Universidade de São Paulo
Instituto de Física de São Carlos

LÍVIA DE BARROS SALUM

**Estudos *in silico* no planejamento de candidatos a novos fármacos
na terapia do câncer de mama e de reposição hormonal**

São Carlos
2007

LÍVIA DE BARROS SALUM

**Estudos *in silico* no planejamento de candidatos a novos fármacos
na terapia do câncer de mama e de reposição hormonal**

Dissertação apresentada ao Instituto de Física de
São Carlos, da Universidade de São Paulo, para
a obtenção do título de Mestre em Ciências.

Área de Concentração: Física Biomolecular.
Orientador: Prof. Dr. Adriano Defini Andricopulo

São Carlos
2007

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Salum, Livia de Barros

Estudos *in silico* no planejamento de candidatos a novos fármacos na terapia do câncer de mama e de reposição hormonal /Livia de Barros Salum; orientador Adriano Defini Andricopulo.-- São Carlos, 2007.

140 f.

Dissertação (Mestrado em Ciências - Área de concentração: Física Biomolecular) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Planejamento de fármacos. 2. Receptor de estrógeno. 3. CoMFA. 4. HQSAR. I. Título.

AGRADECIMENTOS

A Deus, à minha família, e aos meus amigos, pela força e compreensão.

Ao meu orientador, Prof. Adriano D. Andricopulo, pela oportunidade, incentivo, dedicação, compromisso e confiança.

Ao Prof. Igor Polikarpov pelo apoio.

Ao Prof. Carlos Montanari e integrantes do NEQUIM, por me permitirem os primeiros passos nos estudos em Química Medicinal.

Aos amigos que me ajudaram com importantes discussões em Química Medicinal e Computacional, em particular, Profa Káthia Honório, Rafael Guido, Ricardo Nicoluci, Carolina Horta, Matheus Postigo, Napoleão Valadares e Tiago Moda.

Aos demais amigos do Centro de Biotecnologia Molecular Estrutural, sobretudo Nádia Martins, Alessandro Silva, Deise Borchhardt, Karime Stevanatto, Ana Carolina Figueira, Sandra Krauchenco e Diógenes Neto.

Ao Instituto de Física de São Carlos da Universidade de São Paulo, aos professores e funcionários pela colaboração e excelente suporte para realização desse trabalho.

Aos pesquisadores José Martins, Patrícia Baraldi e Gustavo Robello.

Ao apoio do Instituto Internacional de Pesquisas Farmacêuticas e ao CNPq pelo suporte financeiro.

*“O correr da vida embrulha tudo,
A vida é assim: esquentando e esfria,
aperta e daí afrouxa,
sossega e depois desinquieta.
O que ela quer da gente é coragem”.*

João Guimarães Rosa

RESUMO

Os estrógenos exercem importantes efeitos fisiológicos através dos dois subtipos dos receptores de estrógeno humanos (hERs), alfa (hER α) e beta (hER β). Enquanto hER α é um importante alvo macromolecular no desenvolvimento de fármacos para o tratamento do câncer de mama, hER β é um alvo promissor no desenvolvimento de agentes terapêuticos para a terapia de reposição hormonal. O progresso no planejamento de moduladores apresentando maior potência, afinidade e seletividade, entretanto, requer a otimização múltipla de interações intermoleculares fármaco-receptor. A Química Medicinal moderna, de forte caráter multidisciplinar, fornece um arsenal de alternativas e estratégias úteis no processo de planejamento de novos fármacos. As ferramentas de modelagem molecular e de estudos das relações quantitativas entre a estrutura e atividade (QSAR) estão integradas a esse processo, sendo de extremo valor na busca por moléculas bioativas com propriedades múltiplas otimizadas. Para a realização deste trabalho, conjuntos padrões de dados foram organizados para diferentes classes químicas de potentes moduladores dos ERs. Esses conjuntos padronizados para os subtipos do hER, contendo a informação qualificada sobre a estrutura química dos ligantes associada a medida da propriedade farmacológica correspondente, estabeleceram as bases para o desenvolvimento de modelos empregando os métodos holograma QSAR, CoMFA e GRID/PCA. Os modelos finais de HQSAR e CoMFA possuem elevada consistência interna e externa, apresentando bom poder de correlação e predição das propriedades alvo. Juntamente com as informações obtidas pelos mapas de contribuição 2D e de contorno 3D, os modelos de QSAR e GRID/PCA construídos são guias químico-medicinais úteis no planejamento de novos moduladores seletivos do ER possuindo maior afinidade e potência.

