

Pedro Rafael Mendes Reis

# Comparação de métodos de fingerprints para o rastreio virtual de inibidores da 5 $\alpha$ -redutase

Dissertação de Mestrado em Química Farmacêutica Industrial, orientada pela Doutora Cândida G. Silva e pelo Professor Doutor Jorge António Ribeiro Salvador e apresentada à Faculdade de Farmácia da Universidade de Coimbra

Setembro 2014



## Agradecimentos

A realização desta dissertação marca o fim de uma etapa importante da minha vida. Gostaria, por isso, de agradecer a todos aqueles que contribuíram de forma determinante para a sua concretização.

À Faculdade de Farmácia e ao Departamento de Química da Universidade de Coimbra manifesto apreço pela possibilidade de realização do presente trabalho e por todos os meios colocados à disposição. Agradeço igualmente os conhecimentos transmitidos, que foram úteis para esta dissertação.

À Doutora Cândida G. Silva, pela colaboração, conhecimentos transmitidos e capacidade de estímulo ao longo de todo o trabalho. Pela paciência e disponibilidade no esclarecimento de qualquer dúvida. Por todas as opiniões, críticas e orientações fundamentais para a concretização desta dissertação.

Ao Professor Doutor Rui Brito, por permitir a minha integração no grupo de trabalho que dirige, pela orientação, conselhos, compreensão e disponibilidade demonstrada.

Ao Professor Doutor Jorge Salvador, pela orientação, acompanhamento, compreensão e disponibilidade demonstradas ao longo deste trabalho.

Ao Doutor Carlos Simões, por todo o material científico disponibilizado, pela disponibilidade e conselhos dados.

A todos os restantes membros do grupo RMBLab em particular à Raquel, pelos conselhos, disponibilidade, boa disposição e ajuda.

A todos os meus amigos, que estiveram ao meu lado durante esta fase, pelo apoio, incentivo e companheirismo demonstrados, mesmo nos momentos em que o ânimo já era uma raridade.

À minha madrinha Fátima e padrinho António Faria e aos meus afilhados Eduardo e Nuno pela companhia, incentivo e apoio.

Aos meus pais, por todo o apoio, incentivo, amizade, carinho, força e paciência demonstrados e total ajuda na superação dos obstáculos que ao longo desta caminhada foram surgindo. Sem eles não seria possível chegar até ao fim. A eles dedico este trabalho!



# Índice

I.	Introdução.....	1
1.	5 $\alpha$ -redutase .....	2
1.1	Mecanismo de acção e tipos de inibidores da enzima 5 $\alpha$ -redutase .....	4
2.	Rastreio virtual.....	8
2.1	Rastreio virtual baseado na estrutura .....	9
2.2	Rastreio virtual baseado em ligandos .....	9
2.2.1	Similaridade e pesquisa de subestruturas.....	10
2.3	Descritores moleculares .....	11
2.3.1	Estudos quantitativos da relação estrutura-actividade.....	13
II.	Dados, métodos e protocolo experimental.....	17
1.	Dados.....	18
2.	Métodos .....	19
2.1	Cálculo de descritores moleculares .....	19
2.1.1	<i>Fingerprints</i> moleculares.....	19
2.1.2	Descritores moleculares 3D .....	23
2.2	Métodos de <i>data mining</i> e <i>machine learning</i> .....	23
2.2.1	<i>Clustering</i> .....	24
2.2.1.1	<i>Clustering</i> com base na estrutura dos compostos.....	25
2.2.1.2	<i>Clustering</i> com base em <i>fingerprints</i> .....	27
2.2.2	Modelos de classificação .....	27
2.2.2.1	<i>Naïve Bayes</i> .....	29
2.2.2.2	Máquinas de vectores de suporte.....	30
2.2.2.3	Medidas de desempenho .....	32
3.	Protocolo computacional .....	34
3.1	<i>Clustering</i> .....	34
3.2	Modelos de classificação.....	34
III.	Resultados.....	39
1.	Comparação dos <i>clusters</i> obtidos a partir das diferentes <i>fingerprints</i> .....	40
2.	Análise estrutural .....	55
3.	Classificadores.....	69
3.1	SVM.....	69
3.1.1	<i>Fingerprints</i> .....	69

3.1.1.1	5 $\alpha$ -R1.....	69
3.1.1.2	5 $\alpha$ -R2.....	84
3.1.1.3	Inibidores duais .....	99
3.1.2	Descritores 3D .....	104
3.1.2.1	5 $\alpha$ -R1.....	104
3.1.2.2	5 $\alpha$ -R2.....	106
3.1.2.3	Inibidores duais .....	107
3.2	<i>Naïve Bayes</i> .....	108
3.2.1	<i>Fingerprints</i> .....	108
3.2.1.1	5 $\alpha$ -R1.....	108
3.2.1.2	5 $\alpha$ -R2.....	123
3.2.1.3	Inibidores duais .....	138
3.2.2	Descritores 3D .....	143
3.2.2.1	5 $\alpha$ -R1.....	143
3.2.2.2	5 $\alpha$ -R2.....	145
3.2.2.3	Inibidores duais .....	146
4.	Métricas de desempenho.....	148
4.1	5 $\alpha$ -R1 .....	148
4.1.1	SVM.....	148
4.1.1.1	<i>Fingerprints</i> .....	148
4.1.1.2	Descritores 3D .....	152
4.1.2	<i>Naïve Bayes</i> .....	153
4.1.2.1	<i>Fingerprints</i> .....	153
4.1.2.2	Descritores 3D .....	157
4.2	5 $\alpha$ -R2.....	158
4.2.1	SVM.....	158
4.2.1.1	<i>Fingerprints</i> .....	158
4.2.1.2	Descritores 3D .....	162
4.2.1	<i>Naïve Bayes</i> .....	163
4.2.1.1	<i>Fingerprints</i> .....	163
4.2.1.2	Descritores 3D .....	167
4.3	Inibidores duais.....	168
4.3.1	SVM.....	168
4.3.1.1	<i>Fingerprints</i> .....	168

4.3.1.2	Descritores 3D .....	169
4.3.2	<i>Naïve Bayes</i> .....	169
4.3.2.1	<i>Fingerprints</i> .....	169
4.3.2.2	Descritores 3D .....	170
IV.	Discussão .....	171
V.	Conclusão.....	183
VI.	Referências bibliográficas.....	187
VII.	Anexos.....	195
1.	ChEMBL ID dos inibidores da 5 $\alpha$ -RI.....	196
2.	ChEMBL ID dos inibidores da 5 $\alpha$ -R2.....	198
3.	ChEMBL ID dos inibidores duais da 5 $\alpha$ -Redutase .....	200
4.	Padrões SMART da <i>fingerprint</i> FP3 .....	202
5.	Resultados da tese anterior dos classificadores SVM, para a 5 $\alpha$ -RI .....	203
6.	Resultados da tese anterior dos classificadores SVM, para a 5 $\alpha$ -R2 .....	204



## Índice de figuras

Figura 1. Mecanismo de acção da 5 $\alpha$ -redutase (Adaptado de Aggarwal et al., 2010).....	4
Figura 2. Mecanismo de redução da testosterona (Adaptado de Aggarwal et al., 2010).....	5
Figura 3. Estruturas químicas dos compostos finasterida (à esquerda) e dutasterida (à direita).....	6
Figura 4. Mecanismo de inibição da 5 $\alpha$ -redutase pela finasterida (Adaptado de Bull et al., 1996).....	6
Figura 5. Tipos de rastreio virtual.....	8
Figura 6. Esquema representativo da chave estrutural (a), <i>hashed fingerprints</i> (b) e <i>fingerprints</i> circulares (c). Adaptado de Indiana Chemoinformatics Education Portal, Daylight Chemical Information Systems, Inc, 2011 e ChEMBioInfo.....	19
Figura 7. Representação esquemática de uma análise de <i>clustering</i> (Adaptado da página web: <a href="http://scikit-learn.org/0.9/modules/clustering.html">http://scikit-learn.org/0.9/modules/clustering.html</a> ). .....	24
Figura 8. Exemplo de dendrograma. ....	25
Figura 9. Exemplo esquemático de um grafo (Adaptado de Diestel, 2000).....	26
Figura 10. Algoritmos populares de <i>data mining</i> . (a) SVM, (b) Classificadores Bayesianos e (c) Árvores de decisão (Adaptado de Geppert et al., 2009). .....	28
Figura 11. Representação esquemática de um hiperplano óptimo de separação de padrões linearmente separáveis. Adaptado de Kotsiantis, 2007.....	31
Figura 12. Representação esquemática do mapeamento de características (Adaptado de Sousa et al., 2013).....	32
Figura 13. Análise de <i>clustering</i> para as várias <i>fingerprints</i> utilizando o KNIME. ....	34
Figura 14. <i>Workflow</i> do KNIME para o método SVM.....	36
Figura 15. <i>Workflow</i> do KNIME para o método Naïve Bayes, quando se utiliza <i>fingerprints</i> . .....	36
Figura 16 <i>Workflow</i> do KNIME para o método Naïve Bayes, quando se utilizam descritores 3D.....	37
Figura 17. Dendrogramas obtidos do <i>clustering</i> hierárquico das <i>fingerprints</i> CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), <i>Standard</i> (F), <i>Extended</i> (G), <i>EState</i> (H), <i>PubChem</i> (I) e <i>MACCS</i> (J) que representam os 32 inibidores da 5 $\alpha$ -R1 classificados como pertencentes à classe de IC50 excelente, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos <i>clusters</i> foi 0,30.....	42
Figura 18. Dendrogramas obtidos do <i>clustering</i> hierárquico das <i>fingerprints</i> CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), <i>Standard</i> (F), <i>Extended</i> (G), <i>EState</i> (H), <i>PubChem</i> (I) e <i>MACCS</i> (J) que representam os 141 inibidores da 5 $\alpha$ -R2 classificados como pertencentes à classe de IC50 excelente, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos <i>clusters</i> foi 0,30.....	45

Figura 19. Dendrogramas obtidos do <i>clustering</i> hierárquico das <i>fingerprints</i> CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), <i>Standard</i> (F), <i>Extended</i> (G), <i>EState</i> (H), <i>PubChem</i> (I) e MACCS (J) que representam os 58 inibidores duais da 5 $\alpha$ -Redutase, com classificação Forte/Forte de IC50, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos <i>clusters</i> foi 0,30.....	48
Figura 20. Dendrogramas obtidos do <i>clustering</i> hierárquico das <i>fingerprints</i> CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), <i>Standard</i> (F), <i>Extended</i> (G), <i>EState</i> (H), <i>PubChem</i> (I) e MACCS (J) que representam os 46 inibidores duais da 5 $\alpha$ -Redutase, com classificação Forte/Fraco de IC50, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos <i>clusters</i> foi 0,30.....	51
Figura 21. Dendrogramas obtidos do <i>clustering</i> hierárquico das <i>fingerprints</i> CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), <i>Standard</i> (F), <i>Extended</i> (G), <i>EState</i> (H), <i>PubChem</i> (I) e MACCS (J) que representam os 23 inibidores duais da 5 $\alpha$ -Redutase, com classificação Fraco/Forte de IC50, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos <i>clusters</i> foi 0,30.....	54
Figura 22. Dendrograma obtido do <i>clustering</i> hierárquico das estruturas químicas dos 32 inibidores da 5 $\alpha$ -RI, com IC50 Excelente.....	55
Figura 23. Dendrograma obtido do <i>clustering</i> hierárquico das estruturas químicas dos 141 inibidores da 5 $\alpha$ -R2, com IC50 Excelente.....	57
Figura 24. Dendrograma obtido do <i>clustering</i> hierárquico das estruturas químicas dos 58 inibidores duais da 5 $\alpha$ -Redutase, com IC50 Forte/Forte. ....	62
Figura 25. Dendrograma obtido do <i>clustering</i> hierárquico das estruturas químicas dos 46 inibidores duais da 5 $\alpha$ -Redutase, com IC50 Forte/Fraco. ....	64
Figura 26. Dendrograma obtido do <i>clustering</i> hierárquico das estruturas químicas dos 46 inibidores duais da 5 $\alpha$ -Redutase, com IC50 Fraco/Forte. ....	66

## Índice de tabelas

Tabela 1. Descrição dos compostos em estudo para a 5 $\alpha$ -RI e 5 $\alpha$ -R2 obtidos da ChEMBL, versão 18.....	18
Tabela 2. Tabela resumo das <i>fingerprints</i> estudadas e programas utilizados para o seu cálculo.....	22
Tabela 3. Atribuição de 4 tipos classes aos valores de IC50.....	35
Tabela 4. Atribuição de 5 tipos de classes aos valores de IC50.....	35
Tabela 5. Atribuição de classificação binária aos valores de IC50.....	35
Tabela 6. Divisão por classes dos inibidores duais. ....	36
Tabela 7. Análise das MCS dos 32 inibidores da 5 $\alpha$ -RI com IC50 Excelente. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade. ....	55
Tabela 8. Análise das MCS dos 141 inibidores da 5 $\alpha$ -R2 com IC50 Excelente. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade. ....	58
Tabela 9. Análise das MCS dos 58 inibidores duais da 5 $\alpha$ -Redutase com IC50 Forte/Forte. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade. ....	63
Tabela 10. Análise das MCS dos 46 inibidores duais da 5 $\alpha$ -Redutase com IC50 Forte/Fraco. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.....	64
Tabela 11. Análise das MCS dos 23 inibidores duais da 5 $\alpha$ -Redutase com IC50 Fraco/Forte. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.....	67
Tabela 12. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	70
Tabela 13. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	70
Tabela 14. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). ....	71
Tabela 15. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	71
Tabela 16. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	72
Tabela 17. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).....	72

Tabela 18. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	73
Tabela 19. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	73
Tabela 20. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). ....	74
Tabela 21. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	74
Tabela 22. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	75
Tabela 23. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). ....	75
Tabela 24. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	76
Tabela 25. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	76
Tabela 26. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). ....	77
Tabela 27. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	77
Tabela 28. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	78
Tabela 29. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	78
Tabela 30. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	79

Tabela 31. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	79
Tabela 32. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	80
Tabela 33. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	80
Tabela 34. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	81
Tabela 35. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	81
Tabela 36. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	82
Tabela 37. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	82
Tabela 38. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	83
Tabela 39. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	83
Tabela 40. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	84
Tabela 41. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).....	84
Tabela 42. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	85
Tabela 43. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	85

Tabela 44. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	86
Tabela 45. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	86
Tabela 46. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	87
Tabela 47. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c). .....	87
Tabela 48. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	88
Tabela 49. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	88
Tabela 50. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	89
Tabela 51. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	89
Tabela 52. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	90
Tabela 53. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	90
Tabela 54. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	91
Tabela 55. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	91
Tabela 56. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	92

Tabela 57. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	92
Tabela 58. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	93
Tabela 59. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	93
Tabela 60. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	94
Tabela 61. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	94
Tabela 62. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	95
Tabela 63. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	95
Tabela 64. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	96
Tabela 65. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	96
Tabela 66. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	97
Tabela 67. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b). .....	97
Tabela 68. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c). .....	98
Tabela 69. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	98

Tabela 70. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	99
Tabela 71. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).....	99
Tabela 72. Reclassificação dos compostos em análise, representados por CHF, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	100
Tabela 73. Reclassificação dos compostos em análise, representados por ECFP, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.	100
Tabela 74. Reclassificação dos compostos em análise, representados por FP2, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	101
Tabela 75. Reclassificação dos compostos em análise, representados por FP3, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	101
Tabela 76. Reclassificação dos compostos em análise, representados por FP4, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	102
Tabela 77. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	102
Tabela 78. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	103
Tabela 79. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	103
Tabela 80. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	104
Tabela 81. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais. ....	104

Tabela 82. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	105
Tabela 83. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	105
Tabela 84. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	106
Tabela 85. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).....	106
Tabela 86. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).....	107
Tabela 87. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	107
Tabela 88. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais. ....	108
Tabela 89. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	109
Tabela 90. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	109
Tabela 91. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	110
Tabela 92. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	110
Tabela 93. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	111
Tabela 94. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (c).....	111

Tabela 95. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	112
Tabela 96. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	112
Tabela 97. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	113
Tabela 98. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	113
Tabela 99. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	114
Tabela 100. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	114
Tabela 101. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	115
Tabela 102. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	115
Tabela 103. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	116
Tabela 104. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	116
Tabela 105. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	117
Tabela 106. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	117
Tabela 107. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	118

Tabela 108. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	118
Tabela 109. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	119
Tabela 110. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	119
Tabela 111. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	120
Tabela 112. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	120
Tabela 113. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	121
Tabela 114. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	121
Tabela 115. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	122
Tabela 116. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	122
Tabela 117. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	123
Tabela 118. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (c).....	123
Tabela 119. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	124
Tabela 120. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	124

Tabela 121. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	125
Tabela 122. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	125
Tabela 123. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	126
Tabela 124. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (c).....	126
Tabela 125. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	127
Tabela 126. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	127
Tabela 127. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	128
Tabela 128. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	128
Tabela 129. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	129
Tabela 130. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	129
Tabela 131. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a). .....	130
Tabela 132. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	130
Tabela 133. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	131

Tabela 134. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	131
Tabela 135. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	132
Tabela 136. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	132
Tabela 137. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	133
Tabela 138. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	133
Tabela 139. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	134
Tabela 140. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	134
Tabela 141. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	135
Tabela 142. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	135
Tabela 143. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	136
Tabela 144. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	136
Tabela 145. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	137
Tabela 146. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	137

Tabela 147. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	138
Tabela 148. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (c).....	138
Tabela 149. Reclassificação dos compostos em análise, representados por CHF, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	139
Tabela 150. Reclassificação dos compostos em análise, representados por ECFP, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	139
Tabela 151. Reclassificação dos compostos em análise, representados por FP2, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais. ....	140
Tabela 152. Reclassificação dos compostos em análise, representados por FP3, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais. ....	140
Tabela 153. Reclassificação dos compostos em análise, representados por FP4, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais. ....	141
Tabela 154. Reclassificação dos compostos em análise, representados por <i>Standard fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	141
Tabela 155. Reclassificação dos compostos em análise, representados por <i>Extended fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	142
Tabela 156. Reclassificação dos compostos em análise, representados por <i>EState fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	142
Tabela 157. Reclassificação dos compostos em análise, representados por <i>PubChem fingerprints</i> , que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	143

Tabela 158. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	143
Tabela 159. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	144
Tabela 160. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	144
Tabela 161. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	145
Tabela 162. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (a).....	145
Tabela 163. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação (b).....	146
Tabela 164. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	146
Tabela 165. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em <i>Naïve Bayes</i> , utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.....	147
Tabela 166. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	148
Tabela 167. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	149
Tabela 168. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	149
Tabela 169. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	149
Tabela 170. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	150

Tabela 171. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>standard fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	150
Tabela 172. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>extended fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	150
Tabela 173. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>EState fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	151
Tabela 174. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>PubChem fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	151
Tabela 175. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	152
Tabela 176. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -RI .....	152
Tabela 177. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	153
Tabela 178. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	153
Tabela 179. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	154
Tabela 180. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	154
Tabela 181. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	154
Tabela 182. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>standard fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	155
Tabela 183. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>extended fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI .....	155

Tabela 184. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>EState fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	156
Tabela 185. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>PubChem fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	156
Tabela 186. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	156
Tabela 187. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -R1.....	157
Tabela 188. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	158
Tabela 189. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	158
Tabela 190. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	159
Tabela 191. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	159
Tabela 192. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	159
Tabela 193. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>standard fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	160
Tabela 194. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>standard fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	160
Tabela 195. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>extended fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	161
Tabela 196. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>EState fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	161

Tabela 197. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e <i>PubChem fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	161
Tabela 198. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	162
Tabela 199. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -R2.....	163
Tabela 200. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	163
Tabela 201. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	164
Tabela 202. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	164
Tabela 203. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	164
Tabela 204. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	165
Tabela 205. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>standard fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	165
Tabela 206. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>extended fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	166
Tabela 207. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>EState fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	166
Tabela 208. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e <i>PubChem fingerprints</i> para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	166
Tabela 209. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	167

Tabela 210. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -R2.....	168
Tabela 211. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação dos inibidores duais e <i>fingerprints</i> para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.....	168
Tabela 212. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação dos inibidores duais e descritores 3D para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.....	169
Tabela 213. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação dos inibidores duais e <i>fingerprints</i> para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.....	169
Tabela 214. Métricas de desempenho dos algoritmos de aprendizagem <i>Naïve Bayes</i> , que utilizaram critérios de classificação dos inibidores duais e descritores 3D para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.....	170
Tabela 215. ChEMBL ID dos 310 inibidores da 5 $\alpha$ -R1 com IC50 reportado na ChEMBL.....	196
Tabela 216. ChEMBL ID dos 355 inibidores da 5 $\alpha$ -R2 com IC50 reportado na ChEMBL.....	198
Tabela 217. ChEMBL ID dos 181 inibidores duais 5 $\alpha$ -Redutase com IC50 reportado na ChEMBL.....	200
Tabela 218. 55 padrões SMART da <i>fingerprint</i> FP3.....	202
Tabela 219. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (a).....	203
Tabela 220. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	203
Tabela 221. Métricas de desempenho dos algoritmos de aprendizagem SVM da tese anterior, que utilizaram critérios de classificação (a) e (c) e descritores ID, para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.....	203
Tabela 222. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (a).....	204
Tabela 223. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).....	204
Tabela 224. Métricas de desempenho dos algoritmos de aprendizagem SVM da tese anterior, que utilizaram critérios de classificação (a) e (c) e descritores ID, para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.....	204



## Lista de siglas e abreviaturas

0D – Adimensional

1D – Unidimensional

2D - Bidimensional

3D – Tridimensional

4D – Tetradimensional

5 $\alpha$ -R1 – 5 $\alpha$ -Redutase 1

5 $\alpha$ -R2 – 5 $\alpha$ -Redutase 2

CHF – *Chemical hashed fingerprint*

ECFP – *Extended-connectivity fingerprint*

IC50 – Concentração inibitória de 50%

MACCS – *Molecular access system*

MCS – Subestrutura Máxima Comum (no Inglês, *Maximum Common Substructure*)

NADP<sup>+</sup> – Fosfato de dinucleótido de nicotinamida e adenina (*Nicotinamide adenine dinucleotide phosphate*)

NADPH – Forma reduzida da *Nicotinamide adenine dinucleotide phosphate*

QSAR – Estudos quantitativos da relação estrutura-atividade (no Inglês, *Quantitative Structure-Activity Relationship*)

SVM – Máquinas de vetores de suporte (no Inglês, *Support Vector Machines*)



## Resumo

A hiperplasia benigna da próstata e o cancro da próstata são duas doenças proliferativas relacionadas com a idade com morbilidade e mortalidade significativas. O controlo da proliferação e da actividade celular na próstata possibilita minorar os sintomas e o avanço destas patologias, melhorando a qualidade de vida. Este controlo pode ser assegurado pelo bloqueio da acção da testosterona e da di-hidrotestosterona ou pela inibição da conversão da testosterona em di-hidrotestosterona pela enzima 5 $\alpha$ -redutase. Os fármacos disponíveis no mercado, dutasterida e finasterida inibem a 5 $\alpha$ -Redutase 1 e 5 $\alpha$ -Redutase 2, no primeiro caso, e a 5 $\alpha$ -Redutase 2, no segundo. Com vista a minimizar os efeitos adversos destes fármacos e aumentar a potência e a selectividade dos inibidores tem-se investido no desenvolvimento de novos inibidores da 5 $\alpha$ -redutase.

Como até ao momento não foi possível determinar a estrutura tridimensional da proteína 5 $\alpha$ -Redutase, o presente trabalho incide sobre o rastreio virtual baseado em ligandos. Neste sentido o trabalho realizado centrou-se em dois objectivos principais: (i) comparar e avaliar eventuais diferenças de representações moleculares através de *clustering*; e (ii) comparar e avaliar o desempenho da combinação de métodos de aprendizagem de máquina com diferentes descritores moleculares, de modo a contribuir para a descoberta e desenvolvimento de novos inibidores com potência e selectividade aumentada para a 5 $\alpha$ -Redutase.

Foram seleccionados diferentes descritores moleculares 2D e 3D, em particular 10 tipos diferentes de *fingerprints* 2D e 128 descritores 3D. Para a análise de *clustering* foram utilizadas as 10 *fingerprints* para verificar se esta análise permite obter resultados semelhantes ao *clustering* das estruturas químicas dos compostos baseada na máxima subestrutura comum. Numa segunda etapa, analisaram-se as subestruturas máximas comuns entre os inibidores da 5 $\alpha$ -Redutase 1, 5 $\alpha$ -Redutase 2 e inibidores duais. Por fim, aplicaram-se métodos de aprendizagem de máquina – *Support Vector Machines* e *Naïve Bayes* – com diferentes critérios de classificação sobre os inibidores da 5 $\alpha$ -Redutase 1, 5 $\alpha$ -Redutase 2 e inibidores duais descritos através das 10 *fingerprints* utilizadas e dos descritores tridimensionais para avaliar qual a combinação que permite a melhor identificação de inibidores da 5 $\alpha$ -Redutase.

Na primeira etapa do trabalho, as *hashed fingerprints* obtiveram os resultados mais semelhantes ao *clustering* das estruturas químicas. A segunda etapa revelou que estruturas 4-azasteróides e 6-azasteróides estão presentes em grande parte dos inibidores. Por último, a terceira etapa determinou que os classificadores *Support Vector Machines* com critérios de classificação específicos e utilizando *fingerprints* circulares proporcionam bons classificadores que poderão, no futuro, ser utilizados em bases de dados mais vastas para detectar novos inibidores da 5 $\alpha$ -Redutase.

Palavras-chave: 5 $\alpha$ -Redutase, QSAR, *fingerprints*, descritores moleculares, *clustering*, MCS, SVM e *Naïve Bayes*.

## **Abstract**

The benign prostatic hyperplasia and prostate cancer are two age related proliferative diseases, with significant morbidity and mortality. Cell proliferation and activity control in prostate allows to lowering the symptoms and the progress of these diseases, improving life quality. This control can be achieved by blocking testosterone and dihydrotestosterone, or through inhibition of the conversion of testosterone into dihydrotestosterone by 5 $\alpha$ -reductase enzyme. The available drugs in the market, finasteride and dutasteride that inhibit both 5 $\alpha$ -reductase 1 and 5 $\alpha$ -reductase 2, and 5 $\alpha$ -reductase 2, respectively. In order to minimize adverse effects and to increase the potency and selectivity of 5 $\alpha$ -reductase inhibitors, the quest to develop new inhibitors has increased.

The three dimensional structure of 5 $\alpha$ -reductase protein has not been determined yet, thus the present work focuses on ligand-based virtual screening. This work has focused on two major goals: (i) compare and evaluate possible differences in molecular representation through clustering; and (ii) compare and evaluate the performance of different combinations of machine learning methods and molecular descriptors, in order to discover and develop new 5 $\alpha$ -Reductase inhibitors with increased potency and selectivity.

Different 2D and 3D molecular descriptors were selected, in particular 10 different fingerprints and 128 3D descriptors. Clustering analysis was performed using 10 fingerprints, and the results compared with clustering analysis based on the maximum common substructure. In a second step, we analysed the maximum common substructures among the 5 $\alpha$ -reductase 1 inhibitors, 5 $\alpha$ -reductase 2 inhibitors and dual inhibitors. Finally, we applied machine learning methods - Support Vector Machines and Naïve Bayes - with different classification criteria on 5 $\alpha$ -reductase 1, 5 $\alpha$ -Reductase 2 and dual inhibitors described through 10 fingerprints and 3D molecular descriptors to assess which combination allows a better identification of 5 $\alpha$ -Reductase inhibitors.

In the first point, hashed fingerprints obtained the most similar results to the clustering of chemical structures. Second, structures revealed that 4-azasteroids and 6-azasteroids are present in many 5 $\alpha$ -Reductase inhibitors. Finally, the Support Vector Machines classifiers with a specific classification criteria and using circular fingerprints provide good classifiers that may, hereafter, be used in wider databases to detect new 5 $\alpha$ -Reductase inhibitors.

Keywords: 5 $\alpha$ -Reductase, QSAR, fingerprints, molecular descriptors, clustering, MCS, SVM and Naïve Bayes.

## **I. Introdução**

## I. 5 $\alpha$ -redutase

A testosterona é o androgénio mais abundante na circulação sanguínea. Nos fetos masculinos, a testosterona estimula a diferenciação do canal deferente nos órgãos genitais internos masculinos (epidídimo, canal deferente e vesículas seminais) bem como o desenvolvimento da libido, cordas vocais, músculos esqueléticos, pênis, escroto e a iniciação da espermatogénese na puberdade. Quando absorvida pelas células a testosterona é, seguidamente, convertida em di-hidrotestosterona pela enzima 5 $\alpha$ -redutase. A di-hidrotestosterona é importante na diferenciação uterina e no crescimento da próstata, dos órgãos genitais externos masculinos e dos pêlos faciais e corporais durante a puberdade. A di-hidrotestosterona representa um papel importante em diversas patologias humanas, nas quais se incluem acne, hirsutismo, calvície masculina, hiperplasia benigna da próstata e cancro da próstata (Azzouni *et al.*, 2012).

Na próstata, a testosterona e a di-hidrotestosterona exercem os seus efeitos através da ligação ao receptor androgénico. Deste modo, é iniciada a transcrição e, conseqüentemente, a produção de proteínas específicas e o crescimento celular. (Salvador *et al.*, 2013) A hiperplasia benigna da próstata e o cancro da próstata são duas doenças proliferativas relacionadas com a idade com morbidade e mortalidade significativas (Salvador *et al.*, 2013).

A hiperplasia benigna da próstata, mais comum que o cancro da próstata, resulta do aumento da proliferação celular quer no estroma quer no epitélio da próstata, e/ou comprometimento do mecanismo de apoptose (Aggarwal *et al.* 2010). Deste modo pode ocorrer obstrução uretral e uma disfunção do músculo detrusor, resultando num esvaziamento incompleto da bexiga (Azzouni *et al.*, 2012). Estas alterações do funcionamento normal do sistema urinário podem levar a uma diminuição da qualidade de vida, sendo que os sintomas típicos da hiperplasia benigna da próstata podem ser agrupados em obstrutivos (ex: menor jacto urinário e maior volume residual de urina) e irritativos (ex: hesitação e aumento da frequência urinária) (Salvador *et al.*, 2013). A hiperplasia benigna da próstata pode ser encontrada em cerca de 50% dos homens com idade a rondar os 50 anos e em 90% dos homens com idade a rondar os 80 anos (Azzouni *et al.*, 2012).

Por sua vez, o cancro da próstata é o cancro mais frequente, exceptuando os cancros da pele, nos homens europeus e dos Estados Unidos da América (Azzouni *et al.*, 2012). Tal como a hiperplasia benigna da próstata, o desenvolvimento inicial do cancro da próstata

depende da acção dos androgénios, sendo útil por este motivo reduzir os níveis de androgénios em circulação que possam estimular a proliferação celular na próstata (Salvador *et al.*, 2013).

Assim, ao controlar a proliferação e a actividade celular na próstata é possível minorar os sintomas e avanço da hiperplasia benigna da próstata e do cancro da próstata, melhorando a qualidade de vida. Este controlo pode ser assegurado quer pelo bloqueio da acção da testosterona, da di-hidrotestosterona ou pela inibição da conversão da testosterona em di-hidrotestosterona pela enzima 5 $\alpha$ -redutase (Oliveira *et al.*, 2005).

A família 5 $\alpha$ -redutase é composta por 3 subfamílias e 5 membros (isoenzimas) no total:

- 5 $\alpha$ -redutase 1 (5 $\alpha$ -R1) e 5 $\alpha$ -redutase 2 (5 $\alpha$ -R2);
- 5 $\alpha$ -redutase 3;
- Proteínas GPSN2 (glicoproteína sináptica 2) e GPSN2L (glicoproteína sináptica 2-like) (Azzouni *et al.*, 2012).

As isoenzimas 5 $\alpha$ -R1 e 5 $\alpha$ -R2, em estudo, são enzimas associadas à membrana (microssomais) NADPH-dependentes (NADPH, forma reduzida do inglês *nicotinamide adenine dinucleotide phosphate*), compostas por 259 e 254 aminoácidos, e com um peso molecular de 29,5 e 28,4 kDa, respectivamente. Estas isoenzimas possuem um elevado conteúdo de aminoácidos hidrofóbicos distribuídos pelas suas sequências, sugerindo que são proteínas intrínsecas profundamente enraizadas na bicamada lipídica (Azzouni *et al.*, 2012).

Nos estudos realizados em lisados de células transfectadas, verificou-se que a 5 $\alpha$ -R1 possui um espectro de pH óptimo, que varia de 6,0 a 8,5, enquanto a 5 $\alpha$ -R2 apresenta um espectro mais apertado de pH óptimo (pH 5-5,5). Contudo, existem evidências que sugerem que dentro de células humanas intactas, o pH óptimo a que a isoenzima 5 $\alpha$ -R2 exerce a sua actividade é mais neutro (pH 6,0-7,0) (Azzouni *et al.*, 2012).

A 5 $\alpha$ -R1 encontra-se predominantemente no fígado e nas glândulas sebáceas da pele. Algumas linhagens celulares humanas com cancro da próstata (DU-145 e LNCaP) expressam unicamente 5 $\alpha$ -R1 evidenciando um papel mais importante desta enzima no cancro da próstata do que na hiperplasia benigna da próstata (Hosoda & Hashimoto, 2007; Salvador *et al.*, 2013).

Por sua vez, a 5 $\alpha$ -R2 pode ser encontrada, maioritariamente, na próstata, vesículas seminais, epidídimo, fígado e folículos capilares apresentando um papel muito importante na hiperplasia benigna da próstata (Hosoda & Hashimoto, 2007; Salvador *et al.*, 2013).

Tendo em conta estes aspectos, vários compostos foram desenvolvidos com o objectivo de inibir o sistema enzimático da 5 $\alpha$ -reductase, representando, assim, um papel importante na prevenção e tratamento de doenças como a hiperplasia benigna da próstata e o cancro da próstata (Hosoda & Hashimoto, 2007).

## 1.1 Mecanismo de acção e tipos de inibidores da enzima 5 $\alpha$ -reductase

O mecanismo químico proposto para a redução da testosterona a di-hidrotestosterona catalisado pela 5 $\alpha$ -reductase (Fig. 1) envolve a formação de um complexo binário entre a enzima e o NADPH, seguido pela formação de um complexo ternário com o substrato testosterona (Aggarwal *et al.*, 2010).

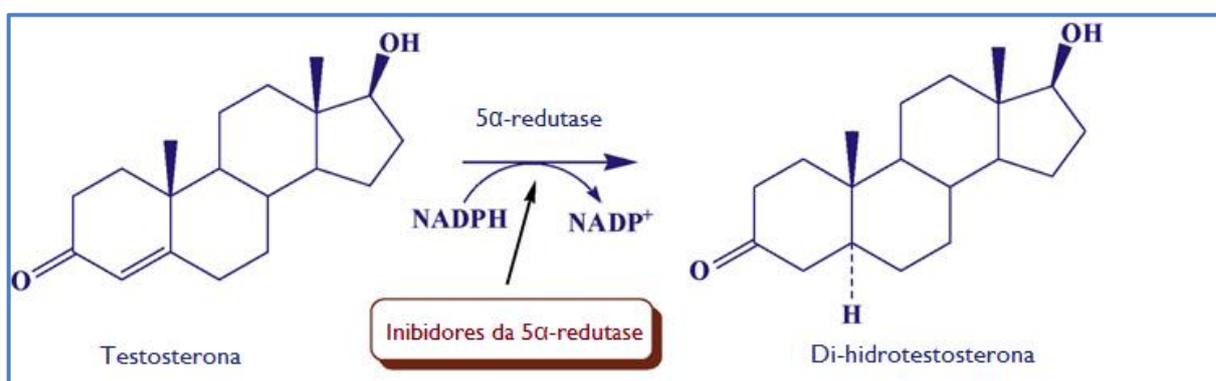


Figura 1. Mecanismo de acção da 5 $\alpha$ -reductase (Adaptado de Aggarwal *et al.*, 2010).

Depois, um resíduo electrofílico no local activo da enzima activa o sistema  $\Delta^4$ -3-cetona do substrato, originando um carbocátion deslocalizado em C<sub>5</sub> (Fig. 2) (Salvador *et al.*, 2013). Seguidamente, forma-se o enolato de di-hidrotestosterona por transferência do hidreto do NADPH para a face  $\alpha$  do carbocátion deslocalizado levando a uma redução selectiva em C<sub>5</sub>. Este enolato, coordenado com NADP<sup>+</sup> na face  $\alpha$ , é, de seguida, protonado na face  $\beta$  em C<sub>4</sub> originando um complexo ternário composto pela enzima-NADP<sup>+</sup>-dihidrotestosterona. Após a di-hidrotestosterona e o NADP<sup>+</sup> partirem, a enzima fica livre para novos ciclos catalíticos (Aggarwal *et al.*, 2010).

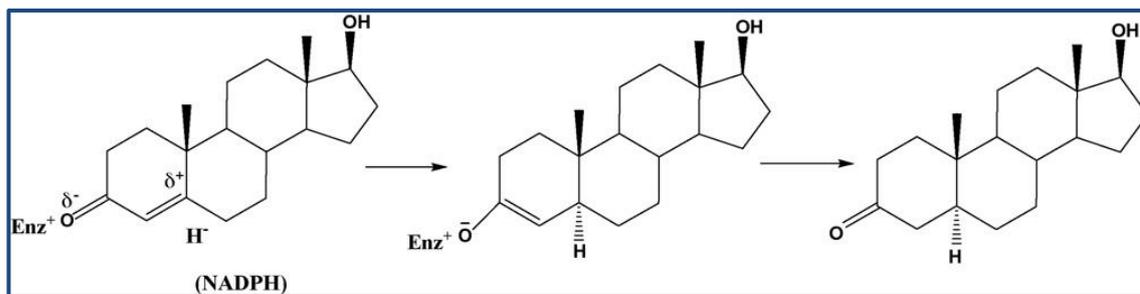


Figura 2. Mecanismo de redução da testosterona (Adaptado de Aggarwal et al., 2010).

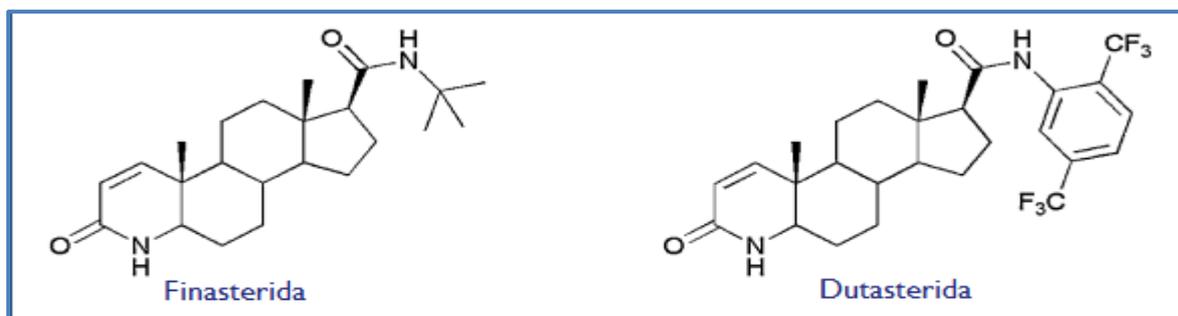
Os inibidores da 5 $\alpha$ -redutase podem ser classificados em três grupos, tendo em conta o mecanismo de inibição (Aggarwal et al., 2010):

- 1) inibidores do tipo A que competem com o cofactor (NADPH) e com o substrato (testosterona) e interagem com a enzima livre;
- 2) inibidores do tipo B que competem com o substrato (testosterona) e ligam-se ao complexo enzima-NADPH;
- 3) inibidores do tipo C que encaixam no complexo enzima-NADP<sup>+</sup> exibindo um mecanismo de inibição não-competitivo.

Do ponto estrutural, os inibidores da 5 $\alpha$ -redutase podem ser agrupados em esteróides e não esteróides, sendo que a classe esteróide contém mais compostos do que a classe não esteróide (Salvador et al., 2013).

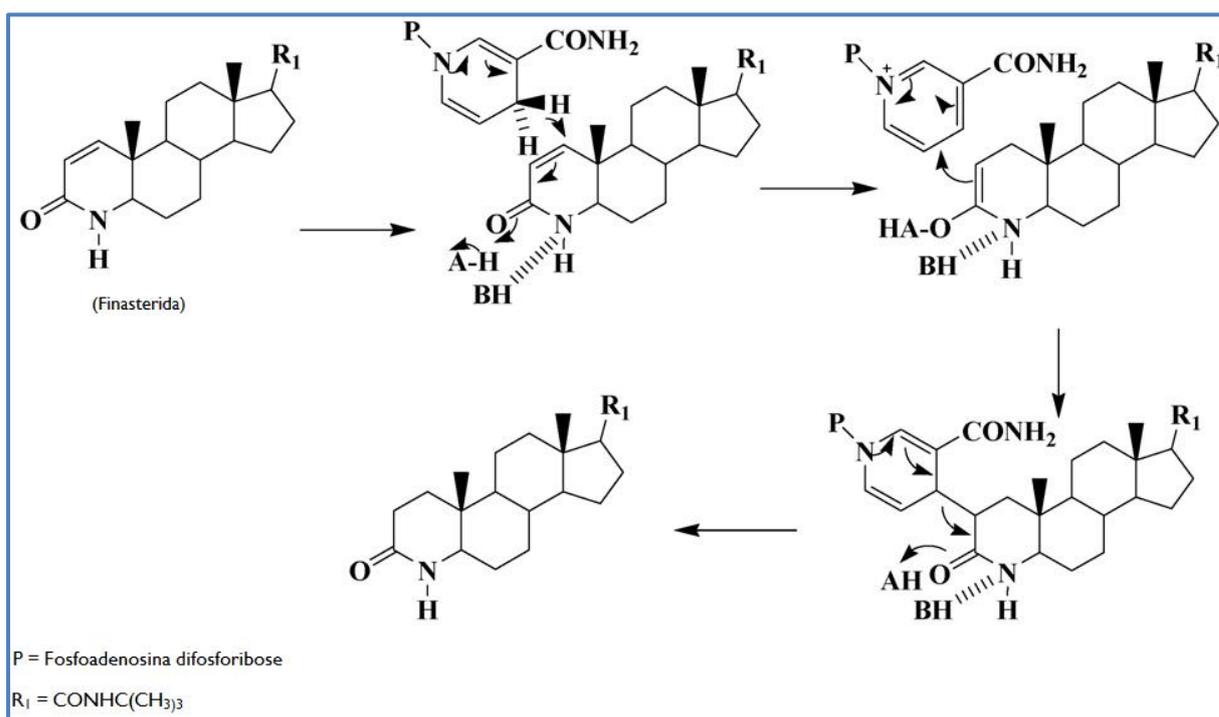
Na classe dos inibidores não esteróides encontram-se, principalmente, análogos azasteróides, preparados através da remoção de um ou mais anéis da estrutura primária, ou triterpenóides isolados do fungo *Ganoderma lucidum*, como o ácido ganodérico. Estes compostos foram estudados e desenhados com o objectivo de ultrapassar os efeitos adversos dos compostos esteróides (Salvador et al., 2013).

Relativamente aos compostos pertencentes à classe esteróide, estes podem ser agrupados em três subclasses, os azasteróides, ácidos 3-carboxilícos e outros derivados do pregnano/androstano. Os compostos do grupo mais importante - os azasteróides - foram desenvolvidos com vista a mimetizar o enolato intermediário que se liga à enzima através de uma mudança isostérica entre um carbono e um azoto. Neste grupo encontramos os compostos finasterida e dutasterida (Fig. 3) (Salvador et al., 2013).



**Figura 3. Estruturas químicas dos compostos finasterida (à esquerda) e dutasterida (à direita).**

A finasterida, o primeiro inibidor da 5 $\alpha$ -reductase a ser comercializado, inibe selectivamente a 5 $\alpha$ -R2 e actua como um substrato alternativo, que após se ligar à enzima possibilita a transferência do hidreto do cofactor NADPH para a dupla ligação  $\Delta^1$  da finasterida (Fig. 4). O enolato intermediário tautomeriza no local activo da enzima formando um bisubstrato análogo, onde a di-hidrofinasterida está covalentemente ligada ao NADP<sup>+</sup>. De seguida, forma-se um aducto extremamente estável através de uma ligação não covalente entre a enzima e o complexo NADP-di-hidrofinasterida (Salvador *et al.*, 2013).



**Figura 4. Mecanismo de inibição da 5 $\alpha$ -reductase pela finasterida (Adaptado de Bull *et al.*, 1996)**

Apesar da redução do antigénio específico da próstata e da melhoria global do estado de saúde dos doentes com hiperplasia benigna da próstata, a administração de finasterida implica efeitos adversos tais como a perda de apetite sexual, disfunção eréctil e depressão. Por este motivo, foram desenhados outros inibidores da 5 $\alpha$ -reductase com vista a minimizar os efeitos adversos da finasterida, melhorar o poder inibitório e selectividade realizando

estudos quantitativos da relação estrutura-atividade (QSAR, do Inglês *Quantitative Structure-Activity Relationship*) (Salvador *et al.*, 2013).

A dutasterida, constitui um importante avanço nos inibidores da 5 $\alpha$ -redutase, uma vez que se trata de um inibidor duplo, ou seja, inibe ambas as isoenzimas (5 $\alpha$ -R1 e 5 $\alpha$ -R2), permitindo obter um bloqueio mais geral da síntese de di-hidrotestosterona. Apesar da diferença estrutural entre a dutasterida e a finasterida residir na cadeia lateral, na ligação C<sub>17</sub>, a dutasterida, como inibidor competitivo irreversível, apresenta uma potência 60 vezes superior à finasterida e forma um complexo estável, de baixa taxa de dissociação, com as duas isoenzimas. Esta inibição permite reduzir em mais de 90% os níveis de di-hidrotestosterona em circulação e melhora os resultados clínicos de doentes com hiperplasia benigna da próstata. Para além destes aspectos, a dutasterida é bem tolerada e possui apenas efeitos secundários transientes (Salvador *et al.*, 2013).

Em suma, a terapia com inibidores da 5 $\alpha$ -redutase melhora os sintomas urinários, diminui a frequência e o risco de cirurgias relacionadas com a hiperplasia benigna da próstata e do cancro da próstata. Contudo pode provocar efeitos adversos significativos como a perda de libido, disfunção erétil e ejaculatória e depressão (Traish *et al.*, 2011). Com vista a minimizar estes efeitos adversos, mas também para aumentar a potência e a selectividade dos inibidores tem-se investido no desenvolvimento de novos inibidores da 5 $\alpha$ -redutase. Porém existem alguns obstáculos neste caminho. Até ao momento, apenas foi possível aferir a sequência primária da 5 $\alpha$ -redutase, por estimativa a partir do c-DNA (do Inglês *complementary-DeoxiriboNucleic Acid*), não sendo conhecida a estrutura tridimensional da enzima, o que condiciona o desenho e desenvolvimento de novos inibidores. (Aggarwal *et al.*, 2010; Salvador *et al.*, 2013).

## 2. Rastreo virtual

O processo de desenho e desenvolvimento de fármacos pretende identificar compostos com interesse farmacológico. Estes compostos são, na sua maioria, pequenas moléculas orgânicas (ligandos) que interagem com biomoléculas específicas (receptores/alvos) e são compilados em bases de dados (Bielska *et al.*, 2011). Apesar de tecnologias como a química combinatória e o rastreo de alto rendimento aumentarem substancialmente a quantidade de compostos identificados, não é razoável, do ponto de vista económico, experimental e temporal, sintetizar tantos compostos (Walters *et al.*, 1998; Bielska *et al.*, 2011).

Assim, o rastreo virtual surge como uma ferramenta importante, uma vez que utiliza o poder computacional para pesquisar e analisar milhões de compostos químicos, visando encontrar novos compostos ou quimiotipos com a actividade biológica pretendida (Bielska *et al.*, 2011; Lavecchia & Di Giovanni 2013). Esta metodologia permite, também, incluir nas bases de dados compostos que existam apenas em teoria e não na realidade. Desta forma, reduz-se substancialmente o número de compostos a comprar/sintetizar, bem como o tempo e os custos do estudo (Bielska *et al.*, 2011; Reddy *et al.*, 2007).

O rastreo virtual pode ser dividido em duas categorias (Fig. 5): rastreo virtual baseado em ligandos e rastreo virtual baseado na estrutura (Lavecchia & Di Giovanni, 2013).

