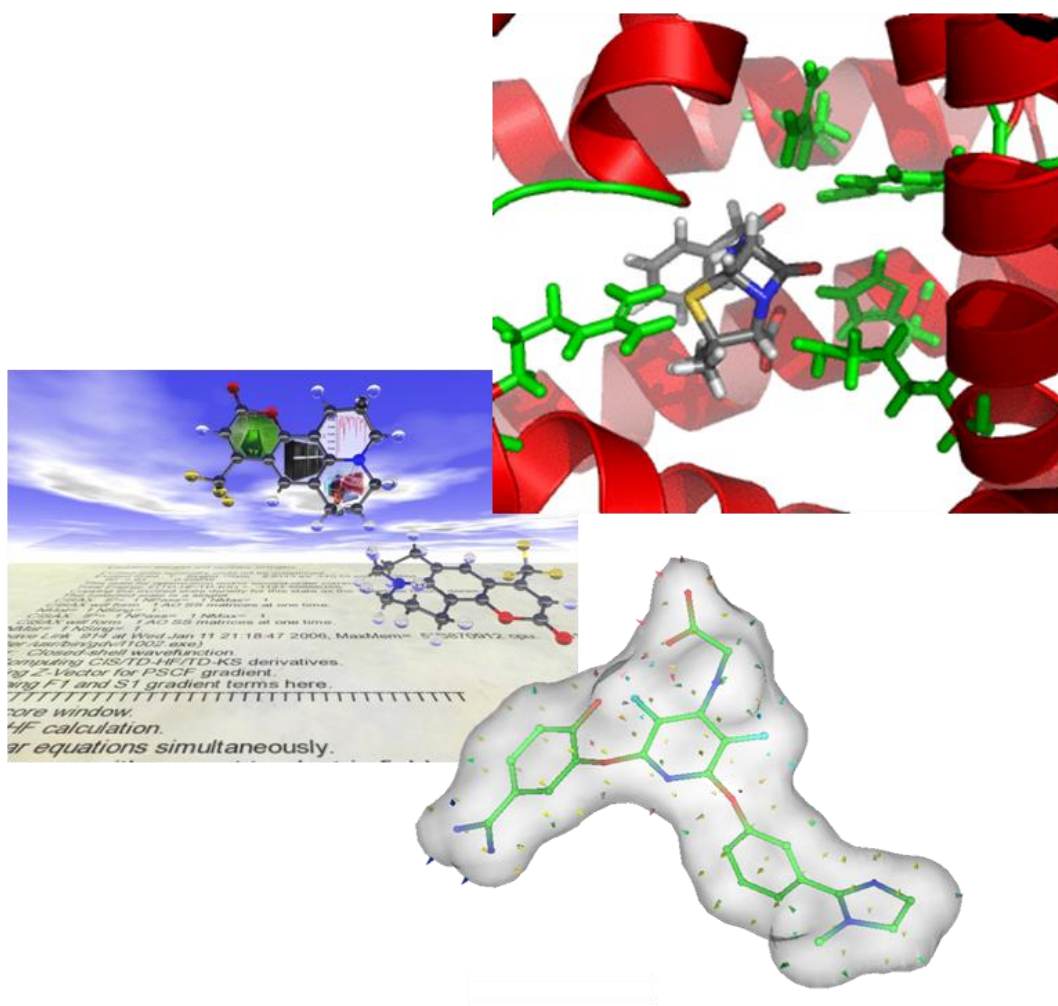
Os estudos realizados utilizando as redes neuronais artificiais (ANN) têm vindo a aumentar sendo, no entanto, uma metodologia ainda relativamente recente. Existem várias áreas às quais as ANN podem ser aplicadas, desde estudos sobre as relações entre a estrutura e a actividade, a análise de compostos, a análise alimentar, entre outras.

Na análise alimentar esta metodologia é utilizada para efectuar a análise sensorial, a rastreabilidade, para mapear as preferências dos consumidores e para identificar produtos alimentares adulterados, entre outras aplicações. Neste âmbito, as redes neuronais artificiais são utilizadas conjuntamente com técnicas utilizadas vulgarmente para a análise alimentar, como a cromatografia gasosa acoplada a um espectrómetro de massa, para determinar os compostos presentes na amostra e seleccionar os marcadores dos alimentos em análise. Este tipo de aplicação é particularmente importante no que diz respeito à identificação de compostos adulterados e para avaliar se os produtos alimentares estão de acordo com as normas em vigor.<sup>[41][42]</sup>

Alguns investigadores utilizam as ANN conjuntamente com técnicas espectroscópicas para efectuar a análise quantitativa de componentes presentes em amostras farmacêuticas, a informação quantitativa é obtida através do processamento das técnicas espectroscópicas utilizando os modelos das redes neuronais artificiais. A utilização das ANN em conjunção com as técnicas espectroscópicas pode ser aplicada para a determinação de índices de viscosidade, densidade, pontos de ebulição, número de determinados grupos químicos, entre outros parâmetros.<sup>[43][44][45]</sup>

As ANN têm vindo a ser aplicadas na medicina, no diagnóstico, no prognóstico e na orientação terapêutica, de forma a auxiliar os médicos a focarem-se nos pacientes reais e diminuindo o custo monetário empregue na medicação. As ANN conseguem, a partir de diversos padrões e atributos que caracterizam as células em benignas ou malignas, identificar a presença ou não de células cancerígenas. A construção de modelos de ANN que sejam de confiança constitui um passo importante, porque a detecção precoce de cancro é um factor importante para o seu tratamento.<sup>[46][47][48]</sup>

# CAPÍTULO IV



# METODOLOGIA

---

## Capítulo IV - Metodologia

### 4.1. *Docking*

O *docking* é uma aplicação importante para a modelação molecular assistida por computador, baseando-se no estudo da complementaridade molecular, de forma a investigar os possíveis fenómenos de interacção molecular.

No *docking* existem dois princípios que são importantes para o reconhecimento e ligação dos elementos, receptor e ligando, que participam neste processo. O primeiro princípio é o da complementaridade da forma, este indica que as formas das moléculas que constituem o complexo de *docking* são complementares, isto é, existe um grande ajuste entre as superfícies dos componentes do *docking*. O segundo princípio é o da complementaridade química que demonstra que existe uma forte interacção química (em relação às ligações de hidrogénio, interacções electrostáticas, hidrofobicidade, etc) entre os componentes envolvidos no *docking*.<sup>[49]</sup>

Os procedimentos de *docking* têm como objectivo identificar as posições correctas dos ligandos no centro activo da proteína e prever a afinidade entre o ligando e a proteína. Desta forma, o *docking* descreve um processo em que duas moléculas são ajustadas entre si no espaço tridimensional.

A estrutura cristalina dos ligandos presentes no receptor alvo é uma das fontes mais importantes para obter informação acerca dos mecanismos básicos de interacção entre as partes constituintes da estrutura do complexo tridimensional. As interacções podem produzir diversos acontecimentos, como por exemplo a reacção catalítica com a clivagem do substrato ou estabilização do estado de transição, inibição de enzimas ou bloqueio do centro activo das proteínas devido à ligação muito próxima com um inibidor. Este tipo de acontecimentos representa a base molecular dos efeitos farmacológicos.<sup>[2]</sup>

#### 4.1.1. Requisitos Básicos para a Realização do *Docking*

Existem diversos requisitos que têm de ser cumpridos para ser possível a realização do *docking*. Estes requisitos são o acesso à estrutura da proteína alvo com ou

sem um ligando no complexo proteína-ligando, às moléculas de interesse ou à base de dados que contenha compostos existentes ou virtuais para realizar o processo de *docking* e um sistema computacional que permita a implementação do *docking* e os procedimentos de *scoring*.

Na maioria dos algoritmos de *docking*, a proteína é considerada como sendo rígida devido ao elevado custo computacional que implicaria considerar a flexibilidade da mesma. No entanto, o ligando geralmente é considerado como sendo flexível.

O *docking* pode ser realizado colocando moléculas ou fragmentos rígidos no centro activo da proteína utilizando diferentes aproximações como o *clique-search*, *geometric hashing* ou *pose clustering*. Na *clique-search* os parâmetros são utilizados para descrever as características compatíveis (padrão de forma ou de interacção) entre o ligando e a proteína através da média da distância do gráfico de compatibilidade. A função *geometric hashing* é criada para descrever as características geométricas como as distâncias em dois passos, a fase de pré-processamento e a de reconhecimento. Esta aproximação é vantajosa do ponto de vista de tempo e de eficácia, e também devido à opção do *matching* parcial do ligando no centro activo da proteína. Relativamente ao *pose clustering*, este é um algoritmo baseado no *matching* dos tripletos das características do ligando com o triplete de características da proteína. As características representam as zonas de interacção entre o ligando e o receptor. <sup>[2]</sup>

### 4.1.2. Aplicações do *Docking*

O *docking* tem vindo a tornar-se uma ferramenta essencial no desenho de ligandos baseado na estrutura, apresentando uma vasta gama de aplicações. A principal aplicação deste método é a identificação de novos compostos activos para uma proteína alvo em particular. O comportamento das moléculas pequenas no centro activo das proteínas alvo pode ser descrito através do *docking* molecular, sendo esta técnica muito utilizada na descoberta de compostos orientadores e para a optimização. <sup>[2]</sup>

Esta metodologia funciona como um filtro rápido e seguro no rastreio virtual de alto débito, fornecendo assim uma variedade de opções para novas estruturas orientadoras. Por vezes, o *docking* e os métodos de *scoring* são realizados quando a síntese e os testes experimentais foram previamente efectuados, de forma a

correlacionar os dados obtidos computacionalmente com os da actividade biológica. O *docking* também é aplicado quando é necessário explicar ou identificar o modo de ligação para uma determinada classe de compostos.

A maioria das aproximações de *docking* têm como objectivo encontrar a metodologia específica e/ou que funções de *scoring* que são mais apropriadas para um determinado sistema alvo. O *docking* também é utilizado para rastrear base de dados constituídas por moléculas activas, quando os métodos de *docking/ scoring* conseguem discriminar eficazmente entre compostos activos e inactivos. Existem alguns alvos críticos que requerem a criação de modelos de proteínas ou a personalização dos processos de *docking*, sendo necessário aplicar o *docking* utilizando ligandos conhecidos para testar os novos sistemas. <sup>[2]</sup>

O *docking* tem contribuído grandemente para a descoberta de fármacos, sendo que uma das motivações principais na descoberta de fármacos é a identificação de pequenos *scaffolds* moleculares que apresentem elevada afinidade de ligação e elevada selectividade para o alvo e que apresentem um perfil ADME (absorção, distribuição, metabolismo, excreção) razoável. <sup>[2]</sup>

### 4.1.3. Programa eHiTS

Os programas de *docking* estão divididos em duas categorias, os programas que realizam aproximações estocásticas ou aleatórias e aqueles que efectuem aproximações sistemáticas ou directas. No primeiro caso encontram-se, por exemplo, os programas AutoDock2, DockVision, GOLD, ProLeads e no segundo caso, programas como FlexX, DOCK, FLOG, FRED, entre outros. No caso dos métodos sistemáticos, estes podem ser a construção incremental, a pesquisa conformacional, entre outros. No caso dos métodos estocásticos ou aleatórios, estes podem ser os algoritmos de Monte Carlo e genéticos.

O eHiTS, *electronic High-Throughput Screening*, é um programa de *docking* do ligando flexível. O programa utiliza um algoritmo de procura exaustiva, que enumera rapidamente as correspondências da interacção dos átomos entre o receptor e o ligando e considera todas as conformações sem um conflito estérico significativo. Este programa fornece uma optimização das coordenadas tridimensionais das conformações resultantes do *docking* das moléculas de ligando no centro activo do receptor. <sup>[5][50]</sup>

O eHiTS é indicado para aplicações de rastreio de alto débito, sendo capaz de reproduzir experimentalmente os modos de ligação observados. Teoricamente, uma procura exaustiva explora o espaço rotacional e translacional continuamente. Aquando do desenvolvimento do eHiTS, um dos principais objectivos foi a criação de um método exaustivo “inteligente” que conseguisse limitar a amostragem do espaço de procura às zonas de interesse, onde se encontram as boas soluções de *scoring*, eliminando os grandes espaços de procura onde seja garantido que não se encontre boas posições de *scoring*.<sup>[5]</sup>

O eHiTS possui um algoritmo de detecção que pode ajudar os químicos computacionais a encontrar o centro de ligação correcto no receptor. O algoritmo de *docking* realiza a interacção entre os fragmentos rígidos derivados do ligando independentemente do centro do receptor.

Numa primeira fase, o ligando é dividido em fragmentos rígidos e em cadeias de ligações flexíveis. Os fragmentos rígidos interagem independentemente no centro do receptor e depois um algoritmo gráfico de rápida correspondência encontra as conformações para cada fragmento, estes fragmentos podem ser novamente ligados formando assim o ligando de partida. Os fragmentos são ligados através de uma cadeia de ligação flexível apropriada, sendo esta posteriormente otimizada (utilizando um algoritmo de minimização local da energia) de forma a ser obtida a conformação final do ligando. Para formar o registo final, é guardado o número de conformações que foi previamente definido.<sup>[50]</sup>

O eHiTS realiza ainda a interacção entre todos os fragmentos independentemente do centro do receptor, utilizando um algoritmo hipergráfico de correspondência para enumerar todos os conjuntos de conformações compatíveis. Este conjunto de conformações é o conjunto das conformações dos fragmentos (uma conformação para cada fragmento) que são depois capazes de reconstruir o ligando usado inicialmente. Esta aproximação executa uma pesquisa exaustiva, isto é, encontra todas as soluções possíveis. A escolha de quais as conformações que devem ser guardadas para posterior processamento e optimização baseia-se no resultado global de todo o ligando e não apenas nas estruturas parciais.<sup>[50]</sup>

#### 4.1.3.1. eHiTS: o método

A avaliação de todas as posições e conformações possíveis não é viável tendo em conta o tempo de utilização do CPU, desta forma, é necessário reduzir o espaço de procura. O eHiTS efectua esta redução limitando a procura a conformações e posições que evitem conflitos estéricos significativos entre o receptor e o ligando, isto é, onde o ajuste geométrico é possível. <sup>[5]</sup>

O eHiTS, ao contrário de programas como o DOCK ou FlexX, não utiliza o método de construção incremental, pois efectua tentativas de modo a encontrar o “ótimo global” através da enumeração das combinações do *docking* de estruturas parciais independentes. O algoritmo utilizado neste programa produz os principais modos de *docking* que são compatíveis com os impedimentos estéricos e químicos.

Numa primeira fase de realização do *docking* utilizando o eHiTS, é determinado o local de ligação através da construção da grelha estérica para todo o receptor, as regiões são divididas em grupos separados, sendo identificados os possíveis sítios de interacção. Seguidamente, é obtida a descrição da cavidade que consiste em milhares de formas geométricas (poliedro). <sup>[5]</sup>

A aproximação realizada pelo eHiTS envolve a quebra dos ligandos em fragmentos rígidos, a ligação de cadeias flexíveis e a interacção sistemática, independentemente para cada fragmento rígido em todo o espaço disponível da cavidade. Este programa realiza o *docking* entre os fragmentos rígidos e todos os locais possíveis que se encontram na cavidade independentemente um dos outros. <sup>[5]</sup>

Uma correspondência exhaustiva dos conjuntos de conformações dos fragmentos rígidos é realizada através de um algoritmo gráfico de detecção rápida, resultando em vários milhares de combinações de conformações aceitáveis. Os resultados são fornecidos para cada componente avaliada e é possível efectuar uma decisão global sobre quais combinações das conformações dos fragmentos são as melhores. Isto significa que mesmo que os fragmentos tenham maus resultados de interacção com o receptor, poderá ser que parte do conjunto da conformação tenha resultados muito bons e assim, o conjunto total da conformação pode ser aceitável. <sup>[5][50]</sup>

Numa fase final, as cadeias flexíveis são ajustadas para as posições específicas dos fragmentos rígidos, constituindo o resultado final. As soluções reconstruídas definem a posição de ligação e a conformação do ligando, estas posições são

aperfeiçoadas através da minimização da energia local no centro activo do receptor, impulsionadas pela função de *scoring*.<sup>[5]</sup>

Devido ao facto do eHiTS realizar a interacção de todos os fragmentos rígidos independentemente do ligando, em qualquer local no interior do centro activo do receptor, é possível reutilizar a informação do *docking* para ligandos subsequentes se o fragmento for repetido. Esta é uma característica extremamente útil do eHiTS poupando-se, desta forma, tempo. Guardando a informação obtida através do *docking* para os fragmentos da base de dados, o eHiTS é capaz de simplesmente ler essa informação em vez de recalculá-la.<sup>[50]</sup>

### **4.1.3.2. Forma geométrica e características químicas**

A fragmentação do ligando foca-se na separação de fragmentos rígidos a partir de conectores flexíveis. O eHiTS identifica as ligações com rotação livre dentro do ligando inicial e estas ligações são removidas originando um conjunto de fragmentos rígidos. Sempre que uma ligação é partida durante o processo de fragmentação, ambos os átomos da ligação são duplicados, um é mantido no fragmento rígido e o outro na cadeia flexível.

Todos os sistemas de anéis são considerados rígidos e as suas conformações são preservadas, no entanto, é preferível utilizar conformações múltiplas dos anéis para uma amostragem conformacional completa. Os fragmentos acíclicos, com ligações duplas ou de ressonância e todos os átomos com hibridização  $sp^2$  são considerados rígidos.<sup>[50]</sup>

A cavidade e os ligandos candidatos são descritos através da forma geométrica e do gráfico de características químicas.<sup>[5]</sup>

### **4.1.3.3. Reconstrução e optimização**

Após a realização do ajuste de todas as cadeias flexíveis em relação à posição do fragmento rígido, é efectuada a reconstrução do ligando completo a partir dos fragmentos.

Cada hipergráfico define uma solução separada, cada solução é construída através da comparação dos pares de fragmentos rígidos na posição seleccionada, aos quais estão ajustados as cadeias flexíveis. A posição identificada de cada fragmento rígido e a conformação resultante do ajuste da cadeia flexível são sobrepostas, utilizando os dois átomos que formam a ligação que foi quebrada. Estes dois átomos são replicados, estando quer no fragmento rígido, quer na cadeia flexível podendo ser utilizados para orientar a reconstrução. <sup>[5][50]</sup>

A minimização contínua da energia local, que apenas permite alterações torsionais e as transformações de corpo rígido (rotações e translações), é aplicada ao ligando completo para aperfeiçoar as geometrias de ligação e soluciona qualquer amostragem irregular que existisse no poliedro inicial, com base no posicionamento do fragmento rígido.

A optimização termina quando o valor para a função de *scoring* não melhora em qualquer direcção no espaço de transformação dimensional ( $6+n$ ; em que  $n$  é o número de ligações com rotação livre), significando que o mínimo local foi atingido. Nesta técnica de optimização não existe nenhum elemento estocástico, uma vez que o objectivo é encontrar o mínimo local da função objectiva para todas as soluções particulares. A cobertura global do espaço de procura é garantida pela cobertura total da cavidade do fragmento rígido no *docking* e pelo resultado final das posições obtidas através do algoritmo exaustivo. <sup>[5][50]</sup>

#### 4.1.3.4. Protonação

O estado de protonação é um factor muito importante para a realização do *docking*, sendo este um dos aspectos que são tratados superficialmente durante o processo de *docking* pela maioria dos programas. No entanto, o programa eHiTS efectua uma aproximação única através da avaliação sistemática de todos os estados de protonação possíveis, quer para o receptor, quer para o ligando, num único passo. <sup>[5][50]</sup>

As situações ambíguas são atribuídas a posições que podem ser protonadas ou desprotonadas. Durante o *docking*, ambos os estados de alguns pontos da superfície são avaliados e ordenados, sendo seleccionado o melhor estado de protonação para cada

interacção individual independentemente, evitando assim o efeito combinatório dos grupos funcionais múltiplos com estados de protonação variáveis.

Os resultados num único passo, utilizando as propriedades ambíguas, contêm os resultados cumulativos que devem ser obtidos através do processamento de diversos passos de *docking* individuais com estados de protonação fixos, considerando todos os estados de protonação dos ligandos contra todos os estados de protonação do receptor.<sup>[5]</sup>

### 4.1.3.5. Função de Scoring

Existem três funções diferentes de *scoring* que são utilizadas no eHiTS. Primeiro é utilizado um *flag* químico simples e rápido baseado na função estatística de *scoring* ( $SF_s$ ) durante o *docking* do fragmento rígido e fases de *pose-matching*. Esta função não é muito sensível a pequenas variações na geometria, na distância de interacção e no ângulo de ligação de hidrogénio.

Uma função mais sensível é a função de *scoring* empírica ( $SF_e$ ), esta é utilizada durante a fase de minimização de energia local final. Esta função de *scoring* possui curvas regulares representando a dependência da distância e do ângulo das interacções, enquanto suporta a optimização baseada no gradiente de eficiência.

As posições finais são avaliadas através de uma terceira função de *scoring*, que consome mais tempo ( $SF_c$ ), esta combina quer os componentes estatísticos, quer os empíricos e ainda uma grelha adicional baseada nos termos geométricos. Considera ainda a estimativa de perda de entropia e outro novo elemento de *scoring* baseado na cobertura da área superficial do receptor. Esta função final de *scoring* tenta estimar a energia livre de ligação, sendo que o resultado preciso do *scoring* final é usado para ordenar as soluções obtidas.<sup>[5]</sup>

### 4.1.3.6. Vantagens do eHiTS

O programa de *docking* eHiTS oferece diversas vantagens, entre as quais o facto de apresentar um método de procura exaustivo, que é importante para minimizar a possibilidade de obtenção de falsos negativos durante o estudo de rastreio virtual. Este

programa consegue reproduzir as conformações obtidas por cristalografia de raios-X dos ligandos com uma precisão extremamente elevada, para uma vasta variedade de famílias. <sup>[5][50]</sup>

O algoritmo do eHiTS, ao realizar o *docking* dos fragmentos rígidos, reutiliza a informação proveniente de estudos das interações que tenham sido realizados anteriormente utilizando fragmentos iguais aumentando, desta forma, a velocidade de execução do *docking*. O manuseamento automático do estado de protonação permite ao eHiTS testar todos os possíveis estados de protonação do par receptor-ligando e, assim, fornecer a forma mais apropriada de acordo com a função de energias.

Este programa apresenta uma forma diferente de realização do *docking*, uma vez que é completamente automático, não é necessária a preparação das estruturas do receptor e do ligando (não é necessário atribuir os estados de protonação, as cargas parciais e a minimização da energia) e não compromete os resultados com a função de energia já que esta não é preparada para uma família específica de proteínas. <sup>[5][50]</sup>

#### **4.1.4. Realização do *Docking***

##### **4.1.4.1. Enzimas**

As enzimas utilizadas neste estudo são de diversas famílias e actuam em diversas fases do ciclo celular. A maioria das enzimas utilizadas para a realização do *docking* são alvos utilizados na terapia anti-cancerígena, tendo sido também usadas enzimas que são alvos na terapia anti-inflamatória e anti-bacteriana. Algumas das enzimas utilizadas como alvos na terapia anti-bacteriana que foram seleccionadas para este estudo são provenientes de organismos diferentes (como por exemplo bactérias e fungos). As enzimas utilizadas neste estudo encontram-se descritas nas seguintes tabelas.

**Tabela 1:** Enzimas utilizadas para o *docking*, sendo alvos utilizados na terapia anti-cancerígena. <sup>[51][52]</sup>

Alvos	Fases do ciclo celular	Família	Alvos	Fases do ciclo celular	Família
Actina	M	Transferase	FNTB	G0/G1	Transferase
AKT1	G1;G2/M;G0/G1	Transferase	GSK-3B	G1;G0/G1	Transferase
AMD1	G1;G0/G1	Liase	GTP	G0/G1	Transferase
Asp	G2	Isomerase	Importina B	G2/M	
AURKA	G2;G2/M;M	Transferase	KITLG	S	
AURKB	G2/M;M	Transferase	MAD2	M	
CAP-G	M		MAP2K1	G1;G0/G1	Transferase
CDC2	G1/S;G2;G2/M;M	Transferase	MAX	G0/G1	
CDC25B	G1;S;G2/M	Hidrolase	MCM2	G1/S	Hidrolase
CDC25C	G2/M	Hidrolase	MYC	G0/G1	
CDC42	G0/G1	Transferase	NEK2	G2	Transferase
CDC6	S;G2	Transferase	p107	G1	
CDC7	G1;G1/S;G0/G1	Transferase	p21 CIP1	G1;G1/S	
CDH1	M		p27 KIP1	G1;G1/S	Transferase
CDK2	G1/S;S	Transferase	p53	G1;G2	
CDK4	G1; G0/G1	Transferase	PAK1	G0/G1	Transferase
CDK7	G1;G2/M	Transferase	PIK3CG	G0/G1	Transferase
CENP-A	M		PIN1	M;G0/G1	Isomerase
CHK1	S;G2	Transferase	PKA	G2/M	Transferase
CHK2	G1;G2	Transferase	Plk1	G2;G2/M;M	Transferase
CCNH	G1;G2/M	Transferase	ROCK	G0/G1	Transferase
DP	G1		SKP1	G1;G1/S	Transferase
E2F	G1	Transferase	TOP1	G1/S;M	Isomerase
Eg5	G2;M	Hidrolase	TOP2A	G1/S;M	Isomerase
Erk1	G1;G2/M;G0/G1	Transferase	Tubulina	M	Transferase
Erk2	G1;G2/M;G0/G2	Transferase	WEE1	S/G2;G2/M;M	Transferase

**Tabela 2:** Enzimas utilizadas para o *docking*, sendo alvos da terapia anti-inflamatória.

Alvos	Família
COX	Oxidoreductase
IMPDH	Oxidoreductase
HO-1	Oxidoreductase
mPGES-1	Isomerase
iNOS	Oxidoreductase
MAP2K1	Transferase
5-LOX	Oxidoreductase

**Tabela 3:** Enzimas utilizadas para o *docking*, sendo alvos da terapia anti-bacteriana.

Alvos	Família	Características
DHFR	Oxidoreductase	Bactéria
DHFR	Oxidoreductase	Fungo
TOPOII	Isomerase	Fungo
TOPOIV	Isomerase	Bactéria
RNAP	Transferase	Bactéria
RNAP	Transferase	Fungo
DHPS	Transferase	Bactéria
DD-Transpeptidase	Hidrolase	Bactéria
Racemase da alanina	Isomerase	Bactéria
DD-Ligase	Ligase	Bactéria
EF-G		Bactéria
EF-G		Fungo
NA	Hidrolase	Complexado com o Zanamivir
NA	Hidrolase	Influenza A vírus
M2TM		Proteína Viral

As estruturas das enzimas foram obtidas da base de dados de proteínas, PDB. <sup>[51]</sup> As enzimas têm diversas estruturas associadas, tendo sido escolhidas para a realização do *docking* aquelas que possuíam um substrato com características mais semelhantes aos compostos em estudo e também as estruturas com melhor resolução. No entanto, não foi possível encontrar nesta base de dados algumas das estruturas para as enzimas em estudo, utilizando-se para estes casos os modelos obtidos por homologia para estas enzimas.

#### 4.1.4.2. Modelação por homologia

A modelação por homologia é um método utilizado para prever a estrutura das proteínas a partir da sequência de resíduos que a constitui, e baseia-se no facto de proteínas com sequências semelhantes também apresentarem estruturas semelhantes. Esta metodologia é importante para diversas áreas como o desenho de fármacos baseado na estrutura, o desenho racional de proteínas, a análise de interacções, entre outras.

Com base numa estrutura obtida experimentalmente para uma dada proteína, que irá servir como molde, é possível construir modelos para uma sequência homóloga. <sup>[53][54]</sup>

A modelação por homologia é constituída por várias etapas, como a identificação de estruturas tridimensionais conhecidas de proteínas relacionadas com aquela que se pretende modelar a estrutura e que possam ser utilizadas como moldes para a construção do modelo; o alinhamento da sequência da proteína utilizada como molde com a da proteína alvo; a construção do modelo para a proteína alvo com base na estrutura tridimensional da proteína usada como molde e no alinhamento das sequências; e o aperfeiçoamento/ validação/ avaliação dos modelos obtidos. <sup>[54]</sup>

Os modelos obtidos através deste método podem conter erros, podendo estes ser originados pela escolha do molde, por alinhamentos imprecisos e pela utilização de métodos de aperfeiçoamento ineficazes. <sup>[54]</sup>

#### **4.1.4.3. Criação dos modelos por homologia**

Para a criação de modelos por homologia das enzimas, recorreu-se ao servidor Web UniProtKB ([uniprot.org/uniprot/](http://uniprot.org/uniprot/)) para efectuar a pesquisa da sequência de resíduos da enzima através do seu nome. Esta pesquisa fornece vários resultados para cada enzima e para diversos organismos.

A escolha da sequência a utilizar é efectuada com base no organismo, neste caso *Homo sapiens*, a sequência obtida é depois transferida para o programa Swiss-PdbViewer 3.7. Através deste programa submete-se um pedido para a obtenção de um molde apropriado recorrendo ao servidor *Find appropriate SWISS-MODEL templates* obtendo-se vários resultados, o molde escolhido é o que apresentar a sequência de resíduos mais idêntica com a sequência pesquisada.

Após a comparação e o alinhamento das sequências no programa Swiss-PdbViewer 3.7, é submetido um pedido para obter um modelo com base nas sequências alinhadas, o modelo obtido é depois enviado para o correio electrónico do utilizador.

#### 4.1.4.4. *Ligandos*

Os ligandos estudados neste trabalho são compostos de origem natural que em ensaios celulares apresentaram actividade anti-cancerígena. Estes compostos pertencem a diversas famílias químicas.

As estruturas dos ligandos foram optimizadas utilizando o programa HyperChem 7.5, tendo sido realizada uma optimização com Mecânica Molecular, utilizando o algoritmo Polak-Ribiere, (gradiente conjugado). Os critérios utilizados para que as estruturas dos ligandos fossem definidas como optimizadas foram: o gradiente RMS, que calcula o declive da energia total como a raiz quadrada da média do valor, apresentar um valor inferior a 0,1kcal/(Å mol) ou que o número de ciclos realizados atingisse o valor máximo de 120 ciclos. Esta optimização foi realizada em condições de vácuo.

#### 4.1.4.5. *Optimização das condições de docking*

Para a optimização das condições de *docking* seleccionamos ligandos e enzimas representativos do conjunto em estudo. Nesta optimização procedemos à variação de alguns parâmetros para a realização do *docking*, estudando deste modo o efeito da variação desses parâmetros no tempo necessário para a realização do *docking* e nas energias de interacção que são obtidas.

Os parâmetros alterados no âmbito desta optimização foram a margem, utilizando os valores 3, 5 e 7Å, o número de conformações, usando os valores 15, 20, 25 e 32, e a opção *fast* para uma maior rapidez na realização do *docking*.

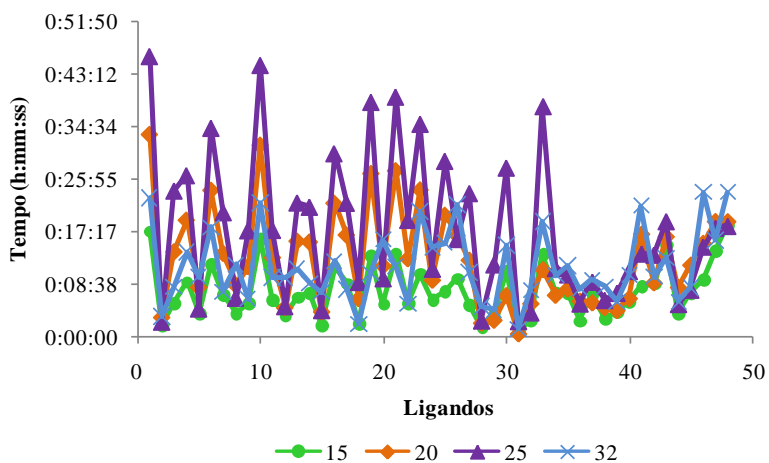
O *docking* decorre no centro activo do receptor, no local onde ocorre a interacção existe uma “caixa” que envolve o centro activo e o ligando. Desta forma, ao ser alterado o valor da margem é alterada a distância entre o local onde ocorre a interacção e os extremos da caixa que a envolve, sendo assim alterado o volume onde irá ocorrer a interacção.

Durante a realização do *docking*, são produzidas diversas conformações de forma a serem obtidos os melhores valores de energias de interacção, sendo apenas registados os melhores resultados para o número de conformações escolhido.

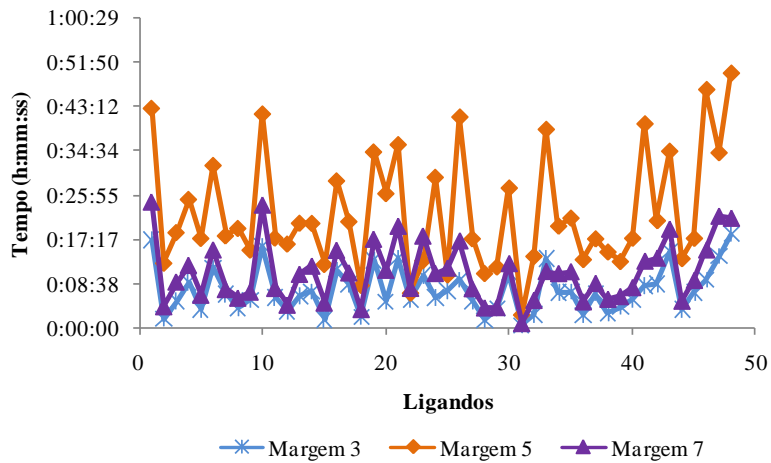
Na optimização das condições de *docking* também foi comparado o tempo necessário para a realização do *docking* e os valores de energia obtidos com e sem a utilização da opção *fast*. O parâmetro *fast* faz com que seja analisado um pequeno conjunto de soluções para a obtenção de energias de interacção mais favoráveis, fazendo desta forma com que o programa eHiTS execute o *docking* com uma maior velocidade.

Num primeiro passo, foi estudada a influência da variação do número de conformações requeridas no tempo de execução do *docking* e nos valores das energias obtidas. Conhecido o número de conformações que realiza o *docking* mais rapidamente sem afectar os valores de energia, é efectuada a variação dos valores das margens para esse número de conformações. Posteriormente, também foi estudada a influência da utilização da opção *fast*, que provoca um aumento da velocidade com que ocorre o *docking*.

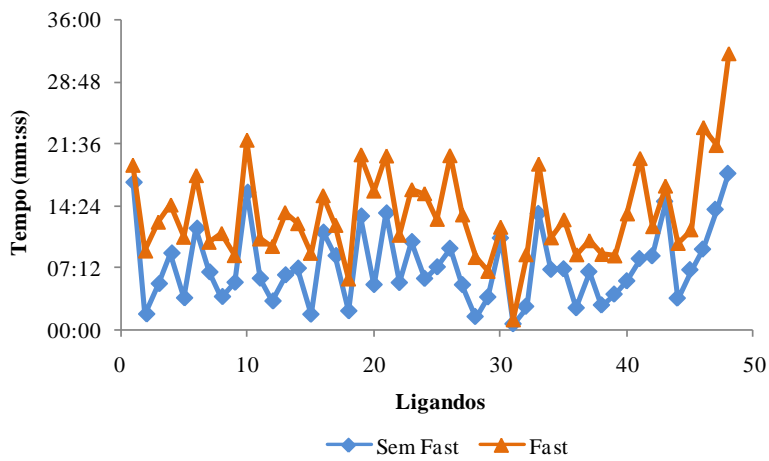
Nas figuras que se seguem encontram-se representados os gráficos da optimização de cada um dos parâmetros para a Tubulina, os gráficos referentes às restantes enzimas utilizadas para a optimização encontram-se no Anexo A.



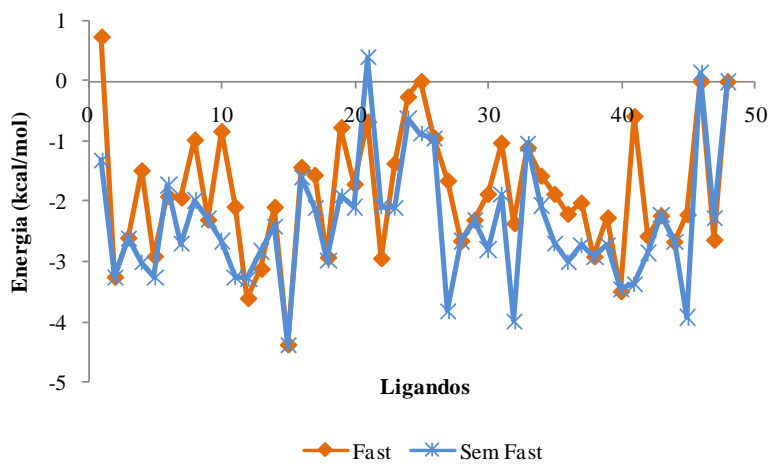
**Figura 3:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a proteína Tubulina.



**Figura 4:** Influência da variação da margem no tempo necessário para a realização do *docking* para a proteína Tubulina.



**Figura 5:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a proteína Tubulina.



**Figura 6:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a proteína Tubulina.

Realizando a optimização foi verificado que a variação das margens para o local onde decorre o *docking* influencia apenas o tempo necessário para a realização do mesmo, não influenciando as energias de interacção obtidas. Dos resultados obtidos observa-se que quanto menor é a margem utilizada menor é o tempo de *docking*. Desta forma, o valor de margem 3 Å é aquele para o qual o tempo de *docking* é menor.

No estudo da influência do número de conformações nos valores de energias obtidos e no tempo de *docking*, verificou-se que este factor apenas influencia o tempo de *docking*. O número de conformações 15 é o que tem um menor tempo de *docking*.

Estudando a influência da utilização da opção *fast* nos valores de energia obtidos e no tempo necessário para a realização do *docking*, constatou-se que o tempo de *docking* é menor quando não é utilizada a opção *fast*, tendo-se também obtido melhores valores de energia quando esta opção não é utilizada.

#### 4.1.4.6. Correção dos Resultados

Os resultados obtidos no *docking* foram depois rectificadas utilizando um factor que corrige os valores obtidos para cada um dos ligandos em relação a todas as enzimas. Para um ligando (*i*) que interagiu com todas as enzimas (*j*) é calculado o desvio padrão e a média das energias obtidas, da seguinte forma:

$$\mu_i = \sum_j (S_{ij})/N \dots\dots\dots j = 1, N \quad \text{Equação 1}$$

$$(\sigma_i)^2 = \sum_j (S_{ij} - \mu_i)^2 / (N-1) \dots\dots\dots j = 1, N \quad \text{Equação 2}$$

As energias obtidas através do *docking* são corrigidas utilizando a seguinte equação:

$$S_{ij}' = (S_{ij} - \mu_i) / \sigma_i \quad \text{Equação 3}$$

Em que  $S_{ij}'$  é a energia corrigida para o composto *i* no centro activo *j* e  $S_{ij}$  é o valor original de energia. O valor de energia corrigido é denominado de energia de correcção de múltiplos centros activos ou energia MASC (*Multiple Active Site*

---

*Corrections for Docking and Virtual Screening*). Esta correcção aumenta a exactidão dos resultados do ligando e reduz os valores de erro.

É importante que esta correcção seja efectuada, uma vez que existem compostos que apresentam bons valores de energia de interacção com todas as enzimas o que, no entanto, não significa que sejam bons ligandos. Isto acontece porque estes compostos possuem na sua constituição algum grupo funcional que lhes permitem interagir favoravelmente com todas as enzimas. Sendo assim, a correcção corrige este tipo de situações e evita-se assim a ocorrência de enviesamentos. <sup>[55]</sup>

#### **4.1.4.7. Testes de Enriquecimento**

Na realização dos testes de enriquecimento utilizou-se ligandos falsos obtidos a partir da base de dados DUD, *Directory of Useful Decoys*. <sup>[56]</sup> Os conjuntos de ligandos falsos que se encontram nesta base de dados estão organizados por famílias de enzimas, não se encontrando disponíveis ligandos falsos para todas as famílias estudadas no presente trabalho.

Para os casos em que não foi possível encontrar ligandos falsos através desta base de dados, recorreu-se ao programa ROCS, *Rapid Overlay of Chemical Structures*, para efectuar o rastreio da base de dados NCI, *National Cancer Institute*. Este rastreio foi efectuado com o intuito de obter um conjunto de compostos com características semelhantes aos ligandos conhecidos para cada enzima em questão e, que pudessem ser utilizados como ligandos falsos nos testes de enriquecimento.

O programa ROCS foi concebido para realizar rastreios de base de dados tridimensionais em larga escala utilizando um método de sobreposição que encontra a similaridade entre compostos de uma forma não intuitiva. Este programa baseia-se no método de sobreposição baseada na estrutura. Desta forma, as moléculas são alinhadas através de um processo de optimização de corpo sólido que maximiza o volume de sobreposição entre estas. O ROCS utiliza apenas os átomos pesados do ligando, sendo os hidrogénios ignorados. Como, neste contexto, o tamanho e o volume estão muito relacionados, o procedimento de maximização do volume de sobreposição do volume constitui um excelente método para aumentar a compreensão acerca das formas semelhantes.

O programa ROCS é primeiramente um método baseado na forma, utilizando definições especificadas da química que podem ser incluídas na sobreposição e na análise da similaridade, o que facilita a identificação dos compostos que são semelhantes, quer em termos de forma como de química.

Para realizar o rastreio utilizando o ROCS é necessário escolher uma base de dados para efectuar a procura dos compostos. Neste caso utilizou-se a base de dados do NCI que é constituída por um elevado número de compostos (260071), pertencentes a diversas famílias químicas. É também necessário um composto para servir de referência, ou seja, um composto com base no qual se pretenda efectuar a procura. Para esta situação, utilizou-se como composto de referência um dos ligandos conhecidos para cada enzima em questão.

Os resultados que se obtêm através do ROCS são ordenados de acordo com a similaridade, sendo possível ao utilizador seleccionar o número máximo de melhores resultados que pretende obter. <sup>[57]</sup>

## 4.2. Relação Quantitativa entre a Estrutura e a Actividade (QSAR)

O QSAR é a relação quantitativa entre a estrutura química e as suas actividades físicas, químicas ou biológicas. Este tipo de estudos é de grande importância para as áreas da química e da bioquímica e constitui uma das mais importantes aplicações das técnicas de modelação. <sup>[58][59]</sup>

Esta metodologia baseia-se na transformação da procura de compostos com propriedades desejadas utilizando a intuição química e a experiência numa forma quantificada matematicamente e computadorizada. Uma vez obtida correlação entre a estrutura e a actividade, é possível efectuar o rastreio de forma a escolher estruturas com as propriedades desejadas. Após esta selecção, os compostos que se apresentarem como sendo os mais promissores podem ser escolhidos para serem sintetizados e testados laboratorialmente.

A obtenção de boas correlações entre a estrutura e a actividade não é fácil, sendo importante utilizar os descritores moleculares adequados. Muitos dos descritores reflectem propriedades moleculares simples, podendo assim melhorar a compreensão acerca da natureza físico-química da actividade em questão. A qualidade dos modelos

de QSAR depende ainda da qualidade e do tipo de dados e é apenas válido para estruturas análogas à dos compostos usados para construir o modelo. <sup>[58][59]</sup>

É importante correlacionar as estruturas químicas dos fármacos com as suas actividades farmacológicas de interesse, uma vez que os custos do desenvolvimento de novos fármacos são elevados, a previsão segura da actividade dos compostos antes da sua síntese é de grande interesse para os laboratórios de síntese. <sup>[59]</sup>

O método QSAR envolve o reconhecimento de que a molécula é realmente a distribuição tridimensional de propriedades. As propriedades mais importantes são as estéricas (como a forma e o volume), as electrónicas (como a carga eléctrica e o potencial electrostático) e as propriedades lipofílicas (como são as secções polares e não polares das moléculas, sendo geralmente representadas pelo coeficiente de partição octanol-água, log P).

Esta metodologia envolve diversos passos-chave: 1) Conversão das estruturas moleculares em descritores matemáticos que abrangem as propriedades principais das moléculas e que sejam relevantes para a actividade em estudo. 2) Selecção dos melhores descritores a partir de um grande número de descritores acessíveis. 3) Relacionar os descritores moleculares com as propriedades. 4) Validação do modelo para determinar a sua capacidade de previsão e se este apresenta uma boa previsão quando aplicado a novas moléculas, que não estejam incluídas no conjunto de dados utilizado para criar o modelo (o conjunto de teste). <sup>[60]</sup>

O QSAR aplica diversos métodos como a Regressão Linear Múltipla (MLR), *Partial Least Squares* (PLS), Redes Neurais (NN), *Support Vector Machine* (SVM) e a Programação da Expressão de Genes (GEP), entre outros.

A Regressão Linear Múltipla (MLR) é um dos métodos mais antigos que têm sido utilizados na construção de modelos de QSAR, continuando a ser muito utilizado na actualidade. A vantagem deste método é o facto de ser uma expressão matemática de forma simples e facilmente interpretável. No entanto, este método é vulnerável em relação aos descritores que estão intercorrelacionados, fazendo com que seja incapaz de decidir que conjuntos podem ser mais significativos para o modelo. <sup>[61]</sup>

Com base na regressão linear múltipla foram desenvolvidas novas metodologias como o método da melhor regressão linear múltipla (BMLR), o método heurístico (HM), o algoritmo genético baseado na regressão linear múltipla (GA-MLR), entre outras. <sup>[61]</sup>

### 4.2.1. Descritores Moleculares

Os descritores moleculares representam a forma como a informação química, contida na estrutura molecular é transformada e codificada para relacionar com os problemas químicos, farmacológicos e toxicológicos nos estudos de QSAR, e têm em conta diferentes aspectos da informação química. A aproximação para obter esta informação pode ser obtida através de experiências, cálculos teóricos ou operações simples de contagem sobre toda a molécula, fragmentos das moléculas ou grupos funcionais e é necessário o conhecimento da estrutura tridimensional da molécula ou do seu gráfico molecular, ou simplesmente a sua fórmula para designar a informação definida por valores escalares, vectores ou campos escalares. <sup>[62]</sup>

Existem diversos grupos de descritores moleculares, entre estes os descritores constitucionais, topológicos, geométricos, electrostáticos, químico-quânticos e termodinâmicos. <sup>[63]</sup>

- Descritores Constitucionais: estes descritores moleculares apenas reflectem a composição dos compostos sem a utilização da estrutura geométrica ou electrónica da molécula. Alguns dos descritores incluídos neste grupo são o número de átomos, o número de ligações, peso molecular, entre outros.
  
- Descritores Topológicos: têm como função descrever a conectividade atómica existente na molécula. Alguns destes descritores representam o valor aditivo dos incrementos atribuídos a fragmentos estruturais específicos, nomeadamente o parâmetro de ligação de comprimento  $n$  (ordem do descritor). Neste grupo estão incluídos descritores como o índice de Wiener, os índices de Hier & Hall, o índice de flexibilidade de Kier, entre outros.
  
- Descritores Geométricos: Estes descritores descrevem a geometria das moléculas, necessitando portanto, das coordenadas tridimensionais dos átomos presentes nas moléculas. Alguns dos descritores moleculares que constam neste grupo são os momentos de inércia, os índices de sombra, o volume molecular, entre outros.

- Descritores Electrostáticos: estes descritores reflectem as características da distribuição de carga existente na molécula. As cargas parciais empíricas são calculadas usando a aproximação proposta por Zefirov. Este método baseia-se na escala de electronegatividade de Sanderson e utiliza o conceito que representa a electronegatividade molecular como a média geométrica das electronegatividades atómicas. Neste grupo de descritores estão incluídas as cargas parciais máximas e mínimas para a molécula, as cargas parciais máximas e mínimas para determinados tipos de átomos, o parâmetro de polaridade, entre outros.

- Descritores Químico-Quânticos: estes descritores adicionam informações importantes para os descritores convencionais e podem ser classificados em diferentes classes. Existe a classe de descritores relacionados com a distribuição de carga que representam ou dependem directamente da distribuição de carga calculada com base na química e na quântica e descrevem as interacções polares entre as moléculas ou a sua reactividade química. Outra classe incluída neste grupo de descritores é a classe referente aos descritores relacionados com a valência, estes estão relacionados com a força das interacções das ligações intramoleculares e caracterizam a estabilidade das moléculas, a sua flexibilidade conformacional e outras propriedades relacionadas com a valência. Os descritores relacionados com a energia da mecânica quântica caracterizam a energia total da molécula em diferentes escalas de energia e a distribuição da energia intramolecular utilizando diferentes esquemas de particionamento. A classe de descritores da mecânica quântica relacionados com a rotação e vibração molecular descreve as propriedades vibracionais das ligações individuais da molécula e a energia total de vibração à temperatura de 0K da molécula. Existe ainda uma outra classe pertencente a este grupo de descritores, sendo esta referente aos descritores de solvatação molecular calculados com base nas características quânticas e mecânicas. Exemplos de alguns dos descritores que pertencem a este grupo são o momento dipolar total da molécula, energia total da molécula, energia de solvatação de Born, entre outros.

- Descritores moleculares termodinâmicos: estes descritores são calculados com base na mecânica-quântica e na função de partição total da molécula e nas suas componentes electrónicas, translacionais, rotacionais e vibracionais. Os descritores referentes à

entropia translacional da molécula, à capacidade calorífica vibracional da molécula e à entalpia translacional da molécula são alguns dos descritores incluídos neste grupo de descritores. <sup>[63]</sup>

#### 4.2.2. Aplicações

O QSAR é uma das áreas mais bem desenvolvidas da química computacional. A sua aplicação tem vindo a aumentar nas últimas décadas, provando ser uma técnica de baixo custo e um investimento com elevado retorno. A utilização do QSAR é de elevada importância para o entendimento de diversos aspectos das interacções químicas e biológicas, em relação à investigação de fármacos, pesticidas e na área da toxicologia. Este método é importante para elucidar os mecanismos das interacções químicas e biológicas em diversas biomoléculas, membranas, células, entre outras.

A utilização de modelos de QSAR para rastrear as bases de dados químicas ou bibliotecas virtuais antes da sua síntese é um factor importante para as companhias farmacêuticas, agências governamentais e fabricantes de produtos químicos. <sup>[64]</sup>

Na área farmacêutica, o QSAR também é frequentemente utilizado para estabelecer correlações entre as propriedades estruturais e electrónicas de potenciais candidatos a fármacos e a sua afinidade de ligação a uma macromolécula alvo, sendo também utilizado para prever as propriedades de absorção, distribuição, metabolismo, eliminação e toxicidade (ADMET) ou a biodisponibilidade oral dos compostos. <sup>[3]</sup>

#### 4.2.3. Realização do QSAR

Numa primeira fase dos estudos de QSAR realizados neste trabalho, separou-se o conjunto de 800 compostos em grupos de acordo com a estrutura base dos mesmos. Seguidamente, dividiu-se cada grupo de compostos em dois conjuntos, o de treino e o de teste de acordo com os valores experimentais das actividades biológicas em estudo (IC<sub>50</sub>, GI<sub>50</sub>, TGI, LC<sub>50</sub> e média), tendo sido escolhidos os compostos de forma aleatória de entre cada subconjunto de dados, de forma a obter compostos representativos de cada subconjunto. A selecção dos compostos para os conjuntos de treino e de teste foi

efectuada de modo a garantir a diversidade do conjunto de treino e assegurar que os compostos do conjunto de teste são representativos da base de dados. Os compostos foram classificados em conjunto de treino e de teste numa razão de 4:1.

Para a realização destes estudos lineares de QSAR utilizou-se o programa CODESSA, recorrendo também ao E-Dragon (cálculo do log P e do número de átomos dadores e aceptores de ligações de hidrogénio) e ao MOPAC incluído no programa VEGA ZZ 2.2.0. (cálculo dos parâmetros quânticos e termodinâmicos). Para o cálculo dos parâmetros termodinâmicos utilizou-se as palavras-chave FORCE PRECISE THERMO ROT=X no MOPAC. Quanto aos cálculos quânticos, estes foram efectuados utilizando as palavras-chave VECTORS BONDS PI POLAR PRECISE ENPART.

No programa CODESSA é utilizado um ficheiro de entrada onde constam os nomes das estruturas, dos respectivos ficheiros de estruturas, dos ficheiros com os parâmetros quânticos e termodinâmicos e ainda os valores dos descritores moleculares externos.

Após a introdução deste ficheiro de entrada no programa CODESSA, efectuou-se o cálculo dos descritores moleculares. Seguidamente definiu-se quais as estruturas que pertencem ao conjunto de treino e de teste, e utilizou-se o método heurístico para a selecção dos descritores e para estabelecer correlações entre a actividade biológica em estudo e os descritores moleculares.

Os resultados são obtidos após alguns segundos de processamento. Nos resultados constam a informação acerca da equação dos modelos obtidos, os valores dos testes F e t, os valores da validação cruzada, do RMSE,  $R^2$  dos conjuntos de treino e de teste. Sendo ainda possível aceder aos valores calculados das actividades biológicas calculados através dos modelos obtidos e também os valores residuais (diferença entre os valores calculados e experimentais).

Nestes estudos de QSAR utilizou-se uma *forward feature selection* com MLR para estabelecer os modelos. O valor referente ao teste F foi utilizado para a análise da variância e os parâmetros  $R^2$  e RMSE para o conjunto de treino foram utilizados como critérios de selecção. Utilizou-se um número de descritores que fosse cinco vezes menor que o número de compostos pertencentes ao conjunto de treino. A análise dos modelos foi efectuada utilizando os parâmetros  $R^2$ , RMSE e o valor do teste F.

#### 4.2.4. Método Heurístico

Como referido anteriormente, para a selecção dos descritores moleculares utilizou-se o método heurístico. Este método verifica todos os descritores de modo a assegurar que estão disponíveis os valores para cada descritor e para cada estrutura, e para verificar se existe uma variação nos valores dos descritores. Este método efectua a eliminação dos descritores, descartando os que satisfazem uma das seguintes condições: (a) o descritor não está disponível para todas as estruturas, (b) o descritor possui um valor constante para todas as estruturas. Após este passo, são calculadas as equações de correlação de um parâmetro para cada descritor.

Para reduzir o número de descritores do conjunto inicial foram aplicados os seguintes critérios e a eliminação dos descritores é efectuada quando: (a) o valor do teste F para cada correlação de um parâmetro com o descritor é menor que 1.0; (b) o coeficiente de correlação quadrático para a equação de um parâmetro é menor que  $R_{\min}^2$  (0,01); (c) o parâmetro para o teste t é menor do que  $t_1$  (0,1) (em que  $R_{\min}^2$  e  $t_1$  são valores especificados); (d) o descritor é altamente intercorrelacionado com outro descritor e este outro descritor apresenta um elevado coeficiente de correlação nas equações de um parâmetro baseado nestes descritores. Todos os restantes descritores são ordenados por ordem decrescente de acordo com o coeficiente de correlação da equação de correlação de um parâmetro.<sup>[63]</sup>

O método heurístico é vulgarmente utilizado nos estudos lineares de QSAR e é uma excelente ferramenta para a selecção dos descritores antes da construção dos modelos linear e não linear. As vantagens deste método são a elevada velocidade de processamento e a ausência de restrições de software em relação ao tamanho do conjunto de dados e a sua estratégia única da selecção de variáveis.

Este método fornece uma boa estimativa de uma forma muito rápida acerca da correlação esperada a partir dos dados, ou deriva alguns modelos com melhor regressão. O referido método origina correlações, cerca de 2 a 5 vezes mais rapidamente do que os outros métodos de semelhante qualidade. Adicionalmente, o número máximo de parâmetros no modelo resultante pode ser fixado de acordo com a situação em causa, poupando assim tempo. No entanto, este método está limitado a modelos lineares.<sup>[61]</sup>

### 4.3. Redes Neurais Artificiais (ANN)

As redes neuronais artificiais foram criadas como uma alternativa aos métodos existentes, como o PLS, MLR, entre outros. Este método processa a informação inicial e gera modelos “escondidos” das relações. Algumas das vantagens das redes neuronais artificiais são o facto de estas serem naturalmente capazes de modelar sistemas não lineares, não sendo necessário especificar profundamente. As desvantagens que estas apresentam prendem-se com a tendência de sobreajustar os dados, com o nível significativo de dificuldade em determinar quais descritores são mais significantes no modelo resultante e com a dificuldade em extrair a relação estabelecida na modelação.<sup>[4]</sup>

As redes neuronais artificiais têm sido utilizadas no QSAR desde os anos 80, estas actuam como motores estatísticos, tendo sido usadas para ultrapassar algumas das limitações dos métodos estatísticos tradicionais, já que funcionam de uma forma não linear. Os métodos das redes neuronais artificiais são aplicados para maximizar a forma como os descritores moleculares estão relacionados com a actividade biológica para as séries de compostos.<sup>[4]</sup>

As redes neuronais artificiais são sistemas computacionais e de processamento de informação constituídos por um elevado número de elementos simples e de processamento que estão altamente interconectados e que simulam a estrutura e o modo de funcionamento do sistema nervoso biológico. Quer as redes neuronais biológicas, quer as artificiais, podem consistir num número ilimitado de neurónios. A função das redes neuronais é definida através de diversos factores, como o número e arranjo dos neurónios, as suas interconexões, entre outros.<sup>[65][66]</sup>

O conceito do neurónio artificial baseia-se no neurónio biológico. Cada neurónio artificial possui um determinado número de “entradas”, estas “entradas” apresentam transmissões diferentes, tendo a si atribuído um peso que indica a importância das mesmas. No neurónio, a soma dos pesos das “entradas” é calculado e quando esta soma ultrapassa determinado limiar, a soma é processada utilizando a função de transferência e o resultado é distribuído através das “saídas” para o próximo neurónio artificial.<sup>[65]</sup>

As redes neuronais artificiais estão organizadas vulgarmente em camadas: de entrada, “escondidas” e de saída. A camada de entrada comunica com uma ou mais camadas “escondidas”, onde o actual processamento é efectuado via os pesos das

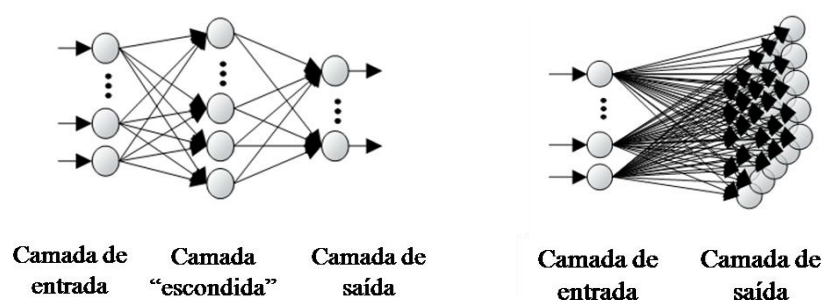
conexões. Todos os neurónios presentes nas camadas “escondidas” conectam-se a todos os neurónios da camada de saída. <sup>[66]</sup>

A aprendizagem efectuada nas redes neuronais artificiais é acompanhada através de algoritmos de treino que são desenvolvidos com base nas regras de aprendizagem, tentando imitar os mecanismos de aprendizagem dos sistemas biológicos. <sup>[66]</sup>

Neste método existem diversos valores (pesos, *bias*) que têm de ser estabelecidos, por esta razão, foram desenvolvidos diversos algoritmos de adaptação para este propósito. Estes algoritmos estão divididos em dois grupos básicos: supervisionados e não supervisionados. <sup>[65]</sup>

O algoritmo supervisionado requer o conhecimento da “saída” desejada e consegue calcular a “saída” com os correntes pesos e *bias*. A “saída” é comparada com a saída alvo e os pesos e *bias* são ajustados pelo algoritmo. Este ciclo é repetido até à diferença entre os valores alvos e calculados ser suficientemente pequena. Os algoritmos supervisionados mais utilizados baseiam-se nos métodos de gradiente (como a *back-propagation*) e nos genéticos (algoritmos genéticos). As redes neuronais artificiais supervisionadas mais utilizadas são as redes neuronais artificiais de *feed-forward* e as redes neuronais *radial basis function*.

O algoritmo não supervisionado não necessita de conhecer os valores de “saída”, produzindo os seus próprios valores de “saída”, sendo estes avaliados posteriormente. Um exemplo deste tipo de algoritmo é a aprendizagem de Kohonen utilizada nos *self-organising maps*. <sup>[65]</sup>



**Figura 7:** Exemplos de estruturas de redes neuronais artificiais, *Multilayer Perceptrons* (à esquerda) e *Kohonen self-organising maps* (à direita). <sup>[65]</sup>

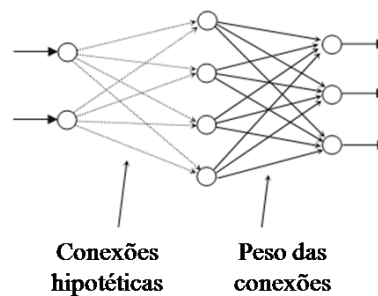
As redes neuronais artificiais utilizam diversos métodos, como *Radial Basis Function* (RBF) e *Multilayer Perceptrons* (MLP), encontrando-se alguns destes métodos explicados seguidamente. <sup>[4]</sup>

### 4.3.1. Radial Basis Function (RBF)

As redes RBF foram propostas e utilizadas como sendo uma alternativa às redes MLP para muitos problemas de engenharia. A arquitectura das redes RBF é muito semelhante às redes MLP apresentando, tal como estas, três camadas designadas de entrada, “escondida” e saída. Os neurónios dentro de cada camada estão completamente conectados aos neurónios da camada anterior. <sup>[66]</sup>

A camada de entrada não processa a informação, apenas distribui os vectores de entrada para a camada “escondida”. As conexões entre as camadas de entrada e “escondida” não possuem qualquer coeficiente de peso. <sup>[61][66]</sup>

Os neurónios presentes na camada “escondida” recebem as variáveis de entrada inalteradas. Cada neurónio presente na camada “escondida” aplica a *radial basis function* como a função de transferência não linear para trabalhar os dados de entrada. Um exemplo deste tipo de redes neuronais está representado na figura que se segue. <sup>[61][66]</sup>



**Figura 8:** Estrutura típica de uma rede neural do tipo RBF. <sup>[66]</sup>

Em geral, existem diversas *radial basis functions* (RBF): linear, cúbica, *thin plate spline*, Gaussiana, multi-quadrática e multi-quadrática inversa. A função mais utilizada na RBF é a função Gaussiana que é caracterizada pelo centro ( $c_j$ ) e pelo raio ( $r_j$ ). A transformação não linear que ocorre através da RBF na camada não linear é dada através da seguinte expressão:

$$h_j(x) = \left( \frac{-\|x - c_j\|^2}{r_j^2} \right) \quad \text{Equação 4}$$

Em que  $h_j$  é a notação para o resultado de saída da unidade de RBF para  $j$ ,  $c_j$  e  $r_j$  são o centro e o raio do RBF  $j$ , respectivamente. O funcionamento da camada de saída é linear e é obtida através da seguinte equação:

$$y_k(x) = \sum w_{kj} h_j(x) + b_k \quad \text{Equação 5}$$

Onde  $y_k$  é unidade de saída  $k$  para o vector de entrada  $x$ ,  $w_{kj}$  é o peso da conexão entre a unidade de saída  $k$  e a unidade da camada “escondida”  $j$  e  $b_k$  é o *bias*. O procedimento de treino quando é utilizada a RBF envolve a selecção de centros, raio e pesos. <sup>[61]</sup>

As redes *Radial Basis Function* (RBF) combinam uma camada “escondida” radial única com uma camada de saída de produto interno. A camada “escondida” de neurónios actua como centros de clusters, agrupando casos de treino semelhantes e a camada de “saída” forma uma função discriminante ou uma regressão. A transformação de *clustering* é não linear e a camada de saída é linear, formando assim uma função não linear. <sup>[67]</sup>

O processo de treino deste tipo de redes é constituído por duas etapas, numa primeira fase é efectuada a atribuição dos centros radiais e dos seus desvios, numa segunda fase é efectuada a optimização da camada de “saída”. Uma RBF clássica utiliza a função de activação identidade na camada de saída, em que pode ser utilizada a optimização linear (pseudo-inversa, SVD), sendo relativamente mais rápida do que o treino de MLP. Nos problemas de classificação, uma função de erro de entropia é combinada com a função de activação não linear, sendo utilizado o algoritmo gradiente conjugado descendente que é mais lento.

O procedimento padrão de treino envolve a selecção de algoritmos para a atribuição dos centros e dos desvios, com a etapa de optimização da “saída” assumida. <sup>[67]</sup>

#### 4.3.2. *Multilayer Perceptrons* (MLP)

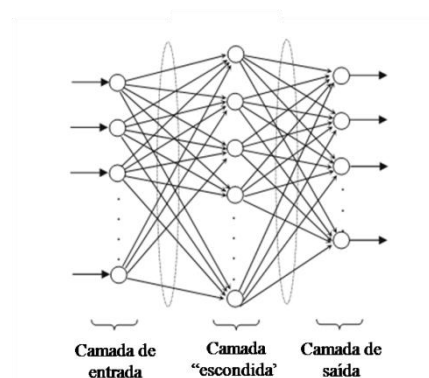
Este é um dos mais populares tipos de redes e em muitos problemas este método oferece o melhor desempenho. Estas redes são treinadas utilizando algoritmos iterativos, sendo o mais conhecido a *back-propagation*.

Tem sido desenvolvida muita investigação no sentido de melhorar os algoritmos para o treino de MLP, os mais usados destes algoritmos são os algoritmos de optimização de segunda ordem (gradiente conjugado descendente, *quasi-Newton* e

Levenberg-Marquardt). Estes algoritmos são vulgarmente descritos como convergindo mais rapidamente do que a *back-propagation* (uma ou duas ordens de grandeza mais rápido).<sup>[67]</sup>

Se os dados do treino são dispersos para a complexidade da função de base, seguidamente a MLP pode passar por outro problema das redes neuronais, que é o sobreajuste (as redes neuronais estimam uma função “sobre-complexa”, modelando o ruído no conjunto de dados preferencialmente do que na função de base). O sobreajuste pode ser minimizado aplicando a regularização dos pesos (que penaliza os pesos grandes que correspondem a funções complexas), e pela paragem precoce (verificação dupla do desempenho da rede contra a selecção do subconjunto de dados durante o treino). Os algoritmos de treino de segunda ordem são mais eficazes do que a *back-propagation*, estando mais inclinados para o sobreajuste dos dados, que é a segunda razão porque algumas vezes o algoritmo *back-propagation* simples prova ser superior.<sup>[67]</sup>

As MLP cobrem um grande grupo de redes neuronais *feed-forward* com uma ou mais camadas de neurónios. Na maioria das aplicações as redes MLP apresentam três camadas para além das camadas de entrada e de saída utilizadas. Os neurónios presentes na camada de entrada apresentam uma função de activação linear, mas algumas funções de activação não lineares como as funções logarítmicas e as tangentes da função sigmoideal são utilizadas nos neurónios das camadas “escondida” e de saída. Um esquema da arquitectura para a rede MLP possuindo uma camada “escondida” está seguidamente representado.<sup>[66]</sup>



**Figura 9:** Estrutura de uma rede neural do tipo MLP com uma camada “escondida”.<sup>[66]</sup>

### 4.3.3. Aplicações

Este método é aplicado em diversas áreas como a previsão da toxicidade de misturas complexas, na química para a optimização de processos de separação, a classificação das reacções químicas, a optimização do método de HPLC na análise de vinho, as relações quantitativas entre a estrutura e a actividade, para a previsão da estrutura secundária das proteínas, para a correlação entre o espectro de infravermelho e a estrutura dos compostos, para a previsão das propriedades das superfícies moleculares, entre outras.

Relativamente à aplicação das redes neuronais artificiais no estabelecimento de relações quantitativas entre a estrutura e a actividade, estas têm vindo a ganhar proeminência uma vez que estes podem ser desenvolvidos em modelos complexos. Particularmente na investigação farmacêutica e no desenvolvimento de muitas investigações nas relações entre a estrutura e a actividade biológica. <sup>[59][65][66]</sup>

As redes neuronais artificiais também têm sido muito aplicadas a problemas de engenharia como a classificação, reconhecimento, estimativa e controlo. <sup>[66]</sup>

### 4.3.5. Realização das Redes Neuronais Artificiais

Neste trabalho, para a aplicação das redes neuronais artificiais, foi utilizado o programa Statistica 7. Para desenvolver as redes neuronais utilizou-se os valores experimentais das actividades biológicas em estudo ( $IC_{50}$ ,  $GI_{50}$ , TGI e  $LC_{50}$ ), os valores dos descritores moleculares que apresentaram uma maior correlação com a propriedade quando aplicado o método heurístico e, utilizaram-se ainda os mesmos compostos nos conjuntos de treino e de teste.

Através do programa Statistica aplicou-se um método de regressão rápido, utilizando como análise o *Intelligent Problem Solver*. Para prosseguir é necessário definir os tipos de variáveis: as variáveis de saída (*Continuous Output*) são as actividades biológicas em estudo, as variáveis de entrada (*Continuous Input*) correspondem aos descritores moleculares e a variável do subconjunto corresponde à indicação dos conjuntos existentes (conjunto de treino e de teste).

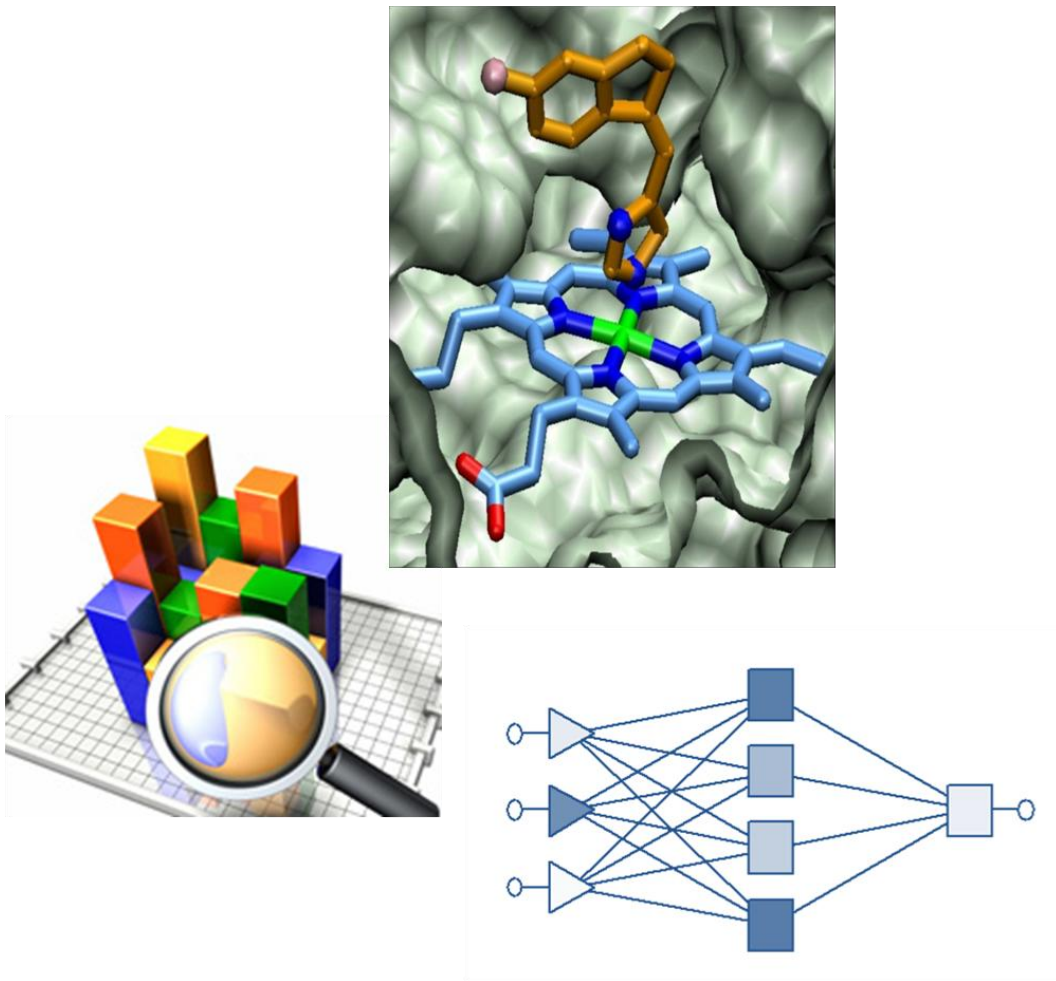
Numa fase seguinte é necessário seleccionar que tipos de redes neuronais artificiais irão ser testadas e o nível de complexidade que irá ser usado para as redes neuronais testadas, estando este relacionado com o número de unidades presentes na camada “escondida” que irão ser testados. Nesta situação seleccionou-se os tipos Linear, *Radial Basis Function* e *Multilayer Perceptrons*, tendo sido testadas as redes neuronais com a variação do número de unidades na camada “escondida” de 1 a 10 para a RBF e de 1 a 9 para a MLP e seleccionou-se ainda a opção para reter os cinco melhores modelos.