Palavras-chave: Planejamento de fármacos, Receptor de estrógeno, CoMFA, HQSAR

ABSTRACT

Estrogens exert important physiological effects through two human estrogen receptor subtypes (hERs), alpha (hER α) and beta (hER β). While hER α is a macromolecular target of great importance for breast cancer therapy, hER β is an attractive drug target for the development of novel therapeutic agents for hormone replacement therapy. Progress towards the design of modulators having improved potency, affinity and selectivity requires the optimization of multiple ligand-receptor interactions. The strong multidisciplinary character of modern Medicinal Chemistry supplies a rich arsenal of useful rational strategies for the design of new drug candidates. Molecular modeling tools and quantitative structure-activity relationships (QSAR) are integrated into the drug design process in the search of bioactive molecules having optimized properties. In this study, standard data sets were organized for different chemical classes of ER modulators, integrating the qualified information about chemical structure associated to the corresponding pharmacological property. The data sets established the scientific basis for the development of models employing the hologram QSAR, CoMFA and GRID/PCA methods. The final HQSAR and CoMFA models possess high internal and external consistency, with good correlative and predictive power. The generated QSAR and GRID/PCA models as well as the information gathered from the 3D contour maps provide useful guidelines for the design of new selective ER modulators having improved affinity and potency.

Keywords: Drug design, Estrogen receptor, CoMFA, HQSAR.

LISTA DE FIGURAS

Figura 1	Etapas do processo de descoberta e desenvolvimento de novos fármacos.	20
Figura 2	Etapas de identificação, seleção e otimização de substâncias bioativas durante o processo de descoberta de novos fármacos (fase pré-clínica).	21
Figura 3	Tipos de cânceres mais incidentes, estimados para 2006, exceto pele não melanoma, na população brasileira, nos gêneros masculino e feminino.	23
Figura 4	Moduladores do receptor de estrógeno comercialmente disponíveis.	26
Figura 5	Estruturas dos fitoestrógenos mais importantes utilizados na terapia de reposição hormonal.	28
Figura 6	Localização dos tecidos-alvo dos estrógenos no corpo da mulher.	29
Figura 7	Estrutura química do estradiol, principal estrógeno feminino.	29
Figura 8	Estruturas de domínios dos subtipos do receptor de estrógeno.	30
Figura 9	Representação dos LBDs de hER α (azul) e hER β (magenta) em complexo com o estradiol (verde).	31
Figura 10	Ativação do receptor de estrógeno mediada pelo ligante endógeno estradiol.	32
Figura 11	Ligação do estradiol no bolsão hidrofóbico de hER α .	33
Figura 12	Mudanças conformacionais no LBD do ER provocadas pela interação com o estradiol. A. Na ausência do estradiol. B. Reposicionamento da hélice 12 na presença do estradiol.	34
Figura 13	A. Dímero do LBD de hER α com estradiol (azul). Note conformação fechada da hélice 12 (magenta). B. Esta conformação é muito diferente da conformação aberta da mesma hélice em complexo com o antagonista raloxifeno (vermelho).	35
Figura 14	Modelo hipotético para a ligação estrogênica e antiestrogênica ao ER. O estradiol é ancorado no sítio fenólico (SF) com alta afinidade. O monohidroxitamoxifeno tem alta afinidade pelo sítio de ligação, mas possui uma cadeia lateral que interage com uma região antiestrogênica.	36
Figura 15	Complementaridade multiestrutural na interação do agonista dietilbestrol e antagonistas tamoxifeno e raloxifeno com hER α , e do agonista genisteína com hER β .	37
Figura 16	Esquema ilustrativo do processo de geração dos hologramas moleculares.	42
Figura 17	Etapas envolvidas na geração de um modelo de HQSAR.	44
Figura 18	Etapas envolvidas na geração de um modelo de CoMFA.	46