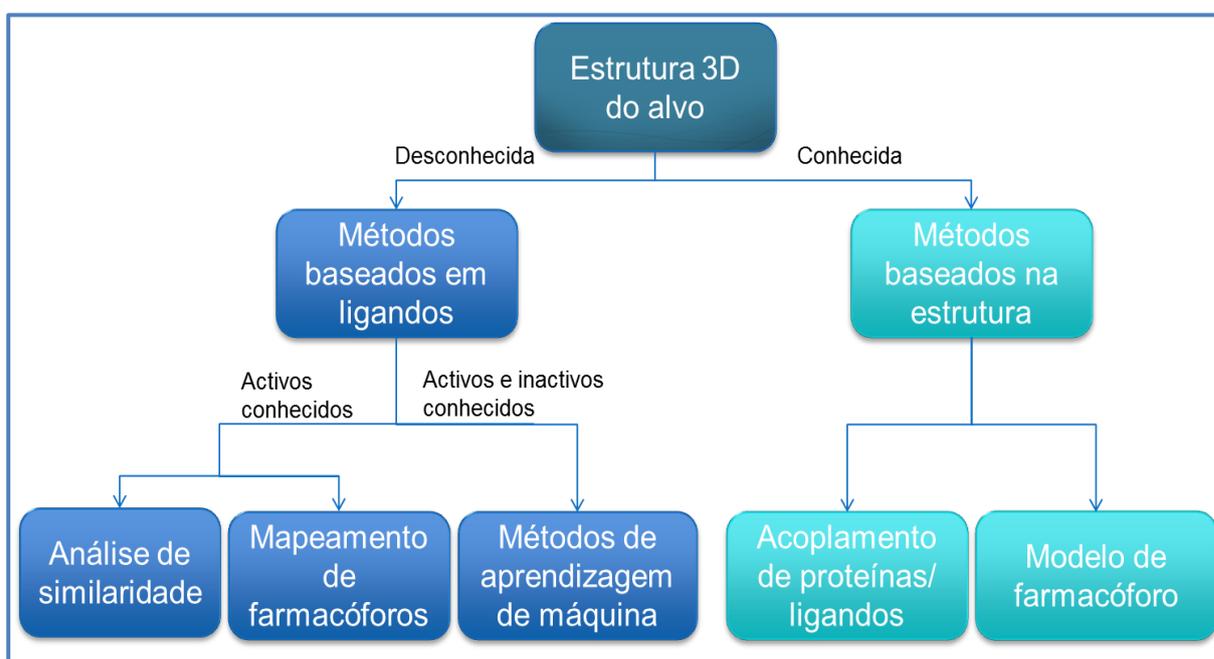


Figura 5. Tipos de rastreo virtual.

## 2.1 Rastreo virtual baseado na estrutura

O rastreo virtual baseado na estrutura necessita que a estrutura tridimensional (3D) do alvo biológico esteja disponível para que seja possível estabelecer, por exemplo, modelos de acoplamento entre o alvo e os ligandos candidatos, com vista a identificar os ligandos com melhor afinidade e complementaridade para com o alvo. As estruturas 3D podem ser determinadas experimentalmente, por cristalografia de raio-X ou por ressonância magnética nuclear, ou computacionalmente através da modelação molecular por homologia estrutural (Sanders *et al.*, 2012; Lavecchia & Di Giovanni, 2013).

Para se estabelecerem os modelos de acoplamento é necessário um método que faça uma previsão do espaço conformacional do ligando respeitando o espaço conformacional do alvo, gerando hipóteses de ligação entre o alvo e o ligando – avaliação da complementaridade. É ainda necessário uma função de pontuação, que avalie e ordene as hipóteses de ligação propostas de acordo com a força de ligação – avaliação da afinidade (Bielska *et al.*, 2011; Lyne, 2002).

Caso sejam conhecidas estruturas de complexos proteína-ligando, é possível aplicar uma análise de similaridade do ligando ou construir um modelo de farmacóforo baseado no receptor, para gerar um conjunto de compostos que possua as características físico-químicas dos ligandos já existentes (Bielska *et al.*, 2011). Um farmacóforo é definido como o conjunto de características, bem como a orientação espacial relativa que define a interação da ligação entre um receptor e um ligando. Nestas características incluem-se dadores e aceitadores de hidrogénio, grupos carregados positiva e negativamente, regiões hidrofóbicas e anéis aromáticos (Leach & Gillet, 2007). Devido ao aumento dos dados tridimensionais de estruturas proteicas, este tipo de rastreo virtual tem sido cada vez mais utilizado no desenho e desenvolvimento de novos fármacos (Reddy *et al.*, 2007; Sanders *et al.*, 2012).

## 2.2 Rastreo virtual baseado em ligandos

O rastreo virtual baseado em ligandos tenta explorar dados sobre a relação estrutura-actividade de um conjunto de compostos activos que se ligam ao alvo em estudo, de modo a identificar outros candidatos com propriedades semelhantes para analisar experimentalmente (Lavecchia & Di Giovanni, 2013). Este tipo de rastreo é frequentemente utilizado quando não se conhece a estrutura tridimensional e é possível dividi-lo em três classes:

- 1) Análise de similaridade e pesquisa de subestruturas, onde se tem como referência as estruturas bidimensionais e/ou tridimensionais dos compostos sob rastreio. A análise de similaridade assenta no princípio que moléculas estruturalmente relacionadas são susceptíveis de apresentar propriedades semelhantes (Reddy *et al.*, 2007; Guido *et al.*, 2008; Lavecchia & Di Giovanni, 2013)
- 2) Modelos de farmacóforo baseado em ligandos, definidos como arranjo tridimensional de características ou fragmentos moleculares que são necessários, mas não necessariamente suficientes, para que um composto seja capaz de se ligar a um alvo. Estes métodos envolvem a identificação do padrão farmacofórico comum a um conjunto de compostos activos conhecidos, bem como a utilização desse padrão na pesquisa de subestruturas tridimensionais.
- 3) Métodos de aprendizagem de máquina, onde se constrói uma regra de classificação e/ou regressão a partir de um conjunto de dados de treino que contém moléculas activas e inactivas. Este tipo de rastreio permite construir modelos de QSAR robustos e precisos utilizando um vasto conjunto de compostos (Guido *et al.*, 2008; Lavecchia & Di Giovanni, 2013).

### 2.2.1 Similaridade e pesquisa de subestruturas

A análise de similaridade é uma das técnicas mais utilizadas nos rastreios virtuais. Esta técnica compara várias moléculas entre si, podendo ter como referência um composto activo, que serve para indicar uma medida de similaridade estrutural e depois classificar as moléculas consoante essa medida (Hert *et al.*, 2004). Esta análise assenta no pressuposto de que moléculas semelhantes tendem a possuir actividades semelhantes. (Bender *et al.*, 2009; Cai *et al.*, 2013).

Uma vantagem deste princípio é que não faz suposições sobre as características físicas e químicas relacionadas com a actividade molecular. Contudo, esta vantagem pode também ser considerada uma desvantagem uma vez que a presença de características irrelevantes diminui a qualidade dos resultados (Cai *et al.*, 2013). Porém é também sabido que uma pequena mudança na estrutura química pode diminuir ou aumentar drasticamente a actividade de uma substância, promovendo uma boa base para optimizações de compostos com actividade farmacológica ou biológica possivelmente útil a nível terapêutico (*leads*) (Eckert & Bajorath, 2007).

A pesquisa de subestruturas e superestruturas consiste numa análise de similaridade local muito utilizada, em que se procura grupos funcionais ou padrões de átomos que possam estar repetidos nos compostos a analisar (Cai *et al.*, 2013). Assim, se em duas estruturas A e B, a estrutura A estiver totalmente contida na estrutura B, então A é uma subestrutura de B e B é uma superestrutura de A. De acordo com o princípio da similaridade, A e B poderão possuir actividade semelhante uma vez que possuem uma subestrutura comum. Posto isto, uma subestrutura que está associada a determinadas propriedades pode ser utilizada numa pesquisa em bases de dados para identificar compostos que partilhem esta subestrutura (ou superestrutura) e, possivelmente, a sua actividade. Contudo esta estratégia de comparação é bastante rígida, tem uma elevada taxa de falsos-negativos e necessita de um modelo (subestrutura) bem definido, caso contrário os resultados serão de baixa qualidade e, como não gera uma medida quantitativa de similaridade, torna-se difícil classificar os compostos (Cao *et al.*, 2008).

Neste contexto, surge um método de pesquisa de subestruturas, a designado pesquisa da Subestrutura Máxima Comum (MCS, do Inglês *Maximum Common Substructure*), que permite determinar a maior subestrutura comum entre dois compostos. Este método é vantajoso, uma vez que é intuitivo (a maior subestrutura comum de compostos estruturalmente parecidos tem, provavelmente, importância na actividade dos compostos), permite destacar a MCS entre dois compostos e permite gerar uma medida de similaridade (Cao *et al.*, 2008).

Para uma análise de similaridade por vezes não é suficiente utilizar as estruturas moleculares por si só, pelo que é necessário caracterizar as moléculas através de descritores moleculares para depois se proceder à comparação (Hert *et al.*, 2004; Willett, 2005).

## 2.3 Descritores moleculares

Um descritor molecular consiste no resultado final de um procedimento matemático lógico que transforma a informação química representada através de estruturas químicas num número ou resultado de um procedimento padrão (Todeschini & Consonni, 2000).

Existem, actualmente, vários tipos de representações estruturais de moléculas, como por exemplo descritores constitucionais, de contagem, *fingerprints* e descritores geométricos, entre outros. Um bom descritor deve apresentar as seguintes características (Leszczynski *et al.*, 2010; Verma, 2010; Sousa *et al.*, 2013):

- ter uma interpretação estrutural;
- apresentar boa correlação com pelo menos uma propriedade;
- permitir, preferencialmente, a discriminação de isómeros;
- ser aplicável a estruturas locais;
- ser generalizável a descritores maiores;
- ser independente;
- simplicidade;
- não ser construído com base nas propriedades;
- não deve ser trivialmente relacionado com outros descritores;
- permitir uma construção eficiente;
- usar conceitos estruturais familiares;
- apresentar uma dependência correcta em relação ao tamanho;
- apresentar uma mudança gradual com uma mudança gradual das estruturas.

De acordo com a dimensionalidade, estes descritores moleculares podem ser divididos em 5 classes (Todeschini & Consonni, 2000):

- Adimensionais (0D) – Forma mais simples de descrever uma molécula. Envolve a contagem de ligações, de átomos e do peso molar;
- Unidimensionais (1D) – Integra a contagem de fragmentos e de dadores/aceitadores de ligações de hidrogénio;
- Bidimensionais (2D) – Descritores topológicos, nos quais se incluem *fingerprints*, têm em conta propriedades como a densidade, logP, refractividade molecular, área superficial;
- Tridimensionais (3D) – Descritores geométricos, que inclui as coordenadas 3D e os graus de rotação da molécula, por exemplo;
- Tetradimensionais (4D) – Descritores que juntam coordenadas 3D com as possíveis conformações.

Os descritores 0D e 1D permitem uma análise computacional mais rápida, mas não discriminam suficientemente as moléculas de forma a permitir uma análise mais concreta. Os descritores 2D descrevem as estruturas de acordo com o tamanho e forma global. As *fingerprints* 2D são descritas como ferramentas poderosas na análise de similaridade, reconhecendo compostos activos através de sequências de *bits*, não existindo nenhuma *fingerprint* que possa ser definida como *standard*. A sensibilidade de uma *fingerprint* torna-se

importante para perceber qual será a melhor a utilizar em cada caso (Godden *et al.*, 2005; Eckert & Bajorath, 2007). Os descritores 3D fornecem uma visão mais global do que as *fingerprints* 2D, uma vez que têm em conta propriedades que dependem das coordenadas ou da orientação da molécula (por exemplo, momento dipolar de ligação), permitem distinguir diferentes conformações e quiralidade de uma molécula e presume-se que não sejam tão afectados por erros analógicos (Eckert & Bajorath, 2007; Venkatraman *et al.*, 2010). Contudo, os descritores 3D não integram a mudança de conformação das moléculas aquando da sua ligação ao receptor e não permitem testar, por enquanto, um grande número de moléculas devido ao aumento do tempo de análise computacional. No caso dos descritores 4D, como já se inclui as energias de interacção, é possível eventuais mudanças de conformação ultrapassando o limite da mudança de conformação presente nos descritores 3D (Todeschini & Consonni, 2000).

### 2.3.1 Estudos quantitativos da relação estrutura-actividade

O estudo quantitativo da relação estrutura-actividade (QSAR) permite criar modelos computacionais ou matemáticos que visam encontrar uma correlação estatisticamente significativa entre as propriedades estruturais e/ou físico-químicas e a actividade biológica de compostos activos (Verma, 2010; Lavecchia & Di Giovanni, 2013). O conhecimento de valores de actividade biológica dos compostos, como afinidade de ligação, constante de dissociação ( $K_d$ ) ou concentração inibitória de 50% (IC50), é um requisito para os estudos de QSAR. As estruturas moleculares são normalmente representadas por conjuntos específicos de propriedades estruturais e físico-químicas (descritores moleculares) das moléculas consideradas mais relevantes para a actividade em estudo (Lavecchia & Di Giovanni, 2013). Os modelos QSAR providenciam previsões úteis, permitindo ultrapassar limitações como o elevado tempo e custo dos métodos de síntese e ensaios biológicos convencionais de moléculas que posteriormente apresentam fracos perfis de absorção, distribuição, metabolização e toxicidade (Verma *et al.*, 2010). A qualidade do estudo QSAR é afectada, entre outros aspectos, pela representatividade dos dados estrutura-actividade, que tal como a selecção dos descritores moleculares a utilizar, deve descrever eficazmente os parâmetros relevantes para a actividade e alvo em estudo. A adequação das relações quantitativas desenvolvidas, bem como a análise estatística desenvolvida e a correspondente validação visam garantir a confiança e a utilidade do modelo (Venkatraman *et al.*, 2010; Lavecchia & Di Giovanni, 2013; Sousa *et al.*, 2013). As metodologias QSAR podem ser baseadas na dimensionalidade ou no tipo de métodos quimiométricos utilizados. De acordo

com a classificação de Verma *et al.*, publicada em 2010, dentro da dimensionalidade podemos categorizar os métodos QSAR em:

- 1D-QSAR: correlaciona a actividade com propriedades moleculares gerais como o  $pK_a$ ,  $\log P$ , etc;
- 2D-QSAR: correlaciona a actividade com padrões estruturais, como índices de conectividade, sem ter em conta a representação 3-D dessas propriedades;
- 3D-QSAR: correlaciona a actividade com campos de interacção moleculares não covalentes;
- 4D-QSAR: inclui, adicionalmente, o conjunto das configurações das ligações em 3D-QSAR;
- Pentadimensional (5D)-QSAR: representa modelos ajustados do 4D-QSAR;
- Hexadimensional (6D)-QSAR: Incorpora características de solvação nos modelos de 5D-QSAR.

Os métodos utilizados para os estudos de QSAR são categorizados consoante a técnica que estabelece a correlação entre propriedades estruturais e actividade biológica. Estas técnicas podem ser lineares (ex. regressão linear, regressão linear múltipla) e não-lineares (ex. redes neurais artificiais, redes neurais de Bayes) (Verma *et al.*, 2010).

Pretende-se com o presente trabalho atingir os seguintes objectivos:

- Aplicar métodos computacionais de análise de similaridade com vista a detectar possíveis subestruturas comuns a inibidores duais da 5 $\alpha$ -redutase;
- Comparar e avaliar eventuais diferenças de representação molecular através de *clustering*;
- Comparar e avaliar o desempenho da combinação de métodos de aprendizagem de máquina com diferentes descritores (*fingerprints* e descritores 3D), de modo a contribuir para a descoberta e desenvolvimento de novos inibidores com potência e selectividade aumentada para a 5 $\alpha$ -Redutase e/ou com menor toxicidade.



## **II. Dados, métodos e protocolo experimental**

## I. Dados

A ChEMBL<sup>(1)</sup> é uma base de dados de acesso livre que contem mais de 1,2 milhões de moléculas com características conhecidas, como a sua actividade sobre milhares de alvos terapêuticos, entre outras informações funcionais ou relacionadas com a absorção, distribuição, metabolização, excreção e toxicidade. Estes dados foram extraídos de publicações científicas e cuidadosamente verificados de forma a eliminar qualquer ambiguidade, tornando a ChEMBL numa base de dados muito fiável (Cai *et al.*, 2013).

No seguimento de uma tese de mestrado anterior que tinha como objectivo a utilização de descritores moleculares 0D e 1D para o treino de modelos de classificação para a identificação de compostos com capacidade para inibir a actividade da 5 $\alpha$ -R1 e 5 $\alpha$ -R2, realizou-se uma actualização da biblioteca de compostos com bioactividade conhecida que inibem a 5 $\alpha$ -R1 e 5 $\alpha$ -R2 que foram extraídos da base de dados ChEMBL (Sousa *et al.*, 2013). A actualização realizada visou retirar compostos cuja actividade deixou de estar comprovada ou que foram retirados da base de dados, e incluir novos compostos cuja actividade surge descrita na versão 18 da ChEMBL. Desta forma, para a 5 $\alpha$ -R1 foram encontrados 310 compostos com IC50 reportado (Anexo I) e para a 5 $\alpha$ -R2 foram encontrados 355 compostos com IC50 reportado (Anexo II). De todos estes inibidores, 181 são inibidores duais com IC50 reportado (Anexo III) para as duas isoenzimas (Tabela I).

**Tabela I. Descrição dos compostos em estudo para a 5 $\alpha$ -R1 e 5 $\alpha$ -R2 obtidos da ChEMBL, versão 18.**

	Número de compostos estudados
5 $\alpha$ -redutase 1	310
5 $\alpha$ -redutase 2	355
Comuns	181

<sup>(1)</sup> <https://www.ebi.ac.uk/chembl/>

## 2. Métodos

### 2.1 Cálculo de descritores moleculares

No presente estudo foram utilizados descritores 2D e 3D para descrever todas as moléculas em análise.

#### 2.1.1 Fingerprints moleculares

As *fingerprints* moleculares são representações das estruturas químicas desenhadas inicialmente para facilitar a pesquisa de subestruturas em grandes bases de dados, bem como para auxiliar tarefas de análise de similaridade, *clustering* e classificação (Rogers & Hahn, 2010). Esta representação estrutural é feita através de uma sequência de 0 e 1 (no Inglês, *bit strings*) (Venkatraman *et al.*, 2010). As *fingerprints* utilizadas podem ser divididas em 3 grupos com particularidades diferentes: chaves estruturais, *hashed fingerprints* e *fingerprints* circulares (Fig. 6).

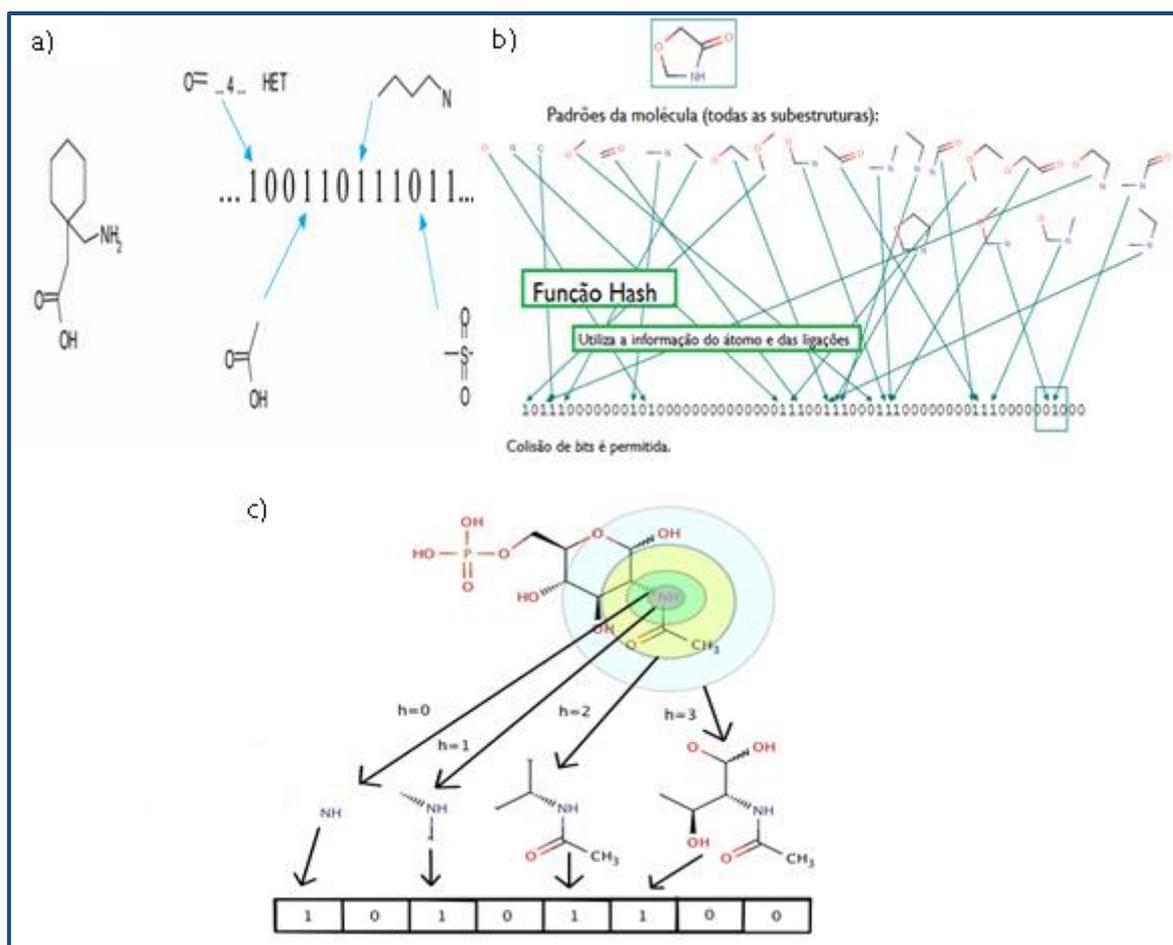


Figura 6. Esquema representativo da chave estrutural (a), *hashed fingerprints* (b) e *fingerprints* circulares (c). Adaptado de Indiana Chemoinformatics Education Portal, Daylight Chemical Information Systems, Inc, 2011 e Chembioinfo.

Numa chave estrutural, cada *bit* (0 ou 1) representa a presença ou a ausência de uma estrutura ou padrão químico predefinido, como por exemplo, sequências de átomos, configurações electrónicas e sistemas de anéis. Este tipo de *fingerprints* é muito utilizado na pesquisa de subestruturas (Hert *et al.*, 2004; Rogers & Hahn, 2010).

As *hashed fingerprints* representam padrões/caminhos de átomos e ligações até um determinado comprimento, normalmente de 2 a 7 átomos (por exemplo, átomo-ligação-átomo), podendo incluir alguns sistemas de anéis. É, assim, gerado um vector binário, que é sujeito a um método de *hashing* (mapeamento), como por exemplo, um algoritmo que transforma dados de comprimento variável em dados de comprimento fixo para facilitar a manipulação, uma vez que o conjunto de padrões é muito grande e varia de molécula para molécula (Hert *et al.*, 2004; Rogers & Hahn, 2010; Venkatraman *et al.*, 2010). Nas *hashed fingerprints*, não é possível atribuir padrões específicos a um determinado *bit* ou conjunto de *bits*, uma vez que um certo número de *bits* é atribuído a cada padrão e um *bit* pode até representar vários padrões (colisão de *bits*). Caso este fenómeno seja muito frequente a qualidade da *fingerprint* decresce. Este tipo de *fingerprints* é bastante utilizado na análise de bases de dados de compostos, na pesquisa de estruturas, subestruturas e na análise de similaridade, devida à sua rapidez e eficácia, podendo contudo gerar alguns falsos *hits* (Daylight Chemical Information Systems, Inc, 2011; Hert *et al.*, 2004; Rogers & Hahn, 2010).

Por último, as *fingerprints* circulares, que representam a estrutura através da vizinhança dos átomos, foram desenhadas para captar características moleculares relevantes para a actividade da molécula e não com vista a pesquisar subestruturas (Glem *et al.*, 2006; Rogers & Hahn, 2010).

As *extended-connectivity fingerprints* (ECFP) são um exemplo de *fingerprints* circulares e resultam de uma derivação do algoritmo de Morgan (Morgan, 1965) para resolver o problema do isomorfismo (duas moléculas iguais com diferentes números atómicos) (Hert *et al.*, 2004; Rogers & Hahn, 2010). No algoritmo de Morgan, um processo iterativo atribui identificadores numéricos a cada átomo, inicialmente utilizando uma regra que codifica a informação atómica num identificador e, depois, utilizando esses identificadores iniciais para as iterações futuras. O processo de iteração continua até que os identificadores atómicos sejam únicos (ou o mais perto possível permitido), sendo que os identificadores intermédios são descartados. O algoritmo das ECFP faz algumas alterações ao algoritmo de Morgan, uma vez que a geração de ECFP termina após um número predeterminado de iterações (neste

caso 4, de acordo com as especificações predefinidas do programa). Os identificadores iniciais, intermédios e finais são reunidos num conjunto que define a ECFP. Estas alterações permitem reduzir os recursos computacionais necessários e comparar ECFP entre moléculas diferentes. Desta forma, o processo de geração da ECFP tem três passos sequenciais (Hattori *et al.*, 2008; Rogers & Hahn, 2010):

1. Uma fase de atribuição inicial, na qual é atribuído a cada átomo um identificador inicial, que é baseado no número de ligações da molécula, número atómico, carga e massa do átomo;
2. Uma fase de actualização iterativa, em que cada identificador é actualizado para reflectir os vizinhos de cada átomo, incluindo a identificação de se tratar de uma repetição de outras características;
3. Uma fase de remoção de duplicados, na qual múltiplas ocorrências da mesma característica são reduzidas a uma na lista final de características.

Segundo Rogers & Hahn (2010), as ECFP possuem vantagens sobre as chaves estruturais, uma vez que as últimas, descrevem um conjunto de características predefinido não tendo em conta a variação estrutural. Além deste facto, as chaves estruturais são desenhadas para pesquisa de subestruturas, limitando a sua utilidade nas pesquisas relacionadas com actividade. Por último, mesmo o conjunto de chaves estruturais mais indicado é limitado na sua cobertura, ao contrário das ECFP. Por sua vez, as *hashed fingerprints* comparadas com as ECFP também apresentam vantagens nas análises computacionais relacionadas com a actividade, uma vez que alguns dos caminhos descritos pelas *hashed fingerprints* podem não corresponder a um grupo funcional ou subestrutura facilmente identificável. A estereoquímica também é apenas descrita nas ECFP ao contrário das *hashed fingerprints* (Rogers & Hahn, 2010).

**Tabela 2. Tabela resumo das fingerprints estudadas e programas utilizados para o seu cálculo.**

Fingerprints	Programas		
	ChemAxon®	OpenBabel®	CDK®
Hashed fingerprints	Chemical hashed fingerprints	FP2	Standard fingerprint, Extended fingerprint
Chaves estruturais		FP3, FP4	EState fingerprint, PubChem fingerprint, MACCS
Fingerprints circulares	Extended-connectivity fingerprints		

As fingerprints 2D foram calculadas recorrendo a três programas diferentes: ChemAxon<sup>®(2)</sup>, OpenBabel<sup>®(3)</sup> e CDK<sup>(4)</sup> (Tabela 2). Através do software da ChemAxon<sup>®</sup> foram calculadas duas fingerprints com 1024 bits de comprimento cada: as *chemical hashed fingerprints* (CHF) que pertencem ao grupo das hashed fingerprints e as *extended-connectivity fingerprints* (ECFP) que são as únicas fingerprints circulares utilizadas no presente estudo. Com o software Open Babel<sup>®</sup> foram calculadas mais três fingerprints, designadas FP2, FP3 e FP4. As FP2 são *hashed fingerprints* que identificam padrões lineares e subestruturas circulares na molécula com comprimento de 1 a 7 átomos (excluindo subestruturas de 1 átomo de C e N) e mapeia estes padrões em *bit strings* de 1024 bits de comprimento utilizando uma função *hash*. As restantes duas fingerprints são chaves estruturais. No caso da FP3, é verificada a presença ou ausência de 55 padrões SMART (Anexo IV) e a FP4 verifica a presença ou ausência de 307 padrões SMART (descritos no ficheiro SMARTS\_IntelLigand.txt) (O'Boyle et al., 2011). Com recurso às funcionalidades do CDK<sup>®</sup> disponíveis no software KNIME foram calculadas mais cinco fingerprints. Inseridas no grupo das *hashed fingerprints* e com um comprimento de 1024 bits, utilizaram-se a *standard fingerprint* e a *extended fingerprint*, que é similar à *standard fingerprint*, mas tem em conta os anéis presentes nas moléculas. Do grupo das chaves estruturais foram utilizadas a *EState fingerprint* que verifica a presença de 79<sup>6</sup> padrões químicos, a *PubChem fingerprint* que consiste em 881<sup>7</sup> fragmentos estruturais, e a MACCS (no Inglês, *Molecular ACCess System*) que consiste em 166<sup>8</sup> fragmentos estruturais de 1 a 10 átomos não-hidrogénio (Hattori et al., 2008; Venkatraman et al., 2010; Dimova et al., 2013).

<sup>(2)</sup> <http://www.chemaxon.com/>

<sup>(3)</sup> [http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)

<sup>(4)</sup> <http://sourceforge.net/projects/cdk/>

<sup>(5)</sup> [https://github.com/baoilleach/openbabel-svn-mirror/blob/master/data/SMARTS\\_IntelLigand.txt](https://github.com/baoilleach/openbabel-svn-mirror/blob/master/data/SMARTS_IntelLigand.txt)

<sup>(6)</sup> <http://pele.farmbio.uu.se/nightly-1.3.1/cdk-javadoc-1.4.0/org/openscience/cdk/qsar/descriptors/molecular/KierHallSmartsDescriptor.html>

<sup>(7)</sup> [ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem\\_fingerprints.txt](ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt)

<sup>(8)</sup> <http://rdkit.org/docs/api/rdkit.Chem.MACCSkeys-pysrc.html>

## 2.1.2 Descritores moleculares 3D

Os descritores moleculares 3D são gerados através da estrutura 3D dos compostos e descrevem várias características da estrutura molecular. A metodologia mais aplicada é a Análise Comparativa do Campo Molecular (CoMFA) e outras abordagens similares. Nestes descritores 3D, a energia de interacção é descrita computacionalmente em milhares de pontos o que implica que estes modelos possuam um grau de complexidade elevado. Os descritores livres de alinhamento, como os do Volsurf<sup>®</sup> representam uma segunda geração de descritores 3D desenvolvidos para ultrapassar as desvantagens associadas aos anteriores. Os descritores do Volsurf<sup>®</sup> representam as propriedades globais da molécula como o tamanho, a forma, a interactividade, a polaridade, as distâncias moleculares, a hidrofília e a lipofília. Uma descrição pormenorizada dos 128 descritores Volsurf<sup>®</sup> encontra-se manual do programa<sup>(9)</sup> (Fontaine & Pastor, 2004; Molecular Discovery Ltd, 2008).

## 2.2 Métodos de *data mining* e *machine learning*

O rastreio virtual evoluiu, nos últimos anos, da tradicional análise de similaridade, utilizando um composto de referência, para a utilização de métodos de mineração de dados (*data mining*) e métodos de aprendizagem de máquina (*machine learning*) que requerem um conjunto de compostos de referência. Este facto resulta do aumento de informação sobre os dados de bioactividade dos compostos.

As técnicas de mineração de dados permitem explorar grandes quantidades de dados à procura de padrões consistentes e relações sistemáticas entre variáveis, auxiliando na tomada de decisões.

Os métodos de aprendizagem de máquina têm como objectivo construir modelos que permitam identificar relações entre estruturas químicas e a actividade observada, tornando-o num método muito popular nos rastreios virtuais (Geppert et al., 2009; Leach & Gillet, 2007).

---

<sup>(9)</sup> Molecular Discovery Ltd, 2008

## 2.2.1 Clustering

A análise de grupos (*clusters*) permite dividir um conjunto de objectos em *clusters*, de forma que objectos semelhantes entre si pertençam ao mesmo *cluster* enquanto objectos diferentes pertençam a outros *clusters* (Fig. 7).

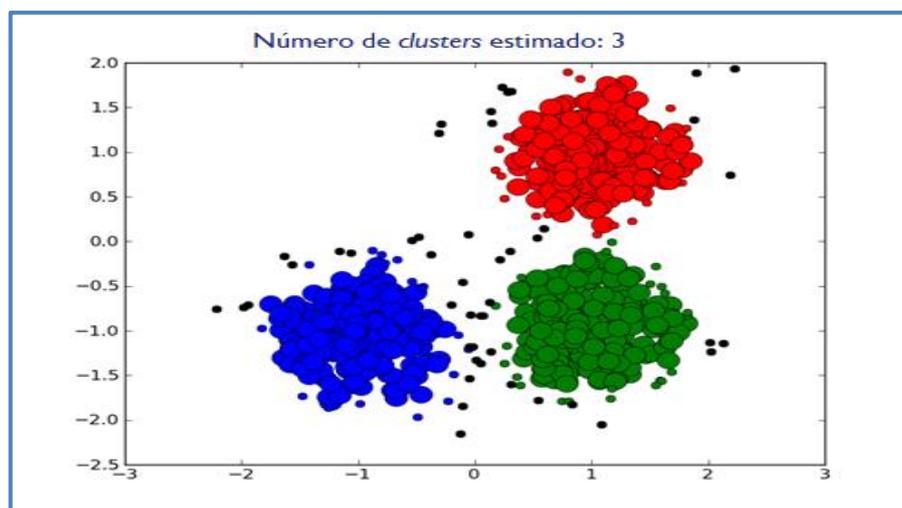


Figura 7. Representação esquemática de uma análise de *clustering* (Adaptado da página web: <http://scikit-learn.org/0.9/modules/clustering.html>).

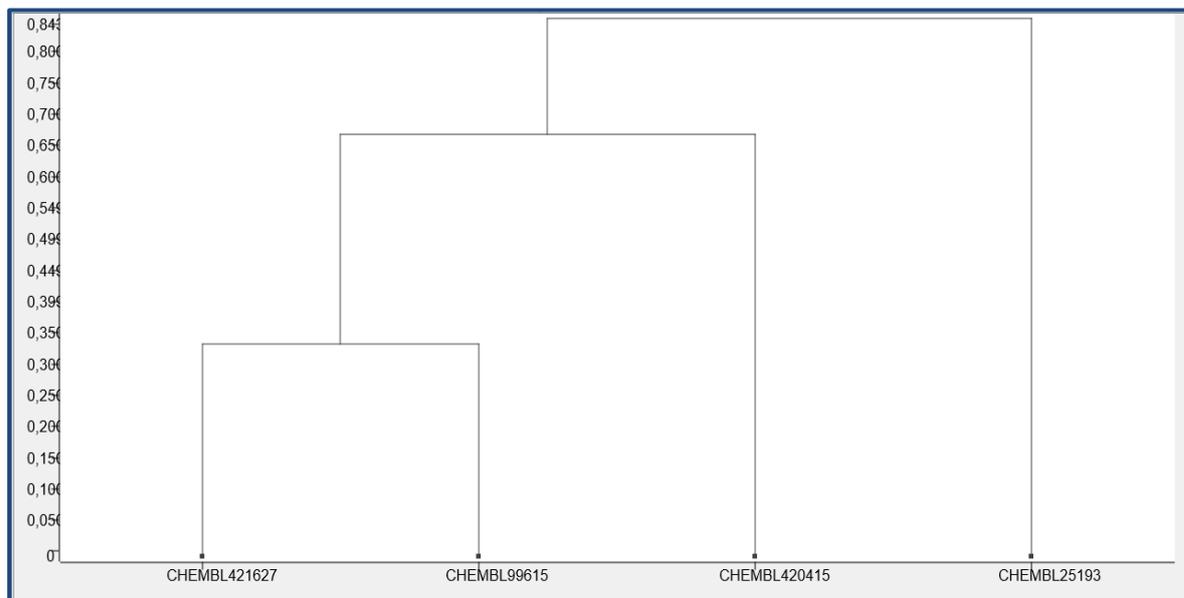
Para se efectuar esta análise é necessário:

1. Gerar os descritores moleculares para cada composto da base de dados;
2. Calcular a similaridade ou distância entre todos os compostos da base de dados com base nos descritores calculados;
3. Utilizar um algoritmo de *clustering* para agrupar os compostos.

Normalmente, a distância é calculada aplicando a fórmula  $1 - S$  ( $S$  corresponde ao coeficiente de similaridade) quando as moléculas são representadas por descritores moleculares binários, e pela distância euclidiana quando as moléculas são representadas por propriedades físico-químicas (ou derivados escalares).

A maioria dos métodos de *clustering* não implicam sobreposição, isto é, cada objecto pertence apenas a um único *cluster*. Os métodos de *clustering* podem ser divididos consoante a metodologia base do algoritmo (aglomerativo ou particional), consoante a estrutura do resultado (hierárquico e não hierárquico), consoante as características do espaço onde operam (abordagens sobre as propriedades ou sobre a similaridade) e consoante o tipo de *clusters* que são descobertos (globular ou transitivo) (Zhao & Karypis, 2005).

Os algoritmos particionais, como *K-means*, *K-medoids* e probabilístico encontram os *clusters* por particionamento dos dados num número predeterminado (ou derivado computacionalmente) de *clusters*. Os algoritmos aglomerativos hierárquicos, em que os resultados são apresentados através de dendrogramas (Fig. 8), encontram os *clusters*, atribuindo inicialmente cada objecto ao seu *cluster* e fundindo repetidamente pares de *clusters* até que um determinado ponto de paragem seja atingido (Zhao & Karypis, 2005).



**Figura 8. Exemplo de dendrograma.**

Nos algoritmos de *clustering* que operam sobre as propriedades, como o *K-means*, cada objecto é representado como um vector multidimensional de propriedades e o *clustering* é obtido otimizando repetidamente a distância entre o objecto e o centróide do *cluster*. Por outro lado, nos algoritmos baseados na similaridade, como o *K-medoids*, o *clustering* é obtido através do cálculo inicial das semelhanças entre todos os objectos que determinam, secundariamente, o resultado do *clustering*.

Nos *clusters* globulares, verifica-se um forte padrão de conservação nos sub-*clusters*. Por sua vez, nos *clusters* transitivos este padrão de conservação não é tão forte, o que leva que um objecto de um sub-*cluster* partilhe propriedades que ocorrem noutra sub-*cluster* (Zhao & Karypis, 2005; Leach & Gillet, 2007)

### 2.2.1.1 *Clustering* com base na estrutura dos compostos

A pesquisa da subestrutura máxima comum (MCS) entre duas moléculas baseia-se em algoritmos centrados em grafos. Estes algoritmos verificam a similaridade entre duas

estruturas e encontram o máximo grafo/subestrutura comum entre as duas moléculas (Raymond & Willet, 2002; Cao *et al.*, 2008).

Um grafo é um conceito usado para descrever a relação entre “objectos”, sendo representado graficamente por uma figura com vértices, os “objectos”, unidos por um traço/aresta que representa a relação entre eles (Fig. 9) (Diestel, 2000; Sousa *et al.*, 2013).

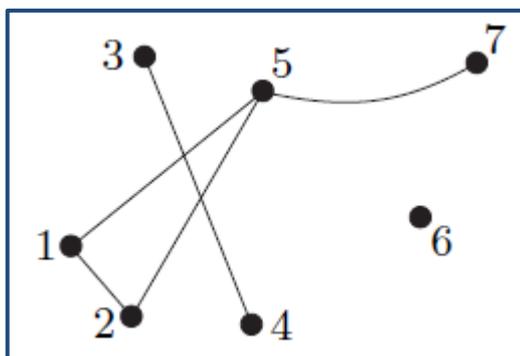


Figura 9. Exemplo esquemático de um grafo (Adaptado de Diestel, 2000).

Na Química, se considerarmos a estrutura 2D de uma molécula, os vértices do grafo representam os átomos, e as arestas revelam as ligações covalentes entre eles (Raymond & Willet, 2002). Dois grafos são considerados isomorfos se existir uma correspondência de 1 para 1 entre os seus vértices e se a aresta existir entre os vértices em causa de ambos os grafos. As estratégias de pesquisa de MCS podem seguir dois tipos de critérios: maior número de vértices (átomos) em comum (MCIS, do inglês *maximum common induced subgraph*), ou maior número de arestas em comum (MCES, do inglês *maximum common edge subgraph*).

Comparando a pesquisa de MCS com os métodos baseados em descritores moleculares, a primeira possui a vantagem de identificar localmente quais as semelhanças existentes ao contrário dos métodos baseados em descritores, sendo útil na descoberta e desenvolvimento de fármacos, nomeadamente em estudos de *docking* de proteínas-ligando, na previsão de actividade biológica e na modelação do sítio activo (Raymond & Willet, 2002; Cao *et al.*, 2008).

A LibMCS é uma ferramenta de pesquisa de MCS de forma hierárquica, que se encontra disponível através do *software* da ChemAxon. Esta biblioteca de pesquisa segue o critério MCES para encontrar as MCS.

### 2.2.1.2 Clustering com base em fingerprints

A medida mais comum para comparar *fingerprints* 2D é o coeficiente de Tanimoto que é dado pela equação 1:

$$T_{AB} = \frac{c}{a+b-c} \quad (\text{Eq. 1})$$

onde  $a$  e  $b$  são o número de *bits* com valor 1 definidos nas *fingerprints* das moléculas A e B, respectivamente, e onde  $c$  é o número de *bits* comuns em ambas as *fingerprints*, providenciando uma maneira rápida de calcular similaridade molecular (Godden *et al.*, 2005; Venkatraman *et al.*, 2010). Este coeficiente é sensível ao tamanho das moléculas uma vez que, geralmente, moléculas pequenas produzem coeficientes de Tanimoto menores do que moléculas maiores (Hattori *et al.*, 2008)

A escolha de linha de corte (no Inglês, *threshold*) para avaliar a similaridade é um factor crítico quando se procura por *fingerprints* em grandes bases de dados. Existem vários valores que são apontados como o limiar ideal (ex. 0,85 ou 0,80), mas cada vez mais estudos indicam que não um valor ideal e que este muda consoante as *fingerprints* (Godden *et al.*, 2005; Dimova *et al.*, 2013).

### 2.2.2 Modelos de classificação

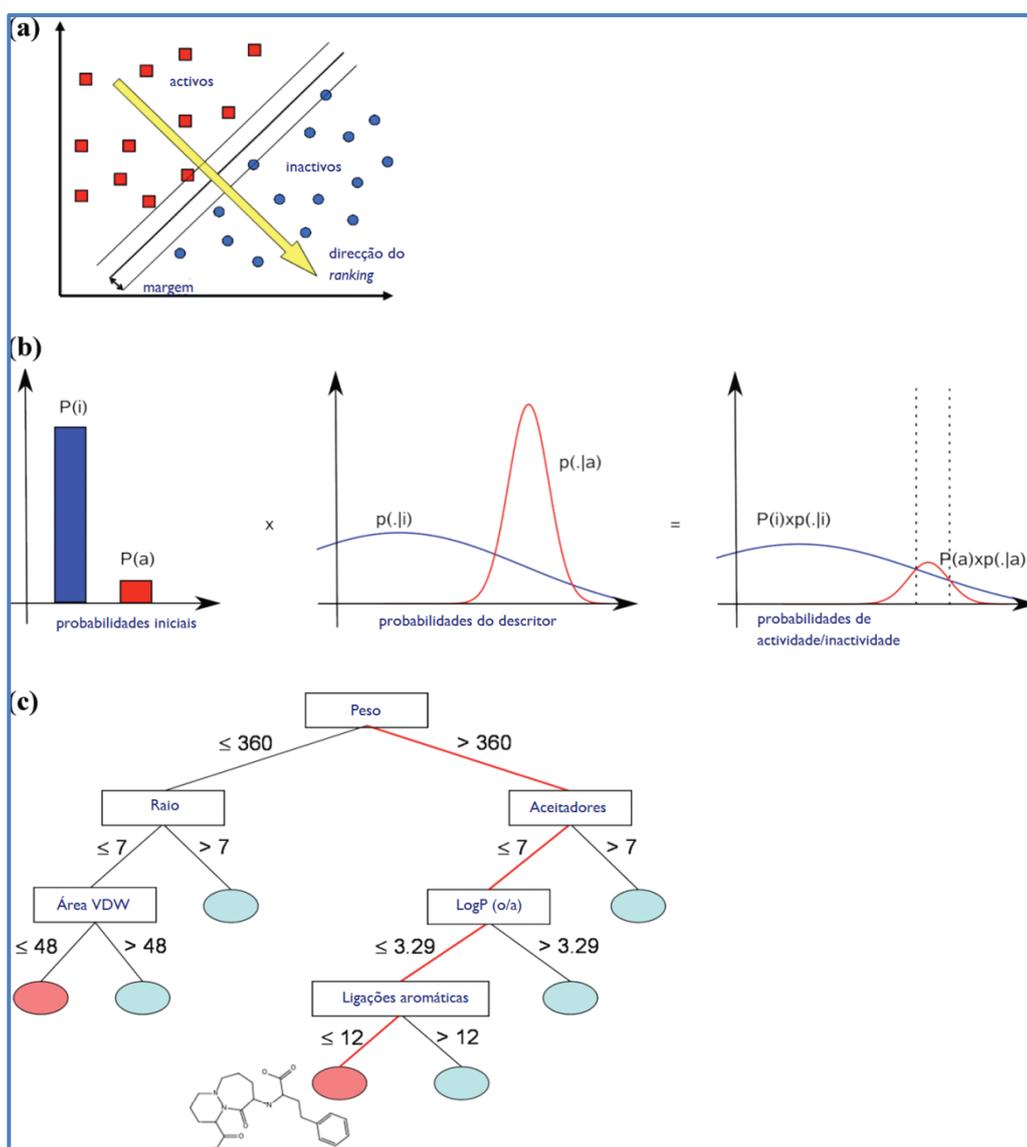
Para construir um modelo de classificação, as moléculas são classificadas como “activas” ou “inactivas”, ou divide-se a sua actividade em pequenos grupos (ex. “muito boa”, “média”, “baixa”) em vez de utilizar o valor numérico da actividade obtido experimentalmente. Contudo, no caso dos problemas de regressão, utilizam-se valores numéricos de actividade, para prever valores de actividade de outras moléculas.

Globalmente, para qualquer método de aprendizagem de máquina, seja ele de classificação ou de regressão, o processo inicia-se com a análise dos compostos cuja actividade se conhece, seguindo-se o processo de treino a partir dos compostos conhecidos, de modo a estabelecer uma relação entre os valores dos atributos previsão e as classes (Prompramote *et al.*, 2005; Sousa *et al.*, 2013). O classificador resultante permite então prever classes de compostos que estejam descritos por valores de atributos de previsão, mas cuja classe não é conhecida (Kotsiantis, 2007; Cunningham *et al.*, 2008). De seguida, são avaliados os resultados dos métodos de aprendizagem de máquina e, caso sejam relevantes, o método é aplicado em novas bases de dados com vista a detectar potenciais moléculas

com a actividade pretendida (Prompramote *et al.*, 2005; Sousa *et al.*, 2013). O objectivo desta técnica é, assim, extrair um modelo computacional que permita prever a classe ou valor de actividade de novas estruturas (Leach & Gillet, 2007).

Dos diferentes tipos de aprendizagem que têm sido explorados, três abordagens têm dominado este campo (Fig. 10) (Geppert *et al.*, 2009):

- Máquinas de vectores de suporte (SVM, do inglês, *support vector machines*);
- Classificadores Bayesianos;
- Árvores de decisão.



**Figura 10. Algoritmos populares de *data mining*. (a) SVM, (b) Classificadores Bayesianos e (c) Árvores de decisão (Adaptado de Geppert *et al.*, 2009).**

O desempenho dos métodos de aprendizagem de máquina depende de vários factores como a diversidade do conjunto de dados de treino e a capacidade dos métodos para lidar com bases de dados desequilibradas (número de compostos inactivos superam largamente o número de compostos activos). Desta forma, modelos de aprendizagem de máquina capazes de rastrear grandes bases de dados com bom rendimento e baixa taxa de falsos-positivos podem ser desenvolvidos utilizando um conjunto de dados de treino bastante diversificado e com compostos inactivos na sua constituição (Lavecchia & Di Giovanni, 2013).

No presente trabalho foram utilizados os algoritmos de classificação *Naïve Bayes* e SVM para verificar qual deles possui melhor capacidade preditiva de futuros inibidores da 5 $\alpha$ -R1 e 5 $\alpha$ -R2, utilizando como conjunto de treino as moléculas obtidas na ChEMBL. Esta escolha prende-se com a grande utilização e bons resultados que estes dois métodos proporcionam na classificação de bases de dados de compostos facilitando o desenho e desenvolvimento de novos fármacos (Eckert & Bajorath, 2007; Geppert et al., 2009).

### 2.2.2.1 *Naïve Bayes*

O teorema de Bayes foi o ponto de partida para o desenvolvimento dos métodos de classificação Bayesianos. Este teorema relaciona a probabilidade condicional e o seu inverso (Eq. 2):

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \quad (\text{Eq. 2})$$

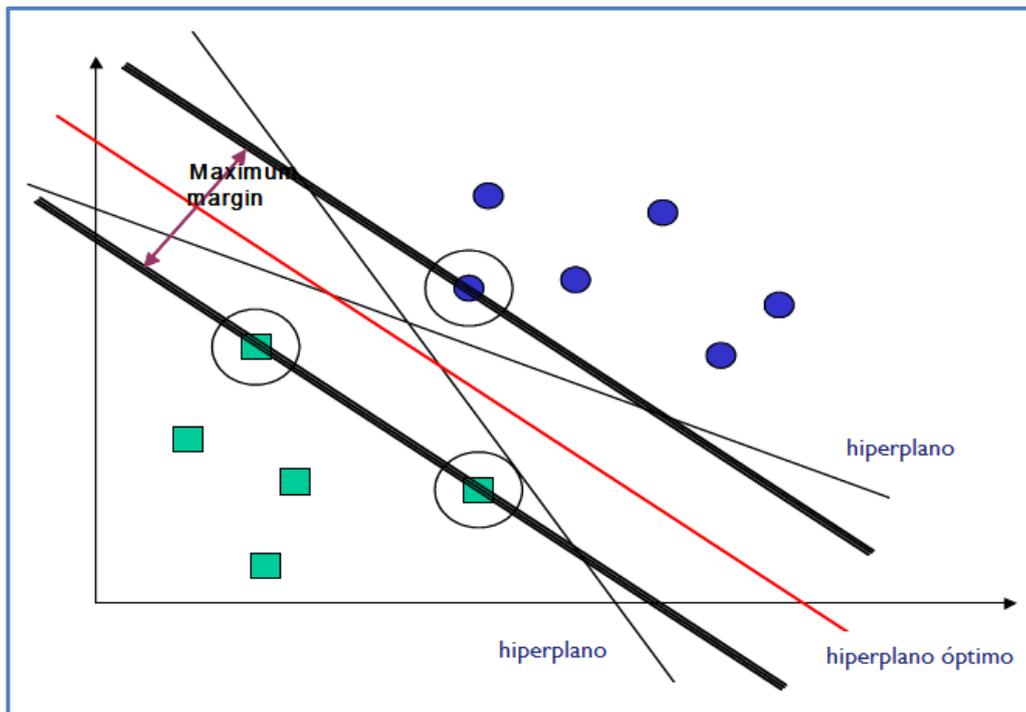
onde, P(A) indica a probabilidade *a priori* (ou marginal) de A, isto é, não tem em conta qualquer informação de B; P(A|B) é a probabilidade condicional de A, dado B; P(B|A) é a probabilidade condicional de B, dado A; e P(B) é a probabilidade *a priori* de B (Okeh & Ugwu, 2008).