O programa processa rapidamente, demorando apenas alguns segundos para indicar os resultados. Através deste método pode-se aceder a um resumo dos modelos criados, em que é indicado cada um dos melhores modelos, o perfil da rede neuronal, o desempenho e o erro dos conjuntos utilizados, os métodos utilizados para realizar o treino das respectivas redes, o número de unidades de entrada e o número de unidades existentes nas camadas “escondidas”. Outro dos resultados obtidos é os valores das previsões e dos residuais (diferença entre o valor de actividade biológica experimental e calculada) das actividades biológicas em estudo para cada um dos compostos obtidas através dos diversos modelos construídos. Também é possível aceder à análise de sensibilidade de cada um dos descritores moleculares utilizados, sendo possível, através destes resultados, determinar a significância de cada descritor para cada modelo.

Existem ainda outros parâmetros que podem ser obtidos através deste método como a média dos dados, o desvio padrão, os valores de erro para a média e para o desvio padrão, entre outros. Para além dos dados numéricos também é possível obter algumas representações, como as representações das redes neuronais obtidas e representações gráficas da influência dos descritores moleculares na actividade, entre outras. As correlações entre os dados experimentais e previstos são obtidas através da construção das representações gráficas utilizando os valores retirados do programa Statistica.

Os métodos utilizados para treinar as redes neuronais artificiais obtidas neste trabalho foram: *K-Means* (Atribuição de Centro); *K-Nearest Neighbour* (Atribuição de desvio); Pseudo-Inversa (optimização linear por mínimos quadrados); *Back-Propagation* e gradiente conjugado descendente.

# CAPÍTULO V



# RESULTADOS E DISCUSSÃO

## Capítulo V – Resultados e Discussão

### 5.1. Docking

#### 5.1.1. Distribuição de Energia

Após a realização do *docking* dos 807 compostos com cada uma das proteínas, os resultados obtidos foram ordenados de acordo com as suas energias de interação, tendo sido obtidas representações gráficas com um perfil típico. As figuras seguidamente apresentadas representam alguns exemplos da distribuição de energias.

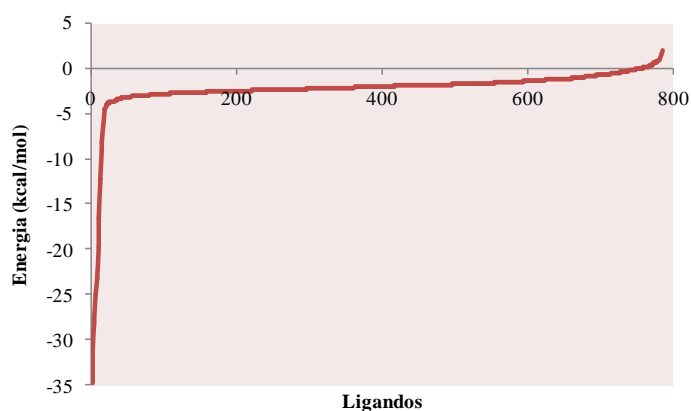


Figura 10: Distribuição das energias obtidas para a enzima CDC25C.

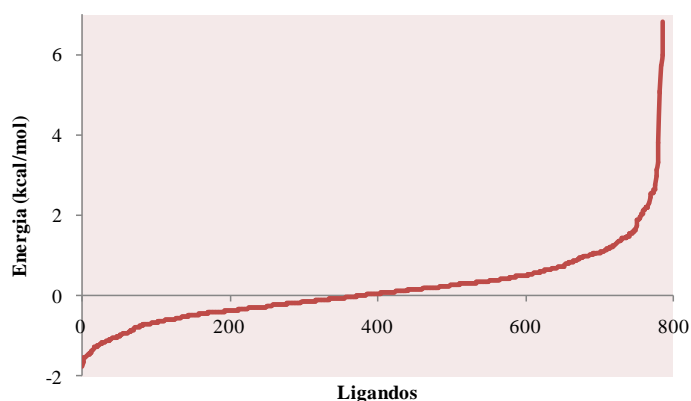


Figura 11: Distribuição das energias obtidas para a enzima ERK1.

Através da representação da distribuição das energias obtidas para o *docking* é possível observar quais os ligandos que apresentam melhores energias de interação. Quando na distribuição da energia existe um declive acentuado da linha de distribuição para a parte negativa da representação, significa que para a enzima se obtiveram bons

valores de energias de interacção, encontrando-se os melhores ligandos nesta zona de maior acentuação do declive. Um exemplo desta situação pode ser observado para a distribuição das energias de interacção para a enzima CDC25C, representada na figura 10, verificando-se também nas distribuições de energia obtidas para as proteínas CDC25B, p53, SKP1, CHK1, Tubulina, FNTB, HO-1, 5-LOX, EF-G (bactéria), entre outras proteínas.

Quando, pelo contrário, não existe uma acentuação da linha de distribuição para a parte negativa da representação gráfica, a enzima apresenta valores de energias de interacção mais desfavoráveis. Algumas das proteínas que apresentam este tipo de distribuição de energia são, por exemplo, as proteínas ERK1, AKT1, CCNH, p27 KIP1, CENP-A, Actina, iNOS.

### 5.1.2. Constante de Dissociação

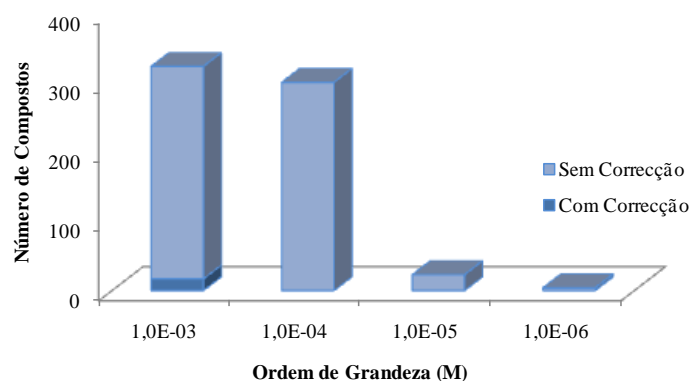
A constante de dissociação ( $K_D$ ) pode ser obtida a partir dos valores de energia obtidos ( $\Delta G^0$ ) com a realização do *docking*. A variação da energia livre de Gibbs padrão ( $\Delta G^0$ ) fornece informação sobre a estabilidade da ligação do complexo ligando-receptor, quanto menor for a variação da energia livre de Gibbs padrão, maior é a estabilidade do complexo.

Para a dissociação de um complexo ligando-receptor ( $R + L \leftrightarrow RL$ ) é possível relacionar a  $K_D$  com a variação da energia livre de Gibbs através das seguintes equações:<sup>[7]</sup>

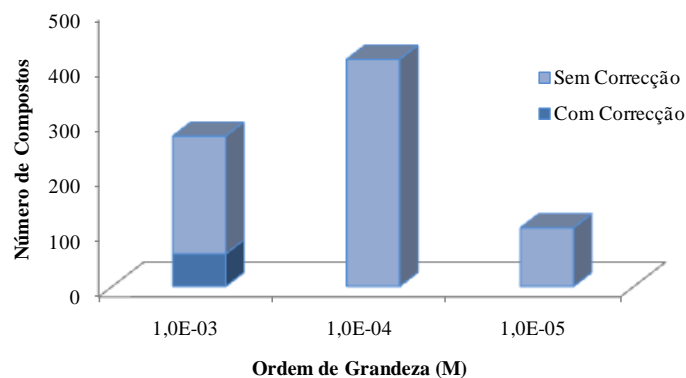
$$K_D = \frac{[R][L]}{[RL]} \quad \text{Equação 6}$$

$$\Delta G^0 = RT \ln K_D \quad \text{Equação 7}$$

Utilizando os valores de energia de interacção obtidos através do *docking* ( $\Delta G^0$ ), efectuou-se o cálculo da constante de dissociação e comparou-se os resultados obtidos antes e após a correcção dos valores de *docking* obtidos em relação a todas as proteínas utilizadas. Nas figuras que se seguem estão representados alguns dos resultados obtidos.



**Figura 12:** Representação gráfica do número de compostos que apresentam valores de constante de dissociação de diferentes ordens de grandeza para a enzima GSK-3B.



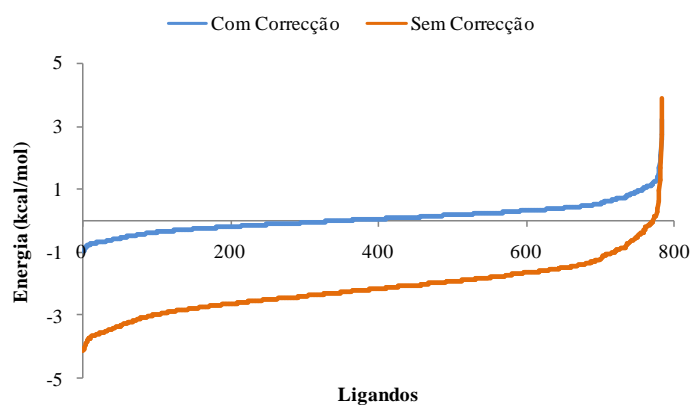
**Figura 13:** Representação gráfica do número de compostos que apresentam valores de constante de dissociação de diferentes ordens de grandeza para a enzima ERK2.

Nas representações gráficas anteriormente apresentadas, é possível observar que após a correção, os valores da constante de dissociação alteram-se significativamente, diminuindo o número de compostos que apresentam valores para a constante de dissociação na ordem dos micromolar ou inferiores. Os valores de constante de dissociação antes da correção encontram-se maioritariamente na ordem dos micromolar, apresentando alguns compostos valores abaixo desta grandeza. No entanto, quando analisados os valores após a correção existe um baixo número compostos que apresentem valores de constante de dissociação abaixo dos micromolar. Quer antes, quer após a correção, os compostos que interagem com a enzima CDC25C são aqueles que apresentam os valores mais baixos para a constante de dissociação. Para os casos

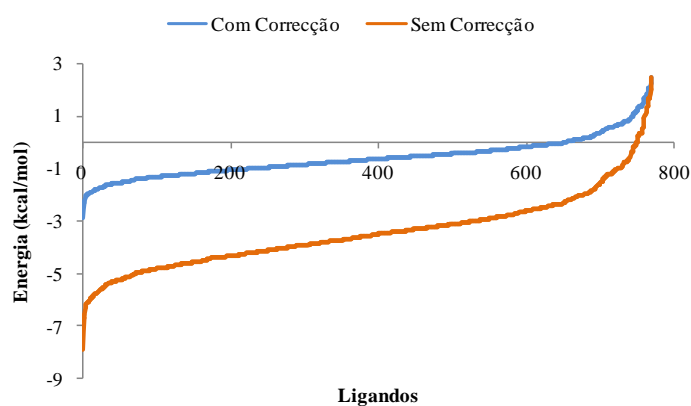
em que os valores da constante de dissociação são superiores aos micromolar, os estudos não devem ser prosseguidos.

### 5.1.3. Correção dos Resultados de *Docking*

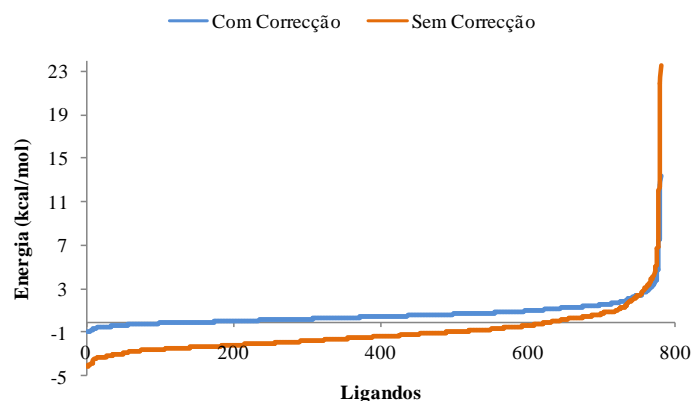
Após a obtenção dos resultados do *docking*, foi realizada a sua correção utilizando as equações 1-3, de modo a efectuar a correção de cada ligando em relação a todas as enzimas. As figuras seguidamente apresentadas, representam a comparação entre as distribuições de energia antes e após a correção.



**Figura 14:** Distribuição da energia antes e após a correção para a enzima AKT1.



**Figura 15:** Distribuição da energia antes e após a correção para a enzima CHK1.



**Figura 16:** Distribuição da energia antes e após a correção para a enzima CDC2.

Pela observação das representações gráficas anteriores, verifica-se que ao ser efectuada a correção o intervalo dos valores de energia diminui. No entanto, não se verificou uma reordenação significativa dos resultados, ou seja, os melhores ligandos antes da correção continuam a ser os melhores ligandos após a correção, verificando-se o mesmo com os piores ligandos.

Os gráficos da distribuição das energias obtida após a correção para os diversos alvos em estudo encontram-se representados no Anexo B.

#### 5.1.4. Cálculo do Erro (MRE)

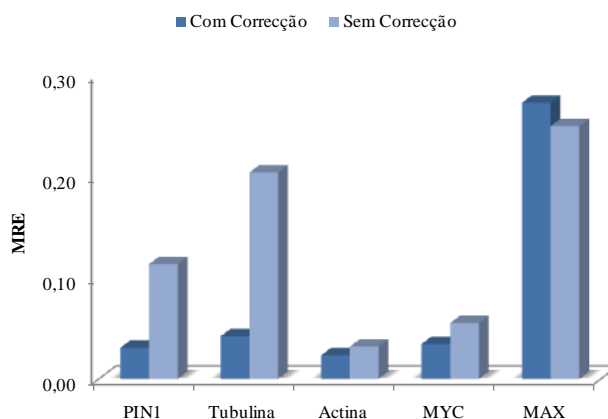
O erro MRE (*Mean Ranking Error*) é calculado com base nas energias obtidas  $S_{ij}$  para diversos ligandos ( $i$ ) no centro activo ( $j$ ) através da seguinte equação:

$$Err_j = (S_{jmelhor} - S_{jj}) / (S_{jmelhor} - S_{jpior}) \quad \text{Equação 8}$$

Em que o  $S_{jj}$  é a energia para o ligando cognato ( $j$ ) no centro activo ( $j$ ),  $S_{jmelhor}$  é a melhor energia de qualquer ligando no centro ( $j$ ) e  $S_{jpior}$  é a pior energia para qualquer ligando. A diferença entre a melhor e a pior energia para qualquer ligando fornece a gama de energias para todos os ligandos no centro activo e  $(S_{jmelhor} - S_{jj})$  é a diferença entre a melhor energia obtida para o centro activo e a energia para o ligando cognato. Assim,  $Err_j = 0$  para o ligando cognato que tem a melhor energia de todos os ligandos testados e  $Err_j = 1$  para a pior energia apresentada. Este erro também pode ser calculado

usando o valor médio de  $Err_j$  ao longo de todos os centros activos ( $j$ ). Um valor de erro de 0,0 significa que os valores de energias são perfeitos, enquanto que um valor de 0,5 indica um valor de energia médio. [55]

Na figura seguinte está representada a comparação entre os valores referentes ao erro calculado antes e após a correcção dos valores de energia para algumas das enzimas em estudo. Os gráficos representativos dos valores de erro obtidos para as restantes enzimas em estudo constam do Anexo C.



**Figura 17:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correcção.

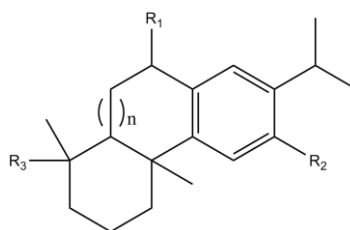
Com base na representação gráfica anterior, constata-se que os valores de erro obtido antes da correcção são, em geral, superiores aos calculados utilizando os valores de energia corrigidos. Existem alguns casos em que os valores de erro são superiores quando calculados utilizando os valores de energia com a correcção.

Para a maioria das enzimas, os valores de erro são baixos sendo, por vezes, muito próximos de zero. No entanto, nalguns casos, os valores de erro encontram-se próximos do valor de energia médio (0,5).

### 5.1.5. Preferências das Enzimas

Após a observação dos resultados obtidos para o *docking* entre as diversas enzimas e os compostos em estudo verificamos que existia preferência de interacção entre algumas enzimas e determinado tipo de compostos. Desta forma, foram determinadas estruturas base de alguns compostos com os quais as enzimas interagem preferencialmente.

Analisando os resultados obtidos para a enzima AMD1, observou-se que esta interage preferencialmente com compostos que possuem a estrutura base que seguidamente é apresentada.

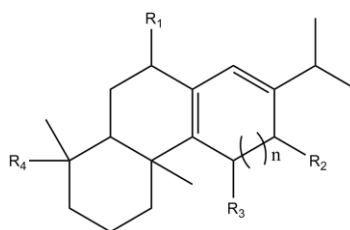


**Figura 18:** Estrutura base dos compostos com os quais a enzima AMD1 interage preferencialmente.

**Tabela 4:** Substituintes para a estrutura base representada na figura 18.

<b>R<sub>1</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>3</sub></b>	<b>n</b>
H <sub>2</sub>	H	COOH	Simples
O	COCH <sub>3</sub>	COOCH <sub>3</sub>	Simples
H <sub>2</sub>	COCH <sub>3</sub>	COOCH <sub>3</sub>	Simples
H <sub>2</sub>	OCOCH <sub>3</sub>	COOCH <sub>3</sub>	Simples
O	OCH <sub>3</sub>	CH <sub>3</sub>	Simples
O	OCH <sub>3</sub>	CH <sub>3</sub>	Dupla

Relativamente à enzima WEE1, esta exibiu uma interacção preferencial ao interagir com compostos com a estrutura base que se segue.

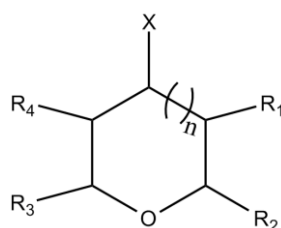


**Figura 19:** Estrutura base dos compostos com os quais a enzima WEE1 interage preferencialmente.

**Tabela 5:** Substituintes para a estrutura base representada na figura 19.

$R_1$	$R_2$	$R_3$	$R_4$	n
O	$C_9H_{11}N_4O_2$	H	$CH_3$	Dupla
$H_2$	$OCH_3$	H	$CH_2OH$	Dupla
O	OH	H	$CH_3$	Dupla
$H_2$	O	O	$CH_3$	Simple
O	$OCH_3$	H	$CH_3$	Dupla

Por sua vez, a enzima CDC25B apresentou uma preferência de interação com os compostos que possuem a estrutura base que se segue.



**Figura 20:** Estrutura base dos compostos com os quais a enzima CDC25B interage preferencialmente.

**Tabela 6:** Substituintes para a estrutura base representada na figura 20.

$R_1$	$R_2$	$R_3$	$R_4$	X	n
H	$C_4H_5$	H	H	Cl	Dupla
$H_2$	$C_6H_5$	H	H	Cl	Simple
$H_2$	$C_6H_5NO_2H$	H	H	Cl	Simple
$Si(CH_3)_3$	$C_4H_5$	$CH_3$	$CH_3$	Cl	Dupla
$H_2$	$C_6H_5OCH_3$	$C_4H_9$	H	Cl	Simple
$Si(CH_3)_3$	$C_4H_9$	H	H	Cl	Dupla
H	$C_6H_{11}$	$CH_2CN$	H	Cl	Dupla

Si(CH <sub>3</sub> ) <sub>3</sub>	C <sub>4</sub> H <sub>7</sub>	CH <sub>3</sub>	CH <sub>3</sub>	Cl	Dupla
Si(CH <sub>3</sub> ) <sub>3</sub>	C <sub>4</sub> H <sub>9</sub>	CH <sub>3</sub>	H	Br	Dupla
Si(CH <sub>3</sub> ) <sub>3</sub>	C <sub>4</sub> H <sub>5</sub> Br <sub>2</sub>	CH <sub>3</sub>	CH <sub>3</sub>	Cl	Dupla

### 5.1.6. Estudo da Especificidade

Para o estudo da especificidade seleccionamos uma das famílias de enzimas que actua quer nas células saudáveis quer nas células cancerígenas, as cinases. Efectuamos a comparação das constantes de dissociação apresentadas pelas cinases que actuam quer nas células saudáveis, quer cancerígenas e uma das cinases que actuam especificamente nas células cancerígenas, a AURKA. Para esta comparação foram escolhidos os ligandos que ao interagirem com a AURKA apresentaram melhores energias de interacção, e realizamos o *docking* entre estes ligandos e algumas das cinases que actuam em células cancerígenas e em células normais.

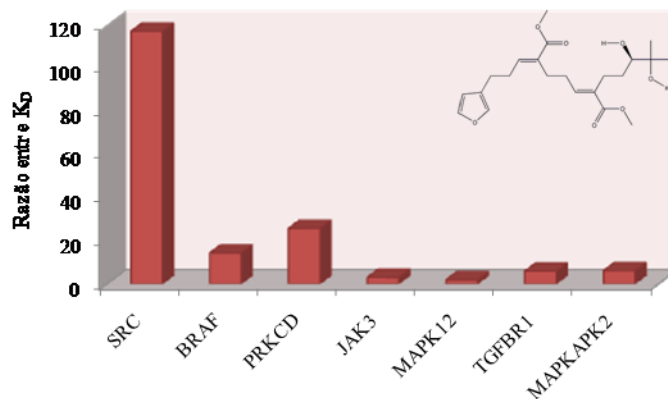
As cinases utilizadas para estabelecer uma comparação em termos de especificidade com a cinase AURKA encontram-se representadas na seguinte tabela juntamente com a família à qual pertencem.

**Tabela 7:** Cinases utilizadas no estudo da especificidade e respectiva família. <sup>[68]</sup>

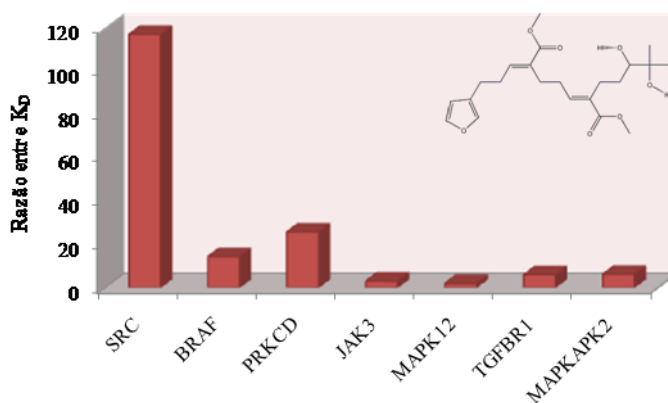
Família	Cinase
JAK	JAK3
Src	SRC
RAF	BRAF
TGBFR	TGFBR1
Sub-família p38	MAPK12
PKC	PRKCD
MAPKAP	MAPKAPK2

Após a realização do *docking*, foi calculada a proporção entre os valores das constantes de dissociação das cinases que actuam quer nas células saudáveis quer nas

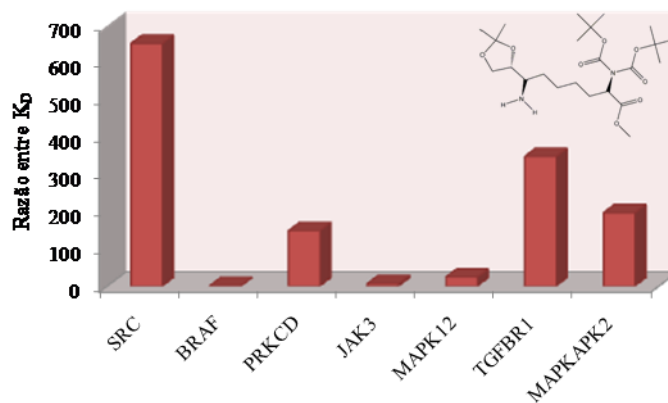
células cancerígenas e da cinase AURKA. Desta forma, determinamos a especificidade dos compostos em relação às cinases estudadas, os resultados obtidos encontram-se representados nas figuras seguintes.



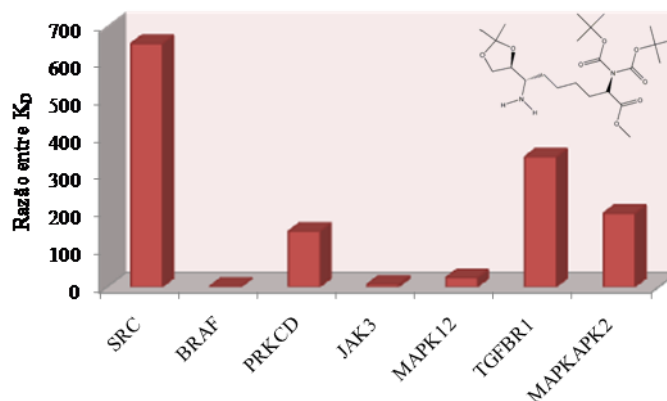
**Figura 21:** Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para VJH-6.



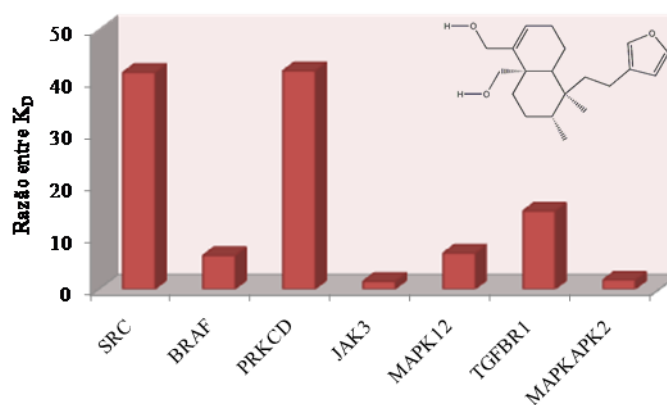
**Figura 22:** Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para VJH-8.



**Figura 23:** Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para NHR-38.



**Figura 24:** Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para NHR-47.



**Figura 25:** Valores da proporção entre as constantes de dissociação da cinase AURKA e as famílias de cinases estudadas para VJH-21.

Neste tipo de estudos é importante determinar a especificidade dos compostos em relação a alvos diferentes daqueles que são objecto de estudo, utilizando por exemplo enzimas que actuem tanto em células saudáveis como em células cancerígenas. O estudo da especificidade é realizado para poder prever a ocorrência de efeitos indesejáveis, devendo os compostos apresentar uma maior especificidade em relação ao alvo desejado do que em relação a outros alvos.

Com base nos resultados obtidos, verifica-se que os compostos estudados apresentam uma especificidade muito superior para com a cinase AURKA quando comparados com determinadas famílias de cinases estudadas, como são o caso das famílias Src, PKC, TGBFR e MAPKAP. Apresentando os compostos em média uma especificidade 316, 78, 144 e 81 vezes superior para a cinase AURKA do que para as cinases das famílias Src, PKC, TGBFR e MAPKAP, respectivamente.

Em relação a algumas das famílias de cinases estudadas, os compostos apresentaram uma especificidade semelhante àquela que apresentam para a cinase AURKA. Esta situação verifica-se em relação às famílias RAF, JAK e p38, uma vez que os compostos apresentam em média, respectivamente, uma especificidade de 7, 4 e 12 vezes superior para a cinase AURKA do que para as famílias de cinases referidas anteriormente.

### 5.1.7. Testes de Enriquecimento

Foram realizados testes de enriquecimento, construindo para este fim uma base de dados de ligandos para cada uma das enzimas submetidas a este teste. Esta base de dados é constituída por 1000 ligandos falsos e por um conjunto de 20 ligandos (ligandos conhecidos e os melhores ligandos determinados por *docking* em relação a cada enzima). Os ligandos falsos foram obtidos a partir da base de dados DUD, *Directory of Useful Decoys*, tendo sido seleccionados com base nas suas propriedades (peso molecular, o número de aceitadores de ligações de hidrogénio, o número de doadores de ligações de hidrogénio e o número de ligações com rotação livre), devendo estas serem semelhantes às dos ligandos conhecidos. [69]

Realizou-se o *docking* entre as enzimas e os compostos da base de dados, tendo sido posteriormente classificados os compostos de acordo com as energias de interacção. Os resultados obtidos para algumas das enzimas estudadas encontram-se seguidamente representados, no entanto, no Anexo D encontram-se as representações gráficas para os restantes testes de enriquecimento realizados.

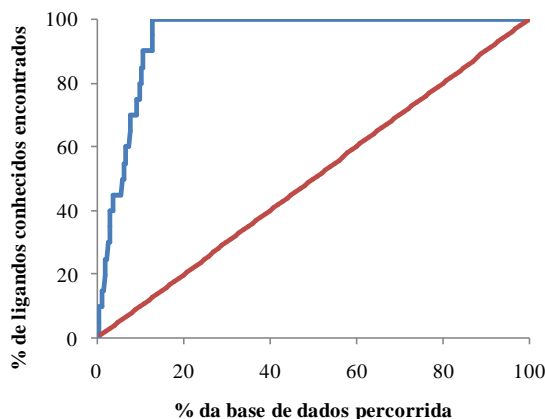
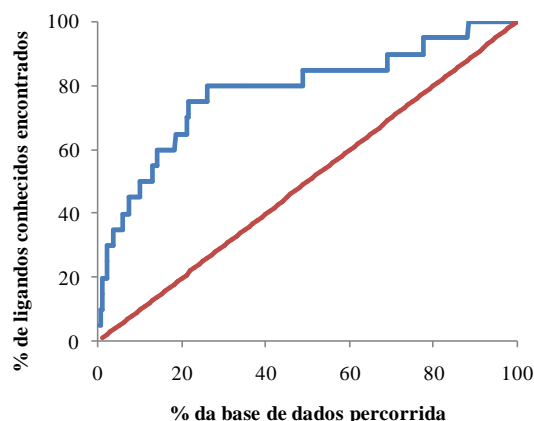


Figura 26: Teste de Enriquecimento realizado para a enzima CDC7.



**Figura 27:** Teste de Enriquecimento realizado para a enzima CDK4.

Nos gráficos acima representados a linha vermelha exemplifica uma situação na qual a selecção dos ligandos é realizada de uma forma aleatória, enquanto que a linha azul indica a representação obtida para cada caso estudado. Os resultados obtidos realizando o teste de enriquecimento são tanto melhores quanto mais para a zona esquerda do gráfico se situar a linha azul.

Através da análise das representações obtidas, verifica-se que a realização do teste de enriquecimento provocou um enriquecimento de aproximadamente seis vezes para a enzima CDC7, já que foram seleccionados 80% dos ligandos conhecidos percorrendo cerca de 15% da base de dados. No entanto, o enriquecimento obtido para a enzima CDK4 foi muito menos significativo, tendo provocado um enriquecimento de cerca de duas vezes até atingir 50% da base de dados, uma vez que ao percorrer esta percentagem da base de dados seleccionou 80% dos ligandos conhecidos. A partir dos 50% da base de dados percorrida, o enriquecimento sofre um decréscimo acentuado.

Analisando os resultados dos testes de enriquecimento efectuados para as enzimas em estudo é possível observar que existem alguns casos em que não se verificou um enriquecimento elevado, como é o caso das enzimas NA, RNAP (bactéria) e TOPOII.

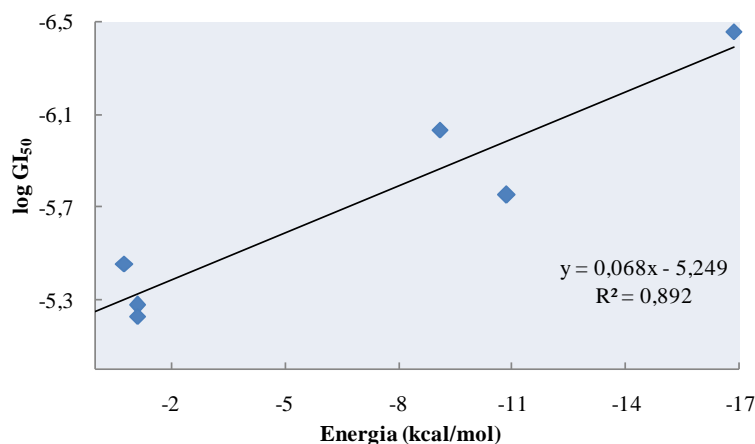
### 5.1.8. Correlação entre os Dados Computacionais e Experimentais

Após a realização do *docking*, procedemos à verificação da existência ou não de correlação entre os resultados computacionais e os dados experimentais.

Experimentalmente foram estudados diversos factores, como a concentração necessária para a inibição de 50% da amostra ( $IC_{50}$ ), a concentração necessária para induzir a morte de 50% da amostra ( $LC_{50}$ ), a concentração necessária para inibir 50% do crescimento da amostra ( $GI_{50}$ ) e a concentração necessária para a inibição total do crescimento (TGI). Estes testes celulares foram realizados para diversas linhas celulares relacionadas com diversos tipos de cancro, entre os quais a leucemia (HL60), o cancro no ovário (A2780), o cancro no pulmão (SW1573), o carcinoma mamário (T-47D) e o cancro no cólon (WiDr).

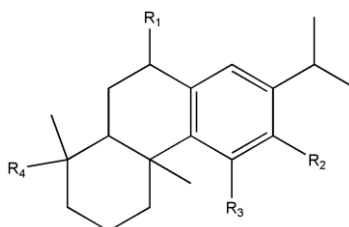
A correlação entre os resultados experimentais e os dados computacionais é de difícil estabelecimento. Computacionalmente estudamos apenas a interacção entre os compostos e as proteínas, enquanto que experimentalmente existem vários factores que influenciam os resultados, como a solubilidade dos compostos e a difusão para o interior das células. Apesar de ser difícil estabelecer uma correlação entre os resultados experimentais e os resultados obtidos computacionalmente, foi realizada uma tentativa de correlacionar estes dados. Nesta tentativa foram seleccionados os 20 ligandos que apresentaram melhores energias de interacção com cada uma das enzimas estudadas. Se a correlação entre os dados experimentais e as energias de *docking* produzisse um valor para o coeficiente de correlação quadrático superior a 0,6 aceitava-se esse valor, caso contrário eram eliminados os pontos que apresentavam valores de  $\log IC_{50}$  (ou outra medida), nos testes celulares, superiores a -4 (este limite era reduzido em fracções de -0.1 se não fosse obtido um valor para o coeficiente de correlação quadrático superior a 0.6) até ser estabelecidas as correlações. Se o valor de  $R^2$  não fosse alcançado com um mínimo de 5 pontos, considerava-se que não existia correlação.

Seguidamente são apresentadas algumas das representações gráficas obtidas aquando do estudo da correlação e as estruturas base identificadas. As restantes correlações obtidas entre os dados computacionais e os experimentais encontram-se representadas no Anexo E.



**Figura 28:** Correlação entre os dados computacionais e experimentais para a enzima CDC25C para a linha celular A2780.

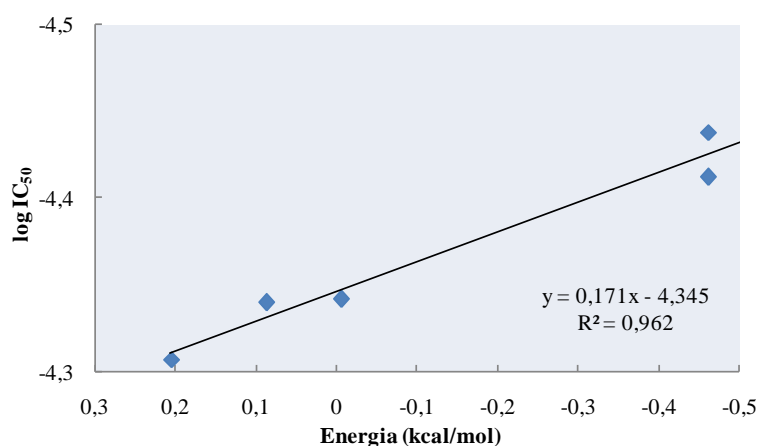
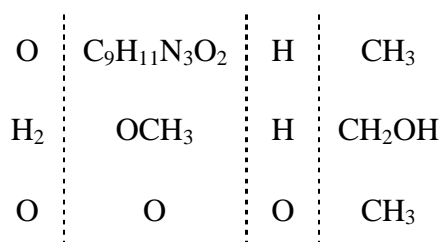
Calculando as correlações entre os dados experimentais e computacionais para a linha celular A2780, utilizando os diversos parâmetros de actividade ( $IC_{50}$ ,  $GI_{50}$ , TGI e  $LC_{50}$ ), obteve-se correlações para diversas enzimas como a CDC25C, E2F, p53, CDC2, DD-Ligase, TOP2A, entre outras. Quanto aos coeficientes de correlação obtidos, estes encontram-se entre 0,609 e 0,892. Nestas correlações participam compostos com a estrutura base que se encontra representada na figura seguinte.



**Figura 29:** Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular A2780.

**Tabela 8:** Substituintes para a estrutura base representada na figura 29.

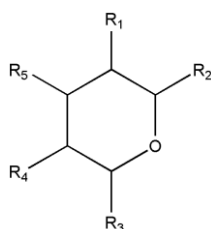
<b>R<sub>1</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>3</sub></b>	<b>R<sub>4</sub></b>
O	C <sub>3</sub> H <sub>5</sub> O <sub>2</sub>	H	CH <sub>3</sub>
O	C <sub>11</sub> H <sub>16</sub> NO <sub>3</sub>	H	CH <sub>3</sub>
O	C <sub>17</sub> H <sub>20</sub> NO <sub>3</sub>	H	CH <sub>3</sub>



**Figura 30:** Correlação entre os dados computacionais e experimentais para a enzima PAK1 para a linha celular HL60.

Para a linha celular HL60, estabelecemos diversas correlações entre os dados experimentais (IC<sub>50</sub>, GI<sub>50</sub>, TGI e LC<sub>50</sub>) e computacionais com base nos dados obtidos para o *docking* para as enzimas PAK1, CDK7, PIN1, ROCK, COX, DHPS, entre outras. Os coeficientes de correlação obtidos para esta linha celular encontram-se entre 0,612 e 0,962.

A estrutura base dos compostos que estão envolvidos nestas correlações está seguidamente representada.



**Figura 31:** Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular HL60.

Tabela 9: Substituintes para a estrutura base representada na figura 31.

R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>
C <sub>9</sub> H <sub>11</sub> SO <sub>4</sub>	C <sub>10</sub> H <sub>14</sub> SO <sub>4</sub>	H	H	H
OH	CH <sub>2</sub> OH	C <sub>8</sub> H <sub>8</sub> O	C <sub>7</sub> H <sub>7</sub> O	C <sub>7</sub> H <sub>7</sub> O
OH	C <sub>3</sub> H <sub>7</sub>	C <sub>8</sub> H <sub>8</sub> OF	C <sub>7</sub> H <sub>6</sub> OF	C <sub>7</sub> H <sub>6</sub> OF
O	CH <sub>3</sub>	C <sub>8</sub> H <sub>9</sub> O	C <sub>7</sub> H <sub>7</sub> O	C <sub>7</sub> H <sub>7</sub> O
C <sub>6</sub> H <sub>15</sub> OSi	C <sub>5</sub> H <sub>7</sub> O <sub>2</sub>	H	H	H
OH	C <sub>2</sub> H <sub>4</sub> OH	C <sub>8</sub> H <sub>9</sub> O	C <sub>7</sub> H <sub>7</sub> O	C <sub>7</sub> H <sub>7</sub> O
O	C <sub>2</sub> H <sub>5</sub>	C <sub>8</sub> H <sub>9</sub> O	C <sub>7</sub> H <sub>7</sub> O	C <sub>7</sub> H <sub>7</sub> O

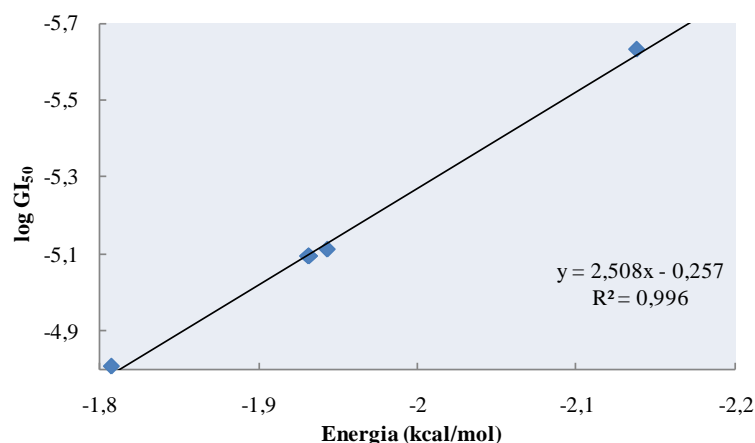
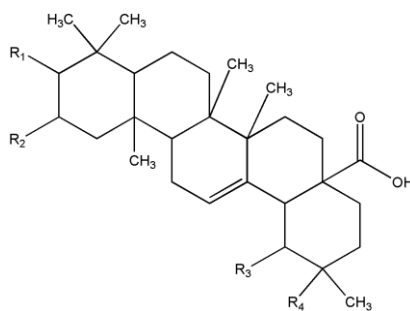


Figura 32: Correlação entre os dados computacionais e experimentais para a enzima 5-LOX para a linha celular SW1573.

Para a linha celular SW1573 foram estabelecidas correlações entre os dados computacionais e experimentais (IC<sub>50</sub>, GI<sub>50</sub>, TGI e LC<sub>50</sub>), tendo-se obtido coeficientes de correlação entre 0,612 e 0,996. Nas correlações obtidas para esta linha celular, encontram-se envolvidas as enzimas 5-LOX, Asp, PIK3CG, PAK1, p107, DD-Ligase, entre outras.

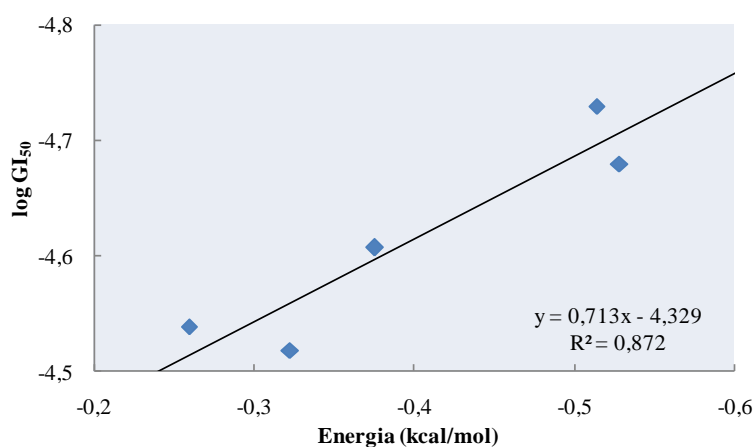
Para a referida linha celular foi identificada a estrutura base representada na figura seguinte.



**Figura 33:** Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular SW1573.

**Tabela 10:** Substituintes para a estrutura base representada na figura 33.

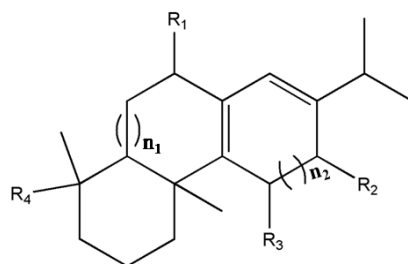
R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>
C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	H	H	CH <sub>3</sub>
OH	OH	H	CH <sub>3</sub>
OH	H	CH <sub>3</sub>	H
C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	H	CH <sub>3</sub>
OH	H	H	CH <sub>3</sub>



**Figura 34:** Correlação entre os dados computacionais e experimentais para a enzima NA para a linha celular T-47D.

Para a linha celular T-47D estabeleceu-se correlações entre os resultados dos testes celulares ( $IC_{50}$ ,  $GI_{50}$ , TGI e  $LC_{50}$ ) e os dados computacionais, tendo sido obtidas correlações para as enzimas CCNH, MAP2K1, CHK1, NA, DD-Ligase, AURKA. Para as correlações estabelecidas, os coeficientes encontram-se entre 0,609 e 0,872.

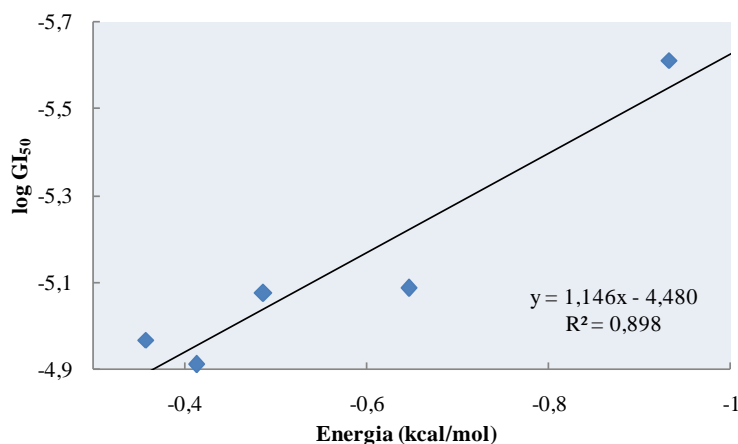
A estrutura base dos compostos que estão envolvidos nestas correlações está seguidamente representada.



**Figura 35:** Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular T-47D.

**Tabela 11:** Substituintes para a estrutura base representada na figura 35.

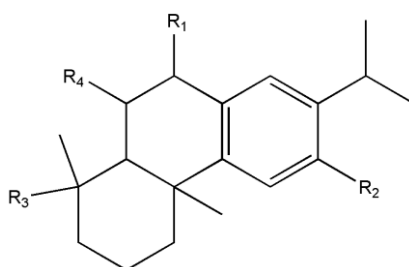
$R_1$	$R_2$	$R_3$	$R_4$	$n_1$	$n_2$
O	H	H	$C_2H_3O_2$	Simples	Dupla
$H_2$	$C_2H_3O_2$	H	$C_2H_3O_2$	Simples	Dupla
$H_2$	$OCH_3$	H	$CH_2OH$	Simples	Dupla
$H_2$	$OCH_3$	H	$CH_3$	Simples	Dupla
$H_2$	OH	H	$CH_3$	Simples	Dupla
O	O	O	$CH_3$	Simples	Simples
O	$OCH_3$	H	$CH_3$	Dupla	Dupla



**Figura 36:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular WiDr.

Em relação às correlações obtidas para a linha celular WiDr, verificou-se que a existência de um menor número de correlações entre os dados computacionais e os experimentais quando comparado com o número de correlações obtidas em relação às restantes linhas celulares em estudo. Os valores dos coeficientes de correlação encontram-se entre 0,607 e 0,898. Nestas correlações estão envolvidas várias enzimas, com por exemplo, DD-Ligase, CDC7, WEE1, ERK1, PIK3CG, p107, entre outras.

Na figura seguinte encontra-se representada a estrutura base dos compostos que participam nestas correlações.



**Figura 37:** Estrutura base dos compostos envolvidos nas melhores correlações para a linha celular WiDr.

Tabela 12: Substituintes para a estrutura base representada na figura 37.

R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>
O	C <sub>11</sub> H <sub>16</sub> NO <sub>3</sub>	CH <sub>3</sub>	H
H <sub>2</sub>	H	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	H
H <sub>2</sub>	OCH <sub>3</sub>	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	H
H <sub>2</sub>	OCH <sub>3</sub>	CH <sub>2</sub> OH	H
O	OCH <sub>3</sub>	CH <sub>3</sub>	OH
H <sub>2</sub>	OH	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	H
H <sub>2</sub>	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub>	H
O	C <sub>3</sub> H <sub>5</sub> O <sub>2</sub>	CH <sub>3</sub>	H

Ao serem estabelecidas as correlações entre os dados adquiridos através dos testes celulares e os dados obtidos computacionalmente, obteve-se algumas correlações negativas, estando seguidamente apresentados alguns exemplos.

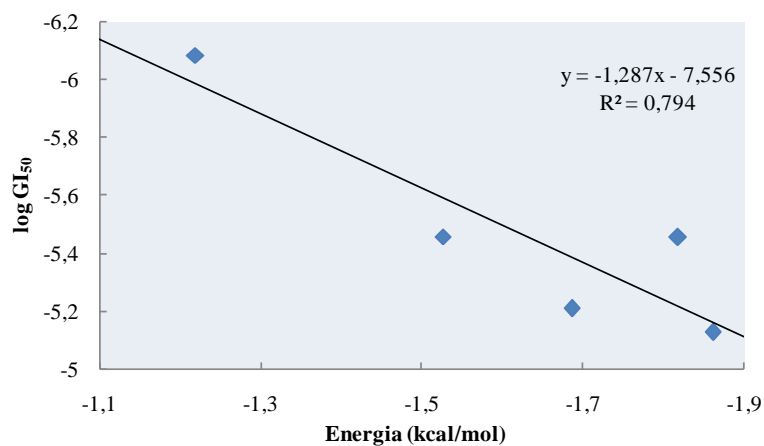
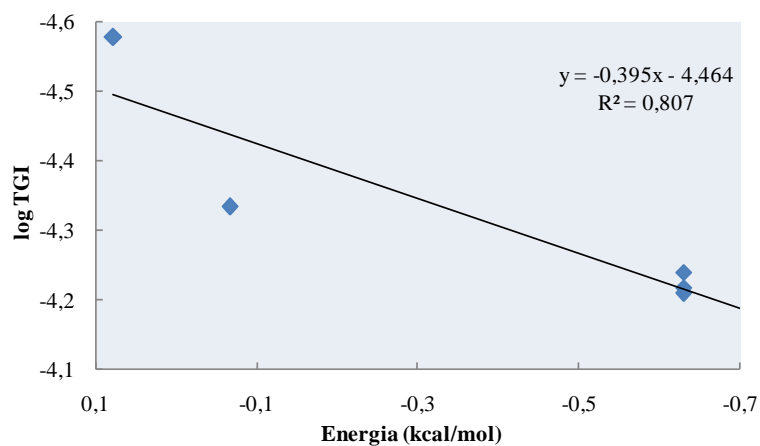
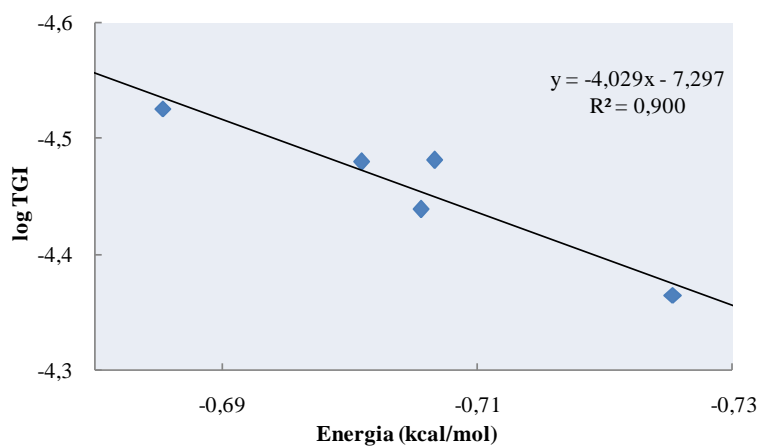


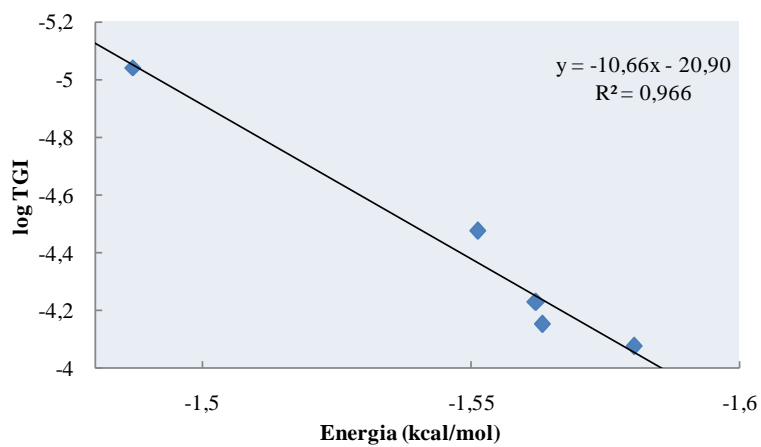
Figura 38: Correlação entre os dados computacionais e experimentais para a enzima HO-1 para a linha celular A2780.



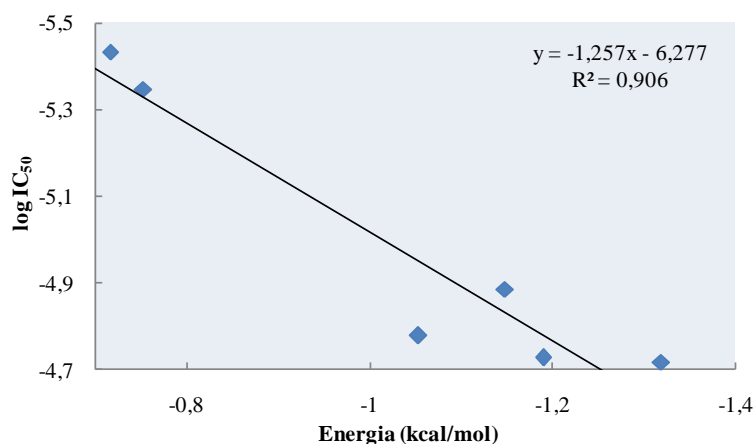
**Figura 39:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular HL60.



**Figura 40:** Correlação entre os dados computacionais e experimentais para a proteína CAP-G para a linha celular SW1573.



**Figura 41:** Correlação entre os dados computacionais e experimentais para a enzima NEK2 para a linha celular T-47D.



**Figura 42:** Correlação entre os dados computacionais e experimentais para a proteína MAX para a linha celular WiDr.

Estas correlações negativas poderão estar relacionadas com alguns casos em que duas enzimas se encontram ligadas e desta forma não têm qualquer efeito no crescimento das células cancerígenas, mas que ao estarem separadas uma irá provocar a proliferação das células e outra não. Deste modo, o que poderá ter acontecido é que computacionalmente terá sido inibida a enzima que não está relacionada com o crescimento das células cancerígenas e, experimentalmente, terá sido inibida a enzima associada ao crescimento das células.

## 5.2. Relação Quantitativa entre a Estrutura e a Actividade (QSAR)

### 5.2.1. Série 1

#### 5.2.1.1. Linha Celular do Cancro do Ovário (A2780)

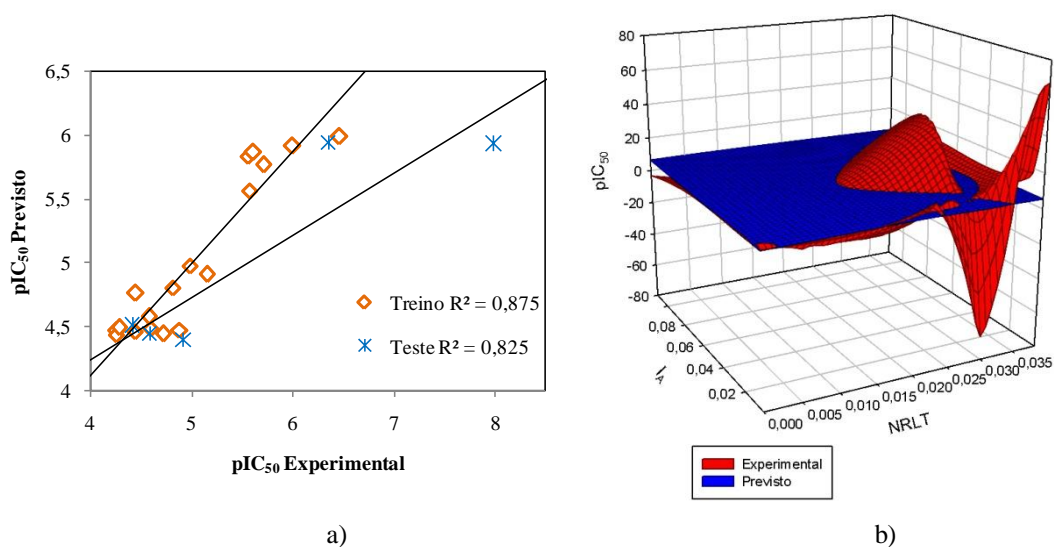
Para realizar os estudos de QSAR de compostos com actividade anti-tumoral na linha celular do cancro do ovário, foram utilizados os dados experimentais obtidos para os diferentes parâmetros de actividade estudadas ( $IC_{50}$ ,  $GI_{50}$ , TGI e  $LC_{50}$ ) e a média destes valores (Tabela 1F - Anexo F). Estes estudos foram realizados utilizando 23 compostos, 18 no conjunto de treino e 5 no conjunto de teste, à excepção do modelo obtido para o parâmetro  $LC_{50}$  que foi obtido utilizando 18 compostos no conjunto de treino e 4 no de teste, tendo sido eliminado um *outlier* utilizando os valores de *Z-score*.

Obtiveram-se modelos válidos para os parâmetros  $IC_{50}$ ,  $GI_{50}$ ,  $LC_{50}$  e para a média dos vários parâmetros de actividade.

A equação de QSAR que descreve a relação entre o  $IC_{50}$  e os descritores moleculares é:

$$pIC_{50} = 2,864 + 38,381(NRLT) - 8,480(I_A) + 7,720 \times 10^{-2}(DCPAS) \quad \text{Equação 9}$$

$$n_{\text{treino}} = 18; R^2_{\text{treino}} = 0,875; R^2_{cv} = 0,817; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,825$$



**Figura 43:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $IC_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

Analisando este modelo é possível verificar que o aumento da actividade biológica pode ser obtido aumentando o número relativo de ligações triplas e a diferença da carga parcial das áreas superficiais, e diminuindo o momento de inércia A.

Com base nos valores do teste t obtidos, verifica-se que o descritor que apresenta uma contribuição mais significativa é o número relativo de ligações triplas (NRLT). Este é um descritor constitucional que reflecte a composição molecular dos compostos.<sup>[63]</sup>

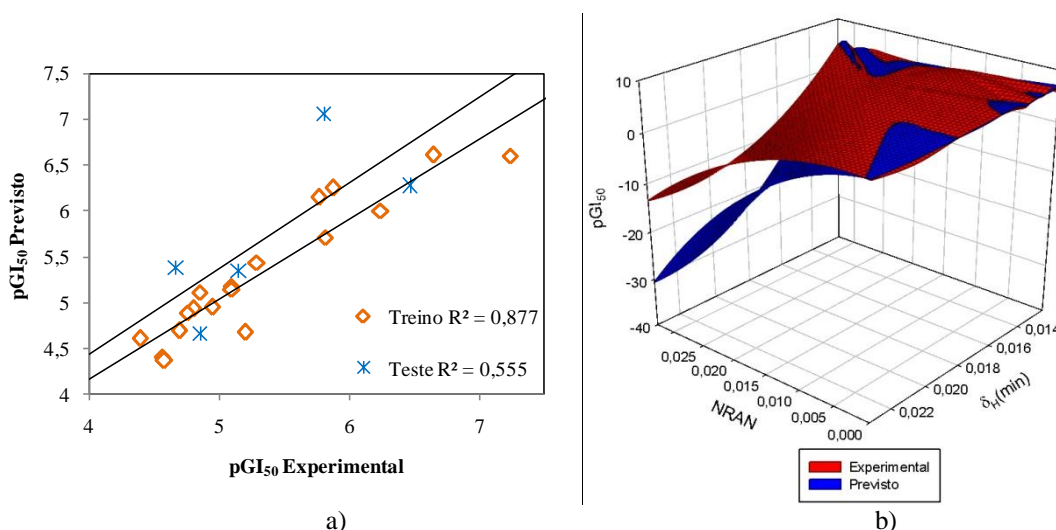
O descritor menos significativo para este modelo é o momento de inércia A ( $I_A$ ), este descritor geométrico contém informação acerca da distribuição de massa existente na molécula em torno do eixo x e está relacionado com a forma das moléculas.<sup>[63][70][71]</sup>

A diferença da carga parcial das áreas superficiais (DCPAS) baseia-se na diferença entre as áreas superficiais com carga positiva e as que possuem carga negativa, estando relacionado, deste modo, com a distribuição de carga negativa e positiva e a respectiva superfície. Este descritor possui ainda informação acerca das características polares dos compostos e das interacções entre as moléculas.<sup>[72][73]</sup>

A relação entre o parâmetro de actividade  $GI_{50}$  e os descritores moleculares é descrita pela equação que segue

$$pGI_{50} = 7,000 - 58,759(NRAN) - 0,189(CAPNS) - 1,545 \times 10^2 (\delta_H(\text{min})) \quad \text{Equação 10}$$

$$n_{\text{treino}} = 18; R^2_{\text{treino}} = 0,877; R^2_{\text{cv}} = 0,783; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,555$$



**Figura 44:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $GI_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $GI_{50}$ , em função dos descritores mais e menos significativos.

No modelo de QSAR obtido utilizando os valores de  $GI_{50}$  estão envolvidos três descritores moleculares, o número relativo de átomos de azoto, a contribuição da área superficial parcial negativa e a carga parcial mínima para o átomo de hidrogénio. Para aumentar a actividade biológica é necessário diminuir o valor destes descritores.

Neste modelo, o descritor que contribui mais significativamente é o número relativo de átomos de azoto (NRAN), sendo este um descritor constitucional que pode explicar a capacidade da molécula para formar ligações de hidrogénio. [74]

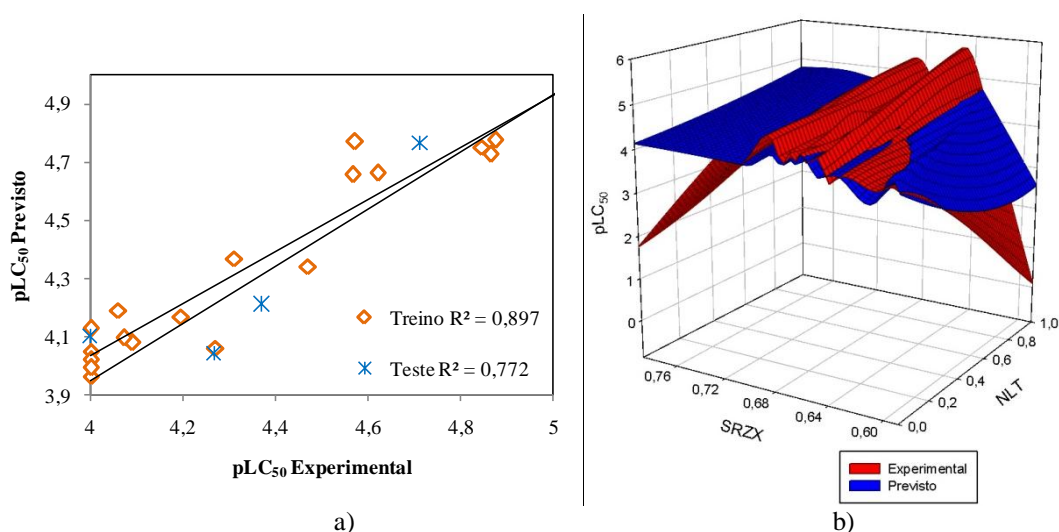
O segundo descritor mais significativo é a contribuição da área parcial negativa da superfície (CAPNS), este descritor pertence ao grupo de descritores relacionados com as cargas parciais das áreas superficiais e com a distribuição de carga que ocorre na molécula, estando também relacionado com a polaridade das áreas superficiais. [75][76]

A carga parcial mínima do átomo de hidrogénio ( $\delta_H(\min)$ ) reflecte a distribuição de carga neste átomo e está relacionada com as ligações de hidrogénio que se estabelecem e com as interações entre catiões e aniões. [77]

A equação de QSAR que descreve a relação entre  $LC_{50}$  e os descritores moleculares é a que está seguidamente representada

$$pLC_{50} = 4,590 + 0,466(NLT) - 0,426(MCI1) + 1,365(SRZX) \quad \text{Equação 11}$$

$$n_{\text{treino}} = 18; R^2_{\text{treino}} = 0,897; R^2_{\text{cv}} = 0,791; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,772$$



**Figura 45:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $LC_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $LC_{50}$ , em função dos descritores mais e menos significativos.

Com base neste modelo, verifica-se que para aumentar a actividade biológica dos compostos é necessário aumentar o número de ligações triplas e a razão entre a sombra projectada no plano ZX e o rectângulo no plano ZX, e diminuir a média do conteúdo de informação (ordem 1).

Para o modelo acima representado, o descritor mais significativo é o número de ligações triplas (NLT), sendo este um descritor constitucional que está associado à reactividade dos compostos. [78]

O descritor menos significativo é a razão entre a sombra projectada no plano ZX e o rectângulo no plano ZX (SRZX), sendo este é um descritor geométrico que descreve a forma e a amplitude da molécula em termos de coordenadas tridimensionais. [79]

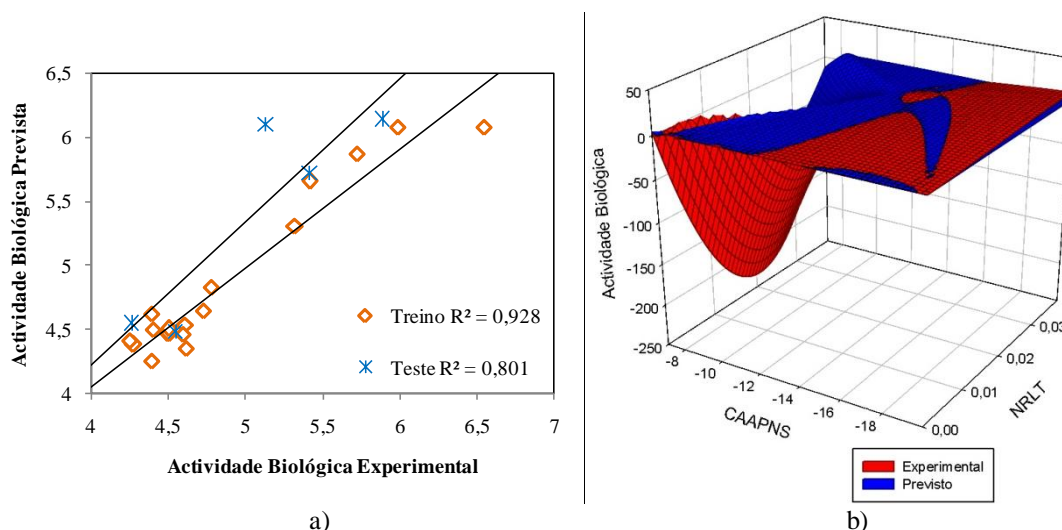
A média do conteúdo de informação (ordem 1) (MCI1) representa o facto da esfera de coordenação cobrir apenas o primeiro nível de valência, assumindo os átomos directamente ligados ao átomo considerando como sendo de primeira ordem. Este descritor fornece informação acerca de quantos tipos de átomos diferentes existem nas moléculas e do nível de diversidade da ramificação presente no primeiro nível de valência. [80]

A relação entre a média das actividades biológicas e os descritores moleculares é descrita pela seguinte equação

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 3,899 + 43,138(\text{NRLT}) - 14,361(\text{I}_A) + 0,777(\text{ASPR}) - 3,775 \times 10^{-2}(\text{CAAPNS})$$

**Equação 12**

$$n_{\text{treino}} = 18; R^2_{\text{treino}} = 0,928; R^2_{\text{cv}} = 0,848; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,801$$



**Figura 46:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

A equação anteriormente representada indica que o aumento do número relativo de ligações triplas e área superficial positiva relativa, e a diminuição do momento de inércia  $A$  e da carga atómica da área parcial negativa da superfície causam um aumento da actividade biológica.

Os valores obtidos referentes ao teste  $t$  indicam que o descritor molecular que apresenta uma maior contribuição para este modelo é o número relativo de ligações triplas (NRLT), enquanto que o descritor que menos contribui para este modelo é a carga atómica da área parcial negativa da superfície (CAAPNS). Este último descritor pertence ao grupo dos descritores electrostáticos, estando relacionado com a assimetria de carga induzida na molécula e com a área total da superfície molecular. <sup>[79]</sup>

A área superficial positiva relativa (ASPR) é um descritor molecular relacionado com a carga parcial total da área superficial, baseia-se na área da superfície de toda a molécula e na distribuição da carga existente na mesma. Combina, deste modo, a informação electrónica e a referente à forma que caracteriza a molécula e as características responsáveis pelas interacções polares entre moléculas. <sup>[81]</sup>

Os modelos de QSAR obtidos utilizando os dados experimentais para os diferentes parâmetros de actividade e para a média dos mesmos apresentam coeficientes de correlação acima de 0,8 para os conjuntos de treino, apresentando um valor máximo

de 0,928. Para os conjuntos de teste, foram obtidos coeficientes de correlação entre 0,555 e 0,825.

Os valores calculados de RMSE,  $s^2$  e SEE encontram-se muito próximos de zero. Os valores do factor de inflação da variância (VIF) dos descritores são inferiores a 4, à excepção do descritor molecular do número de ligações triplas envolvido no modelo obtido para o parâmetro  $LC_{50}$  e para a média das actividades biológicas. Valores do VIF superiores a 4 indicam que existe colinearidade entre os descritores que participam nos modelos.

Os parâmetros de validação cruzada obtidos para os modelos encontram-se entre 0,783 e 0,848, indicando que os modelos obtidos são estatisticamente válidos. Os valores obtidos para o teste F e t são superiores aos valores de referência, o que significa que os modelos e os descritores neles envolvidos são estatisticamente válidos.

Os dados estatísticos e das equações obtidas para os modelos referidos anteriormente são apresentados no Anexo F (F.5. e F.6.).

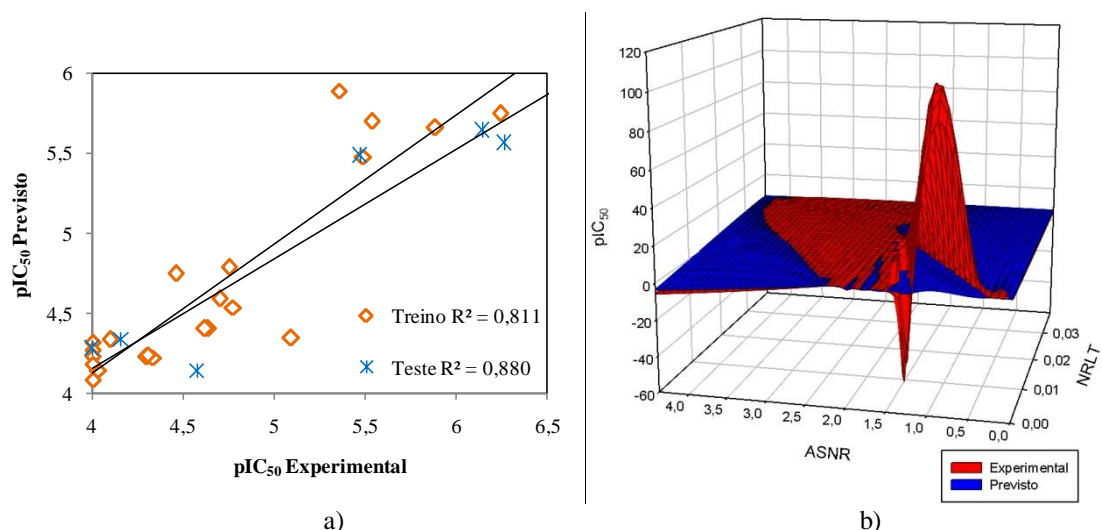
### 5.2.1.2. Linha Celular do Cancro do Pulmão (SW1573)

Para obter modelos de QSAR de compostos com actividade anti-tumoral na linha celular do cancro do pulmão, utilizaram-se os valores experimentais para os parâmetros  $IC_{50}$ ,  $GI_{50}$ , TGI,  $LC_{50}$  e a média destes (Tabela 2F - Anexo F). Nestes estudos foram utilizados 28 compostos, 22 para o conjunto de treino e 6 para o conjunto de teste. Foram obtidos modelos de QSAR válidos para os parâmetros  $IC_{50}$ ,  $GI_{50}$  e para a média de todos os parâmetros.

A equação de QSAR que descreve a relação entre o parâmetro de actividade  $IC_{50}$  e os descritores moleculares é:

$$pIC_{50} = 3,989 + 40,709(NRLT) - 0,187(ASNR) - 6,143 \times 10^{-2}(CAAPNS) \quad \text{Equação 13}$$

$n_{\text{treino}} = 22; R^2_{\text{treino}} = 0,811; R^2_{\text{cv}} = 0,712; n_{\text{teste}} = 6; R^2_{\text{teste}} = 0,880$



**Figura 47:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $IC_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

Com base no modelo anterior, verifica-se que para aumentar a actividade biológica é necessário aumentar o número relativo de ligações triplas, e diminuir a área superficial negativa relativa e a carga atómica da área parcial negativa da superfície.