Figura 19	Etapas envolvidas na geração de um modelo GRID/PCA.	49
Figura 20	Estrutura representativa da diversidade química dos conjuntos de dados.	62
Figura 21	Representação gráfica da distribuição dos valores de pIC_{50} para o conjunto de dados 1.	77
Figura 22	Representação gráfica da distribuição dos valores de pIC_{50} para o conjunto de dados 2.	77
Figura 23	Gráfico de HQSAR dos valores experimentais e preditos para as 127 moléculas do conjunto de dados (modelo para afinidade).	84
Figura 24	Gráfico de HQSAR dos valores experimentais e preditos para as 69 moléculas do conjunto de dados (modelo para potência).	84
Figura 25	Mapas de contribuição de HQSAR. A. Modulador com maior potência (43). B. modulador com maior afinidade (67).	86
Figura 26	Complementaridade multifuncional entre o modulador 33 e hER α (código PDB: 1XP1).	87
Figura 27	Alinhamento molecular estrutural do conjunto de dados.	89
Figura 28	Gráfico CoMFA de afinidade de ligação dos valores experimentais e preditos para as 127 moléculas do conjunto de dados.	94
Figura 29	Gráfico CoMFA de potência dos valores experimentais e preditos para as 69 moléculas do conjunto de dados.	94
Figura 30	Mapas de contorno 3D CoMFA eletrostáticos e estereoquímicos (contorno 80:20 padrão). A. modulador de maior afinidade (67). B. modulador mais potente (43).	96
Figura 31	Representação esquemática do modo de ligação do modulador apresentando mais alta afinidade no interior da estrutura cristalográfica de hER α .	97
Figura 32	Mapa de contorno estereoquímico (modulador 67 na cavidade de interação do hER α).	98
Figura 33	Impedimento estérico ao posicionamento de grupos volumosos ligados ao átomo de enxofre do anel benzoxatiina (modulador 86 na cavidade de interação do hER α).	98
Figura 34	Mapa de contorno eletrostático (modulador 67 na cavidade de interação do hER α).	99
Figura 35	Diferenças nas cavidades de interação com ligantes de hER α (verde) e hER β (magenta), em complexos com a genisteína.	102
Figura 36	Representação gráfica da distribuição dos valores de pIC_{50} para hER β .	110
Figura 37	Representação gráfica da distribuição dos valores de pIC_{50} para hER α .	110