Os métodos Bayesianos utilizados nos rastreios virtuais, salvo diferenças algorítmicas, derivam a probabilidade da actividade dos compostos em estudos com base nas suas representações numéricas (descritores de propriedades ou *fingerprints*) (Fig. 8). Uma vez que os classificadores Bayesianos estimam a probabilidade da actividade, tornam-se úteis para classificar compostos presentes em bases de dados de acordo com essa actividade. Um dos métodos Bayesianos mais populares é o *Naïve Bayes* (Geppert et al., 2009).

O classificador *Naive Bayes* assume que todas as características dos compostos são independentes entre si no contexto da sua classe, daí a denominação de *naïve* (Kim et al., 2006; Kotsiantis, 2007, Mitchell, 2010)

### 2.2.2.2 Máquinas de vectores de suporte

A metodologia de máquinas de vectores de suporte (SVM), mais recente que os classificadores Bayesianos, tornou-se popular na década de 1990 tendo sido desenvolvida por Cortes e Vapnik (1995). Esta técnica baseia-se na ideia de construir um hiperplano num espaço hiperdimensional de modo a separar e classificar os dados em duas margens (Fig. 11). A metodologia SVM assenta na teoria de aprendizagem estatística, uma vez que tem em consideração dois objectivos que colidem nos métodos de aprendizagem de máquina: (i) encaixar correctamente o conjunto de treino e (ii) generalizar o suficiente para que o conjunto de teste externo possa ser classificado. Ou seja, quanto mais complexa a estrutura de um modelo de treino melhor encaixará num conjunto de treino, mas mais dificilmente generalizará para um conjunto de teste (Geppert et al., 2009). A construção de um hiperplano que maximize a margem de separação,  $\rho$ , entre o hiperplano e o ponto mais próximo de ambas as classes possibilita otimizar a performance de generalização do classificador e diminuir o erro, obtendo-se um hiperplano óptimo (Kotsiantis, 2007; Geppert et al., 2009).



**Figura 11. Representação esquemática de um hiperplano ótimo de separação de padrões linearmente separáveis. Adaptado de Kotsiantis, 2007.**

Assim, para um problema de classificação binária, onde os compostos são divididos entre activos e inactivos. Um conjunto de treino diz-se linearmente separável se for possível separar os padrões dos compostos de classes diferentes por pelo menos um hiperplano. Se considerarmos um conjunto de treino  $\{x_i, y_i\}$  ( $i = 1, \dots, n$ ), onde  $x_i \in \mathbb{R}^d$  corresponde à representação da molécula através de uma *fingerprint* (vector de *bits*) e  $y_i \in \{-1, +1\}$  corresponde à identificação da classe (activo ou inactivo), para derivar um hiperplano  $H$  que melhor separe os positivos dos negativos no conjunto de treino, temos a equação 3:

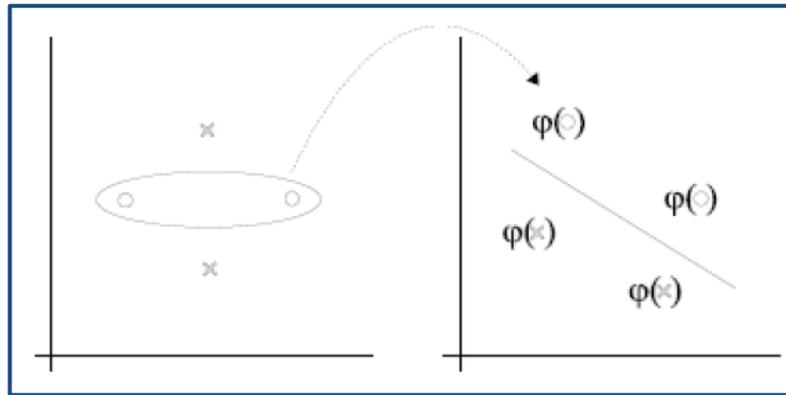
$$H: (x, w) + b = 0 \quad (\text{Eq. 3})$$

onde, está definido o hiperplano  $H$  é definido pelo produto escalar entre os vectores  $x$  e  $w$  ( $x, w$ ), em que  $x$  é um vector que representa os padrões de entrada do conjunto de treino,  $w$  é um vector normal e  $b$  é um limiar conhecido como *bias* (Wasserman *et al.*, 2009). A equação 3 pode ser reescrita como (Eq. 4 e 5):

$$w^T x_i + b \geq 1, \text{ para todos os } x_i \in P \quad (\text{Eq. 4})$$

$$w^T x_i + b \leq -1, \text{ para todos os } x_i \in N \quad (\text{Eq. 5})$$

No caso das classes do conjunto de treino não serem linearmente separáveis no espaço original é possível introduzir dimensões adicionais para permitir a separação linear do conjunto de treino (Wasserman *et al.*, 2009). Estas dimensões adicionais são introduzidas com um mapeamento não linear (Fig. 12) através de um produto interno (*kernel*) (Eckert & Bajorath, 2007; Wasserman *et al.*, 2009; Sousa *et al.*, 2013).



**Figura 12. Representação esquemática do mapeamento de características (Adaptado de Sousa *et al.*, 2013).**

Para problemas de classificação de multiclasse o SVM também é uma ferramenta poderosa e adaptada a este tipo de desafios uma vez que generaliza bem dados de elevadas dimensões (Prompramote *et al.*, 2005; Geppert *et al.*, 2009).

### 2.2.2.3 Medidas de desempenho

As metodologias para analisar e avaliar estatísticas de resultados gerados por métodos de aprendizagem de máquina podem variar de acordo com os estudos, o que dificulta uma comparação dos resultados entre estudos que utilizem métricas diferentes (Sousa *et al.*, 2013).

No presente estudo foram utilizadas a *accuracy*, coeficiente *Cohen's kappa*, sensibilidade, especificidade, precisão e *F-score* como medidas estatísticas para avaliar o desempenho dos classificadores. A *accuracy* (Eq. 6) indica a eficácia global do método em prever positivos correctamente. O *Cohen's kappa* indica o sucesso do método em prever positivos correctamente para cada classe e depois agrega-o num coeficiente. A sensibilidade (Eq. 7) que também pode ser designada de *Recall* mede a proporção de positivos que se previu correctamente. Ao invés, a especificidade (Eq. 8) indica a percentagem de negativos que foram correctamente previstos. A precisão (Eq. 9) demonstra o poder preditivo do algoritmo ao indicar os positivos e negativos previstos correctamente. A *F-score* (Eq. 10) estabelece uma relação entre a precisão e a sensibilidade (Sokolova et al., 2006; García et al., 2009).

$$Accuracy = \frac{Vp+Vn}{Vp+Fp+Fn+Vn} \quad (\text{Eq. 6})$$

$$Sensibilidade = \frac{Vp}{Vp+Fn} \quad (\text{Eq. 7})$$

$$Especificidade = \frac{Vn}{Fp+Vn} \quad (\text{Eq. 8})$$

$$Precisão = \frac{Vp}{Vp+Fp} \quad (\text{Eq. 9})$$

$$F - score = \frac{(\beta^2+1) \times precisão \times sensibilidade}{\beta^2 \times precisão + sensibilidade} \quad (\text{Eq. 10})$$

Nas equações apresentadas, *Vp* representa o número de verdadeiros positivos (positivos correctamente previstos), *Vn* indica o número de verdadeiros negativos, *Fp* indica o número de falsos positivos e *Fn* representa o número de falsos negativos.

### 3. Protocolo computacional

O programa KNIME (Konstanz Information Miner)<sup>(10)</sup>, uma interface gráfica que permite construir, através de nós, vários processos utilizados em rastreamento virtual, entre os quais criação de *fingerprints*, análise de *clusters*, criação de algoritmos de aprendizagem de máquina, entre outras funcionalidades, foi utilizado para a maior parte das tarefas de análise aqui apresentadas.

#### 3.1 Clustering

O *clustering* baseado na estrutura dos compostos foi realizado recorrendo ao programa LibMCS (ChemAxon®). Esta análise foi realizada para três conjuntos de compostos: inibidores da classe “Excelente” (Tabela 3) da 5 $\alpha$ -R1 e da 5 $\alpha$ -R2, e inibidores comuns à 5 $\alpha$ -R1 e 5 $\alpha$ -R2.

Paralelamente, para estes conjuntos de compostos foi realizada a análise de *clustering* hierárquico com base nas 10 fingerprints calculadas, utilizando o programa KNIME (Fig. 13). A distância entre as fingerprints foi avaliada utilizando o coeficiente de Tanimoto e uma linha de corte de 0,30.

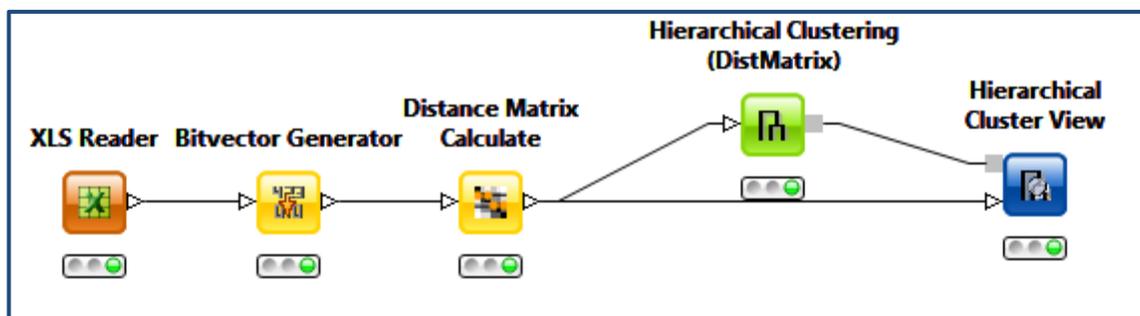


Figura 13. Análise de *clustering* para as várias *fingerprints* utilizando o KNIME.

#### 3.2 Modelos de classificação

Com o objectivo de verificar qual representação – *fingerprints* ou descritores 3D – produz o melhor classificador para os inibidores da 5 $\alpha$ -redutase e melhor permite realizar o rastreamento virtual para encontrar novos inibidores em bases de dados diferentes, foram aplicados dois métodos de aprendizagem de máquina: *Naïve Bayes* e SVM. Para isso foi necessário converter os valores numéricos de IC50 para classes de IC50. Foram testadas várias discretizações dos valores de IC50 (Tabelas 3 a 6) que permitiram produzir

<sup>(10)</sup> <https://www.knime.org/>

classificadores multiclasse (Tabelas 3 e 4) e binários (Tabela 5) para inibidores da 5 $\alpha$ -R1 e 5 $\alpha$ -R2, e classificadores multiclasse (Tabela 6) para inibidores duais da 5 $\alpha$ -Redutase.

**Tabela 3. Atribuição de 4 tipos classes aos valores de IC50.**

Classes de IC50	Valores de IC50 (nM)	Compostos	
		5 $\alpha$ -R1	5 $\alpha$ -R2
Excelente	]0;5]	32	141
Muito bom	]5;20]	64	19
Bom	]20;100]	54	42
Fraco	]100;+ $\infty$ [	160	153
<b>Total</b>		310	355

A discretização apresentada na Tabela 3 difere da apresentada na Tabela 4 apenas na definição das classe Fraco e Outros.

**Tabela 4. Atribuição de 5 tipos de classes aos valores de IC50.**

Classes de IC50	Valores de IC50 (nM)	Compostos	
		5 $\alpha$ -R1	5 $\alpha$ -R2
Excelente	]0;5]	32	141
Muito bom	]5;20]	64	19
Bom	]20;100]	54	42
Fraco	]100;1000]	102	97
Outros	]1000;+ $\infty$ [	58	56
<b>Total</b>		310	355

Para verificar se uma análise de classificação binária permitiria obter melhores resultados que a análise de classificação multiclasse, foram utilizadas as classes descritas na tabela 5:

**Tabela 5. Atribuição de classificação binária aos valores de IC50.**

Classes de IC50	Valores de IC50 (nM)	Compostos	
		5 $\alpha$ -R1	5 $\alpha$ -R2
Forte	]0;100]	150	202
Fraco	]100;+ $\infty$ [	160	153
<b>Total</b>		310	355

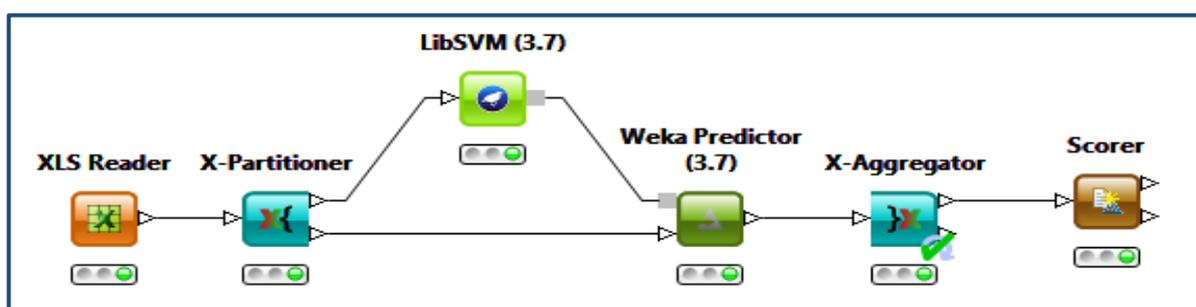
Com vista a detectar potenciais inibidores duais ou selectivos para uma das 5 $\alpha$ -redutases foi realizada uma aprendizagem de máquina utilizando como conjunto de treino os

181 inibidores duais com valores actividade para as duas isoenzimas fornecidos pela ChEMBL e cujas classes são definidas de acordo com a tabela 6:

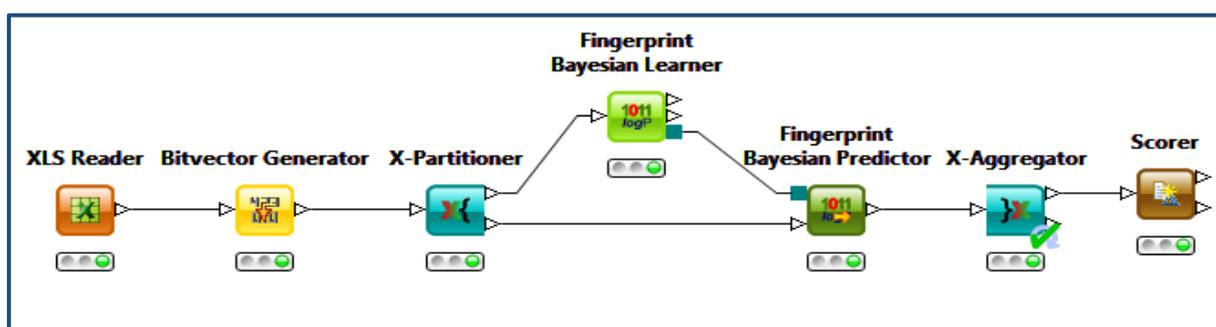
**Tabela 6. Divisão por classes dos inibidores duais.**

Classes de IC50	Classe de IC50 para 5 $\alpha$ -R1	Classe de IC50 para 5 $\alpha$ -R2	Compostos
Forte/Forte	Forte	Forte	58
Forte/Fraco	Forte	Fraco	46
Fraco/Forte	Fraco	Forte	23
Fraco/Fraco	Fraco	Fraco	54
<b>Total</b>			181

Todos os classificadores foram treinados e avaliados utilizando validação cruzada, ou seja, após a divisão do conjunto de dados em dois, o modelo é repetidamente treinado com uma parte dos dados e validado com a outra (Meinl *et al.*, 2009; Refaeilzadeh *et al.*, 2012). Na validação cruzada utilizada, os dados foram divididos em amostragens aleatórias com um número de validações,  $k$  ( $k=5$ ). (Meinl *et al.*, 2009). Nas figuras seguintes estão representados os *workflows* do KNIME para obter os classificadores com os métodos de aprendizagem de máquina SVM e *Naïve Bayes* com validação cruzada quando se utilizam *fingerprints* (Fig. 14 e 15) e quando se utilizam descritores 3D, uma vez que, no caso do método *Naïve Bayes*, os nós utilizados na validação cruzada são diferentes (Fig. 16).



**Figura 14. Workflow do KNIME para o método SVM.**



**Figura 15. Workflow do KNIME para o método Naïve Bayes, quando se utiliza fingerprints.**

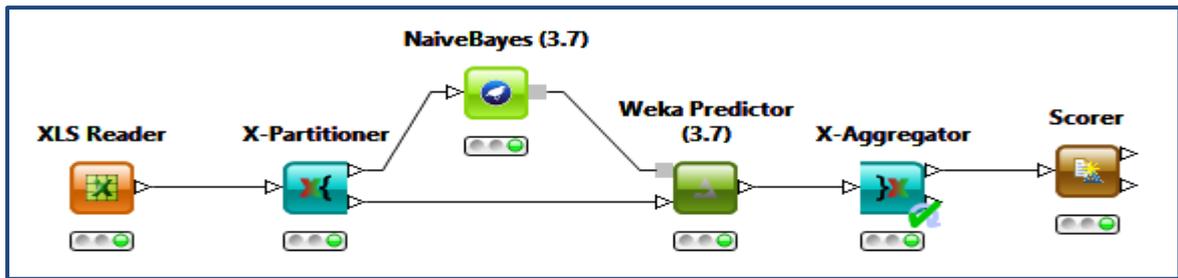


Figura 16 Workflow do KNIME para o método Naïve Bayes, quando se utilizam descritores 3D.

A normalização dos dados foi necessária no caso dos SVMs, e foi calculada utilizando a normalização z-score onde os valores de cada atributo são normalizados utilizando a média e o desvio padrão de cada atributo de acordo com a equação 10:

$$d' = \frac{d - \bar{x}(P)}{\sigma(P)} \quad (\text{Eq. 10})$$

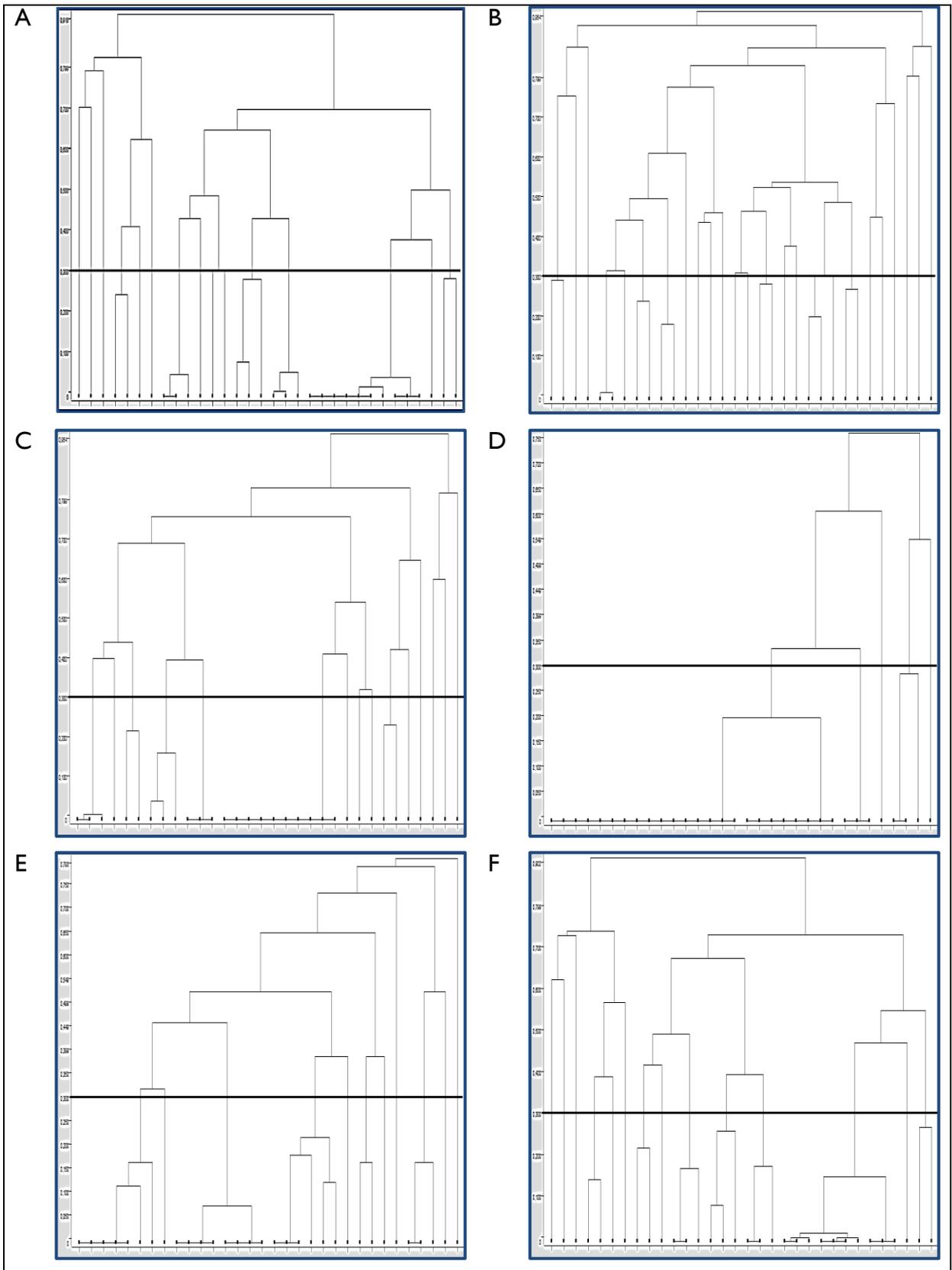
onde,  $\bar{x}(P)$  indica a média de todos os atributos de P e  $\sigma(P)$  indica o desvio padrão de todos os atributos de P (Saranya & Manikandan, 2013). No caso do método de aprendizagem de máquina *Naïve Bayes* não é necessária normalização dos dados.

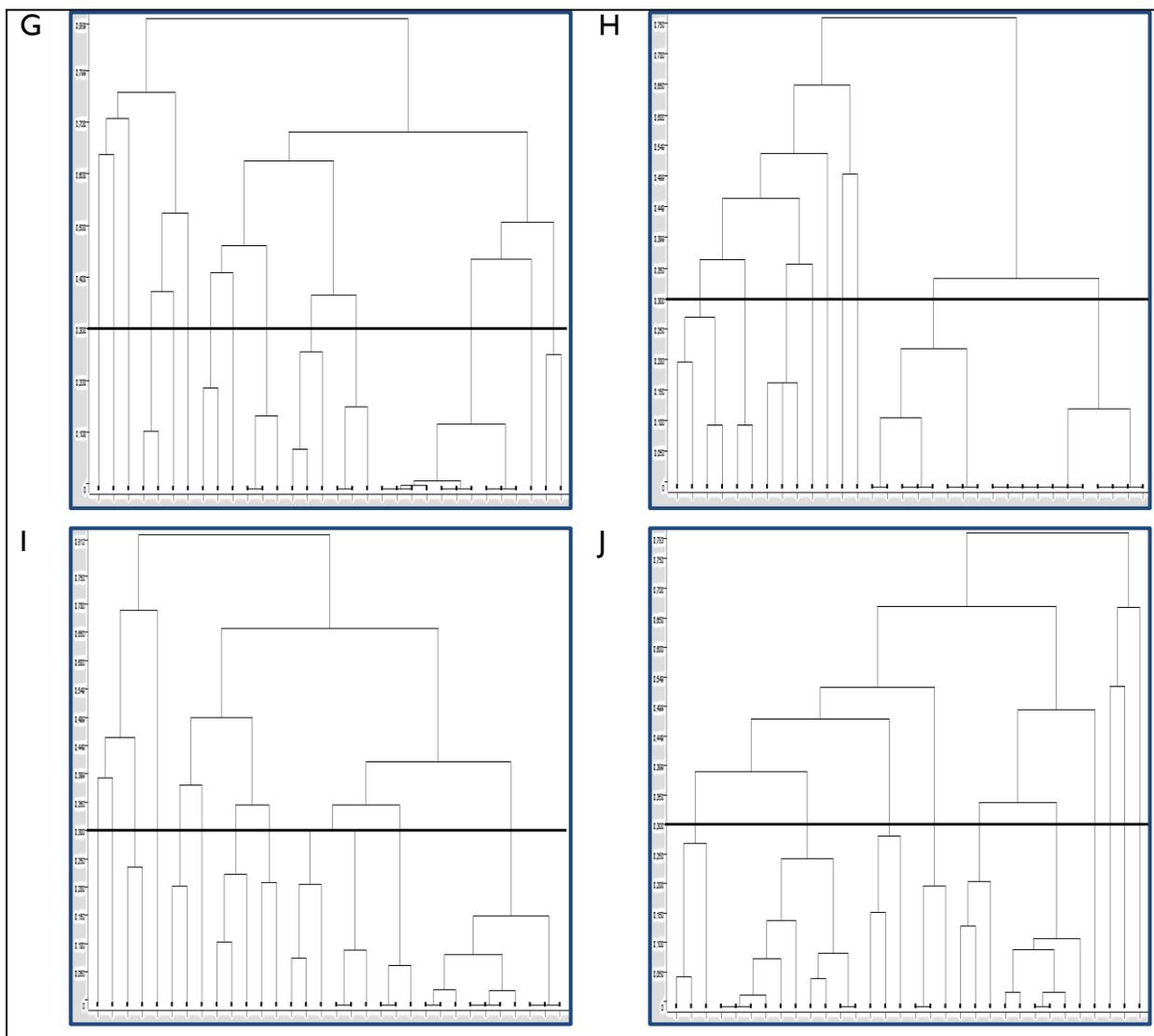


### **III. Resultados**

## I. Comparação dos *clusters* obtidos a partir das diferentes *fingerprints*

Com recurso ao KNIME procedeu-se ao *clustering* hierárquico das *fingerprints*, com respectiva análise dos resultados. Nesta análise foram utilizadas apenas as moléculas com classe de IC50 considerada excelente, na classificação multiclasse. Este motivo relaciona-se com o facto de ser esta a classe teoricamente, com maior potencial de interesse, permitindo verificar se existem diferenças significativas entre os *clusters* obtidos a partir da representação com as *fingerprints* utilizadas. Para os 32 inibidores da 5 $\alpha$ -RI classificados como pertencentes à classe de IC50 excelente obtiveram-se 10 dendrogramas diferentes, correspondendo a cada uma das *fingerprints* (Fig. 17). Os *clusters* foram analisados para uma linha de corte (*threshold*) de 0,30. Para as CHF da ChemAxon<sup>®</sup> foi possível dividir os compostos em 15 *clusters* (Fig. 17, A). Relativamente às ECFP, obtiveram-se 25 *clusters* (Fig. 17, B). No caso das FP2, da OpenBabel<sup>®</sup> verificou-se a existência de 15 *clusters* (Fig. 17, C). Para as *fingerprints* FP3, verificou-se a existência de 5 *clusters* (Fig. 17, D). Relativamente às FP4, os compostos foram separados por 11 *clusters* (Fig. 17, E). No que concerne às *fingerprints* obtidas através do CDK<sup>®</sup>, as *standard fingerprints* indicaram a presença de 14 *clusters* (Fig. 17, F). Para as *Extended fingerprints*, verificou-se a existência de 14 *clusters* (Fig. 17, G). As *EState fingerprints* indicaram a ocorrência de 9 *clusters* (Fig. 17, H). No que respeita às *PubChem fingerprints*, a análise realizada indicou a presença de 12 *clusters* (Fig. 17, I). Por último, nas MACCS, foram detectados 10 *clusters* (Fig. 17, J).

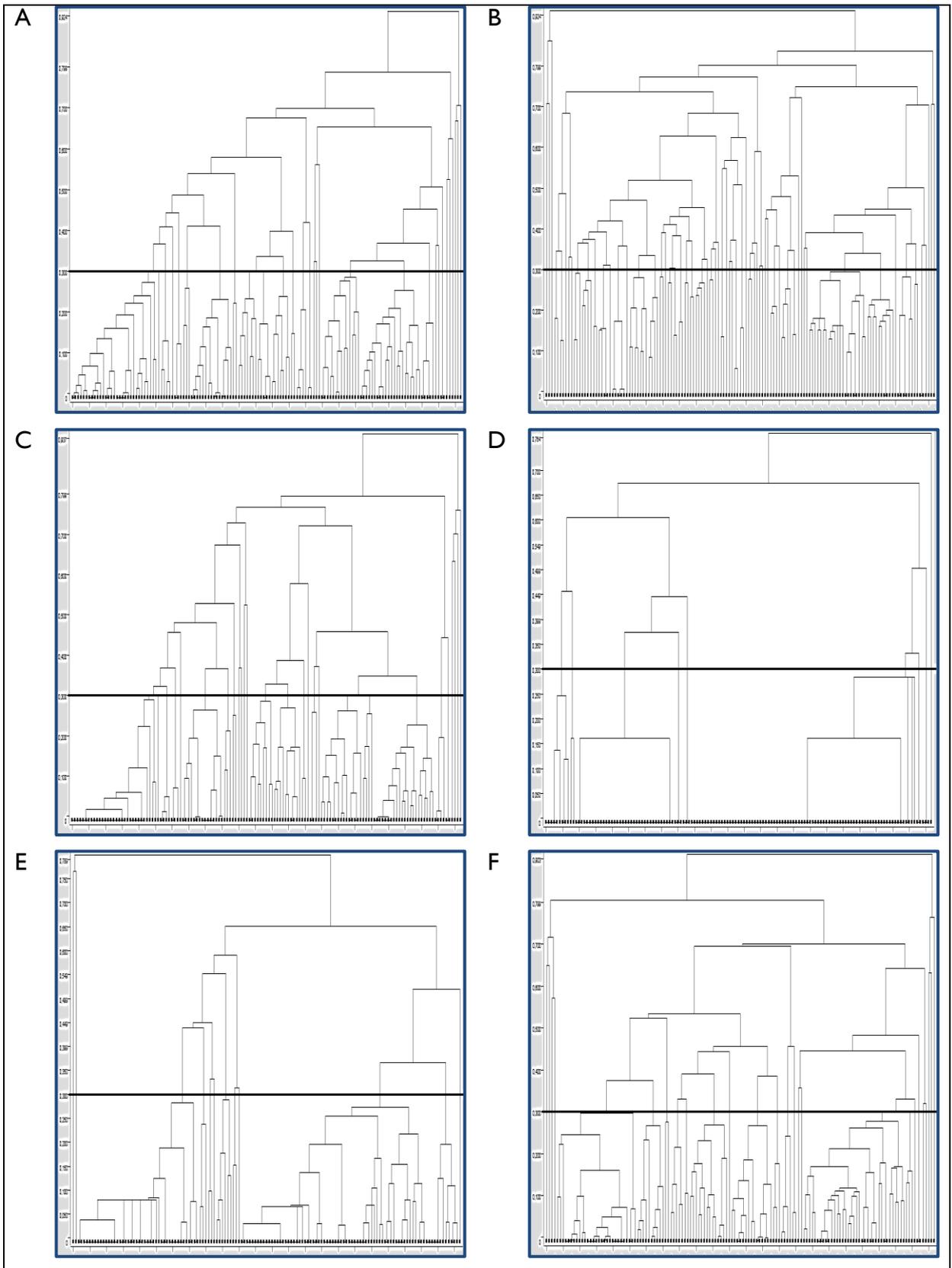


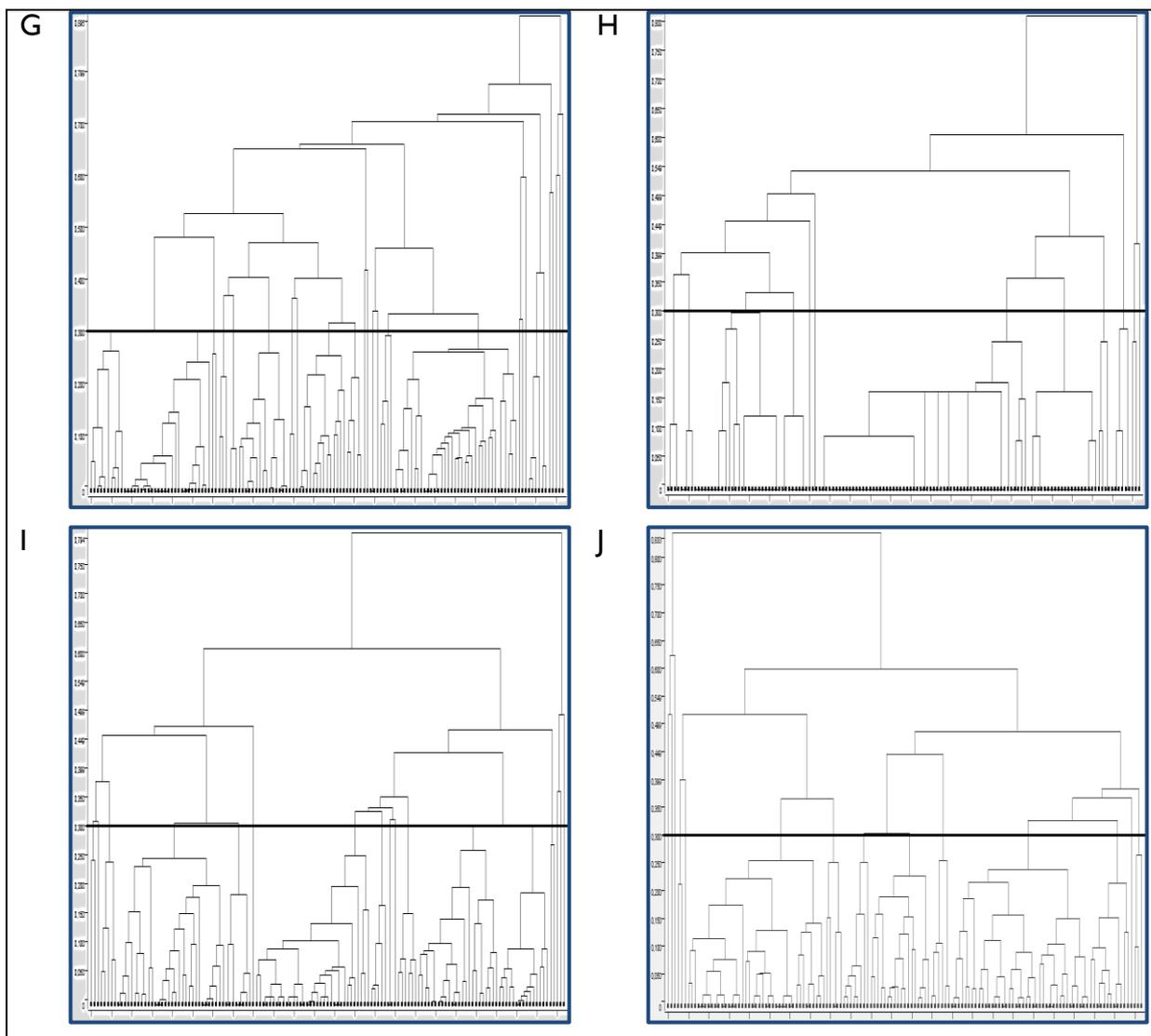


**Figura 17.** Dendrogramas obtidos do *clustering* hierárquico das *fingerprints* CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), *Standard* (F), *Extended* (G), *EState* (H), *PubChem* (I) e *MACCS* (J) que representam os 32 inibidores da 5 $\alpha$ -RI classificados como pertencentes à classe de IC50 excelente, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos *clusters* foi 0,30.

Para os 141 inibidores da 5 $\alpha$ -R2 classificados como pertencentes à classe de IC50 excelente obtiveram-se, também 10 dendrogramas diferentes. Para as CHF da ChemAxon<sup>®</sup> foi possível dividir os compostos em 29 *clusters*, de acordo com a linha de corte 0,30 escolhida para verificar o número de *clusters* formados (Fig. 18, A). No caso das ECFP, a análise separou os 141 compostos em 61 *clusters* (Fig. 18, B). Relativamente aos 141 compostos descritos através da FP2 da OpenBabel<sup>®</sup>, foram apurados 27 *clusters* (Fig. 18, C). No que respeita à análise realizada aos compostos representados pelas *fingerprints* FP3, estes foram divididos em 9 *clusters* (Fig. 18, D). A análise dos 141 compostos representados pelas *fingerprints* FP4 indicou a presença de 12 *clusters* (Fig. 18, E). Para os 141 compostos representados através das *standard fingerprints* da CDK<sup>®</sup>, verificou-se a existência de 26 *clusters* (Fig. 18, F). No caso da análise realizada aos 141 compostos representados pelas

*extended fingerprints*, estes foram agrupados em 26 *clusters* (Fig. 18, G). Para os 141 compostos, representados pelas *EState fingerprints*, analisados, verificou-se uma distribuição dos compostos por 12 *clusters* (Fig. 18, H). Os resultados da análise aos 141 compostos representados pela *PubChem fingerprint* indicaram a presença de 17 *clusters* (Fig. 18, I). Por último, a análise realizada aos compostos representados pela MACCS, indicou a separação dos compostos em 14 *clusters* (Fig. 18, J).

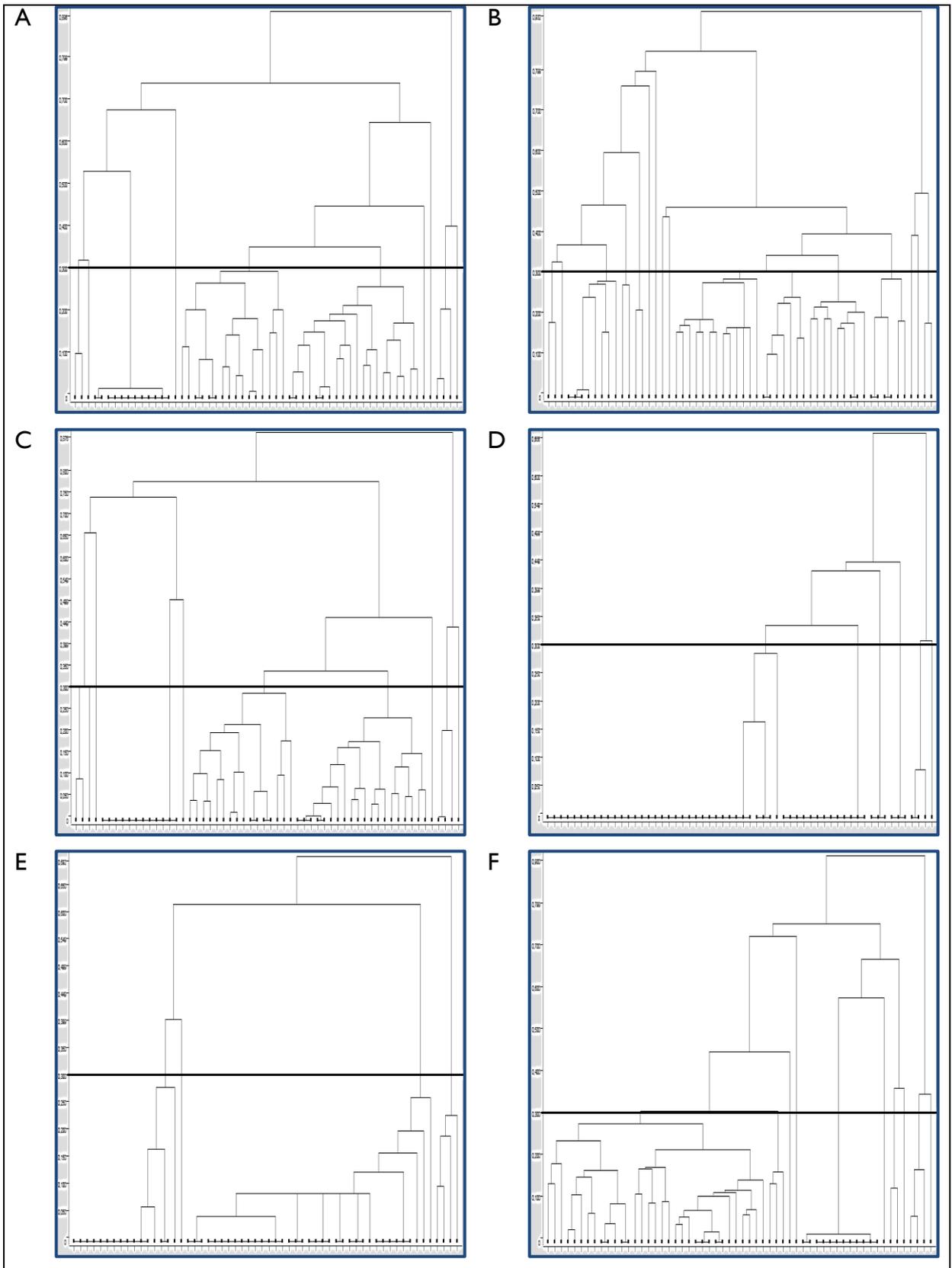


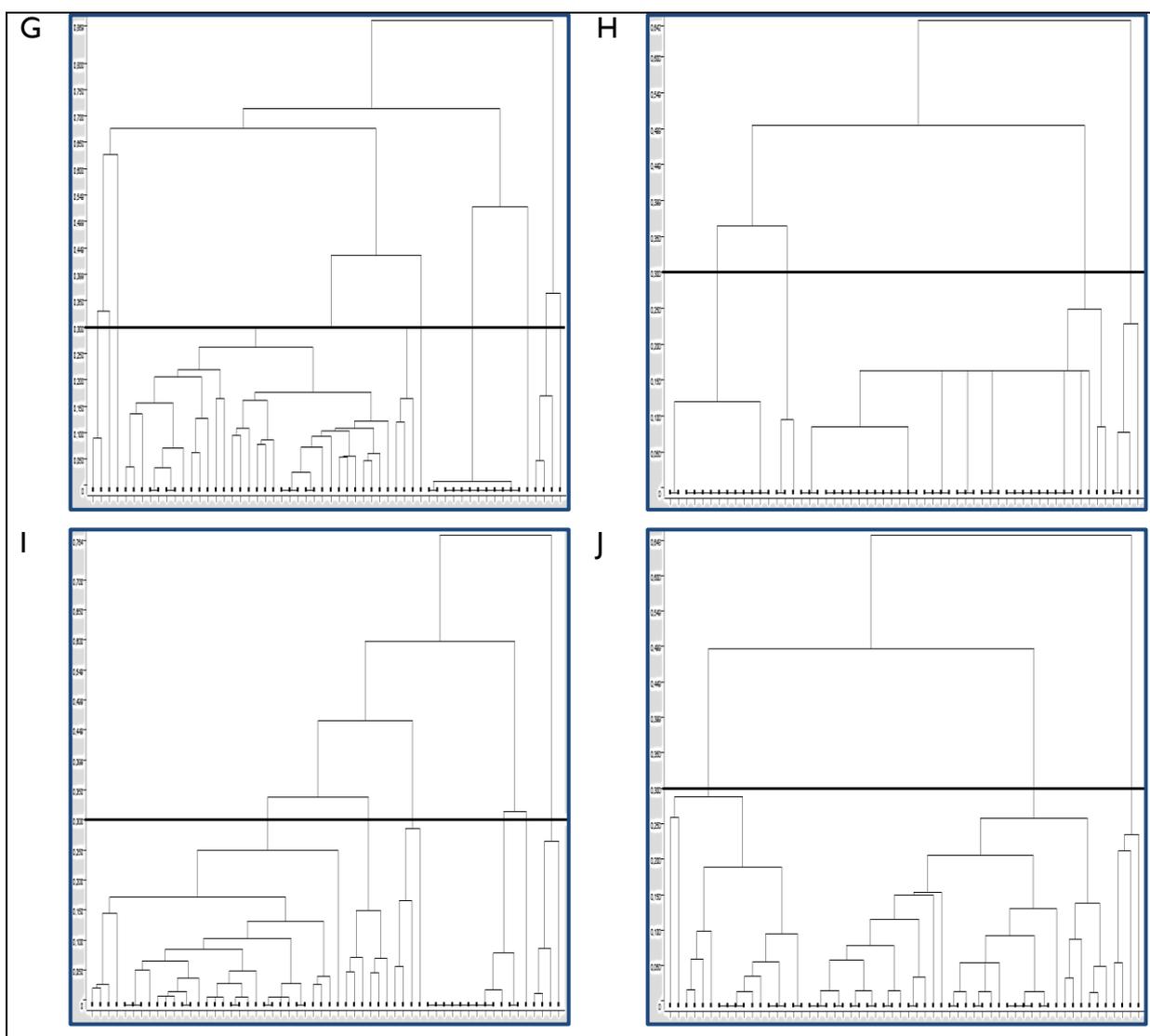


**Figura 18.** Dendrogramas obtidos do *clustering* hierárquico das *fingerprints* CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), *Standard* (F), *Extended* (G), *EState* (H), *PubChem* (I) e *MACCS* (J) que representam os 141 inibidores da 5 $\alpha$ -R2 classificados como pertencentes à classe de IC50 excelente, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos *clusters* foi 0,30.

Para os 58 inibidores das duas isoenzimas da 5 $\alpha$ -Redutase que possuem IC50 Forte/Forte obtiveram-se, também 10 dendrogramas diferentes. Para as CHF da ChemAxon<sup>®</sup> foi possível dividir os compostos em 10 *clusters*, de acordo com a linha de corte 0,30 seleccionada (Fig. 19, A). No caso das ECFP, a análise separou os 141 compostos em 15 *clusters* (Fig. 19, B). Relativamente aos 141 compostos descritos através da FP2 da OpenBabel<sup>®</sup>, foram apurados 10 *clusters* (Fig. 19, C). No que respeita à análise realizada aos compostos representados pelas *fingerprints* FP3, estes foram divididos em 6 *clusters* (Fig. 19, D). A análise dos 141 compostos representados pelas *fingerprints* FP4 indicou a presença de 4 *clusters* (Fig. 19, E). Para os 141 compostos representados através das *standard fingerprints* da CDK<sup>®</sup>, verificou-se a existência de 10 *clusters* (Fig. 19, F). No caso da análise realizada aos 141 compostos representados pelas *extended fingerprints*, estes foram agrupados em 10

*clusters* (Fig. 19, G). Para os 141 compostos, representados pelas *EState fingerprints*, analisados, verificou-se uma distribuição dos compostos por 4 *clusters* (Fig. 19, H). Os resultados da análise aos 141 compostos representados pela *PubChem fingerprint* indicaram a presença de 6 *clusters* (Fig. 19, I). Por último, a análise realizada aos compostos representados pela MACCS, indicou a separação dos compostos em 3 *clusters* (Fig. 19, J).