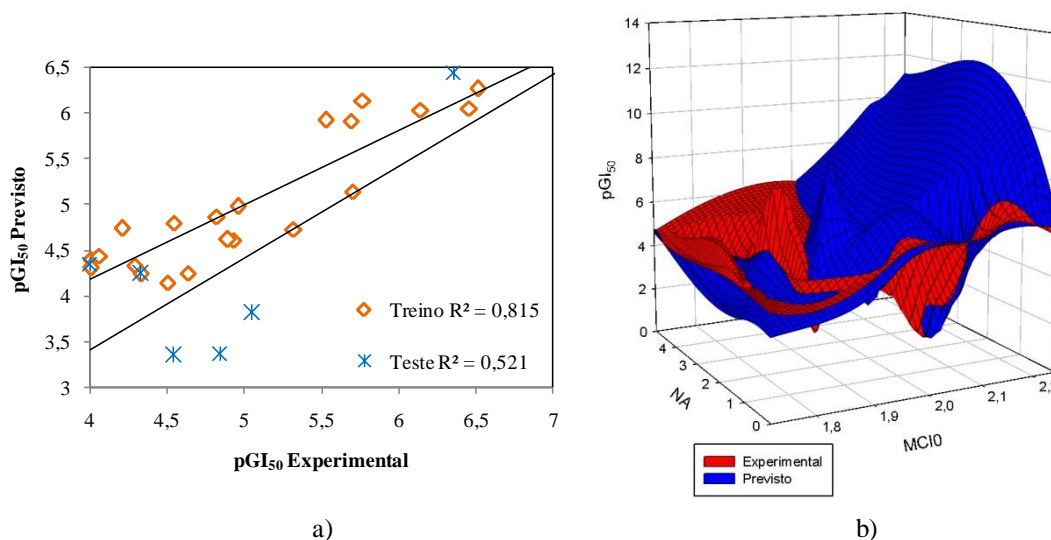
Os valores do teste t mostram que o descritor que apresenta uma contribuição mais significativa é o número relativo de ligações triplas. Por outro lado, o descritor molecular que contribui menos significativamente para este modelo é a área superficial negativa relativa (ASNR). Este descritor baseia-se quer na forma, quer na informação electrónica que caracteriza as moléculas, estando também relacionado com as características responsáveis pelas interacções polares existentes entre as moléculas. [77][82]

A equação seguinte descreve a relação entre os descritores moleculares e o parâmetro  $GI_{50}$

$$pGI_{50} = -15,790 + 8,864(MCIO) - 6,396(SRXY) + 8,288(ASPP/ATSM) + 0,278(NA)$$

**Equação 14**

$$n_{\text{treino}} = 22; R^2_{\text{treino}} = 0,815; R^2_{\text{cv}} = 0,674; n_{\text{teste}} = 6; R^2_{\text{teste}} = 0,521$$



**Figura 48:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $GI_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $GI_{50}$ , em função dos descritores mais e menos significativos.

Através da equação anterior é possível observar que o aumento da actividade biológica pode ser alcançado efectuando um aumento da média do conteúdo de informação (ordem 0), da razão entre a área superficial parcial positiva e a área total da superfície molecular e do número de anéis, e uma diminuição da razão entre a sombra projectada no plano XY e o rectângulo no plano XY.

Pela análise dos valores obtidos para o teste t, constata-se que o descritor molecular mais significativo é a média do conteúdo de informação (ordem 0) (MCI0). Este é um descritor topológico que descreve o tamanho, as ramificações e a composição da molécula e está relacionado com as interacções de dispersão que ocorrem entre as moléculas. [83]

O descritor molecular que apresenta uma menor contribuição é o número de anéis (NA), este pertence ao grupo de descritores constitucionais e diferencia os compostos de cadeia dos compostos que possuem anéis, estando também relacionado com a forma molecular. [84][85]

A razão entre a área superficial parcial positiva e a área total da superfície molecular (ASPP/ATSM) é um descritor relacionado com a distribuição de carga, fornecendo uma medida da polaridade do composto. [86]

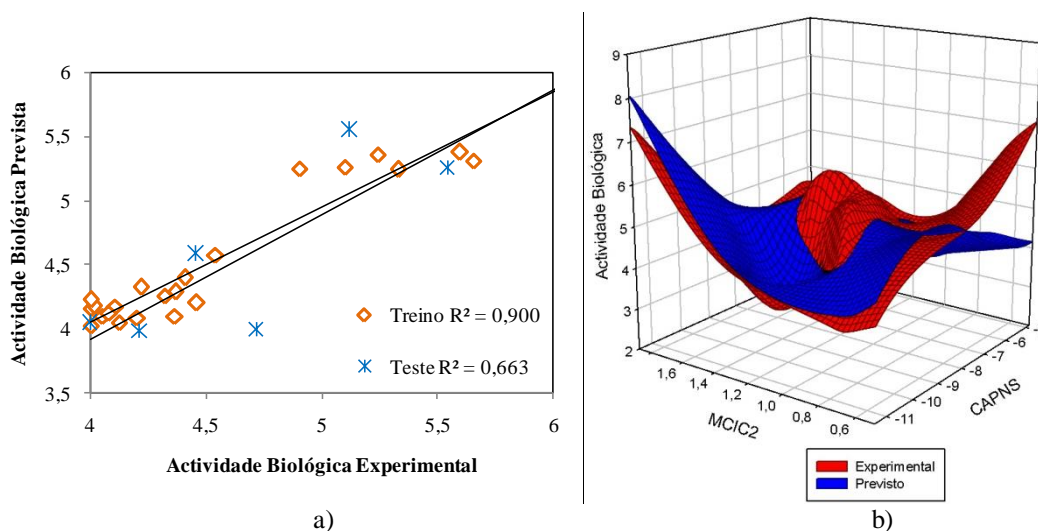
A razão entre a sombra projectada no plano XY e o rectângulo no plano XY (SRXY) pertence ao grupo de descritores geométricos, descrevendo a forma e a extensão da molécula em termos das coordenadas tridimensionais. [79]

A relação entre a média das atividades experimentais e os descritores moleculares é dada através da seguinte equação de QSAR

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 3,225 - 0,170(\text{CAPNS}) + 4,058 \times 10^{-3}(\text{CTAPNS}) + 25,163(\text{NRLT}) + 0,297(\text{MCIC2})$$

Equação 15

$$n_{\text{treino}} = 22; R^2_{\text{treino}} = 0,900; R^2_{\text{cv}} = 0,846; n_{\text{teste}} = 6; R^2_{\text{teste}} = 0,663$$



**Figura 49:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros da atividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de atividade, em função dos descritores mais e menos significativos.

Com base no modelo anterior verifica-se que o aumento da carga total da área parcial negativa da superfície, do número relativo de ligações triplas e da média do conteúdo de informação complementar (ordem 2), e a diminuição da contribuição da área parcial negativa da superfície, causa um aumento da atividade biológica.

Através do teste t é possível constatar que o descritor molecular mais significativo é a contribuição da área parcial negativa da superfície (CAPNS), enquanto que o menos significativo é a média do conteúdo de informação complementar (ordem 2) (MCIC2). Este último descritor pertence ao grupo dos topológicos, sendo definido com base na teoria de informação de Shannon. Este pode ser calculado para diferentes ordens de vizinhanças, no segundo nível, o conjunto de átomos é decomposto em classes equivalentes com base na sua natureza química e no padrão de ligação acima da segunda ordem de ligações das vizinhanças. [87]

A carga total da área parcial negativa da superfície (CTAPNS) é um descritor que fornece informação acerca de diversos aspectos da estrutura molecular e é calculado

através da multiplicação da área superficial parcial negativa acessível ao solvente pela carga negativa total. <sup>[88]</sup>

Os modelos obtidos para os parâmetros individuais de actividade apresentaram coeficientes de correlação para o conjunto de treino entre 0,811 e 0,900. Para o conjunto de teste, foram obtidos coeficientes de correlação acima de 0,5, apresentando um valor máximo de 0,880.

Em termos dos parâmetros de validação cruzada obtidos, estes encontram-se entre 0,674 e 0,846. O valor mais baixo foi obtido para o modelo relativo ao GI<sub>50</sub> e o valor mais elevado foi obtido para o modelo obtido usando a média dos vários parâmetros.

Os valores do factor de inflação da variância foram inferiores a 4, apresentando um valor máximo de 3,569, o que indica que existe uma baixa colinearidade entre os descritores envolvidos nos modelos. Os valores de RMSE,  $s^2$  e SEE obtidos encontram-se muito próximo de zero.

Relativamente aos valores obtidos para o teste t e F, verificou-se que estes eram superiores aos valores de referência, indicando que os modelos obtidos e os descritores moleculares neles envolvidos são estatisticamente válidos.

Os dados estatísticos e das equações obtidas para os modelos referidos anteriormente são apresentados no Anexo F (F.5. e F.6.).

### **5.2.1.3. Linha Celular do Carcinoma Mamário (T-47D)**

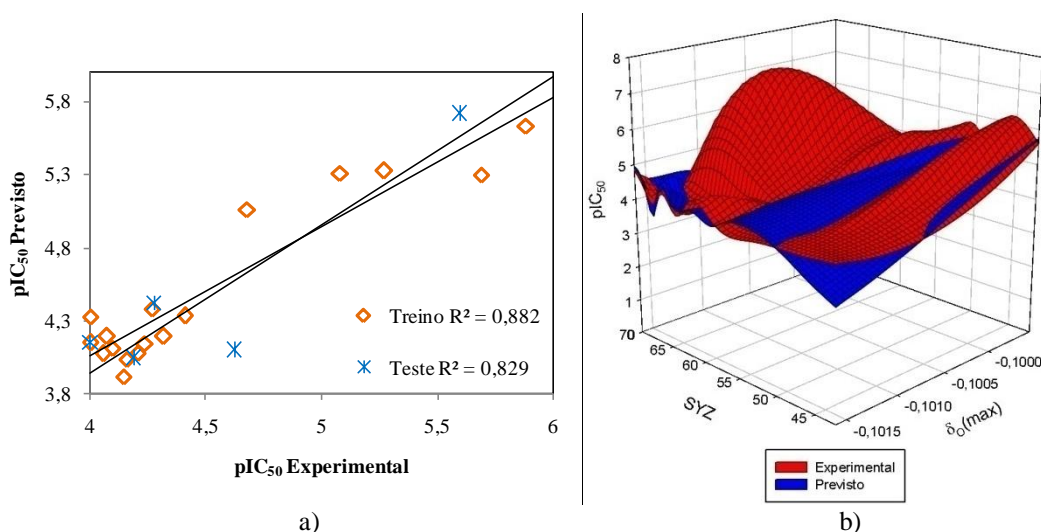
Para estabelecer relações quantitativas entre a estrutura e a actividade anti-tumoral dos compostos para a linha celular do carcinoma mamário, foram utilizados valores de actividade para diversos parâmetros (IC<sub>50</sub>, GI<sub>50</sub>, TGI, LC<sub>50</sub>) e a média destes valores (Tabela 3F - Anexo F).

Para a obtenção dos modelos foram utilizados 22 compostos, sendo o conjunto de treino constituído por 17 compostos e o conjunto de teste por 5 compostos. Não foi possível obter modelos válidos para os parâmetros GI<sub>50</sub> e LC<sub>50</sub>.

A equação de QSAR abaixo representada descreve a relação entre o parâmetro IC<sub>50</sub> e os descritores moleculares

$$pIC_{50} = 64,701 + 5,716 \times 10^2 (\delta_O(\max)) - 1,607(\text{ASPR}) - 2,645 \times 10^{-2}(\text{SYZ}) \quad \text{Equação 16}$$

$$n_{\text{treino}} = 17; R^2_{\text{treino}} = 0,882; R^2_{\text{cv}} = 0,726; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,829$$



**Figura 50:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro IC<sub>50</sub> para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro IC<sub>50</sub>, em função dos descritores mais e menos significativos.

O modelo acima apresentado indica que o aumento da actividade biológica pode ser alcançado aumentando a carga parcial máxima para o átomo de oxigénio e diminuindo a área superficial positiva relativa e a sombra projectada ao longo do plano YZ.

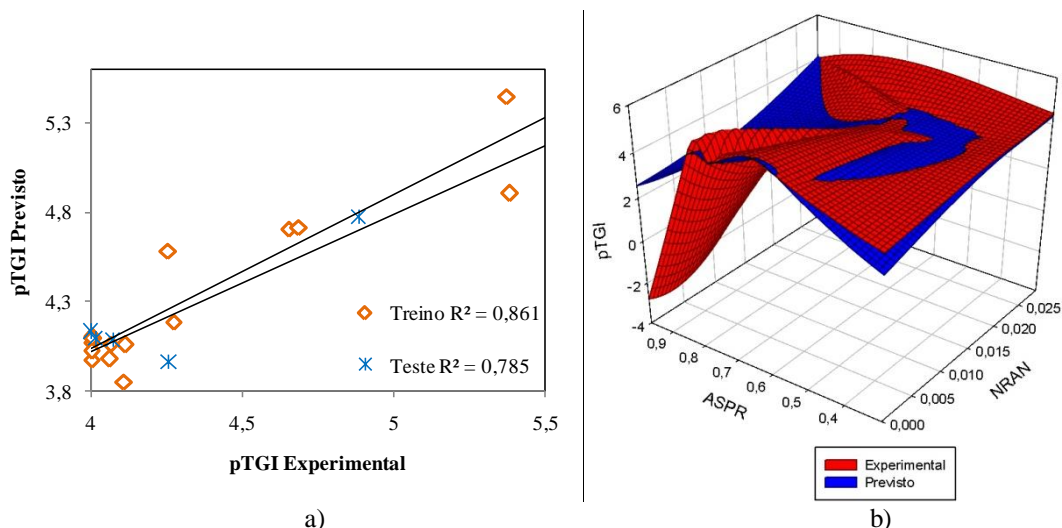
O descritor molecular que apresenta uma contribuição mais significativa é a carga parcial máxima para o átomo de oxigénio ( $\delta_O(\max)$ ), este é um descritor electrostático que descreve as interacções intermoleculares e a polaridade das moléculas. [58]

O descritor que contribui menos significativamente é a sombra projectada ao longo do plano YZ (SYZ), este descritor pertence ao grupo dos descritores geométricos que descrevem o tamanho das moléculas. [76]

A equação abaixo representada descreve a relação entre o TGI e os descritores moleculares

$$pTGI = 4,789 - 34,670(\text{NRAN}) + 52,519(I_A) - 1,002(\text{ASPR}) \quad \text{Equação 17}$$

$$n_{\text{treino}} = 17; R^2_{\text{treino}} = 0,861; R^2_{\text{cv}} = 0,732; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,785$$



**Figura 51:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro TGI para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro TGI, em função dos descritores mais e menos significativos.

O modelo de QSAR utilizando os valores dos parâmetros de TGI, demonstra que o aumento da actividade biológica dos compostos em estudo pode ser efectuado através da diminuição do número relativo de átomos de azoto e da área superficial positiva relativa, e do aumento do momento de inércia A.

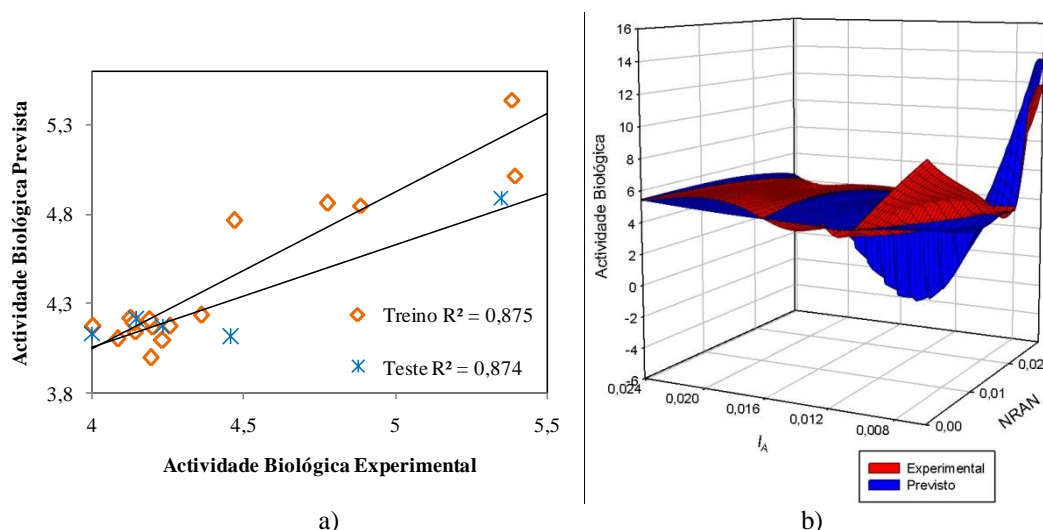
Com base nos valores obtidos através do teste t, é possível verificar que o descritor molecular que contribui mais significativamente é o número relativo de átomos de azoto (NRAN), enquanto que o descritor que apresenta uma menor contribuição é a área superficial positiva relativa (ASPR).

A relação entre a média dos parâmetros de actividade e os descritores moleculares é descrita através da equação seguinte

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 4,857 - 35,505(\text{NRAN}) + 42,164(I_A) - 0,719(\text{ASPR})$$

**Equação 18**

$$n_{\text{treino}} = 17; R^2_{\text{treino}} = 0,875; R^2_{\text{cv}} = 0,793; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,874$$



**Figura 52:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

Com base neste modelo, é possível verificar que o aumento da actividade biológica pode ser obtido aumentando o momento de inércia A, e diminuindo o número relativo de átomos de azoto e a área superficial positiva relativa.

Os valores referentes ao teste t permitem-nos afirmar que os descritores que apresentam maior e menor contribuição para este modelo são o número relativo de átomos de azoto e o momento de inércia A, respectivamente.

Analisando os modelos obtidos, observa-se que os coeficientes de correlação para o conjunto de treino encontram-se entre 0,861 e 0,882, enquanto que para o conjunto de teste os coeficientes de correlação encontram-se entre 0,785 e 0,874. Os parâmetros da validação cruzada encontram-se num intervalo entre 0,726 e 0,793.

Os valores calculados do factor de inflação da variância são maioritariamente inferiores a 4, à excepção da carga parcial máxima para o átomo de oxigénio (modelo para o  $IC_{50}$ ). Este resultado indica que existe uma certa colinearidade entre este e os restantes descritores que participam neste modelo. Os valores para RMSE,  $s^2$  e SEE encontram-se muito próximos de zero.

Quanto aos testes t e F, os valores obtidos são superiores aos valores de referência, indicando que quer os modelos obtidos, quer os descritores que neles participam são estatisticamente válidos.

Os dados estatísticos e das equações obtidas para os modelos referidos anteriormente são apresentados no Anexo F (F.5. e F.6.).

## 5.2.2. Série 7

### 5.2.2.1. Linha Celular da Leucemia (HL60)

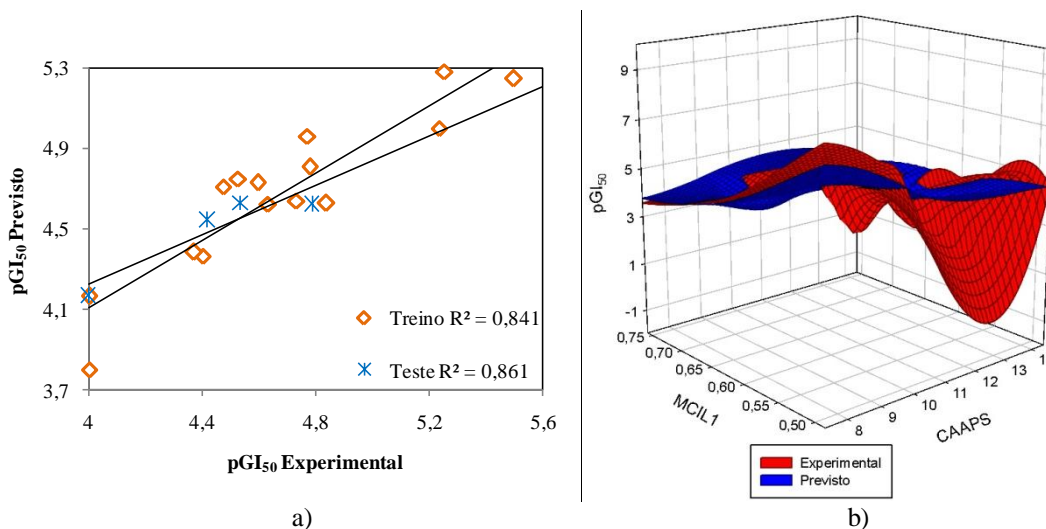
Para estudar a relação entre a estrutura e a actividade dos compostos que constituem esta série para a linha celular da leucemia, foram utilizados os dados experimentais para os parâmetros IC<sub>50</sub>, GI<sub>50</sub>, TGI, LC<sub>50</sub> e a média destes (Tabela 7F - Anexo F). Tendo sido utilizados 19 compostos, 15 compostos que constituem o conjunto de treino e 4 compostos no conjunto de teste.

Como resultado destes estudos, foram obtidos modelos válidos para o GI<sub>50</sub> e para a média das actividades.

A equação de QSAR que descreve a relação entre o parâmetro de actividade GI<sub>50</sub> e os descritores moleculares é a seguinte

$$pGI_{50} = 7,607 + 5,779 \times 10^{-4}(W) - 0,203(CAAPS) - 2,461(MCIL1) \quad \text{Equação 19}$$

$$n_{\text{treino}} = 15; R^2_{\text{treino}} = 0,841; R^2_{\text{cv}} = 0,600; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,861$$



**Figura 53:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro GI<sub>50</sub> para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro GI<sub>50</sub>, em função dos descritores mais e menos significativos.

Com base no modelo referente ao parâmetro de actividade  $GI_{50}$ , verifica-se que o aumento do índice de Wiener e a diminuição da carga atómica da área positiva da superfície e da média do conteúdo de informação sobre as ligações (ordem 1) provoca um aumento da actividade biológica.

O descritor molecular que apresenta a maior contribuição para este modelo é a carga atómica da área positiva da superfície (CAAPS). Este é um descritor electrostático que está relacionado com a área superficial acessível ao solvente dos átomos com carga positiva, podendo ser indicativo da interacção soluto-solvente que surge através da presença de átomos polarizáveis no soluto. <sup>[89][90]</sup>

O descritor que contribui menos significativamente é a média do conteúdo de informação sobre as ligações (ordem 1) (MCIL1). Este descritor pertence ao grupo dos topológicos, sendo definido com base na teoria de informação de Shannon e reflecte a conectividade entre átomos na molécula na primeira esfera de coordenação. <sup>[91]</sup>

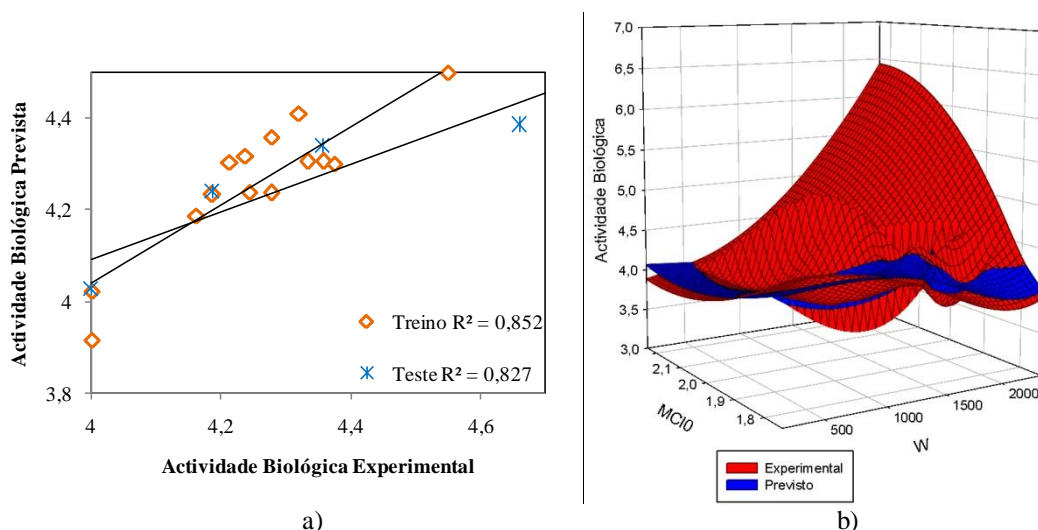
O índice de Wiener (W) é um descritor topológico que descreve a compactação da molécula, apresentando valores elevados para compostos de cadeia linear e valores baixos para compostos ramificados. É também utilizado para medir outras propriedades topológicas como a ciclicidade e a centralidade. <sup>[92][93]</sup>

A relação entre a média dos parâmetros de actividade e os descritores moleculares é descrita pela seguinte equação

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 5,718 + 2,449 \times 10^{-4}(W) - 6,287 \times 10^{-2}(\text{CAAPS}) - 0,555(\text{MCI0})$$

**Equação 20**

$$n_{\text{treino}} = 15; R^2_{\text{treino}} = 0,852; R^2_{\text{cv}} = 0,651; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,827$$



**Figura 54:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $GI_{50}$ , em função dos descritores mais e menos significativos.

Analisando o modelo obtido utilizando a média dos parâmetros em estudo, é possível constatar que o aumento da actividade biológica pode ser obtido aumentando o valor do índice de Wiener e diminuindo a carga atômica da área positiva da superfície e a média do conteúdo de informação (ordem 0).

Os valores obtidos para o teste t indicam que o descritor que apresenta uma maior contribuição é o índice de Wiener (W) e o menos significativo é a média do conteúdo de informação (ordem 0) (MCI0).

Analisando os vários modelos obtidos, verifica-se que os coeficientes de correlação obtidos para o conjunto de treino encontram-se entre 0,841 e 0,852 e para o conjunto de teste entre 0,827 e 0,861. Os valores de validação cruzada apresentam um máximo de 0,651.

Os valores calculados para o VIF são inferiores a 4, apresentando um valor máximo de 1,766. Este resultado indica que não existe colinearidade entre os descritores. Para os parâmetros RMSE,  $s^2$  e SEE, foram obtidos valores muito próximos de zero.

Os valores obtidos para o teste t e para o teste F são superiores aos valores de referência, o que significa que os valores obtidos e os descritores neles envolvidos são estatisticamente válidos.

Os dados estatísticos e das equações obtidas para os modelos referidos anteriormente são apresentados no Anexo F (F.5. e F.6.).

### 5.2.2.2. Linha Celular do Cancro do Ovário (A2780)

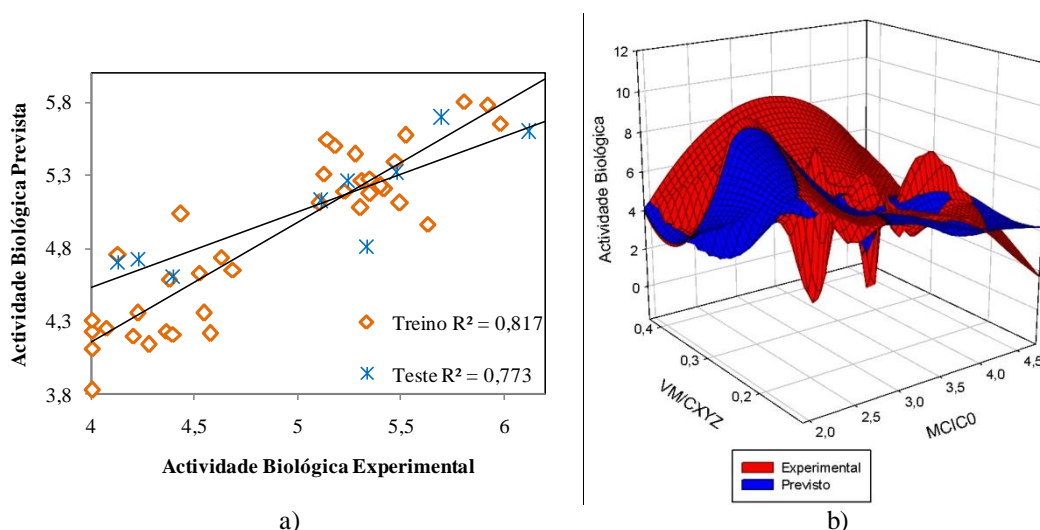
Tendo como objectivo estudar as possíveis relações entre a estrutura e a actividade dos compostos em estudo, foram utilizados os dados experimentais para os diversos parâmetros de actividade ( $IC_{50}$ ,  $GI_{50}$ , TGI e  $LC_{50}$ ) e a média destes, em relação à linha celular do cancro do ovário (Tabela 8F - Anexo F). Foi utilizado um conjunto de 46 compostos, 37 pertencentes ao conjunto de treino e 9 ao conjunto de teste.

Não foi possível encontrar modelos válidos para os parâmetros individualmente, tendo sido apenas encontrado um modelo válido utilizado a média dos valores dos parâmetros.

A relação existente entre a média dos parâmetros e os descritores moleculares é descrita através da equação de QSAR representada seguidamente

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 21,093 - 1,942(\text{MCIC0}) - 42,759(\text{CPR}) - 0,355(\text{NAO}) - 0,699(\text{MCS}) - 0,271(\text{MMR}) - 4,688(\text{VM/CXYZ}) \quad \text{Equação 21}$$

$$n_{\text{treino}} = 37; R^2_{\text{treino}} = 0,817; R^2_{\text{cv}} = 0,752; n_{\text{teste}} = 9; R^2_{\text{teste}} = 0,773$$



**Figura 55:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

Este modelo indica que para se obter um aumento da actividade biológica é necessário diminuir a média do conteúdo de informação complementar (ordem 0), a carga positiva relativa, o número de átomos de oxigénio, a média das cargas da superfície, a massa molecular relativa e o valor da razão entre o volume molecular e a caixa de coordenadas XYZ.

O descritor mais significativo é a média do conteúdo de informação complementar (ordem 0) (MCIC0) que descreve a conectividade e ramificações presentes na molécula e pode ser relacionado com a forma molecular e com a simetria.<sup>[85]</sup>

O segundo descritor mais significativo é a carga positiva relativa (CPR), este contém informação para descrever a molécula e com características responsáveis pelas interacções entre as moléculas. Representa também o efeito das interacções intermoleculares polares.<sup>[92][94]</sup>

A massa molecular relativa (MMR) é um descritor constitucional que contribui para as massas atómicas (volumes) e para a sua distribuição dentro do espaço molecular e quantifica de forma eficaz a coesão dos compostos desde as interacções dispersivas às hidrófobas.<sup>[91]</sup>

A média das cargas da superfície (MCS) é um descritor electrostático que descreve a capacidade para um composto actuar como aceitador de ligações de hidrogénio.<sup>[90]</sup>

O descritor que apresenta a menor contribuição para este modelo é a razão entre o volume molecular e a caixa de coordenadas XYZ (VM/CXYZ), este é calculado como sendo a razão entre o volume molecular e o volume da caixa com as dimensões Xmax, Ymax e Zmax que contém a molécula. Este descritor está relacionado com a linearidade da molécula e fornece informação acerca da compactação da molécula.<sup>[95][96]</sup>

O número de átomos de oxigénio (NAO) é um descritor constitucional que afecta a densidade da nuvem electrónica da molécula e reflecte a capacidade de uma molécula para aceitar ligações de hidrogénio.<sup>[97]</sup>

Os coeficientes de correlação obtidos para o conjunto de treino e de teste foram 0,817 e 0,773, respectivamente. Efectuando o teste da validação cruzada, foi obtido um valor de 0,752.

Os valores para o VIF são inferiores a 4, apresentando um valor máximo de 1,212, indicando que não existe colinearidade entre os descritores. Os valores de RMSE,  $s^2$  e SEE encontram-se muito próximos de zero.

Relativamente aos testes t e F, foram obtidos valores superiores aos de referência, indicando que tanto o modelo obtido, quer os descritores envolvidos são estatisticamente válidos.

Os dados estatísticos e da equação obtida para o modelo referido anteriormente são apresentados no Anexo F (F.5. e F.6.).

### 5.2.3. Série 9

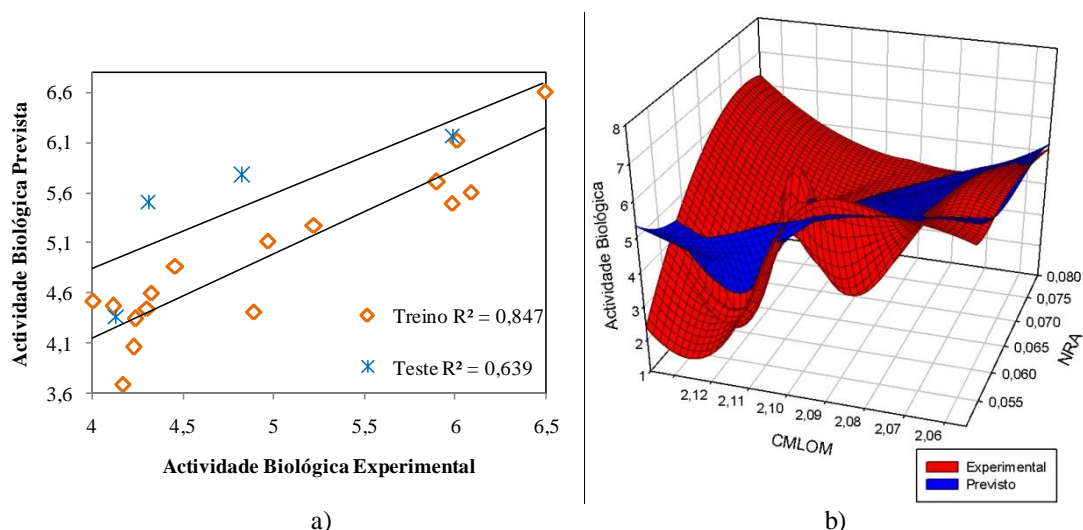
#### 5.2.3.1. *Linha Celular do Cancro do Ovário (A2780)*

Com o objectivo de estabelecer relações quantitativas entre a estrutura e a actividade dos compostos, foram utilizados valores dos parâmetros de actividade IC<sub>50</sub>, GI<sub>50</sub>, TGI, LC<sub>50</sub> e a média destes parâmetros (Tabela 10F - Anexo F). No entanto, apenas foi possível encontrar um modelo válido para a média das actividades biológicas. Neste modelo foram utilizados 20 compostos, 16 que constituem o conjunto de treino e 4 o conjunto de teste.

A equação de QSAR que descreve a relação entre a média das actividades biológicas e os descritores moleculares está seguidamente representada

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 47,062 - 76,759(\text{NRA}) + 0,405(\text{NAO}) - 18,438(\text{CMLOM}) \quad \text{Equação 22}$$

$$n_{\text{treino}} = 16; R^2_{\text{treino}} = 0,847; R^2_{\text{cv}} = 0,707; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,639$$



**Figura 56:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

Este modelo indica que o aumento do número de átomos de oxigénio, e a diminuição do número relativo de anéis e do valor da contribuição máxima para a ligação de uma orbital molecular poderia aumentar a actividade biológica dos compostos em estudo.

Os valores para o teste t indicam que o descritor molecular que apresenta uma maior contribuição é o número relativo de anéis (NRA), sendo este um descritor constitucional estando relacionado com a forma e o tamanho da molécula, contribuindo para a rigidez da cadeia e para o efeito de impedimento estéreo dos compostos. <sup>[85][96]</sup>

O descritor molecular menos significativo é a contribuição máxima para a ligação de uma orbital molecular (CMLOM). Este é um descritor químico-quântico que está relacionado com a força das interações das ligações intermoleculares e caracteriza a estabilidade das moléculas, a sua flexibilidade e outras propriedades relacionadas com a valência. <sup>[63]</sup>

Em termos estatísticos, os coeficientes de correlação obtidos para os conjuntos de treino e de teste foram 0,847 e 0,639, respectivamente. Quanto ao valor do parâmetro da validação cruzada, este foi de 0,707.

Os valores de RMSE,  $s^2$  e SEE calculados são muito próximos de zero. Enquanto que os valores para o VIF são inferiores a 4, apresentando um valor máximo

de 1,298, indicando a ausência de colinearidade entre os descritores envolvidos no modelo.

Relativamente aos valores dos testes t e F, estes são superiores aos valores de referência, o que significa que tanto o modelo como os descritores envolvidos no mesmo são estatisticamente válidos.

Os dados estatísticos e da equação obtida para o modelos referido anteriormente são apresentados no Anexo F (F.5. e F.6.).

### 5.2.3.2. Linha Celular do Cancro do Pulmão (SW1573)

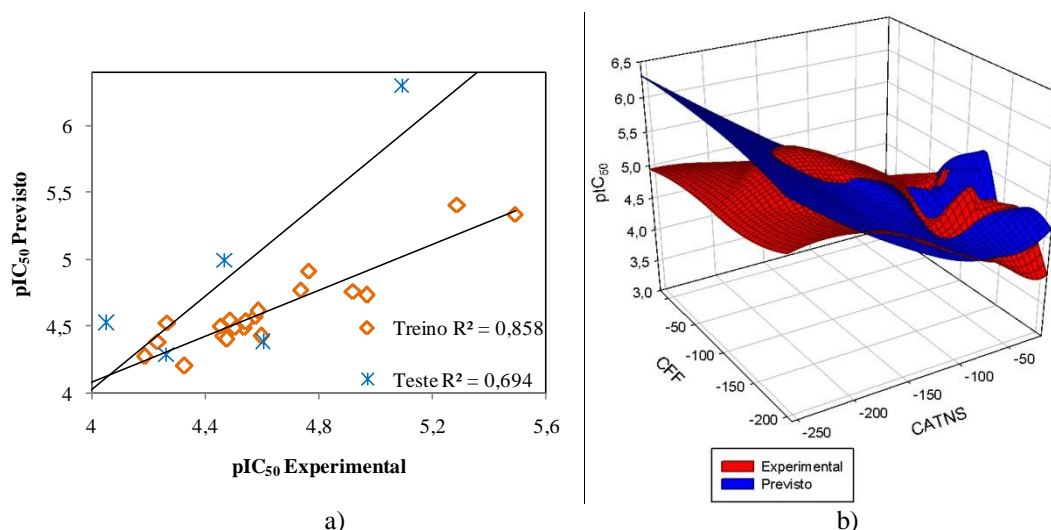
Para desenvolver estudos de QSAR em relação à linha celular do cancro do pulmão utilizando estes compostos, foram usados os valores experimentais dos parâmetros IC<sub>50</sub>, GI<sub>50</sub>, TGI, LC<sub>50</sub> e a média dos mesmos (Tabela 11F - Anexo F).

Obtiveram-se modelos válidos para os parâmetros IC<sub>50</sub>, LC<sub>50</sub> e a média dos parâmetros. Para a obtenção destes modelos foram utilizados 26 compostos, divididos entre o conjunto de treino e de teste. Para o modelo do parâmetro IC<sub>50</sub> utilizou-se 20 compostos no conjunto de treino e 5 no de teste, procedendo-se à eliminação de um *outlier* (*Z-score* de 0,3). No caso do modelo para o LC<sub>50</sub>, foram utilizados conjuntos de treino e de teste constituídos por 21 e 4 compostos, respectivamente, tendo-se eliminado um *outlier* (*Z-score* de 0,07). Para obter o modelo referente à média dos vários parâmetros utilizados foi utilizado um conjunto de treino com 20 compostos e um conjunto de teste com 4 compostos, tendo-se procedido à eliminação de dois *outliers*, detectados com base nos valores de *Z-score* de 18 e de 2, respectivamente.

A relação entre o parâmetro de actividade IC<sub>50</sub> e os descritores moleculares é descrita através da seguinte equação de QSAR

$$pIC_{50} = 4,363 - 1,047 \times 10^{-2}(\text{CATNS}) - 1,242 \times 10^{-2}(\text{SASAALH}) + 22,904(\text{CDLH/ATSM}) + 2,042 \times 10^{-3}(\text{CFF}) \quad \text{Equação 23}$$

$$n_{\text{treino}} = 20; R^2_{\text{treino}} = 0,858; R^2_{\text{cv}} = 0,729; n_{\text{teste}} = 5; R^2_{\text{teste}} = 0,694$$



**Figura 57:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $IC_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

Pela análise do modelo obtido para o parâmetro  $IC_{50}$ , é possível verificar que o aumento da actividade biológica pode ser obtido aumentando a razão entre os centros doadores de ligações de hidrogénio e a área total da superfície molecular e o calor final de formação, e diminuindo a contribuição da área total negativa da superfície e a soma das áreas da superfície dos átomos aceitadores de ligações de hidrogénio.

Com base nos valores obtidos para o teste t, verifica-se que o descritor molecular que contribui mais significativamente é a contribuição da área total negativa da superfície (CATNS). Este pertence ao grupo de descritores da área superficial com carga parcial e descreve as interacções polares entre moléculas. <sup>[98]</sup>

O descritor menos significativo para este modelo é o calor final de formação (CFF), estando este relacionado com a energia mecânica quântica que fornece a energia da molécula na escala padrão termodinâmica e caracteriza a estabilidade das moléculas. <sup>[99]</sup>

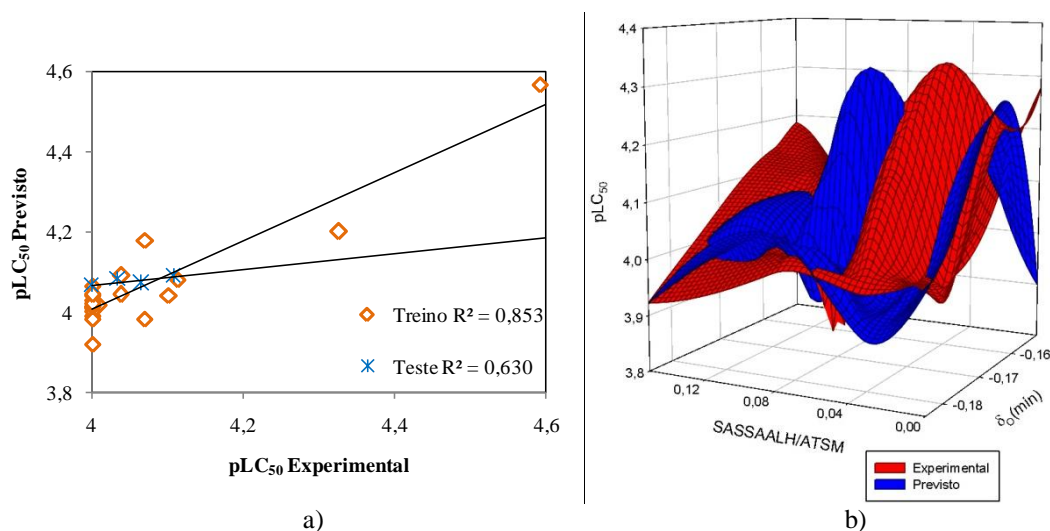
O segundo descritor mais significativo é a soma das áreas da superfície dos átomos aceitadores de ligações de hidrogénio (SASAALH), este descritor descreve a força da interacção electrostática intermolecular. <sup>[100]</sup>

A razão entre os centros doadores de ligações de hidrogénio e a área total da superfície molecular (CDLH/ATSM) é um descritor molecular que está relacionado com a capacidade da molécula estabelecer ligações de hidrogénio. <sup>[101]</sup>

A relação entre o parâmetro  $LC_{50}$  e os descritores moleculares é descrita pela equação que se segue

$$pLC_{50} = 1,649 \times 10^2 + 11,057(\delta_O(\text{min})) - 0,522(\text{EAEMinAO}) + 1,565(\text{SRYZ}) - 0,786(\text{SASAALH/ATSM}) \quad \text{Equação 24}$$

$$n_{\text{treino}} = 21; R^2_{\text{treino}} = 0,853; R^2_{\text{cv}} = 0,606; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,630$$



**Figura 58:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $LC_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $LC_{50}$ , em função dos descritores mais e menos significativos.

Pela análise do modelo obtido utilizando o parâmetro  $LC_{50}$  constata-se que para obter um aumento da actividade biológica é necessário aumentar a carga parcial mínima para o átomo de oxigénio e a razão entre a sombra projectada no plano YZ e o rectângulo no plano YZ. É também necessário diminuir o valor do estado atómico de energia mínima para o átomo de oxigénio e a razão entre a soma das áreas da superfície dos átomos aceitadores de ligações de hidrogénio e a área total da superfície molecular.

A carga parcial mínima para o átomo de oxigénio ( $\delta_O(\text{min})$ ) é o descritor molecular que apresenta uma maior contribuição para este modelo, sendo este um descritor electrostático que descreve as interacções intermoleculares e a polaridade molecular. [58]

O segundo descritor mais significativo é o estado atómico de energia mínima para o átomo de oxigénio (EAEMinAO), sendo que este caracteriza a estrutura electrónica dos compostos. [102]

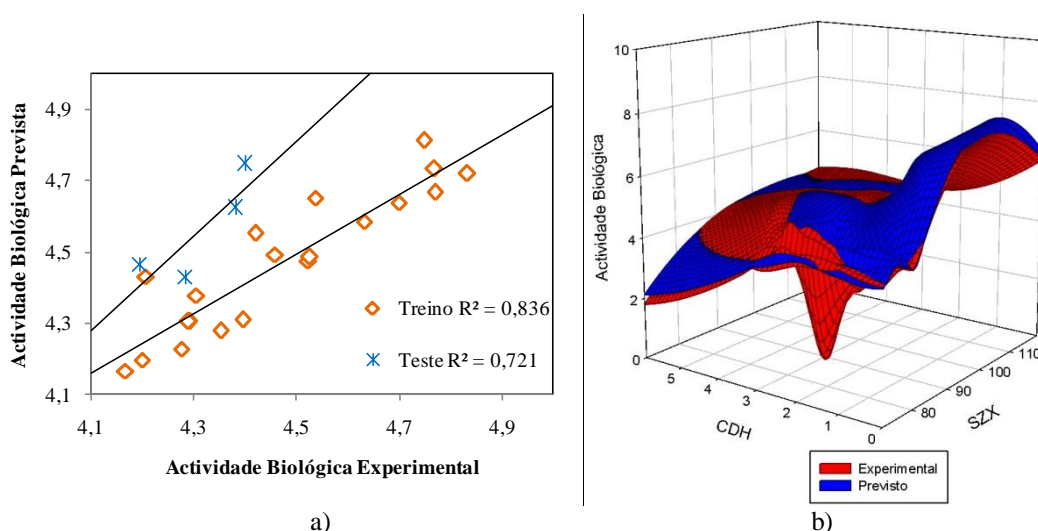
A razão entre a sombra projectada no plano YZ e o rectângulo no plano YZ (SRYZ) pertence ao grupo de descritores geométricos, descrevendo o tamanho e a forma da molécula. [83]

A razão entre a soma das áreas da superfície dos átomos aceitadores de ligações de hidrogénio e a área total da superfície molecular (SASAALH/ATSM) é o descritor menos significativo, estando relacionada com a capacidade dos compostos para aceitar ligações de hidrogénio. [101]

A equação de QSAR que descreve a relação entre a média das actividades biológicas e os descritores moleculares encontra-se seguidamente representada

$$\text{Média (pIC}_{50}, \text{pGI}_{50}, \text{pTGI}, \text{pLC}_{50}) = 0,333 + 1,088 \times 10^{-2}(\text{SZX}) + 4,559(\text{SRYZ}) + 2,176 \times 10^{11}(\text{MinR}_{\text{AO}}) - 7,190 \times 10^{-2}(\text{CDH}) \quad \text{Equação 25}$$

$$n_{\text{treino}} = 20; R^2_{\text{treino}} = 0,836; R^2_{\text{cv}} = 0,724; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,721$$



**Figura 59:** a) Representação gráfica das correlações entre os valores experimentais e previstos da média dos parâmetros de actividade para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

O modelo obtido utilizando a média dos vários parâmetros indica que o aumento da actividade biológica pode ser alcançado através do aumento da sombra projectada no plano ZX, da razão entre a sombra projectada no plano YZ e o rectângulo no plano YZ e do índice de reactividade mínima para 1 electrão para o átomo de oxigénio, e diminuindo o número de centros de doadores de hidrogénio.

Com base nos valores obtidos para o teste t, verifica-se que o descritor molecular mais significativo é a sombra projectada no plano ZX (SZX). Este é um descritor geométrico calculado com base na projecção da sombra no plano bidimensional de acordo com a orientação no espaço. Este reflecte ainda o tamanho e a forma geométrica da molécula. <sup>[103]</sup>

O descritor molecular que apresenta uma menor contribuição é o número de centros de doadores de hidrogénio (CDH), este caracteriza as moléculas de acordo com o número de centros doadores de hidrogénio que são capazes de doar átomos de hidrogénio para o meio circundante. <sup>[99]</sup>

O índice de reactividade mínima para 1 electrão para o átomo de oxigénio (MinR<sub>AO</sub>) depende das energias da orbital de fronteira, efectua uma estimativa da susceptibilidade da molécula para participar em reacções radiculares e está relacionado com as alterações de volume. <sup>[72][95]</sup>

Analisando os vários modelos obtidos, é possível verificar que os coeficientes de correlação obtidos para o conjunto de treino encontram-se entre 0,836 e 0,858, enquanto que para o conjunto de teste estes encontram-se entre 0,630 e 0,721. Quanto aos parâmetros relativos à validação cruzada estes situam-se num intervalo de valores entre 0,606 e 0,729.

Os valores de RMSE,  $s^2$  e SEE são muito próximos de zero. Os valores calculados para o VIF são inferiores a 4, o que indica que não existe colinearidade entre os descritores.

Os valores obtidos para o teste t e F indicam que quer os modelos obtidos, quer os descritores moleculares neles envolvidos são estatisticamente válidos, uma vez que estes são superiores aos valores de referência.

Os dados estatísticos e das equações obtidas para os modelos referidos anteriormente são apresentados no Anexo F (F.5. e F.6.).

### **5.2.3.3. Linha Celular do Cancro do Cólon (WiDr)**

Com o objectivo de estabelecer relações quantitativas entre a estrutura e a actividade dos compostos, foram utilizados os valores experimentais para os parâmetros IC<sub>50</sub>, GI<sub>50</sub>, TGI, LC<sub>50</sub> e a média dos vários parâmetros (Tabela 12F - Anexo F). Apenas

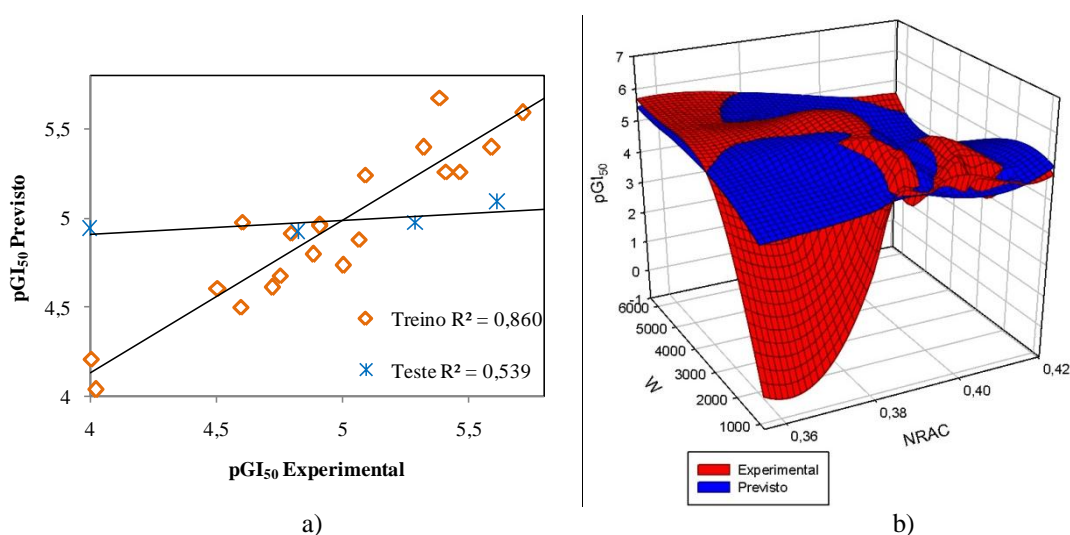
foi possível encontrar um modelo válido para o parâmetro  $GI_{50}$ , tendo sido utilizado um conjunto de treino constituído por 20 compostos e com 4 compostos no conjunto de teste, tendo sido eliminado 1 *outlier* detectado através do valor de *Z-score* de 15.

A relação entre o parâmetro de actividade  $GI_{50}$  e os descritores moleculares é descrita pela seguinte equação

$$pGI_{50} = 28,561 - 21,681(NRAC) + 6,025(SRXY) - 9,280(CMLOM) + 1,459(W)$$

**Equação 26**

$$n_{\text{treino}} = 20; R^2_{\text{treino}} = 0,860; R^2_{\text{cv}} = 0,760; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,539$$



**Figura 60:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $GI_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $GI_{50}$ , em função dos descritores mais e menos significativos.

Pela análise do modelo obtido, verifica-se que o aumento da razão entre a sombra projectada no plano XY e o rectângulo no plano XY, e do índice de Wiener pode provocar um aumento da actividade biológica dos compostos em estudo. Observa-se ainda que a diminuição do número relativo de átomos de carbono e da contribuição máxima para a ligação de uma orbital molecular poderá aumentar a actividade biológica.

Os valores do teste t indicam que o descritor mais significativo para este modelo é o número relativo de átomos de carbono (NRAC), este descritor é calculado pela razão entre o número de átomos de carbono e o número total de átomos da molécula.

Existindo uma dependência entre este descritor e o número de ligações duplas e triplas e a presença de heteroátomos. <sup>[104]</sup>

O descritor molecular que apresenta a menor contribuição é o índice de Wiener.

Analisando o modelo de QSAR obtido para o parâmetro de actividade  $GI_{50}$ , verifica-se que os coeficientes de correlação para os conjuntos de treino e de teste são 0,860 e 0,539, respectivamente. O parâmetro para a validação cruzada é de 0,760.

Os valores obtidos para RMSE,  $s^2$  e SEE estão muito próximos de zero e os valores referentes ao VIF são inferiores a 4, o que indica que não existe colinearidade entre os descritores envolvidos no modelo.

Relativamente aos testes t e F, os valores obtidos são superiores aos valores de referência, significando que os modelos obtidos e os descritores que neles participam são estatisticamente válidos.

Os dados estatísticos e da equação obtida para o modelos referido anteriormente são apresentados no Anexo F (F.5. e F.6.).

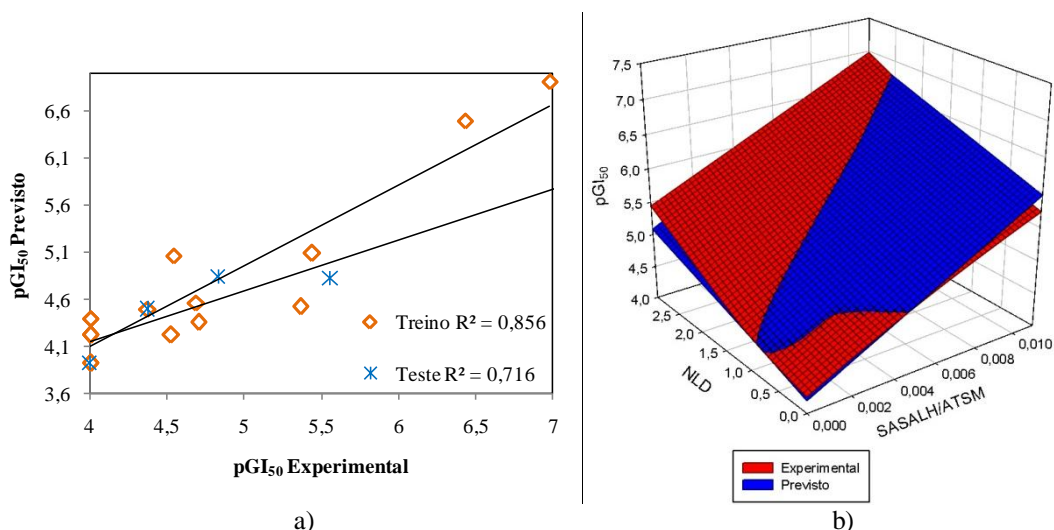
### 5.2.4. Série 12

#### 5.2.4.1. Linha Celular do Cancro do Ovário (A2780)

Com o intuito de encontrar relações quantitativas entre as estruturas e as actividades dos compostos pertencentes a esta série, foram utilizados os valores referentes aos parâmetros  $IC_{50}$ ,  $GI_{50}$ , TGI e  $LC_{50}$  e a média destes (Tabela 16F - Anexo F). Apenas foi possível encontrar um modelo válido, referente ao parâmetro de actividade  $GI_{50}$ , no qual foram utilizados 16 compostos no conjunto de treino e 4 no conjunto de teste.

A equação de QSAR que se segue descreve a relação entre o parâmetro de actividade  $GI_{50}$  e os descritores moleculares

$$pGI_{50} = 3,922 + 1,678 \times 10^2 (SASALH/ATSM) + 0,615(ASPR) + 0,301(NLD) \text{ Equação 27}$$
$$n_{\text{treino}} = 16; R^2_{\text{treino}} = 0,856; R^2_{\text{cv}} = 0,716; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,716$$



**Figura 61:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $GI_{50}$  para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $GI_{50}$ , em função dos descritores mais e menos significativos.

Analisando este modelo, verifica-se que para aumentar a actividade biológica destes compostos é necessário aumentar a razão entre a soma das áreas da superfície com aceptadores de ligações de hidrogénio e a área total da superfície molecular, a área relativa superficial positiva e o número de ligações duplas.

Com base nos valores para o teste t, constata-se que o descritor molecular que apresenta uma maior contribuição para este modelo é a razão entre a soma das áreas da superfície com aceptadores de ligações de hidrogénio e a área total da superfície molecular (SASALH/ATSM). Este descritor está relacionado com as propriedades dos compostos para aceitar ligações de hidrogénio. [84]

O número de ligações duplas (NLD) é o descritor molecular de menor contribuição, este é um descritor molecular que descreve o grau de deslocalização electrónica. [105]

Em termos de resultados, este modelo apresenta um coeficiente de correlação para o conjunto de treino de 0,856 e de 0,716 para o conjunto de teste, sendo o parâmetro para a validação cruzada de 0,716.

Os valores obtidos para RMSE,  $s^2$  e SEE são muito próximos de zero e os valores de VIF são inferiores a 4, apresentando um máximo de 3,722. Estes valores de VIF indicam que não existe colinearidade entre os descritores envolvidos no modelo.

Os valores para os testes t e F resultantes são superiores aos de referência, isto significa que o modelo e os descritores envolvidos no modelo são estatisticamente válidos.

Os dados estatísticos e da equação obtida para o modelos referido anteriormente são apresentados no Anexo F (F.5. e F.6.).

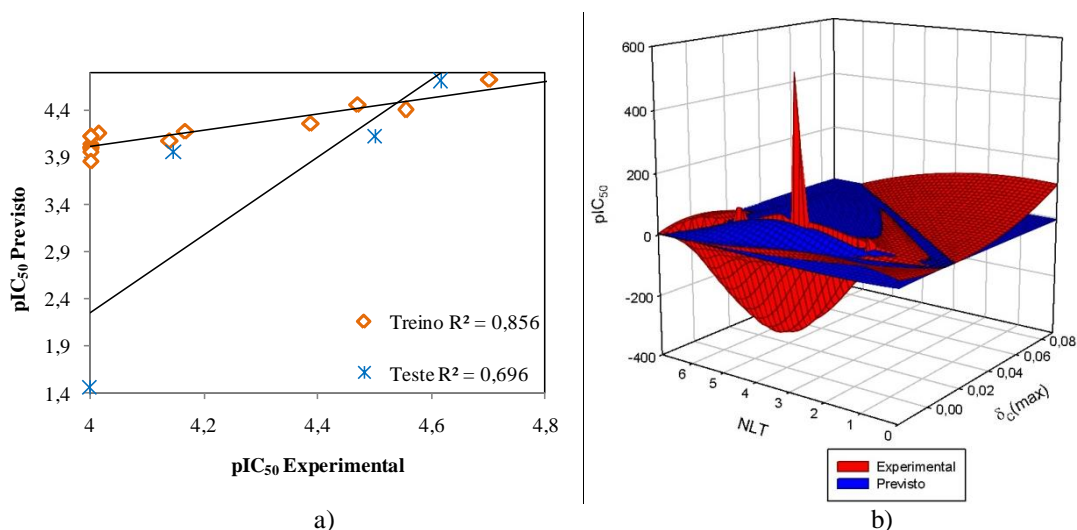
#### 5.2.4.2. Linha Celular do Carcinoma Mamário (T-47D)

Para estabelecer relações quantitativas entre a estrutura e a actividade dos compostos, foram utilizados os valores para os diversos parâmetros de actividades e a média dos mesmos (Tabela 17F - Anexo F). No entanto, apenas foi encontrado um modelo para o parâmetro IC<sub>50</sub>. Para este modelo foram utilizados 19 compostos, 15 compostos no conjunto de treino e 4 no conjunto de teste.

A relação entre o parâmetro de actividade IC<sub>50</sub> e os descritores moleculares é descrita pela equação que se segue

$$pIC_{50} = 3,676 - 32,669(\delta_c(\max)) + 2,114 \times 10^{-2}(NL) - 8,318 \times 10^{-2}(NLT) \quad \text{Equação 28}$$

$$n_{\text{treino}} = 15; R^2_{\text{treino}} = 0,856; R^2_{\text{cv}} = 0,688; n_{\text{teste}} = 4; R^2_{\text{teste}} = 0,696$$



**Figura 62:** a) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro IC<sub>50</sub> para os conjuntos de treino e de teste; b) Gráfico de superfície dos valores experimentais e previstos para o parâmetro GI<sub>50</sub>, em função dos descritores mais e menos significativos.

O aumento da actividade biológica pode ser obtido através do aumento do número de ligações, e da diminuição da carga máxima para o átomo de carbono e do número de ligações triplas.

O descritor molecular que apresenta uma maior contribuição para este modelo é a carga parcial máxima para o átomo de carbono ( $\delta_C(\text{max})$ ), estando relacionado com a distribuição de carga existente na molécula. <sup>[70]</sup>

O segundo descritor mais significativo é o número de ligações (NL), este descritor constitucional está relacionado com o movimento rotacional. <sup>[100]</sup>

Os coeficientes de correlação para os conjuntos de treino e de teste são 0,856 e 0,696, respectivamente.

Os valores para o RMSE, SEE e  $s^2$  são muito próximos de zero e o VIF são inferiores a 4, indicando ausência de colinearidade entre os descritores.

O valor de validação cruzada é superior a 0,6, o que significa que o modelo é válido. Os valores para os testes t e F, superiores aos valores de referência, indicam que o modelo e os descritores moleculares envolvidos no mesmo são estatisticamente válidos.

Os dados estatísticos e da equação obtida para o modelo referido anteriormente são apresentados no Anexo F (F.5. e F.6.).

### 5.3. Redes Neurais Artificiais (ANN)

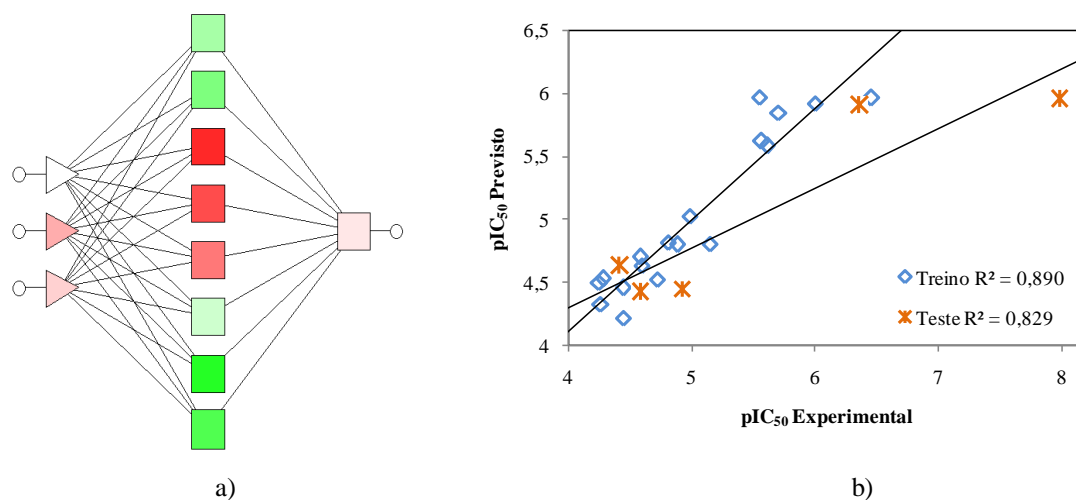
Com o objectivo de estabelecer relações não lineares entre a actividade biológica dos compostos em estudo e os descritores moleculares, utilizaram-se os descritores que, através do método Heurístico anteriormente aplicado, demonstraram estar mais correlacionados com as diversas propriedades em estudo. Para a realização deste trabalho utilizaram-se os mesmos conjuntos de treino e de teste usados no método linear e aplicaram-se os métodos não lineares aos conjuntos com os quais se obtiveram modelos válidos através dos métodos lineares.

#### 5.3.1. Série 1

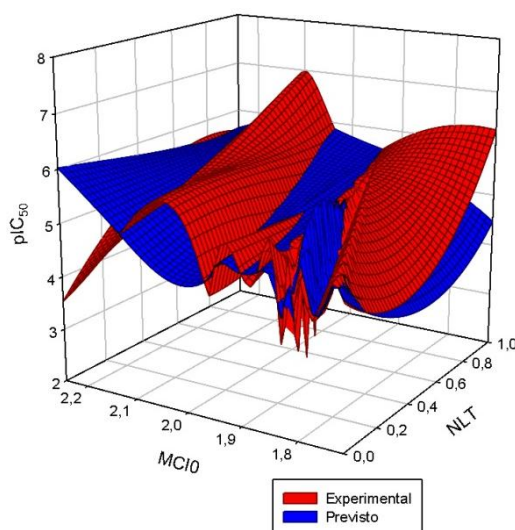
##### 5.3.1.1. Linha Celular do Cancro do Ovário (A2780)

Com base nos valores dos parâmetros de actividade  $IC_{50}$ ,  $GI_{50}$ ,  $LC_{50}$  e a média dos mesmos, aplicaram-se métodos não lineares para relacionar a actividade biológica e os descritores moleculares para a linha celular do cancro do ovário. Como resultado, obtiveram-se redes neuronais para os parâmetros  $IC_{50}$ ,  $LC_{50}$  e para a média dos diversos parâmetros em estudo (Tabela 4F – Anexo F).

A relação não linear para o parâmetro  $IC_{50}$  foi obtida utilizando o método *Multilayer Perceptrons* (MLP), deste resultou uma rede neuronal de estrutura 3-8-1, na qual foram utilizados três descritores moleculares. Os descritores envolvidos nesta rede neuronal são o número de ligações triplas, a carga negativa relativa e a média do conteúdo de informação (ordem 0). Relativamente aos valores dos coeficientes de correlação para os conjuntos de treino e de teste, estes apresentaram valores de 0,890 e de 0,829, respectivamente.



**Figura 63:** a) Representação esquemática da rede neuronal (3-8-1), b) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $IC_{50}$  para os conjuntos de treino e de teste.

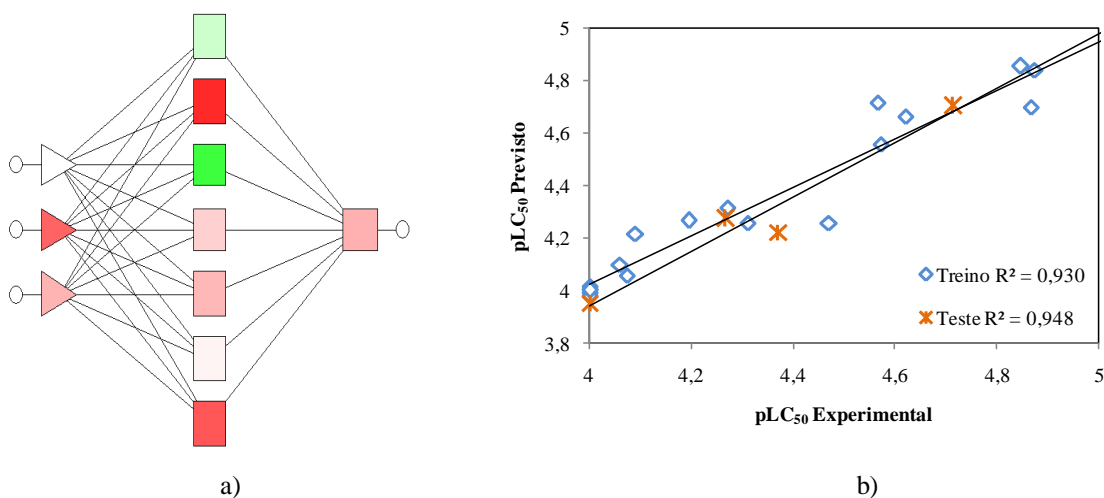


**Figura 64:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

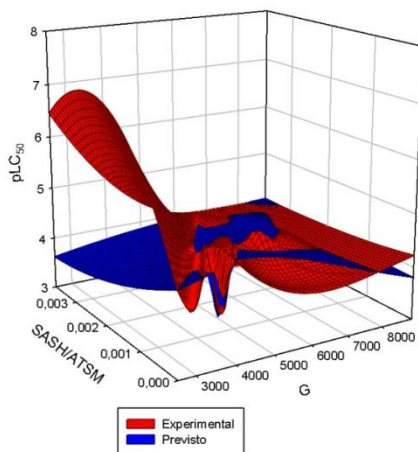
A análise das redes neurais artificiais não permite uma fácil interpretação das contribuições dos descritores, no entanto, com base na razão entre o desempenho das redes neurais antes e depois da remoção de cada descritor (análise de sensibilidade) é possível determinar a significância dos descritores moleculares para o modelo. Com base neste método, para esta rede neuronal, o descritor que apresenta uma maior significância é o número de ligações triplas. O descritor molecular de menor significância para este modelo não linear é a média do conteúdo de informação (ordem 0).

A carga negativa relativa (CNR) é um descritor químico-quântico que representa a distribuição de carga, contribuindo também para descrever as interações electrostáticas. [106]

Utilizando os valores referentes ao parâmetro de actividade  $LC_{50}$ , foi obtida uma rede neuronal de estrutura 3-7-1 através do método *Multilayer Perceptrons* (MLP). Nesta rede neuronal encontram-se envolvidos três descritores moleculares, sendo estes o número de ligações triplas, o índice de gravitação (todos os pares), e a razão entre a soma das áreas das superfícies contendo hidrogénio e a área total da superfície molecular. Efectuando a representação gráfica para os conjuntos de treino e de teste, foram obtidos coeficientes de correlação de 0,930 e de 0,948, respectivamente.



**Figura 65:** a) Representação esquemática da rede neuronal (3-7-1), b) Representação gráfica das correlações entre os valores experimentais e previstos do parâmetro  $LC_{50}$  para os conjuntos de treino e de teste.

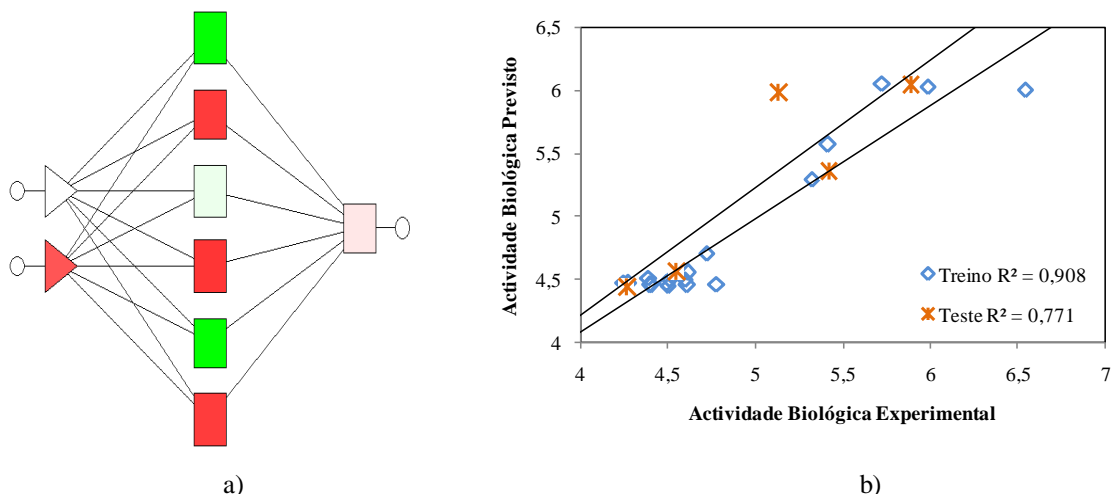


**Figura 66:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $LC_{50}$ , em função dos descritores mais e menos significativos.