Figura 38	Gráfico HQSAR dos valores experimentais e preditos para as 81 moléculas do conjunto de dados (modelo para hER β).	113
Figura 39	Gráfico HQSAR dos valores experimentais e preditos para as 81 moléculas do conjunto de dados (modelo para hER α).	114
Figura 40	Mapa de contribuição de HQSAR para o modulador 1 A) modelo para hER β e B) modelo para hER α .	115
Figura 41	Diferentes tipos de alinhamento estrutural do conjunto de dados (hER α à esquerda e hER β à direita).	116
Figura 42	Gráfico CoMFA do modelo de hER β dos valores experimentais e preditos para as 81 moléculas do conjunto de dados.	118
Figura 43	Gráfico CoMFA do modelo de hER α dos valores experimentais e preditos para as 81 moléculas do conjunto de dados.	118
Figura 44	Mapas de contorno eletrostático (A e C) e estereoquímico (B e D) gerados nos modelos para hER α (A e B) e hER β (C e D).	119
Figura 45	Impedimento estérico ao posicionamento de grupos volumosos em hER α (A). O mesmo não ocorre para hER β .	121
Figura 46	Cadeias laterais de resíduos substituídos em hER α (azul) e hER β (rosa) rodeando as posições 1 e 8 de substituição no anel naftaleno.	121
Figura 47	Posicionamento do modulador 2 na cavidade de interação de hER β .	122
Figura 48	Esquema ilustrativo da definição das dimensões da grade reticulada tridimensional utilizada para os cálculos dos campos moleculares de interação.	124
Figura 49	Resultado do modelo de PCA para os subtipos do hER. As estruturas cristalográficas do hER β , 1QKM, 1X7B e 1X78 (rosa) foram posicionadas em valores positivos de PC1. Estruturas cristalográficas do hER α , 1X7R e 1X7E, foram posicionadas em valores negativos de PC1 (azul). Estruturas em complexo com a genisteína, 1QKM e 1X7R, foram posicionadas em valores positivos de PC2, enquanto que as estruturas em complexo com WAY-244, 1X7E e 1X78, foram posicionadas em valores negativos de PC2.	125
Figura 50	Mapas de contorno obtidos dos pesos da PC1 para as sondas A. DRY e B. OH2. No centro da cavidade está mostrado o ligante ERB041, que foi retirado do sítio durante a análise.	127
Figura 51	Mapas de contorno hidrofóbicos com altos pesos para PC1. Regiões importantes para a seletividade-hER β são mostradas em rosa, e para hER α em azul. No centro da cavidade está mostrado o modulador 2.	128

LISTA DE TABELAS

Tabela 1	Parâmetros de distinção de fragmentos no método HQSAR.	43
Tabela 2	Principais sondas químicas utilizadas no método GRID.	48
Tabela 3	Estruturas químicas e valores associados de afinidade ligação e potência (pIC_{50}) do conjunto completo de moduladores de hER α	63
Tabela 4	Análises de HQSAR para o modelo de afinidade de ligação para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7). Modelo HQSAR	79
Tabela 5	Análises de HQSAR para o modelo de potência para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7).	79
Tabela 6	Análise de HQSAR para o modelo de afinidade de ligação em função da variação do tamanho dos fragmentos (distinção de fragmentos: A/B/C/DA).	80
Tabela 7	Análise de HQSAR para o modelo de potência em função da variação do tamanho dos fragmentos (distinção de fragmentos: A/B/DA).	80
Tabela 8	Valores preditos e experimentais de afinidade de ligação (pIC_{50}), associados aos resíduos, para os 28 compostos do conjunto teste.	82
Tabela 9	Valores preditos e experimentais de potência (pIC_{50}), associados aos resíduos, para os 14 compostos do conjunto teste.	83
Tabela 10	Resultados dos modelos iniciais CoMFA construídos.	90
Tabela 11	Resultados dos modelos finais CoMFA construídos.	91
Tabela 12	Valores preditos e experimentais de afinidade de ligação (pIC_{50}), associados aos resíduos, para os 28 compostos do conjunto teste.	92
Tabela 13	Valores preditos e experimentais de potência (pIC_{50}), associados aos resíduos, para os 14 compostos do conjunto teste.	93
Tabela 14	Estruturas químicas e valores correspondentes de afinidade de ligação para hER α e hER β .	103
Tabela 15	Modelo de HQSAR de hER β para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7).	111
Tabela 16	Modelo de HQSAR de hER α para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7).	111
Tabela 17	Valores preditos e experimentais de afinidade de ligação (βpIC_{50}), associados aos resíduos, para os 16 compostos do conjunto teste.	112