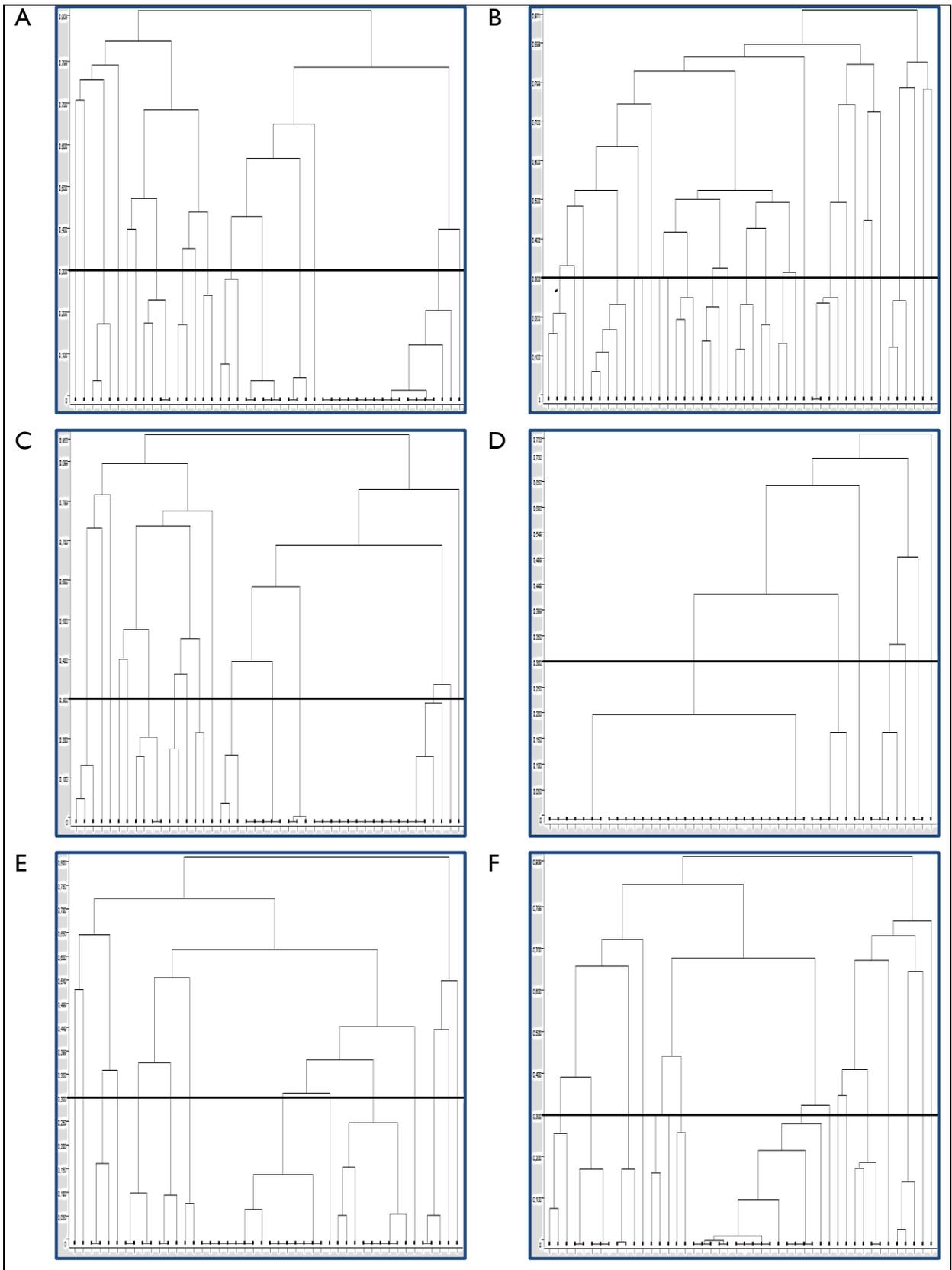


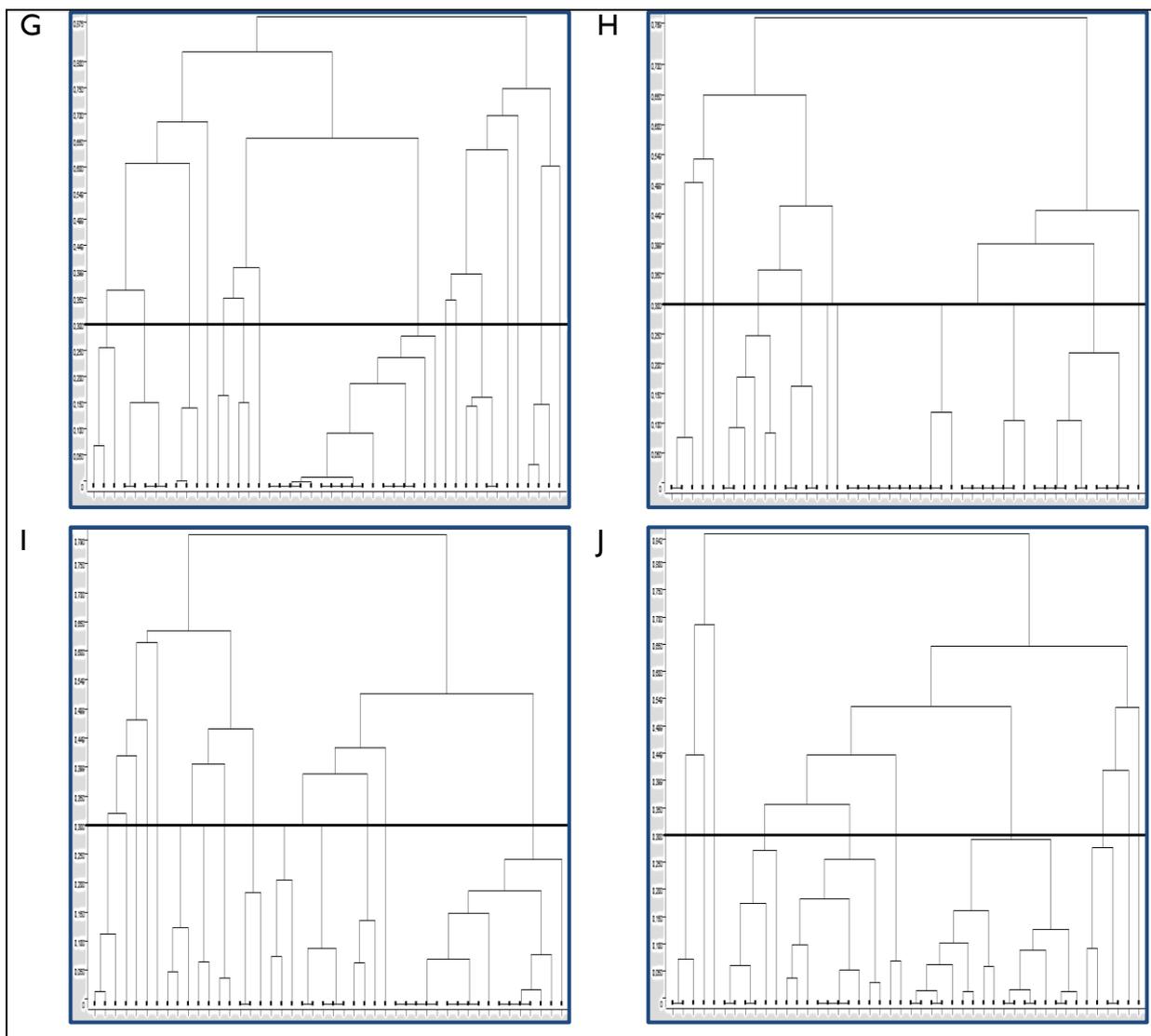


**Figura 19.** Dendrogramas obtidos do *clustering* hierárquico das *fingerprints* CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), *Standard* (F), *Extended* (G), *EState* (H), *PubChem* (I) e *MACCS* (J) que representam os 58 inibidores duais da 5 $\alpha$ -Redutase, com classificação Forte/Forte de IC<sub>50</sub>, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos *clusters* foi 0,30.

No que respeita aos 46 compostos comuns às duas isoenzimas pertencentes à classe de IC<sub>50</sub> Forte/Fraco obtiveram-se 10 dendrogramas diferentes. Para as CHF da ChemAxon<sup>®</sup> foi possível dividir os compostos em 16 *clusters*, de acordo com a linha de corte 0,30 seleccionada (Fig. 20, A). No caso das ECFP, a análise separou os 141 compostos em 26 *clusters* (Fig. 20, B). Relativamente aos 141 compostos descritos através da FP2 da OpenBabel<sup>®</sup>, foram apurados 16 *clusters* (Fig. 20, C). No que respeita à análise realizada aos compostos representados pelas *fingerprints* FP3, estes foram divididos em 7 *clusters* (Fig. 20, D). A análise dos 141 compostos representados pelas *fingerprints* FP4 indicou a presença de 14 *clusters* (Fig. 20, E). Para os 141 compostos representados através das *standard fingerprints* da CDK<sup>®</sup>, verificou-se a existência de 16 *clusters* (Fig. 20, F). No caso da análise realizada aos 141 compostos representados pelas *extended fingerprints*, estes foram agrupados em 15

*clusters* (Fig. 20, G). Para os 141 compostos, representados pelas *EState fingerprints*, analisados, verificou-se uma distribuição dos compostos por 9 *clusters* (Fig. 20, H). Os resultados da análise aos 141 compostos representados pela *PubChem fingerprint* indicaram a presença de 13 *clusters* (Fig. 20, I). Por último, a análise realizada aos compostos representados pela MACCS, indicou a separação dos compostos em 10 *clusters* (Fig. 20, J).

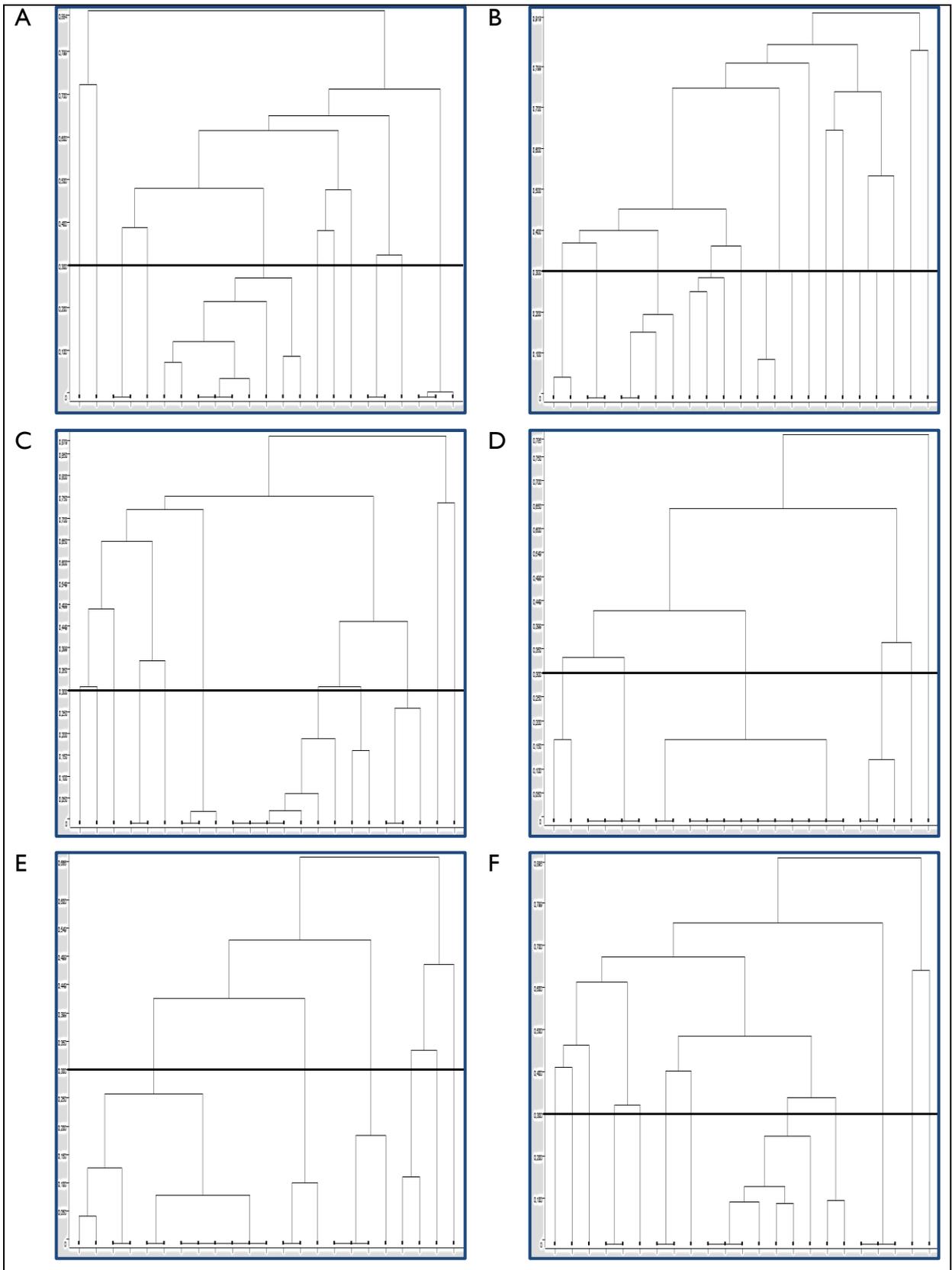


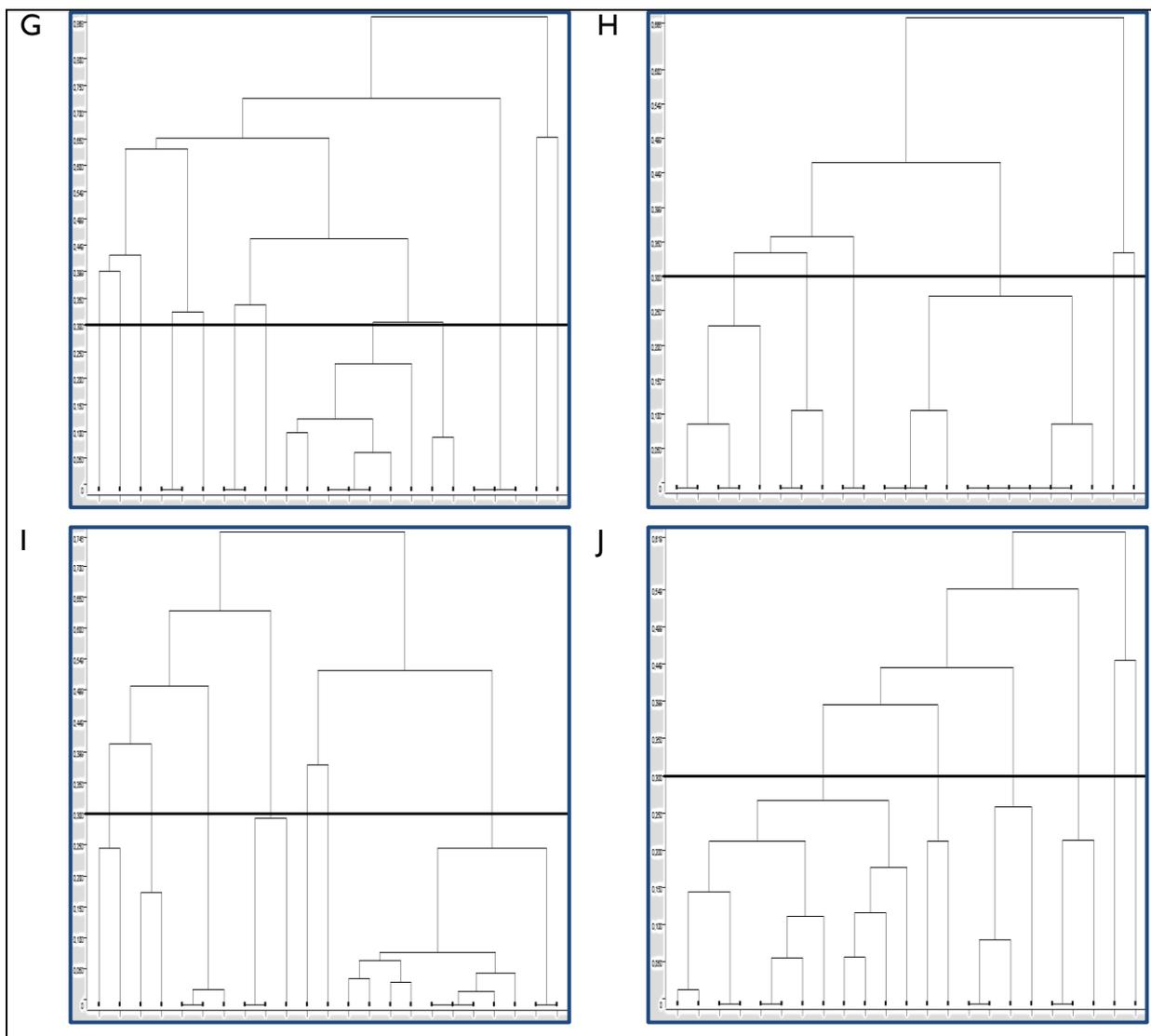


**Figura 20.** Dendrogramas obtidos do *clustering* hierárquico das *fingerprints* CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), *Standard* (F), *Extended* (G), *EState* (H), *PubChem* (I) e *MACCS* (J) que representam os 46 inibidores duais da 5 $\alpha$ -Redutase, com classificação Forte/Fraco de IC<sub>50</sub>, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos *clusters* foi 0,30.

Para os 23 inibidores das duas isoenzimas da 5 $\alpha$ -Redutase que possuem IC<sub>50</sub> Fraco(5 $\alpha$ -R1)/Forte(5 $\alpha$ -R2) obtiveram-se, também 10 dendrogramas diferentes. Para as CHF da ChemAxon<sup>®</sup> foi possível dividir os compostos em 11 *clusters*, de acordo com a linha de corte 0,30 seleccionada (Fig. 21, A). No caso das ECFP, a análise separou os 141 compostos em 15 *clusters* (Fig. 21, B). Relativamente aos 141 compostos descritos através da FP2 da OpenBabel<sup>®</sup>, foram apurados 11 *clusters* (Fig. 21, C). No que respeita à análise realizada aos compostos representados pelas *fingerprints* FP3, estes foram divididos em 6 *clusters* (Fig. 21, D). A análise dos 141 compostos representados pelas *fingerprints* FP4 indicou a presença de 6 *clusters* (Fig. 21, E). Para os 141 compostos representados através das *standard fingerprints* da CDK<sup>®</sup>, verificou-se a existência de 12 *clusters* (Fig. 21, F). No caso da análise realizada aos 141 compostos representados pelas *extended fingerprints*, estes foram agrupados em 12

*clusters* (Fig. 21, G). Para os 141 compostos, representados pelas *EState fingerprints*, analisados, verificou-se uma distribuição dos compostos por 6 *clusters* (Fig. 21, H). Os resultados da análise aos 141 compostos representados pela *PubChem fingerprint* indicaram a presença de 7 *clusters* (Fig. 21, I). Por último, a análise realizada aos compostos representados pela MACCS, indicou a separação dos compostos em 6 *clusters* (Fig. 21, J).





**Figura 21.** Dendrogramas obtidos do *clustering* hierárquico das *fingerprints* CHF (A), ECFP (B), FP2 (C), FP3 (D), FP4 (E), *Standard* (F), *Extended* (G), *EState* (H), *PubChem* (I) e *MACCS* (J) que representam os 23 inibidores duais da 5 $\alpha$ -Redutase, com classificação Fraco/Forte de IC<sub>50</sub>, utilizando o coeficiente de Tanimoto como medida de similaridade. A linha de corte para a definição dos *clusters* foi 0,30.

## 2. Análise estrutural

Recorrendo ao programa LibMCS da ChemAxon<sup>®</sup>, procedeu-se a uma análise de *clustering* com base na estrutura (Fig. 22) e a uma pesquisa de MCS (Tabela 7) para os 32 inibidores da 5 $\alpha$ -RI que apresentam IC50 Excelente.

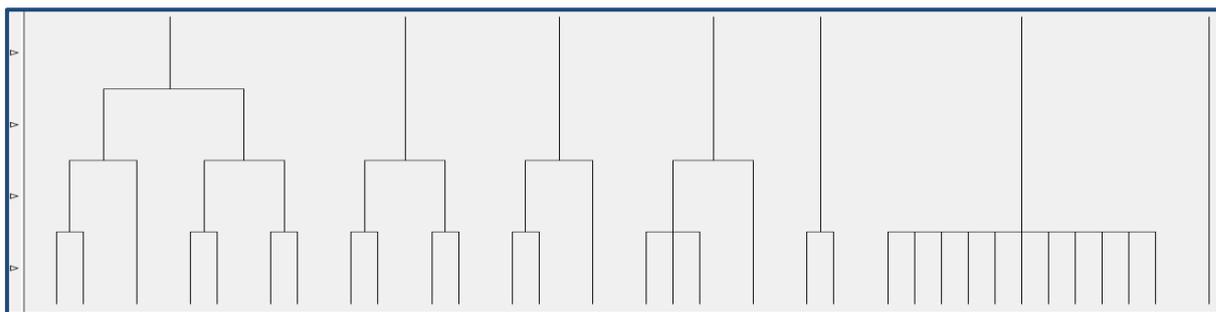
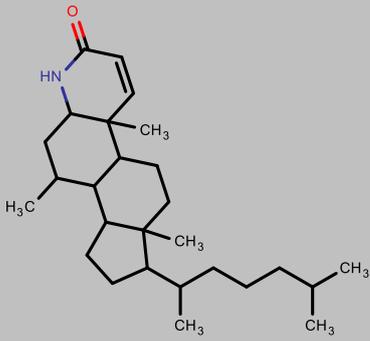
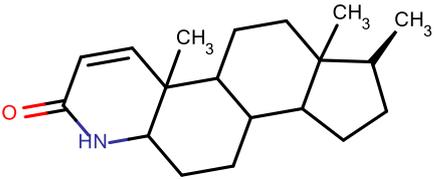
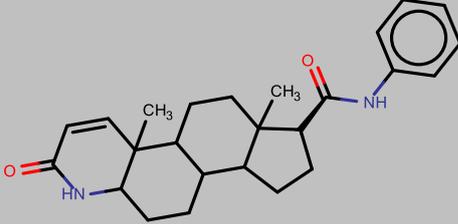
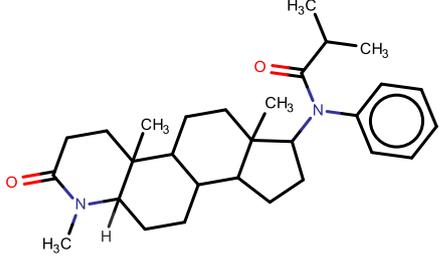
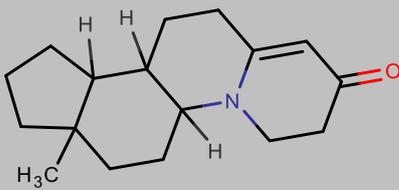
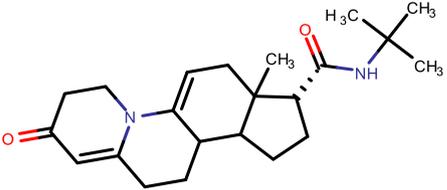
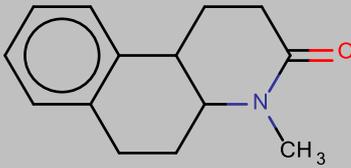
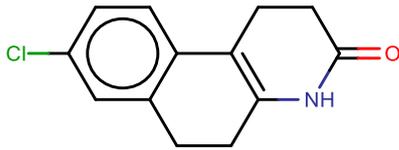
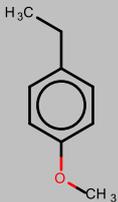
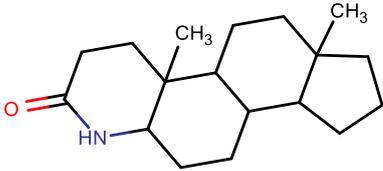
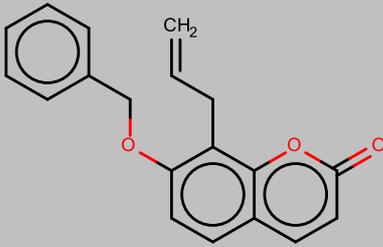


Figura 22. Dendrograma obtido do *clustering* hierárquico das estruturas químicas dos 32 inibidores da 5 $\alpha$ -RI, com IC50 Excelente.

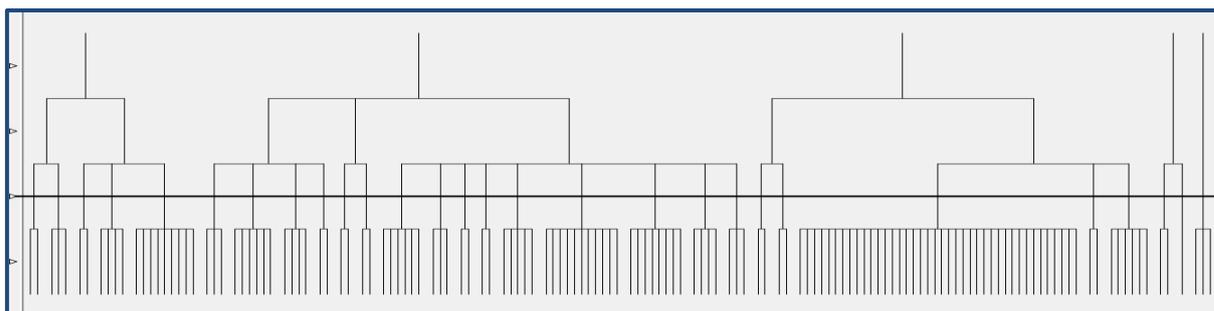
Tabela 7. Análise das MCS dos 32 inibidores da 5 $\alpha$ -RI com IC50 Excelente. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.

Identificação	MCS	# compostos
1		2
2		1

3		2
4		2
5		2
6		2
7		2
8		1
9		3

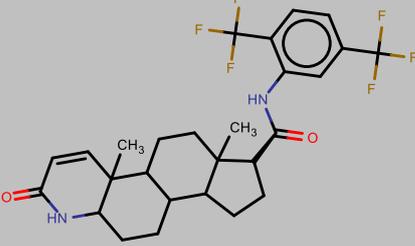
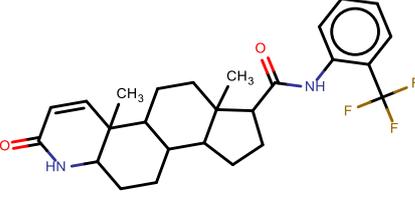
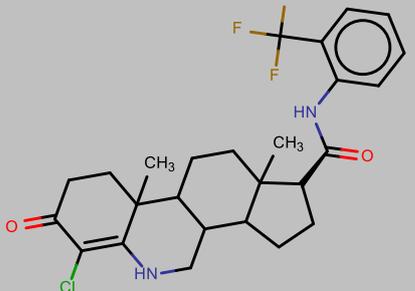
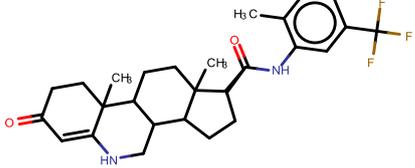
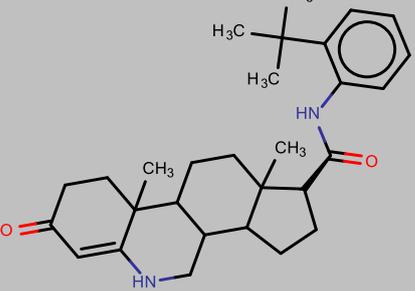
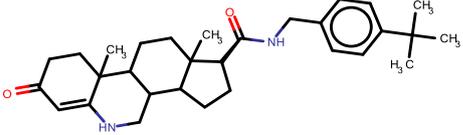
10		I
11		2
12		II
13		I

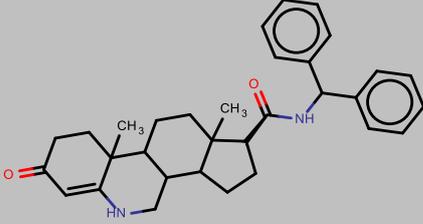
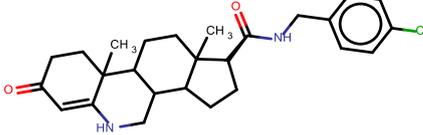
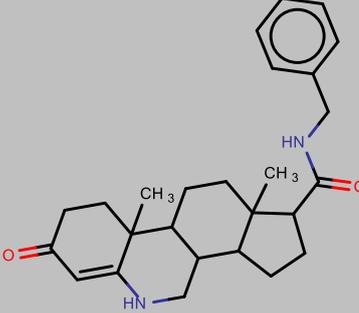
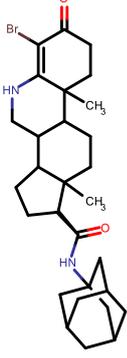
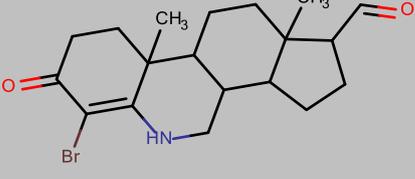
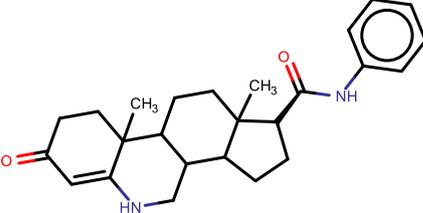
Para os 141 inibidores da 5 $\alpha$ -R2 com IC50 Excelente, foi realizada uma análise de *clustering* com base na estrutura (Fig. 23) e a uma pesquisa de MCS (Tabela 8).

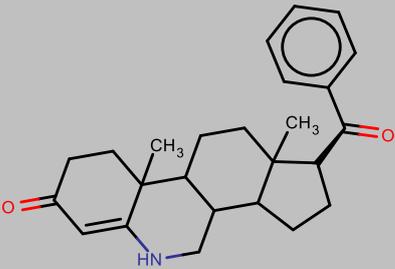
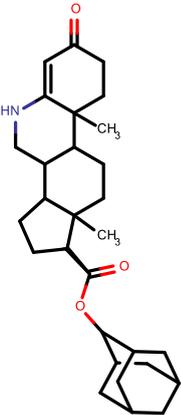
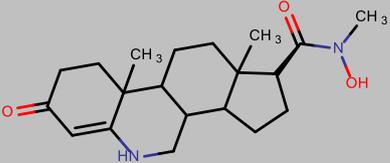
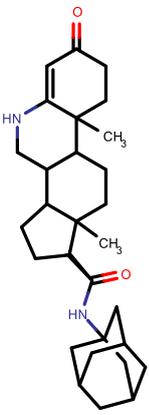
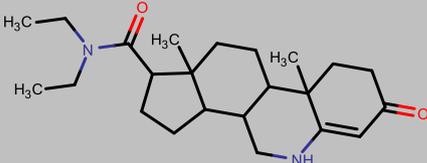


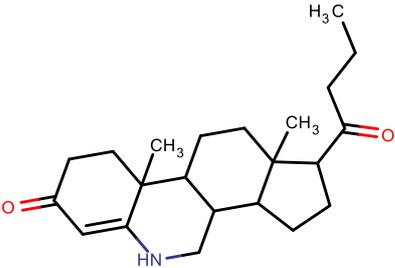
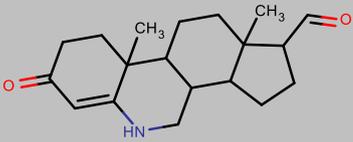
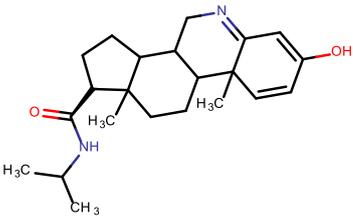
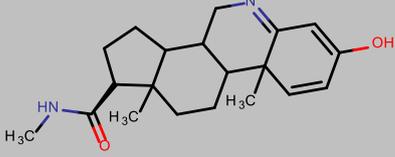
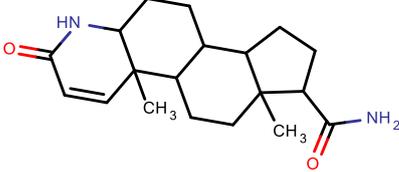
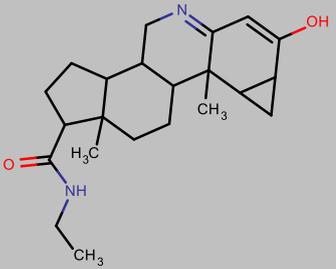
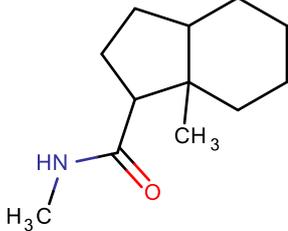
**Figura 23.** Dendrograma obtido do *clustering* hierárquico das estruturas químicas dos 141 inibidores da 5 $\alpha$ -R2, com IC50 Excelente.

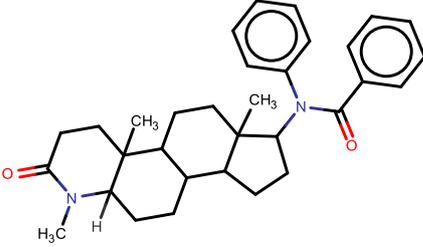
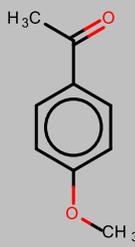
**Tabela 8. Análise das MCS dos 141 inibidores da 5 $\alpha$ -R2 com IC50 Excelente. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.**

Identificação	MCS	# compostos
14		3
15		2
16		2
17		4
18		9
19		3

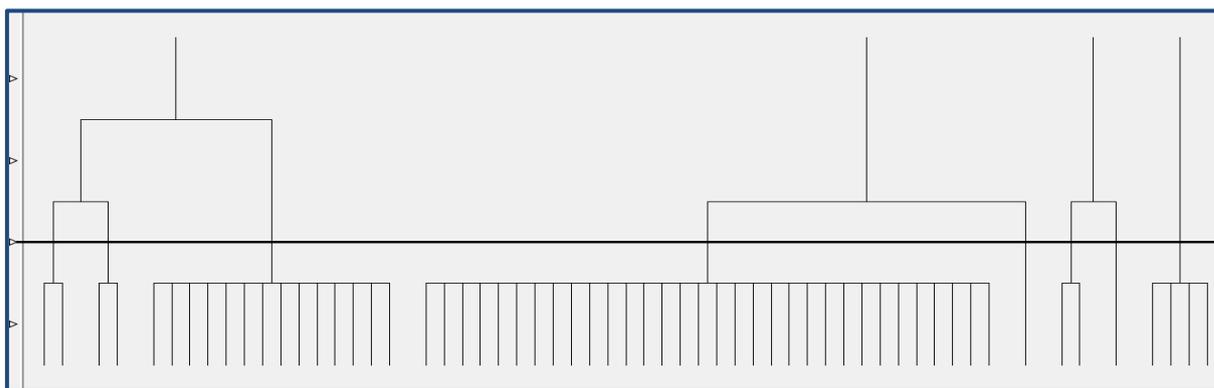
20	 <p>Chemical structure of a pentacyclic alkaloid (pentacyclic tropane derivative) with a benzylamide group attached to the tropane ring system. The structure features a tropane core with two methyl groups (CH<sub>3</sub>) and a benzylamide group (-NH-CH<sub>2</sub>-C<sub>6</sub>H<sub>5</sub>).</p>	6
21	 <p>Chemical structure of a pentacyclic alkaloid (pentacyclic tropane derivative) with a 4-chlorobenzamide group attached to the tropane ring system. The structure features a tropane core with two methyl groups (CH<sub>3</sub>) and a 4-chlorobenzamide group (-NH-CH<sub>2</sub>-C<sub>6</sub>H<sub>4</sub>-Cl).</p>	4
22	 <p>Chemical structure of a pentacyclic alkaloid (pentacyclic tropane derivative) with a benzamide group attached to the tropane ring system. The structure features a tropane core with two methyl groups (CH<sub>3</sub>) and a benzamide group (-NH-CO-C<sub>6</sub>H<sub>5</sub>).</p>	2
23	 <p>Chemical structure of a pentacyclic alkaloid (pentacyclic tropane derivative) with a brominated tropane ring system and a bicyclic amide group attached to the tropane ring system. The structure features a tropane core with two methyl groups (CH<sub>3</sub>) and a bromine atom (Br) on the tropane ring, and a bicyclic amide group (-NH-CO-C<sub>10</sub>H<sub>7</sub>).</p>	2
24	 <p>Chemical structure of a pentacyclic alkaloid (pentacyclic tropane derivative) with a brominated tropane ring system and an aldehyde group attached to the tropane ring system. The structure features a tropane core with two methyl groups (CH<sub>3</sub>) and a bromine atom (Br) on the tropane ring, and an aldehyde group (-CHO).</p>	2
25	 <p>Chemical structure of a pentacyclic alkaloid (pentacyclic tropane derivative) with a benzamide group attached to the tropane ring system. The structure features a tropane core with two methyl groups (CH<sub>3</sub>) and a benzamide group (-NH-CO-C<sub>6</sub>H<sub>5</sub>).</p>	6

26	 <p>A pentacyclic alkaloid core with a benzoyl group attached to one of the methylated carbons. The core consists of a piperidine ring fused to a six-membered ring with a double bond, which is further fused to two five-membered rings. Two methyl groups are attached to the core.</p>	3
27	 <p>A pentacyclic alkaloid core with a bicyclic ester group attached to one of the methylated carbons. The core is similar to structure 26, but the ester group is a bicyclic system.</p>	2
28	 <p>A pentacyclic alkaloid core with a methylamino group attached to one of the methylated carbons. The core is similar to structure 26, but the amine group is a methylamino group.</p>	2
29	 <p>A pentacyclic alkaloid core with a bicyclic amine group attached to one of the methylated carbons. The core is similar to structure 26, but the amine group is a bicyclic system.</p>	5
30	 <p>A pentacyclic alkaloid core with a dimethylamino group attached to one of the methylated carbons. The core is similar to structure 26, but the amine group is a dimethylamino group.</p>	11

31		4
32		3
33		2
34		2
35		40
36		2
37		6

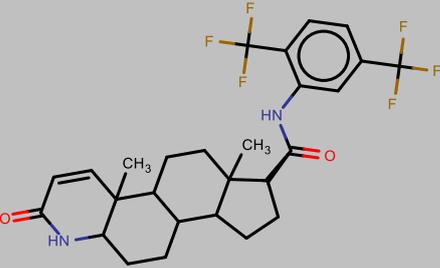
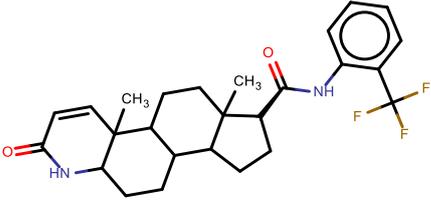
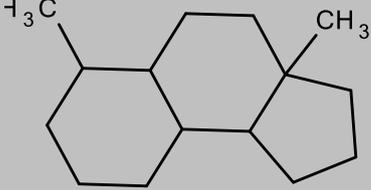
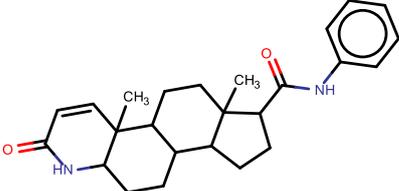
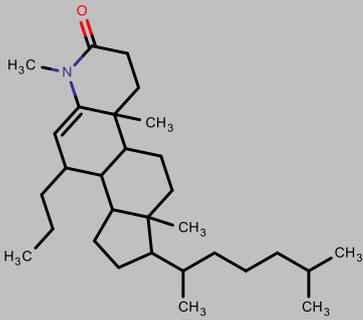
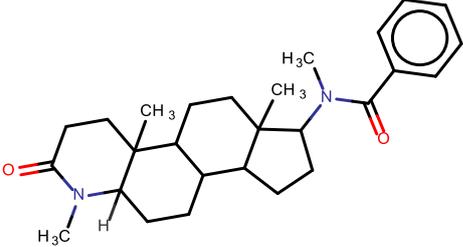
38		2
39		1
40		3

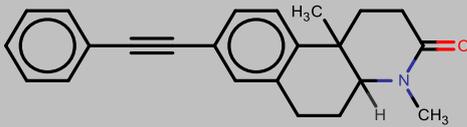
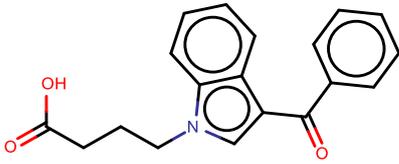
Procedeu-se também a uma análise de *clustering* com base na estrutura para os 58 inibidores duais com IC50 Forte/Forte (Fig. 24). A pesquisa de MCS foi também realizada para os compostos supramencionados (Tabela 9).



**Figura 24.** Dendrograma obtido do *clustering* hierárquico das estruturas químicas dos 58 inibidores duais da 5 $\alpha$ -Redutase, com IC50 Forte/Forte.

**Tabela 9. Análise das MCS dos 58 inibidores duais da 5 $\alpha$ -Redutase com IC50 Forte/Forte. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.**

Identificação	MCS	# compostos
41		2
42		2
43		14
44		32
45		1
46		2

47		1
48		4

Para os 46 compostos classificados com IC50 Forte/Fraco, foi realizada uma análise de *clustering* com base na estrutura (Fig. 25) e uma pesquisa de MCS (Tabela 10).

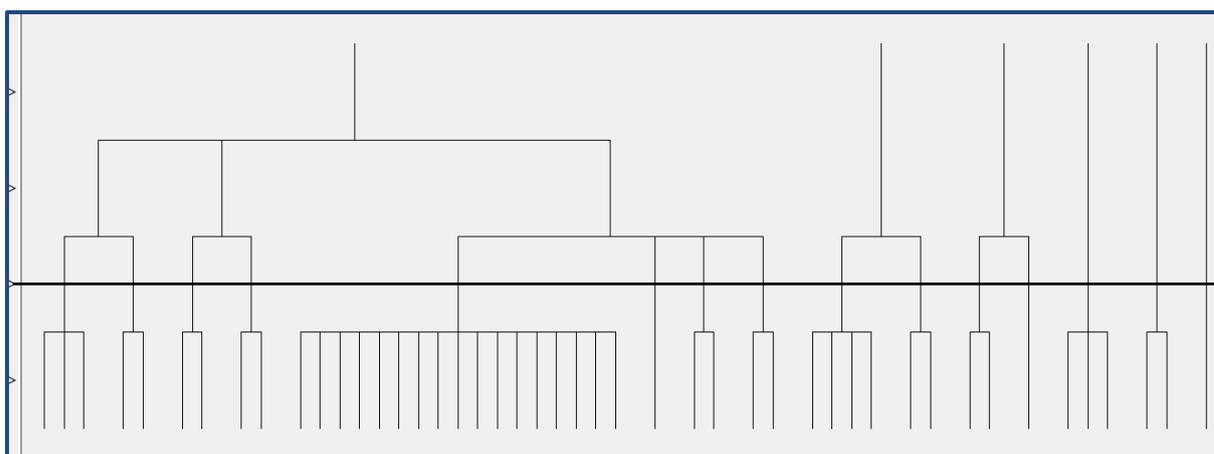
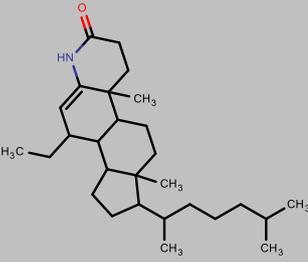
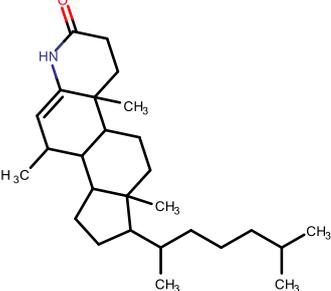
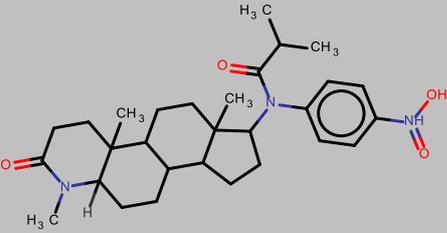
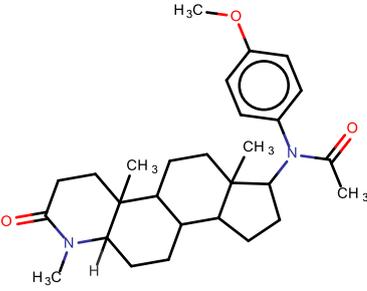
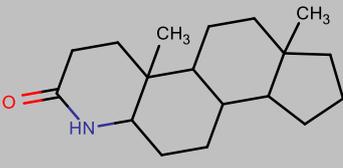
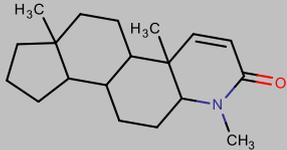
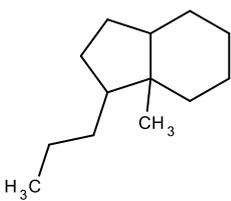
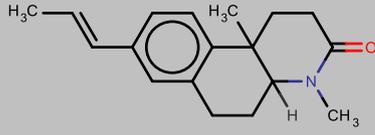
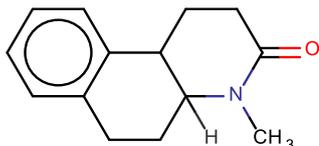
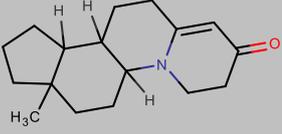
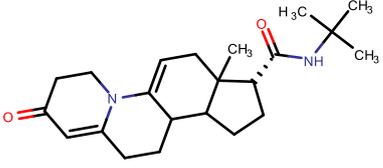
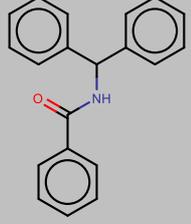
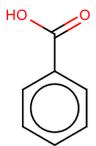
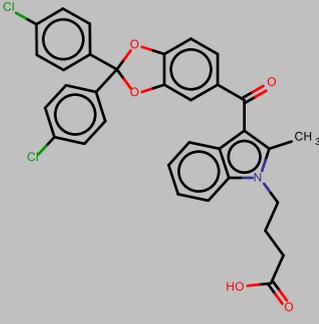


Figura 25. Dendrograma obtido do *clustering* hierárquico das estruturas químicas dos 46 inibidores duais da 5 $\alpha$ -Redutase, com IC50 Forte/Fraco.

Tabela 10. Análise das MCS dos 46 inibidores duais da 5 $\alpha$ -Redutase com IC50 Forte/Fraco. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.

Identificação	MCS	# compostos
49		3
50		2

51		2
52		2
53		17
54		1
55		2
56		2
57		4
58		2

59		2
60		1
61		3
62		2
63		2

Por último, para os 23 compostos classificados com IC50 Fraco/Forte, foi realizada uma análise de *clustering* com base na estrutura (Fig. 26) e uma pesquisa de MCS (Tabela II).

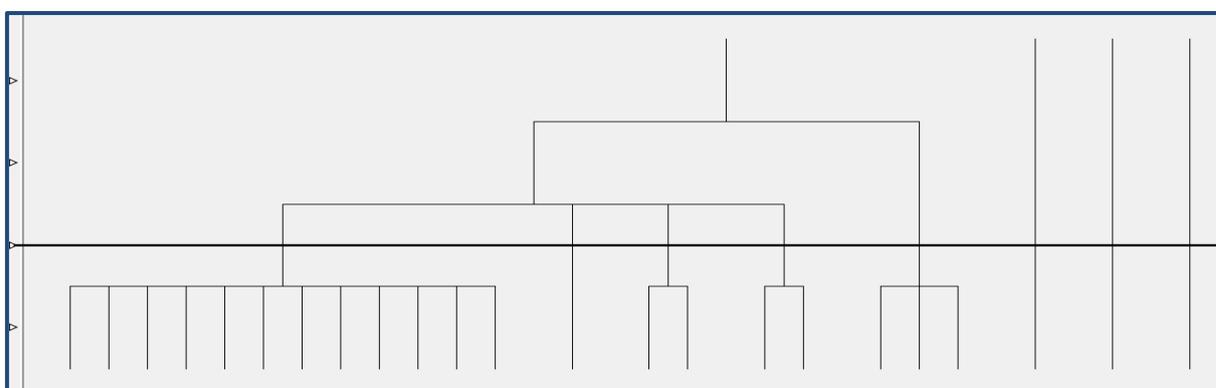
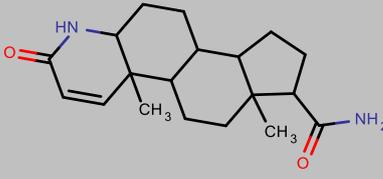
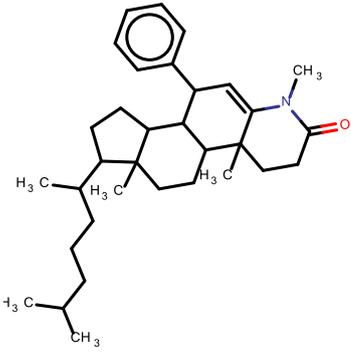
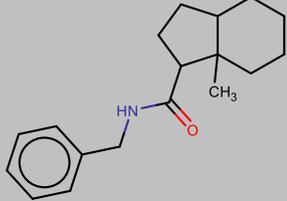
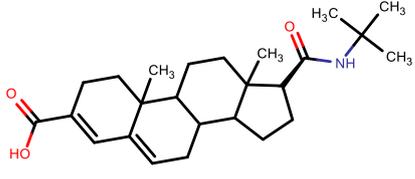
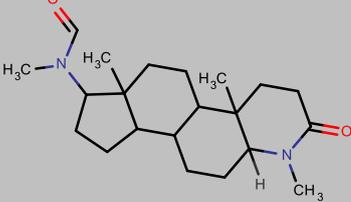
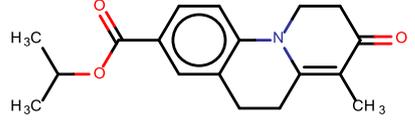
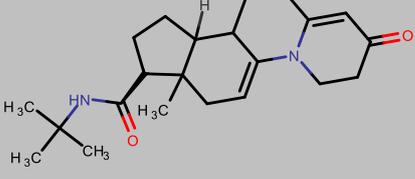
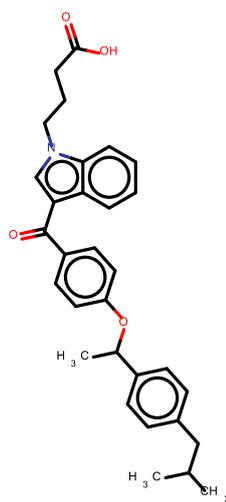


Figura 26. Dendrograma obtido do *clustering* hierárquico das estruturas químicas dos 46 inibidores duais da 5 $\alpha$ -Redutase, com IC50 Fraco/Forte.

**Tabela II. Análise das MCS dos 23 inibidores duais da 5 $\alpha$ -Redutase com IC50 Fraco/Forte. A linha de corte para a selecção das MCS foi o primeiro nível de similaridade.**

Identificação	MCS	# compostos
64		12
65		1
66		2
67		2
68		3
69		1
70		1

71



I

### 3. Classificadores

Através da utilização dos algoritmos de aprendizagem SVM e *Naïve Bayes* com validação cruzada e do programa KNIME analisaram-se as 10 *fingerprints* e os descritores 3D em estudo para os inibidores das isoenzimas 5 $\alpha$ -R1 e 5 $\alpha$ -R2. Esta análise pretende utilizar os compostos referidos para obter o melhor modelo que permita encontrar e classificar potenciais inibidores da 5 $\alpha$ -R1 e 5 $\alpha$ -R2 em bases de dados mais extensas. Assim, nas seguintes tabelas podemos observar quer a reclassificação dos compostos quer as métricas de desempenho que visam avaliar a eficiência da combinação do algoritmo de aprendizagem e *fingerprints*.