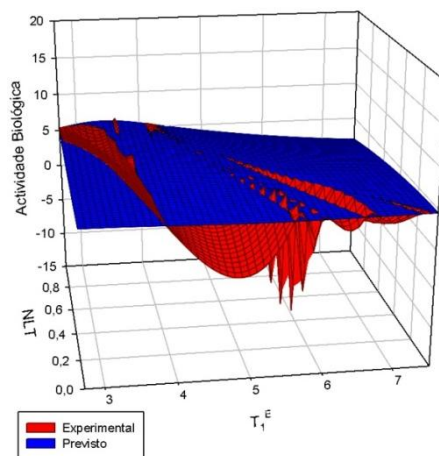
Através do método da análise da sensibilidade é possível constatar que o descritor molecular que apresenta uma maior razão entre o desempenho da rede neuronal antes e após a remoção do mesmo é o que contribui mais significativamente para o modelo. O índice de gravitação (todos os pares) ( $G^1$ ) é o descritor molecular mais significativo para este modelo. Este descritor é calculado tendo em conta todos os pares de átomos, contribuindo para este cálculo as massas atómicas (volumes) e a distribuição dos átomos no interior do espaço molecular e também efectua a quantificação da coesão da estrutura do composto com base nas interações hidrófobas. <sup>[107]</sup>

O descritor molecular que apresenta uma menor significância é a razão entre a soma das áreas das superfícies contendo hidrogénio e a área total da superfície molecular (SASH/ATSM). Este é um descritor electrostático que representa a área da superfície multiplicada pela carga parcial correspondente, reflecte a capacidade de um composto para actuar como doador de átomos de hidrogénios ao interagir com o meio químico, estando também relacionado com a capacidade dos compostos para formar de ligações de hidrogénio. <sup>[108][109]</sup>

Para estabelecer relações não lineares entre a actividade biológica e os descritores moleculares para a linha celular do cancro do ovário, foram ainda utilizados os valores da média dos diversos parâmetros de actividade em estudo. Desta forma, foi obtida uma rede neuronal 2-6-1 recorrendo ao método *Multilayer Perceptrons* (MLP). Os descritores moleculares que participam nesta rede são o número de ligações triplas e o índice electrónico topográfico (todos os pares). O coeficiente de correlação obtido para o conjunto de treino foi de 0,908 e de 0,771 para o conjunto de teste.



**Figura 67:** a) Representação esquemática da rede neuronal (2-6-1), b) Representação gráfica das correlações entre os valores experimentais e previstos da média dos vários parâmetros de actividade para os conjuntos de treino e de teste.



**Figura 68:** Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

Com base nos valores obtidos para a razão entre o desempenho da rede neuronal antes e depois da remoção de cada descritor, verifica-se que o descritor molecular mais significativo é o índice electrónico topográfico (todos os pares) ( $T_1^E$ ), tendo apresentado o valor mais elevado para esta razão. Este descritor molecular pertence ao grupo dos descritores electrostáticos, descreve as características electrostáticas da molécula como um todo, estando também relacionado com as interacções intermoleculares polares.<sup>[110][111]</sup>

O descritor molecular menos significativo para este modelo é o número de ligações triplas.

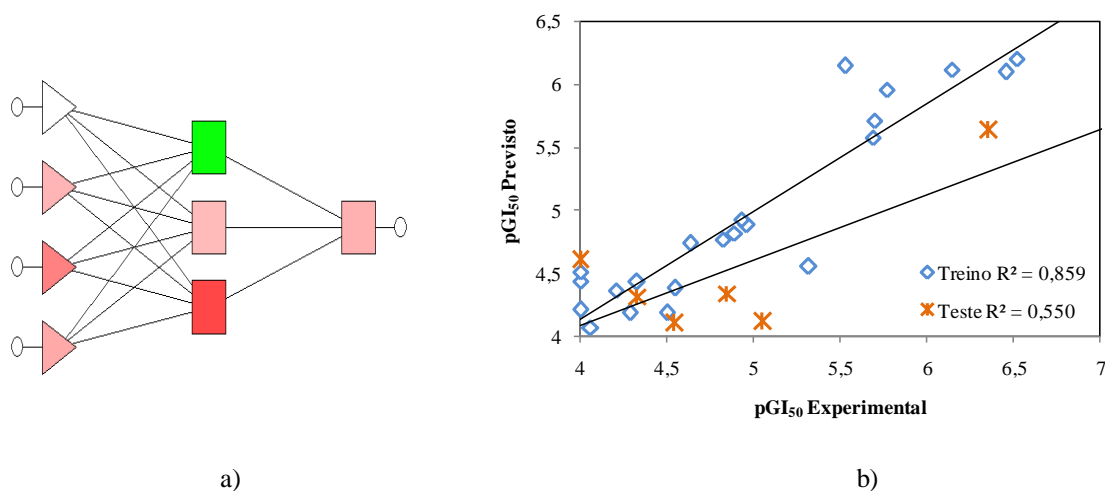
Comparando os resultados obtidos para as relações lineares e as não lineares entre a actividade biológica e os descritores moleculares para os compostos que constituem esta série em relação à linha celular do cancro do ovário, verifica-se que utilizando as relações lineares obtivemos um maior número de modelos, um para cada parâmetro de actividade e para a média dos diversos parâmetros. Por outro lado, através das relações não lineares, apenas obtivemos modelos para os parâmetros  $IC_{50}$ ,  $LC_{50}$  e para a média dos vários parâmetros em estudo.

Relativamente aos coeficientes de correlação, constata-se que nos casos em que se obteve modelos válidos utilizando as redes neuronais artificiais, os coeficientes de correlação quer do conjunto de treino, quer do de teste são superiores aos obtidos utilizando as relações lineares, à excepção do modelo obtido para a média. Para a obtenção do modelo linear foram utilizados quatro descritores, enquanto que para o não linear foram utilizados apenas dois descritores, para os restantes modelos foram utilizados três descritores moleculares quer para as relações lineares, quer para as não lineares.

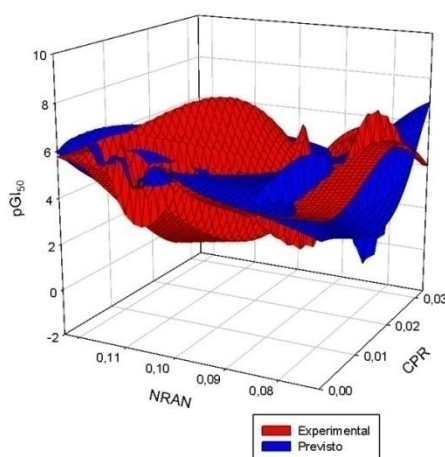
### 5.3.1.2. Linha Celular do Cancro do Pulmão (SW1573)

Para a linha celular do cancro do pulmão, aplicaram-se os métodos não lineares utilizando os valores referentes aos parâmetros  $IC_{50}$  e  $GI_{50}$  e à média das actividades. No entanto, apenas se obtiveram redes neuronais artificiais para o parâmetro  $GI_{50}$  e para a média (Tabela 5F – Anexo F).

Para o parâmetro  $GI_{50}$  obteve-se um modelo não linear, com rede neuronal artificial de estrutura 4-3-1, recorrendo ao método *Multilayer Perceptrons* (MLP). Neste modelo estão envolvidos quatro descritores moleculares, o número de ligações triplas, o índice de Balaban, o número relativo de átomos de azoto e a carga positiva relativa. Os coeficientes de correlação obtidos para os conjuntos de treino e de teste foram 0,859 e 0,550, respectivamente.



**Figura 69:** a) Representação esquemática da rede neuronal (4-3-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro GI<sub>50</sub> para os conjuntos de treino e de teste.



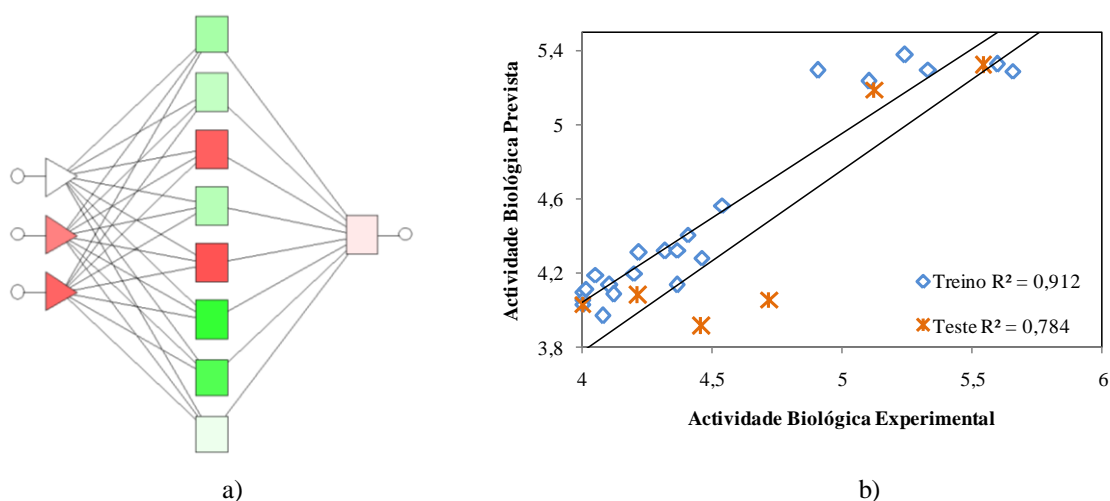
**Figura 70:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro GI<sub>50</sub>, em função dos descritores mais e menos significativos.

Pelo método da análise da sensibilidade constata-se que o descritor molecular que contribui mais significativamente para este modelo não linear é a carga positiva relativa, enquanto que o menos significativo é o número relativo de átomos de azoto.

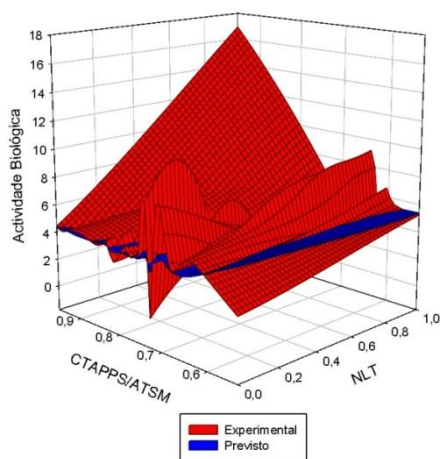
O terceiro descritor molecular mais significativo para este modelo não linear é o índice de Balaban, este é um descritor topológico que descreve a conectividade atômica, fornece informações acerca das ramificações existentes na molécula e estando relacionado com as interações hidrofóbicas que se estabelecem entre as moléculas.<sup>[106]</sup>

Utilizando os valores da média dos diversos parâmetros de actividade, estabeleceu-se uma relação não linear entre a actividade biológica e os descritores

moleculares, obtendo-se uma rede neuronal de estrutura 3-8-1 utilizando o método *Multilayer Perceptrons* (MLP). No referido modelo participam três descritores moleculares, o número de ligações triplas, o número relativo de átomos de azoto, e a fracção entre a carga total da área parcial positiva da superfície e a área total da superfície molecular. Pela representação gráfica da correlação entre os valores experimentais e previstos obteve-se coeficientes de correlação para os conjuntos de treino e de teste de 0,912 e de 0,784, respectivamente.



**Figura 71:** a) Representação esquemática da rede neuronal (3-8-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para a média dos diversos parâmetros de actividade para os conjuntos de treino e de teste.



**Figura 72:** Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

Através da análise das razões entre o desempenho da rede neuronal antes e depois da remoção de cada descritor, conclui-se que o descritor molecular mais

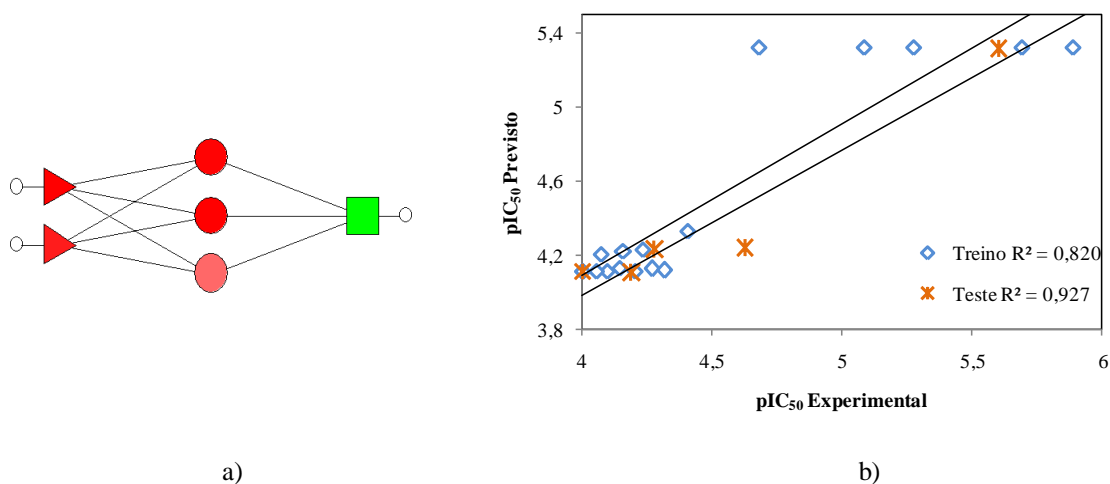
significativo para este modelo é o número de ligações triplas. O descritor que apresenta a menor contribuição é a razão entre a carga total da área parcial positiva da superfície e a área total da superfície molecular (CTAPPS/ATSM), esta caracteriza a área relativa parcial positiva da superfície. <sup>[112]</sup>

Procedendo à comparação dos resultados obtidos utilizando métodos não lineares e lineares para a linha celular do cancro do pulmão, verifica-se que utilizando o método não linear não foi possível obter modelos válidos para o parâmetro  $IC_{50}$ , como se obteve através do método linear. Quanto aos modelos obtidos para o parâmetro  $GI_{50}$  e para a média das actividades, observa-se que os valores dos coeficientes de correlação para os conjuntos de treino e de teste são superiores utilizando as redes neuronais artificiais. Para obter os modelos para o parâmetro  $GI_{50}$ , foi utilizado o mesmo número de descritores, enquanto que para a média das actividades foram utilizados quatro descritores no método linear e três descritores no método não linear.

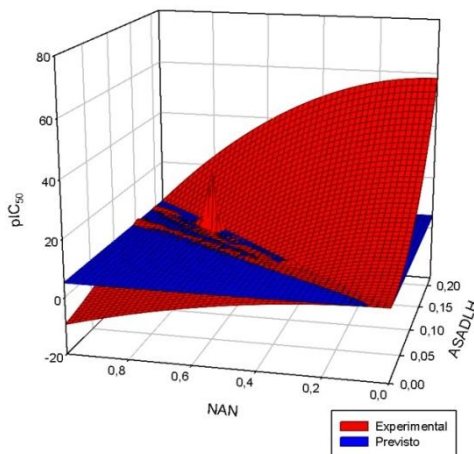
### 5.3.1.3. Linha Celular do Carcinoma Mamário (T-47D)

Com o intuito de estabelecer relações não lineares entre a actividade biológica e os descritores moleculares, utilizaram-se os valores referentes aos parâmetros de actividade  $IC_{50}$ , TGI e à média das actividades. Por aplicação dos métodos não lineares, obteve-se um modelo para o parâmetro  $IC_{50}$  (Tabela 6F – Anexo F).

Para o parâmetro de actividade  $IC_{50}$  obteve-se uma rede neuronal de estrutura 2:3:1 utilizando o método *Radial Basis Function* (RBF). Neste modelo estão envolvidos dois descritores moleculares, o número de átomos de azoto e a área acessível ao solvente dos átomos doadores de ligações de hidrogénio. Os coeficientes de correlação obtidos para os conjuntos de treino e de teste foram de 0,820 e 0,927, respectivamente.



**Figura 73:** a) Representação esquemática da rede neuronal (2-3-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro de actividade  $IC_{50}$  para os conjuntos de treino e de teste.



**Figura 74:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

O método da análise da sensibilidade permite constatar que o descritor mais significativo para este modelo é a área acessível ao solvente dos átomos doadores de ligações de hidrogénio e o menos significativo é o número de átomos de azoto. Este último descritor molecular pertence ao grupo dos descritores constitucionais e está relacionado com a capacidade da molécula para formar ligações de hidrogénio. [74]

A área acessível ao solvente dos átomos doadores de ligações de hidrogénio (ASADLH) é um descritor electrostático que descreve as propriedades associadas com a aceitação de ligações de hidrogénio por parte dos compostos. [106]

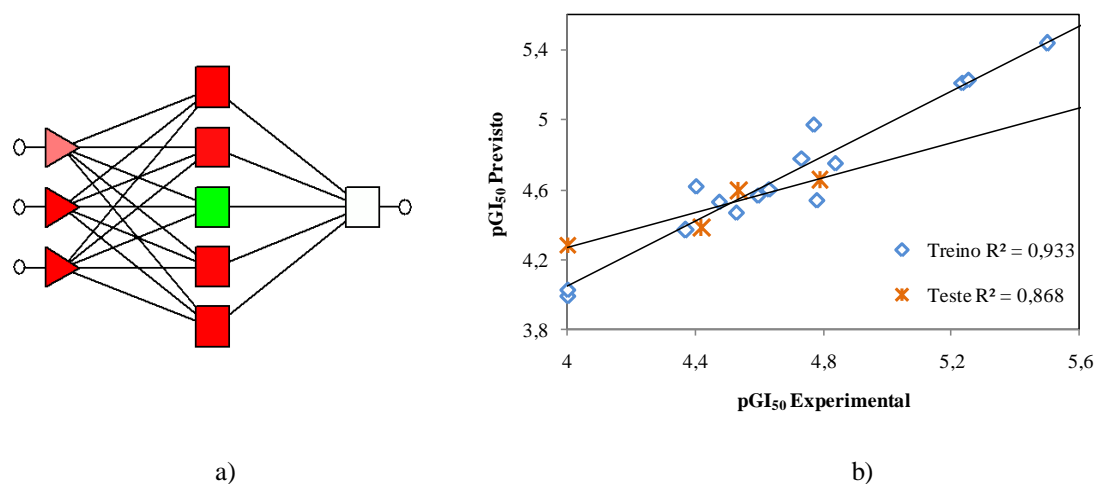
Comparando os resultados obtidos pela utilização de métodos lineares e não lineares para a linha celular do carcinoma mamário, constata-se que utilizando os métodos lineares obtêm-se valores de coeficientes de correlação superiores aos obtidos utilizando os métodos não lineares. Para a obtenção dos modelos utilizando métodos lineares foram utilizados três descritores, enquanto que para obter modelos não lineares foram utilizados apenas dois descritores.

### 5.3.2. Série 7

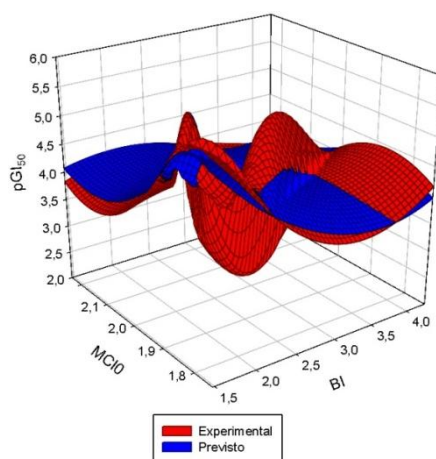
#### 5.3.2.1. Linha Celular da Leucemia (HL60)

Recorrendo aos valores referentes ao parâmetro de actividade  $GI_{50}$  e à média das actividades, aplicou-se métodos não lineares para relacionar a actividade biológica com os descritores moleculares, tendo sido obtidos modelos para cada um destes parâmetros (Tabela 9F – Anexo F).

Em relação ao parâmetro  $GI_{50}$ , obteve-se um modelo não linear com rede neuronal de estrutura 3-5-1 através do método *Multilayer Perceptrons* (MLP). Nesta rede neuronal participam os seguintes descritores, o índice de Balaban, o número relativo de átomos de oxigénio e a média do conteúdo de informação (ordem 0). Os coeficientes de correlação obtidos para os conjuntos de treino e de teste para este modelo são 0,933 e 0,868, respectivamente.



**Figura 75:** a) Representação esquemática da rede neuronal (3-5-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro  $GI_{50}$  para os conjuntos de treino e de teste.

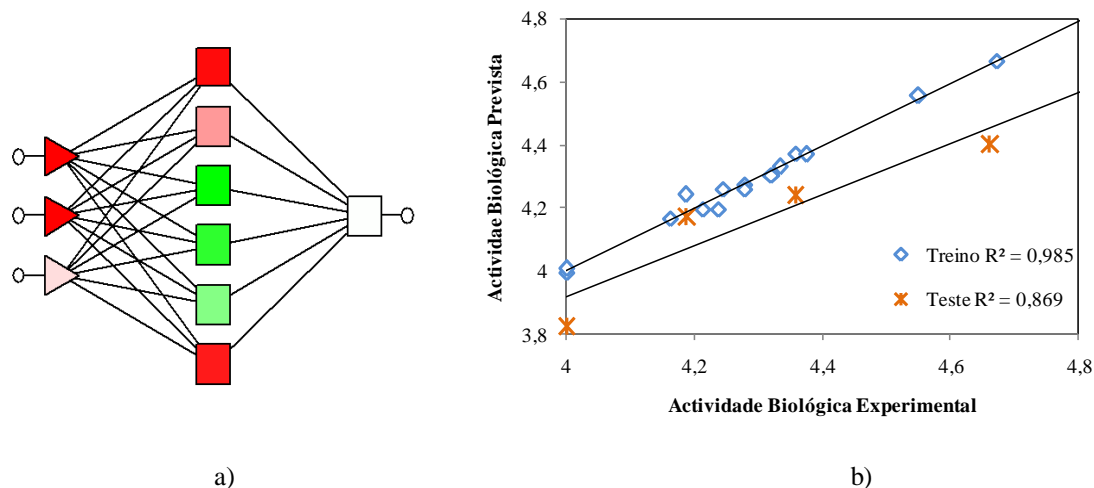


**Figura 76:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $GI_{50}$ , em função dos descritores mais e menos significativos.

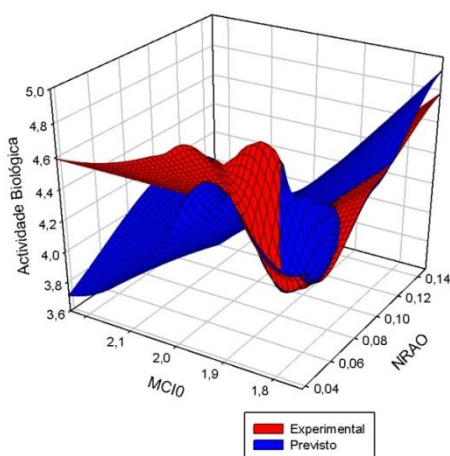
Pela análise da sensibilidade dos descritores moleculares envolvidos neste modelo, é possível constatar que o descritor mais significativo é o índice de Balaban. Por outro lado, o descritor que apresenta uma menor contribuição é a média do conteúdo de informação (ordem 0).

O descritor menos significativo é o número relativo de átomos de oxigénio (NRAO), este é um descritor constitucional que afecta a densidade da nuvem electrónica da molécula, reflectindo ainda a capacidade para o composto aceitar ligações de hidrogénio. <sup>[113]</sup>

Utilizando os valores referentes à média das actividades, obteve-se uma rede neuronal de estrutura 3-6-1 através do método *Multilayer Perceptrons* (MLP). Neste modelo participam três descritores, a média do conteúdo de informação (ordem 0), o número relativo de átomos de oxigénio e o número de átomos de carbono. Através da representação gráfica das correlações entre os dados experimentais e previstos para os conjuntos de treino e de teste, verifica-se que os coeficientes de correlação são de 0,985 e 0,869, respectivamente.



**Figura 77:** a) Representação esquemática da rede neuronal (3-6-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para a média dos parâmetros de actividade para os conjuntos de treino e de teste.



**Figura 78:** Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de actividade, em função dos descritores mais e menos significativos.

A análise da sensibilidade dos descritores moleculares que participam neste modelo, demonstra que o descritor que contribui mais significativamente é o número

relativo de átomos de oxigénio e o menos significativo é a média do conteúdo de informação (ordem 0).

O segundo descritor mais significativo é o número de átomos de carbono, este descritor está relacionado com a constituição e com o tamanho das moléculas.<sup>[83]</sup>

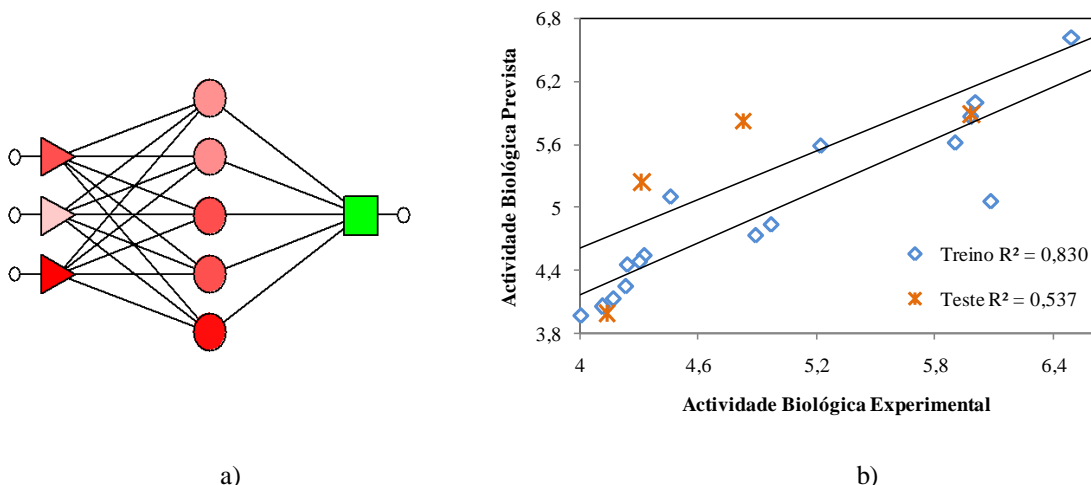
Procedendo à comparação dos modelos lineares e não lineares obtidos para a linha celular da leucemia, é possível concluir que utilizando o mesmo número de descritores, ao aplicar os métodos não lineares obtiveram-se valores de coeficientes de correlação superiores aos obtidos utilizando os métodos lineares.

### 5.3.3. Série 9

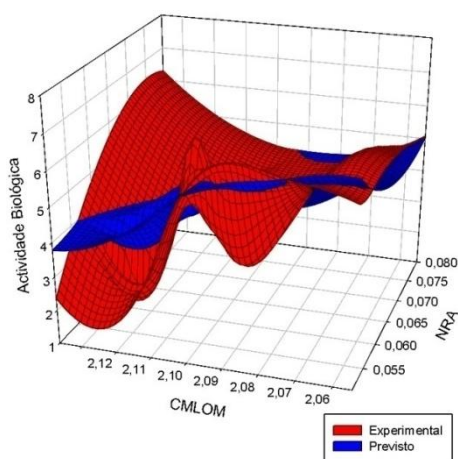
#### 5.3.3.1. Linha Celular do Cancro do Ovário (A2780)

Para a linha celular do cancro do ovário utilizaram-se os valores correspondentes à média dos diversos parâmetros de actividade em estudo com o objectivo de obter relações não lineares entre a actividade biológica dos compostos e os descritores moleculares (Tabela 13F – Anexo F).

A rede neuronal artificial para a média dos diversos parâmetros de actividade foi obtida utilizando o método *Radial Basis Function* (RBF) e apresenta a estrutura 3-5-1. Neste modelo não linear participam os descritores calor final de formação, número relativo de anéis e contribuição máxima para a ligação de uma orbital molecular. Para este modelo obtiveram-se coeficientes de correlação de 0,830 para o conjunto de treino e de 0,537 para o conjunto de teste.



**Figura 79:** a) Representação esquemática da rede neuronal (3-5-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para a média dos parâmetros de atividade para os conjuntos de treino e de teste.



**Figura 80:** Gráfico de superfície dos valores experimentais e previstos para a média dos parâmetros de atividade, em função dos descritores mais e menos significativos.

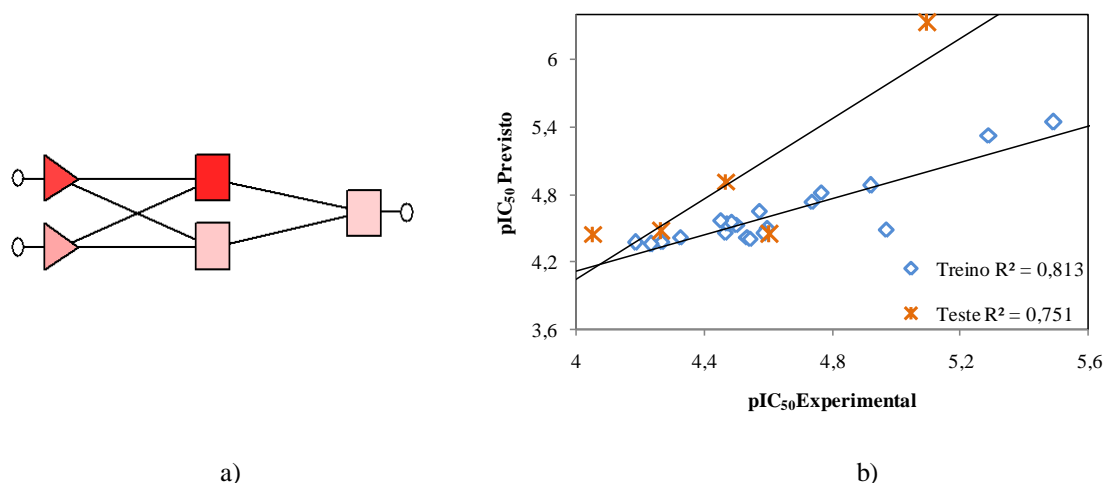
Em termos de significância dos descritores moleculares, com base na análise da sensibilidade, conclui-se que o número relativo de anéis é o descritor mais significativo. O descritor molecular que apresenta a menor contribuição é a contribuição máxima para a ligação de uma orbital molecular.

Comparando os resultados obtidos utilizando o método linear e o não linear, para a linha celular do cancro do ovário, observa-se que através do método linear obtêm-se valores de coeficientes de correlação superiores aos obtidos pelo método não linear, utilizando o mesmo número de descritores moleculares.

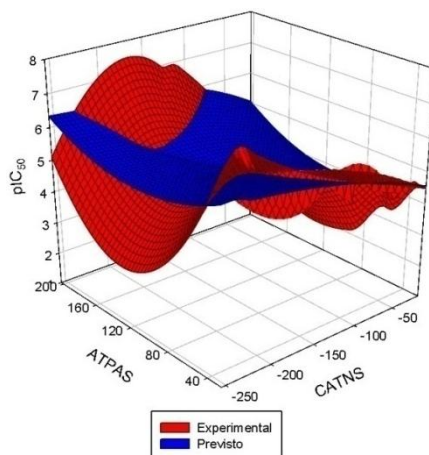
### 5.3.3.2. Linha Celular do Cancro do Pulmão (SW1573)

Com o intuito de estabelecer relações não lineares entre a actividade biológica e os descritores moleculares, utilizaram-se os valores referentes aos parâmetros de actividade  $IC_{50}$ ,  $LC_{50}$  e à média das actividades. Como resultado, obtiveram-se modelos para os parâmetros  $IC_{50}$  e  $LC_{50}$  (Tabela 14F – Anexo F).

No modelo não linear obtido para o parâmetro de actividade  $IC_{50}$ , estão envolvidos dois descritores moleculares, a contribuição da área total negativa da superfície e a área total das cargas parciais das áreas superficiais. A rede neuronal artificial foi obtida utilizando o método *Multilayer Perceptrons* (MLP) e apresenta a estrutura 2-2-1. Os coeficientes de correlação obtidos para os conjuntos de treino e de teste foram de 0,813 e 0,751, respectivamente.



**Figura 81:** a) Representação esquemática da rede neuronal (2-2-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro  $IC_{50}$  para os conjuntos de treino e de teste.

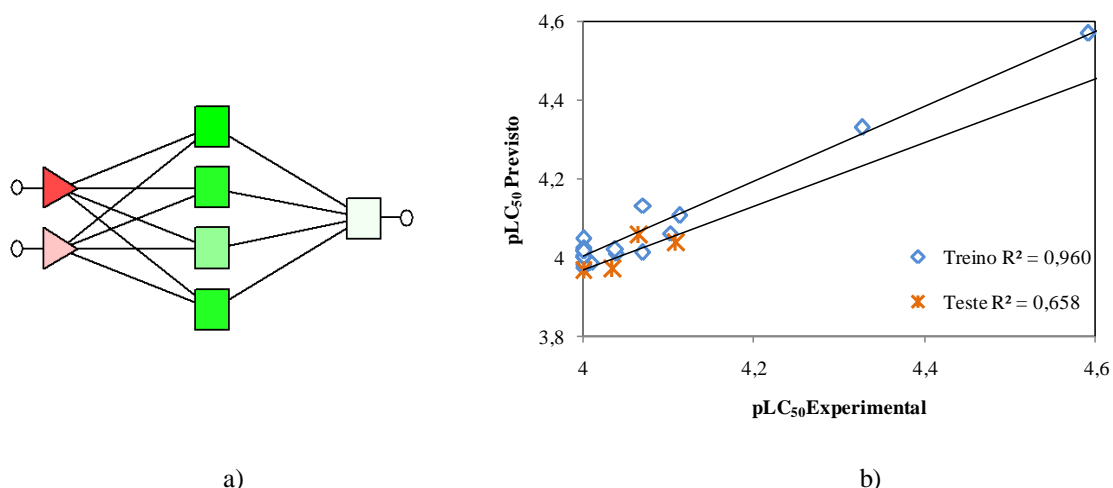


**Figura 82:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

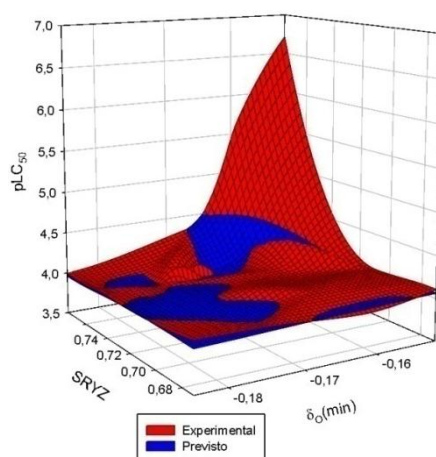
A análise da sensibilidade dos descritores moleculares permite determinar a significância de cada descritor para o modelo não linear, com base neste método, determina-se que, para o modelo em questão, o descritor molecular mais significativo é a contribuição da área total negativa da superfície.

A área total das cargas parciais das áreas superficiais (ATPAS) é um descritor electrostático relacionado com a carga parcial das áreas superficiais, que se baseia na área superficial e na distribuição de carga da molécula, combinando desta forma, a informação electrónica e acerca da forma que caracteriza a molécula. Descreve ainda características responsáveis pelas interacções polares entre as moléculas e indica o efeito da distribuição da carga negativa na molécula. <sup>[81]</sup>

Para o parâmetro  $LC_{50}$  obteve-se um modelo não linear através do método *Multilayer Perceptrons* (MLP), sendo a rede neuronal artificial de estrutura 2-4-1. Nesta rede neuronal estão envolvidos dois descritores, razão entre a sombra projectada no plano YZ e o rectângulo no plano YZ e a carga parcial mínima para o átomo de oxigénio. Os coeficientes de correlação obtidos para os conjuntos de treino e de teste foram de 0,960 e 0,658, respectivamente.



**Figura 83:** a) Representação esquemática da rede neuronal (2-4-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro  $LC_{50}$  para os conjuntos de treino e de teste.



**Figura 84:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $LC_{50}$ , em função dos descritores mais e menos significativos.

Com base na razão entre o desempenho das redes neurais artificiais antes e depois da remoção de cada descritor molecular é possível determinar a significância de cada descritor para as redes neurais. Sendo assim, o descritor molecular que apresenta a maior contribuição é a carga parcial mínima para o átomo de oxigênio.

Procedendo à comparação entre o modelo linear e não linear para a linha celular do cancro do pulmão observa-se que através dos métodos lineares obtêm-se modelos para um maior número de parâmetros, enquanto que aplicando os métodos não lineares apenas se obteve modelos para os parâmetros  $IC_{50}$  e  $LC_{50}$ . Em relação ao parâmetro  $IC_{50}$ , verifica-se que através do modelo linear o valor do coeficiente de correlação para

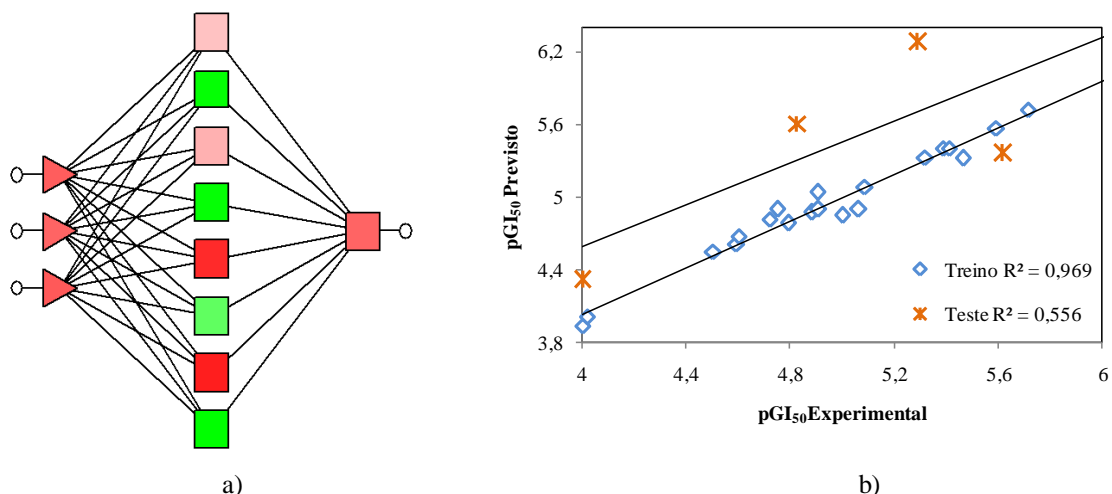
o conjunto de treino é superior ao obtido através do modelo não linear. No entanto, o valor do coeficiente de correlação para o conjunto de teste é superior quando o modelo não linear é utilizado.

No caso dos modelos obtidos para o parâmetro  $LC_{50}$ , constata-se que utilizando o modelo não linear obtêm-se valores dos coeficientes de correlação, para ambos os conjuntos, superiores aos obtidos através do modelo linear. Nos modelos lineares obtidos estão envolvidos quatro descritores moleculares, enquanto que nos não lineares estão envolvidos apenas dois descritores.

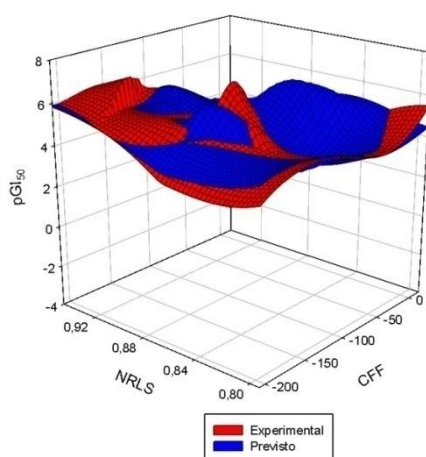
### 5.3.3.3. *Linha Celular do Cancro do Cólon (WiDr)*

Para a linha celular do cancro do cólon, utilizaram-se os valores referentes ao parâmetro de actividade  $GI_{50}$  para estabelecer relações não lineares entre a actividade biológica e os descritores moleculares, tendo-se obtido um modelo para este parâmetro (Tabela 15F – Anexo F).

O modelo não linear para o parâmetro  $GI_{50}$  foi obtido utilizando o método *Multilayer Perceptrons* (MLP), sendo a rede neuronal artificial de estrutura 3-8-1. Neste modelo não linear estão envolvidos os descritores: número relativo de átomos de carbono, número relativo de ligações simples e calor final de formação. Pela representação gráfica obtiveram-se coeficientes de correlação de 0,969 e 0,556 para os conjuntos de treino e de teste, respectivamente.



**Figura 85:** a) Representação esquemática da rede neuronal (3-8-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro GI<sub>50</sub> para os conjuntos de treino e de teste.



**Figura 86:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro GI<sub>50</sub>, em função dos descritores mais e menos significativos.

Com base na análise da sensibilidade dos descritores moleculares, é possível concluir que o descritor mais significativo para este modelo é o calor final de formação, pois apresenta o maior valor da razão entre o desempenho da rede neuronal antes e após a remoção do descritor. O descritor molecular que apresenta a menor contribuição para este modelo é o número relativo de ligações simples (NRLS), este é um descritor constitucional que afecta a densidade da nuvem electrónica da molécula, afectando assim a polaridade da mesma. <sup>[103]</sup>

Comparando os modelos lineares e não lineares obtidos para a linha celular do cancro do cólon, verifica-se que os valores dos coeficientes de correlação obtidos

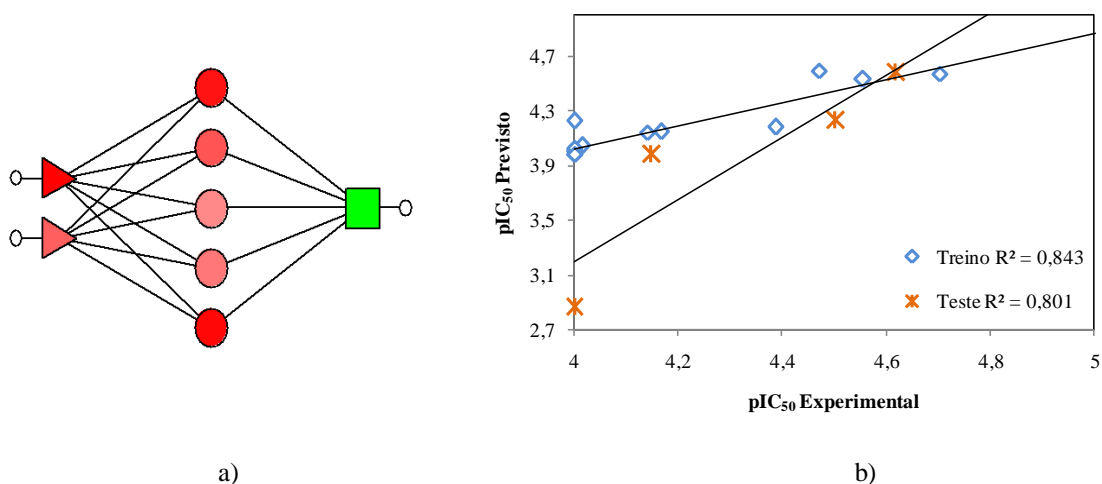
através do método não linear são superiores aos obtidos pelo método linear. Foram utilizados três descritores moleculares para os métodos não lineares e quatro para os métodos lineares.

### 5.3.4. Série 12

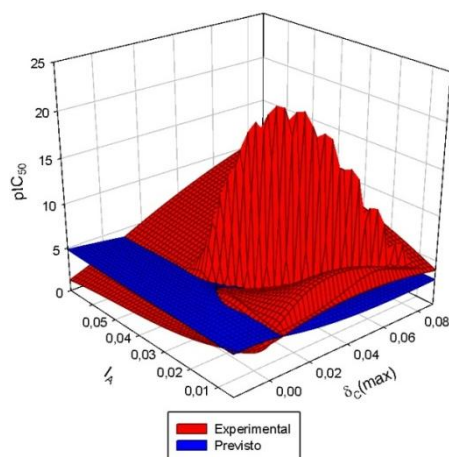
#### 5.3.4.1. Linha Celular do Carcinoma Mamário (T-47D)

Aplicando métodos não lineares aos dados referentes ao parâmetro  $IC_{50}$  para a linha celular do carcinoma mamário, obteve-se uma rede neuronal artificial para relacionar a actividade biológica dos compostos com os descritores moleculares (Tabela 18F – Anexo F).

O modelo obtido para o parâmetro  $IC_{50}$  resultou da aplicação do método *Radial Basis Function* (RBF), obtendo-se uma rede neuronal de estrutura 2-5-1. Neste modelo não linear estão envolvidos dois descritores, a carga parcial máxima para o átomo de carbono e o momento de inércia A. Relativamente aos coeficientes de correlação obtidos para este modelo, estes foram de 0,843 e de 0,801 para os conjuntos de treino e de teste, respectivamente.



**Figura 87:** a) Representação esquemática da rede neuronal (2-5-1), b) Representação gráfica das correlações entre os valores experimentais e previstos para o parâmetro  $IC_{50}$  para os conjuntos de treino e de teste.



**Figura 88:** Gráfico de superfície dos valores experimentais e previstos para o parâmetro  $IC_{50}$ , em função dos descritores mais e menos significativos.

Analisando os descritores moleculares em termos de significância, é possível concluir que o descritor molecular mais significativo para este modelo é a carga parcial máxima para o átomo de carbono. O momento de inércia A é o descritor molecular que apresenta o menor valor para a razão entre o desempenho da rede neuronal antes e após a remoção deste descritor sendo, desta forma, o descritor menos significativo.

Comparando o modelo linear e não linear obtido para a linha celular do carcinoma mamário, verifica-se que o coeficiente de correlação obtido para o conjunto de treino utilizando o modelo linear é superior ao obtido utilizando o modelo não linear. Em relação ao coeficiente de correlação obtido para o conjunto de teste, este é superior quando utilizado o método não linear. O número de descritores utilizados no método linear (três descritores) foi superior ao utilizado no método não linear (dois descritores).

# CAPÍTULO VI



## CONCLUSÕES E PERSPECTIVAS FUTURAS

## Conclusões

O presente trabalho teve como principais objectivos a identificação dos alvos moleculares dos compostos em estudo, a identificação das estruturas base dos compostos com os quais as enzimas interagem preferencialmente e o estabelecimento de correlações entre os dados obtidos computacionalmente e os dados resultantes dos testes celulares. Os objectivos incluíram ainda a identificação das características estruturais relevantes para a actividade anti-cancerígena dos compostos e a determinação dos modelos, lineares ou não lineares, são mais adequados para relacionar os descritores moleculares com a actividade biológica para cada tipo de cancro estudado.

Ao ser realizado o estudo da forma como os compostos interagem com as enzimas seleccionadas foi possível concluir que:

- I. As representações gráficas para a distribuição das energias obtidas através do estudo da interacção entre os compostos e cada uma das enzimas apresentam um perfil típico de distribuição, no qual os melhores ligandos são os que se encontram na zona em que há uma acentuação do declive para a parte negativa do gráfico.
- II. As enzimas que apresentaram melhores energias de interacção ao ser realizado o *docking* entre estas com os compostos em estudo foram as enzimas CDC25C, CDC25B, p53, SKP1, CHK1, Tubulina, HO-1, 5-LOX e EF-G (bactéria).
- III. Os valores de constante de dissociação obtidos encontram-se na ordem dos nanomolar e micromolar, o que significa que o estudo deve ser aprofundado para alguns dos compostos e enzimas estudadas.
- IV. A correcção dos resultados elimina eventuais enviesamentos nos resultados dos ligandos, tendo ocorrido uma reordenação pouco significativa dos ligandos.
- V. Ao ser realizada a correcção os valores de erro são inferiores do que quando o cálculo do erro é efectuado com os resultados não corrigidos.

- VI. Os compostos estudados apresentam uma especificidade superior em relação à cinase AURKA, que actua em células cancerígenas, do que em relação às famílias de cinases que actuaem quer em células cancerígenas, quer em células saudáveis.
- VII. A realização do teste de enriquecimento provocou um enriquecimento dos resultados obtidos entre duas a cinquenta vezes.
- VIII. Foram estabelecidas algumas correlações entre os resultados obtidos computacionalmente e os dados obtidos através dos testes celulares, envolvendo por exemplo as enzimas CDC25C, COX, PAK1, MAP2K1, CHK1, AURKA, entre outras. Os coeficientes de correlação apresentaram valores entre 0,607 e 0,996.
- IX. Com base nas correlações estabelecidas foi verificado que nos casos em que ocorreu a inibição de 50% da amostra é provável que os compostos tenham realizado a inibição das enzimas PAK1, CDC2, CDC25C, DD-Ligase, NA, AKT1, CDC7, PLK1, PIN1, PAK1, p27 KIP1, Asp, COX, ERK1, PIK3CG e NA (complexado com zanamivir). Para a inibição de 50% do crescimento das células é provável que os compostos inibam as enzimas CDC25C, 5-LOX, NA, DD-Ligase, E2F, TOP2A, IMPDH, CDC7, MCM2, PLK1, PIN1, COX, mPGES-1, Racemase da Alanina, CCNH, MAP2K1, Asp, Eg5, p21 CIP1, AURKA, Importina B, NA (complexado com zanamivir) e PIK3CG.
- X. No caso em que ocorre a inibição total do crescimento é provável que as enzimas inibidas pelos compostos em estudos sejam as enzimas CDC2, TOP2A, CDC25C, COX, DD-Transpeptidase, DD-Ligase, CDC7, CDK7, SKP1, KITLG, Eg5, ROCK, FNTB, NA (complexado com zanamivir), CDC25B, PIK3CG, PAK1, CHK1, TOP1, 5-LOX, p53, AURKA, p107 e EF-G (fungo). Para a indução da morte de 50% da amostra é provável que os compostos inibam as enzimas p53, TOP1, CDC25C, ERK2 TOPOIV, DD-Ligase, CDC2, AURKA, CHK2, CCNH, mPGES-1 e WEE1.

Ao ser efectuada a aplicação de métodos lineares e não lineares para relacionar a estrutura e a actividade dos compostos foi possível retirar as conclusões que se seguem:

- I. Os descritores moleculares que se encontraram envolvidos em diversos modelos lineares foram o número de ligações triplas (número total e relativo), o momento de inércia A, os descritores associados com as sombras projectadas nos planos XY, YZ e ZX (estando estes relacionados com o tamanho das moléculas) e o índice de Wiener.
- II. Os descritores moleculares que se encontraram envolvidos em diversos modelos não lineares foram o número de ligações triplas, o número de átomos de azoto (número total e relativo) e o número relativo de átomos de oxigénio.
- III. Os coeficientes de correlação obtidos para os conjuntos de treino através da aplicação de modelos lineares encontraram-se entre 0,811 e 0,928. Para os conjuntos de teste, os coeficientes de correlação obtidos apresentaram valores no intervalo de 0,521 e 0,880.
- IV. Através dos métodos não lineares, os valores dos coeficientes de correlação obtidos para os conjuntos de treino encontraram-se entre 0,813 e 0,985. Relativamente aos coeficientes de correlação referentes aos conjuntos de testes estes situaram-se num intervalo entre os valores 0,537 e 0,948.
- V. Aplicando os métodos lineares obtiveram-se um maior número de modelos do que aplicando os métodos não lineares. No entanto, verificou-se que os métodos não lineares efectuem uma melhor relação entre a estrutura e a actividade para as linhas celulares HL60 e WiDr. Para as linhas celulares A2780 e T-47D, os modelos não lineares revelaram-se mais adequados para estabelecer este tipo de relação. No entanto, em relação à linha celular SW1573 não foi possível determinar qual dos modelos, linear ou não linear, é mais adequado para estabelecer a relação entre a estrutura e a actividade.

## Perspectivas Futuras

Com a realização deste trabalho, foi possível estabelecer os potenciais alvos moleculares para os compostos em estudo, estabelecer correlações entre os dados obtidos computacionalmente e os dados provenientes dos testes celulares, identificar algumas das características relevantes para a actividade anti-cancerígena dos compostos, obter modelos lineares e não lineares para relacionar os descritores moleculares com a actividade biológica em relação a diversas linhas celulares. No entanto, futuramente poderão ser realizados:

- I. Diversos estudos experimentais para comprovar os resultados obtidos neste trabalho.
- II. Alterações ao nível estrutural e das características que se apresentaram como sendo relevantes para a actividade anti-cancerígena dos compostos de forma a ser obtida uma melhor interacção entre estes e os alvos.
- III. Alargar estes métodos a outros conjuntos de compostos com actividade demonstrada experimentalmente em relação a outras doenças como tentativa de identificação dos alvos moleculares e das características relevantes para a actividade biológica em questão.
- IV. Aperfeiçoamentos dos métodos utilizados.

---

**Bibliografia**

- [1] Gad, S. C.; *Drug Discovery Handbook*, John Wiley & Sons, Inc., **2005**.
- [2] Krovat, E. M.; Steindl, T.; Langer T.; *Current Computer-Aided Drug Design*, **2005**, *1*, 93-102.
- [3] Lill, M. A.; *Drug Discov. Today*, **2007**, *12*, 1013-1017.
- [4] Livingstone, D. J.; Manallack, D. T.; *QSAR Comb. Sci.*, **2003**, *22*, 510-518.
- [5] Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P.; *J. Mol. Graph Model.*, **2007**, *26*, 198-212.
- [6] Smith, C.; Marks, A.; Lieberman, M. A.; *Marks' Basic Medical Biochemistry: A Clinical Approach*, 2<sup>nd</sup> ed., Lippincott Williams & Wilkins, **2005**.
- [7] Abraham, D. J.; *Burger's Medicinal Chemistry and Drug Discovery*, 6<sup>th</sup> ed, Volume 5, John Wiley & Sons, Inc., **2003**.
- [8] Koolman, J.; Roehm, K. H.; *Color Atlas of Biochemistry*, 2<sup>nd</sup> ed., Thieme, **2005**.
- [9] Gabriel, J.; *The Biology of Cancer*, 2<sup>nd</sup> ed, John Wiley & Sons Ltd, **2007**.
- [10] Alison, M. R.; *The Cancer Handbook*, 2<sup>nd</sup> ed, John Wiley & Sons, Inc., **2007**.
- [11] Nelson, D. L.; Cox, M. M.; *Lehninger Principles of Biochemistry*, 4th ed., W H Freeman & Co, **2005**.
- [12] Lewin, B.; *Genes VIII*, Pearson Prentice Hall, **2004**.
- [13] <http://www.cancer.gov/cancertopics/factsheet/Therapy/biological> (24/04/2009)
- [14] <http://www.cancerhelp.org.uk/help/default.asp?page=28761> (24/04/2009)
- [15] <http://www.cancer.gov/cancertopics/factsheet/Therapy/radiation> (24/04/2009)
- [16] <http://www.cancerbackup.org.uk/Cancertype/Headneck/Treatment/Chemotherapy> (24/04/2009)
- [17] <http://www.elmhurst.edu/~chm/vchembook/655cancer.html> (27/04/2009)
- [18] <http://www.cancer.gov/cancertopics/factsheet/Therapy/targeted> (27/04/2009)
- [19] <http://www.elmhurst.edu/~chm/vchembook/650drugs.html> (27/04/2009)
- [20] <http://www.elmhurst.edu/~chm/vchembook/651enzymeinhibit.html> (27/04/2009)
- [21] <http://www.elmhurst.edu/~chm/vchembook/660drugreceptor.html> (27/04/2009)
- [22] Blagosklonny, M. V.; *Cell Cycle*, **2005**, *4*, 1518-1521.
- [23] Maddika S., Booy, E. P.; Johar, D.; Gibson, S. B.; Ghavami, S.; Los, M.; *J. Cell Sci.*, **2005**, *118*, 4485-4493.
- [24] Berdis, A. J.; *Biochemistry*, **2008**, *47*, 8253-8260.

- [25] O'Day, S. J.; Hamid, O.; Urba, W. J.; *Cancer*, **2007**, *15*, 2614-2627.
- [26] Goetz, M. P.; Toft, D. O.; Ames, M. M.; Erlichman, C.; *Ann. Oncol.*, **2003**, *14*, 1169-1176.
- [27] <http://www.netsci.org/Science/Compchem/feature12.html> (05/05/2009)
- [28] <http://cnx.org/content/m11113/latest/> (05/05/2009)
- [29] <http://www.cs.rice.edu/CS/Robotics/bioinformatics/drug.html> (05/05/2009)
- [30] Tang, Y.; Zhu, W.; Chen, K.; Jiang, H.; *Drug Discovery Today: Technologies*, **2006**, *3*, 307-313.
- [31] Paul, N.; Kellenberger, E.; Bret, G.; Müller, P.; Rognan, D.; *PROTEINS: Structure, Function and Bioinformatics*, **2004**, *54*, 671-680.
- [32] Frembgen-Kesner, T.; Elcock, A. H.; *J. Mol. Biol.*, **2006**, *359*, 202-214.
- [33] Moro, S., *et al.*; *J. Med. Chem.* **2005**, *48*, 152-162.
- [34] Marzio, W. D.; Galassi, S.; Todeschini, R.; Consolaro, F.; *Chemosphere*, **2001**, *44*, 401-406.
- [35] Bermúdez-Saldaña, J. M.; Escuder-Gilabert, L.; Medina-Hernández M. J.; Villanueva-Camañas, R. M.; Sagrado, S.; *J. Chromatogr. A*, **2005**, *1063*, 153-160.
- [36] Casalegno, M.; Sello, G.; Benfenati, E.; *Chem. Res. Toxicol.*, **2006**, *19*, 1533-1539.
- [37] Chang, J.; Lei B.; Li, J.; Li, S.; Shen, Y.; Yao, X.; *QSAR Comb. Sci.*, **2008**, *27*, 1318 - 1325.
- [38] Bhattacharjee, A. K.; Gordon, J. A.; Marek, E.; Campbell, A.; Gordon; R. K.; *Bioorg. Med. Chem.*, **2009**, *17*, 3999-4012.
- [39] Leonetti, F.; Favia, A.; Rao, A.; Aliano, R.; Paluszczak, A.; Hartmann, R. W.; Carotti, A.; *J. Med. Chem.*, **2004**, *47*, 6792-6803.
- [40] Hu, R.; Barbault, F.; Delamar, M.; Zhang, R.; *Bioorg. Med. Chem.*, **2009**, *17*, 2400-2409.
- [41] Cajka, T.; Hajslova, J.; Pudil, F.; Riddellova, K.; *J. Chromatogr. A*, **2009**, *1216*, 1458-1462.
- [42] Marini, F.; *Anal. Chim. Acta*, **2009**, *635*, 121-131.
- [43] Dou, Y.; Sun, Y.; Ren, Y.; Ren, Y.; *Anal. Chim. Acta*, **2005**, *528*, 55-61.
- [44] Santos Jr., V. O.; Oliveira, F. C.C.; Lima, D. G.; Petry, A. C.; Garcia, E.; Suarez, P. A.Z.; Rubim, J. C.; *Anal. Chim. Acta*, **2005**, *547*, 188-196.
- [45] Marengo, E.; Bobba, M.; Robotti, E.; Lenti, M.; *Anal. Chim. Acta*, **2004**, *511*, 313-322.
- [46] Abbass, H. A.; *Artif. Intell. Med.*, **2002**, *25*, 265-281.

- [47] Lisboa, P. J., Taktak, A. F.G.; *Neural Networks*, **2006**, *19*, 408–415.
- [48] Zhou, Z.; Jiang, Y.; Yang, Y.; Chen, S.; *Artif. Intell. Med.*, **2002**, *24*, 25-36.
- [49] Matsoukas, J.; Mavromoustakos, T.; *Drug Discovery and Design: Medical Aspects*; IOS Press; Netherlands, **2002**.
- [50] Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P.; *Curr. Protein Pept. Sc.*, **2006**, *7*, 1-15.
- [51] <http://www.rcsb.org/pdb/home/home.do> (3/10/2008)
- [52] <http://www.sb-roscoff.fr/CyCell/Frames20.htm> (2/10/2008)
- [53] Bourne, P. E.; Weissig, H.; *Structural Bioinformatics*, Wiley-Liss, Inc., **2003**.
- [54] Cavasotto, C. N.; Phatak, S. S.; *Drug Discov. Today*, **2009**, *14*, 676-683.
- [55] Vigers, G. P. A. ; Rizzi, J. P. ; *J. Med. Chem.*, **2004**, *47*, 80-89.
- [56] <http://dud.docking.org/> (3/10/2008)
- [57] Nicholls, A.; *ROCS*, OpenEye Scientific Software, Inc., **2007**.
- [58] Karelson, M.; Lobanov, V. S.; Katritzky, A. R.; *Chem. Rev.*, **1996**, *96*, 1027-1044.
- [59] Zupan, J.; Gasteiger, J.; *Neural Networks in Chemistry and Drug Design*, 2<sup>nd</sup> Edition, Wiley-VCH, Weinheim, **1999**.
- [60] Winkler, D. A.; *Brief. Bioinform.*, **2002**, *3*, 73-86.
- [61] Liu, P.; Long, W.; *Int. J. Mol. Sci.*, **2009**, *10*, 1978-1998.
- [62] Kubinyi, H.; Folkers, G.; Martin, Y. C.; *3D QSAR in Drug Design, Volume 2*, Kluwer Academic Publishers, London, **2002**.
- [63] Katritzky, A.R.; Lobanov, V.S.; Karelson, M.; *Comprehensive Descriptors for Structural and Statistical Analysis*, Reference Manual, **1994**.
- [64] Verma, R. P.; Hansch, C.; *Chem. Rev.*, **2009**, *109*, 213-235.
- [65] Dohnal, V.; Kuča, K.; Jun, D.; *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.*, **2005**, *149*, 221–224.
- [66] Yilmaz, A. S.; Özer, Z.; *Expert Syst. Appl.*, **2009**, *36*, 9767–9775.
- [67] StatSoft, Inc. (2007). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: <http://www.statsoft.com/textbook/stathome.html> (02/04/2009)
- [68] <http://kinasedb.ontology.ims.u-tokyo.ac.jp:8081/php/DbProteinGroup.php?mode=gview> (29/09/2008)
- [69] Huang, N.; Shoichet, B. K.; Irwin, J. J.; *J. Med. Chem.* **2006**, *49*, 6789-6801.
- [70] Büinz, A.P.; Braun, B.; Janowsky, R.; *Fluid Phase Equilib.*, **1999**, *158–160*, 367–374.

- [71] Katritzky, A. R.; Dobchev, D. A.; Tulp, I.; Karelson, M.; Carlson, D. A.; *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 2306–2311.
- [72] Miller, M. D.; Holder, A. J.; Kilway, K. V.; Giese, G. J.; Finley, J. E.; Travis, D. M.; Iwai, B. T.; Eick, J. D.; *Polymer*, **2006**, *47*, 8595-8603.
- [73] Liu, H.; Yao, X.; Zhang, R.; Liu, M.; Hu, Z.; Fan, B.; *Chemosphere*, **2006**, *63*, 722–733.
- [74] Papa, E.; Villa, F.; Gramatica, P.; *J. Chem. Inf. Model.*, **2005**, *45*, 1256-1266.
- [75] Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E.; *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1806-1814.
- [76] Xue, C. X.; Zhang, R. S.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T.; *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 669-677.
- [77] Sun, N.; He, X.; Dong, K.; Zhang, X.; Lu, X.; He, H.; Zhang, S.; *Fluid Phase Equilib.*, **2006**, *246*, 137–142.
- [78] Katritzky, A. R., Petrukhin, R., Jain, R., Karelson, M.; *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1521-1530.
- [79] Lather, V.; Kairys, V.; Fernandes, M.X.; *Chem. Biol. Drug Des.*, **2009**, *73*, 428-441.
- [80] Liu, H.; Yao, X.; Xue, C.; Zhang, R.; Liu, M.; Hu, Z.; Fan, B.; *Anal. Chim. Acta*, **2005**, *542*, 249–259.
- [81] Luan, F.; Zhang, R.; Zhao, C.; Yao, X.; Liu, M.; Hu, Z.; Fan, B.; *Chem. Res. Toxicol.*, **2005**, *18*, 198-203.
- [82] Patan, E.; Pittal, V.; Guerrero, F.; Salerno, L.; Romeo, G.; Siracusa, M. A.; Russo, F.; Manetti, F.; Botta, M.; Mereghetti, I.; Cagnotto, A.; Mennini, T.; *J. Med. Chem.*, **2005**, *48*, 2420-2431.
- [83] Yao, X. J.; Panaye, A.; Doucet, J. P.; Zhang, R. S.; Chen, H. F.; Liu, M. C.; Hu, Z. D.; Fan, B. T.; *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1257-1266.
- [84] Liu, H.; Yao, X.; Liu, M.; Hu, Z.; Fan, B.; *Environ. Pollut.*, **2007**, *147*, 41–49.
- [85] Katritzky, A. R.; Lomaka, A.; Petrukhin, R.; Jain, R.; Karelson, M.; Visser, A. E.; Rogers, R. D.; *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 71-74.
- [86] Tämm, K.; Fara, D. C.; Katritzky, A. R.; Burk, P.; Karelson, M.; *J. Phys. Chem. A*, **2004**, *108*, 4812–4818 .
- [87] Yuan, Y.; Zhang, R.; Luo, L.; *Chemometr. Intell. Lab.*, **2009**, *96*, 144–148.

- [88] Lü, J.; Shen, Q.; Jiang, J.; Shen, G.; Yu, R.; *J. Pharmaceut. Biomed.*, **2004**, *35*, 679-687.
- [89] Stanton, D. T.; Mattioni, B. E.; Knittel, J. J.; Jurs, P. C.; *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1010-1023.
- [90] Luan, F.; Liu, H.T.; Ma, W.P.; Fan, B.T.; *Ecotox. Environ. Safe*, **2008**, *71*, 731–739.
- [91] Liu, H.; Yao, X.; Zhang, R.; Liu, M.; Hu, Z.; Fan, B.; *J. Phys. Chem. B*, **2005**, *109*, 20565-20571.
- [92] Katritzky, A. R.; Pacureanu, L. M.; Slavov, S. H.; Dobchev, D. A.; Shah, D. O.; Karelson, M.; *Comput. Chem. Eng.*, **2009**, *33*, 321–332.
- [93] García, G. C.; Ruiz, I. L.; Gómez-Nieto, M.A.; Doncel, J. A. C.; Plaza, A. G.; *J. Chem. Inf. Model.*, **2005**, *45*, 231-238.
- [94] Xia, B.; Ma, W.; Zhang, X.; Fan, B.; *Anal. Chim. Acta*, **2007**, *598*, 12–18.
- [95] Colombo, A.; Benfenati, E.; Karelson, M.; Maran, U.; *Chemosphere*, **2008**, *72*, 772–780.
- [96] Coi, A.; Massarelli, I.; Murgia, L.; Saraceno, M.; Calderone, V.; Bianucci, A. M.; *Bioorgan. Med. Chem.*, **2006**, *14*, 3153–3159.
- [97] Fatemi, M.H.; Haghdaei, M.; *J. Mol. Struct.*, **2008**, *886*, 43–50.
- [98] Morrill, J. A.; Jensen, R. E.; Madison, P. H.; Chabalowski, C. F.; *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 912-920.
- [99] Luan, F.; Ma, W.; Zhang, X.; Zhang, H.; Liu, M.; Hu, Z.; Fan, B.T.; *Chemosphere*, **2006**, *63*, 1142–1153.
- [100] Yin, S.; Shuai, Z.; Wang, Y.; *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 970-977.
- [101] Katritzky, A. R.; Stoyanova-Slavova, I. B.; Dobchev, D. A.; Karelson, M.; *J. Mol. Graph. Model.*, **2007**, *26*, 529–536.
- [102] Varnek, A.; Fourches, D.; Solov'ev, V. P.; Baulin, V. E.; Turanov, A. N.; Karandashev, V. K.; Fara, D.; Katritzky, A. R.; *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1365-1382.
- [103] Li, X.; Luan, F.; Si, H.; Hu, Z.; Liu, M.; *Toxicol. Lett.*, **2007**, *175*, 136–144.
- [104] Katritzky, A. R.; Slavov, S. H.; Dobchev, D. A.; Karelson, M.; *Bioorgan. Med. Chem.*, **2008**, *16*, 7055–7069.
- [105] Xue, C.; Liu, H.; Yao, X.; Liu, M.; Hu, Z.; Fan, B.; *J. Chromatogr. A*, **2004**, *1048*, 233–243.

- [106] Xue, C. X.; Zhang, R. S.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T.; *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1693-1700.
- [107] Ren, Y.; Liu, H.; Xue, C.; Yao, X.; Liu, M.; Fan, B.; *Anal. Chim. Acta*, **2006**, *572*, 272–282.
- [108] Massarelli, I.; Coi, A.; Pietra, D.; Nofal, F. A.; Biagi, G.; Giorgi, I.; Leonardi, M.; Fiamingo, F.; Bianucci, A. M.; *Eur. J. Med. Chem.*, **2008**, *43*, 114-121
- [109] Katritzky, A. R.; Slavov, S. H.; Dobchev, D. A.; Karelson, M.; *Comput. Chem. Eng.*, **2007**, *31*, 1123–1130.
- [110] Liu, H.; Zhang, R.; Yao, X.; Liu, M.; Hu, Z.; Fan, B.; *Anal. Chim. Acta*, **2004**, *525*, 31–41.
- [111] Sola, D.; Ferri, A.; Banchemo, M.; Manna, L.; Sicardi, S.; *Fluid Phase Equilib.*, **2008**, *263*, 33–42.
- [112] Guha, R.; Stanton, D. T.; Jurs, P. C.; *J. Chem. Inf. Model.*, **2005**, *45*, 1109-1121.
- [113] Ma, W.; Luan, F.; Zhang, H.; Zhang, X.; Liu, M.; Hu, Z.; Fan, B.; *J. Chromatogr. A*, **2006**, *1113*, 140–147.



IDENTIFICAÇÃO DE ALVOS NA TERAPIA  
ANTI-CANCERÍGENA UTILIZANDO  
TÉCNICAS DE MODELAÇÃO MOLECULAR

**ANEXOS**

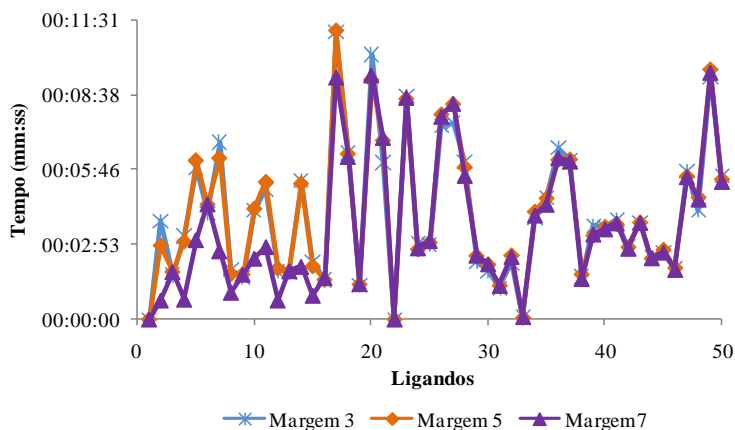
Dissertação para obtenção do Grau de Mestre em  
Bioquímica Aplicada

**Maria Inês Jardim de Sousa**

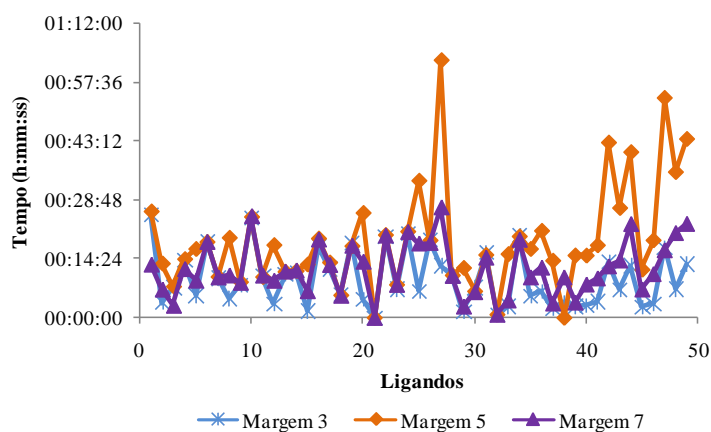
Orientador: Prof. Dr. Miguel Fernandes

Funchal, 2009

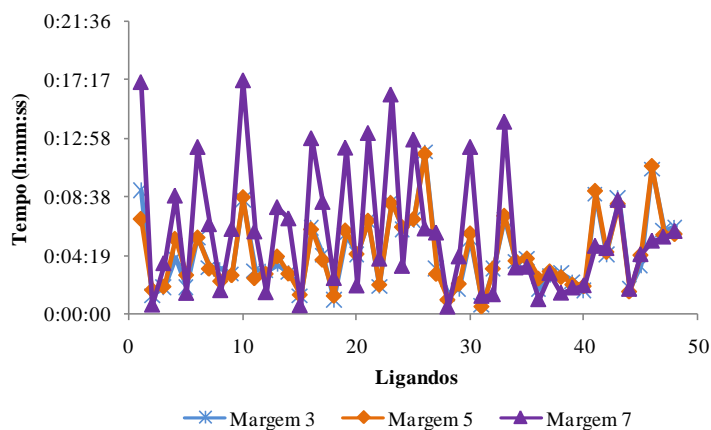
## Anexo A – Otimização das Condições de Docking



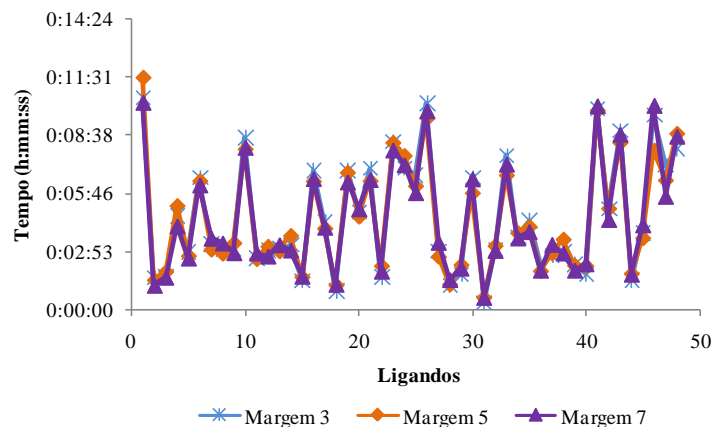
**Figura 1A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima TOP1.