Para a 5 $\alpha$ -R1 e 5 $\alpha$ -R2 os classificadores foram construídos com compostos classificados de três maneiras diferentes de acordo com a tabela 3 – classificação multiclasse (a) com compostos distribuídos por 4 classes de IC50 (Excelente, Muito Bom, Bom e Fraco) –, tabela 4 – classificação multiclasse (b) com compostos distribuídos por 5 classes de IC50 (Excelente, Muito Bom, Bom, Fraco e Outros) –, e tabela 5 – classificação binária (c) com compostos distribuídos por 2 classes (Forte e Fraco).

#### 3.1 SVM

##### 3.1.1 *Fingerprints*

##### 3.1.1.1 5 $\alpha$ -R1

Recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R1 descritos por CHF foram reclassificados como consta na tabela 12, onde se pode verificar que para a classe de IC50 Excelente apenas 7 dos 32 compostos foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 38 dos 64 compostos (59%) foram correctamente reclassificados como pertencentes a esta classe. Para os 54 compostos pertencentes à classe de IC50 Bom, apenas 8 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 160 compostos pertencentes à classe de IC50 Fraco, 109 (68%) foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 12. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	7	13	6	6	32
	22%	41%	19%	19%	
Muito Bom	5	38	14	7	64
	8%	59%	22%	11%	
Bom	3	22	8	21	54
	6%	41%	15%	39%	
Fraco	8	22	21	109	160
	5%	14%	13%	68%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -RI descritos por CHF foram reclassificados como consta na tabela 13, onde se pode verificar que para a classe de IC50 Excelente apenas 4 dos 32 compostos foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 35 dos 64 compostos (55%) foram correctamente reclassificados como pertencentes a esta classe. Para os 54 compostos pertencentes à classe de IC50 Bom, apenas 7 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, 59 (58%) foram reclassificados como pertencentes à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, 29 foram reclassificados nesta classe.

**Tabela 13. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	4	18	5	2	3	32
	13%	56%	16%	6%	9%	
Muito Bom	9	35	16	2	2	64
	14%	55%	25%	3%	3%	
Bom	5	25	7	13	4	54
	9%	46%	13%	24%	7%	
Fraco	2	10	17	59	14	102
	2%	10%	17%	58%	14%	
Outros	1	9	7	12	29	58
	2%	16%	12%	21%	50%	

Para a análise binária (c), que utiliza os critérios de classificação definidos na tabela 5, os inibidores da 5 $\alpha$ -RI descritos por CHF foram reclassificados como consta na tabela 14, onde se pode verificar que para a classe de IC50 Forte 121 dos 150 compostos (81%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 115 dos 160 compostos (72%) foram correctamente reclassificados como pertencentes a esta classe pelo algoritmo de aprendizagem.

**Tabela 14. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	121	29	150
	81%	19%	
Fraco	45	115	160
	28%	72%	

No caso dos compostos representados por ECFP, a mesma análise com classificação multiclasse (a) foi realizada, estando o conjunto de resultados descritos na tabela 15, onde se pode verificar que dos 32 compostos pertencentes à classe de IC50 Excelente, 8 (25%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 26 foram reclassificados pelo classificador como pertencentes à classe de IC50 Muito Bom. Para os 64 compostos pertencentes à classe de IC50 Bom, 12 foram atribuídos à classe de IC50 Bom. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 109 foram atribuídos à classe de IC50 Fraco.

**Tabela 15. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	8	9	3	12	32
	25%	28%	9%	38%	
Muito Bom	9	26	15	14	64
	14%	41%	23%	22%	
Bom	3	17	12	22	54
	6%	31%	22%	41%	
Fraco	12	18	21	109	160
	8%	11%	13%	68%	

Foi, também, realizada uma análise com os critérios de classificação (b) para os compostos representados ECFP, estando o conjunto de resultados descritos na tabela 16, onde se pode verificar que dos 32 compostos pertencentes à classe de IC50 Excelente, 6 foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 33 foram reclassificados pelo classificador como pertencentes à classe de IC50 Muito Bom. Para os 64 compostos pertencentes à classe de IC50 Bom, 16 foram atribuídos à classe de IC50 Bom. Dos 102 compostos pertencentes à classe de IC50 Fraco, 53 foram atribuídos à classe de IC50 Fraco. Relativamente aos restantes 58 compostos pertencentes à classe de IC50 Outros, 32 foram reclassificados nesta classe, pelo classificador.

**Tabela 16. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	6	12	5	7	2	32
	19%	38%	16%	22%	6%	
Muito Bom	8	33	16	4	3	64
	13%	52%	25%	6%	5%	
Bom	6	19	16	9	4	54
	11%	35%	30%	17%	7%	
Fraco	9	15	14	53	11	102
	9%	15%	14%	52%	11%	
Outros	0	6	5	15	32	58
	0%	10%	9%	26%	55%	

Para os compostos representados por ECFP, foi ainda realizada uma análise com os critérios de classificação binária (c), estando o conjunto de resultados descritos na tabela 17, onde se pode verificar que dos 150 compostos pertencentes à classe de IC50 Forte, 115 (77%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 115 foram reclassificados pelo classificador como pertencentes à classe de IC50 Fraco.

**Tabela 17. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
Forte	115	35	150
	77%	23%	
Fraco	45	115	160
	28%	72%	

Nos compostos representados pelas FP2 do OpenBabel®, seguindo os padrões da classificação multiclasse (a), obteve-se os resultados indicados na tabela 18. Apenas 3, dos 32 compostos pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 27, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 8 foram classificados como pertencentes a esta classe. Dos 160 compostos pertencentes à classe de IC50 Fraco, 111 foram atribuídos pelo classificador à classe de IC50 Fraco.

**Tabela 18. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	3	14	6	9	32
	9%	44%	19%	28%	
Muito Bom	7	27	15	15	64
	11%	42%	23%	23%	
Bom	6	16	8	24	54
	11%	30%	15%	44%	
Fraco	12	19	18	111	160
	8%	12%	11%	69%	

Nos compostos representados pelas FP2, seguindo os padrões da classificação multiclasse (b), obteve-se os resultados indicados na tabela 19. Apenas 4, dos 32 compostos pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 26, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 5 foram classificados como pertencentes a esta classe. Dos 102 compostos pertencentes à classe de IC50 Fraco, 56 foram atribuídos pelo classificador à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, 24 compostos foram reclassificados nesta classe.

**Tabela 19. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	4	17	1	8	2	32
	13%	53%	3%	25%	6%	
Muito Bom	13	26	13	10	2	64
	20%	41%	20%	16%	3%	
Bom	3	29	5	14	3	54
	6%	54%	9%	26%	6%	
Fraco	8	15	11	56	12	102
	8%	15%	11%	55%	12%	
Outros	1	9	9	15	24	58
	2%	16%	16%	26%	41%	

Nos compostos representados pelas FP2, seguindo os padrões da análise binária (c), obteve-se os resultados indicados na tabela 20. 114, dos 150 compostos pertencentes à classe de IC50 Forte foram reclassificados como pertencentes à classe de IC50 Forte. 116, dos 160 compostos pertencentes à classe de IC50 Fraco foram atribuídos à classe de IC50 Fraco.

**Tabela 20. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	114	36	150
	76%	24%	
Fraco	44	116	160
	28%	73%	

Na tabela 21 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando a classificação multiclasse (a). Dos 32 compostos pertencentes à classe de IC50 Excelente, 8 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 3 foram identificados pelo classificador como pertencentes à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, nenhum foi atribuído à classe de IC50 Bom. Por sua vez, dos 160 compostos pertencentes à classe de IC50 Fraco, 139 (87%) foram atribuídos à classe de IC50 Fraco.

**Tabela 21. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	8	4	0	20	32
	25%	13%	0%	63%	
Muito Bom	14	3	1	46	64
	22%	5%	2%	72%	
Bom	5	4	0	45	54
	9%	7%	0%	83%	
Fraco	5	13	3	139	160
	3%	8%	2%	87%	

Na tabela 22 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando a classificação multiclasse (b). Dos 32 compostos pertencentes à classe de IC50 Excelente, 5 (16%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 25 foram identificados pelo classificador como pertencentes à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 3 foram atribuídos à classe de IC50 Bom. Por sua vez, dos 102 compostos pertencentes à classe de IC50 Fraco, 52 foram atribuídos à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, 17 foram atribuídos a esta classe.

**Tabela 22. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	5	22	0	4	1	32
	16%	69%	0%	13%	3%	
<b>Muito Bom</b>	16	25	4	14	5	64
	25%	39%	6%	22%	8%	
<b>Bom</b>	2	32	3	13	4	54
	4%	59%	6%	24%	7%	
<b>Fraco</b>	7	30	6	52	7	102
	7%	29%	6%	51%	7%	
<b>Outros</b>	2	18	4	17	17	58
	3%	31%	7%	29%	29%	

Na tabela 23 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando a classificação de análise binária (c). Dos 150 compostos pertencentes à classe de IC50 Forte, 119 (79%) foram atribuídos pelo classificador à classe de IC50 Forte. Para os 160 compostos pertencentes à classe de IC50 Fraco, 86 foram identificados pelo classificador como pertencentes à classe de IC50 Fraco.

**Tabela 23. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	119	31	150
	79%	21%	
<b>Fraco</b>	74	86	160
	46%	54%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação multiclasse (a) encontra-se descrita na tabela 24. 5 dos 32 compostos pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 29 foram atribuídos à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 8 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 118 foram atribuídos a esta classe.

**Tabela 24. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	5	16	0	11	32
	16%	50%	0%	34%	
Muito Bom	10	29	7	18	64
	16%	45%	11%	28%	
Bom	2	17	8	27	54
	4%	31%	15%	50%	
Fraco	9	13	20	118	160
	6%	8%	13%	74%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação multiclasse (b) encontra-se descrita na tabela 25. 3 dos 32 compostos pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 36 foram atribuídos à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 6 foram identificados como pertencentes à classe de IC50 Bom. Dos 102 compostos pertencentes à classe de IC50 Fraco, 59 foram atribuídos a esta classe. Para os restantes 58 compostos pertencentes à classe de IC50 Outros, o classificador atribuiu 23 compostos a esta classe.

**Tabela 25. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	3	20	1	6	2	32
	9%	63%	3%	19%	6%	
Muito Bom	6	36	11	8	3	64
	9%	56%	17%	13%	5%	
Bom	1	23	6	17	7	54
	2%	43%	11%	31%	13%	
Fraco	7	13	12	59	11	102
	7%	13%	12%	58%	11%	
Outros	2	8	6	19	23	58
	3%	14%	10%	33%	40%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação de análise binária (c) encontra-se descrita na tabela 26. 110 dos 150 compostos pertencentes à classe de IC50 Forte foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 111 foram atribuídos à classe de IC50 Fraco.

**Tabela 26. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	110	40	150
	73%	27%	
Fraco	49	111	160
	31%	69%	

Para os compostos descritos pelas *Standard fingerprints* calculadas através do CDK<sup>®</sup>, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 27. Dos 32 compostos pertencentes à classe de IC50 Excelente, 5 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 34 (53%) foram atribuídos à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 6 foram classificados como pertencentes à classe de IC50 Bom. Para os restantes 160 compostos pertencentes à classe de IC50 Fraco, 114 foram atribuídos à classe de IC50 Fraco.

**Tabela 27. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	5	17	1	9	32
	16%	53%	3%	28%	
Muito Bom	9	34	13	8	64
	14%	53%	20%	13%	
Bom	3	23	6	22	54
	6%	43%	11%	41%	
Fraco	10	17	19	114	160
	6%	11%	12%	71%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 28. Dos 32 compostos pertencentes à classe de IC50 Excelente, 7 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 39 (61%) foram atribuídos à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 7 foram classificados como pertencentes à classe de IC50 Bom. Para os 102 compostos pertencentes à classe de IC50 Fraco, 59 foram atribuídos à classe de IC50 Fraco. Relativamente aos restantes 58 compostos pertencentes à classe de IC50 Outros, 27 foram atribuídos a esta classe.

**Tabela 28. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	7	16	3	4	2	32
	22%	50%	9%	13%	6%	
Muito Bom	7	39	10	5	3	64
	11%	61%	16%	8%	5%	
Bom	2	26	7	15	4	54
	4%	48%	13%	28%	7%	
Fraco	5	10	17	59	11	102
	5%	10%	17%	58%	11%	
Outros	1	9	5	16	27	58
	2%	16%	9%	28%	47%	

Para os compostos descritos pelas *Standard fingerprints* da CDK, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 29. Dos 150 compostos pertencentes à classe de IC50 Forte, 112 foram atribuídos pelo classificador a esta classe. Para os restantes 160 compostos pertencentes à classe de IC50 Fraco, 109 foram atribuídos à classe de IC50 Fraco.

**Tabela 29. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	112	38	150
	75%	25%	
Fraco	51	109	160
	32%	68%	

Na tabela 30 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints*, utilizando a classificação multiclasse (a). Dos 32 compostos pertencentes à classe de IC50 Excelente, 7 (22%) foram atribuídos a esta classe. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 34 foram atribuídos a esta classe. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 8 foram atribuídos a esta classe. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 116 foram identificados pelo classificador como pertencentes à classe de IC50 Fraco.

**Tabela 30. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	7	15	1	9	32
	22%	47%	3%	28%	
Muito Bom	7	34	11	12	64
	11%	53%	17%	19%	
Bom	1	19	8	26	54
	2%	35%	15%	48%	
Fraco	8	18	18	116	160
	5%	11%	11%	73%	

Na tabela 31 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints* utilizando a classificação multiclasse (b). Dos 32 compostos pertencentes à classe de IC50 Excelente, 4 foram atribuídos a esta classe. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 28 (44%) foram atribuídos a esta classe. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 7 foram atribuídos a esta classe. Dos 102 compostos pertencentes à classe de IC50 Fraco, 50 foram identificados pelo classificador como pertencentes à classe de IC50 Fraco. Para os restantes 58 compostos pertencentes à classe de IC50 Outros, 32 foram identificados como pertencentes a esta classe.

**Tabela 31. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	4	15	1	10	2	32
	13%	47%	3%	31%	6%	
Muito Bom	14	28	15	4	3	64
	22%	44%	23%	6%	5%	
Bom	1	25	7	16	5	54
	2%	46%	13%	30%	9%	
Fraco	10	9	15	50	18	102
	10%	9%	15%	49%	18%	
Outros	2	5	5	14	32	58
	3%	9%	9%	24%	55%	

Na tabela 32 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints* utilizando a classificação de análise binária (c). Dos 150 compostos pertencentes à classe de IC50 Forte, 111 foram atribuídos a esta classe. Para os 160 compostos pertencentes à classe de IC50 Fraco, 105 foram atribuídos a esta classe.

**Tabela 32. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	111	39	150
	74%	26%	
Fraco	55	105	160
	34%	66%	

No caso dos compostos descritos pelas *EState fingerprints* obtidas através do CDK<sup>®</sup>, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 33. 22% (7 em 32) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 20 foram classificados como pertencentes à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 4 foram identificados como pertencentes à classe de IC50 Bom. 123, dos 160 compostos pertencentes à classe de IC50 Fraco, foram atribuídos à classe de IC50 Fraco.

**Tabela 33. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	7	13	2	10	32
	22%	41%	6%	31%	
Muito Bom	13	20	5	26	64
	20%	31%	8%	41%	
Bom	8	12	4	30	54
	15%	22%	7%	56%	
Fraco	8	21	8	123	160
	5%	13%	5%	77%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 34. 19% (6 em 32) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 24 (38%) foram classificados como pertencentes à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 4 foram identificados como pertencentes à classe de IC50 Bom. 60, dos 102 compostos pertencentes à classe de IC50 Fraco, foram atribuídos à classe de IC50 Fraco. Para os restantes 58 compostos pertencentes à classe de IC50 Outros, 10 foram atribuídos a esta classe.

**Tabela 34. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	6	16	2	6	2	32
	19%	50%	6%	19%	6%	
Muito Bom	12	24	7	19	2	64
	19%	38%	11%	30%	3%	
Bom	5	19	4	22	4	54
	9%	35%	7%	41%	7%	
Fraco	6	14	13	60	9	102
	6%	14%	13%	59%	9%	
Outros	4	6	8	30	10	58
	7%	10%	14%	52%	17%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 35. 61% (91 em 150) dos compostos pertencentes à classe de IC50 Forte foram atribuídos a esta classe pelo classificador. Dos 160 compostos pertencentes à classe de IC50 Fraco, 110 foram classificados como pertencentes à classe de IC50 Fraco.

**Tabela 35. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	91	59	150
	61%	39%	
Fraco	50	110	160
	31%	69%	

Na tabela 36 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* obtidas através do CDK<sup>®</sup> utilizando a classificação multiclasse (a). 19% (6 em 32) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe pelo classificador. 23, dos 64 compostos pertencentes à classe de IC50 Muito Bom foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 9 foram atribuídos a esta classe. Dos 160 compostos pertencentes à classe de IC50 Fraco, 98 foram atribuídos a esta classe.

**Tabela 36. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	6	17	3	6	32
	19%	53%	9%	19%	
Muito Bom	14	23	15	12	64
	22%	36%	23%	19%	
Bom	4	21	9	20	54
	7%	39%	17%	37%	
Fraco	7	20	35	98	160
	4%	13%	22%	61%	

Na tabela 37 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação multiclasse (b). 19% (6 em 32) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe. 23 dos 64 compostos pertencentes à classe de IC50 Muito Bom foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 7 foram atribuídos a esta classe. Dos 102 compostos pertencentes à classe de IC50 Fraco, 53 foram atribuídos a esta classe. 28, dos 58 compostos pertencentes à classe de IC50 Outros, foram identificados como pertencentes a esta classe pelo algoritmo de aprendizagem.

**Tabela 37. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	6	13	10	3	0	32
	19%	41%	31%	9%	0%	
Muito Bom	9	23	22	8	2	64
	14%	36%	34%	13%	3%	
Bom	5	22	7	15	5	54
	9%	41%	13%	28%	9%	
Fraco	4	15	18	53	12	102
	4%	15%	18%	52%	12%	
Outros	2	9	5	14	28	58
	3%	16%	9%	24%	48%	

Na tabela 38 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação de análise binária (c). 78% (117 em 150) dos compostos pertencentes à classe de IC50 Forte foram classificados como pertencentes a esta classe. Para os 160 compostos pertencentes à classe de IC50 Fraco, 113 foram atribuídos a esta classe.

**Tabela 38. Reclassificação dos compostos em análise, representados por PubChem fingerprints, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	117	33	150
	78%	22%	
Fraco	47	113	160
	29%	71%	

Na tabela 39 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS obtidas através do CDK<sup>®</sup> utilizando a classificação multiclasse (a). 4, dos 32 compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. 31, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos a esta classe. Dos 54 compostos pertencentes à classe de IC50 Bom, 12 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 108 foram atribuídos à classe de IC50 Fraco.

**Tabela 39. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	4	15	7	6	32
	13%	47%	22%	19%	
Muito Bom	9	31	14	10	64
	14%	48%	22%	16%	
Bom	2	15	12	25	54
	4%	28%	22%	46%	
Fraco	9	19	24	108	160
	6%	12%	15%	68%	

Na tabela 40 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS utilizando a classificação multiclasse (b). 6, dos 32 compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. 33, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos a esta classe. Dos 54 compostos pertencentes à classe de IC50 Bom, 8 foram identificados como pertencentes à classe de IC50 Bom. Dos 102 compostos pertencentes à classe de IC50 Fraco, 57 foram atribuídos à classe de IC50 Fraco. Relativamente aos 58 compostos pertencentes à classe de IC50 Outros, 24 foram identificados como pertencentes a esta classe.

**Tabela 40. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	6	11	7	6	2	32
	19%	34%	22%	19%	6%	
Muito Bom	8	33	14	4	5	64
	13%	52%	22%	6%	8%	
Bom	6	19	8	18	3	54
	11%	35%	15%	33%	6%	
Fraco	6	9	13	57	17	102
	6%	9%	13%	56%	17%	
Outros	1	10	8	15	24	58
	2%	17%	14%	26%	41%	

Por fim, na tabela 41 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS utilizando a classificação de análise binária (c). Dos 150 compostos pertencentes à classe de IC50 Forte, 113 foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 118 foram atribuídos à classe de IC50 Fraco.

**Tabela 41. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
Forte	113	37	150
	75%	25%	
Fraco	42	118	160
	26%	74%	

### 3.1.1.2 5 $\alpha$ -R2

Recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R2 descritos por CHF foram reclassificados como consta na tabela 42, onde se pode verificar que para a classe de IC50 Excelente 126 dos 141 compostos (89%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 2 dos 19 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 42 compostos pertencentes à classe de IC50 Bom, 11 foram atribuídos à classe de IC50 Bom. Relativamente aos 153 compostos pertencentes à classe de IC50 Fraco, 113 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 42. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	126	6	4	5	141
	89%	4%	3%	4%	
Muito Bom	10	2	1	6	19
	53%	11%	5%	32%	
Bom	8	3	11	20	42
	19%	7%	26%	48%	
Fraco	13	1	26	113	153
	8%	1%	17%	74%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -R2 descritos por CHF foram reclassificados como consta na tabela 43, onde se pode verificar que para a classe de IC50 Excelente 130 dos 141 compostos (92%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 2 dos 19 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 42 compostos pertencentes à classe de IC50 Bom, 15 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 97 compostos pertencentes à classe de IC50 Fraco, 58 foram reclassificados como pertencentes à classe de IC50 Fraco. Dos restantes 56 compostos pertencentes à classe de IC50 Outros, 21 foram reclassificados nesta classe.

**Tabela 43. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	130	2	5	2	2	141
	92%	1%	4%	1%	1%	
Muito Bom	8	2	6	3	0	19
	42%	11%	32%	16%	0%	
Bom	9	0	15	13	5	42
	21%	0%	36%	31%	12%	
Fraco	7	1	17	58	14	97
	7%	1%	18%	60%	14%	
Outros	2	0	11	22	21	56
	4%	0%	20%	39%	38%	

Para a análise binária, que utiliza os critérios de classificação (c) definidos na tabela 5, os inibidores da 5 $\alpha$ -R2 descritos por CHF foram reclassificados como consta na tabela 44, onde se pode verificar que para a classe de IC50 Forte 176 dos 202 compostos (87%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 113 dos 153 compostos (74%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 44. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	176	26	202
	87%	13%	
Fraco	40	113	153
	26%	74%	

No caso dos compostos representados por ECFP, a mesma análise com classificação multiclasse (a) foi realizada, estando o conjunto de resultados descritos na tabela 45, onde se pode verificar que dos 141 compostos pertencentes à classe de IC50 Excelente, 126 (89%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, apenas 1 foi reclassificado pelo classificador como pertencente à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 22 foram atribuídos à classe de IC50 Bom. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 134 foram atribuídos à classe de IC50 Fraco.

**Tabela 45. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	126	4	7	4	141
	89%	3%	5%	3%	
Muito Bom	10	1	6	2	19
	53%	5%	32%	11%	
Bom	9	4	22	7	42
	21%	10%	52%	17%	
Fraco	9	4	6	134	153
	6%	3%	4%	88%	

Para os compostos representados por ECFP, foi, também, realizada uma análise com os critérios de classificação (b), estando o conjunto de resultados descritos na tabela 46, onde se pode verificar que dos 141 compostos pertencentes à classe de IC50 Excelente, 124 foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, 3 foram reclassificados como pertencentes à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 21 foram atribuídos à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, 63 foram atribuídos à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, 23 foram reclassificados nesta classe.

**Tabela 46. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	124	8	5	3	1	141
	88%	6%	4%	2%	1%	
Muito Bom	10	3	4	2	0	19
	53%	16%	21%	11%	0%	
Bom	12	2	21	4	3	42
	29%	5%	50%	10%	7%	
Fraco	2	3	7	63	22	97
	2%	3%	7%	65%	23%	
Outros	3	1	2	27	23	56
	5%	2%	4%	48%	41%	

Foi realizada uma análise com os critérios de classificação binária (c) para os compostos representados por ECFP, estando o conjunto de resultados descritos na tabela 47, onde se pode verificar que dos 202 compostos pertencentes à classe de IC50 Forte, 185 (92%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 135 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 47. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
Forte	185	17	202
	92%	8%	
Fraco	18	135	153
	12%	88%	

Nos compostos representados pelas FP2 obtidas através do OpenBabel<sup>®</sup>, seguindo os padrões da classificação multiclasse (a), obteve-se os resultados indicados na tabela 48. 130, dos 141 compostos (92%) pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 2, dos 19 compostos pertencentes à classe de IC50 Muito Bom foram atribuídos à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 12 foram classificados como pertencentes a esta classe. Dos 153 compostos pertencentes à classe de IC50 Fraco, 133 foram atribuídos pelo classificador como pertencentes à classe de IC50 Fraco.

**Tabela 48. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	130	1	5	5	141
	92%	1%	4%	4%	
Muito Bom	9	2	2	6	19
	47%	11%	11%	32%	
Bom	9	0	12	21	42
	21%	0%	29%	50%	
Fraco	8	3	9	133	153
	5%	2%	6%	87%	

Nos compostos representados pelas FP2, seguindo os padrões da classificação multiclasse (b), obteve-se os resultados indicados na tabela 49. Apenas 132, dos 141 compostos (94%) pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 1 dos 19 compostos pertencentes à classe de IC50 Muito Bom foram atribuídos à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 13 foram classificados como pertencentes a esta classe. Dos 97 compostos pertencentes à classe de IC50 Fraco, 60 foram atribuídos pelo classificador como pertencentes à classe de IC50 Fraco. Dos restantes 56 compostos pertencentes à classe de IC50 Outros, 24 compostos foram reclassificados nesta classe.

**Tabela 49. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	132	1	5	2	1	141
	94%	1%	4%	1%	1%	
Muito Bom	10	1	3	5	0	19
	53%	5%	16%	26%	0%	
Bom	9	0	13	18	2	42
	21%	0%	31%	43%	5%	
Fraco	5	1	16	60	15	97
	5%	1%	16%	62%	15%	
Outros	2	0	6	24	24	56
	4%	0%	11%	43%	43%	

Nos compostos representados pelas FP2 do OpenBabel, seguindo os padrões da análise binária (c), obteve-se os resultados indicados na tabela 50. 167, dos 202 compostos pertencentes à classe de IC50 Forte foram reclassificados como pertencentes à classe de IC50 Forte. 126, dos 153 compostos pertencentes à classe de IC50 Fraco foram atribuídos à classe de IC50 Fraco.

**Tabela 50. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	167	35	202
	83%	17%	
Fraco	27	126	153
	18%	82%	

Na tabela 51 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando a classificação multiclasse (a). Dos 141 compostos pertencentes à classe de IC50 Excelente, 123 (87%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, nenhum foi identificado como pertencente à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 2 foram atribuídos à classe de IC50 Bom. Por sua vez, dos 153 compostos pertencentes à classe de IC50 Fraco, 103 (67%) foram atribuídos à classe de IC50 Fraco.

**Tabela 51. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	123	0	1	17	141
	87%	0%	1%	12%	
Muito Bom	8	0	0	11	19
	42%	0%	0%	58%	
Bom	12	0	2	28	42
	29%	0%	5%	67%	
Fraco	45	0	5	103	153
	29%	0%	3%	67%	

Na tabela 52 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando a classificação multiclasse (b). Dos 141 compostos pertencentes à classe de IC50 Excelente, 123 (87%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, nenhum foi identificado como pertencente à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 6 foram atribuídos à classe de IC50 Bom. Por sua vez, dos 97 compostos pertencentes à classe de IC50 Fraco, 67 foram atribuídos à classe de IC50 Fraco. Dos restantes 56 compostos pertencentes à classe de IC50 Outros, 7 foram atribuídos a esta classe.

**Tabela 52. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	123	0	2	15	1	141
	87%	0%	1%	11%	1%	
<b>Muito Bom</b>	10	0	0	9	0	19
	53%	0%	0%	47%	0%	
<b>Bom</b>	12	0	6	22	2	42
	29%	0%	14%	52%	5%	
<b>Fraco</b>	22	0	6	67	2	97
	23%	0%	6%	69%	2%	
<b>Outros</b>	20	0	4	25	7	56
	36%	0%	7%	45%	13%	

Na tabela 53 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando a classificação de análise binária (c). Dos 202 compostos pertencentes à classe de IC50 Forte, 184 (91%) foram atribuídos pelo classificador à classe de IC50 Forte. Para os 153 compostos pertencentes à classe de IC50 Fraco, 64 foram identificados como pertencentes à classe de IC50 Fraco.

**Tabela 53. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	184	18	202
	91%	9%	
<b>Fraco</b>	89	64	153
	58%	42%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação multiclasse (a) encontra-se descrita na tabela 54. 133, dos 141 compostos pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, nenhum foi atribuído a esta classe de IC50. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 6 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 128 foram atribuídos a esta classe.

**Tabela 54. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	133	0	4	4	141
	94%	0%	3%	3%	
Muito Bom	10	0	0	9	19
	53%	0%	0%	47%	
Bom	11	1	6	24	42
	26%	2%	14%	57%	
Fraco	18	4	3	128	153
	12%	3%	2%	84%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação multiclasse (b) encontra-se descrita na tabela 55. 130, dos 141 compostos (92%) pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, 1 foi atribuído à classe de IC50 Muito Bom. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 8 foram identificados como pertencentes à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, 66 foram atribuídos a esta classe. Para os restantes 56 compostos pertencentes à classe de IC50 Outros, o classificador atribuiu 15 a esta classe.

**Tabela 55. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	130	3	4	4	0	141
	92%	2%	3%	3%	0%	
Muito Bom	10	1	1	6	1	19
	53%	5%	5%	32%	5%	
Bom	10	1	8	21	2	42
	24%	2%	19%	50%	5%	
Fraco	11	2	9	66	9	97
	11%	2%	9%	68%	9%	
Outros	3	0	5	33	15	56
	5%	0%	9%	59%	27%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação de análise binária (c) encontra-se descrita na tabela 56. 171, dos 202 compostos pertencentes à classe de IC50 Forte, foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 116 foram atribuídos à classe de IC50 Fraco.

**Tabela 56. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	171	31	202
	85%	15%	
Fraco	37	116	153
	24%	76%	

Para os compostos descritos pelas *Standard fingerprints* obtidas através do CDK<sup>®</sup>, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 57. Dos 141 compostos pertencentes à classe de IC50 Excelente, 126 (89%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 21 foram classificados como pertencentes à classe de IC50 Bom. Para os restantes 153 compostos pertencentes à classe de IC50 Fraco, 116 foram atribuídos à classe de IC50 Fraco.

**Tabela 57. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	126	4	6	5	141
	89%	3%	4%	4%	
Muito Bom	9	2	5	3	19
	47%	11%	26%	16%	
Bom	11	1	21	9	42
	26%	2%	50%	21%	
Fraco	12	1	24	116	153
	8%	1%	16%	76%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 58. Dos 141 compostos pertencentes à classe de IC50 Excelente, 127 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 3 foram atribuídos à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 20 foram classificados como pertencentes à classe de IC50 Bom. Para os 97 compostos pertencentes à classe de IC50 Fraco, 56 foram atribuídos à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, 19 foram atribuídos a esta classe.

**Tabela 58. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	127	3	7	3	1	141
	90%	2%	5%	2%	1%	
Muito Bom	8	3	5	2	1	19
	42%	16%	26%	11%	5%	
Bom	12	1	20	7	2	42
	29%	2%	48%	17%	5%	
Fraco	7	2	19	56	13	97
	7%	2%	20%	58%	13%	
Outros	4	1	5	27	19	56
	7%	2%	9%	48%	34%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 59. Dos 202 compostos pertencentes à classe de IC50 Forte, 183 (91%) foram atribuídos pelo classificador a esta classe. Para os restantes 153 compostos pertencentes à classe de IC50 Fraco, 118 foram atribuídos à classe de IC50 Fraco.

**Tabela 59. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	183	19	202
	91%	9%	
Fraco	35	118	153
	23%	77%	

Na tabela 60 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints* obtidas através do CDK<sup>®</sup> utilizando a classificação multiclasse (a). Dos 141 compostos pertencentes à classe de IC50 Excelente, 126 (89%) foram atribuídos a esta classe. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos a esta classe. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 25 foram atribuídos a esta classe. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 119 foram identificados pelo classificador como pertencentes à classe de IC50 Fraco.

**Tabela 60. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	126	2	6	7	141
	89%	1%	4%	5%	
Muito Bom	9	2	5	3	19
	47%	11%	26%	16%	
Bom	9	0	25	8	42
	21%	0%	60%	19%	
Fraco	13	2	19	119	153
	8%	1%	12%	78%	

Na tabela 61 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints*, utilizando a classificação multiclasse (b). Dos 141 compostos pertencentes à classe de IC50 Excelente, 126 foram atribuídos a esta classe. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 3 foram atribuídos a esta classe. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 21 foram atribuídos a esta classe. Dos 97 compostos pertencentes à classe de IC50 Fraco, 54 foram identificados pelo classificador como pertencentes à classe de IC50 Fraco. Para os restantes 56 compostos pertencentes à classe de IC50 Outros, 18 foram identificados como pertencentes a esta classe.

**Tabela 61. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	126	2	7	5	1	141
	89%	1%	5%	4%	1%	
Muito Bom	9	3	4	3	0	19
	47%	16%	21%	16%	0%	
Bom	11	1	21	7	2	42
	26%	2%	50%	17%	5%	
Fraco	8	3	14	54	18	97
	8%	3%	14%	56%	19%	
Outros	4	1	6	27	18	56
	7%	2%	11%	48%	32%	

Na tabela 62 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints* utilizando a classificação de análise binária (c). Dos 202 compostos pertencentes à classe de IC50 Forte, 185 foram atribuídos a esta classe. Para os 153 compostos pertencentes à classe de IC50 Fraco, 122 foram atribuídos a esta classe.

**Tabela 62. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	185	17	202
	92%	8%	
Fraco	31	122	153
	20%	80%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 63. 86% (121 em 141) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, nenhum foi classificado como pertencente à classe de IC50 Muito Bom. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, nenhum foi identificado pelo classificador como pertencente à classe de IC50 Bom. 122, dos 153 compostos pertencentes à classe de IC50 Fraco, foram atribuídos à classe de IC50 Fraco.

**Tabela 63. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	121	0	1	19	141
	86%	0%	1%	13%	
Muito Bom	6	0	1	12	19
	32%	0%	5%	63%	
Bom	8	0	0	34	42
	19%	0%	0%	81%	
Fraco	28	0	3	122	153
	18%	0%	2%	80%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 64. 89% (125 em 141) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, nenhum foi classificado como pertencentes à classe de IC50 Muito Bom. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 6 foram identificados como pertencentes à classe de IC50 Bom. 62 dos 97 compostos pertencentes à classe de IC50 Fraco, foram atribuídos à classe de IC50 Fraco. Para os restantes 56 compostos pertencentes à classe de IC50 Outros, 2 foram atribuídos a esta classe.

**Tabela 64. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	125	0	1	9	6	141
	89%	0%	1%	6%	4%	
<b>Muito Bom</b>	6	0	2	9	2	19
	32%	0%	11%	47%	11%	
<b>Bom</b>	10	0	6	17	9	42
	24%	0%	14%	40%	21%	
<b>Fraco</b>	19	0	4	62	12	97
	20%	0%	4%	64%	12%	
<b>Outros</b>	11	0	8	35	2	56
	20%	0%	14%	63%	4%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 65. 77% (155 em 202) dos compostos pertencentes à classe de IC50 Forte foram atribuídos a esta classe. Dos 153 compostos pertencentes à classe de IC50 Fraco, 98 foram classificados como pertencentes à classe de IC50 Fraco.

**Tabela 65. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	155	47	202
	77%	23%	
<b>Fraco</b>	55	98	153
	36%	64%	

Na tabela 66 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* obtidas através do CDK<sup>®</sup> utilizando a classificação multiclasse (a). 93% (131 em 141) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe. 1 dos 19 compostos pertencentes à classe de IC50 Muito Bom, foi atribuído à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 21 foram atribuídos a esta classe. Dos 153 compostos pertencentes à classe de IC50 Fraco, 126 foram atribuídos a esta classe.

**Tabela 66. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	131	0	5	5	141
	93%	0%	4%	4%	
Muito Bom	9	1	6	3	19
	47%	5%	32%	16%	
Bom	10	0	21	11	42
	24%	0%	50%	26%	
Fraco	9	0	18	126	153
	6%	0%	12%	82%	

Na tabela 67 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação multiclasse (b). 92% (130 em 141) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe. 4, dos 19 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 21 foram atribuídos a esta classe. Dos 97 compostos pertencentes à classe de IC50 Fraco, 62 foram atribuídos a esta classe. 21, dos 56 compostos pertencentes à classe de IC50 Outros foram identificados como pertencentes a esta classe pelo classificador.

**Tabela 67. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	130	1	6	3	1	141
	92%	1%	4%	2%	1%	
Muito Bom	8	4	5	1	1	19
	42%	21%	26%	5%	5%	
Bom	8	3	21	7	3	42
	19%	7%	50%	17%	7%	
Fraco	6	1	12	62	16	97
	6%	1%	12%	64%	16%	
Outros	5	0	5	25	21	56
	9%	0%	9%	45%	38%	

Na tabela 68 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação de análise binária (c). 90% (182 em 202) dos compostos pertencentes à classe de IC50 Forte foram classificados como pertencentes a esta classe. Para os 153 compostos pertencentes à classe de IC50 Fraco, 126 foram atribuídos a esta classe.

**Tabela 68. Reclassificação dos compostos em análise, representados por PubChem fingerprints, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	182	20	202
	90%	10%	
Fraco	27	126	153
	18%	82%	

Na tabela 69 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS utilizando a classificação multiclasse (a). 129, dos 141 compostos (91%) pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. Nenhum composto, dos 19 pertencentes à classe de IC50 Muito Bom, foi atribuído a esta classe. Dos 42 compostos pertencentes à classe de IC50 Bom, 13 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 113 foram atribuídos à classe de IC50 Fraco.

**Tabela 69. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	129	2	5	5	141
	91%	1%	4%	4%	
Muito Bom	8	0	3	8	19
	42%	0%	16%	42%	
Bom	9	1	13	19	42
	21%	2%	31%	45%	
Fraco	15	2	23	113	153
	10%	1%	15%	74%	

Na tabela 70 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS utilizando a classificação multiclasse (b). 130, dos 141 compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. Nenhum composto, dos 19 pertencentes à classe de IC50 Muito Bom, foi atribuído a esta classe. Dos 42 compostos pertencentes à classe de IC50 Bom, 12 foram identificados como pertencentes à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, 51 foram atribuídos à classe de IC50 Fraco. Relativamente aos 56 compostos pertencentes à classe de IC50 Outros, 19 foram identificados como pertencentes a esta classe.

**Tabela 70. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	130	4	3	4	0	141
	92%	3%	2%	3%	0%	
<b>Muito Bom</b>	10	0	5	2	2	19
	53%	0%	26%	11%	11%	
<b>Bom</b>	10	1	12	15	4	42
	24%	2%	29%	36%	10%	
<b>Fraco</b>	7	2	20	51	17	97
	7%	2%	21%	53%	18%	
<b>Outros</b>	5	1	3	28	19	56
	9%	2%	5%	50%	34%	

Por fim, na tabela 71 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS utilizando a classificação de análise binária (c). Dos 202 compostos pertencentes à classe de IC50 Forte, 175 foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 120 foram atribuídos à classe de IC50 Fraco.

**Tabela 71. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
<b>Forte</b>	175	27	202
	87%	13%	
<b>Fraco</b>	33	120	153
	22%	78%	

### 3.1.1.3 Inibidores duais

Recorrendo à classificação multiclasse, que utiliza os critérios de classificação definidos na tabela 6, os inibidores com dados de IC50 que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase descritos por CHF foram reclassificados como consta na tabela 72, onde se pode verificar que para a classe de IC50 Forte/Forte 40 dos 58 compostos (69%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Forte/Fraco, 26 dos 46 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, 5 foram atribuídos à classe de IC50 Fraco/Forte. Relativamente aos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 42 foram reclassificados como pertencentes à classe de IC50 Fraco/Fraco.

**Tabela 72. Reclassificação dos compostos em análise, representados por CHF, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	40	14	4	0	58
	69%	24%	7%	0%	
Forte/Fraco	9	26	1	10	46
	20%	57%	2%	22%	
Fraco/Forte	10	4	5	4	23
	43%	17%	22%	17%	
Fraco/Fraco	0	11	1	42	54
	0%	20%	2%	78%	

No caso dos compostos representados por ECFP, a mesma análise, com classificação multiclasse indicada na tabela 6, foi realizada, estando o conjunto de resultados descritos na tabela 73, onde se pode verificar que dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 49 (84%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, 29 foram reclassificados como pertencentes à classe de IC50 Forte/Fraco. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, 8 foram atribuídos à classe de IC50 Fraco/Forte. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 41 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 73. Reclassificação dos compostos em análise, representados por ECFP, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	49	4	4	1	58
	84%	7%	7%	2%	
Forte/Fraco	4	29	1	12	46
	9%	63%	2%	26%	
Fraco/Forte	11	1	8	3	23
	48%	4%	35%	13%	
Fraco/Fraco	0	12	1	41	54
	0%	22%	2%	76%	

Nos compostos representados pelas FP2, seguindo os padrões de classificação indicados na tabela 6, obteve-se os resultados indicados na tabela 74. 42, dos 58 compostos (72%) pertencentes à classe de IC50 Forte/Forte foram reclassificados como pertencentes à classe de IC50 Forte/Forte. 21, dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, foram atribuídos à classe de IC50 Forte/Fraco. Para os 23 compostos pertencentes à classe de IC50 Fraco/Fraco, 2 foram classificados como pertencentes a esta classe. Dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 36 foram atribuídos pelo classificador à classe de IC50 Fraco/Fraco.

**Tabela 74. Reclassificação dos compostos em análise, representados por FP2, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	42	13	3	0	58
	72%	22%	5%	0%	
Forte/Fraco	13	21	1	11	46
	28%	46%	2%	24%	
Fraco/Forte	11	3	2	7	23
	48%	13%	9%	30%	
Fraco/Fraco	2	16	0	36	54
	4%	30%	0%	67%	

Na tabela 75 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos por FP3 utilizando os critérios de classificação dos inibidores duais. Dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 36 (62%) foram atribuídos pelo classificador à classe de IC50 Forte/Forte. Para os 46 compostos pertencentes à classe de IC50 Forte/Fraco, 27 foram identificados como pertencentes à classe de IC50 Forte/Fraco. Dos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi atribuído à classe de IC50 Fraco/Forte. Por sua vez, dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 28 (52%) foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 75. Reclassificação dos compostos em análise, representados por FP3, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	36	16	1	5	58
	62%	28%	2%	9%	
Forte/Fraco	9	27	0	10	46
	20%	59%	0%	22%	
Fraco/Forte	13	8	0	2	23
	57%	35%	0%	9%	
Fraco/Fraco	13	13	0	28	54
	24%	24%	0%	52%	

A avaliação do modelo de classificação SVM para os compostos descritos por FP4 utilizando a classificação dos inibidores duais encontra-se descrita na tabela 76. 54, dos 58 compostos pertencentes à classe de IC50 Forte/Forte foram identificados como pertencentes à classe de IC50 Forte/Forte. Dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, 22 foram atribuídos a esta classe de IC50. Relativamente aos 23 compostos pertencentes à classe de IC50 Fraco/Forte, 10 foram identificados como pertencentes à classe de IC50 Fraco/Forte. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 41 foram atribuídos a esta classe.

**Tabela 76. Reclassificação dos compostos em análise, representados por FP4, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	54	1	1	2	58
	93%	2%	2%	3%	
Forte/Fraco	10	22	2	12	46
	22%	48%	4%	26%	
Fraco/Forte	10	0	10	3	23
	43%	0%	43%	13%	
Fraco/Fraco	4	9	0	41	54
	7%	17%	0%	76%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação relativa aos inibidores duais (tabela 6), a avaliação do modelo de classificação SVM encontra-se descrita na tabela 77. Dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 50 (86%) foram atribuídos à classe de IC50 Forte/Forte. Para os 46 compostos pertencentes à classe de IC50 Forte/Fraco, 21 foram atribuídos à classe de IC50 Forte/Fraco. Dos 23 compostos pertencentes à classe de IC50 Fraco/Forte, 8 foram classificados como pertencentes à classe de IC50 Fraco/Forte. Para os restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 39 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 77. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	50	5	2	1	58
	86%	9%	3%	2%	
Forte/Fraco	13	21	1	11	46
	28%	46%	2%	24%	
Fraco/Forte	8	4	8	3	23
	35%	17%	35%	13%	
Fraco/Fraco	2	12	1	39	54
	4%	22%	2%	72%	

Na tabela 78 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *Extended fingerprints* utilizando a classificação dos inibidores duais. Dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 49 (84%) foram atribuídos a esta classe. Para os 46 compostos pertencentes à classe de IC50 Forte/Fraco, 25 foram atribuídos a esta classe. Relativamente aos 23 compostos pertencentes à classe de IC50 Fraco/Forte, 6 foram atribuídos a esta classe. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 39 foram identificados como pertencentes à classe de IC50 Fraco/Fraco.

**Tabela 78. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	49	5	3	1	58
	84%	9%	5%	2%	
Forte/Fraco	13	25	0	8	46
	28%	54%	0%	17%	
Fraco/Forte	12	1	6	4	23
	52%	4%	26%	17%	
Fraco/Fraco	2	12	1	39	54
	4%	22%	2%	72%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação aplicada aos inibidores duais, a avaliação do modelo de classificação SVM encontra-se descrita na tabela 79. 71% (41 em 58) dos compostos pertencentes à classe de IC50 Forte/Forte foram atribuídos a esta classe pelo classificador. Dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, 21 foram classificados como pertencentes à classe de IC50 Forte/Fraco. Relativamente aos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi identificado como pertencente à classe de IC50 Fraco/Forte. 36, dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 79. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	41	13	1	3	58
	71%	22%	2%	5%	
Forte/Fraco	8	21	1	16	46
	17%	46%	2%	35%	
Fraco/Forte	17	5	0	1	23
	74%	22%	0%	4%	
Fraco/Fraco	6	12	0	36	54
	11%	22%	0%	67%	

Na tabela 80 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação relativa aos inibidores duais. 86% (50 em 58) dos compostos pertencentes à classe de IC50 Forte/Forte foram classificados como pertencentes a esta classe de IC50. 31, dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, foram atribuídos à classe de IC50 Forte/Fraco. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, 2 foram reclassificados nesta classe. Dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 40 foram atribuídos a esta classe.

**Tabela 80. Reclassificação dos compostos em análise, representados por PubChem fingerprints, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	50	4	4	0	58
	86%	7%	7%	0%	
Forte/Fraco	4	31	1	10	46
	9%	67%	2%	22%	
Fraco/Forte	15	1	2	5	23
	65%	4%	9%	22%	
Fraco/Fraco	2	11	1	40	54
	4%	20%	2%	74%	

Por fim, na tabela 81 estão indicados os resultados da avaliação do modelo de classificação SVM para os compostos descritos pelas MACCS utilizando a classificação dos inibidores duais. 51, dos 58 compostos (88%) pertencentes à classe de IC50 Forte/Forte foram atribuídos a esta classe. 28, dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, foram atribuídos a esta classe. Dos 23 compostos pertencentes à classe de IC50 Fraco/Forte, 9 foram identificados como pertencentes à classe de IC50 Fraco/Forte. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 41 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 81. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	51	6	1	0	58
	88%	10%	2%	0%	
Forte/Fraco	9	28	3	6	46
	20%	61%	7%	13%	
Fraco/Forte	11	1	9	2	23
	48%	4%	39%	9%	
Fraco/Fraco	3	8	2	41	54
	6%	15%	4%	76%	

### 3.1.2 Descritores 3D

#### 3.1.2.1 5 $\alpha$ -R1

Recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R1 descritos pelos descritores 3D do Volsurf<sup>®</sup> foram reclassificados como consta na tabela 82, onde se pode verificar que para a classe de IC50 Excelente apenas 1 dos 32 compostos foi correctamente classificado como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 24 dos 64 compostos (38%) foram correctamente reclassificados como pertencentes a esta classe. Para os 54

compostos pertencentes à classe de IC50 Bom, 6 compostos foram atribuídos à classe de IC50 Bom. Relativamente aos 160 compostos pertencentes à classe de IC50 Fraco, 145 (91%) foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 82. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
<b>Excelente</b>	1	16	0	15	32
	3%	50%	0%	47%	
<b>Muito Bom</b>	4	24	1	35	64
	6%	38%	2%	55%	
<b>Bom</b>	2	9	6	37	54
	4%	17%	11%	69%	
<b>Fraco</b>	3	10	2	145	160
	2%	6%	1%	91%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -RI descritos por descritores 3D foram reclassificados como consta na tabela 83, onde se pode verificar que para a classe de IC50 Excelente, 3, dos 32 compostos, foram classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 29 dos 64 compostos (45%) foram correctamente reclassificados como pertencentes a esta classe. Para os 54 compostos pertencentes à classe de IC50 Bom, apenas 4 foram atribuídos à classe de IC50 Bom. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, 74 foram reclassificados como pertencentes à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, 21 foram reclassificados nesta classe.

**Tabela 83. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	3	20	0	9	0	32
	9%	63%	0%	28%	0%	
<b>Muito Bom</b>	7	29	4	21	3	64
	11%	45%	6%	33%	5%	
<b>Bom</b>	0	17	4	28	5	54
	0%	31%	7%	52%	9%	
<b>Fraco</b>	3	12	3	74	10	102
	3%	12%	3%	73%	10%	
<b>Outros</b>	0	4	0	33	21	58
	0%	7%	0%	57%	36%	

Para a análise binária (c), que utiliza os critérios de classificação definidos na tabela 5, os inibidores da 5 $\alpha$ -RI descritos por descritores 3D foram reclassificados como consta na

tabela 84, onde se pode verificar que para a classe de IC50 Forte 103 dos 150 compostos (69%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 114 dos 160 compostos (71%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 84. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R1 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	103	47	150
	69%	31%	
Fraco	46	114	160
	29%	71%	

### 3.1.2.2 5 $\alpha$ -R2

Por sua vez, recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R2 descritos pelos descritores 3D do Volsurf<sup>®</sup> foram reclassificados como consta na tabela 85, onde se pode verificar que para a classe de IC50 Excelente, 125, dos 141 compostos, foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom e na classe de IC50 Bom, nenhum dos compostos pertencentes a estas classes foi correctamente classificado. Relativamente aos 153 compostos pertencentes à classe de IC50 Fraco, 141 (92%) foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 85. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	125	0	0	16	141
	89%	0%	0%	11%	
Muito Bom	7	0	0	12	19
	37%	0%	0%	63%	
Bom	9	0	0	33	42
	21%	0%	0%	79%	
Fraco	12	0	0	141	153
	8%	0%	0%	92%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -R2 descritos por descritores 3D foram reclassificados como consta na tabela 86, onde se pode verificar que para a classe de IC50 Excelente, 132, dos 141 compostos, foram classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, nenhum dos compostos foi reclassificado como pertencentes a esta classe.

Para os 42 compostos pertencentes à classe de IC50 Bom, apenas 3 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 97 compostos pertencentes à classe de IC50 Fraco, 80 foram reclassificados como pertencentes à classe de IC50 Fraco. Dos restantes 56 compostos pertencentes à classe de IC50 Outros, 6 foram reclassificados nesta classe.

**Tabela 86. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	132	0	0	9	0	141
	94%	0%	0%	6%	0%	
<b>Muito Bom</b>	9	0	0	10	0	19
	47%	0%	0%	53%	0%	
<b>Bom</b>	8	0	3	30	1	42
	19%	0%	7%	71%	2%	
<b>Fraco</b>	14	0	1	80	2	97
	14%	0%	1%	82%	2%	
<b>Outros</b>	3	0	0	47	6	56
	5%	0%	0%	84%	11%	

Para a análise binária (c), que utiliza os critérios de classificação definidos na tabela 5, os inibidores da 5 $\alpha$ -R2 descritos por descritores 3D foram reclassificados como consta na tabela 87, onde se pode verificar que para a classe de IC50 Forte 180 dos 202 compostos (89%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 115 dos 153 compostos (75%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 87. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	180	22	202
	89%	11%	
<b>Fraco</b>	38	115	153
	25%	75%	

### 3.1.2.3 Inibidores duais

Recorrendo à classificação multiclasse, que utiliza os critérios de classificação definidos na tabela 6, os inibidores com dados de IC50 que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase descritos por descritores 3D foram reclassificados como consta na tabela 88, onde se pode verificar que para a classe de IC50 Forte/Forte 47 dos 58 compostos (81%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de

IC50 Forte/Fraco, 22 dos 46 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, 5 foram atribuídos à classe de IC50 Fraco/Forte pelo classificador. Relativamente aos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 35 foram reclassificados como pertencentes à classe de IC50 Fraco/Fraco.

**Tabela 88. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
<b>Forte/Forte</b>	47	7	2	2	58
	81%	12%	3%	3%	
<b>Forte/Fraco</b>	11	22	1	12	46
	24%	48%	2%	26%	
<b>Fraco/Forte</b>	9	3	5	6	23
	39%	13%	22%	26%	
<b>Fraco/Fraco</b>	3	15	1	35	54
	6%	28%	2%	65%	

## 3.2 Naïve Bayes

### 3.2.1 Fingerprints

#### 3.2.1.1 5 $\alpha$ -RI

Recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -RI descritos por CHF foram reclassificados como consta na tabela 89, onde se pode verificar que para a classe de IC50 Excelente 18 dos 32 compostos foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 10 dos 64 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 54 compostos pertencentes à classe de IC50 Bom, apenas 34 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco.

**Tabela 89. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	18	3	11	0	32
	56%	9%	34%	0%	
Muito Bom	28	10	26	0	64
	44%	16%	41%	0%	
Bom	13	7	34	0	54
	24%	13%	63%	0%	
Fraco	26	32	102	0	160
	16%	20%	64%	0%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -RI descritos por CHF foram reclassificados como consta na tabela 90, onde se pode verificar que para a classe de IC50 Excelente apenas 17 dos 32 compostos foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 8 dos 64 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 54 compostos pertencentes à classe de IC50 Bom, apenas 36 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 90. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	17	0	15	0	0	32
	53%	0%	47%	0%	0%	
Muito Bom	18	8	38	0	0	64
	28%	13%	59%	0%	0%	
Bom	10	8	36	0	0	54
	19%	15%	67%	0%	0%	
Fraco	11	21	70	0	0	102
	11%	21%	69%	0%	0%	
Outros	5	8	45	0	0	58
	9%	14%	78%	0%	0%	

Para a análise binária (c), que utiliza os critérios de classificação definidos na tabela 5, os inibidores da 5 $\alpha$ -RI descritos por CHF foram reclassificados como consta na tabela 91, onde se pode verificar que para a classe de IC50 Forte 90 dos 150 compostos (60%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 127 dos 160 compostos (79%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 91. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	90	60	150
	60%	40%	
Fraco	33	127	160
	21%	79%	

Para os compostos representados por ECFP obtidos através da ChemAxon<sup>®</sup>, a mesma análise com classificação multiclasse (a) foi realizada, estando o conjunto de resultados descritos na tabela 92, onde se pode verificar que dos 32 compostos pertencentes à classe de IC50 Excelente, 18 (56%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 11 foram reclassificados como pertencentes à classe de IC50 Muito Bom. Para os 64 compostos pertencentes à classe de IC50 Bom, 35 foram atribuídos à classe de IC50 Bom. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 92. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	18	5	9	0	32
	56%	16%	28%	0%	
Muito Bom	22	11	31	0	64
	34%	17%	48%	0%	
Bom	9	10	35	0	54
	17%	19%	65%	0%	
Fraco	20	23	117	0	160
	13%	14%	73%	0%	

Para os compostos representados por ECFP, foi realizada uma análise com os critérios de classificação (b), estando o conjunto de resultados descritos na tabela 93, onde se pode verificar que dos 32 compostos pertencentes à classe de IC50 Excelente, 20 foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 8 foram reclassificados como pertencentes à classe de IC50 Muito Bom. Para os 64 compostos pertencentes à classe de IC50 Bom, 36 foram atribuídos à classe de IC50 Bom. Dos compostos pertencentes à classe de IC50 Fraco e à classe de IC50 Outros, nenhum compostos foi reclassificado nestas classes, pelo classificador.

**Tabela 93. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	20	2	10	0	0	32
	63%	6%	31%	0%	0%	
Muito Bom	21	8	35	0	0	64
	33%	13%	55%	0%	0%	
Bom	9	9	36	0	0	54
	17%	17%	67%	0%	0%	
Fraco	15	18	69	0	0	102
	15%	18%	68%	0%	0%	
Outros	6	12	40	0	0	58
	10%	21%	69%	0%	0%	

Nos compostos representados por ECFP, foi, também, realizada uma análise com os critérios de classificação binária (c), estando o conjunto de resultados descritos na tabela 94, onde se pode verificar que dos 150 compostos pertencentes à classe de IC50 Forte, 88 (59%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 126 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 94. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
Forte	88	62	150
	59%	41%	
Fraco	34	126	160
	21%	79%	

Nos compostos representados pelas FP2 obtidos através do OpenBabel®, seguindo os padrões da classificação multiclasse (a), obteve-se os resultados indicados na tabela 95. 17, dos 32 compostos pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 9, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 31 foram classificados como pertencentes a esta classe. Dos 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 95. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	17	4	11	0	32
	53%	13%	34%	0%	
Muito Bom	25	9	30	0	64
	39%	14%	47%	0%	
Bom	14	9	31	0	54
	26%	17%	57%	0%	
Fraco	27	29	104	0	160
	17%	18%	65%	0%	

Nos compostos representados pelas FP2, seguindo os padrões da classificação multiclasse (b), obteve-se os resultados indicados na tabela 96. Apenas 15, dos 32 compostos pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 12, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 36 foram classificados como pertencentes a esta classe. Dos compostos pertencentes à classe de IC50 Fraco e à classe de IC50 Outros, nenhum compostos foi reclassificado nestas classes.

**Tabela 96. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	15	3	14	0	0	32
	47%	9%	44%	0%	0%	
Muito Bom	18	12	34	0	0	64
	28%	19%	53%	0%	0%	
Bom	9	9	36	0	0	54
	17%	17%	67%	0%	0%	
Fraco	8	14	80	0	0	102
	8%	14%	78%	0%	0%	
Outros	8	8	42	0	0	58
	14%	14%	72%	0%	0%	

Nos compostos representados pelas FP2, seguindo os padrões da análise binária (c), obteve-se os resultados indicados na tabela 97. 103, dos 150 compostos pertencentes à classe de IC50 Forte foram reclassificados como pertencentes à classe de IC50 Forte. 122, dos 160 compostos pertencentes à classe de IC50 Fraco foram atribuídos à classe de IC50 Fraco.

**Tabela 97. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	103	47	150
	69%	31%	
Fraco	38	122	160
	24%	76%	

Na tabela 98 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP3 utilizando a classificação multiclasse (a). Dos 32 compostos pertencentes à classe de IC50 Excelente, 21 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 6 foram identificados como pertencentes à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 23 foram atribuídos à classe de IC50 Bom. Por sua vez, nenhum dos 160 compostos pertencentes à classe de IC50 Fraco foram atribuídos à classe de IC50 Fraco.

**Tabela 98. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	21	8	3	0	32
	66%	25%	9%	0%	
Muito Bom	36	6	22	0	64
	56%	9%	34%	0%	
Bom	23	8	23	0	54
	43%	15%	43%	0%	
Fraco	44	31	85	0	160
	28%	19%	53%	0%	

Na tabela 99 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP3 utilizando a classificação multiclasse (b). Dos 32 compostos pertencentes à classe de IC50 Excelente, 16 (50%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 19 foram identificados como pertencentes à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 22 foram atribuídos à classe de IC50 Bom. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 99. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	16	9	7	0	0	32
	50%	28%	22%	0%	0%	
Muito Bom	24	19	21	0	0	64
	38%	30%	33%	0%	0%	
Bom	20	12	22	0	0	54
	37%	22%	41%	0%	0%	
Fraco	27	16	59	0	0	102
	26%	16%	58%	0%	0%	
Outros	15	6	37	0	0	58
	26%	10%	64%	0%	0%	

Na tabela 100 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP3 utilizando a classificação de análise binária (c). Dos 150 compostos pertencentes à classe de IC50 Forte, 98 (65%) foram atribuídos pelo classificador à classe de IC50 Forte. Para os 160 compostos pertencentes à classe de IC50 Fraco, 106 foram identificados como pertencentes à classe de IC50 Fraco.

**Tabela 100. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	98	52	150
	65%	35%	
Fraco	54	106	160
	34%	66%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação multiclasse (a) encontra-se descrita na tabela 101. 16, dos 32 compostos pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 8 foram atribuídos à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 26 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído a esta classe.

**Tabela 101. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	16	7	9	0	32
	50%	22%	28%	0%	
Muito Bom	36	8	20	0	64
	56%	13%	31%	0%	
Bom	19	9	26	0	54
	35%	17%	48%	0%	
Fraco	52	20	88	0	160
	33%	13%	55%	0%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação multiclasse (b) encontra-se descrita na tabela 102. 15, dos 32 compostos pertencentes à classe de IC50 Excelente, foram identificados como pertencentes à classe de IC50 Excelente. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 12 foram atribuídos à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 27 foram identificados como pertencentes à classe de IC50 Bom. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 102. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	15	6	11	0	0	32
	47%	19%	34%	0%	0%	
Muito Bom	25	12	27	0	0	64
	39%	19%	42%	0%	0%	
Bom	18	9	27	0	0	54
	33%	17%	50%	0%	0%	
Fraco	21	19	62	0	0	102
	21%	19%	61%	0%	0%	
Outros	16	11	31	0	0	58
	28%	19%	53%	0%	0%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação de análise binária (c) encontra-se descrita na tabela 103. 91 dos 150 compostos pertencentes à classe de IC50 Forte foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 118 foram atribuídos à classe de IC50 Fraco.

**Tabela 103. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	91	59	150
	61%	39%	
Fraco	42	118	160
	26%	74%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 104. Dos 32 compostos pertencentes à classe de IC50 Excelente, 20 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 11 foram atribuídos à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 36 foram classificados como pertencentes à classe de IC50 Bom. Para os restantes 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 104. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	20	3	9	0	32
	63%	9%	28%	0%	
Muito Bom	20	11	33	0	64
	31%	17%	52%	0%	
Bom	11	7	36	0	54
	20%	13%	67%	0%	
Fraco	28	26	106	0	160
	18%	16%	66%	0%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 105. Dos 32 compostos pertencentes à classe de IC50 Excelente, 18 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 5 foram atribuídos à classe de IC50 Muito Bom. Dos 54 compostos pertencentes à classe de IC50 Bom, 37 foram classificados como pertencentes à classe de IC50 Bom. Dos compostos pertencentes à classe de IC50 Fraco e à classe de IC50 Outros, nenhum compostos foi reclassificado nestas classes, pelo algoritmo de aprendizagem.

**Tabela 105. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	18	2	12	0	0	32
	56%	6%	38%	0%	0%	
<b>Muito Bom</b>	20	5	39	0	0	64
	31%	8%	61%	0%	0%	
<b>Bom</b>	10	7	37	0	0	54
	19%	13%	69%	0%	0%	
<b>Fraco</b>	17	19	66	0	0	102
	17%	19%	65%	0%	0%	
<b>Outros</b>	12	10	36	0	0	58
	21%	17%	62%	0%	0%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 106. Dos 150 compostos pertencentes à classe de IC50 Forte, 76 foram atribuídos pelo classificador a esta classe. Para os restantes 160 compostos pertencentes à classe de IC50 Fraco, 137 foram atribuídos à classe de IC50 Fraco.

**Tabela 106. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	76	74	150
	51%	49%	
<b>Fraco</b>	23	137	160
	14%	86%	

Na tabela 107 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação multiclasse (a). Dos 32 compostos pertencentes à classe de IC50 Excelente, 19 (59%) foram atribuídos a esta classe. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 6 foram atribuídos a esta classe. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 37 foram atribuídos a esta classe. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi identificado como pertencente à classe de IC50 Fraco.

**Tabela 107. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	19	2	11	0	32
	59%	6%	34%	0%	
Muito Bom	19	6	39	0	64
	30%	9%	61%	0%	
Bom	10	7	37	0	54
	19%	13%	69%	0%	
Fraco	20	29	111	0	160
	13%	18%	69%	0%	

Na tabela 108 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação multiclasse (b). Dos 32 compostos pertencentes à classe de IC50 Excelente, 17 foram atribuídos a esta classe. Para os 64 compostos pertencentes à classe de IC50 Muito Bom, 7 foram atribuídos a esta classe. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 38 foram atribuídos a esta classe. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 108. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	17	3	12	0	0	32
	53%	9%	38%	0%	0%	
Muito Bom	20	7	37	0	0	64
	31%	11%	58%	0%	0%	
Bom	10	6	38	0	0	54
	19%	11%	70%	0%	0%	
Fraco	10	21	71	0	0	102
	10%	21%	70%	0%	0%	
Outros	9	12	37	0	0	58
	16%	21%	64%	0%	0%	

Na tabela 109 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação de análise binária (c). Dos 150 compostos pertencentes à classe de IC50 Forte, 80 foram atribuídos a esta classe. Para os 160 compostos pertencentes à classe de IC50 Fraco, 137 foram atribuídos a esta classe.

**Tabela 109. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	80	70	150
	53%	47%	
Fraco	23	137	160
	14%	86%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 110. 59% (19 em 32) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 9 foram classificados como pertencentes à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 36 foram identificados como pertencentes à classe de IC50 Bom. Nenhum dos 160 compostos pertencentes à classe de IC50 Fraco, foi atribuído à classe de IC50 Fraco.

**Tabela 110. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	19	1	12	0	32
	59%	3%	38%	0%	
Muito Bom	17	9	38	0	64
	27%	14%	59%	0%	
Bom	8	10	36	0	54
	15%	19%	67%	0%	
Fraco	17	26	117	0	160
	11%	16%	73%	0%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 111. 59% (19 em 32) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 64 compostos pertencentes à classe de IC50 Muito Bom, 4 foram classificados como pertencentes à classe de IC50 Muito Bom. Relativamente aos 54 compostos pertencentes à classe de IC50 Bom, 32 foram identificados como pertencentes à classe de IC50 Bom. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 111. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	19	2	11	0	0	32
	59%	6%	34%	0%	0%	
<b>Muito Bom</b>	17	4	43	0	0	64
	27%	6%	67%	0%	0%	
<b>Bom</b>	9	13	32	0	0	54
	17%	24%	59%	0%	0%	
<b>Fraco</b>	9	17	76	0	0	102
	9%	17%	75%	0%	0%	
<b>Outros</b>	9	10	39	0	0	58
	16%	17%	67%	0%	0%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 112. 63% (94 em 150) dos compostos pertencentes à classe de IC50 Forte foram atribuídos a esta classe pelo classificador. Dos 160 compostos pertencentes à classe de IC50 Fraco, 102 foram classificados como pertencentes à classe de IC50 Fraco.

**Tabela 112. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	94	56	150
	63%	37%	
<b>Fraco</b>	58	102	160
	36%	64%	

Na tabela 113 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação multiclasse (a). 59% (19 em 32) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe. 9, dos 64 compostos pertencentes à classe de IC50 Muito Bom foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 39 foram atribuídos a esta classe. Dos 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído a esta classe.

**Tabela 113. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	19	1	12	0	32
	59%	3%	38%	0%	
Muito Bom	17	9	38	0	64
	27%	14%	59%	0%	
Bom	9	6	39	0	54
	17%	11%	72%	0%	
Fraco	20	29	111	0	160
	13%	18%	69%	0%	

Na tabela 114 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação multiclasse (b). 59% (19 em 32) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe. 11 dos 64 compostos pertencentes à classe de IC50 Muito Bom foram atribuídos à classe de IC50 Muito Bom. Para os 54 compostos pertencentes à classe de IC50 Bom, 36 foram atribuídos a esta classe. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, nenhum foi reclassificado como pertencente à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 114. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	19	3	10	0	0	32
	59%	9%	31%	0%	0%	
Muito Bom	17	11	36	0	0	64
	27%	17%	56%	0%	0%	
Bom	9	9	36	0	0	54
	17%	17%	67%	0%	0%	
Fraco	9	20	73	0	0	102
	9%	20%	72%	0%	0%	
Outros	8	10	40	0	0	58
	14%	17%	69%	0%	0%	

Na tabela 115 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação de análise binária (c). 57% (85 em 150) dos compostos pertencentes à classe de IC50 Forte foram classificados como pertencentes a esta classe. Para os 160 compostos pertencentes à classe de IC50 Fraco, 120 foram atribuídos a esta classe.

**Tabela 115. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	85	65	150
	57%	43%	
Fraco	40	120	160
	25%	75%	

Na tabela 116 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação multiclasse (a). 14, dos 32 compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. 5, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos a esta classe. Dos 54 compostos pertencentes à classe de IC50 Bom, 30 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 116. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	16	4	12	0	32
	50%	13%	38%	0%	
Muito Bom	22	5	37	0	64
	34%	8%	58%	0%	
Bom	12	12	30	0	54
	22%	22%	56%	0%	
Fraco	29	24	107	0	160
	18%	15%	67%	0%	

Na tabela 117 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação multiclasse (b). 17, dos 32 compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. 3, dos 64 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos a esta classe. Dos 54 compostos pertencentes à classe de IC50 Bom, 33 foram identificados como pertencentes à classe de IC50 Bom. Dos compostos pertencentes à classe de IC50 Fraco e à classe de IC50 Outros, nenhum compostos foi reclassificado nestas classes.

**Tabela 117. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	17	1	14	0	0	32
	53%	3%	44%	0%	0%	
Muito Bom	31	3	30	0	0	64
	48%	5%	47%	0%	0%	
Bom	17	4	33	0	0	54
	31%	7%	61%	0%	0%	
Fraco	17	9	76	0	0	102
	17%	9%	75%	0%	0%	
Outros	7	11	40	0	0	58
	12%	19%	69%	0%	0%	

Por fim, na tabela 118 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação de análise binária (c). Dos 150 compostos pertencentes à classe de IC50 Forte, 98 foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 160 compostos pertencentes à classe de IC50 Fraco, 116 foram atribuídos à classe de IC50 Fraco.

**Tabela 118. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
Forte	98	52	150
	65%	35%	
Fraco	44	116	160
	28%	73%	

### 3.2.1.2 5 $\alpha$ -R2

Utilizando a classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R2 descritos por CHF foram reclassificados como consta na tabela 119, onde se pode verificar que para a classe de IC50 Excelente 119 dos 141 compostos (84%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, nenhum dos 19 compostos foi correctamente reclassificado como pertencente a esta classe. Para os 42 compostos pertencentes à classe de IC50 Bom, 29 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 153 compostos pertencentes à classe de IC50 Fraco, 28 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 119. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	119	0	17	5	141
	84%	0%	12%	4%	
Muito Bom	9	0	7	3	19
	47%	0%	37%	16%	
Bom	7	0	29	6	42
	17%	0%	69%	14%	
Fraco	12	0	113	28	153
	8%	0%	74%	18%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -R2 descritos por CHF foram reclassificados como consta na tabela 120, onde se pode verificar que para a classe de IC50 Excelente 120 dos 141 compostos (87%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 3 dos 19 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 42 compostos pertencentes à classe de IC50 Bom, 29 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos compostos pertencentes à classe de IC50 Fraco e à classe de IC50 Outros, nenhum foi atribuído a estas classes.

**Tabela 120. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	122	7	12	0	0	141
	87%	5%	9%	0%	0%	
Muito Bom	9	3	7	0	0	19
	47%	16%	37%	0%	0%	
Bom	7	6	29	0	0	42
	17%	14%	69%	0%	0%	
Fraco	7	19	71	0	0	97
	7%	20%	73%	0%	0%	
Outros	2	13	41	0	0	56
	4%	23%	73%	0%	0%	

Para a análise binária, que utiliza os critérios de classificação (c) definidos na tabela 5, os inibidores da 5 $\alpha$ -R2 descritos por CHF foram reclassificados como consta na tabela 121, onde se pode verificar que para a classe de IC50 Forte 144 dos 202 compostos (71%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 130 dos 153 compostos (74%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 121. Reclassificação dos compostos em análise, representados por CHF, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	144	58	202
	71%	29%	
Fraco	23	130	153
	15%	85%	

No caso dos compostos representados por ECFP, a mesma análise com classificação multiclasse (a) foi realizada, estando o conjunto de resultados descritos na tabela 122, onde se pode verificar que dos 141 compostos pertencentes à classe de IC50 Excelente, 132 (94%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, apenas 1 foi reclassificado como pertencente à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 31 foram atribuídos à classe de IC50 Bom. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 122. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	132	2	7	0	141
	94%	1%	5%	0%	
Muito Bom	10	1	8	0	19
	53%	5%	42%	0%	
Bom	7	4	31	0	42
	17%	10%	74%	0%	
Fraco	10	28	115	0	153
	7%	18%	75%	0%	

Para os compostos representados por ECFP, foi realizada uma análise com os critérios de classificação (b), estando o conjunto de resultados descritos na tabela 123, onde se pode verificar que dos 141 compostos pertencentes à classe de IC50 Excelente, 135 foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, nenhum foi reclassificado como pertencente à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 24 foram atribuídos à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 123. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	135	0	6	0	0	141
	96%	0%	4%	0%	0%	
Muito Bom	10	0	9	0	0	19
	53%	0%	47%	0%	0%	
Bom	8	10	24	0	0	42
	19%	24%	57%	0%	0%	
Fraco	9	20	68	0	0	97
	9%	21%	70%	0%	0%	
Outros	3	14	39	0	0	56
	5%	25%	70%	0%	0%	

Para os compostos representados por ECFP, foi, ainda, realizada uma análise com os critérios de classificação binária (c), estando o conjunto de resultados descritos na tabela 124, onde se pode verificar que dos 202 compostos pertencentes à classe de IC50 Forte, 156 (77%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 138 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 124. Reclassificação dos compostos em análise, representados por ECFP, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
Forte	156	46	202
	77%	23%	
Fraco	15	138	153
	10%	90%	

Nos compostos representados pelas FP2, seguindo os padrões da classificação multiclasse (a), obteve-se os resultados indicados na tabela 125. 120, dos 141 compostos (85%) pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 1, dos 19 compostos pertencentes à classe de IC50 Muito Bom foi atribuído à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 28 foram classificados como pertencentes a esta classe. Dos 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído pelo classificador à classe de IC50 Fraco.

**Tabela 125. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
<b>Excelente</b>	120	3	18	0	141
	85%	2%	13%	0%	
<b>Muito Bom</b>	9	1	9	0	19
	47%	5%	47%	0%	
<b>Bom</b>	7	7	28	0	42
	17%	17%	67%	0%	
<b>Fraco</b>	12	21	120	0	153
	8%	14%	78%	0%	

Nos compostos representados pelas FP2, seguindo os padrões da classificação multiclasse (b), obteve-se os resultados indicados na tabela 126. Apenas 123, dos 141 compostos (87%) pertencentes à classe de IC50 Excelente foram reclassificados como pertencentes à classe de IC50 Excelente. 1, dos 19 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 29 foram classificados como pertencentes a esta classe. Dos compostos pertencentes à classe de IC50 Fraco e à classe de IC50 Outros, nenhum foi atribuído a estas classes.

**Tabela 126. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	123	2	16	0	0	141
	87%	1%	11%	0%	0%	
<b>Muito Bom</b>	10	1	8	0	0	19
	53%	5%	42%	0%	0%	
<b>Bom</b>	6	7	29	0	0	42
	14%	17%	69%	0%	0%	
<b>Fraco</b>	10	18	69	0	0	97
	10%	19%	71%	0%	0%	
<b>Outros</b>	7	11	38	0	0	56
	13%	20%	68%	0%	0%	

Nos compostos representados pelas FP2, seguindo os padrões da análise binária (c), obteve-se os resultados indicados na tabela 127. 135, dos 202 compostos pertencentes à classe de IC50 Forte foram reclassificados como pertencentes à classe de IC50 Forte. 140, dos 153 compostos pertencentes à classe de IC50 Fraco foram atribuídos à classe de IC50 Fraco.

**Tabela 127. Reclassificação dos compostos em análise, representados por FP2, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	135	67	202
	67%	33%	
Fraco	13	140	153
	8%	92%	

Na tabela 128 estão indicados os resultados da avaliação do modelo de classificação Naïve Bayes para os compostos descritos por FP3 utilizando a classificação multiclasse (a). Dos 141 compostos pertencentes à classe de IC50 Excelente, 112 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, apenas 1 foi identificado como pertencente à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 27 foram atribuídos à classe de IC50 Bom. Por sua vez, dos 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 128. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	112	6	23	0	141
	79%	4%	16%	0%	
Muito Bom	7	1	11	0	19
	37%	5%	58%	0%	
Bom	11	4	27	0	42
	26%	10%	64%	0%	
Fraco	41	23	89	0	153
	27%	15%	58%	0%	

Na tabela 129 estão indicados os resultados da avaliação do modelo de classificação Naïve Bayes para os compostos descritos por FP3 utilizando a classificação multiclasse (b). Dos 141 compostos pertencentes à classe de IC50 Excelente, 113 (80%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, apenas 1 foi identificado como pertencente à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 23 foram atribuídos à classe de IC50 Bom. Por sua vez, dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 129. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	113	8	20	0	0	141
	80%	6%	14%	0%	0%	
Muito Bom	7	1	11	0	0	19
	37%	5%	58%	0%	0%	
Bom	11	8	23	0	0	42
	26%	19%	55%	0%	0%	
Fraco	21	17	59	0	0	97
	22%	18%	61%	0%	0%	
Outros	20	7	29	0	0	56
	36%	13%	52%	0%	0%	

Na tabela 130 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP3 utilizando a classificação de análise binária (c). Dos 202 compostos pertencentes à classe de IC50 Forte, 123 foram atribuídos pelo classificador à classe de IC50 Forte. Para os 153 compostos pertencentes à classe de IC50 Fraco, 114 foram identificados como pertencentes à classe de IC50 Fraco.

**Tabela 130. Reclassificação dos compostos em análise, representados por FP3, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	123	79	202
	61%	39%	
Fraco	39	114	153
	25%	75%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação multiclasse (a) encontra-se descrita na tabela 131. 125, dos 141 compostos pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos a esta classe de IC50. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 32 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído a esta classe.

**Tabela 131. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	125	1	15	0	141
	89%	1%	11%	0%	
Muito Bom	8	2	9	0	19
	42%	11%	47%	0%	
Bom	5	5	32	0	42
	12%	12%	76%	0%	
Fraco	15	30	108	0	153
	10%	20%	71%	0%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação multiclasse (b) encontra-se descrita na tabela 132. 122, dos 141 compostos (87%) pertencentes à classe de IC50 Excelente foram identificados como pertencentes à classe de IC50 Excelente. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos à classe de IC50 Muito Bom. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 29 foram identificados como pertencentes à classe de IC50 Bom. Por sua vez, dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 132. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	122	6	13	0	0	141
	87%	4%	9%	0%	0%	
Muito Bom	8	2	9	0	0	19
	42%	11%	47%	0%	0%	
Bom	5	8	29	0	0	42
	12%	19%	69%	0%	0%	
Fraco	8	16	73	0	0	97
	8%	16%	75%	0%	0%	
Outros	5	10	41	0	0	56
	9%	18%	73%	0%	0%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação de análise binária (c) encontra-se descrita na tabela 133. 153, dos 202 compostos pertencentes à classe de IC50 Forte, foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 132 foram atribuídos à classe de IC50 Fraco.

**Tabela 133. Reclassificação dos compostos em análise, representados por FP4, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	153	49	202
	76%	24%	
Fraco	21	132	153
	14%	86%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 134. Dos 141 compostos pertencentes à classe de IC50 Excelente, 118 (84%) foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 27 foram classificados como pertencentes à classe de IC50 Bom. Para os restantes 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 134. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	118	6	17	0	141
	84%	4%	12%	0%	
Muito Bom	7	2	10	0	19
	37%	11%	53%	0%	
Bom	7	8	27	0	42
	17%	19%	64%	0%	
Fraco	5	31	117	0	153
	3%	20%	76%	0%	

Para os compostos descritos pelas *Standard fingerprints* da CDK, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 135. Dos 141 compostos pertencentes à classe de IC50 Excelente, 118 foram atribuídos pelo classificador à classe de IC50 Excelente. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos à classe de IC50 Muito Bom. Dos 42 compostos pertencentes à classe de IC50 Bom, 31 foram classificados como pertencentes à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 135. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	118	6	17	0	0	141
	84%	4%	12%	0%	0%	
Muito Bom	7	2	10	0	0	19
	37%	11%	53%	0%	0%	
Bom	7	4	31	0	0	42
	17%	10%	74%	0%	0%	
Fraco	7	20	70	0	0	97
	7%	21%	72%	0%	0%	
Outros	1	11	44	0	0	56
	2%	20%	79%	0%	0%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 136. Dos 202 compostos pertencentes à classe de IC50 Forte, 130 foram atribuídos pelo classificador a esta classe. Para os restantes 153 compostos pertencentes à classe de IC50 Fraco, 135 foram atribuídos à classe de IC50 Fraco.

**Tabela 136. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	130	72	202
	64%	36%	
Fraco	18	135	153
	12%	88%	

Na tabela 137 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação multiclasse (a). Dos 141 compostos pertencentes à classe de IC50 Excelente, 114 (81%) foram atribuídos a esta classe. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 4 foram atribuídos a esta classe. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 29 foram atribuídos a esta classe. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi identificado pelo classificador como pertencente à classe de IC50 Fraco.

**Tabela 137. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	114	2	25	0	141
	81%	1%	18%	0%	
Muito Bom	9	4	6	0	19
	47%	21%	32%	0%	
Bom	7	6	29	0	42
	17%	14%	69%	0%	
Fraco	10	30	113	0	153
	7%	20%	74%	0%	

Na tabela 138 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação multiclasse (b). Dos 141 compostos pertencentes à classe de IC50 Excelente, 111 foram atribuídos a esta classe. Para os 19 compostos pertencentes à classe de IC50 Muito Bom, 2 foram atribuídos a esta classe. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 27 foram atribuídos a esta classe. Dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 138. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	111	3	27	0	0	141
	79%	2%	19%	0%	0%	
Muito Bom	6	2	11	0	0	19
	32%	11%	58%	0%	0%	
Bom	6	9	27	0	0	42
	14%	21%	64%	0%	0%	
Fraco	4	21	72	0	0	97
	4%	22%	74%	0%	0%	
Outros	1	9	46	0	0	56
	2%	16%	82%	0%	0%	

Na tabela 139 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação de análise binária (c). Dos 202 compostos pertencentes à classe de IC50 Forte, 134 foram atribuídos a esta classe. Para os 153 compostos pertencentes à classe de IC50 Fraco, 138 foram atribuídos a esta classe.

**Tabela 139. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	134	68	202
	66%	34%	
Fraco	15	138	153
	10%	90%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (a), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 140. 82% (116 em 141) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, apenas 1 foi classificado como pertencente à classe de IC50 Muito Bom. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 29 foram identificados como pertencentes à classe de IC50 Bom. Nenhum dos 153 compostos pertencentes à classe de IC50 Fraco, foi atribuído à classe de IC50 Fraco.

**Tabela 140. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	116	3	22	0	141
	82%	2%	16%	0%	
Muito Bom	4	1	14	0	19
	21%	5%	74%	0%	
Bom	7	6	29	0	42
	17%	14%	69%	0%	
Fraco	26	28	99	0	153
	17%	18%	65%	0%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação multiclasse (b), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 141. 81% (114 em 141) dos compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe pelo classificador. Dos 19 compostos pertencentes à classe de IC50 Muito Bom, 3 foram classificados como pertencentes à classe de IC50 Muito Bom. Relativamente aos 42 compostos pertencentes à classe de IC50 Bom, 26 foram identificados como pertencentes à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 141. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	114	6	21	0	0	141
	81%	4%	15%	0%	0%	
Muito Bom	4	3	12	0	0	19
	21%	16%	63%	0%	0%	
Bom	8	8	26	0	0	42
	19%	19%	62%	0%	0%	
Fraco	15	17	65	0	0	97
	15%	18%	67%	0%	0%	
Outros	11	11	34	0	0	56
	20%	20%	61%	0%	0%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação de análise binária (c), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 142. 78% (158 em 202) dos compostos pertencentes à classe de IC50 Forte foram atribuídos a esta classe pelo classificador. Dos 153 compostos pertencentes à classe de IC50 Fraco, 108 foram classificados como pertencentes à classe de IC50 Fraco.

**Tabela 142. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	158	44	202
	78%	22%	
Fraco	45	108	153
	29%	71%	

Na tabela 143 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação multiclasse (a). 107, dos 141 dos compostos pertencentes à classe de IC50 Excelente, foram classificados como pertencentes a esta classe. Nenhum dos 19 compostos pertencentes à classe de IC50 Muito Bom, foi atribuído à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 31 foram atribuídos a esta classe. Dos 153 compostos pertencentes à classe de IC50 Fraco, 35 foram atribuídos a esta classe.

**Tabela 143. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	107	0	22	12	141
	76%	0%	16%	9%	
Muito Bom	7	0	10	2	19
	37%	0%	53%	11%	
Bom	7	0	31	4	42
	17%	0%	74%	10%	
Fraco	14	0	104	35	153
	9%	0%	68%	23%	

Na tabela 144 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação multiclasse (b). 74% (104 em 141) dos compostos pertencentes à classe de IC50 Excelente foram classificados como pertencentes a esta classe. 2, dos 19 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos à classe de IC50 Muito Bom. Para os 42 compostos pertencentes à classe de IC50 Bom, 30 foram atribuídos a esta classe. Dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 144. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
Excelente	104	13	24	0	0	141
	74%	9%	17%	0%	0%	
Muito Bom	7	2	10	0	0	19
	37%	11%	53%	0%	0%	
Bom	6	6	30	0	0	42
	14%	14%	71%	0%	0%	
Fraco	9	12	76	0	0	97
	9%	12%	78%	0%	0%	
Outros	3	16	37	0	0	56
	5%	29%	66%	0%	0%	

Na tabela 145 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação de análise binária (c). 115 em 202 compostos pertencentes à classe de IC50 Forte foram classificados como pertencentes a esta classe. Para os 153 compostos pertencentes à classe de IC50 Fraco, 140 foram atribuídos a esta classe.

**Tabela 145. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	115	87	202
	57%	43%	
Fraco	13	140	153
	8%	92%	

Na tabela 146 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação multiclasse (a). 119, dos 141 compostos (84%) pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. 3, dos 19 compostos pertencentes à classe de IC50 Muito Bom, foram atribuídos a esta classe. Dos 42 compostos pertencentes à classe de IC50 Bom, 31 foram identificados como pertencentes à classe de IC50 Bom. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco.

**Tabela 146. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	119	6	16	0	141
	84%	4%	11%	0%	
Muito Bom	6	3	10	0	19
	32%	16%	53%	0%	
Bom	5	6	31	0	42
	12%	14%	74%	0%	
Fraco	24	24	105	0	153
	16%	16%	69%	0%	

Na tabela 147 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação multiclasse (b). 119, dos 141 compostos pertencentes à classe de IC50 Excelente foram atribuídos a esta classe. Nenhum composto, dos 19 pertencentes à classe de IC50 Muito Bom, foi atribuído a esta classe. Dos 42 compostos pertencentes à classe de IC50 Bom, 30 foram identificados como pertencentes à classe de IC50 Bom. Dos 97 compostos pertencentes à classe de IC50 Fraco, nenhum foi atribuído à classe de IC50 Fraco. Relativamente aos restantes 56 compostos pertencentes à classe de IC50 Outros, também nenhum foi reclassificado nesta classe.

**Tabela 147. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	119	2	20	0	0	141
	84%	1%	14%	0%	0%	
<b>Muito Bom</b>	6	0	13	0	0	19
	32%	0%	68%	0%	0%	
<b>Bom</b>	6	6	30	0	0	42
	14%	14%	71%	0%	0%	
<b>Fraco</b>	17	17	63	0	0	97
	18%	18%	65%	0%	0%	
<b>Outros</b>	9	7	40	0	0	56
	16%	13%	71%	0%	0%	

Por fim, na tabela 148 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação de análise binária (c). Dos 202 compostos pertencentes à classe de IC50 Forte, 146 foram identificados como pertencentes à classe de IC50 Forte. Dos restantes 153 compostos pertencentes à classe de IC50 Fraco, 120 foram atribuídos à classe de IC50 Fraco.

**Tabela 148. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação (c).**

	Forte	Fraco	Total
<b>Forte</b>	146	56	202
	72%	28%	
<b>Fraco</b>	33	120	153
	22%	78%	

### 3.2.1.3 Inibidores duais

Recorrendo à classificação multiclasse, que utiliza os critérios de classificação definidos na tabela 6, os inibidores com dados de IC50 que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase descritos por CHF foram reclassificados como consta na tabela 149, onde se pode verificar que para a classe de IC50 Forte/Forte 43 dos 58 compostos (81%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Forte/Fraco, 4 dos 46 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi atribuído à classe de IC50 Fraco/Forte. Relativamente aos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 44 foram reclassificados como pertencentes à classe de IC50 Fraco/Fraco.

**Tabela 149. Reclassificação dos compostos em análise, representados por CHF, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	43	0	0	15	58
	74%	0%	0%	26%	
Forte/Fraco	20	4	0	22	46
	43%	9%	0%	48%	
Fraco/Forte	13	1	0	9	23
	57%	4%	0%	39%	
Fraco/Fraco	1	9	0	44	54
	2%	17%	0%	81%	

No caso dos compostos representados por ECFP, a mesma análise, com classificação multiclasse indicada na tabela 6, foi realizada, estando o conjunto de resultados descritos na tabela 150, onde se pode verificar que dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 54 (93%) foram reclassificados como pertencentes a esta mesma classe de IC50. Dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, 4 foram reclassificados como pertencentes à classe de IC50 Forte/Fraco. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi atribuído à classe de IC50 Fraco/Forte. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 43 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 150. Reclassificação dos compostos em análise, representados por ECFP, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	54	1	0	3	58
	93%	2%	0%	5%	
Forte/Fraco	10	4	0	32	46
	22%	9%	0%	70%	
Fraco/Forte	17	2	0	4	23
	74%	9%	0%	17%	
Fraco/Fraco	1	10	0	43	54
	2%	19%	0%	80%	

Nos compostos representados pelas FP2, seguindo os padrões de classificação indicados na tabela 6, obteve-se os resultados indicados na tabela 151. 52, dos 58 compostos (90%) pertencentes à classe de IC50 Forte/Forte foram reclassificados como pertencentes à classe de IC50 Forte/Forte. 6, dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, foram atribuídos à classe de IC50 Forte/Fraco. Para os 23 compostos pertencentes à classe de IC50 Fraco/Fraco, nenhum foi classificado como pertencentes a esta classe. Dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 39 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 151. Reclassificação dos compostos em análise, representados por FP2, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	52	1	0	5	58
	90%	2%	0%	9%	
Forte/Fraco	24	6	0	16	46
	52%	13%	0%	35%	
Fraco/Forte	14	1	0	8	23
	61%	4%	0%	35%	
Fraco/Fraco	4	11	0	39	54
	7%	20%	0%	72%	

Na tabela 152 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP3 utilizando os critérios de classificação dos inibidores duais. Dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 34 (59%) foram atribuídos à classe de IC50 Forte/Forte. Para os 46 compostos pertencentes à classe de IC50 Forte/Fraco, 6 foram identificados como pertencentes à classe de IC50 Forte/Fraco. Dos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi atribuído à classe de IC50 Fraco/Forte. Por sua vez, dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 36 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 152. Reclassificação dos compostos em análise, representados por FP3, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	34	5	0	19	58
	59%	9%	0%	33%	
Forte/Fraco	7	6	0	33	46
	15%	13%	0%	72%	
Fraco/Forte	10	2	0	11	23
	43%	9%	0%	48%	
Fraco/Fraco	12	6	0	36	54
	22%	11%	0%	67%	

A avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos por FP4 utilizando a classificação dos inibidores duais encontra-se descrita na tabela 153. 35, dos 58 compostos pertencentes à classe de IC50 Forte/Forte foram identificados como pertencentes à classe de IC50 Forte/Forte. Dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, 12 foram atribuídos a esta classe de IC50. Relativamente aos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi identificado como pertencente à classe de IC50 Fraco/Forte. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 41 foram atribuídos a esta classe.

**Tabela 153. Reclassificação dos compostos em análise, representados por FP4, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	35	2	0	21	58
	60%	3%	0%	36%	
Forte/Fraco	3	12	0	31	46
	7%	26%	0%	67%	
Fraco/Forte	6	1	0	16	23
	26%	4%	0%	70%	
Fraco/Fraco	1	12	0	41	54
	2%	22%	0%	76%	

Para os compostos descritos pelas *Standard fingerprints*, utilizando a classificação relativa aos inibidores duais (tabela 6), a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 154. Dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 39 foram atribuídos pelo classificador à classe de IC50 Forte/Forte. Para os 46 compostos pertencentes à classe de IC50 Forte/Fraco, 3 foram atribuídos à classe de IC50 Forte/Fraco. Dos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi classificado como pertencente à classe de IC50 Fraco/Forte. Para os restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 42 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 154. Reclassificação dos compostos em análise, representados por *Standard fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	39	0	0	19	58
	67%	0%	0%	33%	
Forte/Fraco	11	3	0	32	46
	24%	7%	0%	70%	
Fraco/Forte	11	2	0	10	23
	48%	9%	0%	43%	
Fraco/Fraco	4	8	0	42	54
	7%	15%	0%	78%	

Na tabela 155 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *Extended fingerprints* utilizando a classificação dos inibidores duais. Dos 58 compostos pertencentes à classe de IC50 Forte/Forte, 47 (81%) foram atribuídos a esta classe. Para os 46 compostos pertencentes à classe de IC50 Forte/Fraco, 3 foram atribuídos a esta classe. Relativamente aos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi atribuído a esta classe. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 42 foram identificados como pertencentes à classe de IC50 Fraco/Fraco.

**Tabela 155. Reclassificação dos compostos em análise, representados por *Extended fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	47	1	0	10	58
	81%	2%	0%	17%	
Forte/Fraco	16	3	0	27	46
	35%	7%	0%	59%	
Fraco/Forte	15	1	0	7	23
	65%	4%	0%	30%	
Fraco/Fraco	3	9	0	42	54
	6%	17%	0%	78%	

No caso dos compostos descritos pelas *EState fingerprints*, utilizando a classificação aplicada aos inibidores duais, a avaliação do modelo de classificação *Naïve Bayes* encontra-se descrita na tabela 156. 30, dos 58 compostos pertencentes à classe de IC50 Forte/Forte foram atribuídos a esta classe pelo classificador. Dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, 10 foram classificados como pertencentes à classe de IC50 Forte/Fraco. Relativamente aos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi identificado como pertencente à classe de IC50 Fraco/Forte. 44, dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 156. Reclassificação dos compostos em análise, representados por *EState fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	30	6	0	22	58
	52%	10%	0%	38%	
Forte/Fraco	0	10	0	36	46
	0%	22%	0%	78%	
Fraco/Forte	5	5	0	13	23
	22%	22%	0%	57%	
Fraco/Fraco	1	9	0	44	54
	2%	17%	0%	81%	

Na tabela 157 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas *PubChem fingerprints* utilizando a classificação relativa aos inibidores duais. 39 em 58 compostos pertencentes à classe de IC50 Forte/Forte foram classificados como pertencentes a esta classe de IC50. 10 dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, foram atribuídos à classe de IC50 Forte/Fraco. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi reclassificado nesta classe. Dos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 42 foram atribuídos a esta classe.

**Tabela 157. Reclassificação dos compostos em análise, representados por *PubChem fingerprints*, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	39	4	0	15	58
	67%	7%	0%	26%	
Forte/Fraco	11	10	0	25	46
	24%	22%	0%	54%	
Fraco/Forte	12	5	0	6	23
	52%	22%	0%	26%	
Fraco/Fraco	7	5	0	42	54
	13%	9%	0%	78%	

Por fim, na tabela 158 estão indicados os resultados da avaliação do modelo de classificação *Naïve Bayes* para os compostos descritos pelas MACCS utilizando a classificação dos inibidores duais. 40, dos 58 compostos pertencentes à classe de IC50 Forte/Forte foram atribuídos a esta classe. 12, dos 46 compostos pertencentes à classe de IC50 Forte/Fraco, foram atribuídos a esta classe. Dos 23 compostos pertencentes à classe de IC50 Fraco/Forte, nenhum foi identificado como pertencente à classe de IC50 Fraco/Forte. Dos restantes 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 41 foram atribuídos à classe de IC50 Fraco/Fraco.

**Tabela 158. Reclassificação dos compostos em análise, representados por MACCS, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
Forte/Forte	40	6	0	12	58
	69%	10%	0%	21%	
Forte/Fraco	12	12	0	22	46
	26%	26%	0%	48%	
Fraco/Forte	12	3	0	8	23
	52%	13%	0%	35%	
Fraco/Fraco	4	9	0	41	54
	7%	17%	0%	76%	

### 3.2.2 Descritores 3D

#### 3.2.2.1 5 $\alpha$ -R1

Recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R1 descritos pelos descritores 3D do Volsurf<sup>®</sup> foram reclassificados como consta na tabela 159, onde se pode verificar que para a classe de IC50 Excelente apenas 22 dos 32 compostos foi correctamente classificado como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 14 dos 64 compostos (38%) foram correctamente reclassificados como pertencentes a esta classe. Para os 54

compostos pertencentes à classe de IC50 Bom, 12 compostos foram atribuídos à classe de IC50 Bom. Relativamente aos 160 compostos pertencentes à classe de IC50 Fraco, 91 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 159. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
<b>Excelente</b>	22	3	2	5	32
	69%	9%	6%	16%	
<b>Muito Bom</b>	21	14	10	19	64
	33%	22%	16%	30%	
<b>Bom</b>	12	9	12	21	54
	22%	17%	22%	39%	
<b>Fraco</b>	28	7	34	91	160
	18%	4%	21%	57%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -RI descritos por descritores 3D foram reclassificados como consta na tabela 160, onde se pode verificar que para a classe de IC50 Excelente, 21, dos 32 compostos, foram classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 14 dos 64 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 54 compostos pertencentes à classe de IC50 Bom, apenas 12 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 102 compostos pertencentes à classe de IC50 Fraco, 22 foram reclassificados como pertencentes à classe de IC50 Fraco. Dos restantes 58 compostos pertencentes à classe de IC50 Outros, 41 foram reclassificados nesta classe.