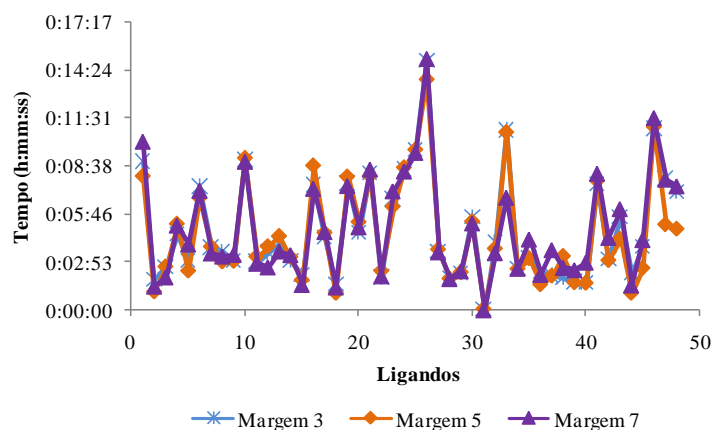
**Figura 2A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima ERK2.



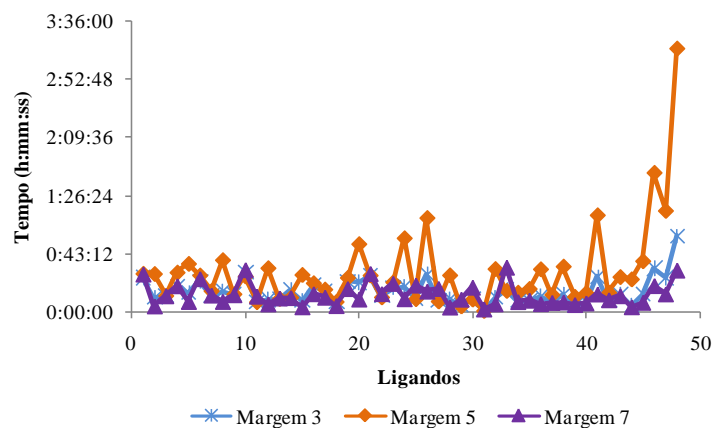
**Figura 3A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima AURKB.



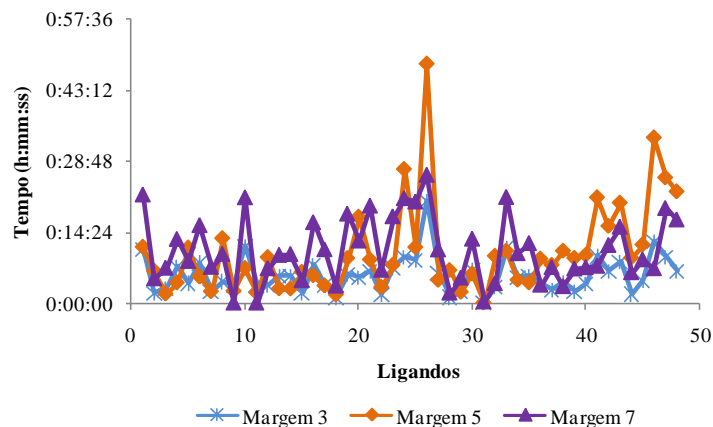
**Figura 4A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima GTP.



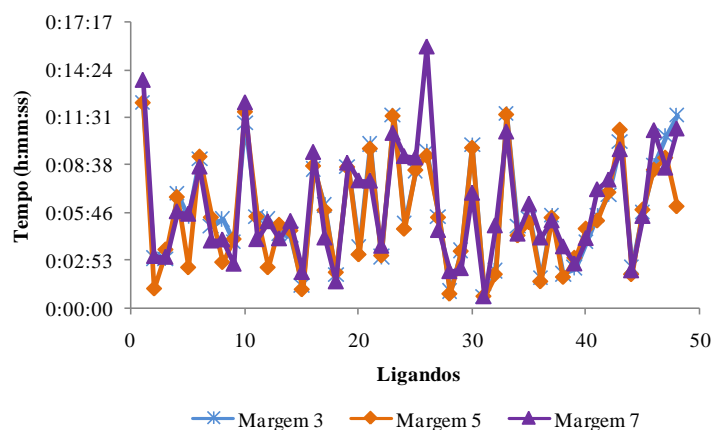
**Figura 5A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima AMD1.



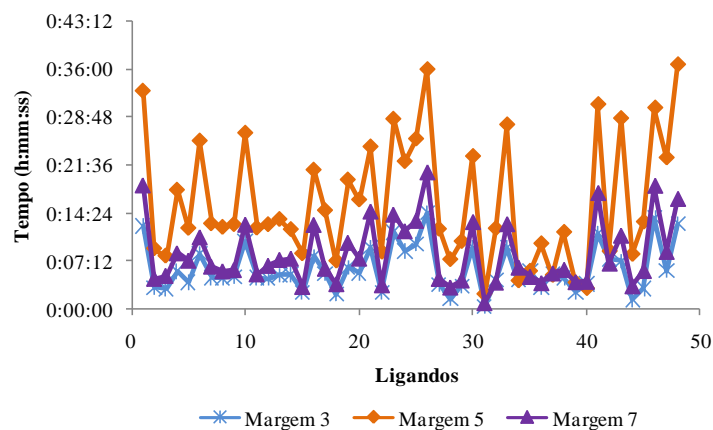
**Figura 6A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima SKP1.



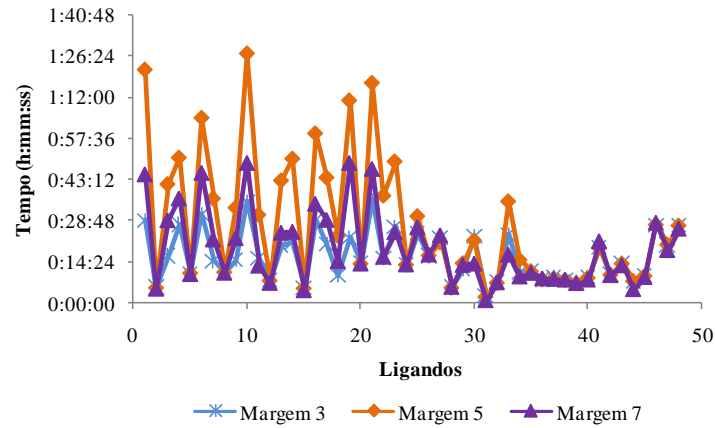
**Figura 7A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima FNTB.



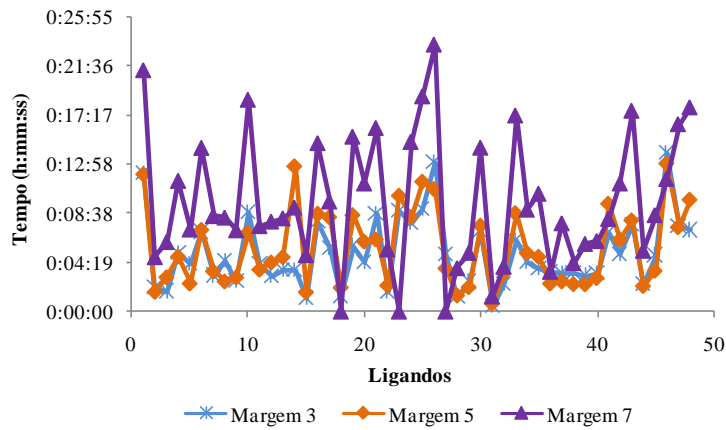
**Figura 8A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima CHK1.



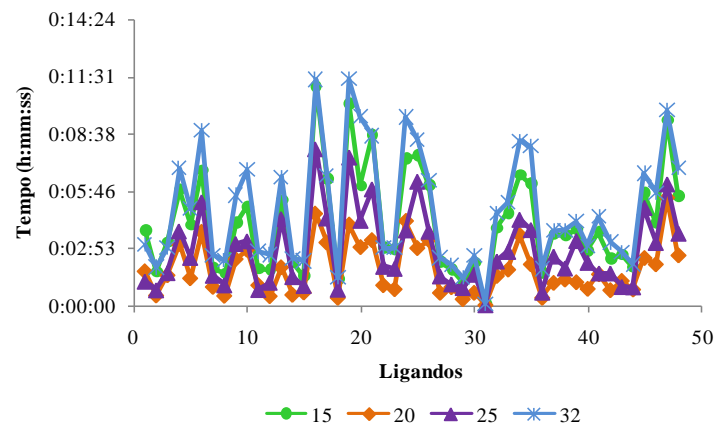
**Figura 9A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima PIN1.



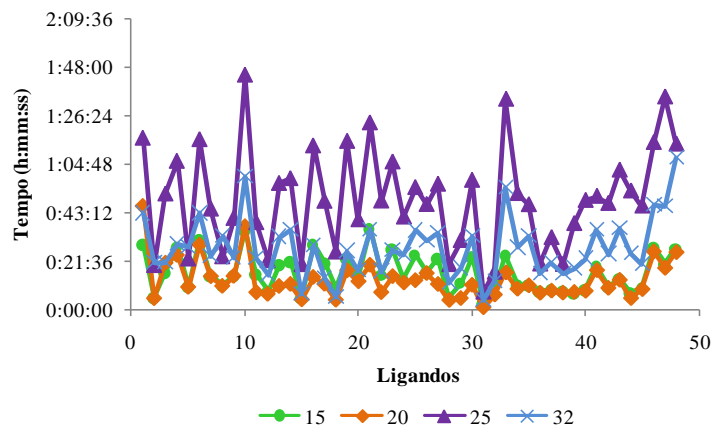
**Figura 10A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima MCM2.



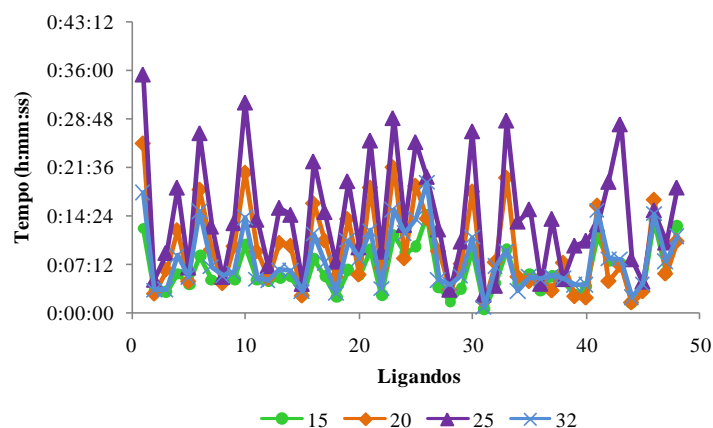
**Figura 11A:** Influência da variação da margem no tempo necessário para a realização do *docking* para a enzima CDC25B.



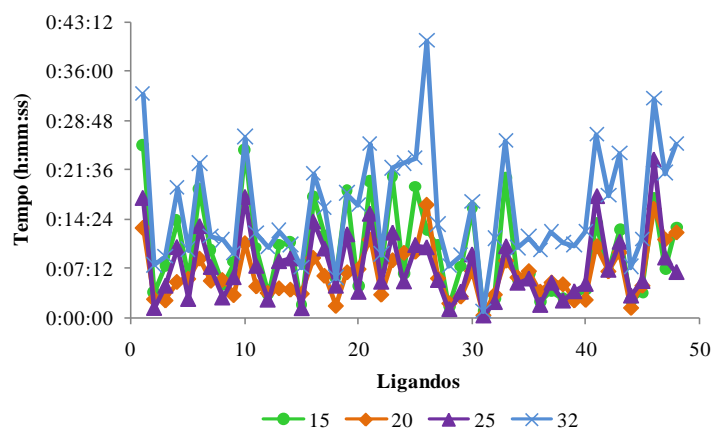
**Figura 12A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima TOPI1.



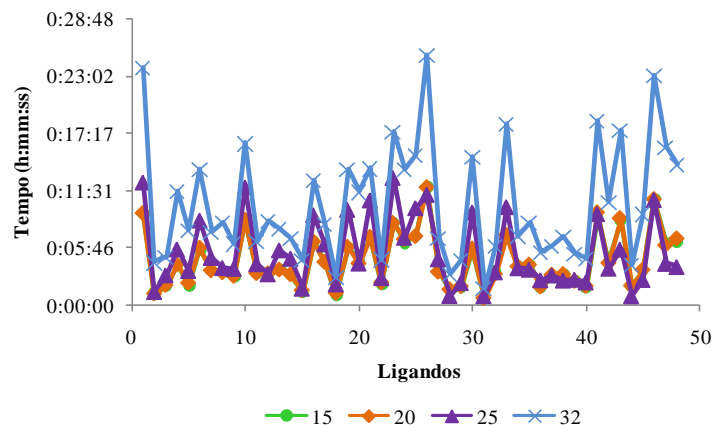
**Figura 13A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima MCM2.



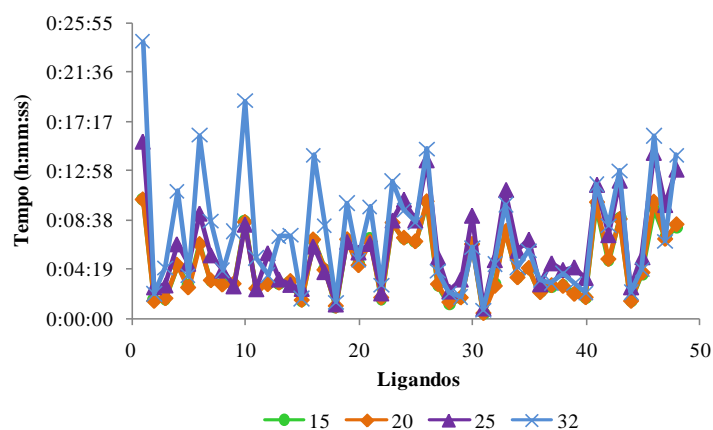
**Figura 14A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima PIN1.



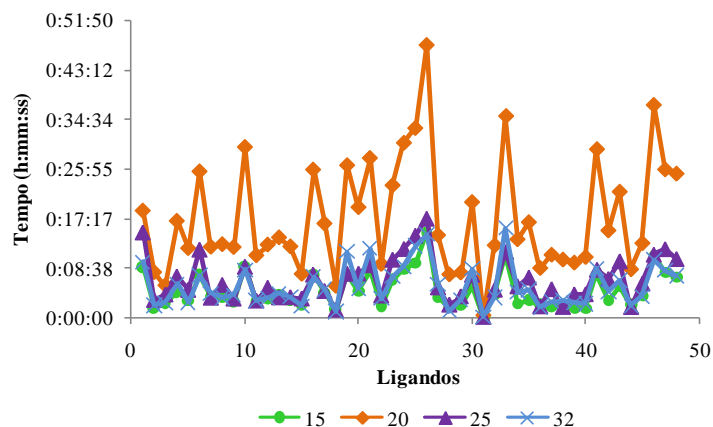
**Figura 15A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima ERK2.



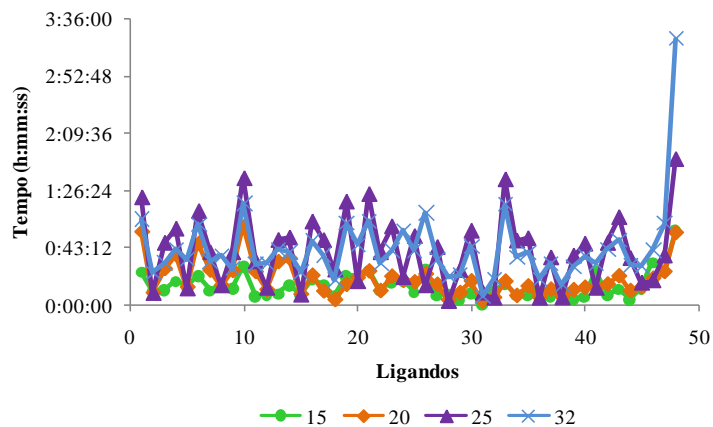
**Figura 16A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima AURKB.



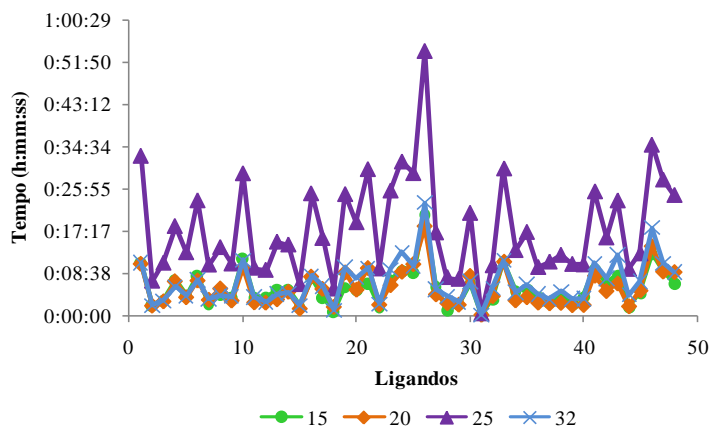
**Figura 17A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima GTP.



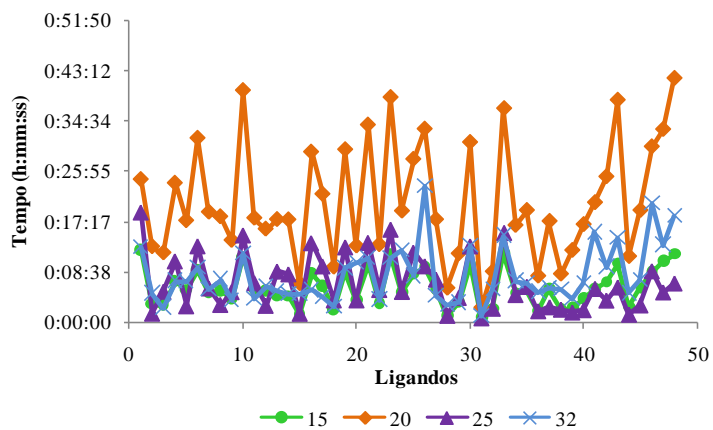
**Figura 18A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima AMD1.



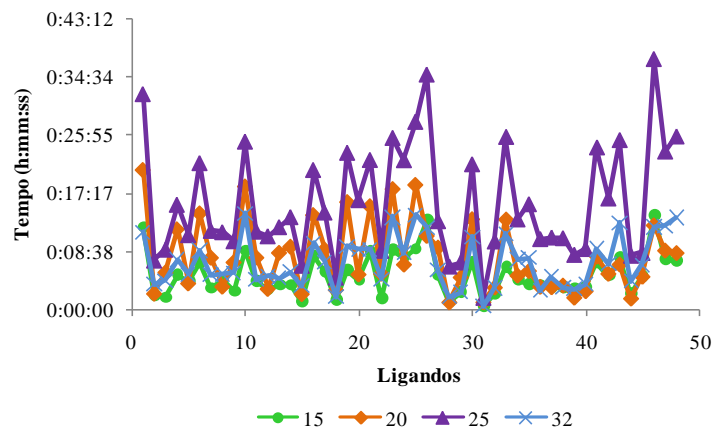
**Figura 19A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima SKP1.



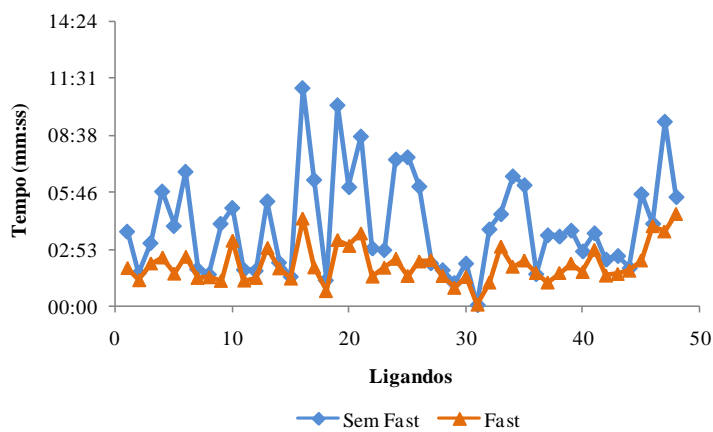
**Figura 20A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima FNTB.



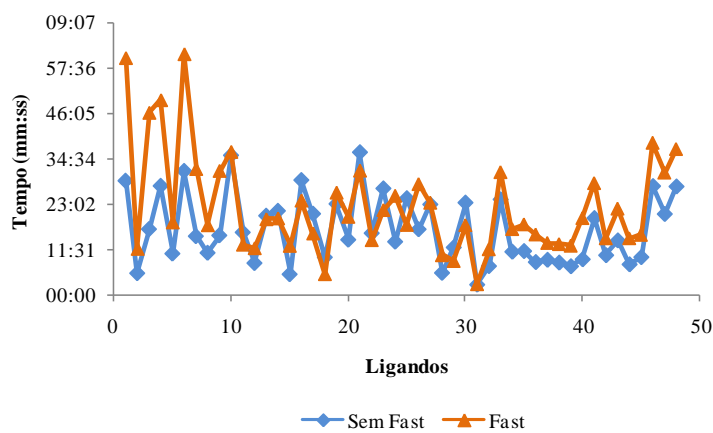
**Figura 21A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima CHK1.



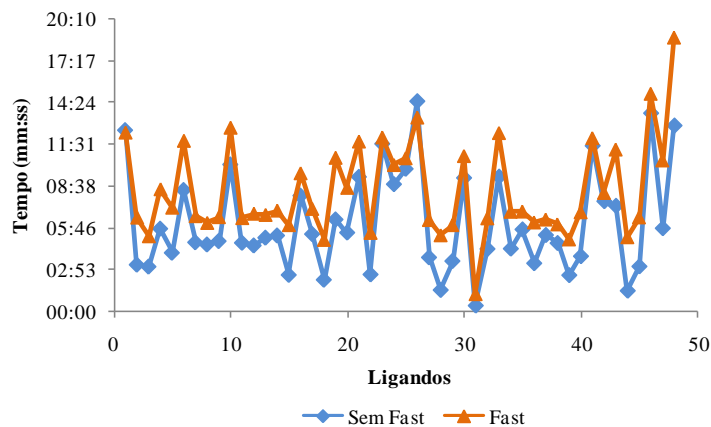
**Figura 22A:** Influência da variação do número de conformações no tempo necessário para a realização do *docking* para a enzima CDC25B.



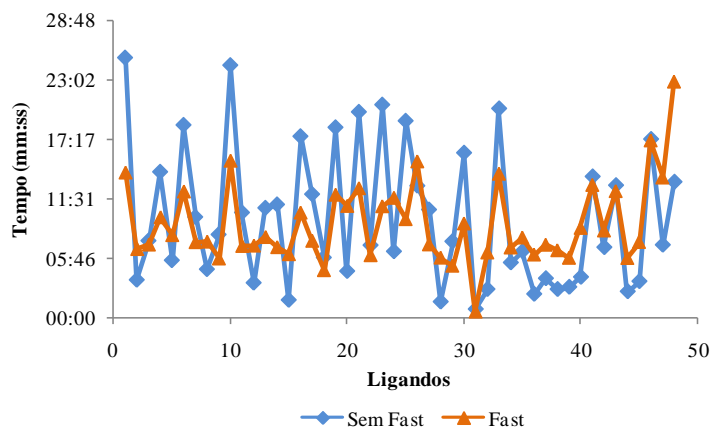
**Figura 23A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima TOP1.



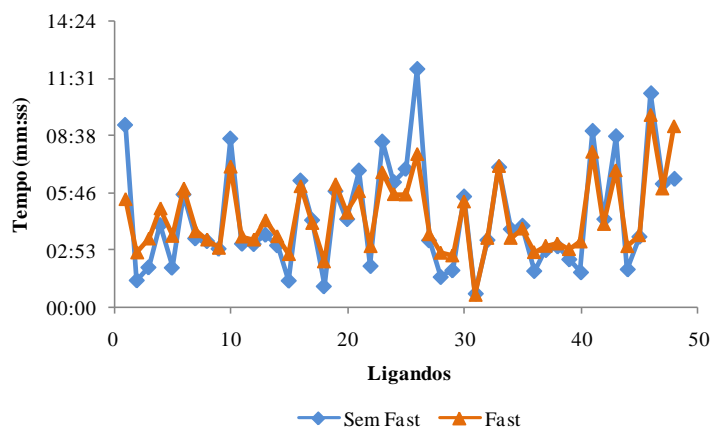
**Figura 24A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima MCM2.



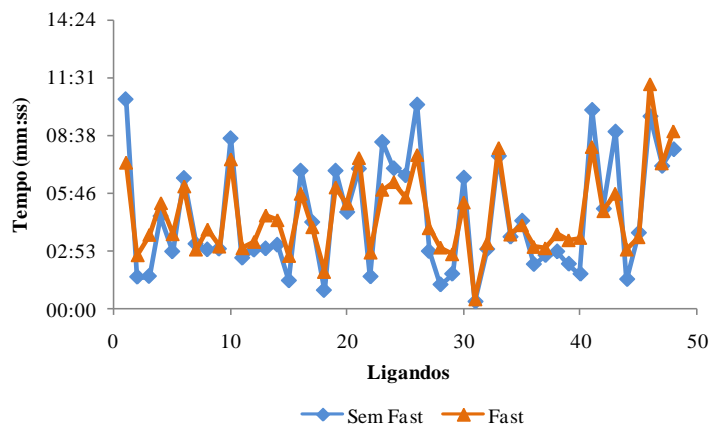
**Figura 25A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima PIN1.



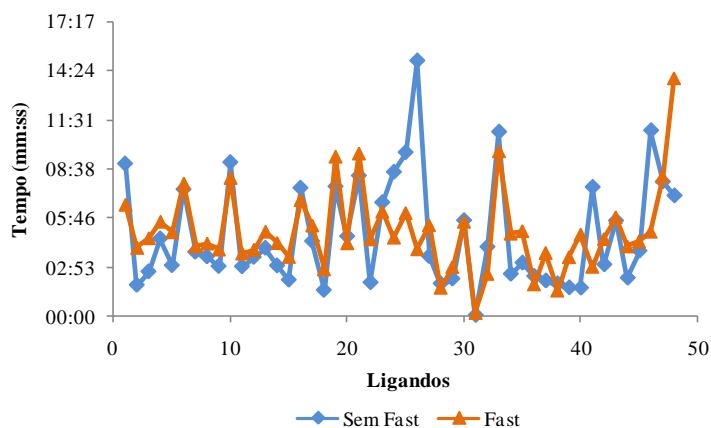
**Figura 26A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima ERK2.



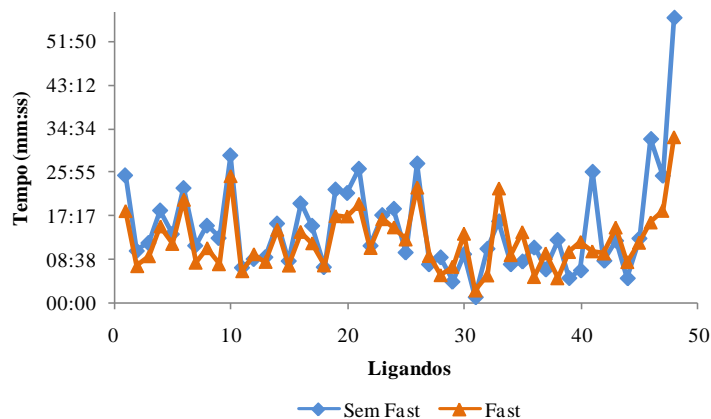
**Figura 27A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima AURKB.



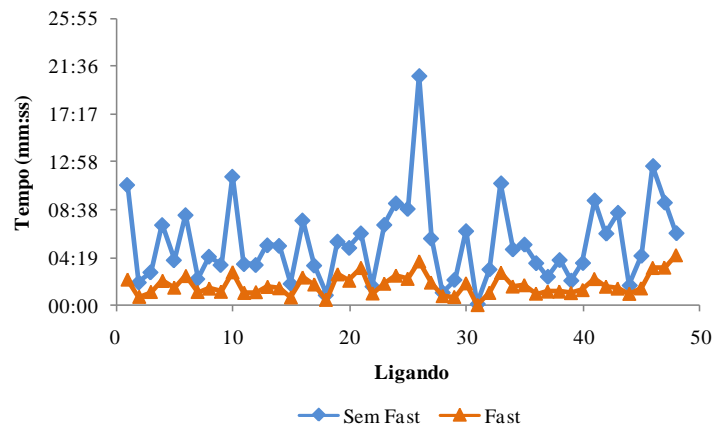
**Figura 28A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima GTP.



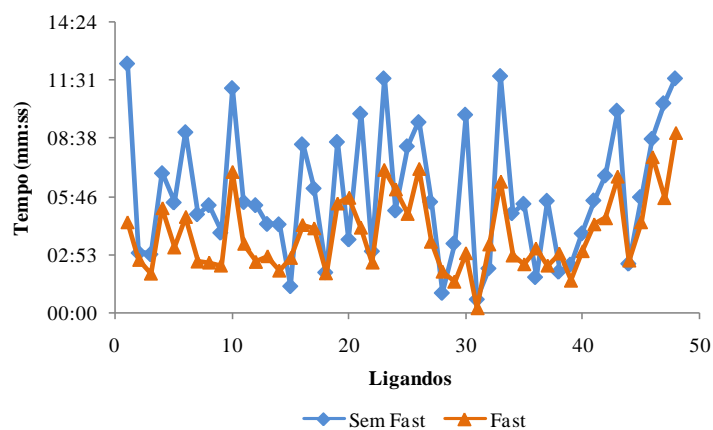
**Figura 29A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima AMD1.



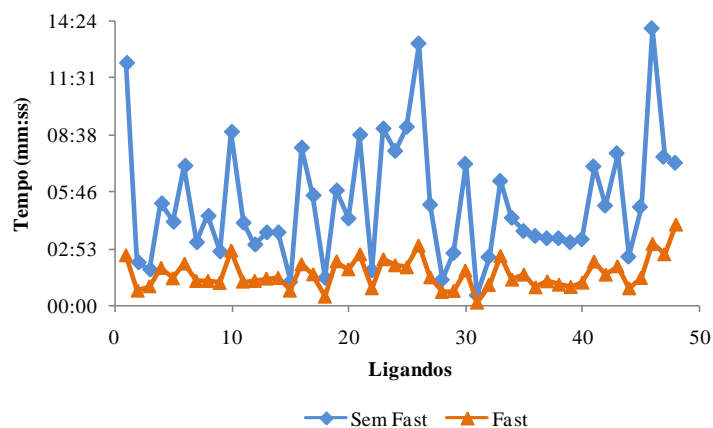
**Figura 30A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima SKP1.



**Figura 31A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima FNTB.



**Figura 32A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima CHK1.



**Figura 33A:** Influência da utilização ou não da opção *fast* no tempo necessário para a realização do *docking* para a enzima CDC25B.

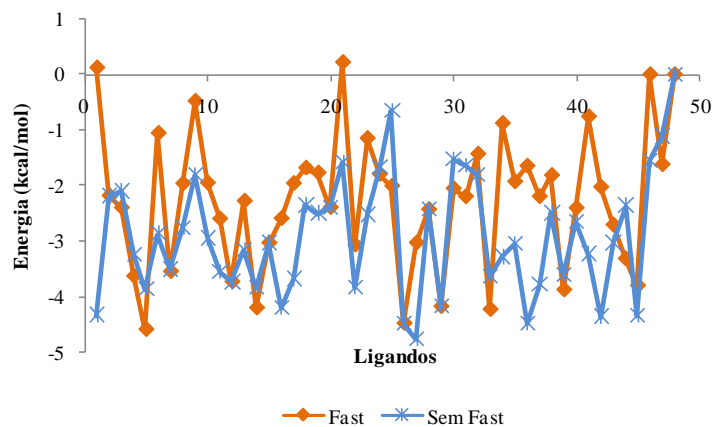


Figura 34A: Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima TOP1.

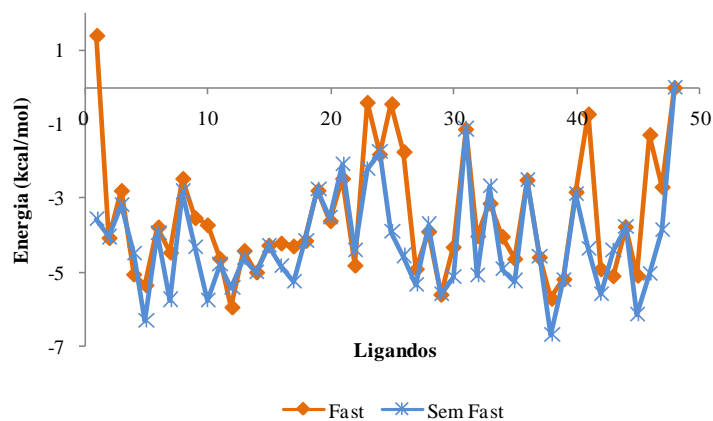


Figura 35A: Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima ERK2.

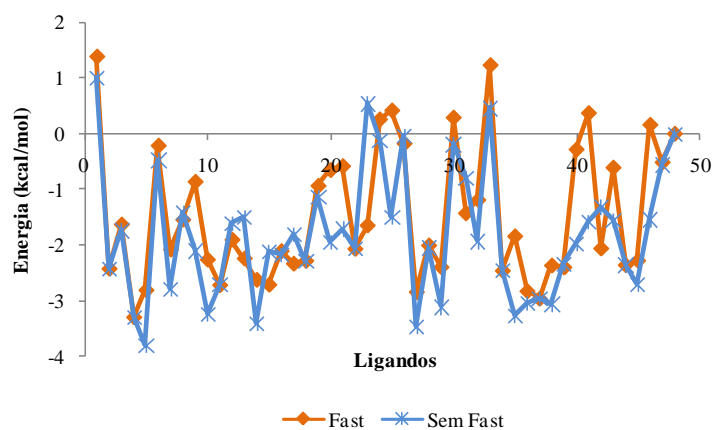
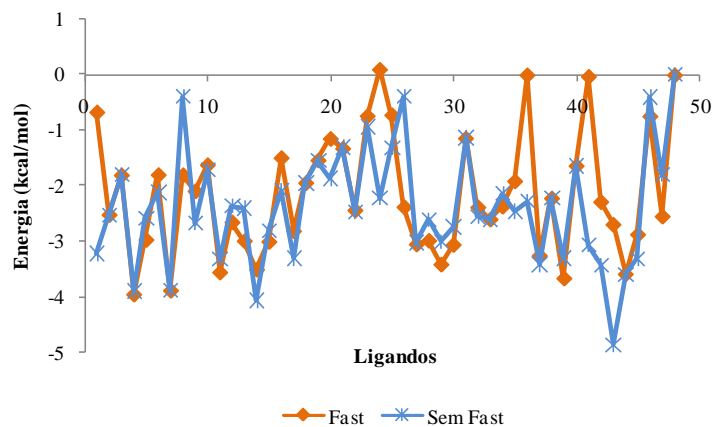
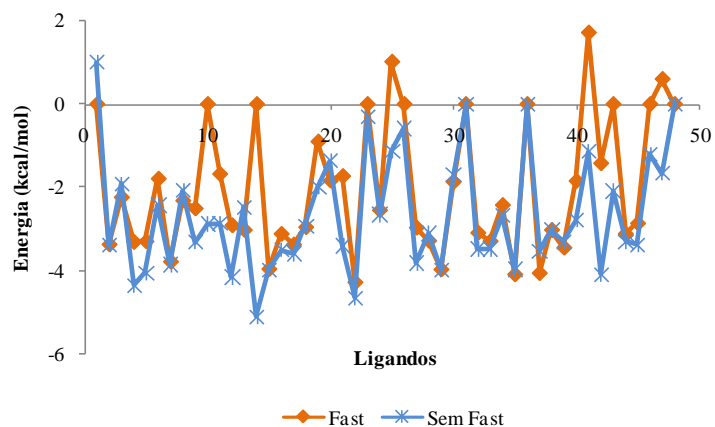


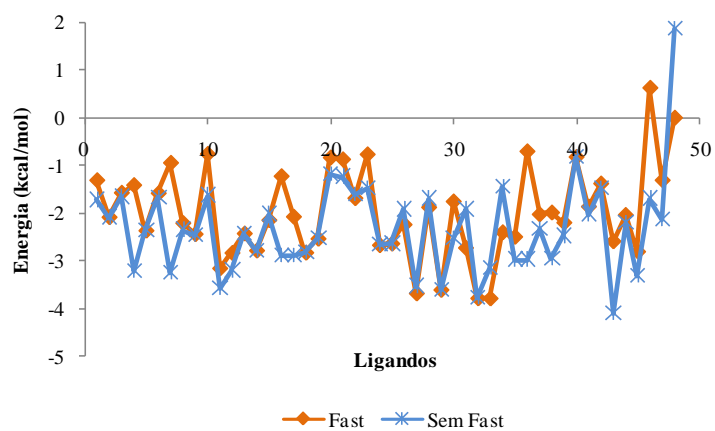
Figura 36A: Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima AURKB.



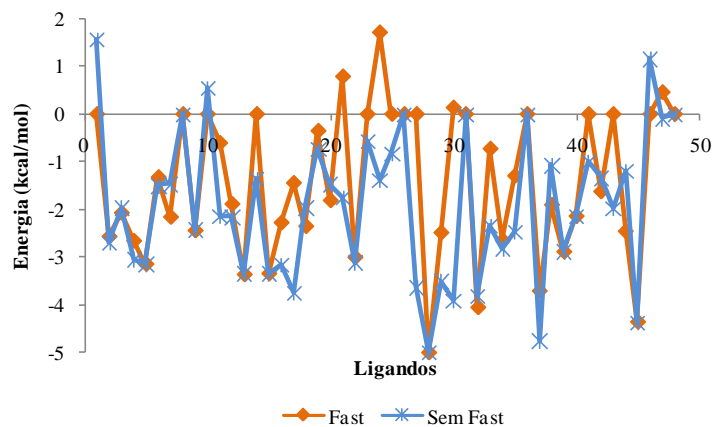
**Figura 37A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima GTP.



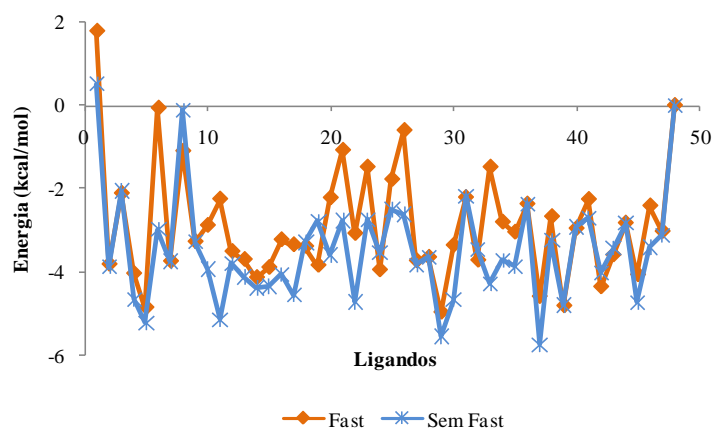
**Figura 38A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima AMD1.



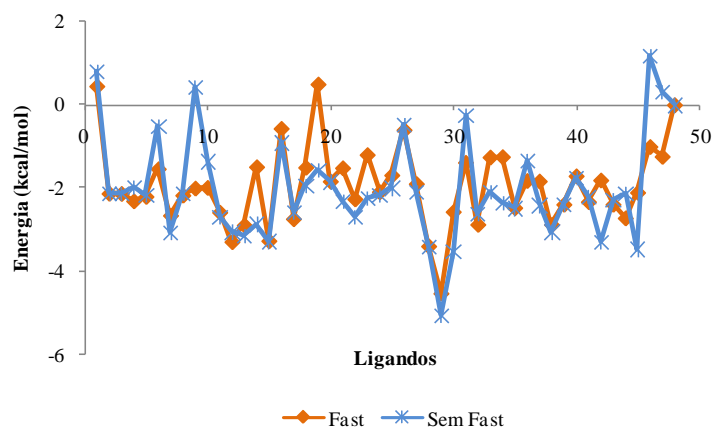
**Figura 39A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima SKP1.



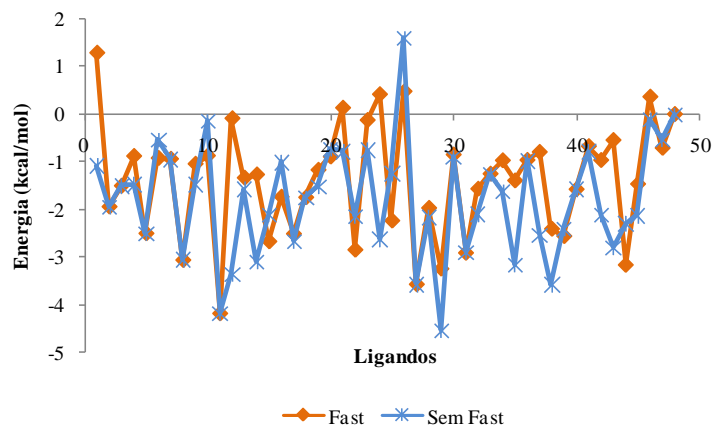
**Figura 40A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima FNTB.



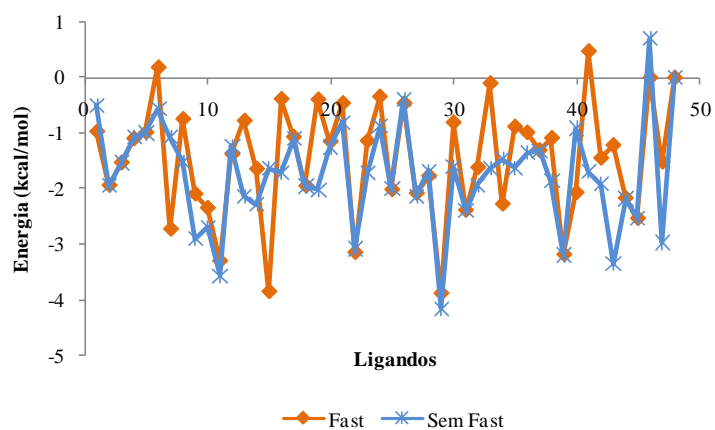
**Figura 41A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima CHK1.



**Figura 42A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima PIN1.

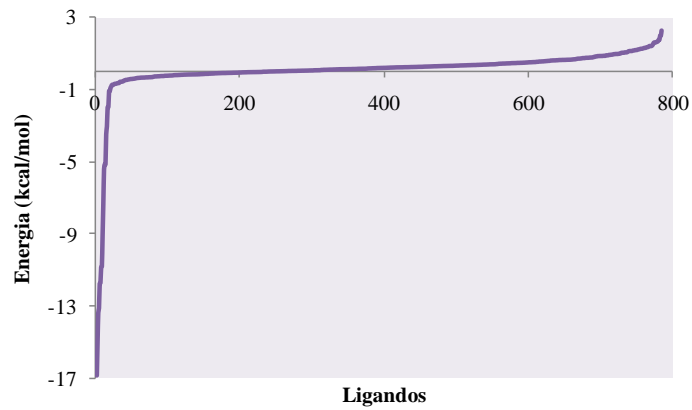


**Figura 43A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima MCM2.

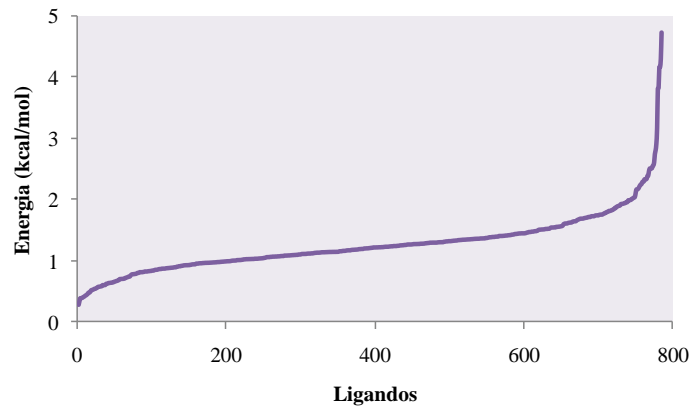


**Figura 44A:** Influência da utilização ou não da opção *fast* nos valores obtidos pelo *docking* para a enzima CDC25B.

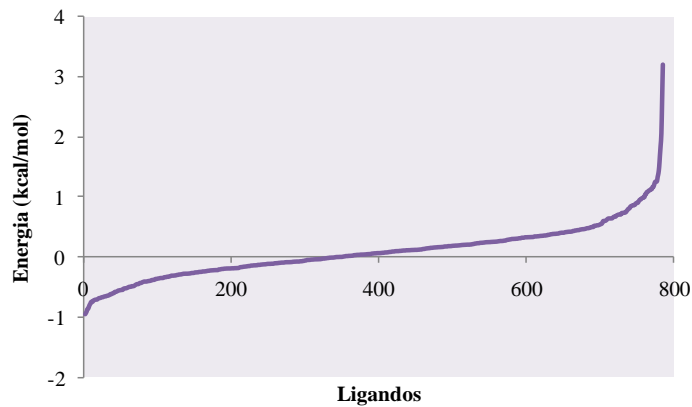
## Anexo B – Distribuição de Energias



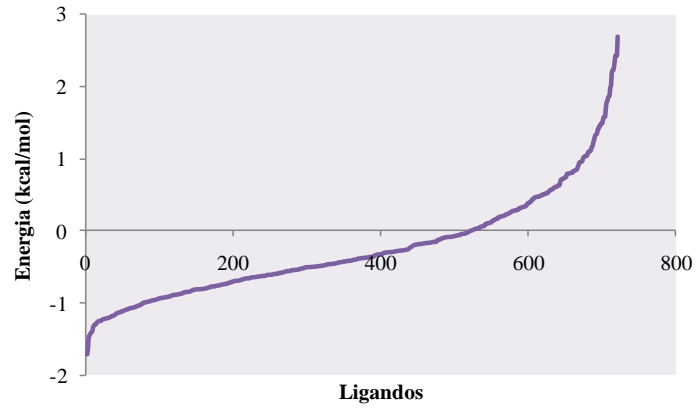
**Figura 1B:** Distribuição das energias obtida após a correção para a enzima CDC25C.



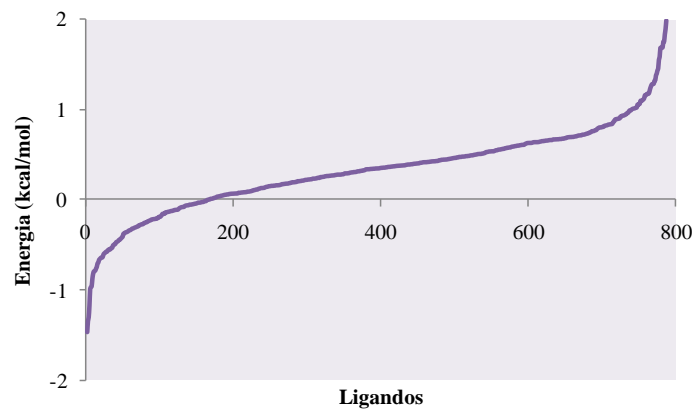
**Figura 2B:** Distribuição das energias obtida após a correção para a enzima ERK1.



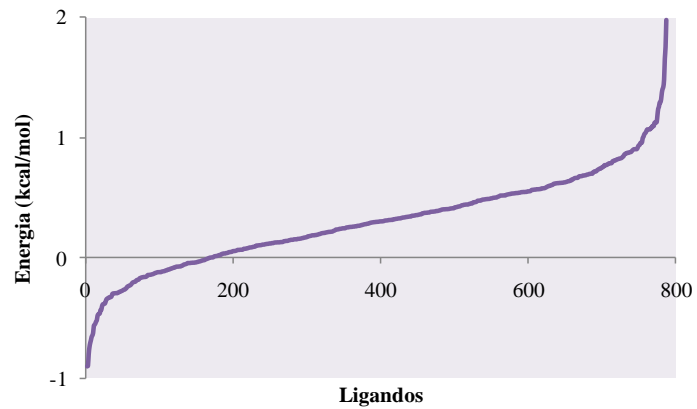
**Figura 3B:** Distribuição das energias obtida após a correção para a enzima AKT1.



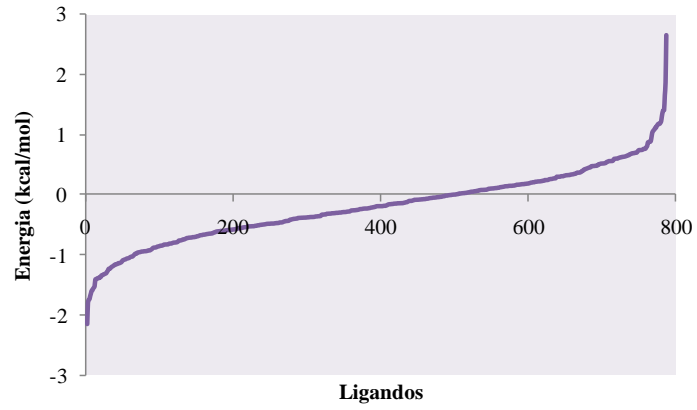
**Figura 4B:** Distribuição das energias obtida após a correção para a enzima AMD1.



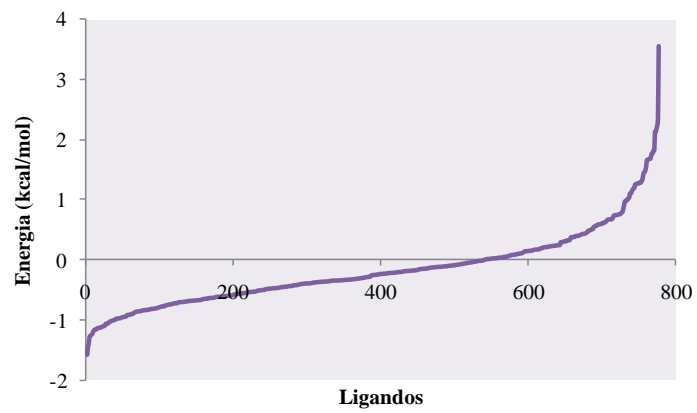
**Figura 5B:** Distribuição das energias obtida após a correção para a enzima CDC25B.



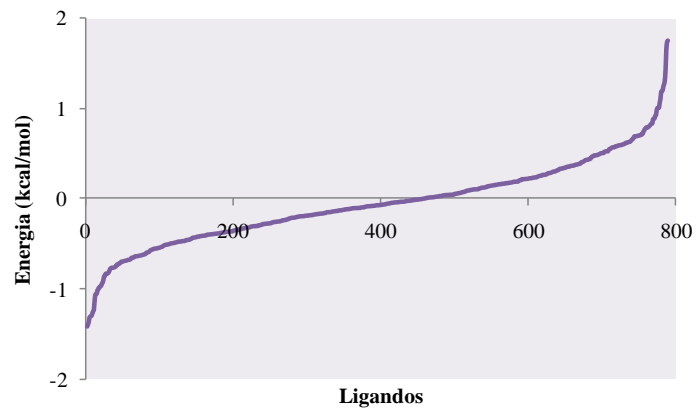
**Figura 6B:** Distribuição das energias obtida após a correção para a enzima CDC7.



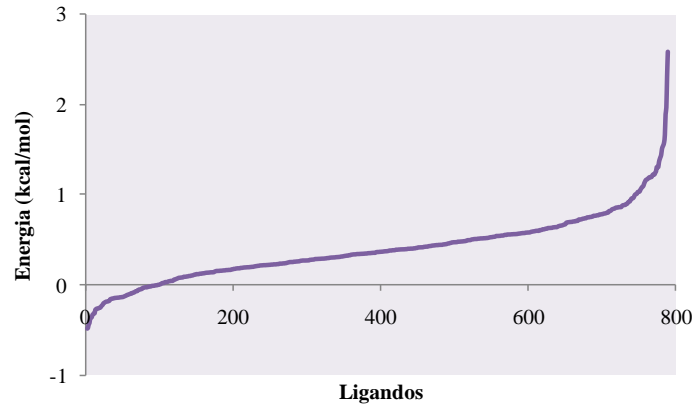
**Figura 7B:** Distribuição das energias obtida após a correção para a enzima CDK4.



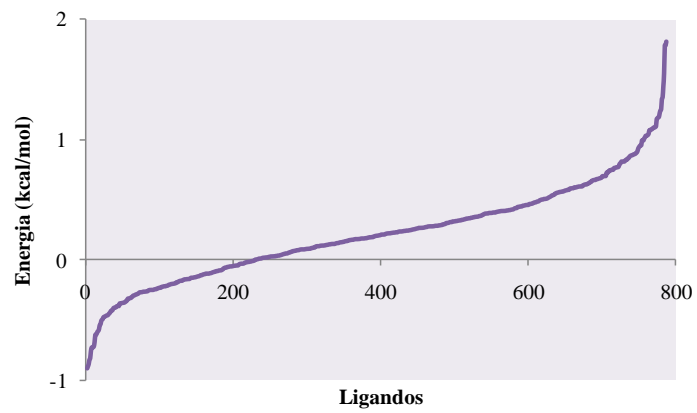
**Figura 8B:** Distribuição das energias obtida após a correção para a enzima CDK7.



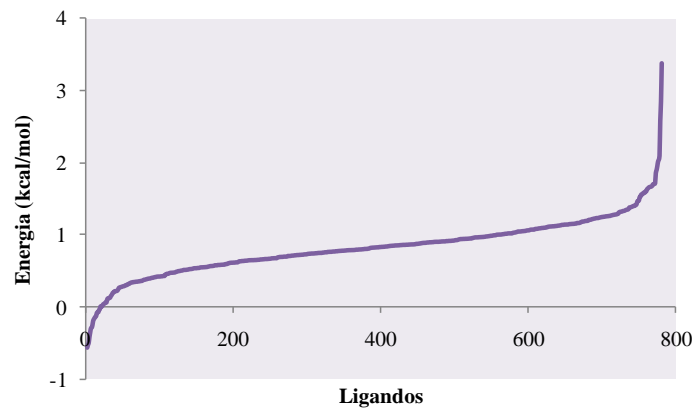
**Figura 9B:** Distribuição das energias obtida após a correção para a enzima CHK2.



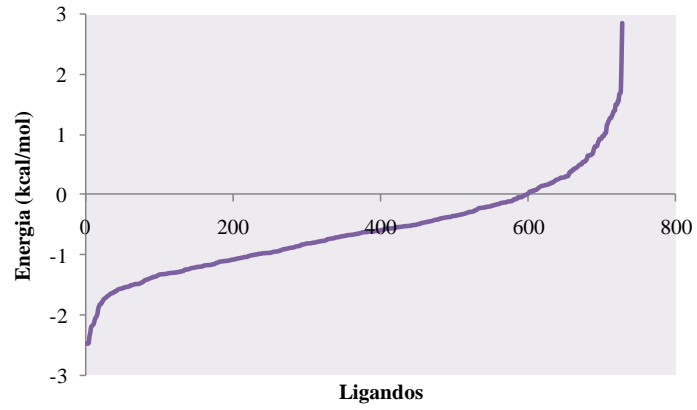
**Figura 10B:** Distribuição das energias obtida após a correção para a enzima CCNH.



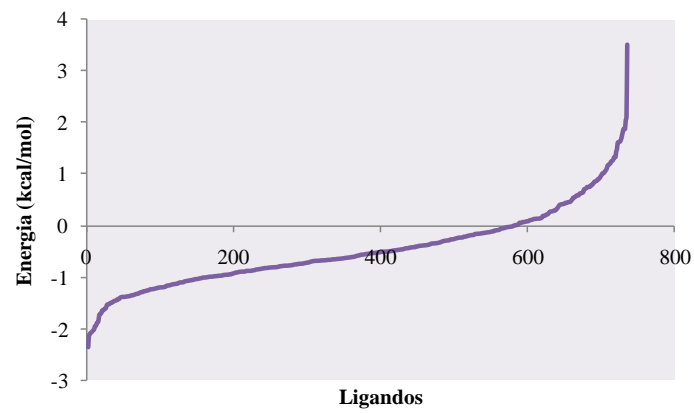
**Figura 11B:** Distribuição das energias obtida após a correção para a proteína DP.



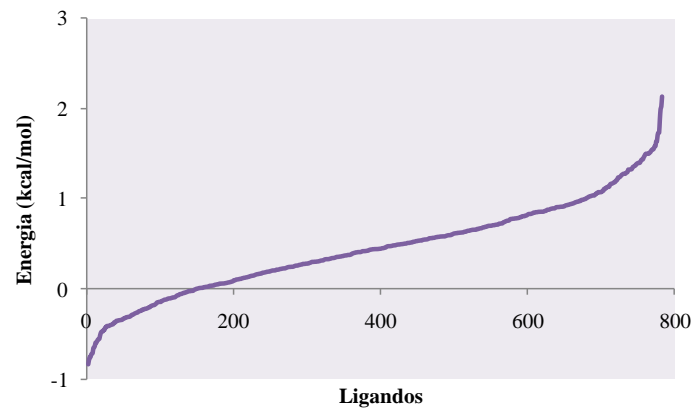
**Figura 12B:** Distribuição das energias obtida após a correção para a enzima E2F.



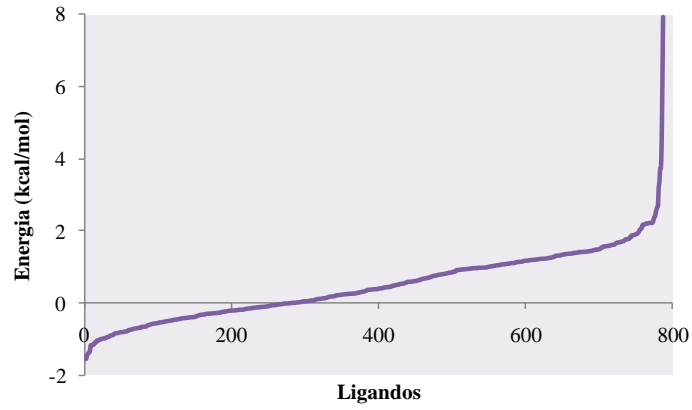
**Figura 13B:** Distribuição das energias obtida após a correção para a enzima GSK-3B.



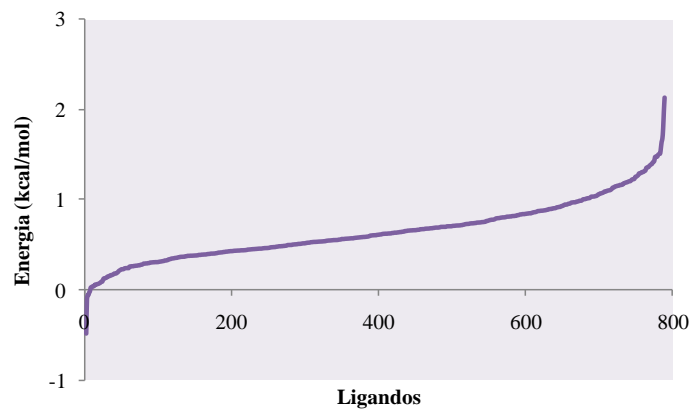
**Figura 14B:** Distribuição das energias obtida após a correção para a enzima MAP2K1.



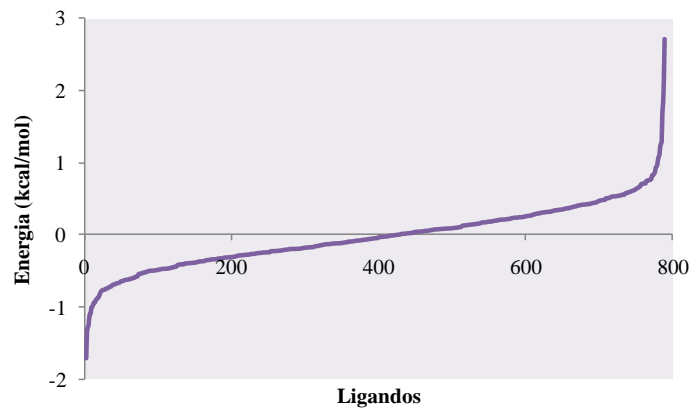
**Figura 15B:** Distribuição das energias obtida após a correção para a proteína p21 CIP1.



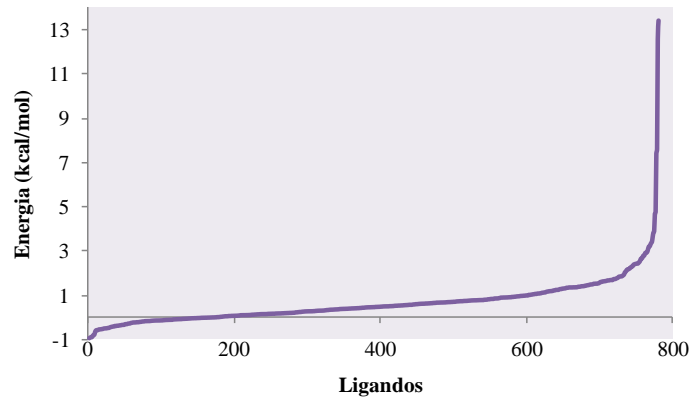
**Figura 16B:** Distribuição das energias obtida após a correção para a enzima p27 KIP1.



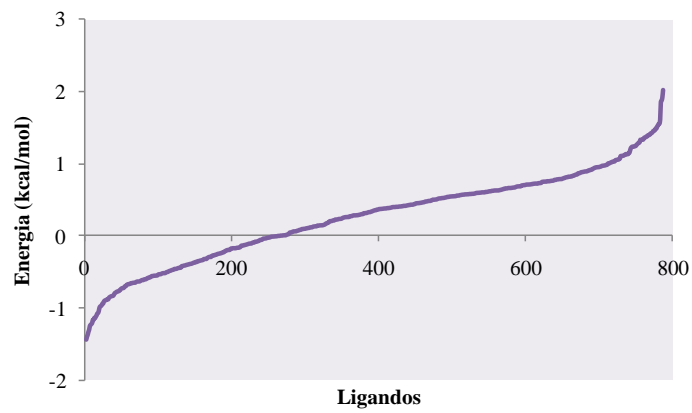
**Figura 17B:** Distribuição das energias obtida após a correção para a proteína p53.



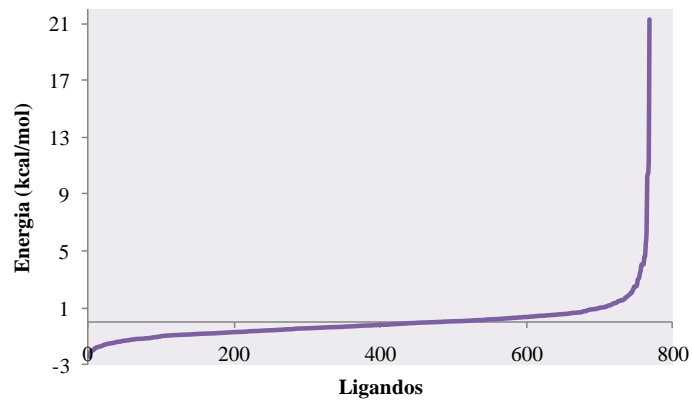
**Figura 18B:** Distribuição das energias obtida após a correção para a enzima SKP1.



**Figura 19B:** Distribuição das energias obtida após a correção para a enzima CDC2.



**Figura 20B:** Distribuição das energias obtida após a correção para a enzima MCM2.



**Figura 21B:** Distribuição das energias obtida após a correção para a enzima CDK2.

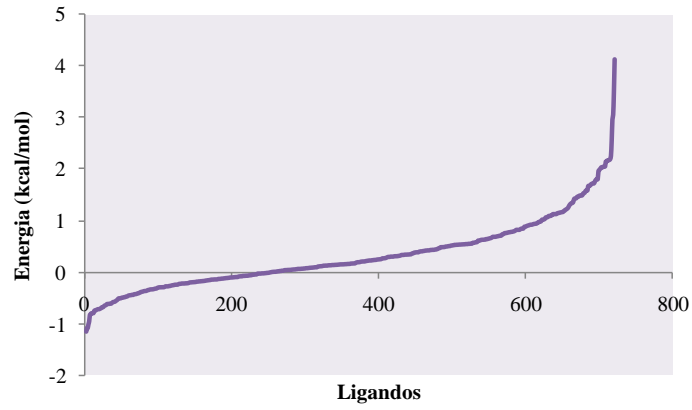


Figura 22B: Distribuição das energias obtida após a correção para a enzima TOP2A.

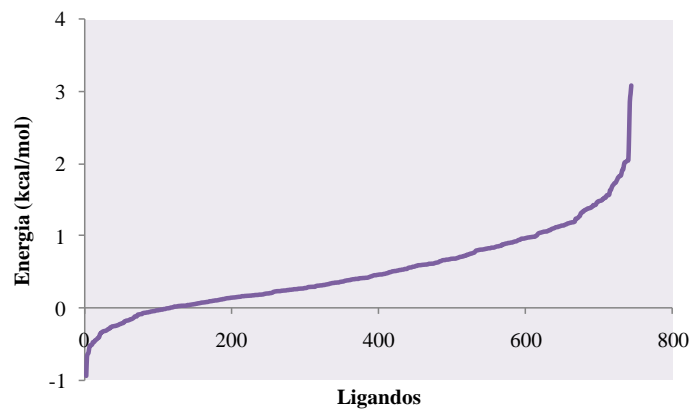


Figura 23B: Distribuição das energias obtida após a correção para a proteína KITLG.

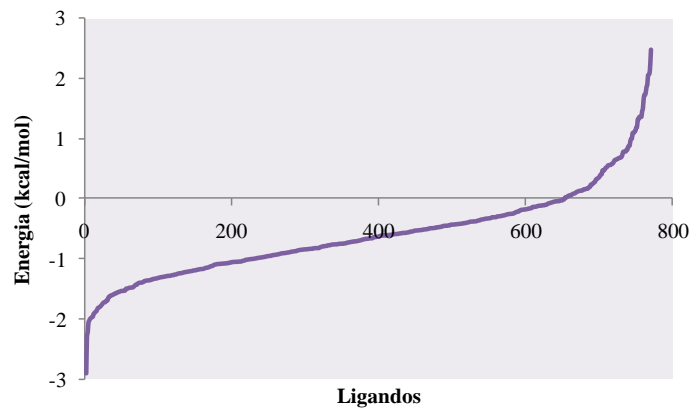
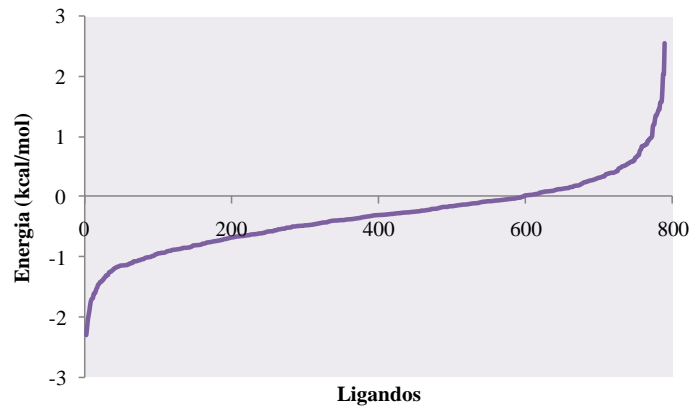
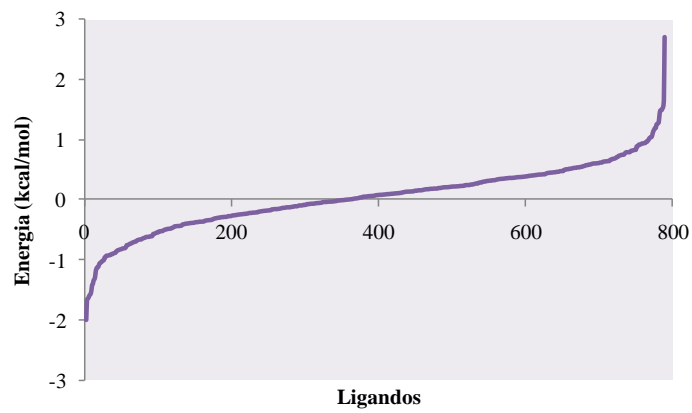


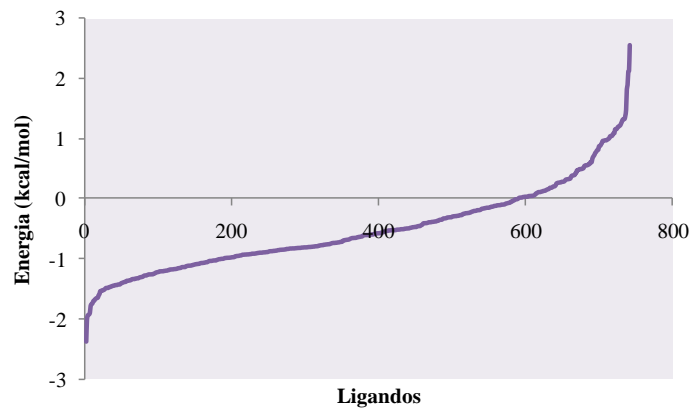
Figura 24B: Distribuição das energias obtida após a correção para a enzima CHK1.



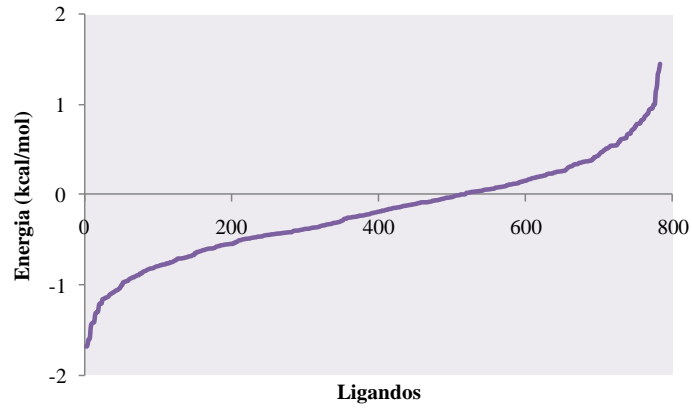
**Figura 25B:** Distribuição das energias obtida após a correção para a enzima TOP1.



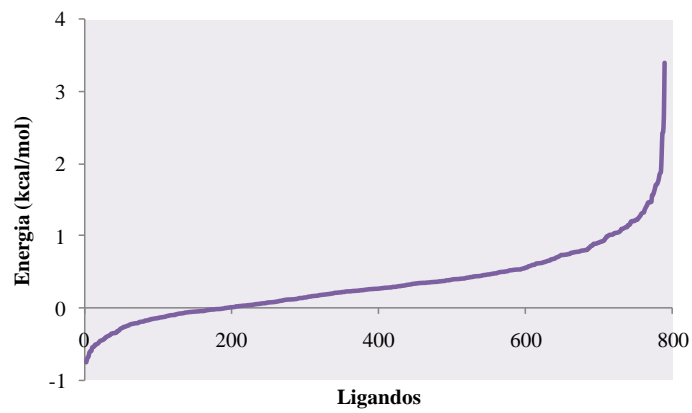
**Figura 26B:** Distribuição das energias obtida após a correção para a enzima CDC6.



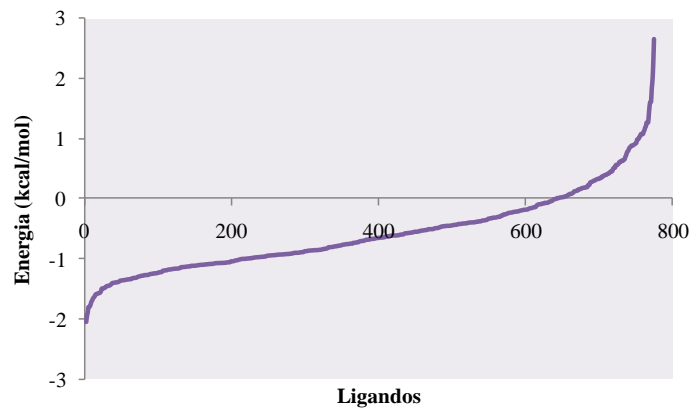
**Figura 27B:** Distribuição das energias obtida após a correção para a enzima WEE1.



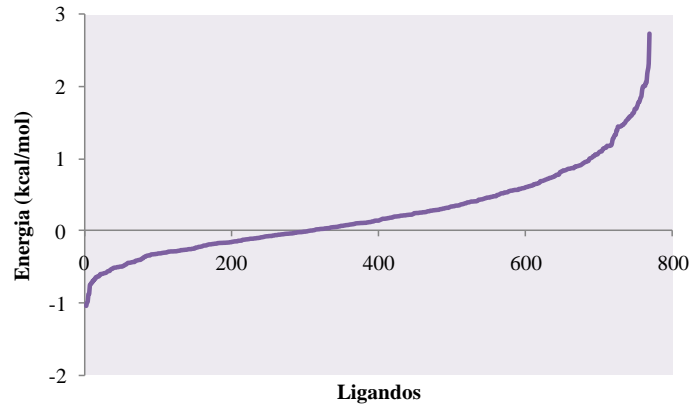
**Figura 28B:** Distribuição das energias obtida após a correção para a enzima PLK1.



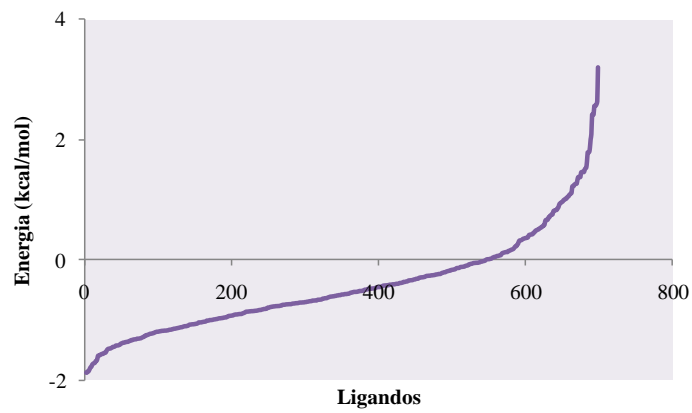
**Figura 29B:** Distribuição das energias obtida após a correção para a enzima Asp.



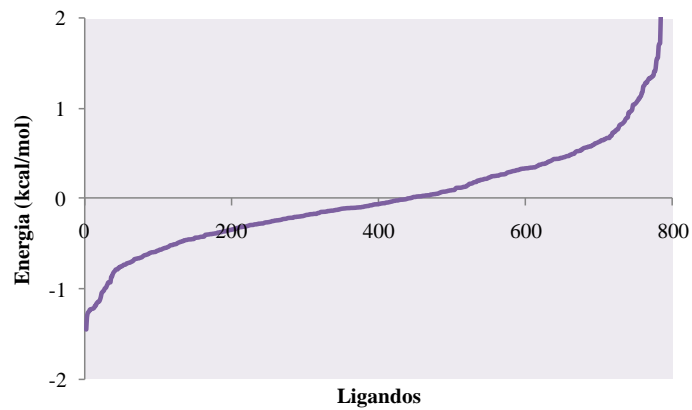
**Figura 30B:** Distribuição das energias obtida após a correção para a enzima NEK2.



**Figura 31B:** Distribuição das energias obtida após a correção para a enzima Eg5.



**Figura 32B:** Distribuição das energias obtida após a correção para a enzima AURKA.



**Figura 33B:** Distribuição das energias obtida após a correção para a enzima PKA.

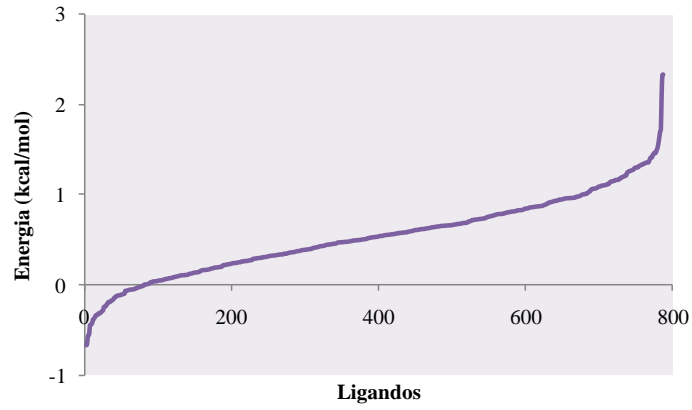


Figura 34B: Distribuição das energias obtida após a correção para a proteína Importina B.

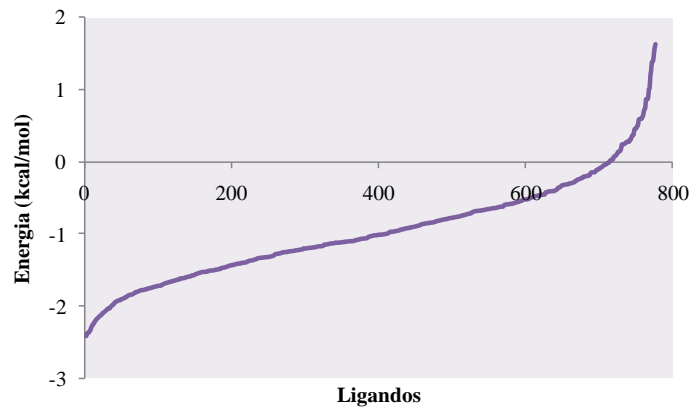


Figura 35B: Distribuição das energias obtida após a correção para a enzima ERK2.

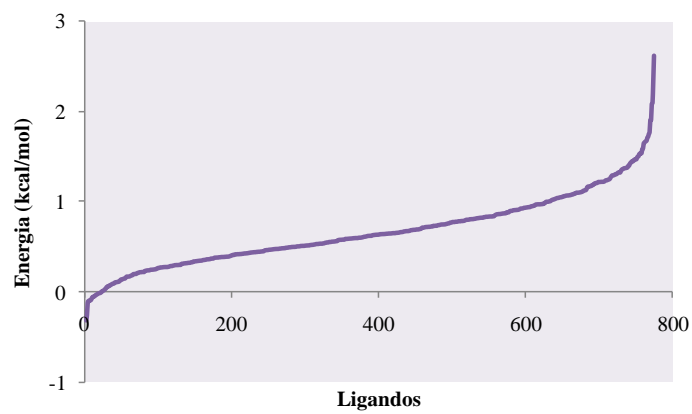
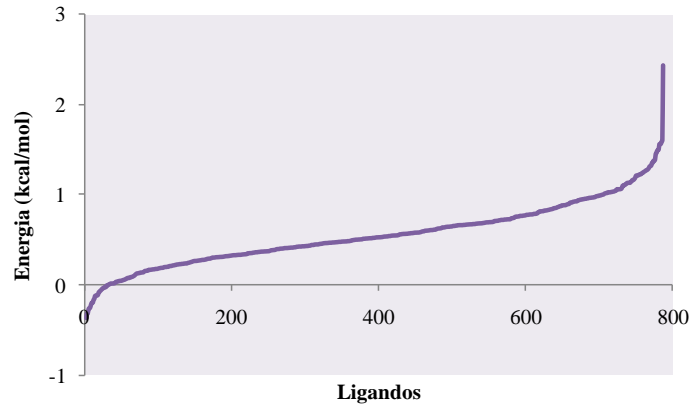
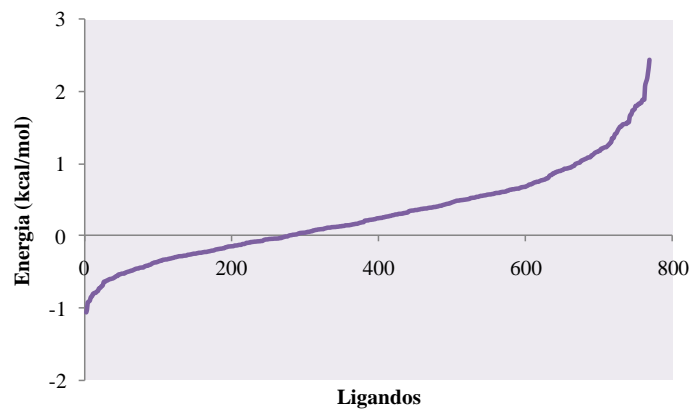


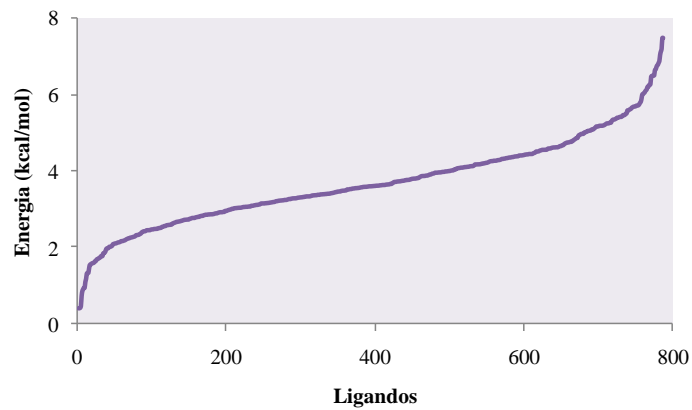
Figura 36B: Distribuição das energias obtida após a correção para a proteína MAD2.



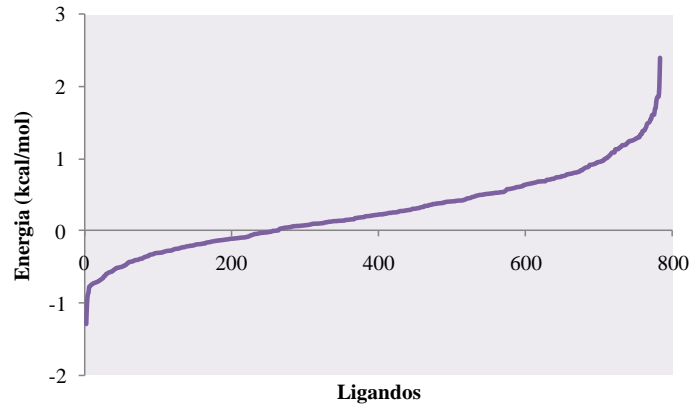
**Figura 37B:** Distribuição das energias obtida após a correção para a proteína CDH1.



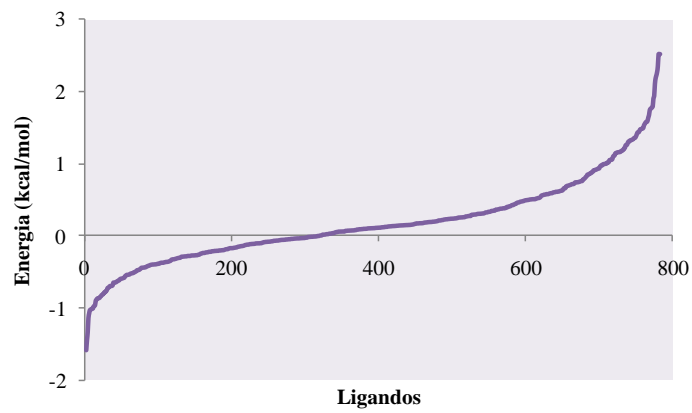
**Figura 38B:** Distribuição das energias obtida após a correção para a enzima AURKB.



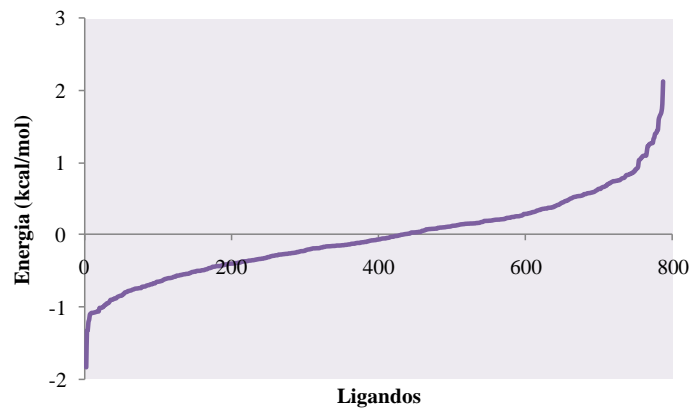
**Figura 39B:** Distribuição das energias obtida após a correção para a proteína CENP-A.



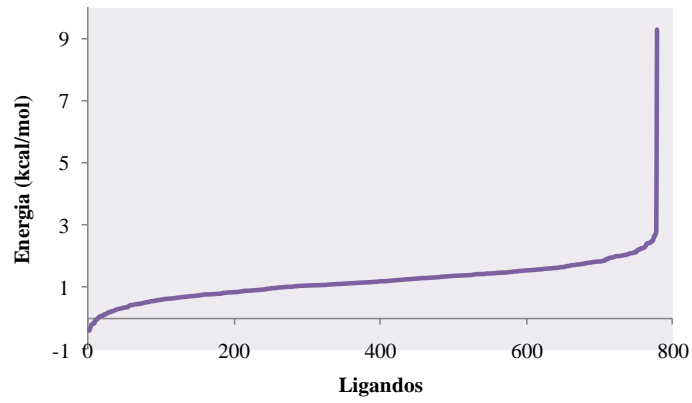
**Figura 40B:** Distribuição das energias obtida após a correção para a proteína CAP-G.



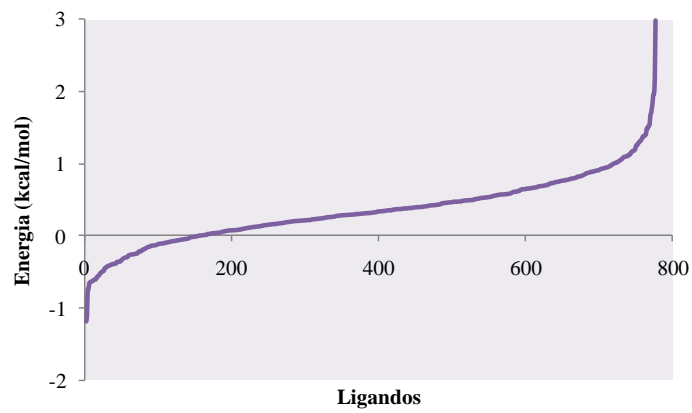
**Figura 41B:** Distribuição das energias obtida após a correção para a enzima PIN1.



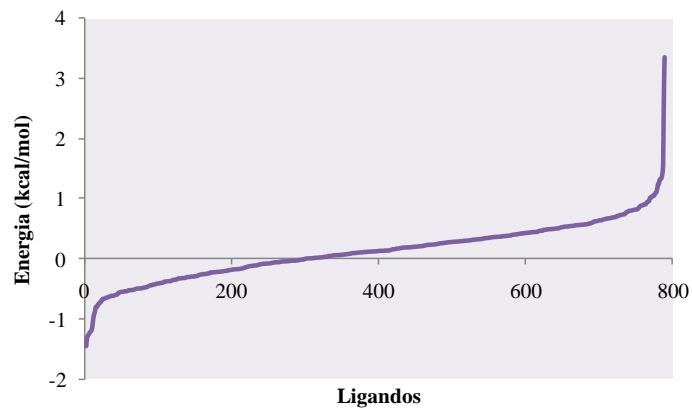
**Figura 42B:** Distribuição das energias obtida após a correção para a proteína Tubulina.



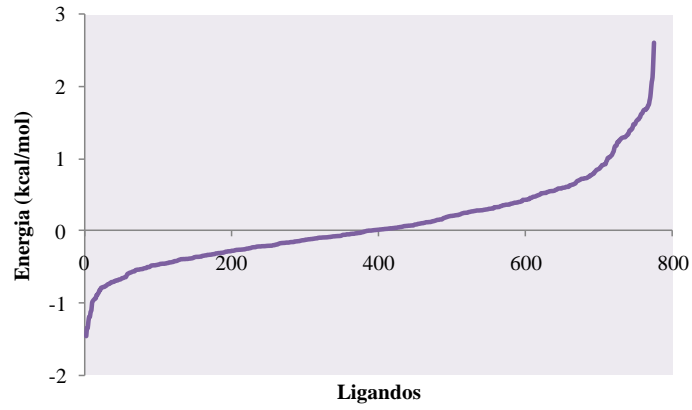
**Figura 43B:** Distribuição das energias obtida após a correção para a enzima Actina.



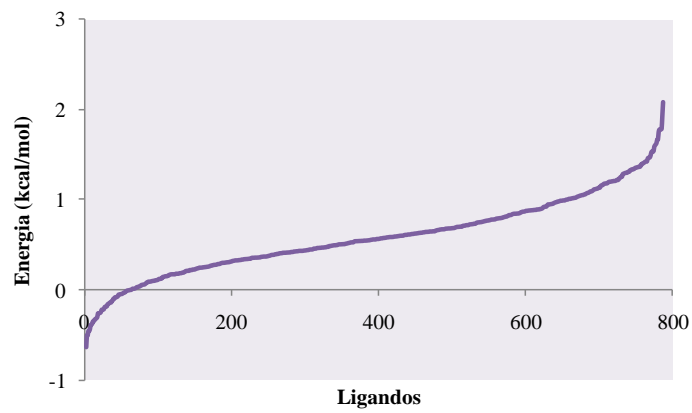
**Figura 44B:** Distribuição das energias obtida após a correção para a proteína MYC.



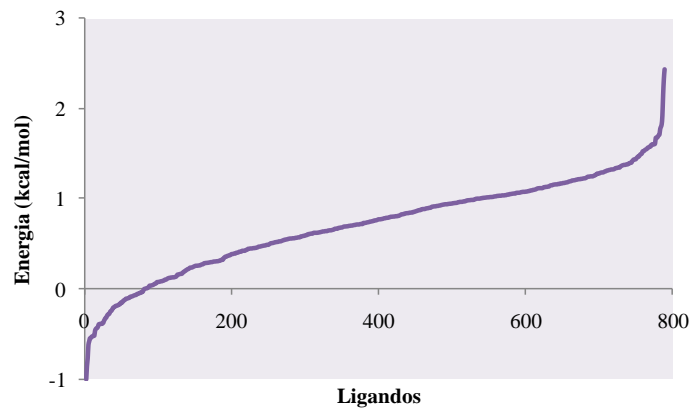
**Figura 45B:** Distribuição das energias obtida após a correção para a proteína MAX.



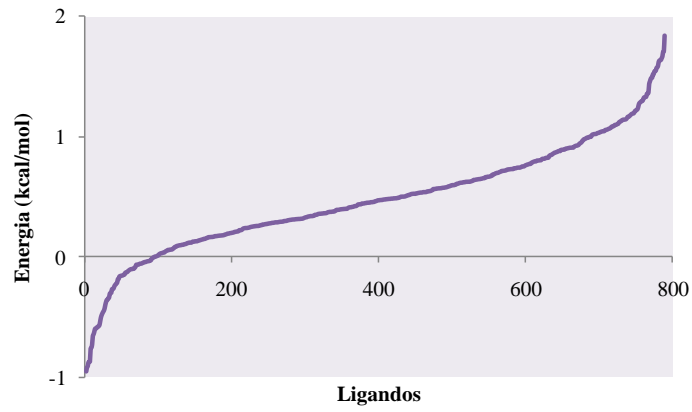
**Figura 46B:** Distribuição das energias obtida após a correção para a enzima PIK3CG.



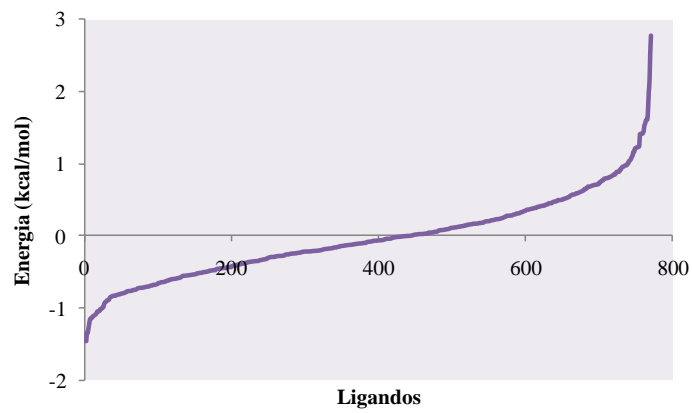
**Figura 47B:** Distribuição das energias obtida após a correção para a enzima PAK1.



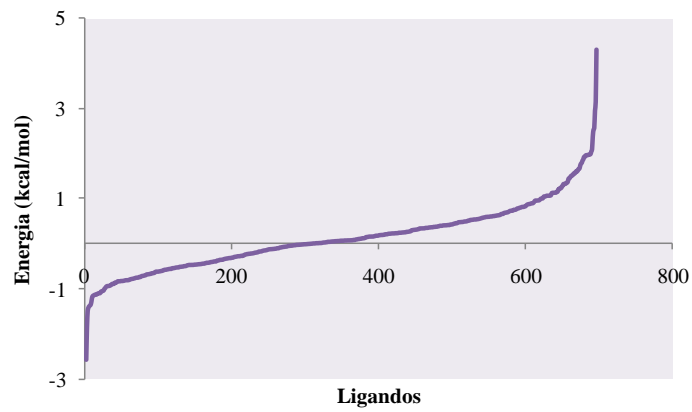
**Figura 48B:** Distribuição das energias obtida após a correção para a enzima ROCK.



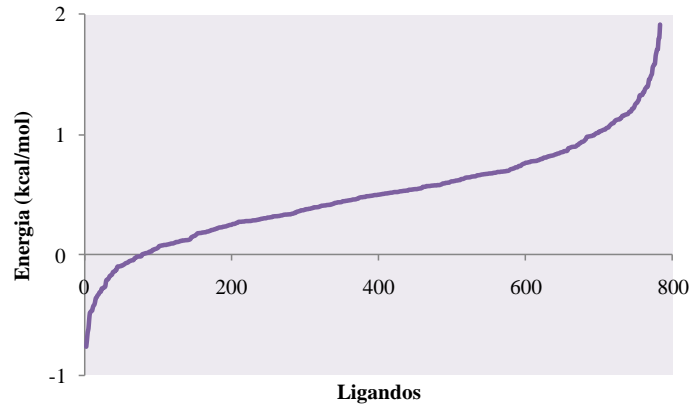
**Figura 49B:** Distribuição das energias obtida após a correção para a enzima CDC42.



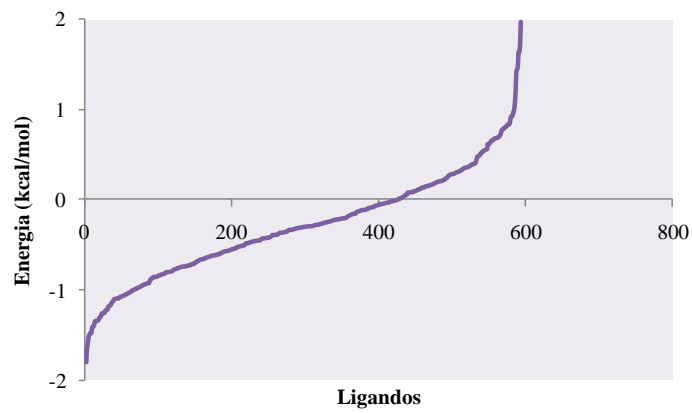
**Figura 50B:** Distribuição das energias obtida após a correção para a enzima GTP.



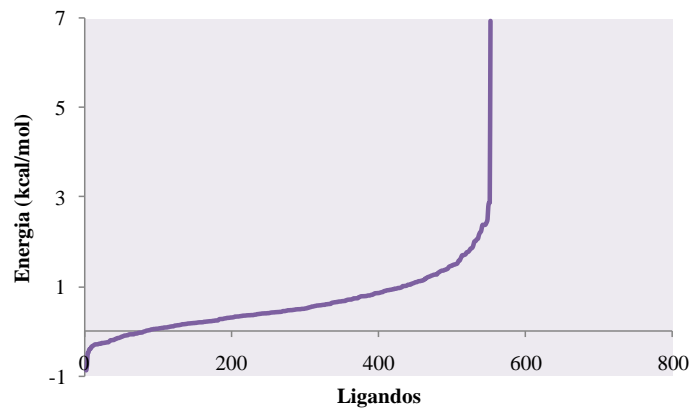
**Figura 51B:** Distribuição das energias obtida após a correção para a enzima FNTB.



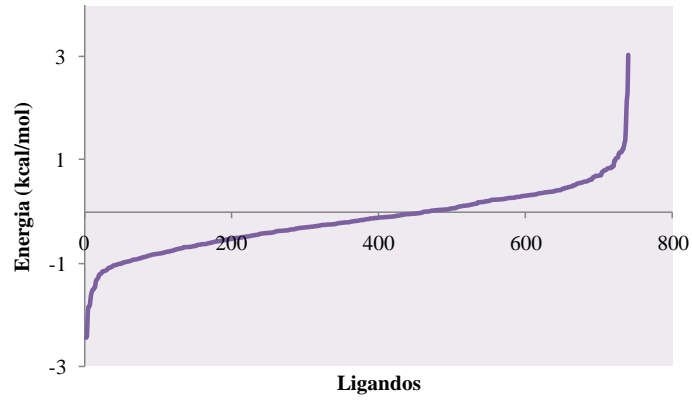
**Figura 52B:** Distribuição das energias obtida após a correção para a proteína p107.



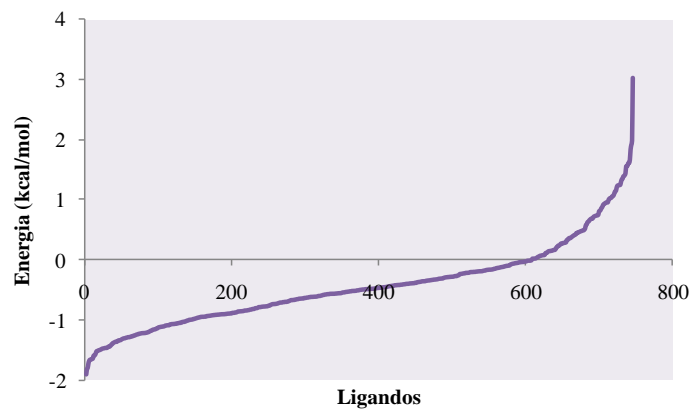
**Figura 53B:** Distribuição das energias obtida após a correção para a enzima COX.



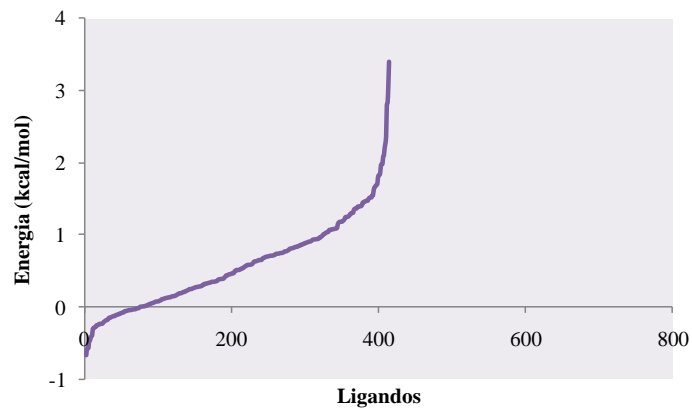
**Figura 54B:** Distribuição das energias obtida após a correção para a enzima IMPDH.



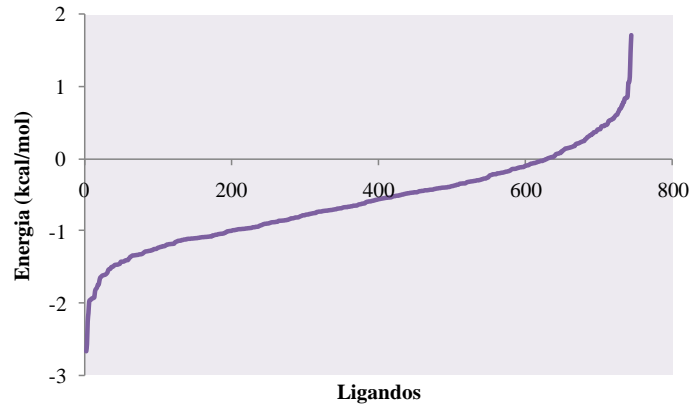
**Figura 55B:** Distribuição das energias obtida após a correção para a enzima HO-1.



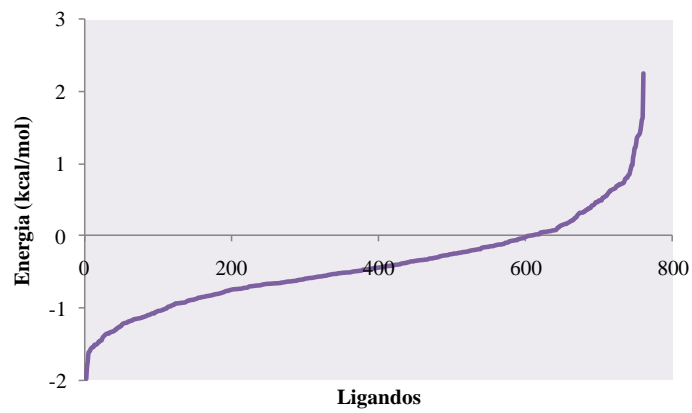
**Figura 56B:** Distribuição das energias obtida após a correção para a enzima mPGES-1-1.



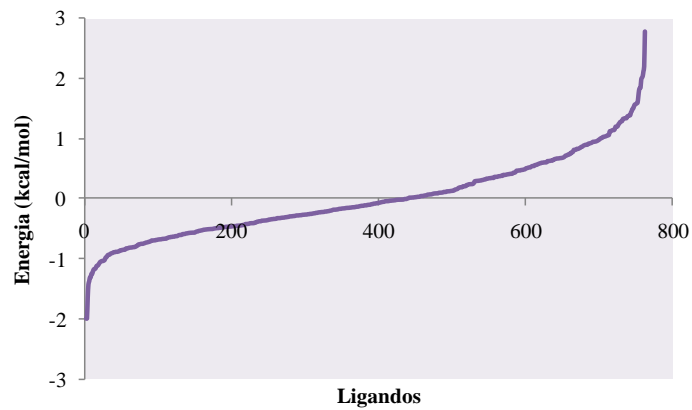
**Figura 57B:** Distribuição das energias obtida após a correção para a enzima iNOS.



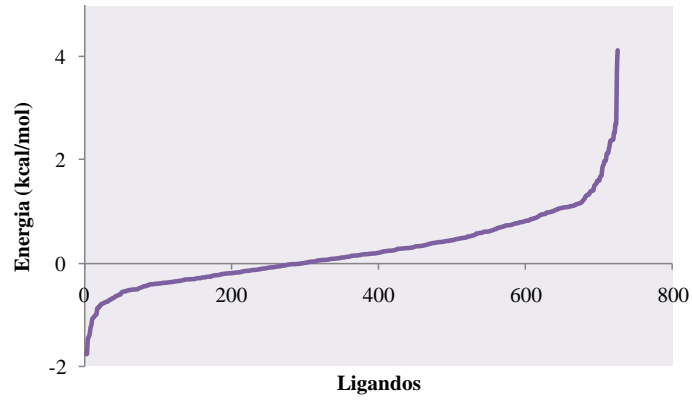
**Figura 58B:** Distribuição das energias obtida após a correção para a enzima 5-LOX.



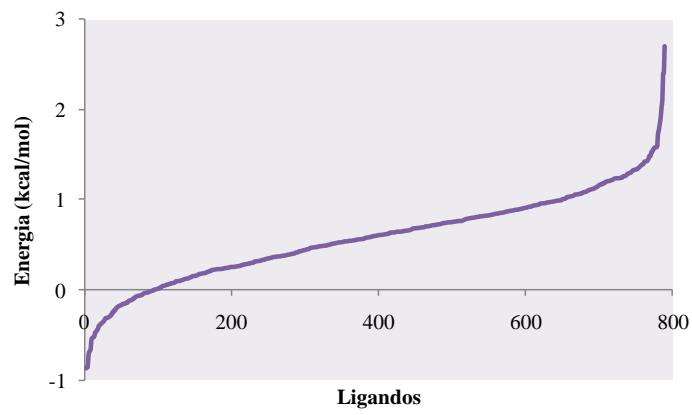
**Figura 59B:** Distribuição das energias obtidas para a enzima DHFR (bactéria).



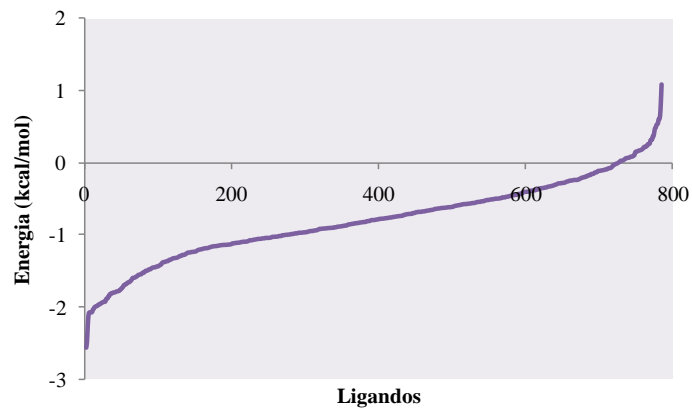
**Figura 60B:** Distribuição das energias obtidas para a enzima DHFR (fungo).



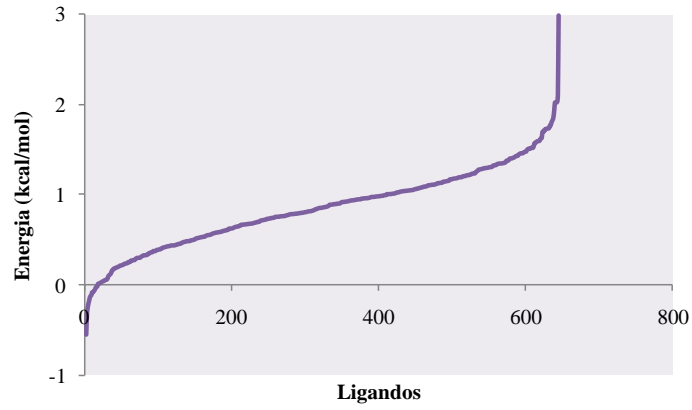
**Figura 61B:** Distribuição das energias obtida após a correção para a enzima TOPOII.



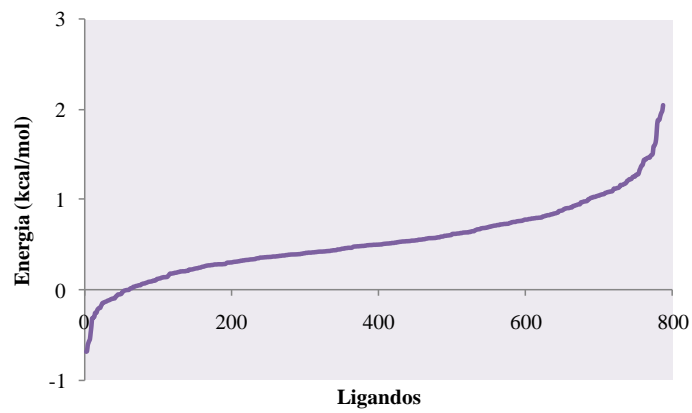
**Figura 62B:** Distribuição das energias obtida após a correção para a enzima TOPOIV.



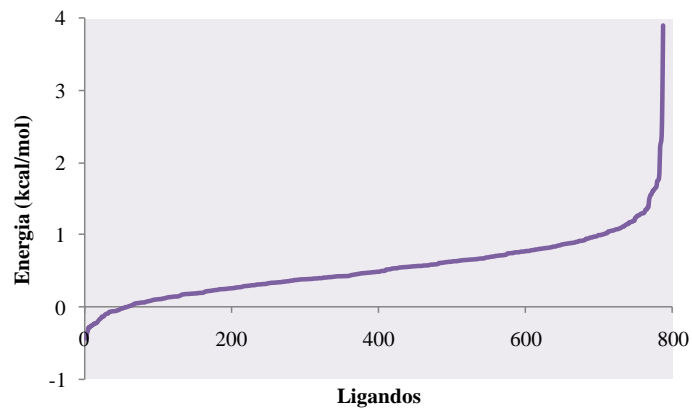
**Figura 63B:** Distribuição das energias obtida após a correção para a enzima RNAP (bactéria).



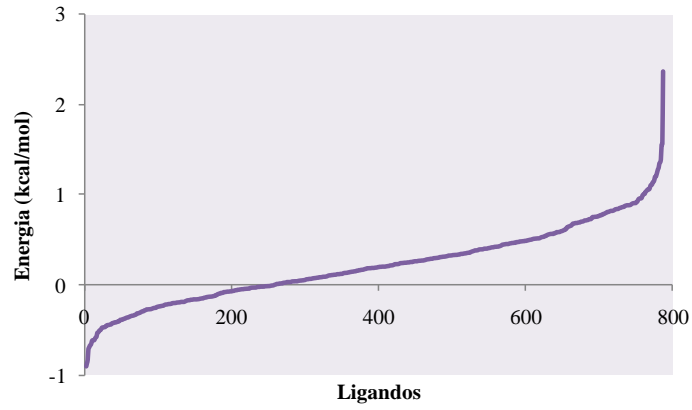
**Figura 64B:** Distribuição das energias obtida após a correção para a enzima RNAP (fungo).



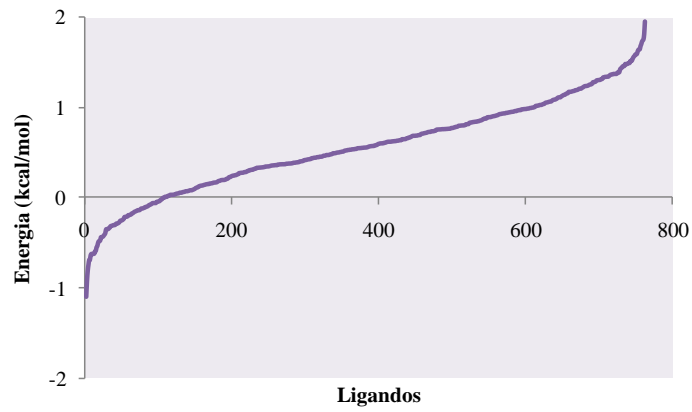
**Figura 65B:** Distribuição das energias obtida após a correção para a enzima DHPS.



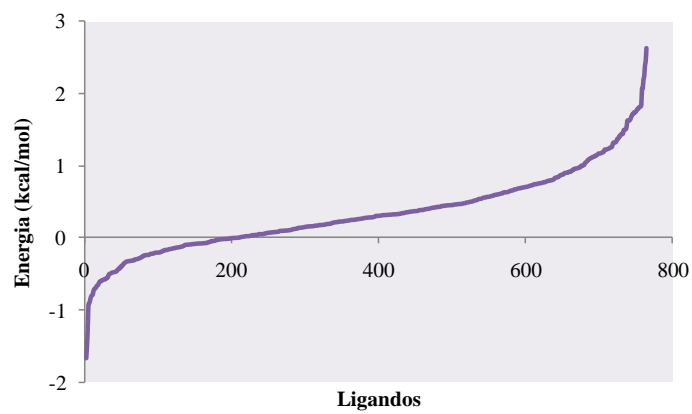
**Figura 66B:** Distribuição das energias obtida após a correção para a enzima DD-Transpeptidase.



**Figura 67B:** Distribuição das energias obtida após a correção para a enzima Racemase da alanina.



**Figura 68B:** Distribuição das energias obtida após a correção para a enzima DD-Ligase.



**Figura 69B:** Distribuição das energias obtida após a correção para a enzima EF-G (bactéria).

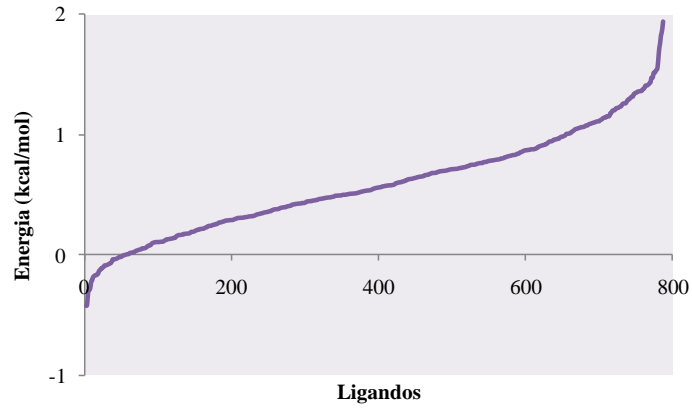


Figura 70B: Distribuição das energias obtida após a correção para a enzima EF-G (fungo).

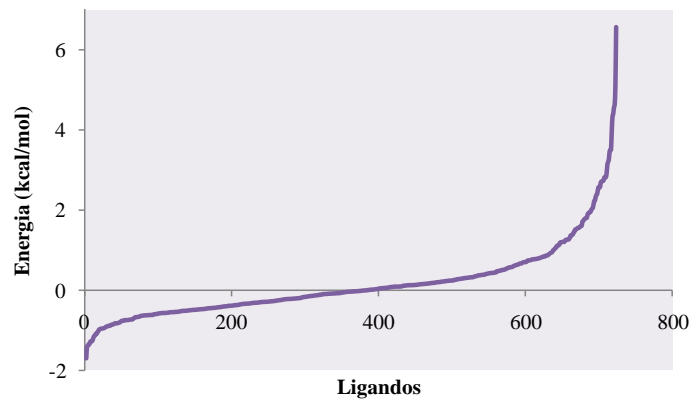


Figura 71B: Distribuição das energias obtida após a correção para a enzima NA (Complexado com o Zanamivir).

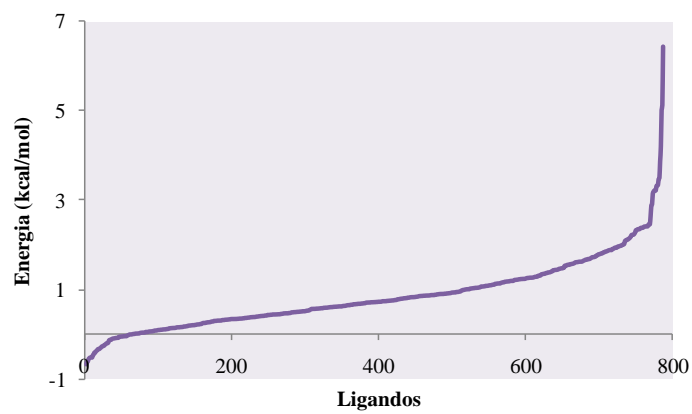
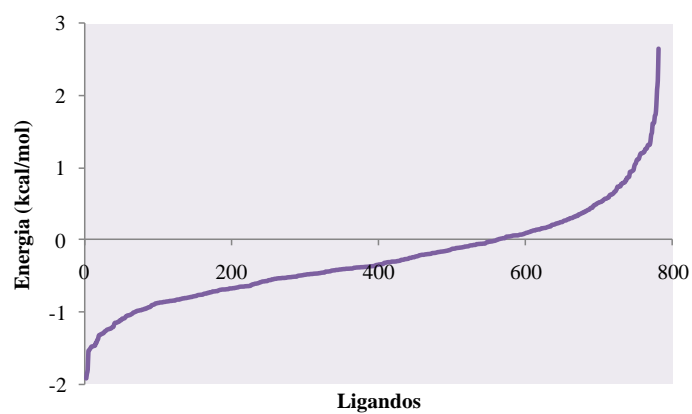
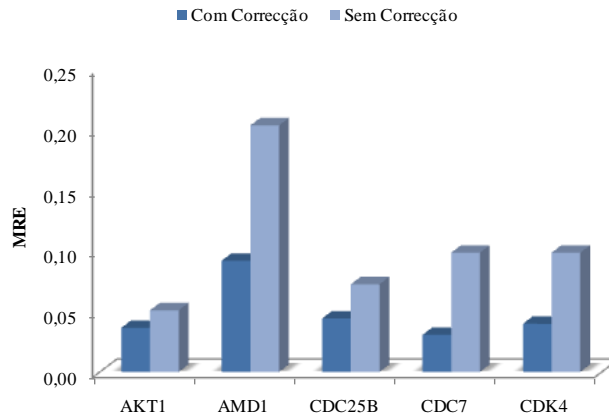


Figura 72B: Distribuição das energias obtida após a correção para a enzima NA.

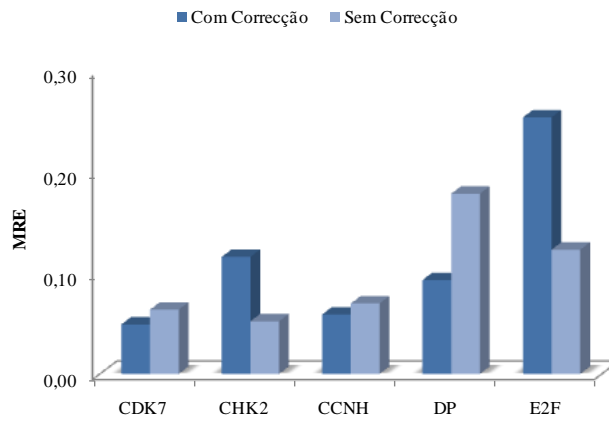


**Figura 73B:** Distribuição das energias obtida após a correção para a enzima M2TM.

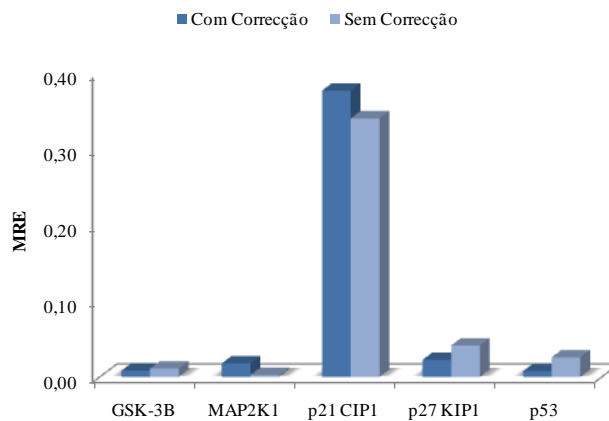
## Anexo C – Cálculo do Erro (MRE)



**Figura 1C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.



**Figura 2C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.



**Figura 3C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

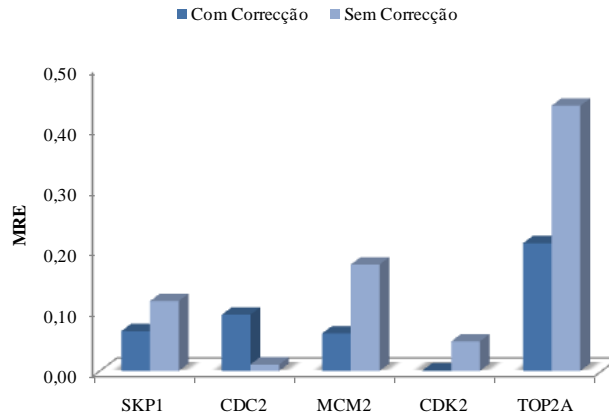


Figura 4C: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

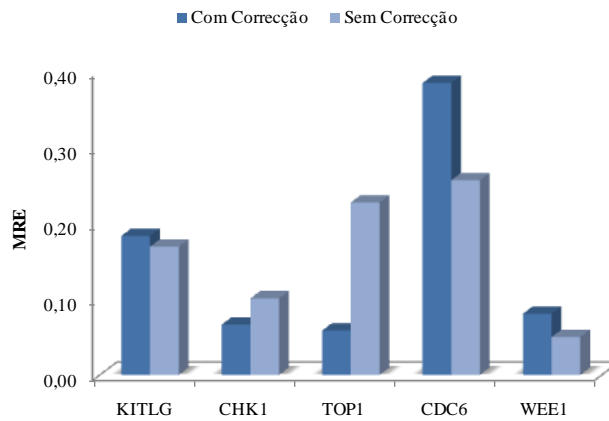


Figura 5C: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

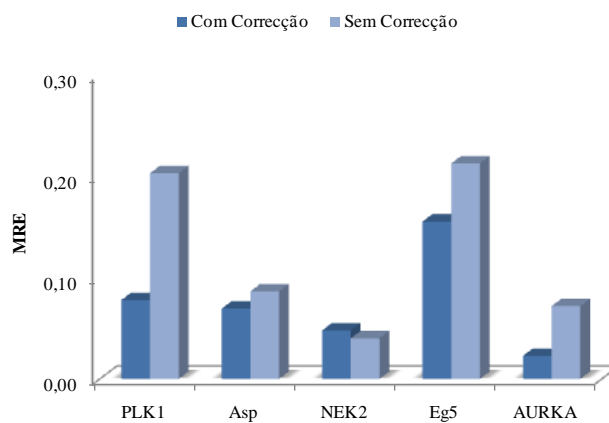
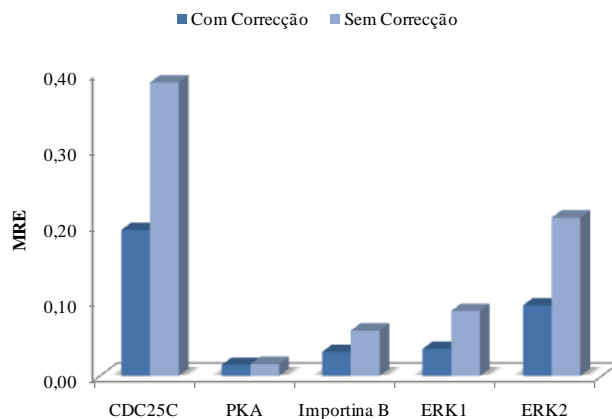
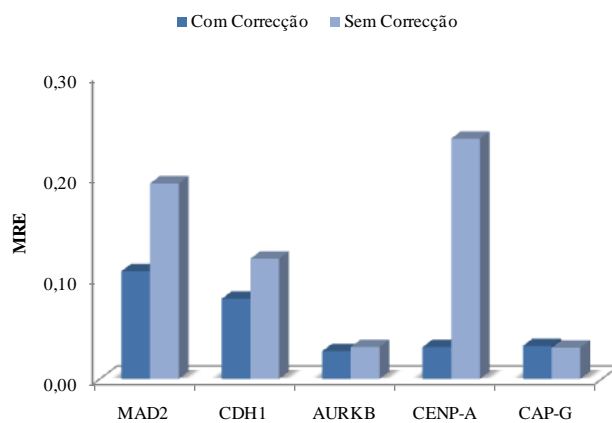


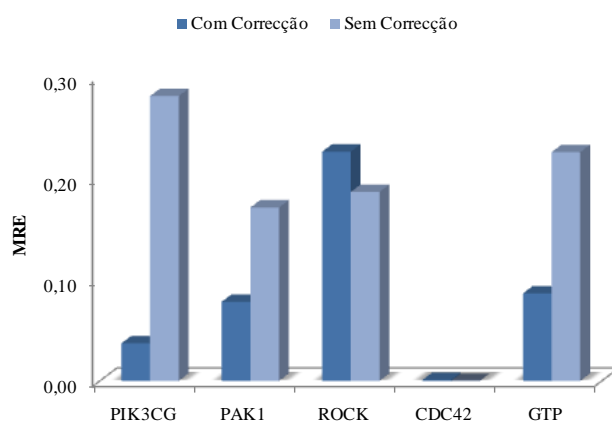
Figura 6C: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.



**Figura 7C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.



**Figura 8C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.



**Figura 9C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

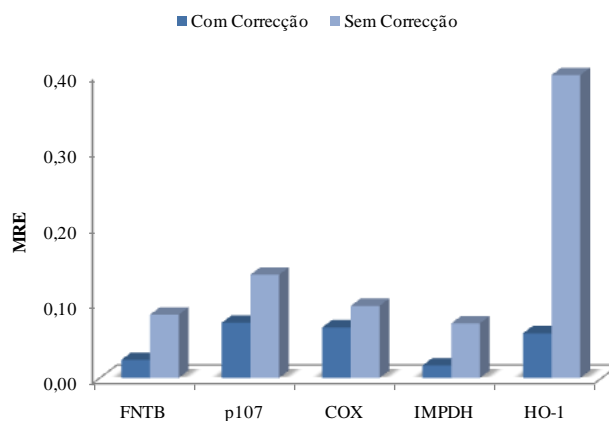


Figura 10C: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

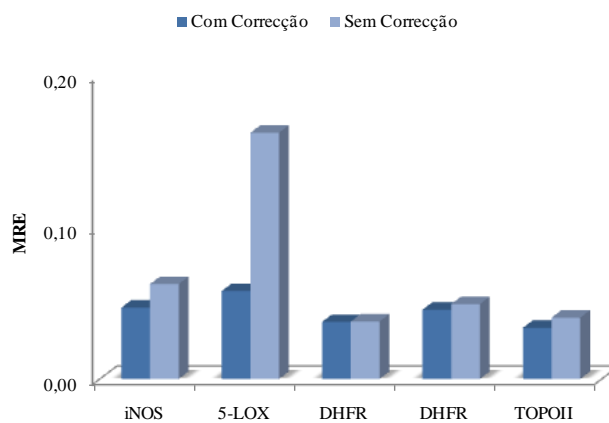


Figura 11C: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

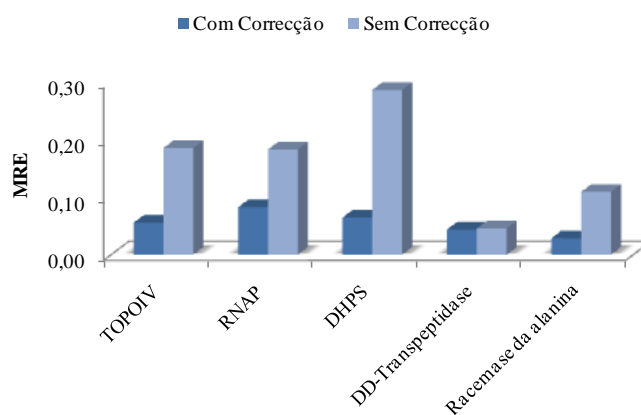
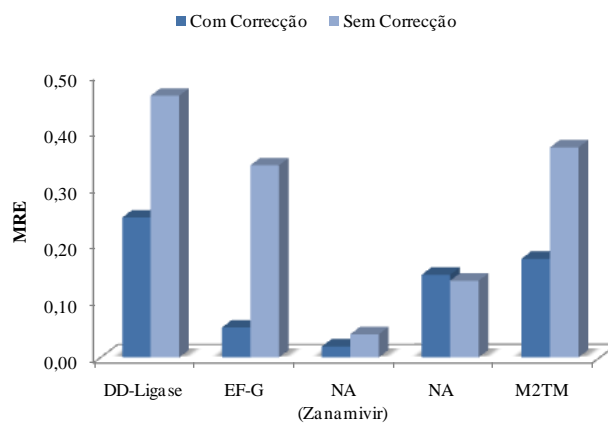
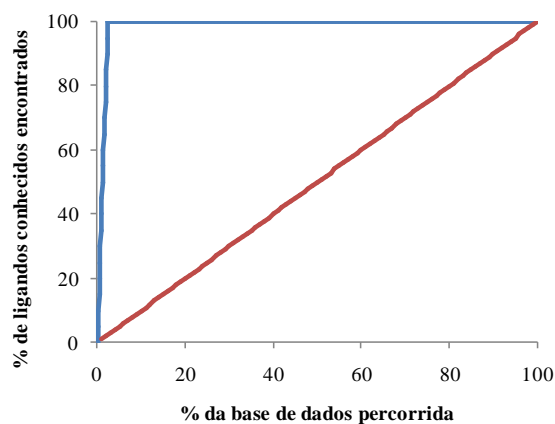


Figura 12C: Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

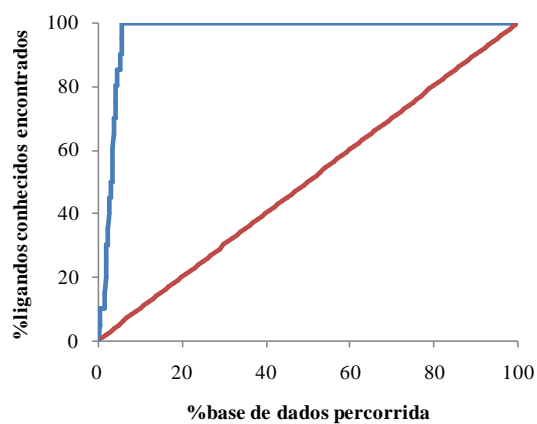


**Figura 13C:** Comparação entre os resultados obtidos para o cálculo do erro antes e após a correção.

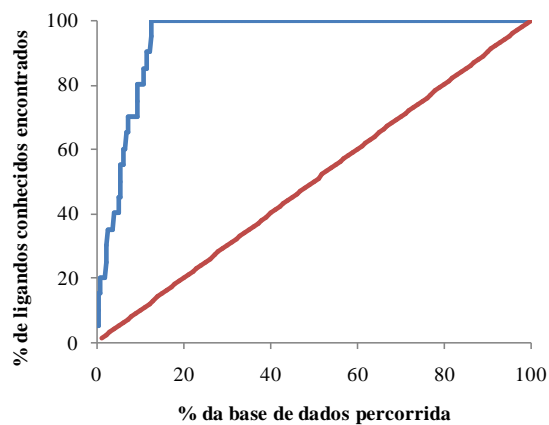
## Anexo D – Testes de Enriquecimento



**Figura 1D:** Teste de Enriquecimento realizado para a enzima AMD1.



**Figura 2D:** Teste de Enriquecimento realizado para a enzima AURKB.



**Figura 3D:** Teste de Enriquecimento realizado para a enzima CDC25B.

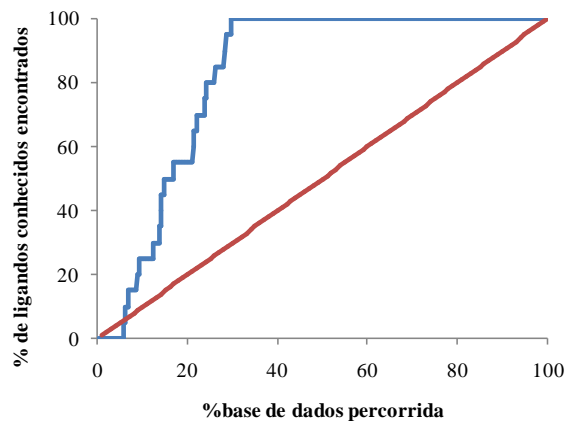


Figura 4D: Teste de Enriquecimento realizado para a enzima Eg5.

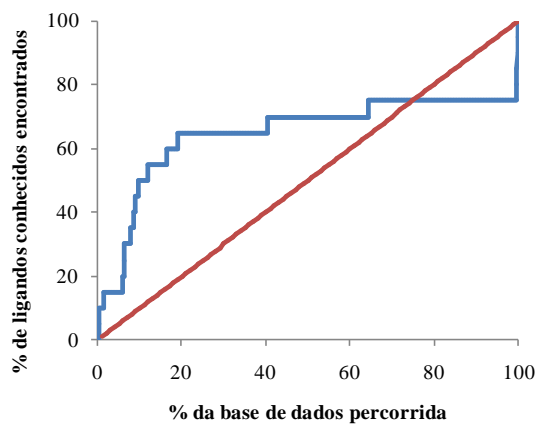


Figura 5D: Teste de Enriquecimento realizado para a enzima FNTB.

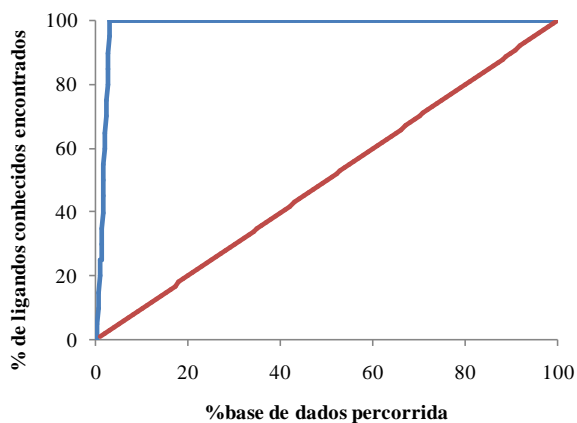
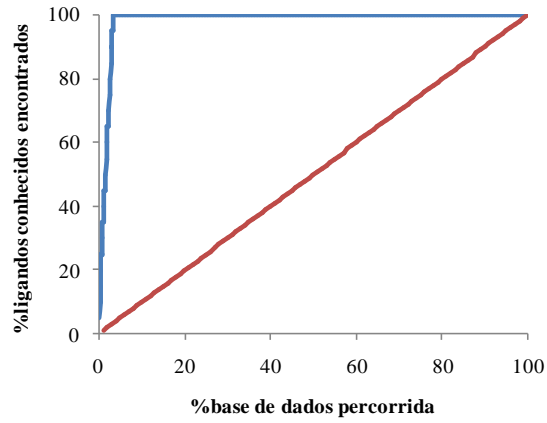
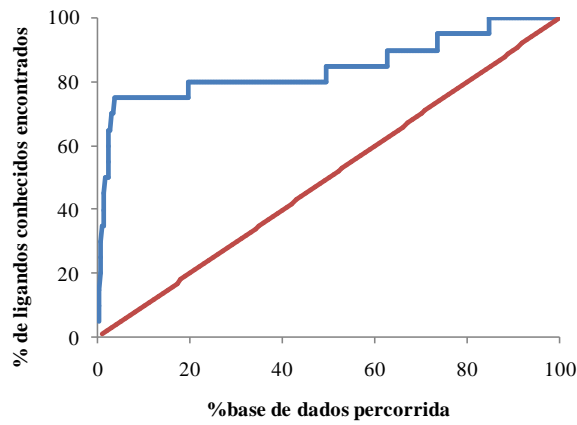


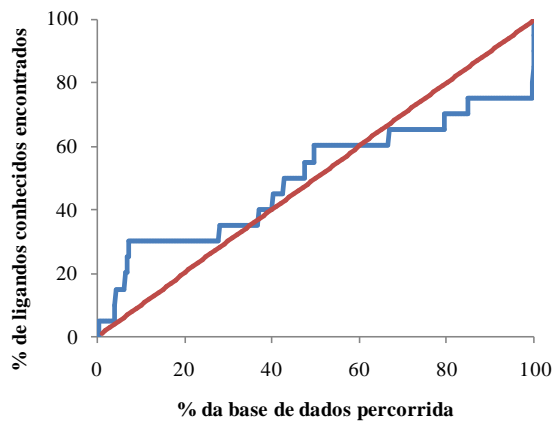
Figura 6D: Teste de Enriquecimento realizado para a enzima PIN1.



**Figura 7D:** Teste de Enriquecimento realizado para a enzima SKP1.



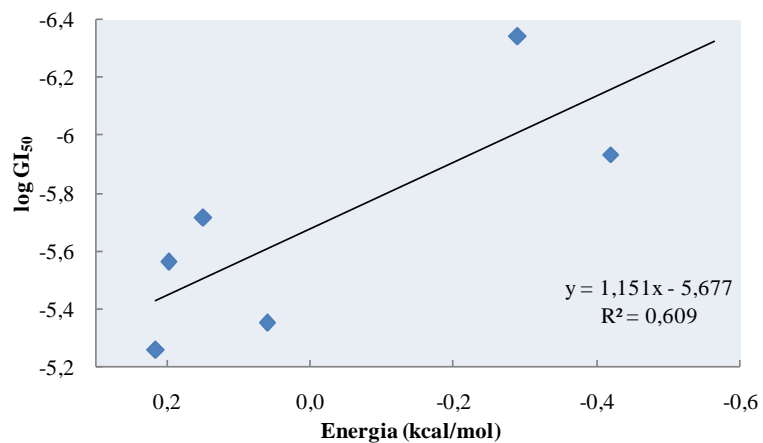
**Figura 8D:** Teste de Enriquecimento realizado para a enzima TOP1.



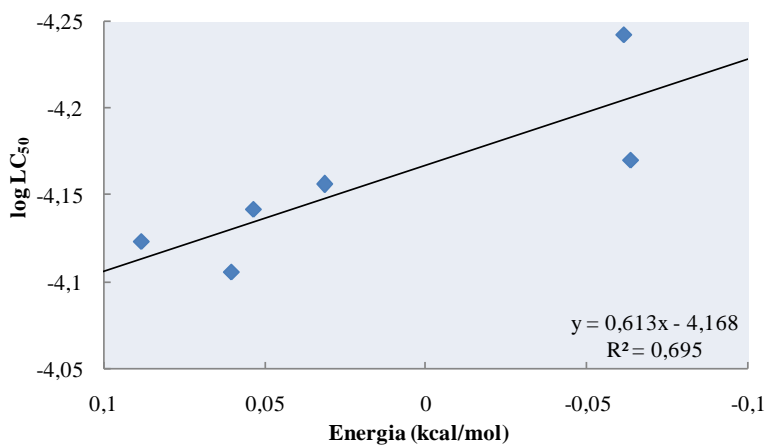
**Figura 9D:** Teste de Enriquecimento realizado para a proteína Tubulina.

## Anexo E – Correlações entre os Dados Computacionais e Experimentais

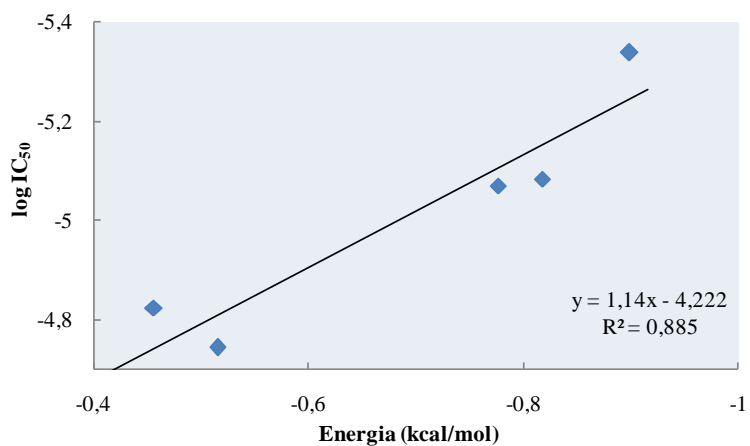
### E.1. Linha celular do cancro do ovário (A2780)



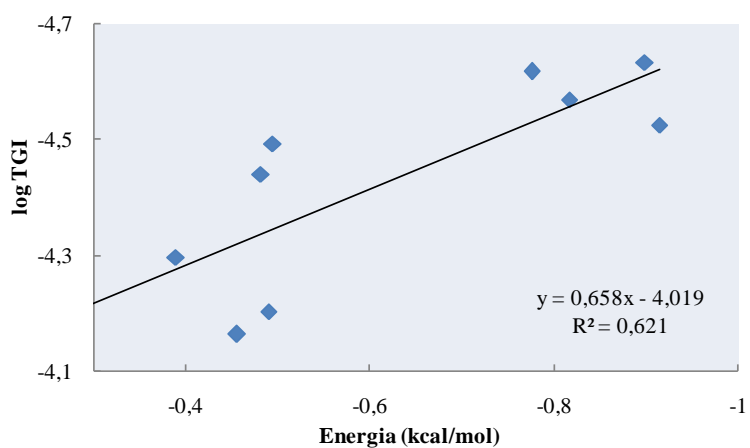
**Figura 1E:** Correlação entre os dados computacionais e experimentais para a enzima E2F para a linha celular A2780.



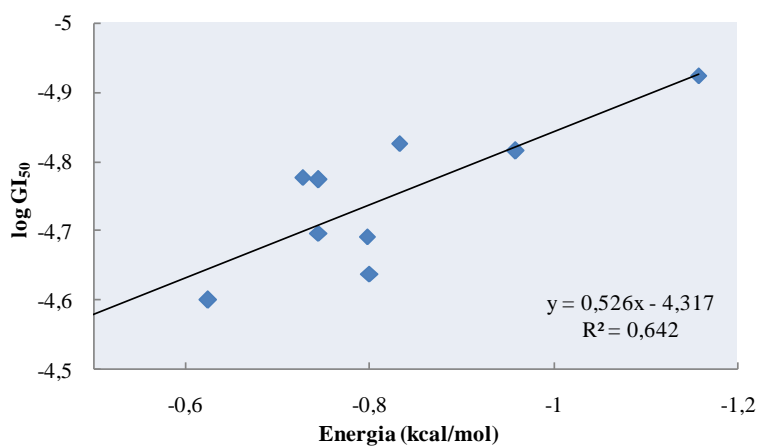
**Figura 2E:** Correlação entre os dados computacionais e experimentais para a proteína p53 para a linha celular A2780.



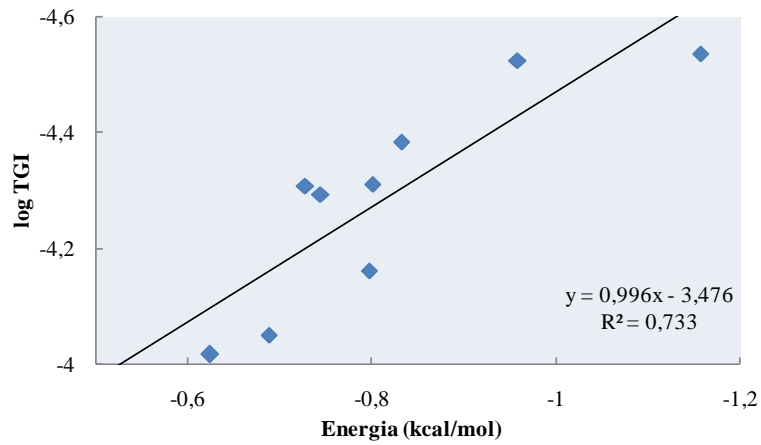
**Figura 3E:** Correlação entre os dados computacionais e experimentais para a enzima CDC2 para a linha celular A2780.



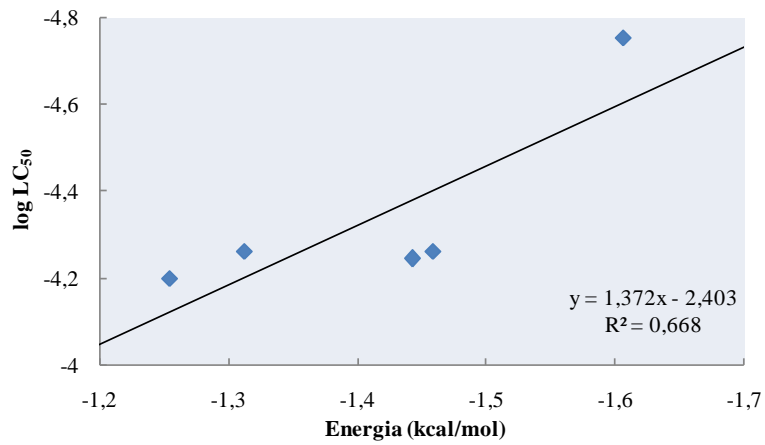
**Figura 4E:** Correlação entre os dados computacionais e experimentais para a enzima CDC2 para a linha celular A2780.



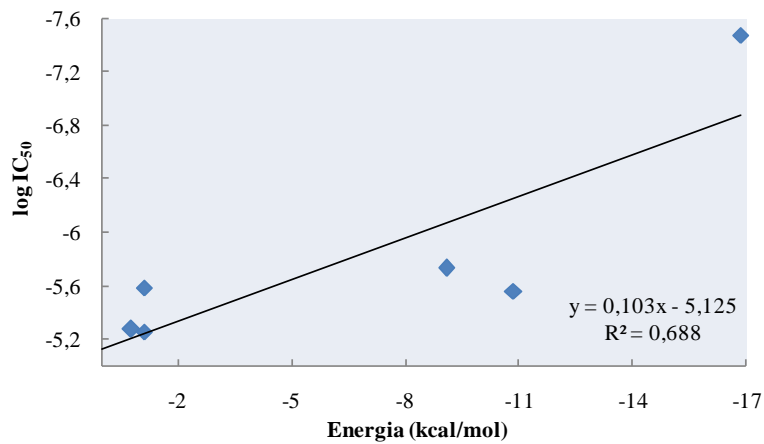
**Figura 5E:** Correlação entre os dados computacionais e experimentais para a enzima TOP2A para a linha celular A2780.



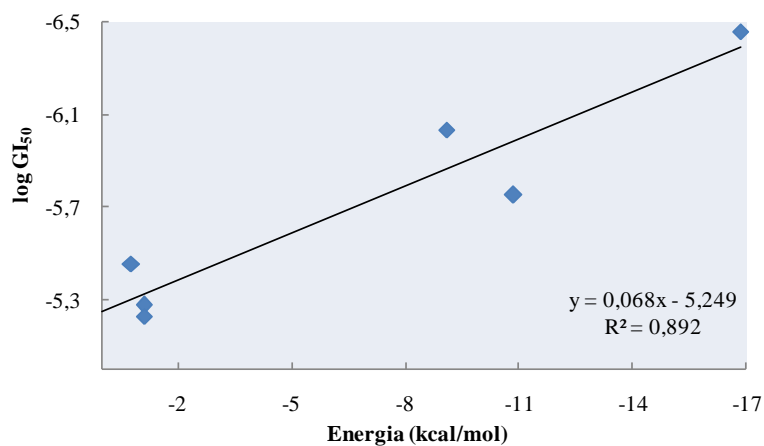
**Figura 6E:** Correlação entre os dados computacionais e experimentais para a enzima TOP2A para a linha celular A2780.



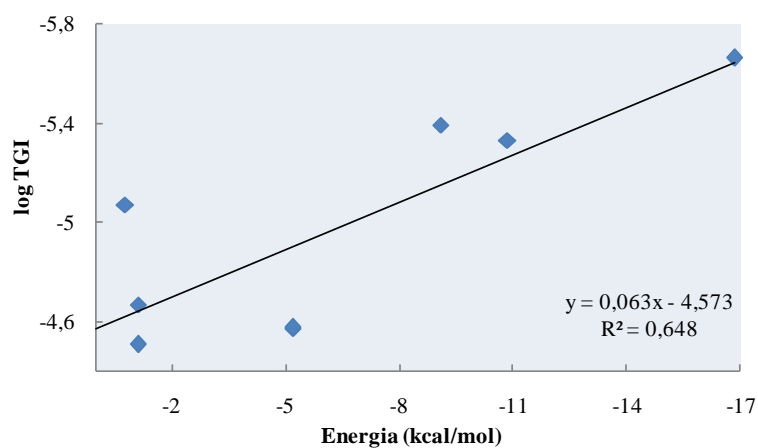
**Figura 7E:** Correlação entre os dados computacionais e experimentais para a enzima TOP1 para a linha celular A2780.