**Tabela 160. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	21	3	2	1	5	32
	66%	9%	6%	3%	16%	
<b>Muito Bom</b>	20	14	10	2	18	64
	31%	22%	16%	3%	28%	
<b>Bom</b>	10	7	12	10	15	54
	19%	13%	22%	19%	28%	
<b>Fraco</b>	15	9	18	22	38	102
	15%	9%	18%	22%	37%	
<b>Outros</b>	8	1	4	4	41	58
	14%	2%	7%	7%	71%	

Para a análise binária (c), que utiliza os critérios de classificação definidos na tabela 5, os inibidores da 5 $\alpha$ -RI descritos por descritores 3D foram reclassificados como consta na

tabela 161, onde se pode verificar que para a classe de IC50 Forte 95 dos 150 compostos (63%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 116 dos 160 compostos (71%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 161. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	95	55	150
	63%	37%	
Fraco	44	116	160
	28%	73%	

### 3.2.2.2 5 $\alpha$ -R2

Por sua vez, recorrendo à classificação multiclasse (a), que utiliza os critérios de classificação definidos na tabela 3, os inibidores da 5 $\alpha$ -R2 descritos pelos descritores 3D do Volsurf<sup>®</sup> foram reclassificados como consta na tabela 162, onde se pode verificar que para a classe de IC50 Excelente, 111, dos 141 compostos, foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Muito Bom, 6 dos 19 compostos foram reclassificados como pertencentes à classe de IC50 Muito Bom. Na classe de IC50 Bom, 10, dos 42 compostos pertencentes a esta classe, foram correctamente atribuídos a esta classe. Relativamente aos 153 compostos pertencentes à classe de IC50 Fraco, 78 foram reclassificados como pertencentes à classe de IC50 Fraco.

**Tabela 162. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	111	20	7	3	141
	79%	14%	5%	2%	
Muito Bom	6	6	5	2	19
	32%	32%	26%	11%	
Bom	7	18	10	7	42
	17%	43%	24%	17%	
Fraco	18	15	42	78	153
	12%	10%	27%	51%	

Para a seguinte análise multiclasse, que utiliza os critérios de classificação (b) definidos na tabela 4, os inibidores da 5 $\alpha$ -R2 descritos por descritores 3D foram reclassificados como consta na tabela 163, onde se pode verificar que para a classe de IC50 Excelente, 113, dos 141 compostos, foram classificados como pertencentes a esta classe de IC50. Na classe de

IC50 Muito Bom, 6, dos 19 compostos pertencentes a esta classe, foram atribuídos a esta classe. Para os 42 compostos pertencentes à classe de IC50 Bom, 7 foram atribuídos à classe de IC50 Bom pelo classificador. Relativamente aos 97 compostos pertencentes à classe de IC50 Fraco, 30 foram reclassificados como pertencentes à classe de IC50 Fraco. Dos restantes 56 compostos pertencentes à classe de IC50 Outros, 25 foram reclassificados nesta classe.

**Tabela 163. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação (b).**

	Excelente	Muito Bom	Bom	Fraco	Outros	Total
<b>Excelente</b>	113	15	7	5	1	141
	80%	11%	5%	4%	1%	
<b>Muito Bom</b>	6	6	4	2	1	19
	32%	32%	21%	11%	5%	
<b>Bom</b>	3	21	7	4	7	42
	7%	50%	17%	10%	17%	
<b>Fraco</b>	9	15	20	30	23	97
	9%	15%	21%	31%	24%	
<b>Outros</b>	3	5	9	14	25	56
	5%	9%	16%	25%	45%	

Para a análise binária (c), que utiliza os critérios de classificação definidos na tabela 5, os inibidores da 5 $\alpha$ -R2 descritos por descritores 3D foram reclassificados como consta na tabela 164, onde se pode verificar que para a classe de IC50 Forte 170 dos 202 compostos (84%) foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Fraco, 112 dos 153 compostos (73%) foram correctamente reclassificados como pertencentes a esta classe.

**Tabela 164. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em Naïve Bayes, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
<b>Forte</b>	170	32	202
	84%	16%	
<b>Fraco</b>	41	112	153
	27%	73%	

### 3.2.2.3 Inibidores duais

Recorrendo à classificação multiclasse, que utiliza os critérios de classificação definidos na tabela 6, os inibidores com dados de IC50 que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase descritos por descritores 3D foram reclassificados como consta na tabela 165, onde se pode verificar que para a classe de IC50 Forte/Forte 41 dos 58 compostos (71%)

foram correctamente classificados como pertencentes a esta classe de IC50. Na classe de IC50 Forte/Fraco, 28 dos 46 compostos foram correctamente reclassificados como pertencentes a esta classe. Para os 23 compostos pertencentes à classe de IC50 Fraco/Forte, 8 foram atribuídos à classe de IC50 Fraco/Forte pelo classificador. Relativamente aos 54 compostos pertencentes à classe de IC50 Fraco/Fraco, 33 foram reclassificados como pertencentes à classe de IC50 Fraco/Fraco.

**Tabela 165. Reclassificação dos compostos em análise, representados por descritores 3D, que inibem ambas as isoenzimas da 5 $\alpha$ -Redutase aplicando modelos de classificação baseados em *Naïve Bayes*, utilizando validação cruzada 5 vezes e critérios de classificação dos inibidores duais.**

	Forte/Forte	Forte/Fraco	Fraco/Forte	Fraco/Fraco	Total
<b>Forte/Forte</b>	41	6	10	1	58
	71%	10%	17%	2%	
<b>Forte/Fraco</b>	6	28	2	10	46
	13%	61%	4%	22%	
<b>Fraco/Forte</b>	8	2	8	5	23
	35%	9%	35%	22%	
<b>Fraco/Fraco</b>	3	11	7	33	54
	6%	20%	13%	61%	

## 4. Métricas de desempenho

Nas seguintes tabelas estão indicadas as métricas de desempenho para todos os treinos reproduzidos. A *accuracy* e o *Cohen's kappa* são métricas globais do algoritmo de aprendizagem. No caso da sensibilidade, especificidade, precisão e *F-Score* estas aparecem para cada classe de IC50, optando-se por indicar as métricas relativas à classe alvo de cada aprendizagem, ou seja, no caso dos critérios de classificação (a) e (b), a classe alvo é a classe de IC50 Excelente; nos critérios de classificação (c), a classe alvo é a classe de IC50 Forte; nos critérios de classificação dos inibidores duais a classe alvo é a classe de IC50 Forte/Forte.

### 4.1 5 $\alpha$ -RI

#### 4.1.1 SVM

##### 4.1.1.1 Fingerprints

Na tabela 166, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por CHF. O critério de classificação (c) obteve as melhores métricas, à excepção da especificidade (critério de classificação (a)).

**Tabela 166. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	<i>F-Score</i>
(a)	52,258%	0,280	0,219	0,942	0,304	0,255
(b)	43,226%	0,267	0,125	0,939	0,190	0,151
(c)	76,129%	0,524	0,807	0,719	0,729	0,766

Na tabela 167, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por ECFP. O critério de classificação (c) obteve as melhores métricas, à excepção da especificidade (critério de classificação (b)).

**Tabela 167. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	50,000%	0,234	0,250	0,914	0,250	0,250
(b)	45,161%	0,295	0,188	0,917	0,207	0,197
(c)	74,194%	0,484	0,773	0,769	0,758	0,766

Na tabela 168, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -R1 descritos por FP2. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critérios de classificação (a) e (b)).

**Tabela 168. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	48,065%	0,200	0,094	0,910	0,107	0,100
(b)	37,097%	0,182	0,125	0,910	0,138	0,131
(c)	74,194%	0,484	0,760	0,725	0,722	0,740

Na tabela 169, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -R1 descritos por FP3. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (a)).

**Tabela 169. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R1.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	48,387%	0,070	0,250	0,914	0,250	0,250
(b)	32,903%	0,127	0,156	0,903	0,156	0,156
(c)	66,129%	0,328	0,793	0,538	0,617	0,694

Na tabela 170, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e

(c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por FP4. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (b)).

**Tabela 170. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	51,613%	0,234	0,156	0,924	0,192	0,172
(b)	40,968%	0,227	0,094	0,942	0,158	0,118
(c)	71,290%	0,426	0,733	0,694	0,692	0,712

Na tabela 171, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *standard fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (b)).

**Tabela 171. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *standard fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	51,290%	0,255	0,156	0,921	0,185	0,169
(b)	44,839%	0,283	0,219	0,946	0,318	0,259
(c)	71,290%	0,427	0,747	0,681	0,687	0,716

Na tabela 172, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *extended fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (a)).

**Tabela 172. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *extended fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	53,226%	0,272	0,219	0,942	0,304	0,255
(b)	39,032%	0,213	0,125	0,903	0,129	0,127
(c)	69,677%	0,395	0,740	0,656	0,669	0,703

Na tabela 173, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *EState fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (b)).

**Tabela 173. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *EState fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	49,677%	0,187	0,219	0,896	0,194	0,206
(b)	33,548%	0,121	0,188	0,903	0,182	0,185
(c)	64,839%	0,295	0,607	0,688	0,645	0,625

Na tabela 174, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *PubChem fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (b)).

**Tabela 174. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *PubChem fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	43,871%	0,168	0,188	0,910	0,194	0,190
(b)	37,742%	0,196	0,188	0,928	0,231	0,207
(c)	74,194%	0,485	0,780	0,706	0,713	0,745

Na tabela 174, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por MACCS. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (a)).

**Tabela 175. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	50,000%	0,241	0,125	0,928	0,167	0,143
(b)	41,290%	0,239	0,188	0,924	0,222	0,203
(c)	74,516%	0,490	0,780	0,706	0,713	0,745

Para os critérios de correcção (a), o classificador onde foram utilizadas as *Extended fingerprints* obteve a melhor *accuracy* (53,226%), especificidade (tal como as CHF – 0,942), precisão (tal como as CHF – 0,304) e *F-score* (tal como as CHF – 0,255). A utilização das CHF para descrever as moléculas permitiu também obter o melhor *Cohen's kappa* (0,280). Os classificadores que obtiveram melhor sensibilidade foram os que utilizaram as ECFP e as FP3 (0,250). Quando se utilizaram os critérios de classificação (b) para classificar os compostos descritos por *fingerprints*. O classificador onde foram utilizadas as ECFP obteve a melhor *accuracy* (45,161%) e *Cohen's kappa* (0,295). Nas *Standard fingerprint* foram obtidas a melhor sensibilidade (0,219), especificidade (0,946), precisão (0,318) e *F-score* (0,259). Para os critérios de classificação (c), o classificador onde os inibidores da 5 $\alpha$ -RI foram descritos por CHF obteve a melhor *accuracy* (76,129%), *cohen's kappa* (0,524), sensibilidade (0,807) e *F-Score* (tal como nas ECFP – 0,766). No classificador onde se utilizaram as ECFP obteve-se a melhor especificidade (0,769) e precisão (0,758).

#### 4.1.1.2 Descritores 3D

Na tabela 169, é possível verificar as métricas de desempenho dos algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por descritores 3D. As melhores métricas foram obtidas com o critério de classificação (c) – classificação binária – à excepção da especificidade (critério de classificação (a)).

**Tabela 176. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	56,774%	0,236	0,031	0,968	0,100	0,048
(b)	42,258%	0,216	0,094	0,964	0,231	0,133
(c)	70,000%	0,399	0,687	0,712	0,691	0,689

## 4.1.2 Naïve Bayes

### 4.1.2.1 Fingerprints

Na tabela 177, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por CHF. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (b)).

**Tabela 177. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	20,000%	0,047	0,562	0,759	0,212	0,308
(b)	19,677%	0,038	0,531	0,842	0,279	0,366
(c)	70,000%	0,396	0,600	0,794	0,732	0,659

Na tabela 178, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por ECFP. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)) e especificidade (critérios de classificação (a) e (b)).

**Tabela 178. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	20,645%	0,051	0,562	0,817	0,261	0,356
(b)	20,645%	0,052	0,625	0,817	0,282	0,388
(c)	69,032%	0,376	0,587	0,788	0,721	0,647

Na tabela 179, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por FP2. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (b)).

**Tabela 179. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	18,387%	0,028	0,531	0,763	0,205	0,296
(b)	20,323%	0,045	0,469	0,845	0,259	0,333
(c)	72,581%	0,450	0,687	0,762	0,730	0,708

Na tabela 180, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por FP3. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (a)) e especificidade (critério de classificação (b)).

**Tabela 180. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	16,129%	0,012	0,656	0,629	0,169	0,269
(b)	18,387%	0,032	0,500	0,691	0,157	0,239
(c)	65,806%	0,316	0,653	0,662	0,645	0,649

Na tabela 181, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por FP4. O critério de classificação (c) obteve as melhores métricas.

**Tabela 181. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	16,129%	0,013	0,500	0,625	0,130	0,206
(b)	17,419%	0,019	0,469	0,712	0,158	0,236
(c)	67,419%	0,345	0,607	0,738	0,684	0,643

Na tabela 182, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação

(a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *standard fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (a)).

**Tabela 182. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *standard fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	21,613%	0,066	0,625	0,788	0,253	0,360
(b)	19,355%	0,039	0,562	0,788	0,234	0,330
(c)	68,710%	0,367	0,507	0,856	0,768	0,610

Na tabela 183, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *extended fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (a)) e da especificidade (critérios de classificação (a) e (b)).

**Tabela 183. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *extended fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	20,000%	0,044	0,594	0,824	0,279	0,380
(b)	20,000%	0,043	0,531	0,824	0,258	0,347
(c)	70,000%	0,393	0,533	0,856	0,777	0,632

Na tabela 184, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *EState fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (a)).

**Tabela 184. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *EState fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	20,645%	0,050	0,594	0,849	0,311	0,409
(b)	17,742%	0,015	0,594	0,842	0,302	0,400
(c)	63,226%	0,264	0,627	0,638	0,618	0,623

Na tabela 185, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por *PubChem fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critérios de classificação (a) e (b) e da especificidade (critério de classificação (b)).

**Tabela 185. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *PubChem fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	21,613%	0,062	0,594	0,835	0,292	0,392
(b)	21,290%	0,057	0,594	0,845	0,306	0,404
(c)	66,129%	0,318	0,567	0,750	0,680	0,618

Na tabela 186, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por MACCS. O critério de classificação (c) obteve as melhores métricas, à exceção da especificidade (critério de classificação (a)).

**Tabela 186. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	16,452%	0,004	0,500	0,773	0,203	0,288
(b)	17,097%	0,017	0,531	0,741	0,191	0,281
(c)	69,032%	0,379	0,653	0,725	0,690	0,671

Quando se utilizaram os critérios de classificação (a) para classificar os 310 inibidores da 5 $\alpha$ -RI, o classificador onde foram utilizadas as *Standard fingerprints* obteve a melhor *accuracy* (tal como as *PubChem fingerprints* – 21,613%) e *Cohen's kappa* (0,066). No classificador onde se utilizaram as FP3 foi obtida a melhor sensibilidade (0,656). No classificador onde se utilizaram as *EState fingerprints* obteve-se a melhor especificidade (0,849), precisão (0,311) e *F-score* (0,409). Nos critérios de classificação (b), o algoritmo de aprendizagem onde foram utilizadas as *PubChem fingerprints* obteve a melhor *accuracy* (21,290%), *Cohen's kappa* (0,057), especificidade (tal como nas FP2 – 0,845), precisão (0,306) e *F-Score* (0,404). No algoritmo onde se utilizaram as ECFP foi obtida a melhor sensibilidade (0,625). Quando se utilizaram os critérios de classificação (c) para classificar os inibidores da 5 $\alpha$ -RI descritos por *fingerprints*, o algoritmo de aprendizagem onde foram utilizadas as FP2 obteve a melhor *accuracy* (72,581%), *cohen's kappa* (0,450), sensibilidade (0,687) e *F-Score* (0,708) No algoritmo onde foram utilizadas as *Standard fingerprints* obteve-se a melhor especificidade (tal como nas *Extended fingerprints* – 0,856) e precisão (0,777).

#### 4.1.2.2 Descritores 3D

Na tabela 187, é possível verificar as métricas de desempenho do algoritmo de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 310 inibidores da 5 $\alpha$ -RI descritos por descritores 3D. As melhores métricas foram obtidas com o critério de classificação (c) – classificação binária – à exceção da sensibilidade (critério de classificação (a)) e especificidade (critério de classificação (b)).

**Tabela 187. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	<i>F-Score</i>
(a)	44,839%	0,202	0,688	0,781	0,265	0,383
(b)	35,484%	0,208	0,656	0,809	0,284	0,396
(c)	68,065%	0,359	0,633	0,725	0,683	0,657

## 4.2 5 $\alpha$ -R2

### 4.2.1 SVM

#### 4.2.1.1 Fingerprints

Na tabela 188, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por CHF. O critério de classificação (b) obteve as melhores métricas, à excepção da *accuracy* (critério de classificação (c)) e do *Cohen's kappa* (critério de classificação (c)).

**Tabela 188. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	F-Score
(a)	70,986%	0,542	0,894	0,855	0,803	0,846
(b)	63,662%	0,490	0,922	0,879	0,833	0,875
(c)	81,408%	0,617	0,767	0,641	0,738	0,752

Na tabela 189, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por ECFP. O critério de classificação (c) obteve as melhores métricas, à excepção da especificidade (critério de classificação (b)).

**Tabela 189. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	F-Score
(a)	79,718%	0,680	0,894	0,869	0,818	0,854
(b)	65,915%	0,525	0,879	0,874	0,821	0,849
(c)	90,141%	0,799	0,916	0,797	0,856	0,885

Na tabela 190, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por FP2. O critério de classificação

(c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)), da especificidade (critérios de classificação (a) e (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 190. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	78,028%	0,643	0,922	0,879	0,833	0,875
(b)	64,789%	0,501	0,936	0,879	0,835	0,883
(c)	82,535%	0,646	0,866	0,784	0,841	0,854

Na tabela 191, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por FP3. O critério de classificação (c) obteve as melhores métricas, à exceção do *Cohen's Kappa* (critério de classificação (a)).

**Tabela 191. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	64,225%	0,398	0,872	0,696	0,654	0,748
(b)	57,183%	0,364	0,872	0,701	0,658	0,750
(c)	69,859%	0,349	0,901	0,824	0,871	0,886

Na tabela 192, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por FP4. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (a)) e da especificidade (critério de classificação (b)).

**Tabela 192. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	75,211%	0,588	0,943	0,818	0,773	0,850
(b)	61,972%	0,452	0,922	0,841	0,793	0,852
(c)	80,845%	0,608	0,906	0,771	0,839	0,871

Na tabela 193, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *standard fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)), da especificidade (critério de classificação (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 193. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *standard fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	74,648%	0,605	0,894	0,850	0,797	0,843
(b)	63,380%	0,487	0,901	0,855	0,804	0,849
(c)	84,789%	0,686	0,871	0,739	0,815	0,842

Na tabela 194, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *standard fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)), da especificidade (critério de classificação (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 194. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *standard fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	74,648%	0,605	0,894	0,850	0,797	0,843
(b)	63,380%	0,487	0,901	0,855	0,804	0,849
(c)	84,789%	0,686	0,871	0,739	0,815	0,842

Na tabela 195, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *extended fingerprints*. O critério de classificação (c) obteve as melhores métricas.

**Tabela 195. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *extended fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	76,620%	0,634	0,894	0,855	0,803	0,846
(b)	62,535%	0,475	0,894	0,850	0,797	0,843
(c)	86,479%	0,721	0,916	0,882	0,911	0,914

Na tabela 196, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *EState fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção do *Cohen's kappa* (critério de classificação (a)) e da sensibilidade (critério de classificação (b)).

**Tabela 196. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *EState fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	68,451%	0,464	0,858	0,804	0,742	0,796
(b)	54,930%	0,343	0,887	0,785	0,731	0,801
(c)	71,268%	0,410	0,827	0,824	0,861	0,843

Na tabela 197, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *PubChem fingerprints*. O critério de classificação (c) obteve a *accuracy* e *Cohen's kappa*. O critério de classificação (a) obteve a melhor sensibilidade e *F-Score* e o critério de classificação (b) obteve a melhor especificidade e precisão.

**Tabela 197. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e *PubChem fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	78,592%	0,660	0,929	0,869	0,824	0,873
(b)	67,042%	0,537	0,922	0,874	0,828	0,872
(c)	86,761%	0,729	0,911	0,418	0,674	0,775

Na tabela 198, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por MACCS. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)), da especificidade (critérios de classificação (a) e (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 198. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	71,831%	0,552	0,915	0,850	0,801	0,854
(b)	59,718%	0,431	0,922	0,850	0,802	0,858
(c)	83,099%	0,654	0,847	0,758	0,822	0,834

Quando se utilizaram os critérios de classificação (a) para classificar os 355 inibidores da 5 $\alpha$ -R2 e, estes foram descritos por *fingerprints*, o algoritmo de aprendizagem onde foram utilizadas as ECFP obteve a melhor *accuracy* (79,718%) e *Cohen's kappa* (0,680). A melhor sensibilidade (0,943) foi obtida quando se utilizaram as FP4. Na utilização da FP2 obteve-se a melhor especificidade (0,879), precisão (0,833) e *F-score* (0,875). Quando se utilizaram os critérios de classificação (b) para classificar os compostos descritos por *fingerprints*, o algoritmo de aprendizagem onde foram utilizadas as *PubChem fingerprints* obteve a melhor *accuracy* (67,042%) e *Cohen's kappa* (0,537). Com a utilização das FP2 obteve-se a melhor sensibilidade (0,936), especificidade (tal como nas CHF – 0,879), precisão (0,835) e *F-score* (0,883). Quando se utilizaram os critérios de classificação (c) para classificar os inibidores da 5 $\alpha$ -R1 descritos por *fingerprints*, o algoritmo de aprendizagem onde foram utilizadas as ECFP obteve a melhor *accuracy* (90,141%), *cohen's kappa* (0,799) e sensibilidade (tal como nas *Extended fingerprints* – 0,916). Na utilização das *extended fingerprints* obteve-se a melhor especificidade (0,882), precisão (0,911) e *F-Score* (0,914).

#### 4.2.1.2 Descritores 3D

Na tabela 199, é possível verificar as métricas de desempenho do algoritmo de aprendizagem SVM, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por descritores 3D. As melhores métricas foram obtidas com o critério de classificação (c) – classificação binária – à exceção da

sensibilidade (critério de classificação (b)), especificidade (critério de classificação (a)) e *F-Score* (critério de classificação (c)).

**Tabela 199. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	74,930%	0,570	0,887	0,869	0,817	0,850
(b)	62,254%	0,440	0,936	0,841	0,795	0,860
(c)	83,099%	0,651	0,891	0,752	0,826	0,857

## 4.2.1 Naïve Bayes

### 4.2.1.1 Fingerprints

Na tabela 200, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por CHF. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)), da especificidade (critério de classificação (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 200. Métricas de desempenho dos algoritmos de aprendizagem Naïve Bayes, que utilizaram critérios de classificação (a), (b) e (c) e CHF para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	49,577%	0,309	0,844	0,869	0,810	0,826
(b)	43,380%	0,269	0,865	0,883	0,830	0,847
(c)	77,183%	0,547	0,713	0,850	0,862	0,780

Na tabela 201, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por ECFP. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 201. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e ECFP para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	46,197%	0,295	0,936	0,874	0,830	0,880
(b)	44,789%	0,274	0,957	0,860	0,818	0,882
(c)	82,817%	0,658	0,772	0,902	0,912	0,836

Na tabela 202, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por FP2. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)) e do *F-Score* (critério de classificação (a)).

**Tabela 202. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e FP2 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	41,972%	0,248	0,851	0,869	0,811	0,830
(b)	43,099%	0,257	0,872	0,846	0,788	0,828
(c)	77,465%	0,559	0,668	0,915	0,912	0,771

Na tabela 203, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por FP3. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 203. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e FP3 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	39,437%	0,196	0,794	0,724	0,655	0,718
(b)	38,592%	0,186	0,801	0,724	0,657	0,722
(c)	66,761%	0,343	0,609	0,745	0,759	0,676

Na tabela 204, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por FP4. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (a)), da especificidade (critério de classificação (b)) e do *F-Score* (critério de classificação (a)).

**Tabela 204. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e FP4 para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	44,789%	0,282	0,887	0,869	0,817	0,850
(b)	43,099%	0,264	0,865	0,879	0,824	0,844
(c)	80,282%	0,607	0,757	0,863	0,879	0,814

Na tabela 205, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *standard fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critérios de classificação (a) e (b)), da especificidade (critério de classificação (a)) e do *F-Score* (critério de classificação (a)).

**Tabela 205. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *standard fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	41,408%	0,251	0,837	0,911	0,861	0,849
(b)	42,535%	0,263	0,837	0,897	0,843	0,840
(c)	74,648%	0,504	0,644	0,882	0,878	0,743

Na tabela 206, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *extended fingerprints*. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critério de classificação (a)), da especificidade (critério de classificação (b)) e do *F-Score* (critério de classificação (b)).

**Tabela 206. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *extended fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's <i>kappa</i>	Sensibilidade	Especificidade	Precisão	F-Score
(a)	41,408%	0,248	0,809	0,879	0,814	0,811
(b)	39,437%	0,233	0,787	0,921	0,867	0,825
(c)	76,620%	0,543	0,663	0,902	0,899	0,764

Na tabela 207, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *EState fingerprints*. O critério de classificação (c) obteve as melhores métricas, à excepção da sensibilidade (critério de classificação (a)), da especificidade (critério de classificação (a)) e do *F-Score* (critério de classificação (a)).

**Tabela 207. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *EState fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's <i>kappa</i>	Sensibilidade	Especificidade	Precisão	F-Score
(a)	41,127%	0,234	0,823	0,827	0,758	0,789
(b)	40,282%	0,225	0,809	0,822	0,750	0,778
(c)	74,930%	0,488	0,782	0,706	0,778	0,780

Na tabela 208, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por *PubChem fingerprints*. O critério de classificação (c) obteve as melhores métricas, à excepção da sensibilidade (critério de classificação (a)) e do *F-Score* (critério de classificação (a)).

**Tabela 208. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e *PubChem fingerprints* para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's <i>kappa</i>	Sensibilidade	Especificidade	Precisão	F-Score
(a)	48,732%	0,297	0,759	0,869	0,793	0,775
(b)	38,310%	0,218	0,738	0,883	0,806	0,770
(c)	71,831%	0,458	0,569	0,915	0,898	0,697

Na tabela 209, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por MACCS. O critério de classificação (c) obteve as melhores métricas, à exceção da sensibilidade (critérios de classificação (a) e (b)), da especificidade (critério de classificação (a)) e do *F-Score* (critério de classificação (a)).

**Tabela 209. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação (a), (b) e (c) e MACCS para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	<i>F-Score</i>
(a)	43,099%	0,259	0,844	0,836	0,773	0,807
(b)	41,972%	0,241	0,844	0,822	0,758	0,799
(c)	74,930%	0,498	0,723	0,784	0,816	0,766

Quando se utilizaram os critérios de classificação (a) para classificar os 355 inibidores da 5 $\alpha$ -R2, o algoritmo de aprendizagem onde foram utilizadas as CHF obteve a melhor *accuracy* (49,577%) e *Cohen's kappa* (0,309). A melhor sensibilidade (0,936) e *F-Score* (0,880) foram obtidas quando se utilizaram as ECFP. Na utilização de *Standard fingerprints* obteve-se a melhor especificidade (0,911) e precisão (0,861). Quando se utilizaram os critérios de classificação (b) para classificar os inibidores da 5 $\alpha$ -R2 descritos por *fingerprints*, o algoritmo de aprendizagem onde foram utilizadas as ECFP obteve a melhor *accuracy* (44,789%), *cohen's kappa* (0,274), sensibilidade (0,957) e *F-Score* (0,882). A utilização de *extended fingerprints* permitiu obter a melhor especificidade (0,921) e precisão (0,867). Por último, quando se utilizaram os critérios de classificação (c) para classificar os inibidores da 5 $\alpha$ -R2 descritos por *fingerprints*, o algoritmo de aprendizagem onde foram utilizadas as ECFP obteve a melhor *accuracy* (82,817%), *cohen's kappa* (0,658), precisão (tal como nas FP2 – 0,912) e *F-Score* (0,836). A utilização de *EState fingerprints* permitiu obter a melhor sensibilidade (0,782). No algoritmo de aprendizagem onde se utilizou as FP2 obteve-se a melhor especificidade (tal como nas *PubChem fingerprints* – 0,915).

#### 4.2.1.2 Descritores 3D

Na tabela 210, é possível verificar as métricas de desempenho do algoritmo de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação (a), (b) e (c) para classificar os 355 inibidores da 5 $\alpha$ -R2 descritos por descritores 3D. As melhores métricas

foram obtidas com o critério de classificação (c) – classificação binária – à exceção da especificidade (critério de classificação (b)) e precisão (critério de classificação (b)).

**Tabela 210. Métricas de desempenho dos algoritmos de aprendizagem Naïve Bayes, que utilizaram critérios de classificação (a), (b) e (c) e descritores 3D para classificar e descrever os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	57,746%	0,398	0,787	0,855	0,782	0,784
(b)	50,986%	0,353	0,801	0,902	0,843	0,822
(c)	79,437%	0,578	0,842	0,732	0,806	0,823

## 4.3 Inibidores duais

### 4.3.1 SVM

#### 4.3.1.1 Fingerprints

Na tabela 211, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação dos inibidores duais para classificar os 181 inibidores duais da 5 $\alpha$ -Redutase descritos por *fingerprints*. O algoritmo de aprendizagem onde foram utilizadas as MACCS obteve a melhor *accuracy* (71,271%) e *Cohen's kappa* (0,599). A melhor sensibilidade (0,931) foi obtida quando se utilizaram as FP4. Nas ECFP obteve-se a melhor especificidade (0,878), precisão (0,766) e *F-score* (0,803).

**Tabela 211. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação dos inibidores duais e *fingerprints* para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.**

<i>Fingerprint</i>	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
CHF	62,431%	0,477	0,690	0,846	0,678	0,684
ECFP	70,166%	0,585	0,845	0,878	0,766	0,803
FP2	55,801%	0,379	0,724	0,789	0,618	0,667
FP3	50,276%	0,299	0,621	0,715	0,507	0,558
FP4	70,166%	0,581	0,931	0,805	0,692	0,794
Standard fingerprint	65,193%	0,513	0,862	0,813	0,685	0,763
Extended fingerprint	65,746%	0,519	0,845	0,780	0,645	0,731
EState fingerprint	54,144%	0,352	0,707	0,748	0,569	0,631
PubChem fingerprint	67,956%	0,550	0,862	0,829	0,704	0,775
MACCS	71,271%	0,599	0,879	0,813	0,689	0,773

### 4.3.1.2 Descritores 3D

Na tabela 212, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem SVM, quando se utilizaram os critérios de classificação dos inibidores duais para classificar os 181 inibidores duais da 5 $\alpha$ -Redutase descritos por descritores 3D.

**Tabela 212. Métricas de desempenho dos algoritmos de aprendizagem SVM, que utilizaram critérios de classificação dos inibidores duais e descritores 3D para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.**

	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	<i>F-Score</i>
Descritores 3D	60,221%	0,442	0,810	0,813	0,671	0,734

### 4.3.2 Naive Bayes

#### 4.3.2.1 Fingerprints

Na tabela 213, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem Naive Bayes, quando se utilizaram os critérios de classificação (a) para classificar os 181 inibidores duais da 5 $\alpha$ -Redutase descritos por *fingerprints*. O algoritmo de aprendizagem onde foram utilizadas as ECFP obteve a melhor *accuracy* (55,801%), *Cohen's kappa* (0,365), sensibilidade (0,931) e *F-Score* (0,771). A melhor especificidade (0,951) e precisão (0,833) foram obtidas quando se utilizaram as *EState fingerprints*.

**Tabela 213. Métricas de desempenho dos algoritmos de aprendizagem Naive Bayes, que utilizaram critérios de classificação dos inibidores duais e fingerprints para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.**

<i>Fingerprint</i>	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	<i>F-Score</i>
CHF	50,276%	0,285	0,741	0,724	0,558	0,637
ECFP	55,801%	0,365	0,931	0,772	0,659	0,771
FP2	53,591%	0,332	0,897	0,659	0,553	0,684
FP3	41,989%	0,170	0,586	0,764	0,540	0,562
FP4	48,619%	0,269	0,603	0,919	0,778	0,680
<i>Standard fingerprint</i>	46,409%	0,231	0,672	0,789	0,600	0,634
<i>Extended fingerprint</i>	50,829%	0,293	0,810	0,724	0,580	0,676
<i>EState fingerprint</i>	46,409%	0,239	0,517	0,951	0,833	0,638
<i>PubChem fingerprint</i>	50,276%	0,289	0,672	0,756	0,565	0,614
MACCS	51,381%	0,306	0,690	0,772	0,588	0,635

### 4.3.2.2 Descritores 3D

Na tabela 214, é possível verificar as métricas de desempenho dos diferentes algoritmos de aprendizagem *Naïve Bayes*, quando se utilizaram os critérios de classificação dos inibidores duais para classificar os 181 inibidores duais da 5 $\alpha$ -Redutase descritos por descritores 3D.

**Tabela 214. Métricas de desempenho dos algoritmos de aprendizagem *Naïve Bayes*, que utilizaram critérios de classificação dos inibidores duais e descritores 3D para classificar e descrever os inibidores duais da 5 $\alpha$ -Redutase.**

	<i>Accuracy</i>	<i>Cohen's kappa</i>	Sensibilidade	Especificidade	Precisão	<i>F-Score</i>
<b>Descritores 3D</b>	60,773%	0,464	0,707	0,707	0,862	0,707

## **IV. Discussão**

Após serem aplicados os métodos de *clustering* às *fingerprints* e às respectivas estruturas químicas podemos verificar que, no caso dos 32 inibidores da 5 $\alpha$ -R1 pertencentes à classe de IC50 Excelente, o programa LibMCS identifica as 13 MCS representativas dos 13 *clusters*. Estabelecendo uma comparação entre este nível de similaridade e a linha de corte utilizada na análise de *clustering* das *fingerprints*, estes 13 *clusters* encontram-se na linha do número de *clusters* da maioria das *fingerprints* (10 *clusters* – MACCS; 11 *clusters* – FP4; 12 *clusters* – PubChem *fingerprints*; 14 *clusters* – standard e extended *fingerprints*; 15 *clusters* – CHF e FP2). De referir, também, que a ECFP foi a *fingerprint* cuja análise revelou mais *clusters* (25), seguida pelas *hashed fingerprints*, o que indicia uma hierarquização mais acentuada dos compostos com este tipo de *fingerprints*, ao contrário das chaves estruturais, onde se obtiveram menos *clusters* por análise. A *fingerprint* que indicou menos *clusters* foi a FP3. Relativamente aos compostos repartidos pelos diferentes *clusters*, verificou-se que no caso das *hashed fingerprints* (CHF, FP2, standard e extended *fingerprints*) e das PubChem *fingerprints* os compostos repartiram-se sensivelmente pelos mesmos *clusters* não existindo uma variação considerável na composição de cada *cluster*. Deve-se ainda salientar a presença de um *cluster* que agrupa os compostos que possuem a subestrutura 12 transmitindo a capacidade das *fingerprints* em reflectir as estruturas das moléculas. No caso da ECFP, as moléculas que possuem a subestrutura 12 encontram-se repartidas por mais *clusters*, mas sem se misturarem com outras moléculas, deste modo pode-se concluir que se fosse seleccionada uma linha de corte (*threshold*) mais abrangente, estas moléculas já pertenceriam ao mesmo *cluster*.

No caso dos 141 inibidores da 5 $\alpha$ -R2 que pertencem à classe de IC50 Excelente, a análise de *clustering* das estruturas químicas indicou a presença de 27 MCS no nível de similaridade em causa. Neste caso, 4 *fingerprints* obtiveram números de *clusters* semelhantes (26 *clusters* – standard e extended *fingerprints*; 27 *clusters* – FP2; 29 *clusters* – CHF). Mais uma vez a ECFP obteve mais *clusters* do que as restantes *fingerprints* (61) e as *hashed fingerprints* seguiram-se com maior número de *clusters*. As chaves estruturais mais uma vez obtiveram menor número de *clusters*, sendo que no caso da FP3 só foram identificados 9 *clusters*. Nesta comparação houve uma maior dispersão de resultados provavelmente motivada pelo maior número de compostos em análise. Relativamente aos compostos repartidos pelos diferentes *clusters*, verificou-se que no caso das *hashed fingerprints* (CHF, FP2, standard e extended *fingerprints*) e da PubChem *fingerprint* os compostos repartiram-se sensivelmente pelos mesmos *clusters* não existindo uma variação considerável na composição de cada *cluster*. A

semelhança entre a *PubChem fingerprint* e as *hashed fingerprint* pode estar relacionada com o facto da chave estrutural da *PubChem* descrever uma elevada quantidade de fragmentos estruturais, permitindo obter resultados semelhantes às *hashed fingerprints*. De referir também que a ECFP indica *clusters* bastante diferentes das restantes *fingerprints*, mesmo a linhas de cortes mais abrangentes uma vez que agrupa moléculas de diferentes *clusters* no mesmo. Os *clusters* relativos às subestruturas 18, 30 e 35 encontram uma melhor associação aos *clusters* indicados pelo LibMCS nas *hashed fingerprints* e na *PubChem fingerprint*.

Relativamente aos 58 inibidores das duas isoenzimas da 5 $\alpha$ -Redutase que possuem IC50 Forte/Forte a análise de *clustering* da estrutura química identificou 8 MCS. Neste caso, a análise de *clustering* de metade das 10 *fingerprints* situou-se perto deste número (6 *clusters* – FP3 e *PubChem fingerprints*; 10 *clusters* – CHF, FP2, *standard* e *extended fingerprints*). A ECFP obteve mais *clusters* (15), seguida pelas *hashed fingerprints* (10 *clusters*) e por fim, as chaves estruturais, com menos *clusters*. No caso dos compostos repartidos pelos diferentes *clusters*, verificou-se que no caso das *hashed fingerprints* (CHF, FP2, *standard* e *extended fingerprints*) os compostos repartiram-se sensivelmente pelos mesmos *clusters* não existindo uma variação considerável na composição de cada *cluster*. Entre as chaves estruturais as FP4 e *EState fingerprints* obtiveram a maior semelhança na constituição dos *clusters*. De referir que o *cluster* da subestrutura 43 aparece bem representado na maioria das *fingerprints*, à excepção da ECFP onde é preciso uma linha de corte mais abrangente para englobar num grupo os compostos pertencentes a este *cluster*. Os compostos que possuem a subestrutura 44 encontram-se distribuídos por dois grandes *clusters* nas *hashed fingerprints*, por um *cluster* nas chaves estruturais e em vários *clusters* nas ECFP.

Para os 46 inibidores das duas isoenzimas da 5 $\alpha$ -Redutase que possuem IC50 Forte(5 $\alpha$ -R1)/Frac(5 $\alpha$ -R2) a análise realizada pelo programa LibMCS identificou 15 MCS representativas. A análise de *clustering* da maioria das *fingerprints* estabeleceu um número semelhante de *clusters* (13 *clusters* – *PubChem fingerprints*; 14 *clusters* – FP4; 15 *clusters* – *extended fingerprints*; 16 *clusters* – CHF, FP2 e *standard fingerprints*). A ECFP foi a *fingerprint* que indicou maior número de *clusters* (26), seguida pelas *hashed fingerprints* (15 e 16 *clusters*). Das chaves estruturais, a FP3 foi a *fingerprint* cuja análise indicou menos *clusters* (7). Relativamente à composição dos diferentes *clusters*, verificou-se que no caso das *hashed fingerprints* (CHF, FP2, *standard* e *extended fingerprints*) e das *EState* e *PubChem fingerprints* a maioria dos *clusters* possui a mesma composição. De referir que o *cluster* da subestrutura 53 aparece bem representado na maioria das *fingerprints*, à excepção da ECFP, onde é preciso

uma linha de corte mais abrangente para englobar num grupo os compostos pertencentes a este *cluster*, e da FP3 que engloba muitos compostos com subestruturas semelhantes à 53. Os resultados obtidos com a FP3 podem ser justificados com a menor diversidade de fragmentos estruturais descritos pela *fingerprint* o que acaba por colocar compostos diferentes no mesmo *cluster*.

No que respeita aos 23 inibidores das duas isoenzimas da 5 $\alpha$ -Redutase que possuem IC50 Fraco(5 $\alpha$ -R1)/Forte(5 $\alpha$ -R2) a análise de *clustering* à estrutura química revelou 8 MCS. A análise feita com *fingerprints* revelou que praticamente todas as *fingerprints* se aproximaram deste número de grupos (6 *clusters* – FP3, FP4, *EState fingerprints* e MACCS; 7 *clusters* – *PubChem fingerprints*; 11 *clusters* – CHF e FP2; 12 *clusters standard e extended fingerprints*). A ECFP foi, como sempre, a *fingerprint* que revelou maior número de *clusters* (15), sendo seguida pelo grupo das *hashed fingerprints*. Relativamente aos compostos repartidos pelos diferentes *clusters*, verificou-se que no caso das *hashed fingerprints* (CHF, FP2, *standard e extended fingerprints*) os compostos repartiram-se na maioria dos casos pelos mesmos *clusters* não existindo uma variação considerável na composição de cada *cluster*. Os *clusters* que englobam a subestrutura 64 encontram-se bem discriminados pelas *fingerprints* FP3 e *PubChem*. Contudo no caso das restantes *fingerprints* são necessários pelo menos dois *clusters* para englobar as mesmas moléculas. Deste modo, a aplicação de uma linha de corte mais abrangente permitiria colocar todas as moléculas no mesmo *cluster*. Os *clusters* relativos à subestrutura 68 encontram uma melhor associação ao *cluster* indicado pelo LibMCS nas *hashed fingerprints* e na *PubChem fingerprint*.

Relativamente à pesquisa de MCS para os 32 inibidores da 5 $\alpha$ -R1 pertencentes à classe de IC50 Excelente, o programa LibMCS indicou a presença de 13 MCS. A subestrutura mais representada é a 12, que está presente em 11 compostos. Esta subestrutura é um 4-azasteróide. Outros 4-azasteróides também foram identificados pelo LibMCS, como a subestrutura 3, 4 e 5, sendo este a base da maioria dos inibidores da 5 $\alpha$ -R1 que pertencem à classe de IC50 Excelente.

No caso dos 141 inibidores da 5 $\alpha$ -R2 com IC50 Excelente foram detectadas 27 MCS, maioritariamente esteróides. A subestrutura com maior representação, entre os diversos compostos, é a subestrutura 35, que corresponde a um esqueleto 4-azasteróide comum a 40 compostos e que também se encontra na estrutura química da finasterida e dutasterida. Foi também encontrada a subestrutura 11, outro 4-azasteróide com cadeia lateral mais

complexa que a da subestrutura 35, que é comum a 11 compostos. A terceira MCS com maior representação neste grupo de composto foi a subestrutura 18, um 6-azasteróide comum a 9 compostos, cuja cadeia lateral possui um grupo *tert*-butilbenzeno ligado ao átomo de azoto do grupo carbamoilo que se liga ao C<sub>17</sub> do esqueleto esteróide.

No que respeita à pesquisa de MCS nos inibidores das isoenzimas 5 $\alpha$ -R1 e 5 $\alpha$ -R2, nos 58 inibidores duais com IC50 Forte/Forte foram detectadas 8 MCS, sendo a subestrutura 44 a mais comum, uma vez que está presente em 32 compostos. Esta subestrutura caracteriza-se por ser um 4-azasteróide, cuja cadeia lateral possui um anel aromático. A segunda subestrutura mais representada, subestrutura 43, é comum a 14 elementos e consiste num segmento do esqueleto esteróide, que pode ser encontrado na larga maioria dos 58 compostos pertencentes a este grupo.

No caso dos 46 inibidores que pertencem à classe de IC50 Forte/Fraco, foram detectadas 15 MCS. A subestrutura 53 encontra-se no maior número de compostos (17) e reflecte o esqueleto de um 4-azaesteróide. Grande parte das restantes MCS encontrada também são 4-azasteróides.

Para os 23 inibidores das duas isoenzimas da 5 $\alpha$ -Redutase que possuem IC50 Fraco(5 $\alpha$ -R1)/Forte(5 $\alpha$ -R2), a pesquisa de MCS indicou a presença de 13 MCS, com maior incidência para a subestrutura 64, um 4-azasteróide com um grupo carbamoilo ligado ao C<sub>17</sub>, que se encontra em 12 dos 23 compostos.

Através dos classificadores SVM e *Naïve Bayes* com validação cruzada foram analisadas as 10 *fingerprints* e os descritores 3D em estudo para verificar qual o algoritmo de aprendizagem que melhor identifica os inibidores das isoenzimas 5 $\alpha$ -R1 e/ou 5 $\alpha$ -R2. Para os 310 inibidores da 5 $\alpha$ -R1, distribuídos por 4 classes de IC50 (Excelente, Muito Bom, Bom e Fraco), classificação (a), os classificadores SVM que identificaram mais compostos de IC50 Excelente foram os que utilizaram as *fingerprints* ECFP, FP2 e FP3 com 8 em 32 compostos correctamente identificados. Após analisar as métricas de desempenho do algoritmo de aprendizagem destaca-se o classificador onde foram utilizadas as *Extended fingerprints* que obteve a melhor eficiência global, especificidade (em conjunto com as CHF), precisão (em conjunto com as CHF) e *F-score* (em conjunto com as CHF). A utilização das CHF permitiu também obter o melhor *Cohen's kappa*, destacando-se também como a melhor alternativa às *extended fingerprints*.

No caso de se utilizar a classificação (b) para descrever os mesmos compostos, o classificador SVM com mais compostos de IC50 Excelente identificados foi o que utilizou as *standard fingerprints* da CDK<sup>®</sup>. Relativamente às métricas de desempenho, os classificadores que mais se destacaram foram os que utilizaram as ECFP e as *standard fingerprints*. As chaves estruturais FP3 e *EState fingerprints* obtiveram as piores métricas de desempenho.

Com a classificação (c) a CHF permitiu ao respectivo classificador SVM identificar 121 em 150 compostos com IC50 Forte. O classificador onde os inibidores da 5 $\alpha$ -RI foram descritos por CHF obteve as melhores métricas, a par das ECFP. As chaves estruturais FP3 e *EState fingerprints*, mais uma vez voltaram a obter o pior desempenho. Por seu turno as chaves estruturais *PubChem fingerprints* obtiveram melhores resultados do que algumas *hashed fingerprints*, indicando que o grande número de fragmentos estruturais que estão representados nestas *fingerprints* permite melhorar o desempenho dos classificadores.

Comparando para o mesmo conjunto de compostos, os resultados obtidos com descritores 3D sob as mesmas condições, podemos verificar que para os critérios de classificação (a), apesar da melhor eficiência global do algoritmo, a melhoria ocorre pelo aumento de compostos correctamente identificados como pertencem à classe de IC50 Fraco, (aumento da especificidade). Nos critérios de classificação (b), também se verifica um ligeiro aumento da especificidade em relação aos algoritmos que utilizaram *fingerprints*, contudo nos restantes factores não se verificou nenhuma melhoria. Nos critérios de classificação (c), a utilização de *fingerprints* permitiu obter melhores resultados do que os descritores 3D.

Nos resultados respeitantes aos 310 inibidores da 5 $\alpha$ -RI, distribuídos por 4 classes de IC50 (Excelente, Muito Bom, Bom e Fraco), os classificadores construídos com o algoritmo de aprendizagem *Naïve Bayes* que identificaram mais compostos de IC50 Excelente foram os que utilizaram as *fingerprints* FP3 com 21 em 32 compostos correctamente identificados. Após analisar as métricas de desempenho do algoritmo de aprendizagem deve-se salientar que as melhores métricas distribuíram-se por várias *fingerprints* (FP3, *standard*, *EState* e *PubChem fingerprints*), apesar de, no caso da FP3, apenas se destacar a maior sensibilidade, uma vez que a eficiência global do método é muito mais baixa quando comparadas com as restantes *fingerprints*.

No caso de se utilizar a classificação (b) para descrever os mesmos compostos, o algoritmo de aprendizagem *Naïve Bayes* com mais compostos de IC50 Excelente identificados foi o que utilizou as ECFP. Relativamente às métricas de desempenho, os classificadores que

mais se destacaram foram os que utilizaram as *PubChem fingerprints*, as ECFP e as FP2. As restantes 4 chaves estruturais obtiveram as piores métricas de desempenho.

Com a classificação (c), a FP2 permitiu ao respectivo classificador *Naïve Bayes* identificar 103 em 150 compostos com IC50 Forte. O classificador onde os inibidores da 5 $\alpha$ -R1 foram descritos por FP2 obteve as melhores métricas, seguido pelas *extended fingerprints*. As chaves estruturais FP3, FP4, *EState* e *PubChem fingerprints* obtiveram os piores desempenhos. A ECFP não superou os resultados das *hashed fingerprints*, pelo que estas são as *fingerprints* que mais se destacaram positivamente nestas condições.

Comparando para o mesmo conjunto de compostos, os resultados obtidos com descritores 3D sob as mesmas condições, podemos verificar que para os critérios de classificação (a), os descritores 3D permitem obter melhores resultados do que as *fingerprints* aumentando a eficiência global do método de 21,613% (verificado com as *standard* e *PubChem fingerprints*) para 44,839%. Nos critérios de classificação (b), apesar da melhoria não ser tão notória, também se verifica um aumento eficiência do classificador quando são utilizados descritores 3D, em vez de *fingerprints*. Quando são utilizados os critérios de classificação (c), a utilização da maioria das *fingerprints*, nomeadamente as FP2, permitiu obter melhores resultados do que os descritores 3D.

Para os 355 inibidores da 5 $\alpha$ -R2, distribuídos através dos critérios de classificação (a), o classificador SVM que identificou correctamente mais compostos pertencentes à classe de IC50 Excelente foi o que utilizou as *fingerprints* FP4 com 133 em 141 compostos correctamente identificados. Após analisar as métricas de desempenho do algoritmo de aprendizagem destaca-se o classificador onde foram utilizadas as ECFP e as FP2 que obtiveram as melhores métricas de desempenho à excepção da sensibilidade (FP4). As chaves estruturais FP3, *EState fingerprints* e MACCS, juntamente com a *hashed fingerprint* CHF obtiveram os piores desempenhos.