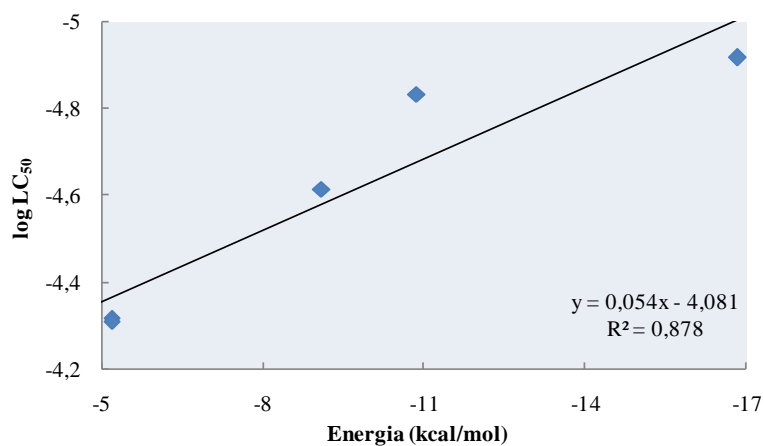
**Figura 8E:** Correlação entre os dados computacionais e experimentais para a enzima CDC25C para a linha celular A2780.



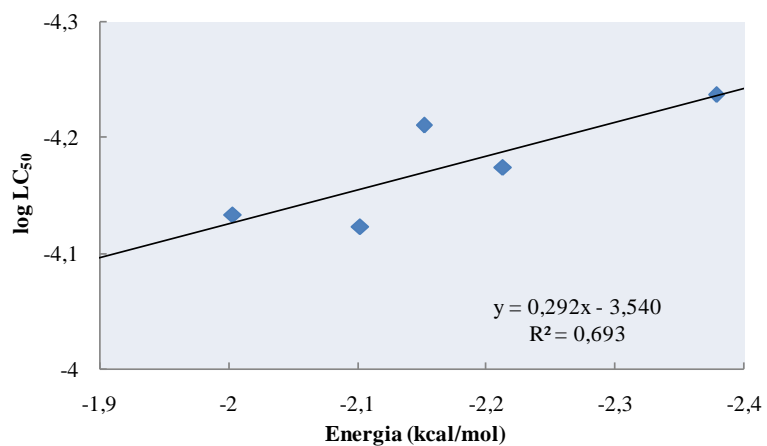
**Figura 9E:** Correlação entre os dados computacionais e experimentais para a enzima CDC25C para a linha celular A2780.



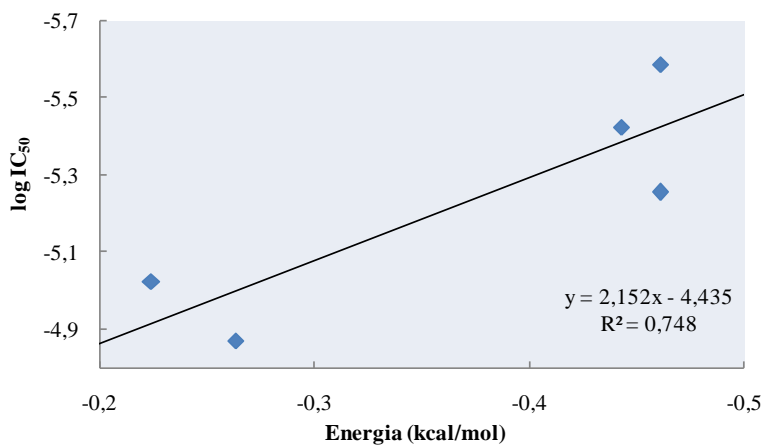
**Figura 10E:** Correlação entre os dados computacionais e experimentais para a enzima CDC25C para a linha celular A2780.



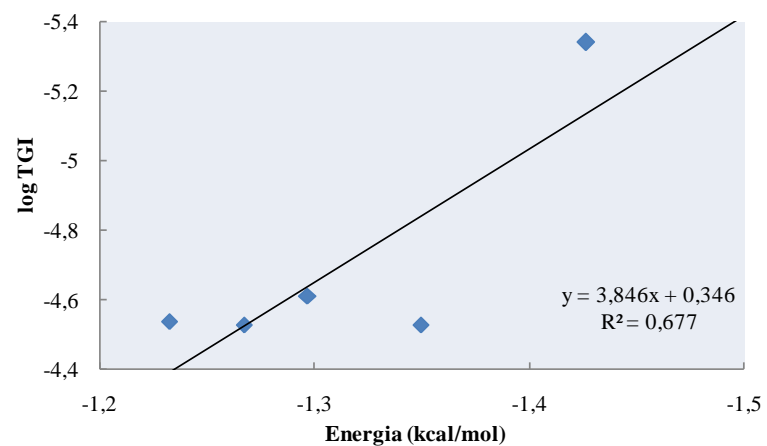
**Figura 11E:** Correlação entre os dados computacionais e experimentais para a enzima CDC25C para a linha celular A2780.



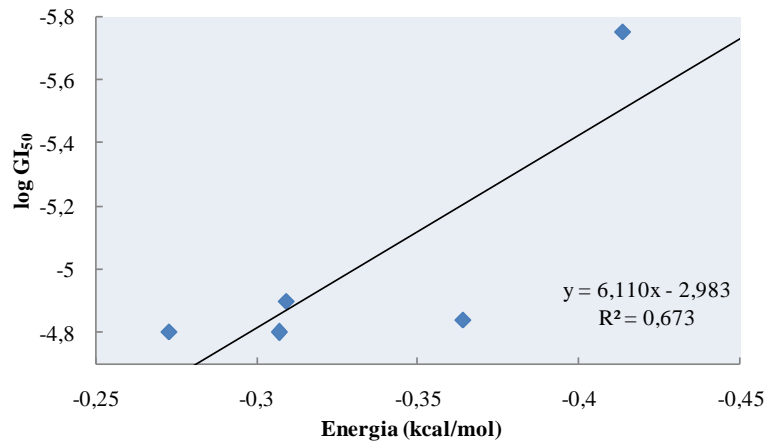
**Figura 12E:** Correlação entre os dados computacionais e experimentais para a enzima ERK2 para a linha celular A2780.



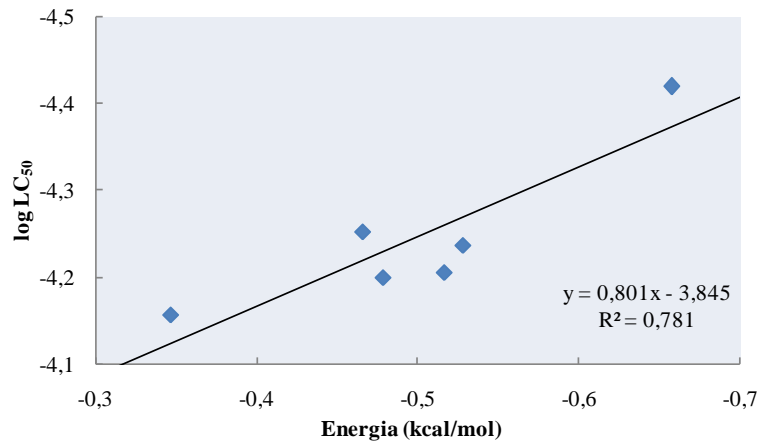
**Figura 13E:** Correlação entre os dados computacionais e experimentais para a enzima PAK1 para a linha celular A2780.



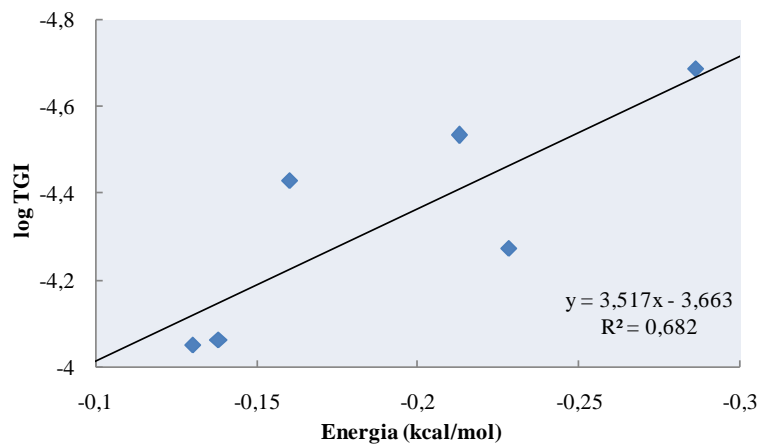
**Figura 14E:** Correlação entre os dados computacionais e experimentais para a enzima COX para a linha celular A2780.



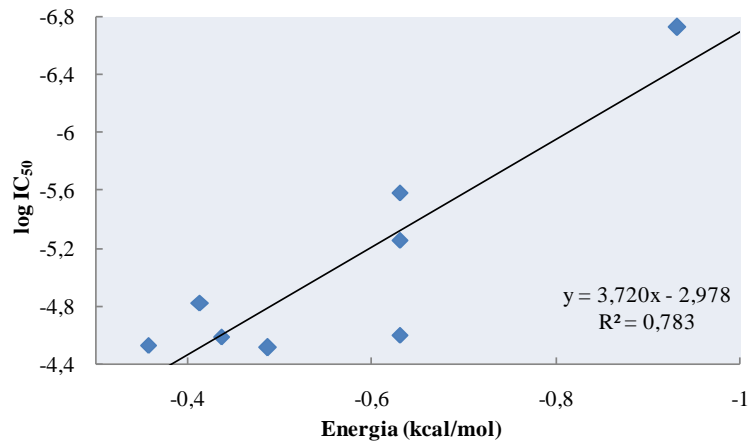
**Figura 15E:** Correlação entre os dados computacionais e experimentais para a enzima IMPDH para a linha celular A2780.



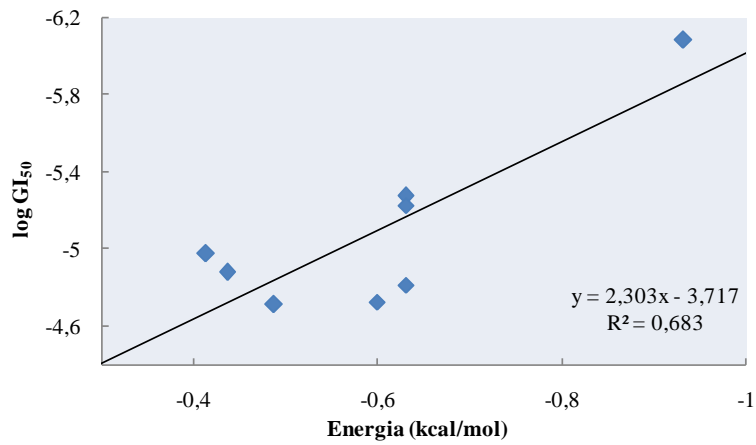
**Figura 16E:** Correlação entre os dados computacionais e experimentais para a enzima TOPOIV para a linha celular A2780.



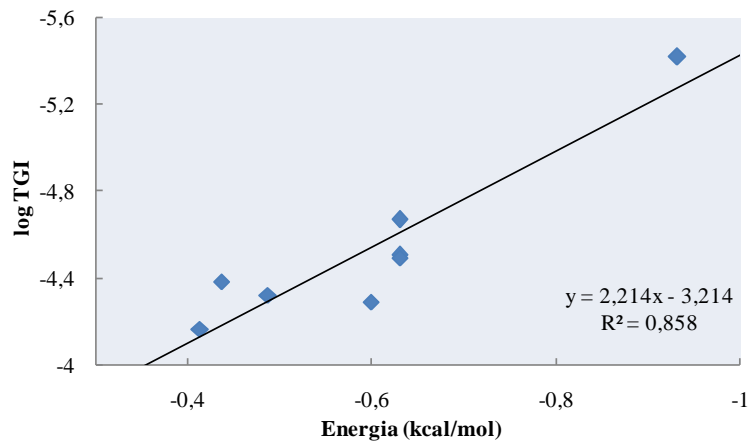
**Figura 17E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Transpeptidase para a linha celular A2780.



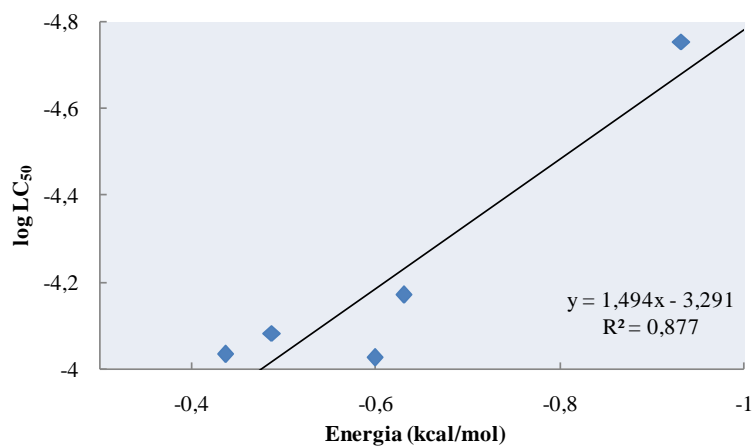
**Figura 18E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular A2780.



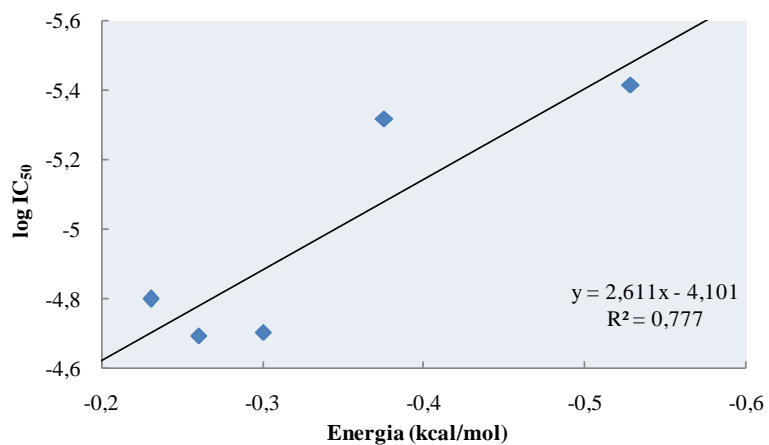
**Figura 19E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular A2780.



**Figura 20E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular A2780.

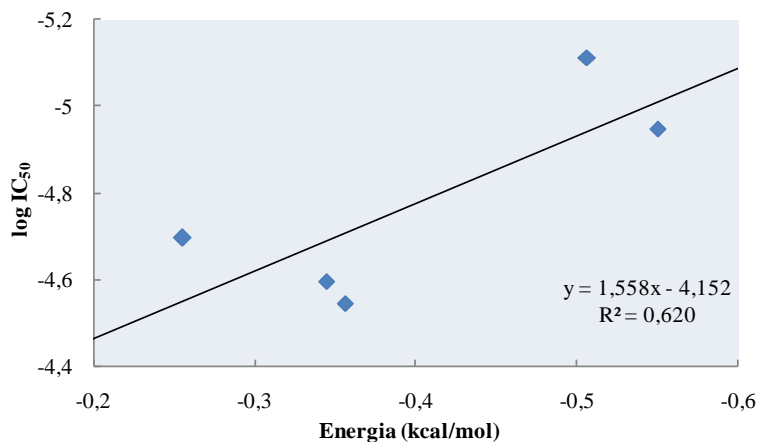


**Figura 21E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular A2780.

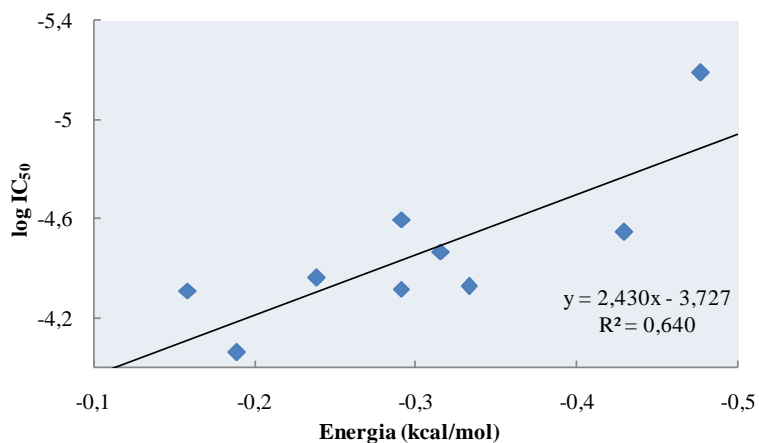


**Figura 22E:** Correlação entre os dados computacionais e experimentais para a enzima NA para a linha celular A2780.

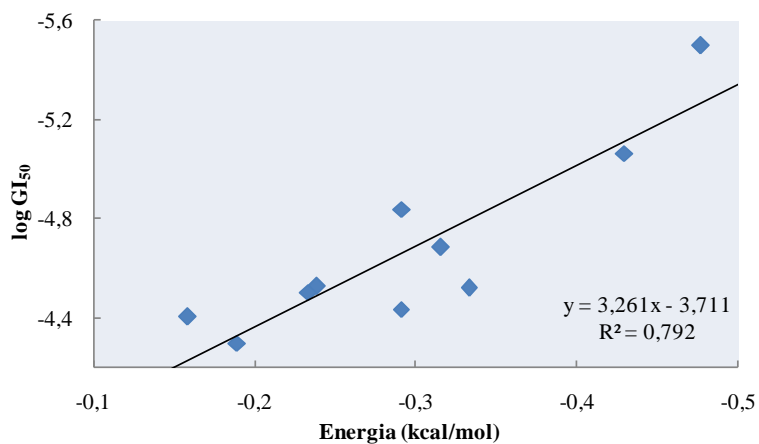
E.2. Linha celular da leucemia (HL60)



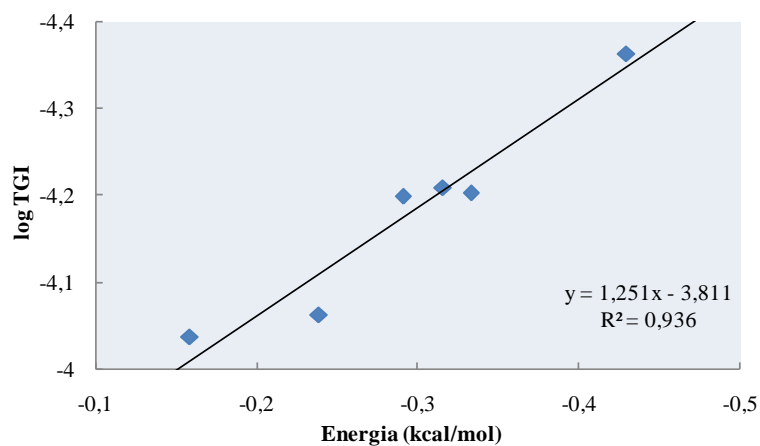
**Figura 23E:** Correlação entre os dados computacionais e experimentais para a enzima AKT1 para a linha celular HL60.



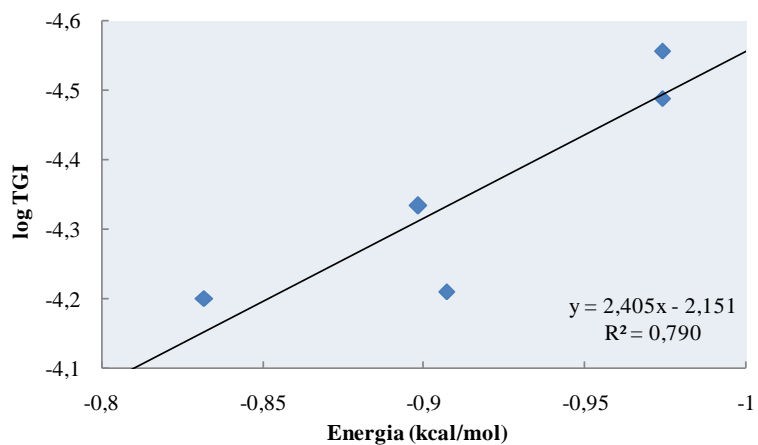
**Figura 24E:** Correlação entre os dados computacionais e experimentais para a enzima CDC7 para a linha celular HL60.



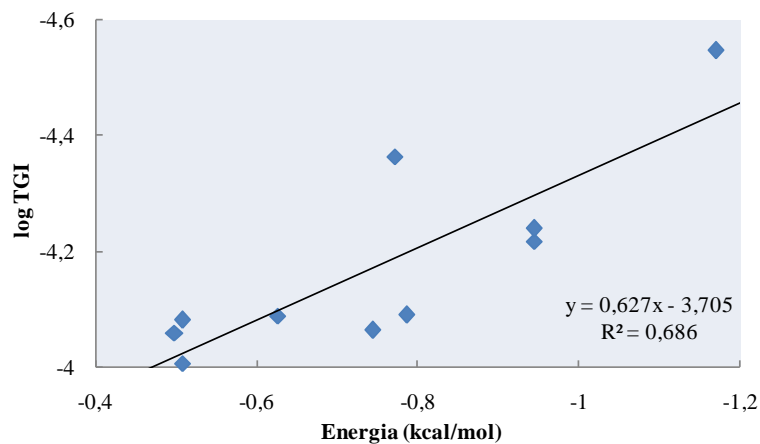
**Figura 25E:** Correlação entre os dados computacionais e experimentais para a enzima CDC7 para a linha celular HL60.



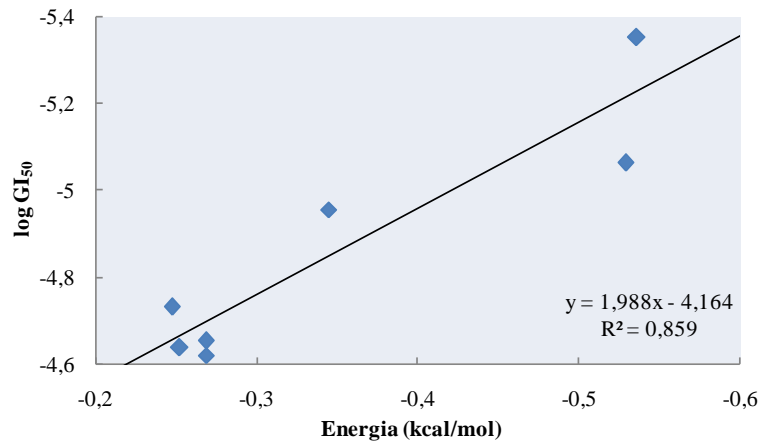
**Figura 26E:** Correlação entre os dados computacionais e experimentais para a enzima CDC7 para a linha celular HL60.



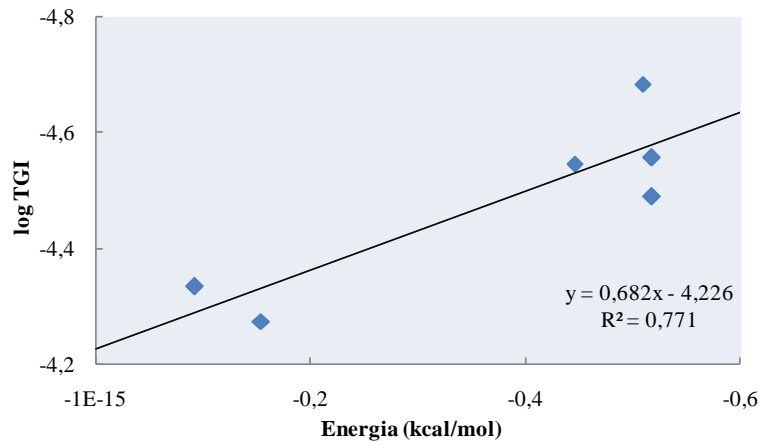
**Figura 27E:** Correlação entre os dados computacionais e experimentais para a enzima CDK7 para a linha celular HL60.



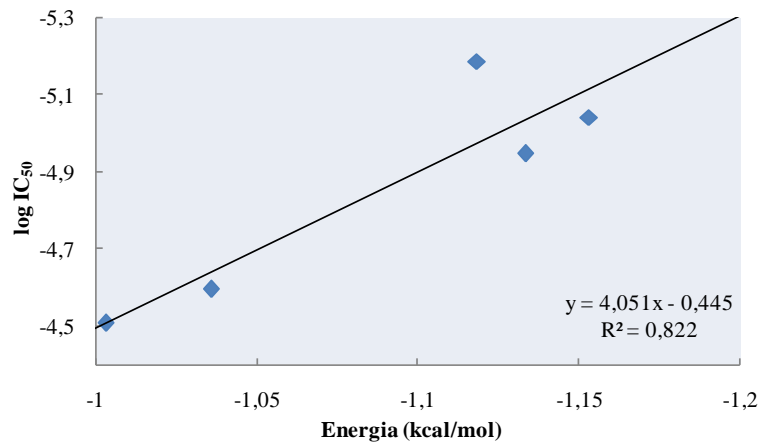
**Figura 28E:** Correlação entre os dados computacionais e experimentais para a enzima SKP1 para a linha celular HL60.



**Figura 29E:** Correlação entre os dados computacionais e experimentais para a enzima MCM2 para a linha celular HL60.

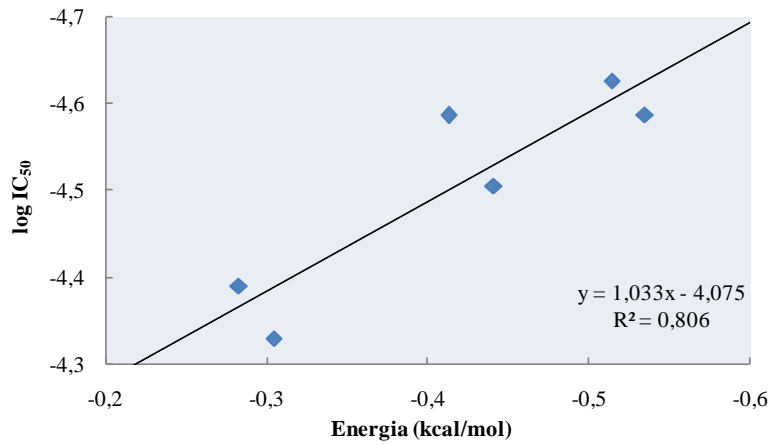


**Figura 30E:** Correlação entre os dados computacionais e experimentais para a proteína KITLG para a linha celular HL60.

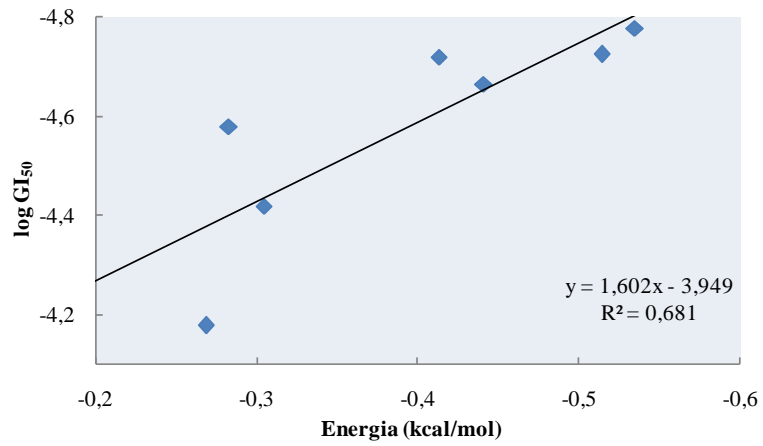


**Figura 31E:** Correlação entre os dados computacionais e experimentais para a enzima PLK1 para a linha celular HL60.

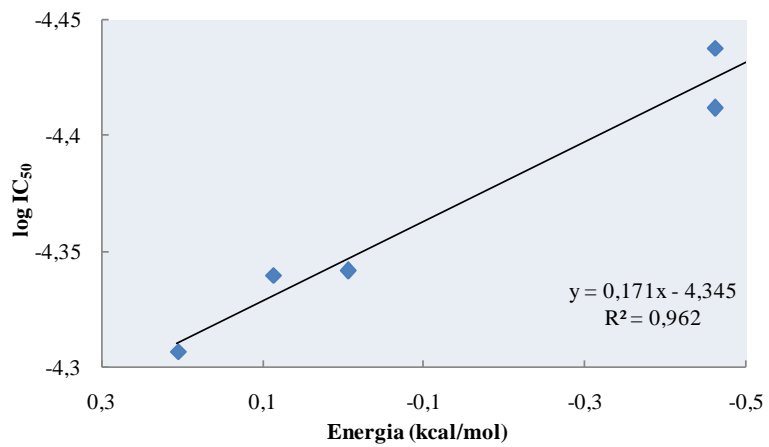




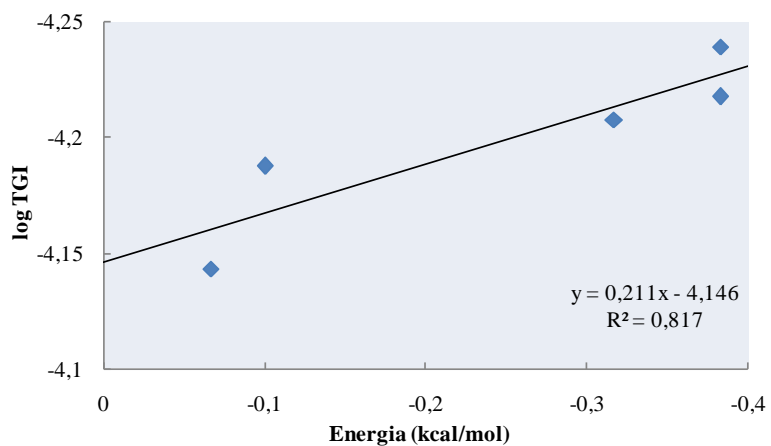
**Figura 35E:** Correlação entre os dados computacionais e experimentais para a enzima PIN1 para a linha celular HL60.



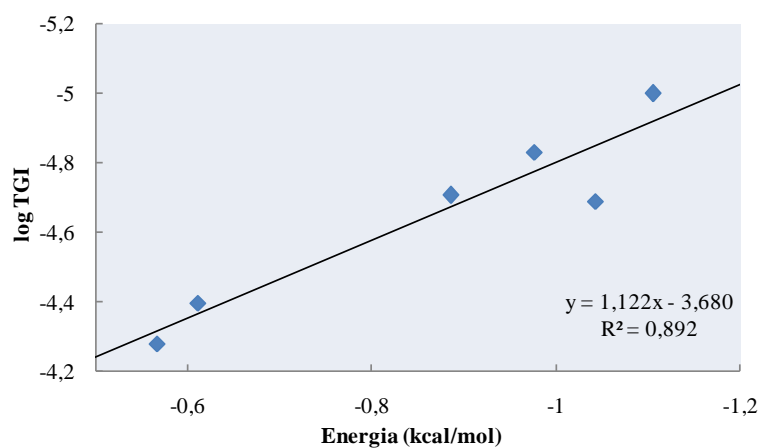
**Figura 36E:** Correlação entre os dados computacionais e experimentais para a enzima PIN1 para a linha celular HL60.



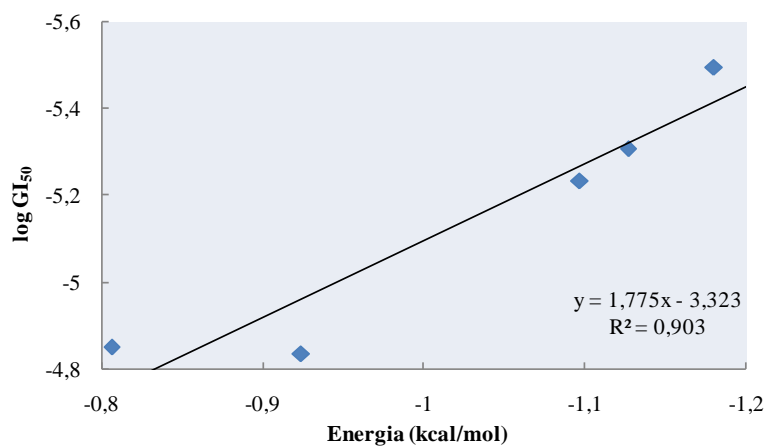
**Figura 37E:** Correlação entre os dados computacionais e experimentais para a enzima PAK1 para a linha celular HL60.



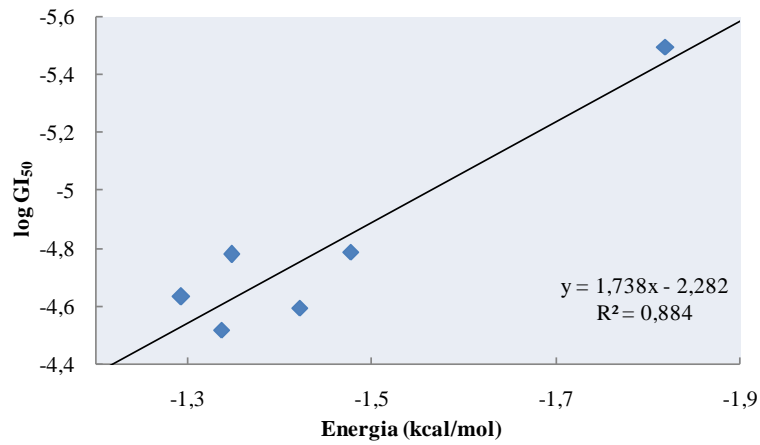
**Figura 38E:** Correlação entre os dados computacionais e experimentais para a enzima ROCK para a linha celular HL60.



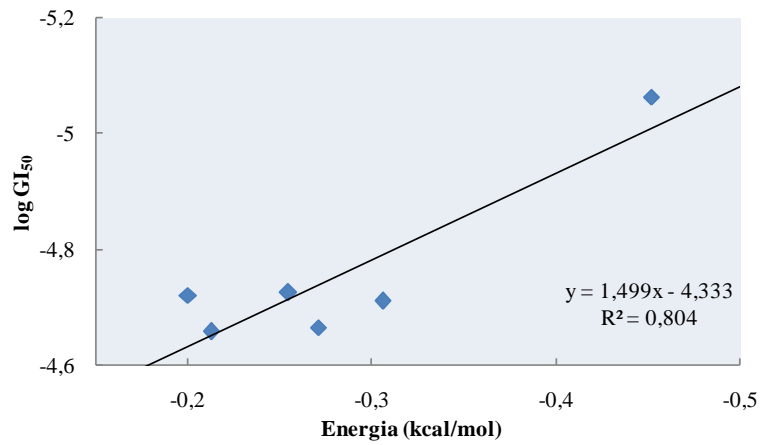
**Figura 39E:** Correlação entre os dados computacionais e experimentais para a enzima FNTB para a linha celular HL60.



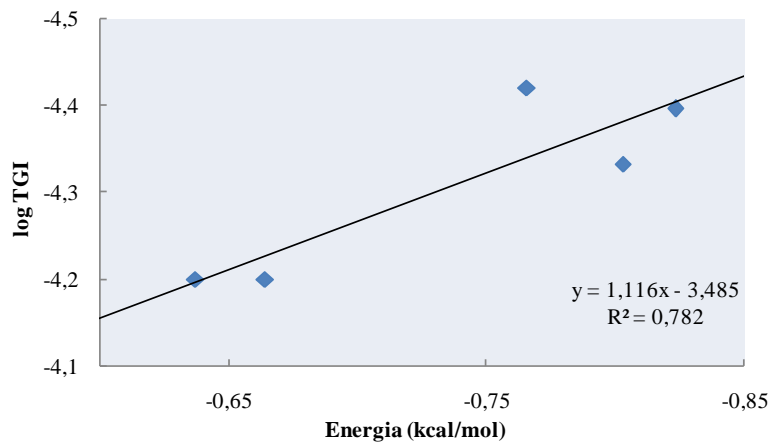
**Figura 40E:** Correlação entre os dados computacionais e experimentais para a enzima COX para a linha celular HL60.



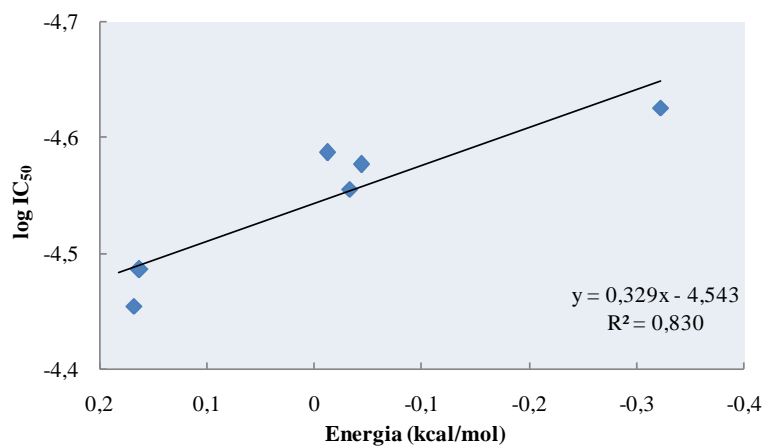
**Figura 41E:** Correlação entre os dados computacionais e experimentais para a enzima mPGES-1 para a linha celular HL60.



**Figura 42E:** Correlação entre os dados computacionais e experimentais para a enzima Racemase da alanina para a linha celular HL60.

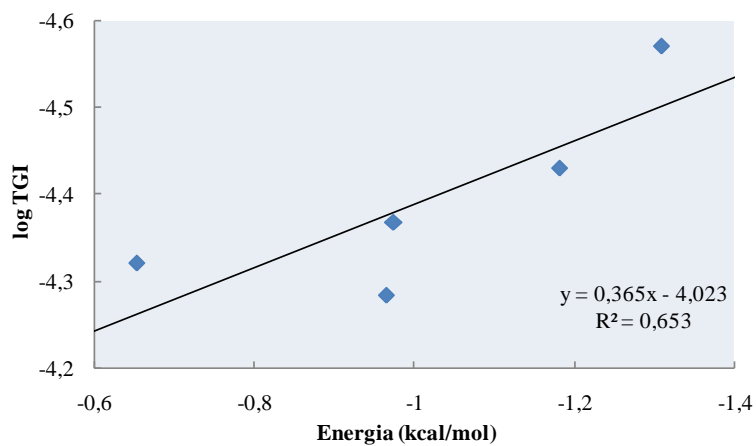


**Figura 43E:** Correlação entre os dados computacionais e experimentais para a enzima NA (Complexado com o Zanamivir) para a linha celular HL60.

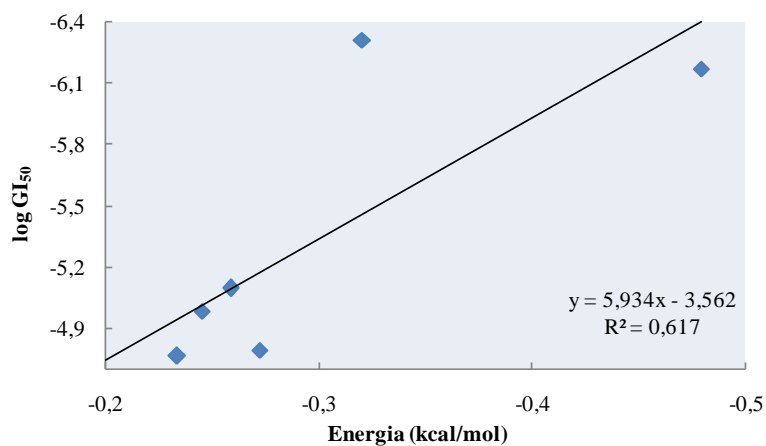


**Figura 44E:** Correlação entre os dados computacionais e experimentais para a enzima NA para a linha celular HL60.

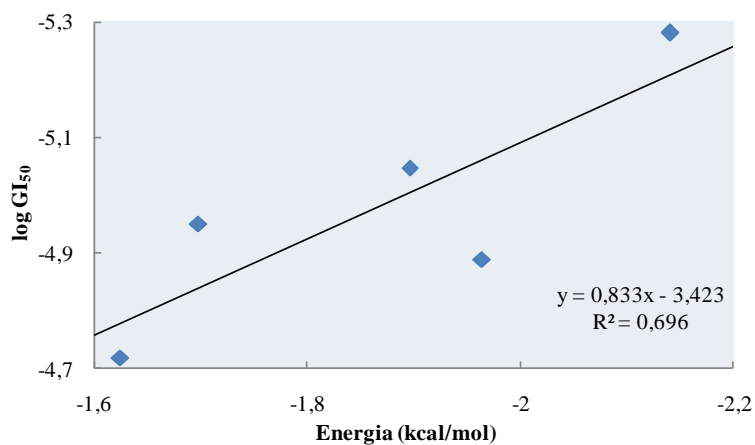
### E.3. Linha celular do cancro do pulmão (SW1573)



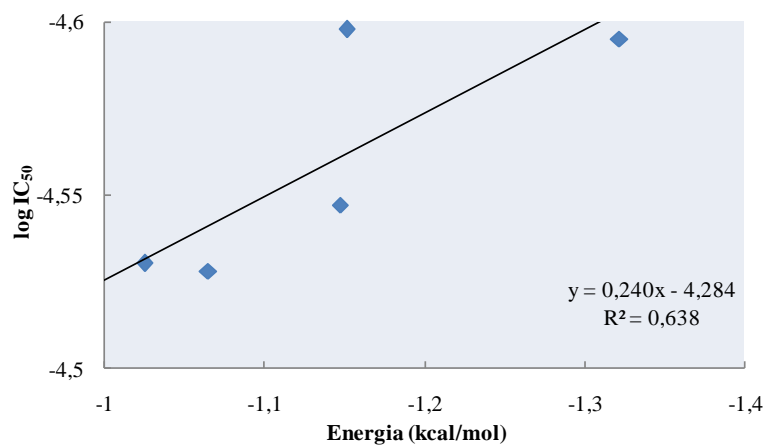
**Figura 45E:** Correlação entre os dados computacionais e experimentais para a enzima CDC25B para a linha celular SW1573.



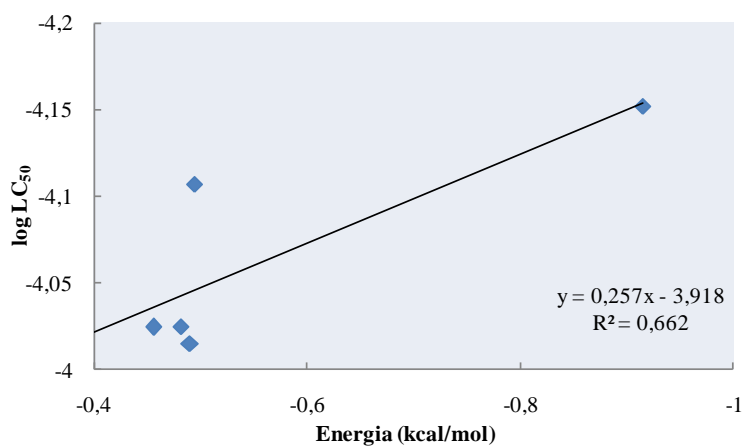
**Figura 46E:** Correlação entre os dados computacionais e experimentais para a enzima CCNH para a linha celular SW1573.



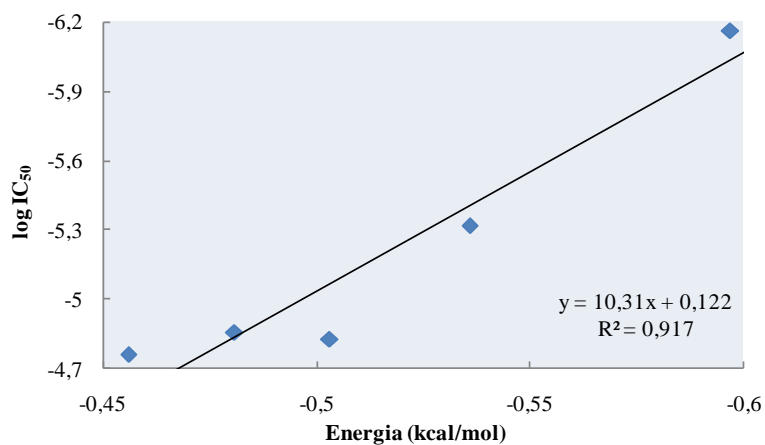
**Figura 47E:** Correlação entre os dados computacionais e experimentais para a enzima MAP2K1 para a linha celular SW1573.



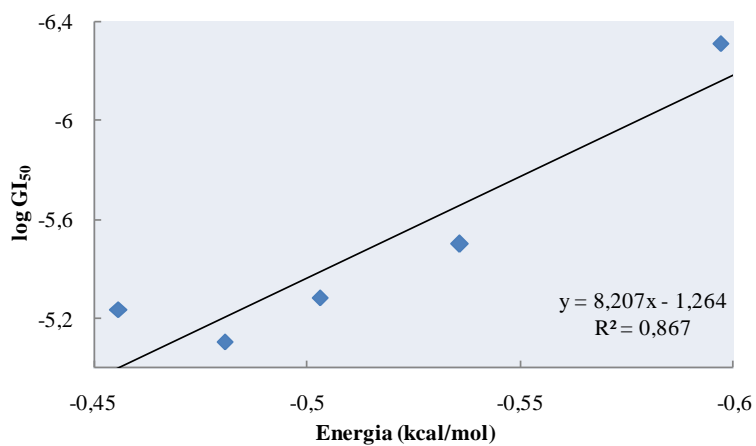
**Figura 48E:** Correlação entre os dados computacionais e experimentais para a enzima p27 KIP1 para a linha celular SW1573.



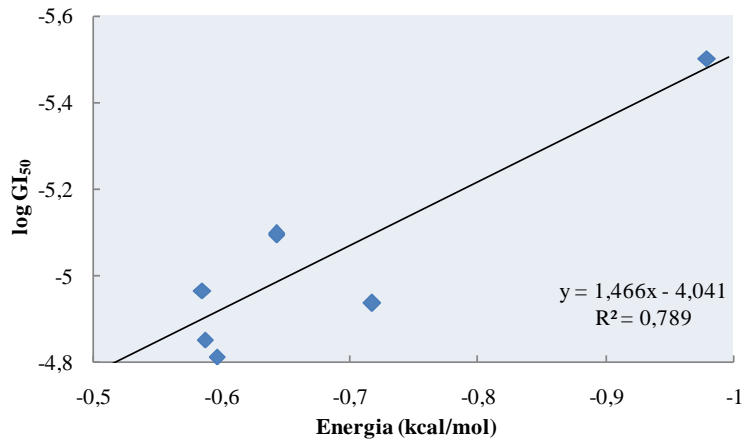
**Figura 49E:** Correlação entre os dados computacionais e experimentais para a enzima CDC2 para a linha celular SW1573.



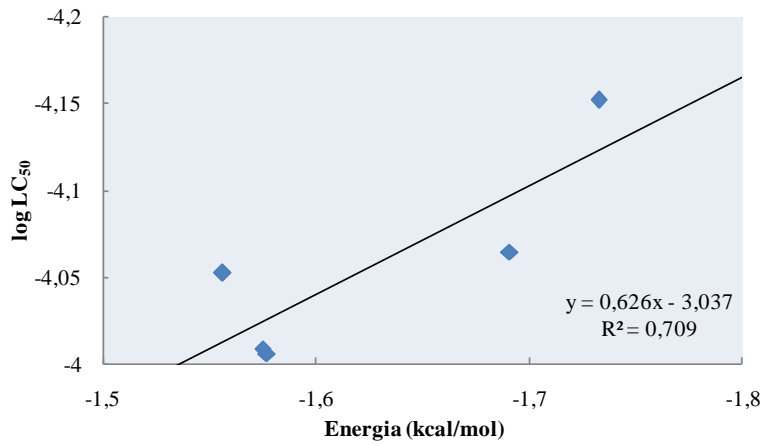
**Figura 50E:** Correlação entre os dados computacionais e experimentais para a enzima Asp para a linha celular SW1573.



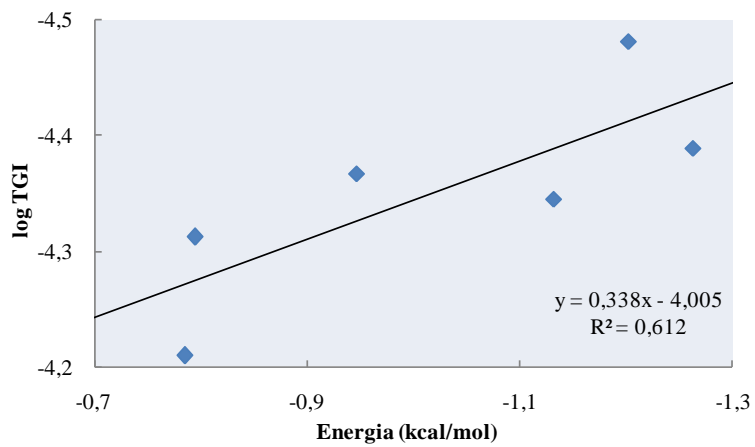
**Figura 51E:** Correlação entre os dados computacionais e experimentais para a enzima Asp para a linha celular SW1573.



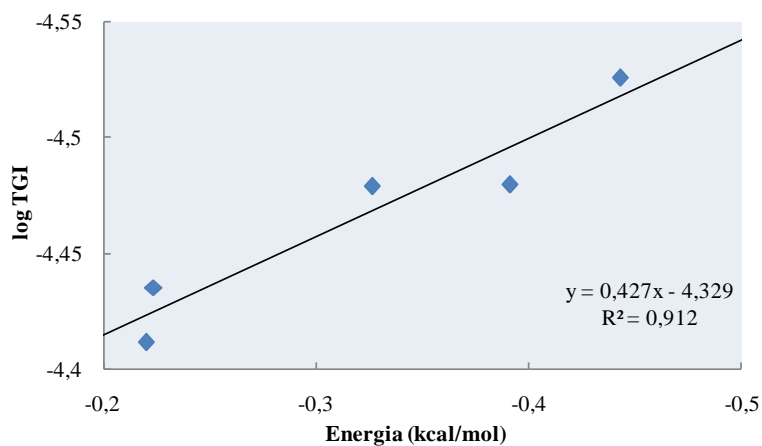
**Figura 52E:** Correlação entre os dados computacionais e experimentais para a enzima Eg5 para a linha celular SW1573.



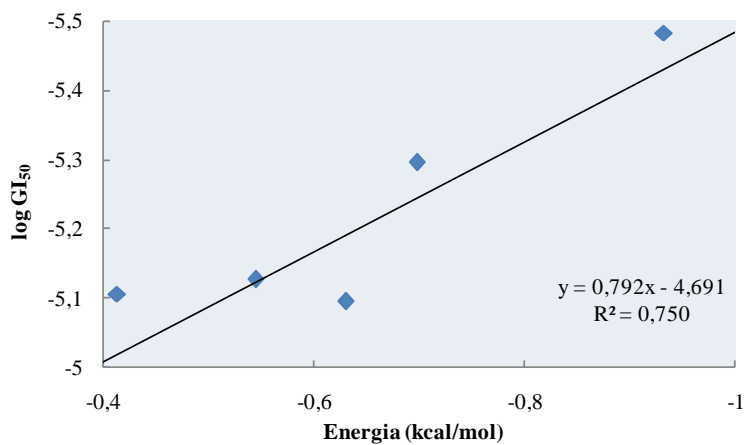
**Figura 53E:** Correlação entre os dados computacionais e experimentais para a enzima AURKA para a linha celular SW1573.



**Figura 54E:** Correlação entre os dados computacionais e experimentais para a enzima PIK3CG para a linha celular SW1573.

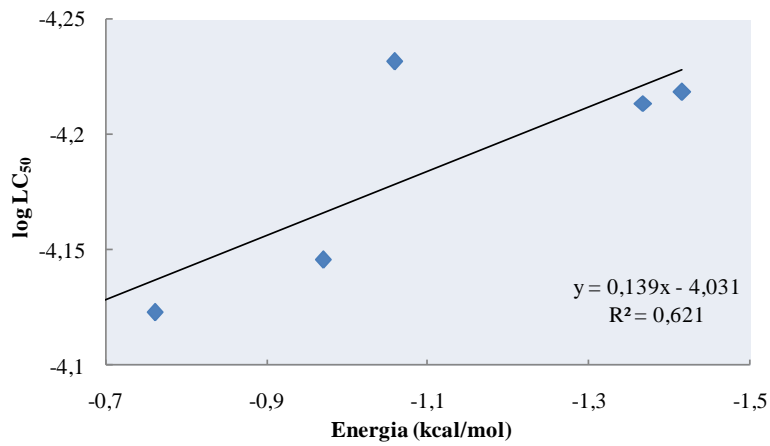


**Figura 55E:** Correlação entre os dados computacionais e experimentais para a enzima PAK1 para a linha celular SW1573.

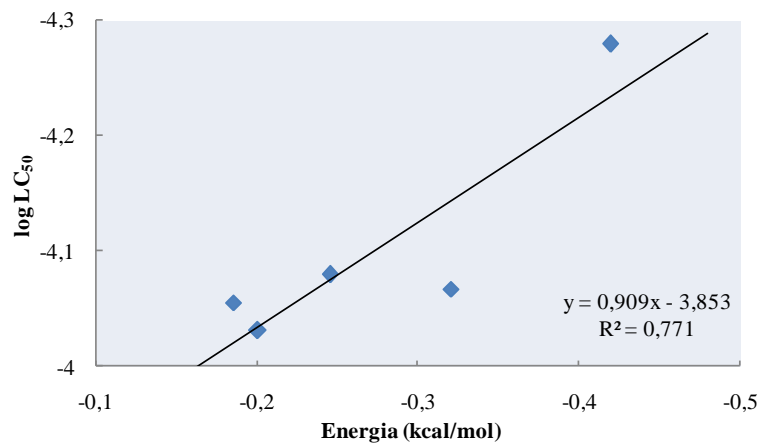


**Figura 56E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular SW1573.

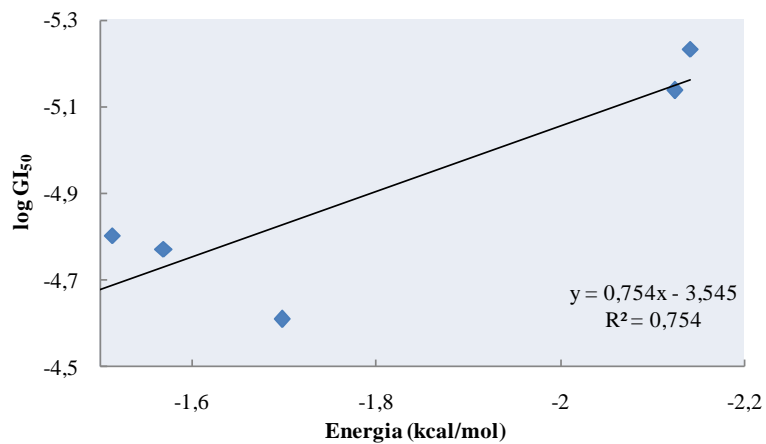
E.4. Linha celular do carcinoma mamário (T-47D)



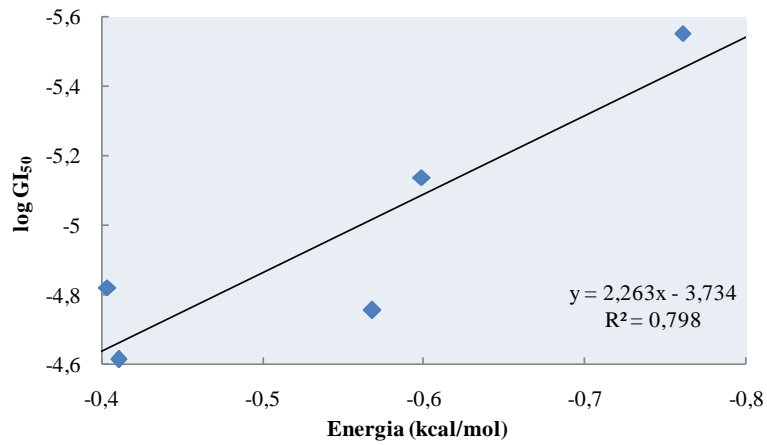
**Figura 57E:** Correlação entre os dados computacionais e experimentais para a enzima CHK2 para a linha celular T-47D.



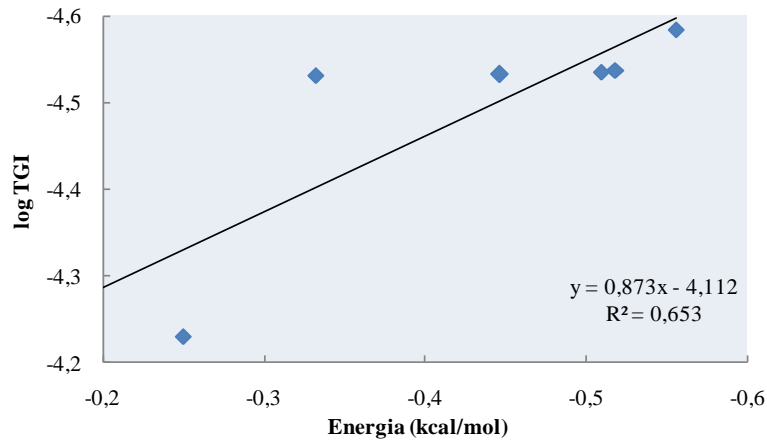
**Figura 58E:** Correlação entre os dados computacionais e experimentais para a enzima CCNH para a linha celular T-47D.



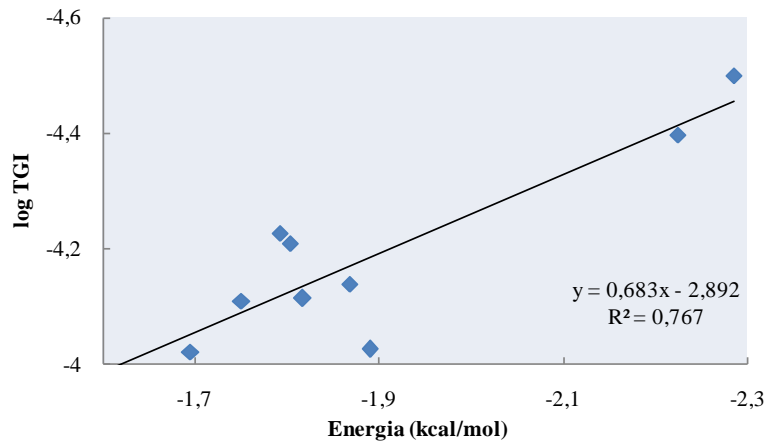
**Figura 59E:** Correlação entre os dados computacionais e experimentais para a enzima MAP2K1 para a linha celular T-47D.



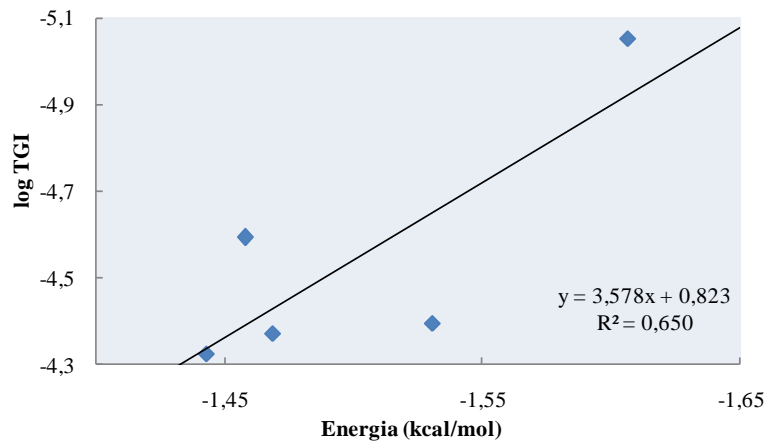
**Figura 60E:** Correlação entre os dados computacionais e experimentais para a proteína p21 CIP1 para a linha celular T-47D.



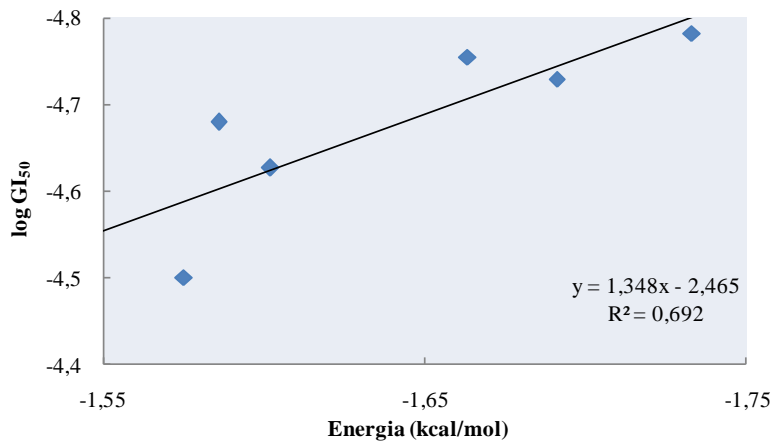
**Figura 61E:** Correlação entre os dados computacionais e experimentais para a proteína KITLG para a linha celular T-47D.



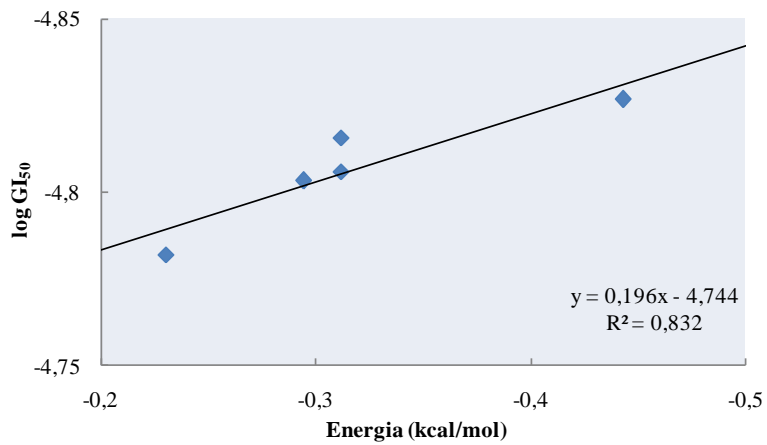
**Figura 62E:** Correlação entre os dados computacionais e experimentais para a enzima CHK1 para a linha celular T-47D.



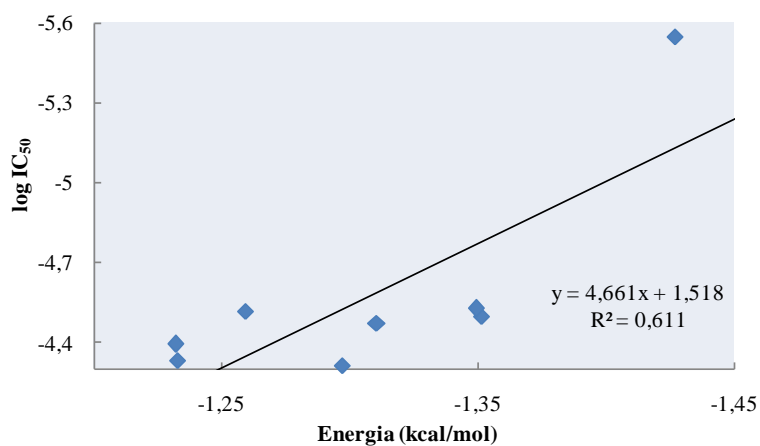
**Figura 63E:** Correlação entre os dados computacionais e experimentais para a enzima TOP1 para a linha celular T-47D.



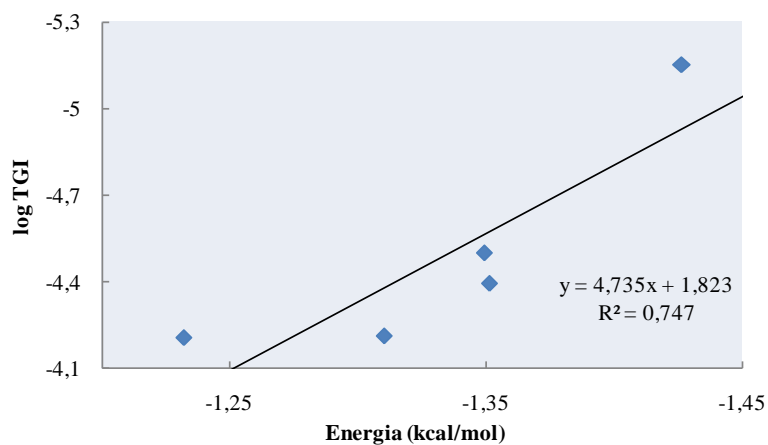
**Figura 64E:** Correlação entre os dados computacionais e experimentais para a enzima AURKA para a linha celular T-47D.



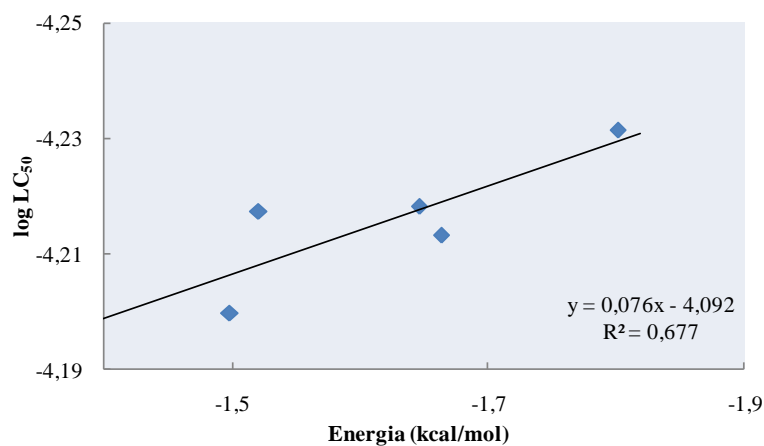
**Figura 65E:** Correlação entre os dados computacionais e experimentais para a proteína Importina B para a linha celular T-47D.



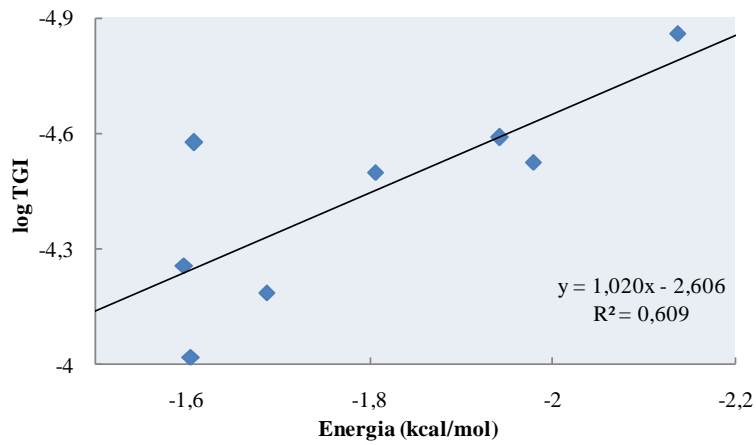
**Figura 66E:** Correlação entre os dados computacionais e experimentais para a enzima COX para a linha celular T-47D.



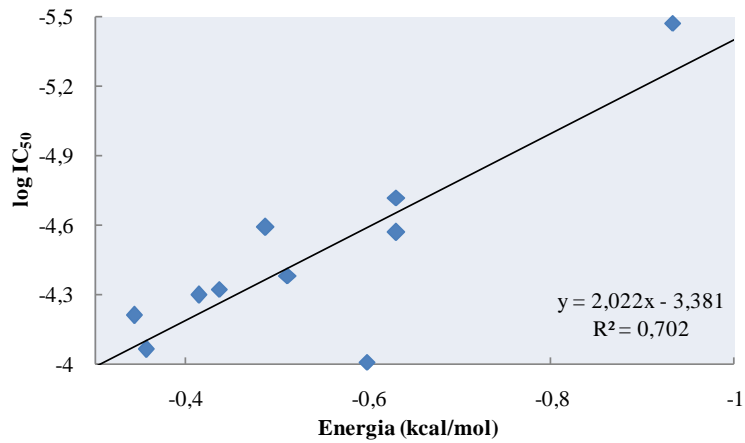
**Figura 67E:** Correlação entre os dados computacionais e experimentais para a enzima COX para a linha celular T-47D.



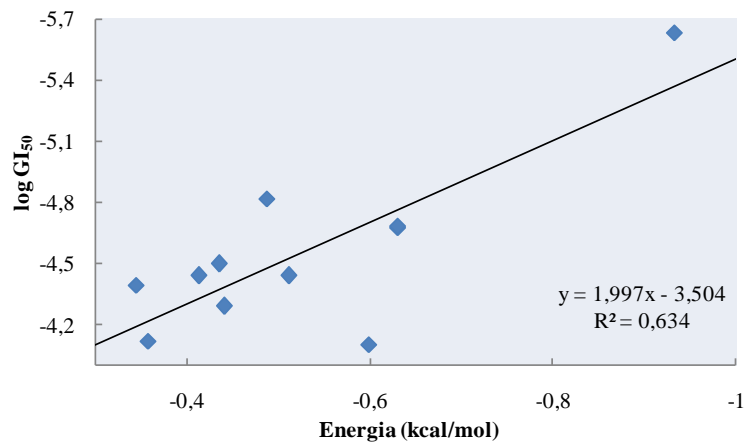
**Figura 68E:** Correlação entre os dados computacionais e experimentais para a enzima mPGES-1-1 para a linha celular T-47D.



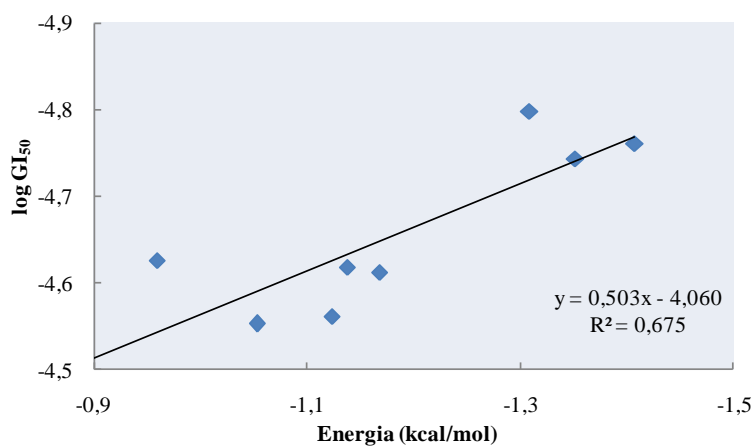
**Figura 69E:** Correlação entre os dados computacionais e experimentais para a enzima 5-LOX para a linha celular T-47D.



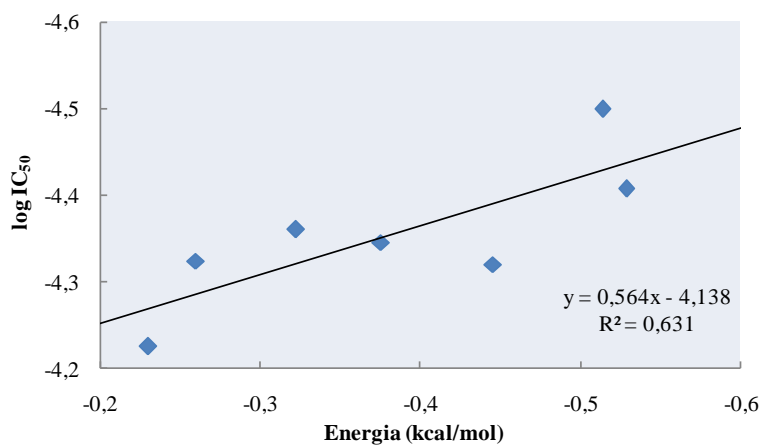
**Figura 70E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular T-47D.



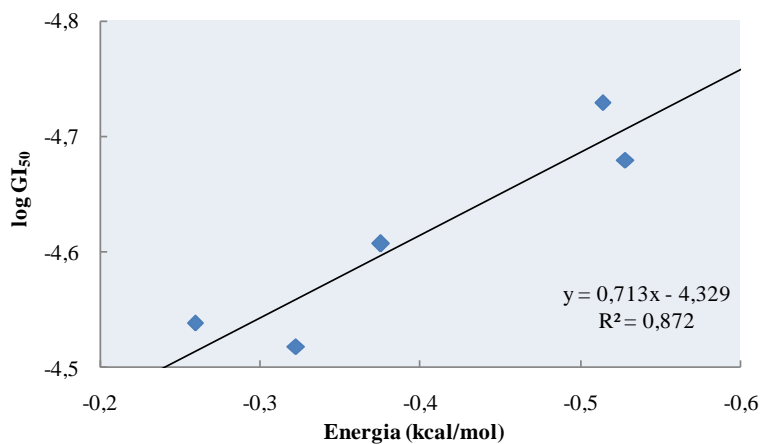
**Figura 71E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular T-47D.



**Figura 72E:** Correlação entre os dados computacionais e experimentais para a enzima NA (complexado com Zanamivir) para a linha celular T-47D.

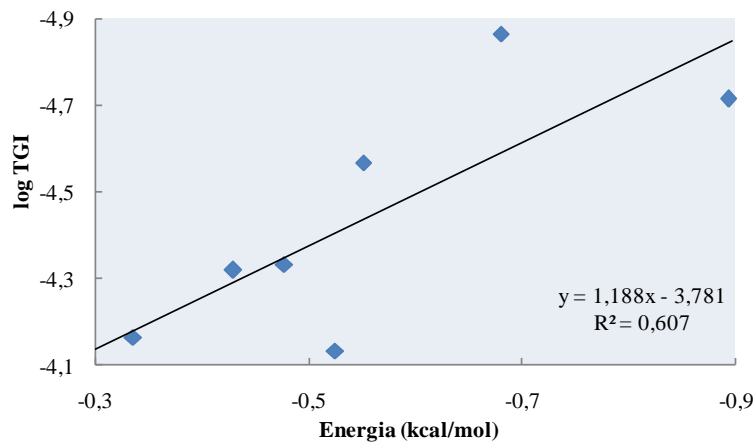


**Figura 73E:** Correlação entre os dados computacionais e experimentais para a enzima NA para a linha celular T-47D.

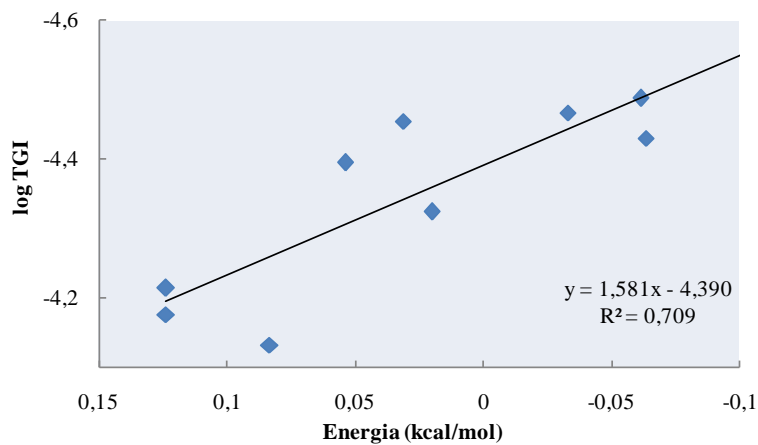


**Figura 74E:** Correlação entre os dados computacionais e experimentais para a enzima NA para a linha celular T-47D.

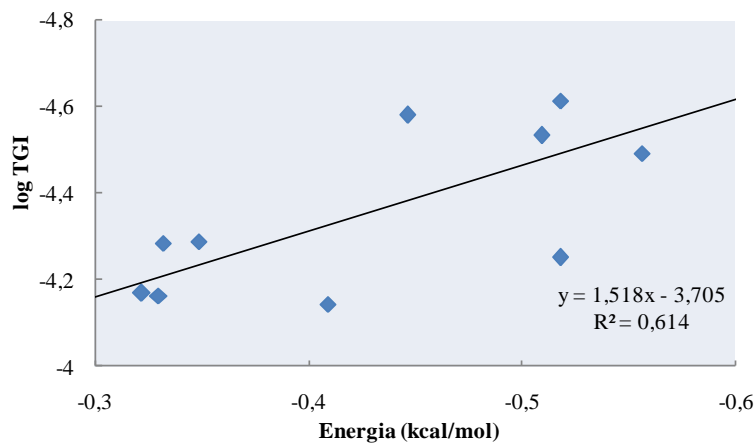
E.5. Linha celular do cancro do cólon (WiDr)



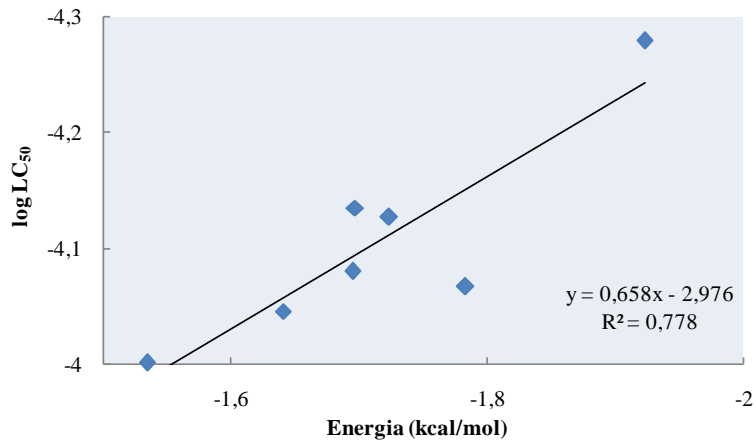
**Figura 75E:** Correlação entre os dados computacionais e experimentais para a enzima CDC7 para a linha celular WiDr.



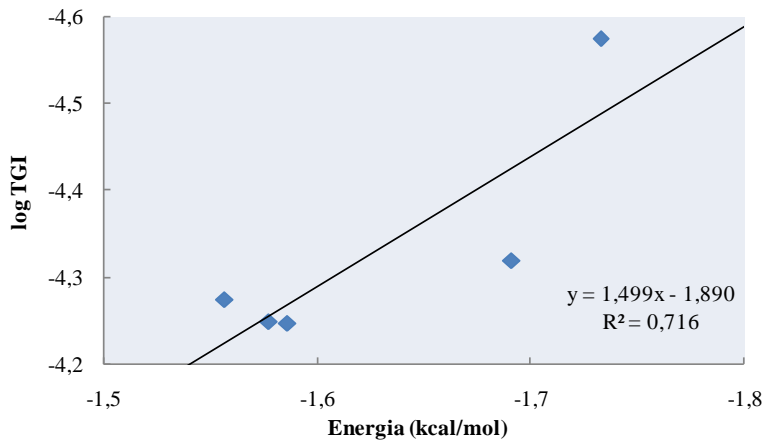
**Figura 76E:** Correlação entre os dados computacionais e experimentais para a proteína p53 para a linha celular WiDr.



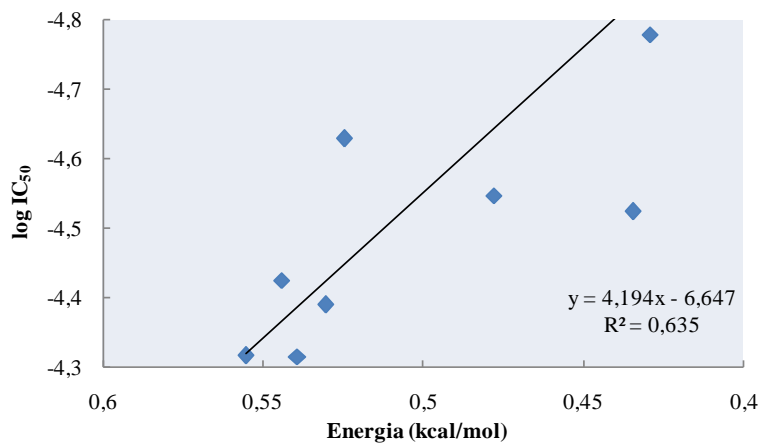
**Figura 77E:** Correlação entre os dados computacionais e experimentais para a proteína KITLG para a linha celular WiDr.



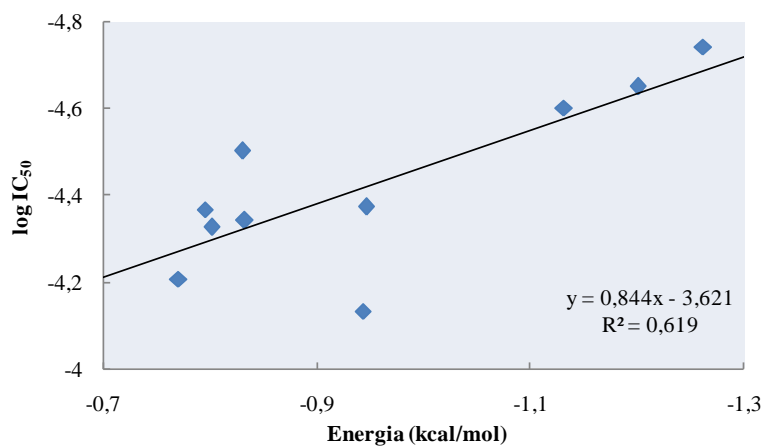
**Figura 78E:** Correlação entre os dados computacionais e experimentais para a enzima WEE1 para a linha celular WiDr.



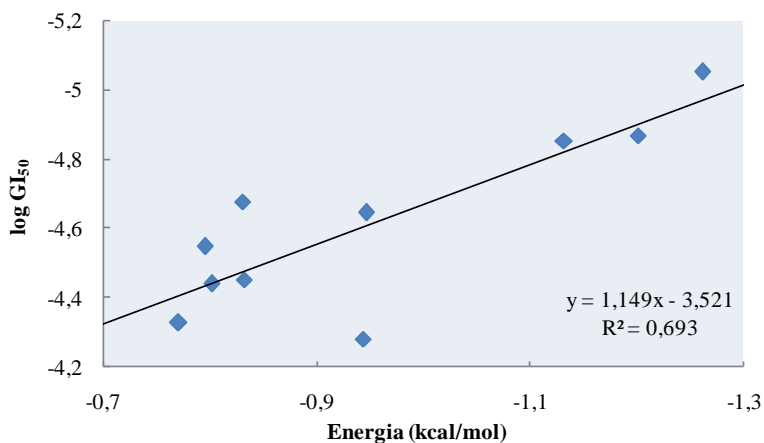
**Figura 79E:** Correlação entre os dados computacionais e experimentais para a enzima AURKA para a linha celular WiDr.



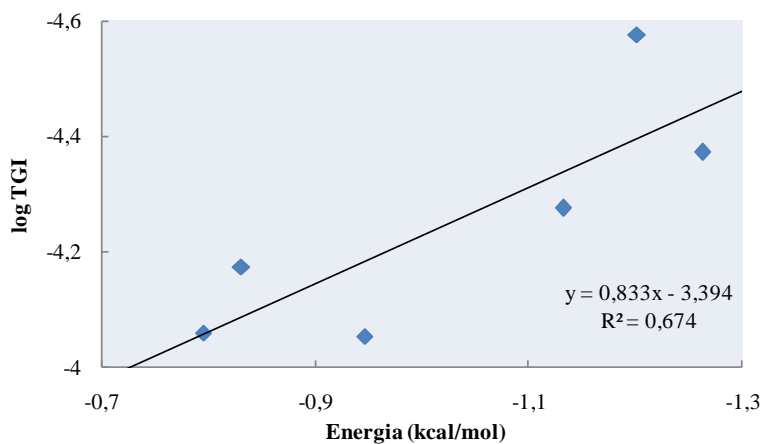
**Figura 80E:** Correlação entre os dados computacionais e experimentais para a enzima ERK1 para a linha celular WiDr.



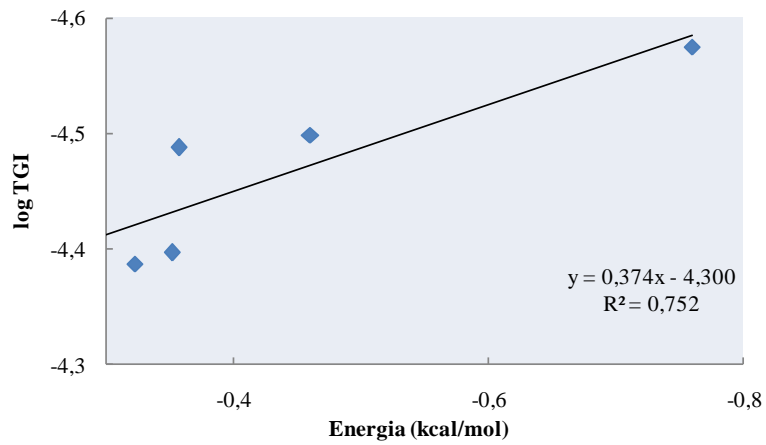
**Figura 81E:** Correlação entre os dados computacionais e experimentais para a enzima PIK3CG para a linha celular WiDr.



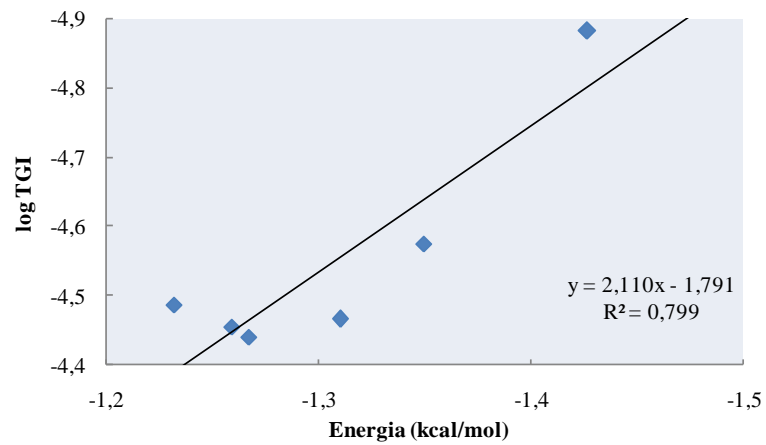
**Figura 82E:** Correlação entre os dados computacionais e experimentais para a enzima PIK3CG para a linha celular WiDr.



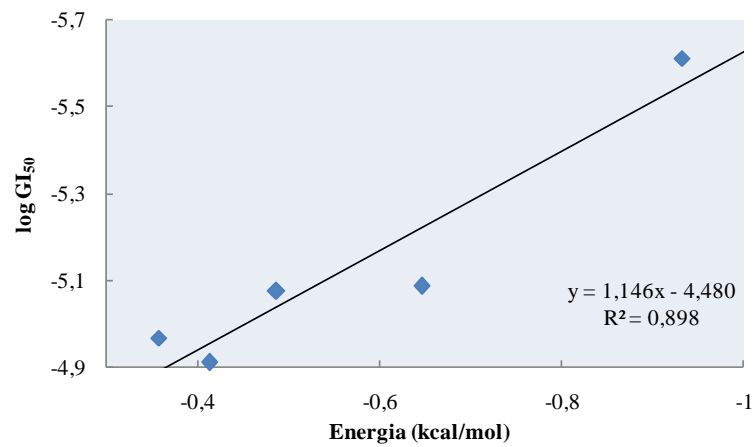
**Figura 83E:** Correlação entre os dados computacionais e experimentais para a enzima PIK3CG para a linha celular WiDr.



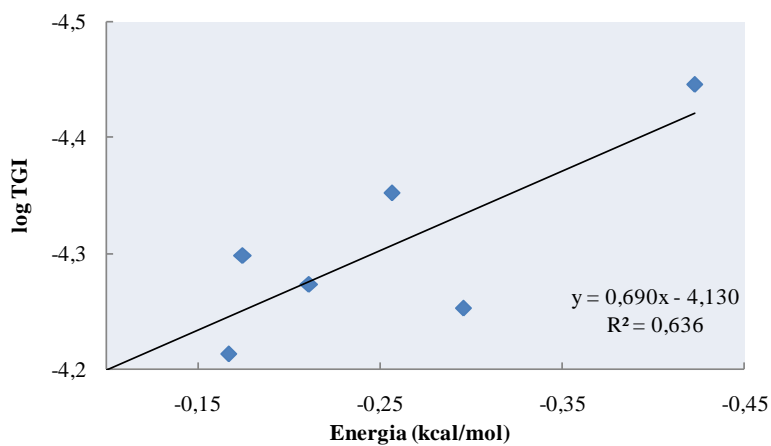
**Figura 84E:** Correlação entre os dados computacionais e experimentais para a proteína p107 para a linha celular WiDr.



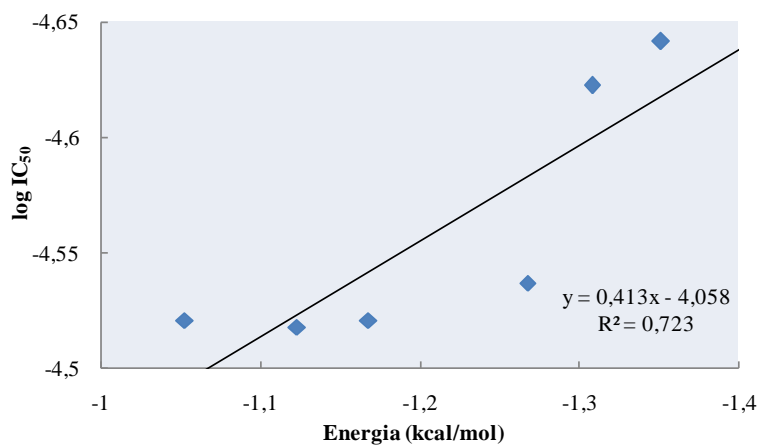
**Figura 85E:** Correlação entre os dados computacionais e experimentais para a enzima COX para a linha celular WiDr.



**Figura 86E:** Correlação entre os dados computacionais e experimentais para a enzima DD-Ligase para a linha celular WiDr.



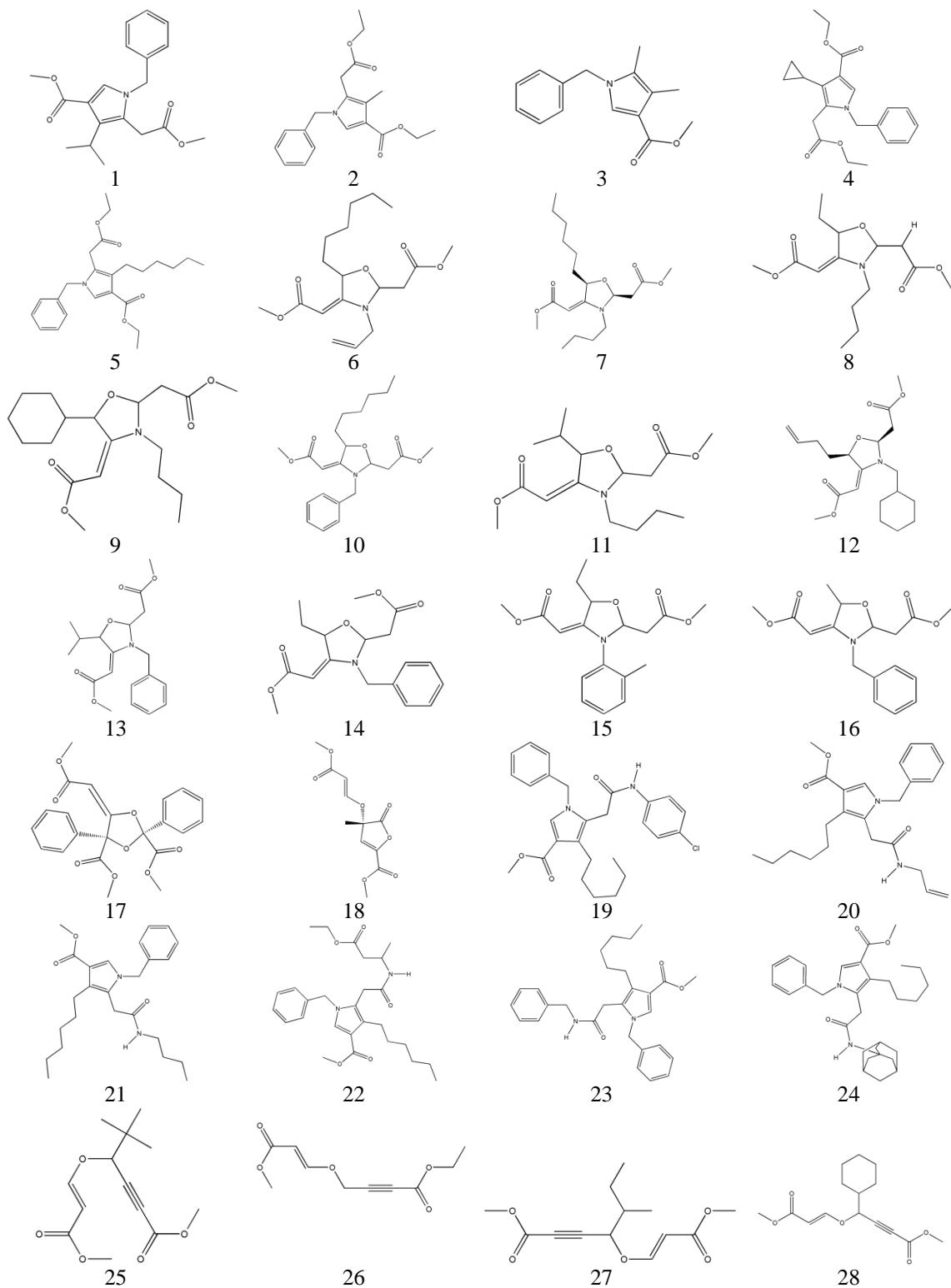
**Figura 87E:** Correlação entre os dados computacionais e experimentais para a enzima EF-G (fungo) para a linha celular WiDr.



**Figura 88E:** Correlação entre os dados computacionais e experimentais para a enzima NA (complexado com Zanamivir) para a linha celular WiDr.

## Anexo F – Relação Quantitativa entre a Estrutura e a Actividade (QSAR) e Redes Neurais Artificiais (ANN)

### F.1. Série 1



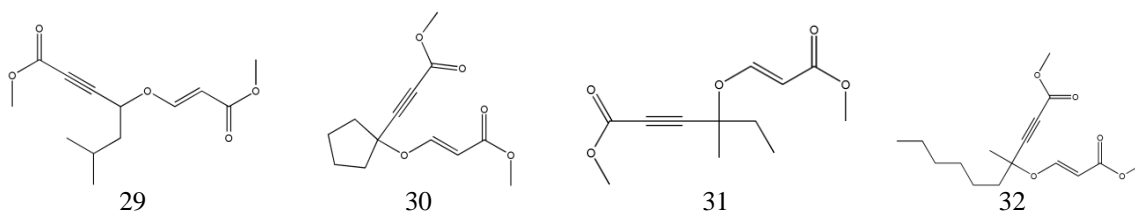


Figura 1F: Estruturas dos compostos que pertencem à série 1.

### F.1.1. QSAR

Tabela 1F: Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do ovário.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.	LC <sub>50</sub> Exp.	LC <sub>50</sub> Prev.	Activ. Biol. Exp.	Activ. Biol. Prev.
5	4,918	4,396	5,088	5,181	4,196	4,169	4,607	4,542
6	4,580	4,446	4,800	4,933	4,370	4,212	4,493	4,472
7	4,590	4,493	4,850	5,113	4,310	4,368	4,505	4,469
9	4,720	4,444	4,940	4,948	4,470	4,343	4,593	4,464
11	4,440	4,462	4,750	4,894	4,060	4,190	4,398	4,498
12	4,580	4,579	4,850	4,653	4,270	4,062	4,498	4,506
13	4,410	4,512	4,690	4,694	4,090	4,081	4,390	4,625
17	4,444	4,765	4,664	5,381	4,073	4,096	4,387	4,257
18	4,982	4,975	5,147	5,348	4,266	4,044	4,721	4,651
19	5,147	4,913	5,287	5,442	4,000	4,050	4,775	4,831
20	4,260	4,471	4,558	4,406	4,000	4,022	4,260	4,550
21	4,247	4,436	4,576	4,367	4,000	4,101	4,272	4,386
22	4,806	4,806	5,094	5,149	4,000	3,965	4,543	4,486
23	4,277	4,495	4,390	4,609	4,000	3,994	4,244	4,413
24	4,880	4,473	5,194	4,674	4,000	4,129	4,611	4,350
25	5,551	5,838	5,759	6,153	4,222*	4,811	5,128	6,098
26	5,565	5,568	5,812	5,719	4,571	4,771	5,317	5,309
27	7,984	5,932	7,231	6,604	4,712	4,766	6,541	6,082
28	5,707	5,770	5,869	6,255	4,846	4,750	5,412	5,657
29	6,458	5,988	6,639	6,615	4,865	4,728	5,980	6,081
30	5,997	5,917	6,239	6,007	4,566	4,656	5,721	5,871
31	6,356	5,944	6,469	6,282	4,621	4,662	5,884	6,143
32	5,609	5,866	5,803	7,059	4,874	4,776	5,415	5,723

\*Composto eliminado

**Tabela 2F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do pulmão.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.	Activ. Biol. Exp.	Activ. Biol. Prev.
5	4,635	4,405	4,821	4,873	4,364	4,096
6	4,299	4,228	4,501	4,142	4,200	4,084
7	4,334	4,224	4,539	3,359	4,218	4,327
8	4,000	4,321	4,000	4,309	4,000	4,163
9	4,618	4,405	4,841	3,370	4,458	4,204
10	4,000	4,271	4,207	4,736	4,052	4,102
11	4,095	4,337	4,327	4,249	4,106	4,168
12	4,302	4,247	4,543	4,803	4,211	3,985
13	4,157	4,336	4,328	4,241	4,121	4,050
14	4,000	4,284	4,000	4,403	4,000	4,022
15	5,092	4,346	5,316	4,734	4,718	3,992
16	4,000	4,236	4,000	4,349	4,000	4,049
17	4,459	4,747	4,633	4,254	4,319	4,256
18	4,752	4,787	4,960	4,980	4,539	4,575
19	4,771	4,539	5,048	3,823	4,455	4,590
20	4,033	4,147	4,287	4,333	4,080	4,120
21	4,000	4,087	4,059	4,440	4,015	4,180
22	4,699	4,593	4,929	4,610	4,407	4,397
23	4,000	4,182	4,000	4,312	4,000	4,227
24	4,576	4,142	4,890	4,618	4,367	4,287
25	5,352	5,891	5,527	5,924	4,903	5,247
26	5,532	5,701	5,765	6,141	5,241	5,357
27	5,883	5,661	6,141	6,034	5,330	5,245
28	5,469	5,490	5,685	5,909	5,099	5,267
29	6,263	5,568	6,454	6,052	5,654	5,313
30	6,140	5,646	6,350	6,435	5,595	5,380
31	6,243	5,750	6,514	6,264	5,540	5,256
32	5,486	5,481	5,698	5,145	5,119	5,556

**Tabela 3F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do carcinoma mamário.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.	TGI Exp.	TGI Prev.	Activ. Biol. Exp.	Activ. Biol. Prev.
1	4,267	4,383	4,061	3,979	4,231	4,096
2	4,159	4,034	4,109	3,854	4,191	3,997
3	4,144	3,914	4,000	4,074	4,140	4,145
4	4,278	4,422	4,073	4,083	4,232	4,173
5	4,625	4,102	4,257	3,960	4,457	4,118
6	4,189	4,048	4,065	4,061	4,196	4,170
7	4,204	4,079	4,016	4,096	4,190	4,209
8	4,000	4,149	4,000	4,099	4,000	4,176

9	4,316	4,192	4,114	4,059	4,255	4,174
10	4,235	4,142	4,000	4,104	4,145	4,212
11	4,096	4,114	4,000	4,134	4,127	4,218
12	4,072	4,192	4,008	4,101	4,137	4,192
13	4,056	4,074	4,000	3,975	4,081	4,103
14	4,000	4,149	4,000	4,023	4,000	4,129
15	4,407	4,334	4,271	4,186	4,360	4,235
16	4,000	4,327	4,000	4,097	4,000	4,173
25	4,678	5,062	4,254	4,587	4,469	4,764
27	5,272	5,326	4,653	4,705	4,886	4,852
29	5,082	5,310	4,686	4,717	4,777	4,862
30	5,599	5,719	5,372	5,450	5,385	5,439
31	5,691	5,300	5,383	4,906	5,395	5,016
32	5,886	5,632	4,883	4,772	5,351	4,889

### F.1.2. ANN

**Tabela 4F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do ovário.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.	LC <sub>50</sub> Exp.	LC <sub>50</sub> Prev.	Activ. Biol. Exp.	Activ. Biol. Prev.
5	4,918	4,453	4,196	4,268	4,607	4,457
6	4,580	4,437	4,370	4,224	4,493	4,472
7	4,590	4,634	4,310	4,258	4,505	4,501
9	4,720	4,525	4,470	4,256	4,593	4,502
11	4,440	4,217	4,060	4,099	4,398	4,457
12	4,580	4,703	4,270	4,317	4,498	4,455
13	4,410	4,644	4,090	4,218	4,390	4,458
17	4,444	4,458	4,073	4,058	4,387	4,514
18	4,982	5,027	4,266	4,281	4,721	4,710
19	5,147	4,809	4,000	4,004	4,775	4,466
20	4,260	4,327	4,000	3,994	4,260	4,449
21	4,247	4,500	4,000	3,953	4,272	4,468
22	4,806	4,823	4,000	3,982	4,543	4,568
23	4,277	4,531	4,000	4,016	4,244	4,474
24	4,880	4,801	4,000	4,001	4,611	4,556
25	5,551	5,967	---	---	5,128	5,990
26	5,565	5,622	4,571	4,559	5,317	5,296
27	7,984	5,965	4,712	4,709	6,541	6,010
28	5,707	5,847	4,846	4,854	5,412	5,575
29	6,458	5,964	4,865	4,701	5,980	6,031
30	5,997	5,920	4,566	4,719	5,721	6,061
31	6,356	5,922	4,621	4,665	5,884	6,049
32	5,609	5,583	4,874	4,839	5,415	5,363

**Tabela 5F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do pulmão.

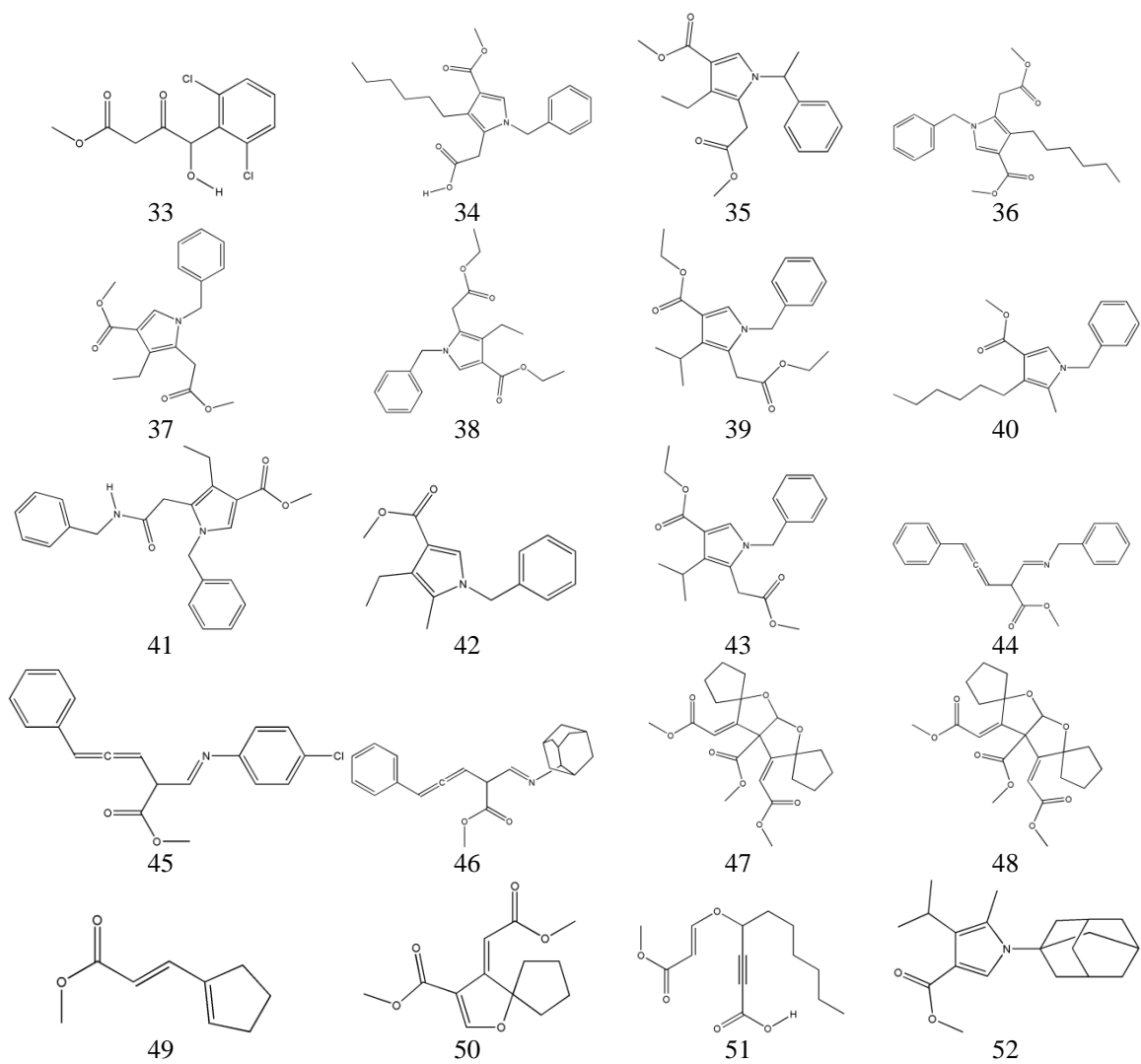
No. Comp.	GI <sub>50</sub> Exp	GI <sub>50</sub> Prev	Activ. Biol. Exp.	Activ. Biol. Prev.
5	4,821	4,773	4,364	4,137
6	4,501	4,190	4,200	4,200
7	4,539	4,113	4,218	4,312
8	4,000	4,436	4,000	4,100
9	4,841	4,343	4,458	4,280
10	4,207	4,368	4,052	4,189
11	4,327	4,318	4,106	4,137
12	4,543	4,383	4,211	4,086
13	4,328	4,433	4,121	4,086
14	4,000	4,514	4,000	4,056
15	5,316	4,556	4,718	4,060
16	4,000	4,620	4,000	4,033
17	4,633	4,739	4,319	4,319
18	4,960	4,887	4,539	4,566
19	5,048	4,129	4,455	3,922
20	4,287	4,190	4,080	3,971
21	4,059	4,072	4,015	4,116
22	4,929	4,926	4,407	4,407
23	4,000	4,214	4,000	4,030
24	4,890	4,821	4,367	4,324
25	5,527	6,148	4,903	5,294
26	5,765	5,961	5,241	5,377
27	6,141	6,122	5,330	5,300
28	5,685	5,576	5,099	5,237
29	6,454	6,099	5,654	5,292
30	6,350	5,649	5,595	5,328
31	6,514	6,203	5,540	5,332
32	5,698	5,713	5,119	5,191

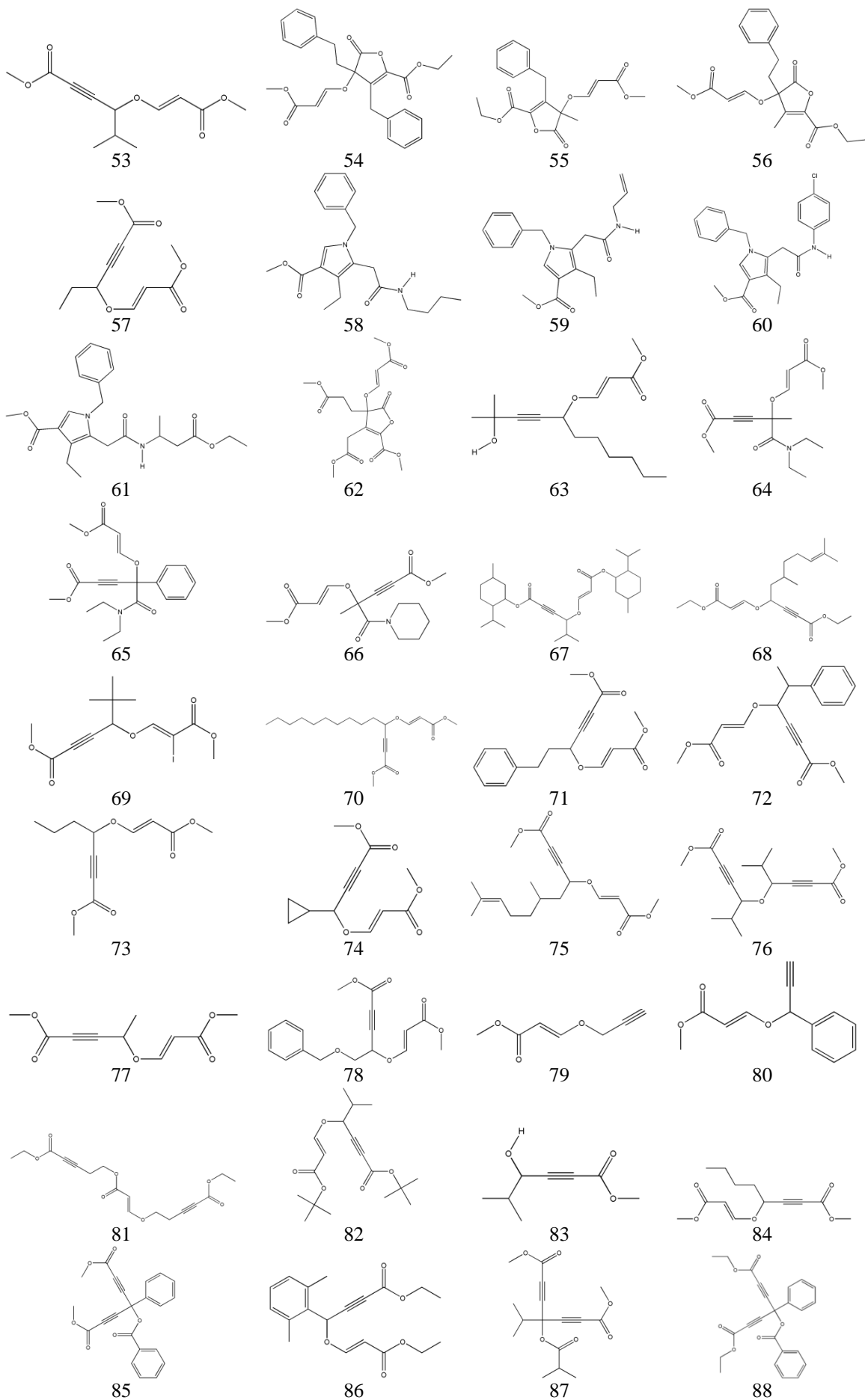
**Tabela 6F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do carcinoma mamário.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.
1	4,267	4,128
2	4,159	4,223
3	4,144	4,134
4	4,278	4,239
5	4,625	4,249
6	4,189	4,114
7	4,204	4,115
8	4,000	4,115

9	4,316	4,120
10	4,235	4,233
11	4,096	4,117
12	4,072	4,206
13	4,056	4,115
14	4,000	4,117
15	4,407	4,332
16	4,000	4,115
25	4,678	5,322
27	5,272	5,322
29	5,082	5,322
30	5,599	5,322
31	5,691	5,322
32	5,886	5,322

## F.2. Série 7





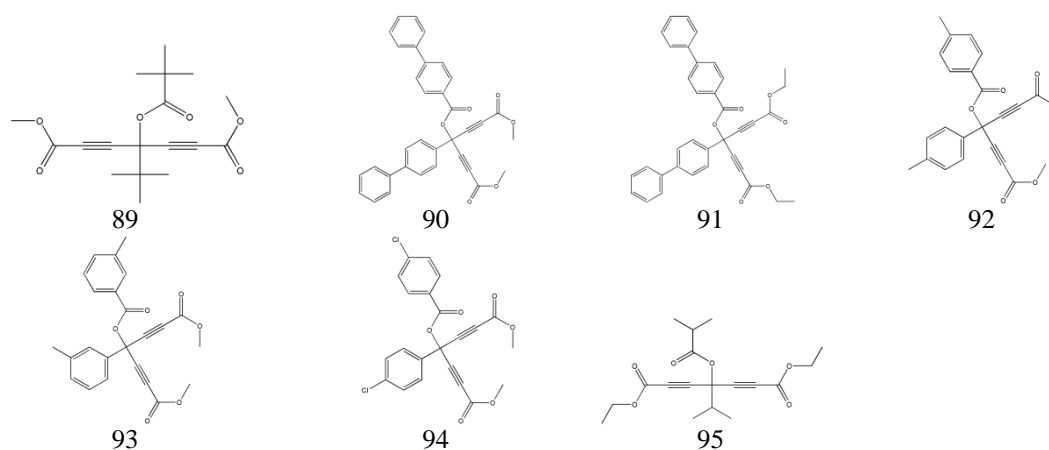


Figura 2F: Estruturas dos compostos que pertencem à série 7.

### F.2.1. QSAR

Tabela 7F: Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro da leucemia.

No. Comp.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.	Activ. Biol. Exp.	Activ. Biol. Prev.
33	4,000	4,166	4,000	4,023
34	4,402	4,368	4,186	4,233
35	4,418	4,546	4,187	4,241
36	4,787	4,624	4,334	4,306
37	4,370	4,384	4,162	4,188
38	4,473	4,712	4,213	4,305
39	4,779	4,808	4,357	4,341
40	4,594	4,731	4,278	4,359
41	5,497	5,247	4,672	4,567
42	4,836	4,628	4,358	4,305
43	4,526	4,747	4,237	4,318
44	4,770	4,962	4,320	4,410
45	5,235	5,002	4,660	4,386
46	5,253	5,281	4,550	4,499
47	4,535	4,627	4,244	4,239
48	4,631	4,624	4,278	4,238
49	4,732	4,638	4,374	4,300
50	4,000	4,170	4,000	4,029
51	4,000	3,800	4,000	3,917

**Tabela 8F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do ovário.

No. Comp.	Activ. Biol. Exp.	Activ. Biol. Prev.	No. Comp.	Activ. Biol. Exp.	Activ. Biol. Prev.
41	4,575	4,215	73	6,122	5,598
51	4,000	4,231	74	5,977	5,655
52	4,276	4,140	75	5,495	5,107
53	5,697	5,697	76	5,336	4,805
54	4,361	4,229	77	5,918	5,779
55	4,394	4,206	78	5,144	5,546
56	4,522	4,628	79	4,000	4,112
57	5,801	5,804	80	4,224	4,355
58	4,543	4,355	81	5,103	5,108
59	4,069	4,249	82	4,128	4,756
60	4,376	4,588	83	4,000	3,826
61	4,225	4,719	84	5,179	5,500
62	4,200	4,194	85	5,226	5,193
63	4,397	4,600	86	5,312	5,264
64	5,125	5,305	87	5,278	5,442
65	5,348	5,172	88	5,417	5,206
66	5,241	5,257	89	5,303	5,077
67	4,000	4,307	90	4,685	4,648
68	4,434	5,035	91	4,125	4,698
69	4,629	4,740	92	5,348	5,276
70	5,628	4,966	93	5,110	5,125
71	5,481	5,320	94	5,520	5,573
72	5,470	5,395	95	5,400	5,226

**F.2.2. ANN****Tabela 9F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro da leucemia.

No. Comp.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.	Activ. Biol. Exp.	Activ. Biol. Prev.
33	4,000	3,991	4,000	3,996
34	4,402	4,618	4,186	4,244
35	4,418	4,389	4,187	4,176
36	4,787	4,660	4,334	4,332
37	4,370	4,369	4,162	4,167
38	4,473	4,531	4,213	4,197
39	4,779	4,542	4,357	4,244
40	4,594	4,564	4,278	4,274
41	5,497	5,434	4,672	4,667

---

42	4,836	4,749	4,358	4,373
43	4,526	4,471	4,237	4,197
44	4,770	4,974	4,320	4,305
45	5,235	5,208	4,660	4,405
46	5,253	5,226	4,550	4,556
47	4,535	4,601	4,244	4,260
48	4,631	4,601	4,278	4,260
49	4,732	4,779	4,374	4,370
50	4,000	4,285	4,000	3,826
51	4,000	4,030	4,000	4,008

---

## F.3. Série 9

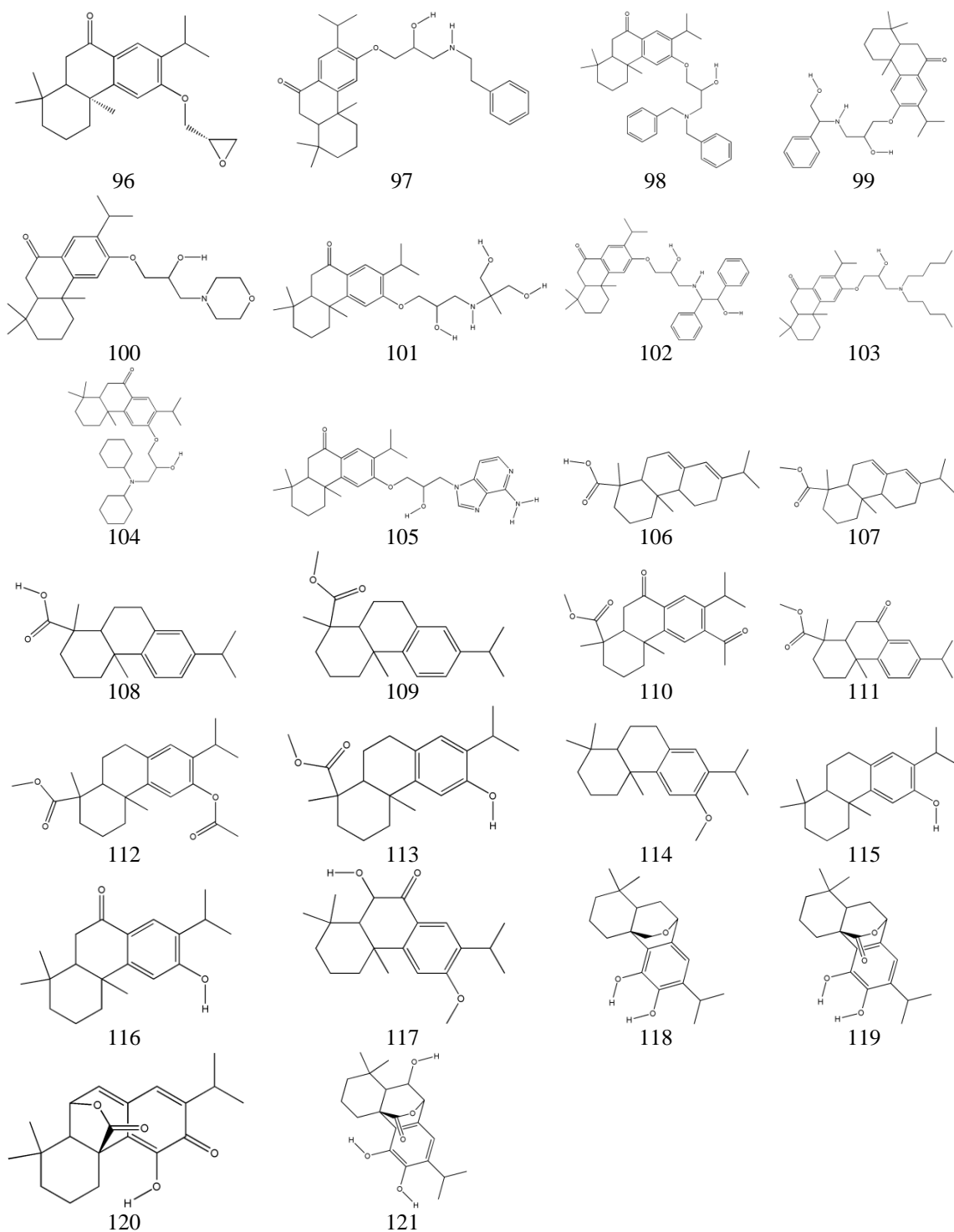


Figura 3F: Estruturas dos compostos que pertencem à série 9.

**F.3.1. QSAR****Tabela 10F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do ovário.

<b>No. Comp.</b>	<b>Activ. Biol. Exp.</b>	<b>Activ. Biol. Prev.</b>
96	4,133	4,357
97	4,310	5,500
98	4,000	4,522
100	4,825	5,778
102	6,082	5,605
104	5,980	5,498
105	4,113	4,466
108	4,889	4,417
109	4,458	4,867
110	5,984	6,160
111	5,898	5,709
112	6,493	6,610
113	5,220	5,276
114	4,235	4,345
115	4,229	4,061
116	4,323	4,599
117	6,007	6,127
118	4,165	3,688
119	4,300	4,446
121	4,966	5,121

**Tabela 11F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do pulmão.

<b>No. Comp.</b>	<b>IC<sub>50</sub> Exp.</b>	<b>IC<sub>50</sub> Prev.</b>	<b>LC<sub>50</sub> Exp.</b>	<b>LC<sub>50</sub> Prev.</b>	<b>Activ. Biol. Exp.</b>	<b>Activ. Biol. Prev.</b>
96	4,535	4,494	4,034	4,084	4,420	4,552
97	5,490	5,343	4,000	4,065	4,833	4,719
98	4,000*	5,853	4,000	4,011	4,010*	4,617
99	5,287	5,407	4,009	4,020	4,749	4,812
100	4,572	4,576	4,000	4,067	4,399	4,748
101	4,919	4,764	4,000	3,989	4,631	4,584
102	5,094	6,294	4,000	4,004	4,699	4,637
103	4,763	4,917	4,000	4,016	4,536	4,651
104	4,734	4,775	4,070	3,983	4,526	4,485
105	4,468	4,996	4,000	3,985	4,305	4,378
106	4,052	4,527	4,102	4,041	4,200	4,195
107	4,594	4,427	4,000	4,042	4,352	4,279
108	4,265	4,532	4,065	4,074	4,291	4,304
109	4,464	4,429	4,070	4,181	4,381	4,625

110	4,326	4,204	4,113	4,083	4,523	4,476
111	4,231	4,384	4,039	4,092	4,207	4,430
112	4,585	4,614	4,591	4,568	4,768	4,732
113	4,602	4,387	4,524*	4,092	4,756*	4,580
114	4,471	4,405	4,000	4,030	4,283	4,428
115	4,500	4,493	4,038	4,044	4,395	4,309
116	4,184	4,275	4,000	4,014	4,167	4,166
117	4,262	4,284	4,000	4,021	4,194	4,464
118	4,484	4,555	4,000	3,922	4,289	4,305
119	4,966	4,732	4,326	4,202	4,770	4,666
120	4,542	4,533	4,107	4,093	4,458	4,491
121	4,451	4,505	4,000	4,045	4,277	4,226

**Tabela 12F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do cólon.

No. Comp.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.
96	4,906	4,970
97	4,000	4,208
98	4,021	4,040
99	4,504	4,609
100	5,385	5,675
101	5,409	5,263
102	4,594	4,500
103	5,591	5,397
104	5,319	5,398
105	4,882	4,797
106	4,907	4,957
107	4,000	4,943
108	4,752	4,675
109	5,000	4,736
110	5,462	5,261
111	4,720	4,612
112	5,714	5,591
113	5,613	5,092
114	4,603	4,980
115	5,062	4,882
117	4,426*	5,404
118	4,824	4,926
119	5,288	4,976
120	4,793	4,914
121	5,086	5,242

\*Composto eliminado

## F.3.2. ANN

**Tabela 13F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do ovário.

No. Comp.	Activ. Biol. Exp.	Activ. Biol. Prev.
96	4,133	4,001
97	4,310	5,250
98	4,000	3,970
100	4,825	5,832
102	6,082	5,060
104	5,980	5,865
105	4,113	4,054
108	4,889	4,737
109	4,458	5,110
110	5,984	5,893
111	5,898	5,624
112	6,493	6,621
113	5,220	5,595
114	4,235	4,460
115	4,229	4,258
116	4,323	4,545
117	6,007	6,002
118	4,165	4,134
119	4,300	4,485
121	4,966	4,837

**Tabela 14F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do pulmão.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.	LC <sub>50</sub> Exp.	LC <sub>50</sub> Prev.
96	4,535	4,424	4,034	3,974
97	5,490	5,443	4,000	3,975
99	5,287	5,329	4,000	4,021
100	4,572	4,659	4,009	3,989
101	4,919	4,882	4,000	3,972
102	5,094	6,341	4,000	4,001
103	4,763	4,817	4,000	4,017
104	4,734	4,737	4,000	4,019
105	4,468	4,917	4,070	4,013
106	4,052	4,450	4,000	3,985
107	4,594	4,506	4,102	4,062
108	4,265	4,382	4,000	4,049
109	4,464	4,465	4,065	4,060
110	4,326	4,423	4,070	4,133
111	4,231	4,358	4,113	4,108

112	4,585	4,457	4,039	4,010
113	4,602	4,456	4,591	4,570
114	4,471	4,548	4,000	4,006
115	4,500	4,530	4,038	4,021
116	4,184	4,375	4,000	4,027
117	4,262	4,482	4,000	4,003
118	4,484	4,558	4,000	4,003
119	4,966	4,490	4,326	4,331
120	4,542	4,409	4,107	4,043
121	4,451	4,573	4,000	4,017

**Tabela 15F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do cólon.

No. Comp.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.
96	4,906	5,039
97	4,000	3,941
98	4,021	4,011
99	4,504	4,554
100	5,385	5,402
101	5,409	5,407
102	4,594	4,617
103	5,591	5,572
104	5,319	5,324
105	4,882	4,877
106	4,907	4,902
107	4,000	4,333
108	4,752	4,901
109	5,000	4,849
110	5,462	5,319
111	4,720	4,814
112	5,714	5,722
113	5,613	5,376
114	4,603	4,677
115	5,062	4,901
118	4,824	5,610
119	5,288	6,294
120	4,793	4,786
121	5,086	5,088

## F.4. Série 12

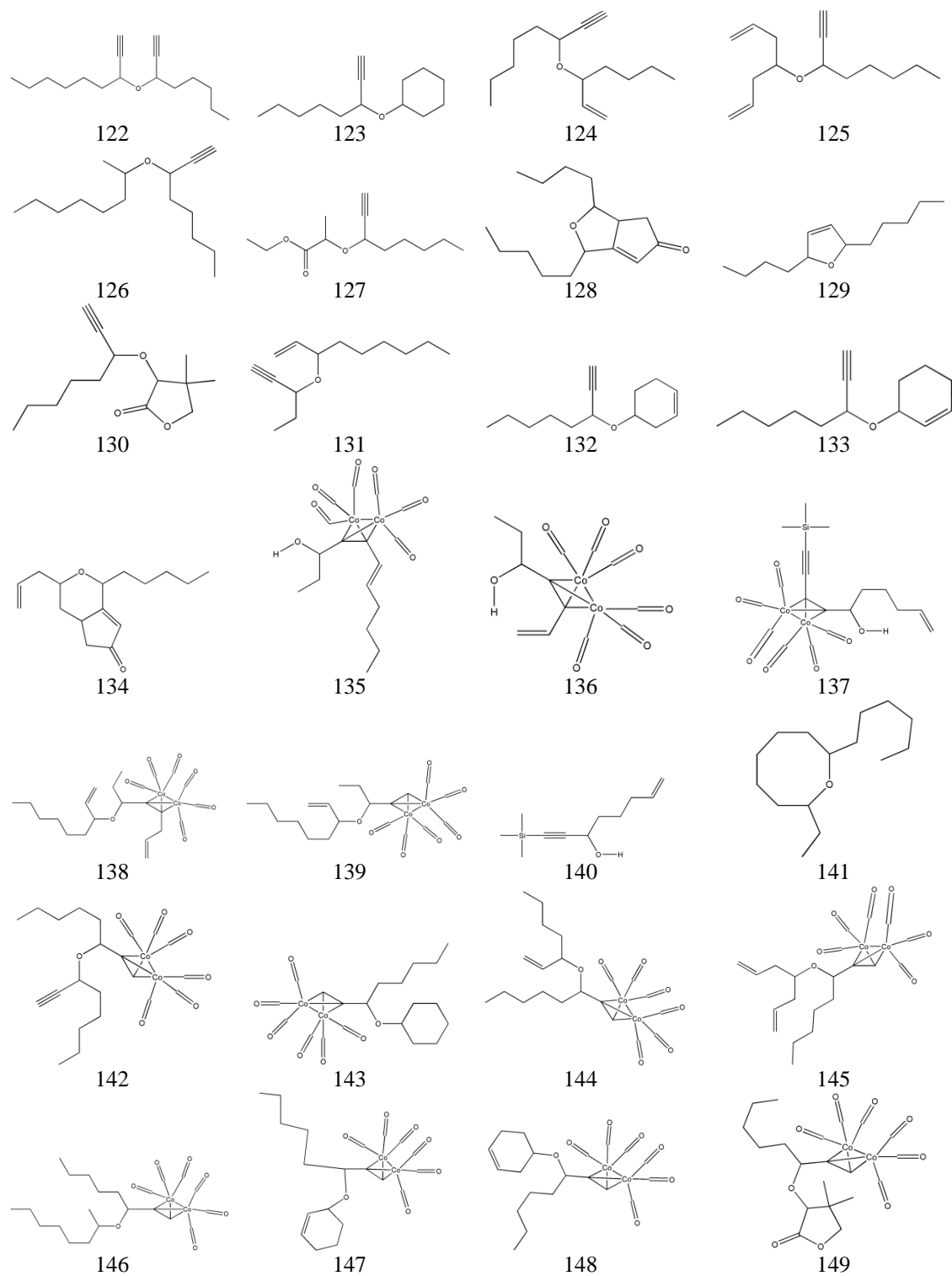


Figura 4F: Estruturas dos compostos que pertencem à série 12.

**F.4.1. QSAR****Tabela 16F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do cancro do ovário.

No. Comp.	GI <sub>50</sub> Exp.	GI <sub>50</sub> Prev.
122	4,373	4,498
123	4,372	4,496
124	4,833	4,834
125	4,539	5,063
126	4,705	4,357
127	4,000	4,400
128	6,427	6,499
129	5,439	5,088
130	4,000	4,403
131	5,556	4,824
133	4,687	4,561
134	6,977	6,909
142	4,000	3,922
143	4,000	3,922
144	4,527	4,223
145	5,364	4,524
146	4,000	3,922
147	4,000	4,223
148	4,000	4,223
149	4,000	4,223

**Tabela 17F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do carcinoma mamário.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.
123	4,000	4,009
124	4,000	4,046
125	4,000	4,007
126	4,015	4,160
129	4,000	4,127
131	4,000	3,993
132	4,000	3,965
133	4,147	3,960
135	4,617	4,713
136	4,470	4,461
137	4,701	4,729
138	4,388	4,265
139	4,500	4,120
140	4,000	3,866

---

141	4,555	4,405
144	4,167	4,183
147	4,140	4,084
148	4,000	4,134
149	4,000	1,453

---

#### F.4.2. ANN

**Tabela 18F:** Valores experimentais e previstos das actividades biológicas obtidas em relação à linha do carcinoma mamário.

No. Comp.	IC <sub>50</sub> Exp.	IC <sub>50</sub> Prev.
123	4,000	3,985
124	4,000	4,006
125	4,000	4,001
126	4,015	4,045
129	4,000	4,016
131	4,000	3,983
132	4,000	4,028
133	4,147	3,993
135	4,617	4,587
136	4,470	4,593
137	4,701	4,573
138	4,388	4,180
139	4,500	4,236
140	4,000	3,982
141	4,555	4,530
144	4,167	4,146
147	4,140	4,141
148	4,000	4,226
149	4,000	2,876

## F.5. Tabelas das Equações de QSAR

### F.5.1. Série 1

Tabela 19F: Modelos de QSAR para a linha celular do cancro do ovário.

<i>pIC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	2,864	0,555	5,165	
NRLT	Número relativo de ligações triplas	38,381	5,601	6,853	3,484
I <sub>A</sub>	Momento de inércia A	-8,480	3,003	-2,824	1,103
DCPAS	Diferença da carga parcial das áreas superficiais	7,720x10 <sup>-2</sup>	2,337x10 <sup>-2</sup>	3,304	1,841
<i>pGI<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	7,000	0,859	8,147	
NRAN	Número relativo de átomos de azoto	-58,759	7,940	-7,400	3,068
CAPNS	Contribuição da área parcial negativa da superfície	-0,189	4,296x10 <sup>-2</sup>	-4,409	1,477
δ <sub>H</sub> (min)	Carga parcial mínima do átomo de hidrogénio	-1,545x10 <sup>2</sup>	57,448	-2,689	1,225
<i>pLC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	4,590	0,491	9,351	
NLT	Número de ligações triplas	0,466	7,051x10 <sup>-2</sup>	6,602	4,787
MCI1	Média do conteúdo de informação (ordem 1)	-0,426	0,116	-3,670	1,487
SRZX	Razão entre a sombra projectada no plano ZX e o rectângulo no plano ZX	1,365	0,697	1,958	1,220
<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	3,899	0,199	19,617	
NRLT	Número relativo de ligações triplas	43,138	4,766	9,051	4,203
I <sub>A</sub>	Momento de inércia A	-14,361	2,720	-5,280	1,072
ASPR	Área superficial positiva relativa	0,777	0,315	2,466	1,372
CAAPNS	Carga atómica da área parcial negativa da superfície	-3,775x10 <sup>-2</sup>	1,650x10 <sup>-2</sup>	-2,287	1,581

Tabela 20F: Modelos de QSAR para a linha celular do cancro do pulmão.

<i>pIC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	3,989	0,259	15,394	
NRLT	Número relativo de ligações triplas	40,709	6,604	6,165	3,569
ASNR	Área superficial negativa relativa	-0,187	8,221x10 <sup>-2</sup>	-2,279	1,057
CAAPNS	Carga atómica da área parcial negativa da superfície	-6,143x10 <sup>-2</sup>	2,528x10 <sup>-2</sup>	-2,430	1,689
<i>pGI<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	-15,790	5,041	-3,133	
MCI0	Média do conteúdo de informação (ordem 0)	8,864	1,557	5,692	1,929
SRXY	Razão entre a sombra projectada no plano XY e o rectângulo no plano XY	-6,396	1,558	-4,104	1,166
ASPP/ATSM	Razão entre a área superficial parcial positiva e a área total da superfície molecular	8,288	2,925	2,833	1,059
NA	Número de anéis	0,278	0,128	2,173	1,643
<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	3,225	0,333	9,674	
CAPNS	Contribuição da área parcial negativa da superfície	-0,170	5,536x10 <sup>-2</sup>	-3,078	1,764
CTAPNS	Carga total da área parcial negativa da superfície	4,058x10 <sup>-3</sup>	2,087x10 <sup>-3</sup>	1,945	1,052
NRLT	Número relativo de ligações triplas	25,163	5,512	4,565	6,053
MCIC2	Média do conteúdo de informação complementar (ordem 2)	0,297	0,154	1,929	1,122

Tabela 21F: Modelos de QSAR para a linha celular do carcinoma mamário.

<i>pIC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	64,701	8,448	7,659	
$\delta_o(\max)$	Carga parcial máxima para o átomo de oxigénio	$5,716 \times 10^2$	87,555	6,529	5,008
ASPR	Área superficial positiva relativa	-1,607	0,536	-2,998	1,104
SYZ	Sombra projectada ao longo do plano YZ	$-2,645 \times 10^{-2}$	$1,362 \times 10^{-2}$	-1,942	1,696
<i>pTGI</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	4,789	0,231	20,768	
NRAN	Número relativo de átomos de azoto	-34,670	5,067	-6,842	2,940
$I_A$	Momento de inércia A	52,519	12,125	4,331	1,532
ASPR	Área superficial positiva relativa	-1,002	0,389	-2,572	1,098
<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	4,857	0,214	22,669	
NRAN	Número relativo de átomos de azoto	-35,505	4,657	-7,624	3,768
$I_A$	Momento de inércia A	42,164	11,038	3,820	1,466
ASPR	Área superficial positiva relativa	-0,719	0,356	-2,020	1,138

## F.5.2. Série 7

Tabela 22F: Modelos de QSAR para a linha celular da leucemia.

<i>pGI<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	7,607	0,979	7,768	
W	Índice de Wiener	5,779x10 <sup>-4</sup>	1,232x10 <sup>-4</sup>	4,691	1,378
CAAPS	Carga atômica da área positiva da superfície	-0,203	3,243x10 <sup>-2</sup>	-6,258	1,125
MCIL1	Média do conteúdo de informação sobre as ligações (ordem 1)	-2,461	1,229	-2,003	1,135
<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	5,718	0,369	15,481	
W	Índice de Wiener	2,449x10 <sup>-4</sup>	4,668x10 <sup>-5</sup>	5,249	1,296
CAAPS	Carga atômica da área positiva da superfície	-6,287x10 <sup>-2</sup>	1,287x10 <sup>-2</sup>	4,886	1,081
MCIO	Média do conteúdo de informação (ordem 0)	-0,555	0,197	-2,817	1,766

Tabela 23F: Modelos de QSAR para a linha celular do cancro do ovário.

<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	21,093	1,689	12,492	
MCIC0	Média do conteúdo de Informação complementar (ordem 0)	-1,942	0,203	-9,579	1,061
CPR	Carga positiva relativa	-42,759	5,239	-8,162	1,058
NAO	Número de átomos de oxigénio	-0,355	5,363x10 <sup>-2</sup>	-6,622	1,055
MCS	Média das cargas da superfície	-0,699	0,179	-3,916	1,176
MMR	Massa molecular relativa	-0,271	6,743x10 <sup>-2</sup>	-4,022	1,044
VM/CXYZ	Razão entre o volume molecular e a caixa de coordenadas XYZ	-4,688	1,214	-3,860	1,212

## F.5.3. Série 9

Tabela 24F: Modelos de QSAR para a linha celular do cancro do ovário.

<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	47,062	11,484	4,098	
NRA	Número relativo de anéis	-76,759	11,360	-6,757	1,298
NAO	Número de átomos de oxigénio	0,405	0,111	3,651	1,235
CMLOM	Contribuição máxima para a ligação de uma orbital molecular	-18,438	5,356	-3,442	1,276

Tabela 25F: Modelos de QSAR para a linha celular do cancro do pulmão.

<i>pIC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	4,363	9,260x10 <sup>-2</sup>	47,117	
CATN	Contribuição da área total negativa da superfície	-1,047x10 <sup>-2</sup>	1,231x10 <sup>-3</sup>	-8,504	2,703
SASAALH	Soma das áreas da superfície dos átomos aceitadores de ligações de hidrogénio	-1,242x10 <sup>-2</sup>	2,728x10 <sup>-3</sup>	-4,554	1,163
CDLH/ATSM	Razão entre os centros doadores de ligações de hidrogénio e a área total da superfície molecular	22,904	8,751	2,617	1,057
CFF	Calor final de formação	2,042x10 <sup>-3</sup>	9,426x10 <sup>-4</sup>	2,167	1,079
<i>pLC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	1,649x10 <sup>2</sup>	28,257	5,837	
δ <sub>o</sub> (min)	Carga parcial mínima para o átomo de oxigénio	11,057	1,9307	5,727	1,274
EAEMinAO	Estado atómico de energia mínima para o átomo de oxigénio	-0,522	9,126x10 <sup>-2</sup>	-5,723	1,262
SRYZ	Razão entre a sombra projectada no plano YZ e o rectângulo no plano YZ	1,565	0,510	3,067	1,336
SASAALH/ATSM	Razão entre a soma das áreas da superfície dos átomos aceitadores de ligações de hidrogénio e a área total da superfície molecular	-0,786	0,334	-2,352	1,161
<i>Média</i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	0,333	0,598	0,5570	
SZX	Sombra projectada no plano ZX	1,088x10 <sup>-2</sup>	1,729x10 <sup>-3</sup>	6,2896	1,405
SRYZ	Razão entre a sombra projectada no plano YZ e o rectângulo no plano YZ	4,559	0,780	5,8461	1,366
MinR <sub>AO</sub>	Índice de reactividade mínima para 1 electrão para o átomo de oxigénio	2,176x10 <sup>11</sup>	5,567x10 <sup>10</sup>	3,9083	1,168
CDH	Número de centros de doadores de hidrogénio	-7,190x10 <sup>-2</sup>	1,986x10 <sup>-2</sup>	-3,6213	1,066

Tabela 26F: Modelos de QSAR para a linha celular do cancro do cólon.

<i>pGI<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	28,561	5.251	5.439	
NRAC	Número relativo de átomos de carbono	- 21,681	2.692	-8.056	1,424
SRXY	Razão entre a sombra projectada no plano XY e o rectângulo no plano XY	6,025	1.107	5.443	1,188
CMLOM	Contribuição máxima para a ligação de uma orbital molecular	- 9,280	2.441	-3.803	1,152
W	Índice de Wiener	1,459	4.184x10 <sup>-5</sup>	3.489	1,069

## F.5.4. Série 12

Tabela 27F: Modelos de QSAR para a linha celular do cancro do ovário.

<i>pGI<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	3,922	0,192	20,467	
SASALH/A TSM	Razão entre a soma das áreas da superfície com aceitadores de ligações de hidrogénio e a área total da superfície molecular	1,678x10 <sup>2</sup>	36,527	4,595	3,722
ASPR	Área superficial positiva relativa	0,615	0,236	2,611	1,081
NLD	Número de ligações duplas	0,301	0,158	1,902	2,085

Tabela 28F: Modelos de QSAR para a linha celular do carcinoma mamário.

<i>pIC<sub>50</sub></i>					
Descritores	Significado	Coefficiente	Erro	Teste t	VIF
Constante	Intercepção	3,676	0,167	22,018	
$\delta_c(\max)$	Carga parcial máxima para o átomo de carbono	-32,669	4,830	-6,763	2,702
NL	Número de ligações	2,114x10 <sup>-2</sup>	5,104x10 <sup>-3</sup>	4,141	1,129
NLT	Número de ligações triplas	-8,318 x10 <sup>-2</sup>	2,322x10 <sup>-2</sup>	-3,582	1,349

## F.6. Tabelas de Estatística dos modelos de QSAR

### F.6.1. Série 1

**Tabela 29F:** Parâmetros estatísticos para a linha celular do cancro do ovário.

Parâmetros	pIC <sub>50</sub>	pGI <sub>50</sub>	pLC <sub>50</sub>	Média
<b>RMSE</b>	0,226	0,268	0,103	0,171
<b>Teste F</b>	32,766	33,181	40,438	42,158
<b>SEE</b>	0,226	0,268	0,103	0,171
<b>s<sup>2</sup></b>	0,066	0,092	0,014	0,040

**Tabela 30F:** Parâmetros estatísticos para a linha celular do cancro do pulmão.

Parâmetros	pIC <sub>50</sub>	pGI <sub>50</sub>	Média
<b>RMSE</b>	0,811	0,340	0,168
<b>Teste F</b>	25,793	18,704	38,204
<b>SEE</b>	0,255	0,298	0,139
<b>s<sup>2</sup></b>	0,099	0,149	0,037

**Tabela 31F:** Parâmetros estatísticos para a linha celular do carcinoma mamário.

Parâmetros	pIC <sub>50</sub>	pTGI	Média
<b>RMSE</b>	0,202	0,167	0,152
<b>Teste F</b>	32,374	26,805	30,440
<b>SEE</b>	0,202	0,167	0,152
<b>s<sup>2</sup></b>	0,053	0,037	0,030

### F.6.2. Série 7

**Tabela 32F:** Parâmetros estatísticos para a linha celular da leucemia.

Parâmetros	pGI <sub>50</sub>	Média
<b>RMSE</b>	0,164	0,065
<b>Teste F</b>	19,376	21,074
<b>SEE</b>	0,164	0,065
<b>s<sup>2</sup></b>	0,037	0,006

**Tabela 33F:** Parâmetros estatísticos para a linha celular do cancro do ovário.

<b>Parâmetros</b>	<b>Média</b>
<b>RMSE</b>	0,260
<b>Teste F</b>	22,308
<b>SEE</b>	0,260
<b>s<sup>2</sup></b>	0,084

**F.6.3. Série 9****Tabela 34F:** Parâmetros estatísticos para a linha celular do cancro do ovário.

<b>Parâmetros</b>	<b>Média</b>
<b>RMSE</b>	0,327
<b>Teste F</b>	22,092
<b>SEE</b>	0,326
<b>s<sup>2</sup></b>	0,142

**Tabela 35F:** Parâmetros estatísticos para a linha celular do cancro do pulmão.

<b>Parâmetros</b>	<b>pIC<sub>50</sub></b>	<b>pLC<sub>50</sub></b>	<b>Média</b>
<b>RMSE</b>	0,123	0,053	0,084
<b>Teste F</b>	22,659	23,117	19,111
<b>SEE</b>	0,123	0,053	0,084
<b>s<sup>2</sup></b>	0,020	0,004	0,009

**Tabela 36F:** Parâmetros estatísticos para a linha celular do cancro do cólon.

<b>Parâmetros</b>	<b>pGI<sub>50</sub></b>
<b>RMSE</b>	0,171
<b>Teste F</b>	23,055
<b>SEE</b>	0,170
<b>s<sup>2</sup></b>	0,039

**F.6.4. Série 12****Tabela 37F:** Parâmetros estatísticos para a linha celular do cancro do ovário.

<b>Parâmetros</b>	<b>pGI<sub>50</sub></b>
<b>RMSE</b>	0,339
<b>Teste F</b>	23,710
<b>SEE</b>	0,339
<b>s<sup>2</sup></b>	0,153

**Tabela 38F:** Parâmetros estatísticos para a linha celular do carcinoma mamário.

<b>Parâmetros</b>	<b>pIC<sub>50</sub></b>
<b>RMSE</b>	0,089
<b>Teste F</b>	21,842
<b>SEE</b>	0,089
<b>s<sup>2</sup></b>	0,011