No caso de se utilizar a classificação (b) para descrever os mesmos compostos, o algoritmo de aprendizagem SVM com mais compostos de IC50 Excelente identificados foi o que utilizou as FP2. As métricas de desempenho confirmaram a utilidade destas *fingerprints* uma vez que foram as FP2 que melhores métricas obtiveram, seguidas pelas *PubChem fingerprints*. As chaves estruturais FP3, MACCS e *EState fingerprints* obtiveram as piores métricas de desempenho.

Com a classificação (c) a ECFP permitiu ao respectivo classificador SVM identificar 185 em 202 compostos com IC50 Forte (92%). O classificador onde os inibidores da 5 $\alpha$ -R2 foram descritos por ECFP obteve as melhores métricas, obtendo uma *accuracy* de 90,141%, a melhor de todos os treinos. As chaves estruturais FP3 e *EState fingerprints*, mais uma vez voltaram a obter o pior desempenho. Por seu turno as chaves estruturais *PubChem fingerprints* obtiveram melhores resultados do que algumas *hashed fingerprints*.

Comparando para o mesmo conjunto de compostos, os resultados obtidos com descritores 3D sob as mesmas condições, podemos verificar que para os critérios de classificação (a) e (b), a maioria das *fingerprints* permite obter melhores desempenhos do que os descritores 3D. Os descritores 3D apenas ultrapassam as *fingerprints* FP3, *EState*, MACCS e CHF no primeiro caso e as *fingerprints* FP3, FP4, *EState* e MACCS no segundo. Nos critérios de classificação (c), a utilização de *fingerprints* permitiu obter melhores resultados do que os descritores 3D, exceptuando nas *fingerprints* CHF, FP2, FP3, FP4 e *EState* em que se obteve piores desempenho do que nos descritores 3D.

Nos resultados respeitantes aos 355 inibidores da 5 $\alpha$ -R2 classificados em 4 classes de IC50 (critério de classificação (a)), os classificadores *Naïve Bayes* que identificaram mais compostos de IC50 Excelente foram os que utilizaram as *fingerprints* ECFP com 132 em 141 compostos correctamente identificados. Após analisar as métricas de desempenho do algoritmo de aprendizagem deve-se salientar que as melhores métricas distribuíram-se por três *fingerprints* (CHF, ECFP e *standard fingerprints*), apesar de, no caso das ECFP e *standard fingerprints* possuírem eficiências globais menores quando comparadas com a CHF. A FP3 foi a *fingerprint* que se destacou negativamente em relação às restantes.

No caso de se utilizar a classificação (b) para descrever os mesmos compostos, o algoritmo de aprendizagem *Naïve Bayes* com mais compostos de IC50 Excelente identificados foi o que utilizou as ECFP (identificou 96% dos compostos pertencentes a esta classe). Relativamente às métricas de desempenho, a ECFP confirmou a sua supremacia em relação às restantes *fingerprints*, à excepção da precisão e especificidade onde se destacou a *extended fingerprint*.

Com a classificação (c), a *EState fingerprint* permitiu ao respectivo classificador *Naïve Bayes* identificar 156 em 202 compostos com IC50 Forte. O classificador onde os inibidores da 5 $\alpha$ -R2 foram descritos por ECFP obteve o melhor desempenho global entre outras métricas. Apesar de possuírem menores desempenhos globais, a FP2 destacou-se no caso da

precisão e especificidade, e *EState fingerprint* destacou-se na sensibilidade. As chaves estruturais FP3 obtiveram os piores desempenhos.

Comparando para o mesmo conjunto de compostos, os resultados obtidos com descritores 3D sob as mesmas condições, podemos verificar que para os critérios de classificação (a), os descritores 3D permitem obter melhores resultados do que as *fingerprints* aumentando a eficiência global do método de 49,577% (verificado com as CHF) para 57,746%. Contudo na sensibilidade, especificidade, precisão e *F-Score* as *fingerprints* permitem obter melhores desempenhos. Nos critérios de classificação (b), apesar do desempenho global também ser superior em relação à melhor *fingerprint* (ECFP) é possível obter melhor sensibilidade, especificidade, precisão e *F-Score* com algumas *fingerprints*. Quando são utilizados os critérios de classificação (c), a utilização de *fingerprints* como a ECFP e FP4 permitem obter melhores resultados que os descritores 3D, apesar destes demonstrarem uma sensibilidade superior a qualquer uma das *fingerprints*.

Para os 181 inibidores duais da 5 $\alpha$ -Redutase, distribuídos através dos critérios de classificação dos inibidores duais, o algoritmo de aprendizagem SVM que identificou correctamente mais compostos pertencentes à classe de IC50 Forte/Forte foi o que utilizou as *fingerprints* FP4 com 54 em 58 compostos correctamente identificados. Após analisar as métricas de desempenho do algoritmo de aprendizagem destaca-se o classificador onde foram utilizadas as *fingerprints* MACCS que obtiveram as melhores métricas de desempenho juntamente com as ECFP e FP4. Nas restantes *fingerprints* obtiveram-se piores desempenhos, com especial relevância para as FP2, FP3 e *EState fingerprints*.

Comparando para o mesmo conjunto de compostos, os resultados obtidos com descritores 3D sob as mesmas condições, podemos verificar que para os inibidores duais, a maioria das *fingerprints* permite obter melhor desempenho do que os descritores 3D. Os descritores 3D apenas ultrapassam as *fingerprints* FP2, FP3 e *Estate*.

Nos resultados respeitantes aos 181 inibidores duais da 5 $\alpha$ -Redutase, classificadores *Naive Bayes* que identificaram mais compostos de IC50 Forte/Forte foram os que utilizaram as *fingerprints* ECFP com 54 em 58 compostos correctamente identificados. Após analisar as métricas de desempenho do algoritmo de aprendizagem deve-se salientar que as melhores métricas distribuíram-se pelas ECFP, apesar de as *EState fingerprints* terem obtido maior precisão e especificidade. O classificador onde se utilizaram as FP3 foi o que obteve pior desempenho.

Comparando para o mesmo conjunto de compostos, os resultados obtidos com descritores 3D sob as mesmas condições, podemos verificar que para os inibidores duais, os descritores 3D permitem obter uma melhor eficiência global, *Cohen's kappa* e precisão do que as *fingerprints*. Contudo os melhores valores de sensibilidade, especificidade e *F-Score* pertencem às *fingerprints*.

A comparação dos resultados obtidos entre os classificadores SVM e *Naïve Bayes*, permite estabelecer que para a 5 $\alpha$ -RI, os SVM permitem alcançar melhores valores nas métricas de desempenho, com diferenças, por vezes, significativas, embora a percentagem de compostos da classe alvo identificados correctamente seja superior nos classificadores *Naïve Bayes*. Este aspecto pode ser útil futuramente para aperfeiçoar a pesquisa, uma vez se a utilização dos classificadores SVM fornecer um leque alargado de compostos estes podem ser reduzidos aplicando um classificador *Naïve Bayes*. No caso da 5 $\alpha$ -R2, apesar da eficiência global dos classificadores SVM ser superior aos algoritmos *Naïve Bayes*, a sensibilidade, especificidade, precisão e *F-Score* encontram-se, geralmente, muito próximos, recaindo também a escolha sobre o SVM como método de eleição.

Em termos de critérios de classificação, o critério (c) permite, na sua maioria, resultados significativamente melhores que os outros dois critérios de classificação. Este aspecto pode dever-se ao facto de existir um equilíbrio entre o número de compostos de ambas as classes, não enviesando o classificador a reconhecer os compostos como pertencentes à classe de IC50 Fraco.

A *fingerprint* que permitiu obter, em traços gerais, melhores resultados foi a ECFP, seguida pelo grupo das *hashed fingerprints*. A diferença ao nível do conceito da *fingerprint*, pode explicar esta diferença em relação às restantes, uma vez que têm em conta a variação estrutural e alguns dos caminhos descritos pelas *hashed fingerprints* podem não corresponder a um grupo funcional ou subestrutura facilmente identificável.

Os descritores 3D utilizados, conseguiram, a espaços superar as *fingerprints*, mas salvo uma outra excepção, as diferenças não justificam a preferência sobre estes descritores em vez das *fingerprints*. Isto pode dever-se ao conjunto de descritores 3D utilizado não ser suficiente para potenciar o desempenho dos algoritmos. Este aspecto pode ser testado com a utilização futura de outros descritores 3D que podem melhorar as métricas de desempenho.

No seguimento da tese de mestrado anterior que utiliza 40 descritores moleculares ID para descrever os inibidores da 5 $\alpha$ -Redutase de acordo com os critérios de classificação (a) e (c) (Anexos V e VI), os resultados que correspondem ao classificador SVM para os inibidores da 5 $\alpha$ -RI para o critério de classificação (a) aproximam-se dos registados para os algoritmos de aprendizagem SVM obtidos nas mesmas condições, utilizando *fingerprints* e descritores 3D. Este facto pode dever-se à boa selecção dos descritores ID utilizados na tese anterior. Contudo para os inibidores da 5 $\alpha$ -RI descritos pelo critério de classificação (c) e para os inibidores da 5 $\alpha$ -R2 descritos pelo critério de classificação (a) e (c), os desempenhos dos classificadores são potenciados com a utilização de *fingerprints*, permitindo obter melhores métricas de desempenho. Este aspecto pode relacionar-se com a melhor representação dos compostos conseguida pelas *fingerprints*, uma vez que englobam mais características que os descritores ID.



## **V. Conclusão**

A mortalidade e morbidade das doenças associadas às isoenzimas 5 $\alpha$ -R1 e 5 $\alpha$ -R2, como a hiperplasia benigna da próstata e o cancro da próstata associados aos efeitos adversos dos parcos inibidores da 5 $\alpha$ -Redutase disponíveis, contribuem para um investimento cada vez maior na busca por novos compostos que possam inibir a 5 $\alpha$ -Redutase. Uma vez que a estrutura 3D da enzima 5 $\alpha$ -Redutase ainda não é conhecida, devido à instabilidade associada ao processo de purificação, é necessário recorrer a outros métodos de rastreio virtual para tentar encontrar novos inibidores desta enzima.

Neste seguimento, o presente trabalho pretende reunir informações sobre a estrutura química dos inibidores da 5 $\alpha$ -Redutase já conhecidos e sobre a melhor maneira de os representar computacionalmente, para permitir novas pesquisas em bases de dados alargadas, de modo a contribuir para o desenvolvimento de inibidores mais potentes, selectivos e com menos toxicidade.

Após reunir os inibidores das isoenzimas 5 $\alpha$ -R1 e 5 $\alpha$ -R2 com IC50 descrito na base de dados ChEMBL, seleccionaram-se os descritores moleculares 2D (*fingerprints*) e 3D (Volsurf<sup>®</sup>) que iriam ser utilizados nas análises de *clustering* e nos métodos de aprendizagem de máquina. Através da análise de *clustering* realizada verificou-se que as *fingerprints* que produzem resultados mais semelhantes à análise de *clustering* das estruturas 2D dos compostos em causa são as *hashed fingerprints* e por vezes, as chaves estruturais, tendo em conta a linha corte utilizada. A ECFP por ser uma *fingerprint* mais complexa divide os compostos em mais *clusters* do que as restantes *fingerprints*, sendo necessário uma linha de corte menos restritiva para se aproximar dos resultados das restantes *fingerprints*.

Por seu turno, a pesquisa de MCS nos inibidores da 5 $\alpha$ -R1 e 5 $\alpha$ -R2 pertencentes à classe de IC50 Excelente e nos inibidores duais pertencentes à classe de IC50 Forte/Forte, Forte/Fraco e Fraco/Forte indicou uma larga predominância de estruturas esteróides, com especial ênfase nos 4-azasteróides, também presentes na estrutura química da finasterida e dutasterida, os dois inibidores comercializados.

Por último, nos métodos de aprendizagem de máquina, verificou-se que, de uma maneira geral, o método SVM apresenta melhores resultados do que o *Naïve Bayes*, sendo preferível utilizá-lo, no caso destas isoenzimas, em pesquisas futuras. As *fingerprints* que, geralmente, proporcionam melhor desempenho são as ECFP, seguidas pelas *hashed fingerprints* e *PubChem fingerprint*. A *PubChem fingerprint* consegue diferenciar-se das restantes estruturais uma vez que descreve 881 fragmentos estruturais ao contrário das restantes, que

descrevem muito menos fragmentos. Em comparação com os descritores 3D do Volsurf<sup>®</sup>, as *fingerprints* permitem obter resultados semelhantes, salvo algumas exceções, sendo necessário estudar outros descritores 3D, para confirmar esta avaliação. Deve-se referir que o número de compostos presentes em cada classe contribui significativamente para o sucesso da previsão pelo classificador, uma vez que os melhores resultados a nível de previsão encontram-se nas classes de IC50 Forte e Fraco, as classes com maior número de compostos.

Como perspectiva de trabalho futuro seria interessante testar alguns algoritmos aqui treinados em bases de dados maiores e, para os compostos daí resultantes que pertençam à classe de IC50 desejável, testá-los laboratorialmente para avaliar a sua actividade.



## **VI. Referências bibliográficas**

AGGARWAL, Saurabh et al. – **Self-organizing molecular field analysis on pregnane derivatives as human steroidal 5 $\alpha$ -reductase inhibitors.** Steroids, 75, 6 (2010) 411–418.

AGGARWAL, Saurabh et al. – **An overview on 5 $\alpha$ -reductase inhibitors.** Steroids, 75, 2 (2010) 109–153.

AZZOUNI, Faris et al. – **The 5 Alpha-Reductase Isozyme Family: A Review of Basic Biology and Their Role in Human Diseases.** Advances in Urology, 2012 (2012) 1-18.

BENDER, Andreas et al. – **How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space.** Journal of Chemical Information and Modeling, 49,1 (2009) 108–119.

BIELSKA, Ewa et al. – **Virtual screening strategies in drug design – methods and applications.** Journal of Biotechnology, Computational Biology and Bionanotechnology, 92, 3 (2011) 249-264.

BULL, Herbert. G. et al. – **Mechanism-Based Inhibition of Human Steroid 5 $\alpha$ -Reductase by Finasteride: Enzyme-Catalyzed Formation of NADP-Dihydrofinasteride, a Potent Bisubstrate Analog Inhibitor.** Journal of the American Chemical Society, 118, 10 (1996) 2359-2365.

CAI, Chaoqian et al. – **Molecular Similarity: Methods and Performance.** Chinese Journal of Chemistry, 31 (2013) 1123-1132.

CAO, Yiqun; JIANG, Tao; GIRKE, Thomas – **A maximum common substructure-based algorithm for searching and predicting drug-like compounds.** Bioinformatics, 24 (2008) 366-374.

CORTES, Corinna; VAPNIK, Vladimir – **Support-Vector Networks.** Machine Learning, 20 (1995) 273-297.

CUNNINGHAM, Pádraig; CORD, Matthieu; DELANY, Sarah Jane – **Supervised Learning.** In: CUNNINGHAM, Pádraig; CORD, Matthieu, Machine Learning Techniques for Multimedia, California: Springer-Verlag TELOS Santa Clara, 2008, ISBN: 354075170X 9783540751700.

DIESTEL, Reinhard – **Graphs**. In: DIESTEL, Reinhard, Graph Theory, New York: Springer-Verlag, 2000, ISBN: 0-387-98976-5.

DIMOVA, Dilyana; STUMPFE, Dagmar; BAJORATH, Jürgen – **Quantifying the Fingerprint Descriptor Dependence of Structure–Activity Relationship Information on a Large Scale**. Journal of Chemical Information and Modeling, 53 (2013) 2275–2281.

ECKERT, Hanna; BAJORATH, Jürgen – **Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches**. Drug Discovery Today, 12, 5/6 (2007) 225-233.

FONTAINE, Fabien; PASTOR, Manuel – **Development and applications of new 3D molecular descriptors**. Barcelona: Pompeu Fabra University, 2004. [Acedido a 10 de Fevereiro de 2014]. Disponível na Internet:  
<http://www.tdx.cat/bitstream/handle/10803/7080/tfdflde1.pdf;jsessionid=EAA884CB358F58CF26C93FB0B5D39A5B.tdx2?sequence=1>

GARCÍA, S. et al. – **A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability**. Applied Soft Computing, 13 (2009) 959-977.

GEPPERT, Hanna; VOGT, Martin; BAJORATH, Jürgen – **Current trends ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation**. Journal of Chemical Information and Modeling, 50 (2010) 205–216.

GLEM, Robert C. et al. – **Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME**. IDrugs, 9, 3 (2006) 199-204.

GODDEN, Jeffrey W.; STAHURA, Florence L.; BAJORATH, Jürgen – **Anatomy of Fingerprint Search Calculations on Structurally Diverse Sets of Active Compounds**. Journal of Chemical Information and Modeling, 45, 6 (2005) 1812-1819.

GUIDO, Rafael V. C.; OLIVA, Glaucius; ANDRICOPULO, Adriano D. – **Virtual Screening and Its Integration with Modern Drug Design Technologies**. Current Medicinal Chemistry, 15, 1 (2008) 37-46.

HATTORI, Kazunari; WAKABAYASHI, Hiroaki; TAMAKI, Kenta – **Predicting Key Example Compounds in Competitors' Patent Applications Using Structural Information Alone.** *Journal of Chemical Information and Modeling*, 48, 1 (2008) 135-142.

HERT, Jérôme et al. – **Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures.** *Organic & Biomolecular Chemistry*, 2, 22 (2004) 3256–66.

HOSODA, Shinnosuke; HASHIMOTO, Yuichi – **3,3-Diphenylpentane skeleton as a steroid skeleton substitute: Novel inhibitors of human 5 $\alpha$ -reductase I.** *Bioorganic & Medicinal Chemistry Letters*, 17 (2007) 5414–5418.

JAMES, C.A.; WEININGER, D. – **Daylight Theory Manual.** California: Daylight Chemical Information Systems. Inc. of Aliso Viejo, 2011. [Acedido a 8 de Dezembro de 2013].

Disponível na Internet:

<http://www.daylight.com/dayhtml/doc/theory/index.pdf>.

KIM, Sang.-Bum et al. – **Some Effective Techniques for Naive Bayes Text Classification.** *IEEE Transactions on knowledge and data engineering* 18, 11 (2006) 1457-1466.

KOTSIANTIS, S. B. – **Supervised Machine Learning: A Review of Classification Techniques.** *Informatica*, 31 (2007) 249-268.

LAVECCHIA, A., DI GIOVANNI, C. – **Virtual Screening Strategies in Drug Discovery: A Critical Review.** *Current Medicinal Chemistry*, 20, 1 (2013).

LEACH, Andrew R.; GILLET, Valerie J. – **An Introduction to Chemoinformatics.** Revised Edition. Dordrecht: Springer, 2007. ISBN 978-1-4020-6291-9

LESZCZYNSKI, Jerzy; PUZYN, Tomasz; CRONIN, Mark T. D. – **Recent Advances in QSAR Studies. Methods and Applications.** Vol. 8. Dordrecht: Springer, 2010.

LYNE, Paul – **Structure-based virtual screening: an overview.** *Drug Discovery Today*, 7, 20 (2002) 1047-1055.

MEINL, Thorsten et al. – **The Konstanz Information Miner 2.0.** In: *Open Source in Data Mining Workshop*, Bangkok. Thailand: 2009.

MITCHELL, Tom M. – **Generative and discriminative classifiers: Naïve Bayes and logistic regression**. In: MITCHELL, Tom M., Machine Learning. New York: McGraw Hill, 2010. ISBN 0070428077.

**VolSurf+ user manual**. In: Molecular Discovery Ltd, 2008.

MORGAN, H. L. – **The Generation of a Unique Machine Description for Chemical Structures – A Technique Developed at Chemical Abstracts Service**. Journal of Chemical Documentation, 2 (1965) 107-113.

O'BOYLE, Noel M. et al. – **Open Babel: An open chemical toolbox**. Journal of Cheminformatics, 3, 33 (2011) 1-14.

OKEH, U. M.; UGWU, A. C. – **Bayes' theorem: A paradigm research tool in biomedical sciences**. African Journal of Biotechnology, 7, 25 (2008) 4807-4818.

OLIVEIRA, Osmar; KOFF, Walter; TRINDADE, Vera – **Avaliação da actividade da enzima esteróide 5 $\alpha$ -Redutase tipo 2 em biópsias de próstata**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2005. [Acedido a 9 de Janeiro de 2014]

Disponível na Internet:

<http://www.lume.ufrgs.br/bitstream/handle/10183/6091/000524855.pdf?sequence=1>

PROMPRAMOTE, Supawan; CHEN, Yan; CHEN, Yi-Ping – **Machine Learning in Bioinformatics**. In: CHEN, Yi-Ping, Bioinformatics Technologies. Australia: Springer Science & Business Media, 2005. ISBN: 354026888X.

RAYMOND, John W.; WILLET, Peter – **Maximum common subgraph isomorphism algorithms for the matching of chemical structures**. Journal of Computer-Aided Molecular Design 16 (2002) 521–533.

REDDY, A. Srinivas et al. – **Virtual Screening in Drug Discovery – A Computational Perspective**. Current Protein and Peptide Science 8 (2007) 329-351.

REFAEILZADEH, Payam; TANG, Lei; LIU, Huan – **Cross-validation**. Encyclopedia of Database Systems, (2012) 532-538.

ROGERS, David; HAHN, Mathew – **Extended-Connectivity Fingerprints**. Journal of Chemical Information and Modeling, 50, 5 (2010) 742–754.

SALVADOR, Jorge A. R., PINTO, Rui, SILVESTRE, Samuel – **Steroidal 5 $\alpha$ -reductase and 17 $\alpha$ -hydroxylase/17,20-lyase (CY17) inhibitors useful in the treatment of prostatic diseases.** Journal of Steroid Biochemistry and Molecular Biology, 137 (2013) 199–222.

SANDERS, Marijn P. A. et al. – **Comparative Analysis of Pharmacophore Screening Tools.** Journal of Chemical Information and Modeling, 52 (2012) 1607–1620.

SARANYA, C., MANIKANDAN, G. – **A Study on Normalization Techniques for Privacy Preserving Data Mining.** International Journal of Engineering and Technology, 5, 3 (2013) 2701-2704.

SOKOLOVA, Marina; JAPKOWICZ, Nathalie; SZPAKOWICZ, Stan – (2006). **Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation.** [Acedido a 20 de Abril de 2014]

Disponível na Internet:

<http://aaaipress.org/Papers/Workshops/2006/WS-06-06/WS06-06-006.pdf>

SOUSA, Joana F. M., SILVA, Cândida. G., SALVADOR, Jorge. A. R. – **Rastreo Virtual na Descoberta de Possíveis Inibidores de 5 $\alpha$ -Redutase.** Coimbra: Faculdade de Farmácia da Universidade de Coimbra, 2013.

TODESCHINI, Roberto; CONSONNI, Viviana – **Handbook of Molecular Descriptors.** Weinheim: Wiley-VCH Verlag, 2000. ISBN: 3-52-29913-0.

TRAISH, Abdulmaged. M. et al – **Adverse Side Effects of 5 $\alpha$ -Reductase Inhibitors Therapy: Persistent Diminished Libido and Erectile Dysfunction and Depression in a Subset of Patients.** The Journal of Sexual Medicine, 8 (2011) 872–884.

VENKATRAMAN, Vishwesh et al. – **Comprehensive Comparison of Ligand-Based Virtual Screening Tools Against the DUD Data set Reveals Limitations of Current 3D Methods.** Journal of Chemical Informations and Modeling, 50, 12 (2010) 2079-2093.

VERMA, Jitander; KHEDKAR, Vijay M.; COUTINHO, Evans C. – **3D-QSAR in Drug Design - A Review.** Current Topics in Medicinal Chemistry, 10, 1 (2010) 95-115.

WALTERS, W. Patrick; STAHL, Matthew T.; MURCKO, Mark A. – **Virtual screening – an overview**. Drug Discovery Today, 3 (1998) 160-178.

WASSERMANN, Anne Mai; GEPPERT, Hanna; BAJORATH, Jürgen – **Searching for Target-Selective Compounds Using Different Combinations of Multiclass Support Vector Machine Ranking Methods, Kernel Functions, and Fingerprint Descriptors**. Journal of Chemical Informations and Modeling, 49, 3 (2009) 582–592.

WILLET, Peter – **Searching Techniques for Databases of Two- and Three-Dimensional Chemical Structures**. Journal of Medicinal Chemistry, 48, 13 (2005) 4183-4199.

ZHAO, Ying; KARYPIS, George – **Data Clustering in Life Sciences**. Molecular Biotechnology, 31 (2005) 55-80.



## **VII. Anexos**

## I. ChEMBL ID dos inibidores da 5 $\alpha$ -RI

Na tabela 215 podemos observar ChEMBL IDs dos 310 inibidores da 5 $\alpha$ -RI com IC50 reportado.

**Tabela 215. ChEMBL ID dos 310 inibidores da 5 $\alpha$ -RI com IC50 reportado na ChEMBL.**

ChEMBL ID	ChEMBL ID	ChEMBL ID	ChEMBL ID	ChEMBL ID
CHEMBL107565	CHEMBL118319	CHEMBL16248	CHEMBL290823	CHEMBL326743
CHEMBL107654	CHEMBL118390	CHEMBL16494	CHEMBL335919	CHEMBL327042
CHEMBL107719	CHEMBL118416	CHEMBL16330	CHEMBL336854	CHEMBL327098
CHEMBL109170	CHEMBL118446	CHEMBL16456	CHEMBL340961	CHEMBL331253
CHEMBL109560	CHEMBL118447	CHEMBL16533	CHEMBL414428	CHEMBL332407
CHEMBL114245	CHEMBL118452	CHEMBL16690	CHEMBL419602	CHEMBL332413
CHEMBL116432	CHEMBL118543	CHEMBL16927	CHEMBL56908	CHEMBL332781
CHEMBL116830	CHEMBL118573	CHEMBL16700	CHEMBL710	CHEMBL333183
CHEMBL117159	CHEMBL118605	CHEMBL16784	CHEMBL76959	CHEMBL333684
CHEMBL118100	CHEMBL118822	CHEMBL16863	CHEMBL77117	CHEMBL334649
CHEMBL118255	CHEMBL118824	CHEMBL17025	CHEMBL86418	CHEMBL336200
CHEMBL118389	CHEMBL119138	CHEMBL17193	CHEMBL99615	CHEMBL338207
CHEMBL118943	CHEMBL119164	CHEMBL17194	CHEMBL275153	CHEMBL338376
CHEMBL119255	CHEMBL119259	CHEMBL17206	CHEMBL276320	CHEMBL338707
CHEMBL119361	CHEMBL119722	CHEMBL17222	CHEMBL276355	CHEMBL339912
CHEMBL124750	CHEMBL1201841	CHEMBL17473	CHEMBL276871	CHEMBL340027
CHEMBL125443	CHEMBL120310	CHEMBL17245	CHEMBL277053	CHEMBL340181
CHEMBL125599	CHEMBL120923	CHEMBL17268	CHEMBL277091	CHEMBL340305
CHEMBL125855	CHEMBL122027	CHEMBL17422	CHEMBL277403	CHEMBL340684
CHEMBL125899	CHEMBL122290	CHEMBL17506	CHEMBL277664	CHEMBL340715
CHEMBL126968	CHEMBL122359	CHEMBL17514	CHEMBL278490	CHEMBL341004
CHEMBL128023	CHEMBL1237306	CHEMBL180235	CHEMBL278787	CHEMBL341088
CHEMBL132086	CHEMBL1237307	CHEMBL180451	CHEMBL278981	CHEMBL344363
CHEMBL134357	CHEMBL124803	CHEMBL181196	CHEMBL279134	CHEMBL362379
CHEMBL135288	CHEMBL124849	CHEMBL182391	CHEMBL279206	CHEMBL36573
CHEMBL136097	CHEMBL125073	CHEMBL182399	CHEMBL279417	CHEMBL400059
CHEMBL155094	CHEMBL125209	CHEMBL182787	CHEMBL279420	CHEMBL411431
CHEMBL16027	CHEMBL125256	CHEMBL182945	CHEMBL279131	CHEMBL414427
CHEMBL16743	CHEMBL125257	CHEMBL183007	CHEMBL280387	CHEMBL415016
CHEMBL24955	CHEMBL125260	CHEMBL183167	CHEMBL280954	CHEMBL417109
CHEMBL25083	CHEMBL125442	CHEMBL183185	CHEMBL281204	CHEMBL417627
CHEMBL25578	CHEMBL125481	CHEMBL183395	CHEMBL283245	CHEMBL420415
CHEMBL25664	CHEMBL125579	CHEMBL183692	CHEMBL283508	CHEMBL421627
CHEMBL266519	CHEMBL125668	CHEMBL183448	CHEMBL283981	CHEMBL424084
CHEMBL108028	CHEMBL125769	CHEMBL183500	CHEMBL285439	CHEMBL425878
CHEMBL108151	CHEMBL126686	CHEMBL183820	CHEMBL285599	CHEMBL432439

<b>CHEMBL108181</b>	CHEMBL127702	CHEMBL24191	CHEMBL29166	CHEMBL432745
<b>CHEMBL108305</b>	CHEMBL127760	CHEMBL24193	CHEMBL294630	CHEMBL432977
<b>CHEMBL108559</b>	CHEMBL128303	CHEMBL24464	CHEMBL300446	CHEMBL439908
<b>CHEMBL108612</b>	CHEMBL128728	CHEMBL24465	CHEMBL30223	CHEMBL439910
<b>CHEMBL108625</b>	CHEMBL129017	CHEMBL248682	CHEMBL308610	CHEMBL441287
<b>CHEMBL108875</b>	CHEMBL129130	CHEMBL24958	CHEMBL308819	CHEMBL441471
<b>CHEMBL109088</b>	CHEMBL130845	CHEMBL25072	CHEMBL311013	CHEMBL443995
<b>CHEMBL109089</b>	CHEMBL131341	CHEMBL251109	CHEMBL311624	CHEMBL51628
<b>CHEMBL109232</b>	CHEMBL131803	CHEMBL251297	CHEMBL312466	CHEMBL55260
<b>CHEMBL110058</b>	CHEMBL134114	CHEMBL251506	CHEMBL321442	CHEMBL55261
<b>CHEMBL110288</b>	CHEMBL134810	CHEMBL251507	CHEMBL321617	CHEMBL55402
<b>CHEMBL110398</b>	CHEMBL134819	CHEMBL25193	CHEMBL321867	CHEMBL56517
<b>CHEMBL111142</b>	CHEMBL135014	CHEMBL25448	CHEMBL321923	CHEMBL56518
<b>CHEMBL111407</b>	CHEMBL135351	CHEMBL264316	CHEMBL322592	CHEMBL56581
<b>CHEMBL111476</b>	CHEMBL135651	CHEMBL269632	CHEMBL322749	CHEMBL57789
<b>CHEMBL114211</b>	CHEMBL135993	CHEMBL274238	CHEMBL323031	CHEMBL58989
<b>CHEMBL116370</b>	CHEMBL137040	CHEMBL274269	CHEMBL323033	CHEMBL76373
<b>CHEMBL116493</b>	CHEMBL138580	CHEMBL274276	CHEMBL323788	CHEMBL76687
<b>CHEMBL116604</b>	CHEMBL14575	CHEMBL27549	CHEMBL323856	CHEMBL77167
<b>CHEMBL116999</b>	CHEMBL152401	CHEMBL276527	CHEMBL323996	CHEMBL77169
<b>CHEMBL118091</b>	CHEMBL153258	CHEMBL277563	CHEMBL324210	CHEMBL77649
<b>CHEMBL118118</b>	CHEMBL155568	CHEMBL278049	CHEMBL324687	CHEMBL77650
<b>CHEMBL118283</b>	CHEMBL15917	CHEMBL278957	CHEMBL325572	CHEMBL77789
<b>CHEMBL138225</b>	CHEMBL16152	CHEMBL283564	CHEMBL326439	CHEMBL80175
<b>CHEMBL2311155</b>	CHEMBL280015	CHEMBL359360	CHEMBL42258	CHEMBL96006
<b>CHEMBL279745</b>	CHEMBL288893	CHEMBL41778	CHEMBL422865	CHEMBL99448

## 2. ChEMBL ID dos inibidores da 5 $\alpha$ -R2

Na tabela 216 podemos observar ChEMBL IDs dos 355 inibidores da 5 $\alpha$ -R2 com IC50 reportado.

**Tabela 216. ChEMBL ID dos 355 inibidores da 5 $\alpha$ -R2 com IC50 reportado na ChEMBL.**

ChEMBL ID	ChEMBL ID	ChEMBL ID	ChEMBL ID	ChEMBL ID
CHEMBL107565	CHEMBL16533	CHEMBL280954	CHEMBL77290	CHEMBL310711
CHEMBL107654	CHEMBL16690	CHEMBL283123	CHEMBL77378	CHEMBL311045
CHEMBL107719	CHEMBL16700	CHEMBL283245	CHEMBL77789	CHEMBL311577
CHEMBL108028	CHEMBL16743	CHEMBL283564	CHEMBL80163	CHEMBL311783
CHEMBL108151	CHEMBL16784	CHEMBL283981	CHEMBL86418	CHEMBL311818
CHEMBL108181	CHEMBL16863	CHEMBL285439	CHEMBL87021	CHEMBL311874
CHEMBL108305	CHEMBL16927	CHEMBL285599	CHEMBL87805	CHEMBL312143
CHEMBL108559	CHEMBL17025	CHEMBL290823	CHEMBL89240	CHEMBL312531
CHEMBL108612	CHEMBL17193	CHEMBL298709	CHEMBL99615	CHEMBL314069
CHEMBL108625	CHEMBL17194	CHEMBL300446	CHEMBL114594	CHEMBL314309
CHEMBL108875	CHEMBL17206	CHEMBL306289	CHEMBL114948	CHEMBL314953
CHEMBL109088	CHEMBL17222	CHEMBL306507	CHEMBL115019	CHEMBL314977
CHEMBL109089	CHEMBL17245	CHEMBL306722	CHEMBL1159458	CHEMBL327702
CHEMBL109170	CHEMBL17268	CHEMBL306948	CHEMBL117643	CHEMBL352434
CHEMBL109232	CHEMBL17422	CHEMBL308532	CHEMBL117881	CHEMBL353524
CHEMBL109560	CHEMBL17473	CHEMBL314725	CHEMBL1237294	CHEMBL355567
CHEMBL110058	CHEMBL17506	CHEMBL321442	CHEMBL128342	CHEMBL36392
CHEMBL110288	CHEMBL17514	CHEMBL321617	CHEMBL135313	CHEMBL370016
CHEMBL110398	CHEMBL180235	CHEMBL321867	CHEMBL135780	CHEMBL382214
CHEMBL111142	CHEMBL181196	CHEMBL321923	CHEMBL136863	CHEMBL420020
CHEMBL111407	CHEMBL182391	CHEMBL322592	CHEMBL137691	CHEMBL421735
CHEMBL111476	CHEMBL182399	CHEMBL322749	CHEMBL137733	CHEMBL424066
CHEMBL113726	CHEMBL182787	CHEMBL323031	CHEMBL138173	CHEMBL430617
CHEMBL114114	CHEMBL182945	CHEMBL323033	CHEMBL138225	CHEMBL432641
CHEMBL115123	CHEMBL183167	CHEMBL323788	CHEMBL1627395	CHEMBL432729
CHEMBL115298	CHEMBL183185	CHEMBL323996	CHEMBL1642919	CHEMBL436215
CHEMBL115438	CHEMBL183448	CHEMBL324210	CHEMBL1642920	CHEMBL442747
CHEMBL116370	CHEMBL183500	CHEMBL324687	CHEMBL166881	CHEMBL445627
CHEMBL116621	CHEMBL183820	CHEMBL325572	CHEMBL169621	CHEMBL44785
CHEMBL117793	CHEMBL1908332	CHEMBL326439	CHEMBL1794821	CHEMBL45241
CHEMBL118283	CHEMBL201425	CHEMBL327098	CHEMBL201138	CHEMBL45671
CHEMBL118389	CHEMBL201762	CHEMBL328670	CHEMBL201699	CHEMBL47679
CHEMBL118390	CHEMBL24191	CHEMBL332781	CHEMBL201736	CHEMBL47785
CHEMBL118416	CHEMBL24193	CHEMBL334649	CHEMBL202871	CHEMBL47959
CHEMBL119722	CHEMBL24291	CHEMBL338376	CHEMBL203352	CHEMBL73816
CHEMBL1201841	CHEMBL24464	CHEMBL339912	CHEMBL2111947	CHEMBL75320

<b>CHEMBL120923</b>	CHEMBL24465	CHEMBL340027	CHEMBL2111948	CHEMBL75926
<b>CHEMBL122359</b>	CHEMBL24929	CHEMBL340181	CHEMBL2112735	CHEMBL75982
<b>CHEMBL1237306</b>	CHEMBL24955	CHEMBL340305	CHEMBL24088	CHEMBL76030
<b>CHEMBL1237307</b>	CHEMBL24958	CHEMBL340684	CHEMBL24470	CHEMBL76192
<b>CHEMBL124750</b>	CHEMBL25072	CHEMBL340715	CHEMBL25516	CHEMBL76260
<b>CHEMBL124803</b>	CHEMBL25083	CHEMBL340961	CHEMBL25593	CHEMBL76648
<b>CHEMBL125209</b>	CHEMBL25183	CHEMBL341004	CHEMBL262635	CHEMBL76671
<b>CHEMBL125260</b>	CHEMBL25193	CHEMBL341088	CHEMBL277224	CHEMBL76693
<b>CHEMBL125579</b>	CHEMBL25196	CHEMBL344363	CHEMBL278860	CHEMBL76803
<b>CHEMBL126686</b>	CHEMBL25282	CHEMBL352942	CHEMBL280155	CHEMBL76892
<b>CHEMBL126968</b>	CHEMBL25448	CHEMBL362379	CHEMBL282037	CHEMBL77024
<b>CHEMBL127702</b>	CHEMBL25578	CHEMBL36573	CHEMBL283430	CHEMBL77124
<b>CHEMBL127760</b>	CHEMBL25664	CHEMBL382649	CHEMBL284645	CHEMBL77129
<b>CHEMBL128023</b>	CHEMBL264316	CHEMBL411431	CHEMBL295396	CHEMBL77144
<b>CHEMBL128303</b>	CHEMBL266519	CHEMBL412425	CHEMBL296685	CHEMBL77328
<b>CHEMBL128728</b>	CHEMBL269632	CHEMBL414427	CHEMBL297174	CHEMBL77765
<b>CHEMBL129017</b>	CHEMBL274238	CHEMBL414428	CHEMBL297524	CHEMBL78064
<b>CHEMBL129130</b>	CHEMBL274269	CHEMBL415016	CHEMBL297697	CHEMBL78161
<b>CHEMBL130845</b>	CHEMBL274276	CHEMBL417627	CHEMBL301725	CHEMBL78275
<b>CHEMBL131341</b>	CHEMBL275153	CHEMBL420415	CHEMBL305730	CHEMBL78382
<b>CHEMBL135351</b>	CHEMBL276320	CHEMBL421627	CHEMBL305855	CHEMBL80160
<b>CHEMBL135651</b>	CHEMBL276355	CHEMBL421696	CHEMBL305932	CHEMBL80520
<b>CHEMBL135993</b>	CHEMBL276527	CHEMBL425878	CHEMBL306554	CHEMBL80669
<b>CHEMBL136097</b>	CHEMBL276871	CHEMBL432439	CHEMBL306781	CHEMBL86832
<b>CHEMBL137040</b>	CHEMBL277053	CHEMBL432977	CHEMBL307181	CHEMBL87082
<b>CHEMBL138580</b>	CHEMBL277403	CHEMBL435332	CHEMBL307398	CHEMBL88032
<b>CHEMBL14220</b>	CHEMBL277563	CHEMBL443995	CHEMBL307626	CHEMBL88485
<b>CHEMBL14575</b>	CHEMBL278490	CHEMBL50113	CHEMBL307745	CHEMBL88494
<b>CHEMBL152401</b>	CHEMBL278787	CHEMBL55402	CHEMBL308107	CHEMBL89327
<b>CHEMBL153258</b>	CHEMBL279134	CHEMBL56518	CHEMBL308258	CHEMBL89557
<b>CHEMBL155094</b>	CHEMBL279206	CHEMBL56908	CHEMBL309261	CHEMBL90440
<b>CHEMBL155568</b>	CHEMBL279745	CHEMBL710	CHEMBL309585	CHEMBL2115222
<b>CHEMBL1642918</b>	CHEMBL280015	CHEMBL76536	CHEMBL310022	CHEMBL2172650
<b>CHEMBL2172643</b>	CHEMBL280387	CHEMBL76637	CHEMBL310710	CHEMBL2172647
<b>CHEMBL2172644</b>	CHEMBL2172645	CHEMBL359360	CHEMBL40831	CHEMBL422865

### 3. ChEMBL ID dos inibidores duais da 5 $\alpha$ -Redutase

Na tabela 317 podemos observar ChEMBL IDs dos 181 inibidores duais da 5 $\alpha$ -Redutase com IC50 reportado.

Tabela 217. ChEMBL ID dos 181 inibidores duais 5 $\alpha$ -Redutase com IC50 reportado na ChEMBL.

ChEMBL ID	ChEMBL ID	ChEMBL ID	ChEMBL ID	ChEMBL ID
CHEMBL107565	CHEMBL118283	CHEMBL16863	CHEMBL276527	CHEMBL323996
CHEMBL107654	CHEMBL118390	CHEMBL17025	CHEMBL277563	CHEMBL324210
CHEMBL107719	CHEMBL118416	CHEMBL17193	CHEMBL283564	CHEMBL324687
CHEMBL109170	CHEMBL119722	CHEMBL17194	CHEMBL290823	CHEMBL325572
CHEMBL109560	CHEMBL1201841	CHEMBL17206	CHEMBL340961	CHEMBL326439
CHEMBL118389	CHEMBL120923	CHEMBL17222	CHEMBL414428	CHEMBL327098
CHEMBL124750	CHEMBL122359	CHEMBL17473	CHEMBL56908	CHEMBL332781
CHEMBL126968	CHEMBL1237306	CHEMBL17245	CHEMBL710	CHEMBL334649
CHEMBL128023	CHEMBL1237307	CHEMBL17268	CHEMBL86418	CHEMBL338376
CHEMBL136097	CHEMBL124803	CHEMBL17422	CHEMBL99615	CHEMBL339912
CHEMBL155094	CHEMBL125209	CHEMBL17506	CHEMBL275153	CHEMBL340027
CHEMBL16743	CHEMBL125260	CHEMBL17514	CHEMBL276320	CHEMBL340181
CHEMBL24955	CHEMBL125579	CHEMBL180235	CHEMBL276355	CHEMBL340305
CHEMBL25083	CHEMBL126686	CHEMBL181196	CHEMBL276871	CHEMBL340684
CHEMBL25578	CHEMBL127702	CHEMBL182391	CHEMBL277053	CHEMBL340715
CHEMBL25664	CHEMBL127760	CHEMBL182399	CHEMBL277403	CHEMBL341004
CHEMBL266519	CHEMBL128303	CHEMBL182787	CHEMBL278490	CHEMBL341088
CHEMBL108028	CHEMBL128728	CHEMBL182945	CHEMBL278787	CHEMBL344363
CHEMBL108151	CHEMBL129017	CHEMBL183167	CHEMBL279134	CHEMBL362379
CHEMBL108181	CHEMBL129130	CHEMBL183185	CHEMBL279206	CHEMBL36573
CHEMBL108305	CHEMBL130845	CHEMBL183448	CHEMBL280387	CHEMBL411431
CHEMBL108559	CHEMBL131341	CHEMBL183500	CHEMBL280954	CHEMBL414427
CHEMBL108612	CHEMBL135351	CHEMBL183820	CHEMBL283245	CHEMBL415016
CHEMBL108625	CHEMBL135651	CHEMBL24191	CHEMBL283981	CHEMBL417627
CHEMBL108875	CHEMBL135993	CHEMBL24193	CHEMBL285439	CHEMBL420415
CHEMBL109088	CHEMBL137040	CHEMBL24464	CHEMBL285599	CHEMBL421627
CHEMBL109089	CHEMBL138580	CHEMBL24465	CHEMBL300446	CHEMBL425878
CHEMBL109232	CHEMBL14575	CHEMBL24958	CHEMBL321442	CHEMBL432439
CHEMBL110058	CHEMBL152401	CHEMBL25072	CHEMBL321617	CHEMBL432977
CHEMBL110288	CHEMBL153258	CHEMBL25193	CHEMBL321867	CHEMBL443995
CHEMBL110398	CHEMBL155568	CHEMBL25448	CHEMBL321923	CHEMBL55402
CHEMBL111142	CHEMBL16533	CHEMBL264316	CHEMBL322592	CHEMBL56518
CHEMBL111407	CHEMBL16690	CHEMBL269632	CHEMBL322749	CHEMBL77789
CHEMBL111476	CHEMBL16927	CHEMBL274238	CHEMBL323031	CHEMBL138225
CHEMBL116370	CHEMBL16700	CHEMBL274269	CHEMBL323033	CHEMBL279745
CHEMBL359360	CHEMBL16784	CHEMBL274276	CHEMBL323788	CHEMBL280015

<b>CHEMBL422865</b>				
---------------------	--	--	--	--

## 4. Padrões SMART da *finger*print FP3

Na tabela 218 encontra-se a lista dos 55 padrões SMART que a *finger*print FP3 verifica se existem nas moléculas.

**Tabela 218. 55 padrões SMART da *finger*print FP3.**

Padrões SMART	Padrões SMART	Padrões SMART	Padrões SMART
<b>Catião</b>	Anião	Aldeído ou cetona	Aldeído
<b>Cetona</b>	Tioaldeído ou tiocetona	Tioaldeído	Tiocetona
<b>Imina</b>	Hidrazona	Semicarbazona	Tiosemicarbazona
<b>Oxima</b>	Éter de oxima	Ceteno	Derivado acetil de ceteno
<b>Hidrato de carbono</b>	Hemiacetal	Acetal	Hemiaminal
<b>Aminal</b>	Tiohemiaminal	Tioacetal	Enamina
<b>Enol</b>	Éter enol	Grupo hidroxilo	Alcool
<b>Alcool primário</b>	Alcool secundário	Alcool terciário	1,2-diol
<b>1,2-aminoálcool</b>	Fenol	1,2-difenol	Enediol
<b>Éter</b>	Éter dialquílico	Éter alquilarílico	Éter diarílico
<b>Tioéter</b>	Dissulfido	Peróxido	Hidroperóxido
<b>Grupo aril</b>	Heteroátomo	Aceitadores de hidrogénio	Dadores de hidrogénio
<b>Sistema cíclico (anel)</b>	Ácido carboxílico	Éster	Nitroderivado
<b>Nitrilos</b>	Anilina	Ureia	

## 5. Resultados da tese anterior dos classificadores SVM, para a 5 $\alpha$ -RI

Nas tabelas 219, 220 e 221 encontram-se os resultados da tese anterior relativos aos classificadores SVM para os 310 inibidores da 5 $\alpha$ -RI representados por 40 descritores moleculares ID e classificados de acordo com o critério de classificação (a) e (c) (Sousa et al., 2013). Estes resultados incluem as alterações à classificação dos compostos e uma actualização do conjunto de compostos analisados.

**Tabela 219. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	0 0%	10 31%	0 0%	22 69%	32
Muito Bom	1 2%	11 17%	0 0%	52 81%	64
Bom	0 0%	5 9%	0 0%	49 91%	54
Fraco	0 0%	8 5%	0 0%	152 95%	160

**Tabela 220. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -RI aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	68 45%	82 55%	150
Fraco	56 35%	104 65%	160

**Tabela 221. Métricas de desempenho dos algoritmos de aprendizagem SVM da tese anterior, que utilizaram critérios de classificação (a) e (c) e descritores ID, para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -RI.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	52,581%	0,087	0,000	0,996	0,000	Não aplicável
(c)	55,484%	0,104	0,453	0,650	0,548	0,496

## 6. Resultados da tese anterior dos classificadores SVM, para a 5 $\alpha$ -R2

Nas tabelas 222, 223 e 224 encontram-se os resultados da tese anterior relativos aos classificadores SVM para os 355 inibidores da 5 $\alpha$ -R2 representados por 40 descritores moleculares ID e classificados de acordo com o critério de classificação (a) e (c) (Sousa et al., 2013). Estes resultados incluem as alterações à classificação dos compostos e uma actualização do conjunto de compostos analisados.

**Tabela 222. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (a).**

	Excelente	Muito Bom	Bom	Fraco	Total
Excelente	106	0	0	35	141
	75%	0%	0%	25%	
Muito Bom	5	0	0	14	19
	26%	0%	0%	74%	
Bom	9	0	1	32	42
	21%	0%	2%	76%	
Fraco	18	0	1	134	153
	12%	0%	1%	88%	

**Tabela 223. Reclassificação dos compostos em análise na tese anterior, representados por descritores moleculares ID, que inibem a 5 $\alpha$ -R2 aplicando modelos de classificação baseados em SVM, utilizando validação cruzada 5 vezes e critérios de classificação binária (c).**

	Forte	Fraco	Total
Forte	168	34	202
	83%	17%	
Fraco	83	7	90
	92%	8%	

**Tabela 224. Métricas de desempenho dos algoritmos de aprendizagem SVM da tese anterior, que utilizaram critérios de classificação (a) e (c) e descritores ID, para classificar e descrever, respectivamente, os inibidores da 5 $\alpha$ -R2.**

Metodologias de classificação	Accuracy	Cohen's kappa	Sensibilidade	Especificidade	Precisão	F-Score
(a)	67,887%	0,450	0,752	0,850	0,768	0,760
(c)	67,042%	0,301	0,831	0,457	0,669	0,742