Tabela 18	Valores preditos e experimentais de afinidade de ligação (α plC ₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.	112
Tabela 19	Resultados dos modelos finais CoMFA construídos.	117
Tabela 20	Valores preditos e experimentais de afinidade de ligação (α plC ₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.	117
Tabela 21	Valores preditos e experimentais de afinidade de ligação (β plC ₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.	117
Tabela 22	Características dos receptores de estrógeno humano utilizados nos estudos.	123

LISTA DE ABREVIATURAS E SIGLAS

2D	Bidimensional
3D	Tridimensional
ADME	Administração, Distribuição, Metabolismo e Excreção
AF	Função de Ativação
CBME	Centro de Biotecnologia Molecular Estrutural
CoMFA	Análise Comparativa dos Campos Moleculares
CRC	Algoritmo Cíclico de Verificação de Redundância
DBD	Domínio de Ligação ao DNA
DNA	Ácido Desoxirribonucléico
ER	Receptor de Estrógeno
FDA	Administração Federal de Alimentos e Medicamentos dos Estados Unidos
HCA	Análise Hierárquica de Agrupamento
hER	Receptor de Estrógeno Humano
HQSAR	Holograma QSAR
HRT	Terapia de Reposição Hormonal
HTS	Ensaio em Larga Escala
IND	Investigação de Novo Fármaco
LBD	Domínio de Ligação ao Ligante
LBDD	Planejamento de Fármacos Baseado no Ligante
LQMC	Laboratório de Química Medicinal e Computacional
MIF	Campo de Interação Molecular
MSS	Planilha Molecular
NCE	Nova Entidade Química
NIH	Instituto Nacional de Saúde dos Estados Unidos
NDA	Requisição de um Novo Fármaco
OMS	Organização Mundial da Saúde
P&D	Pesquisa e Desenvolvimento
PCA	Análise de Componentes Principais
PDB	Banco de Dados de Proteínas
PLS	Regressão Multivariada por Mínimos Quadrados Parciais
QSAR	Relações Quantitativas Estrutura-Atividade
SAR	Relações Estrutura-Atividade
SBDD	Planejamento de Fármacos Baseado na Estrutura do Receptor
SERM	Modulador Seletivo do Receptor de Estrógeno
VS	Ensaio Virtual

SUMÁRIO

CAPÍTULO 1	17
INTRODUÇÃO	
1 Fármacos e Indústria Farmacêutica	18
1.1 O Novo Paradigma	18
2 Química Medicinal e a Gênese Planejada de Fármacos	19
2.1 Estratégias para o Planejamento de Candidatos a Fármacos	23
3 Câncer de Mama	24
3.1 Tratamento do Câncer de Mama	24
4 Menopausa	27
5 Os Receptores de Estrógeno	28
5.1 Domínios Estruturais dos Subtipos do Receptor de Estrógeno	30
5.2 Mecanismo de Ação	31
5.3 Hélice 12	33
5.4 Interações Intermoleculares	35
6 Os Receptores de Estrógeno como Alvos Macromoleculares Promissores	38
7 Química Medicinal e as Técnicas de QSAR	40
8 Holograma QSAR	41
9 Análise Comparativa dos Campos Moleculares	45
10 O Método GRID/PCA	47
CAPÍTULO 2	50
OBJETIVOS	

CAPÍTULO 3	52
PROCEDIMENTO EXPERIMENTAL	
1 Modelagens de QSAR e GRID/PCA	53
2 Conjuntos de Dados	53
3 Identificação das Estruturas Tridimensionais de ERs	55
4 Alinhamento Estrutural	55
5 Modelos de HQSAR	56
6 Modelos CoMFA	57
7 Método GRID/PCA	58
 CAPÍTULO 4	 60
RESULTADOS E DISCUSSÃO	
 PARTE A - Bases Químicas e Estruturais para a Otimização da Afinidade e Potência de Moduladores do Receptor de Estrógeno	 61
1 Conjuntos Padrões de Dados	61
2 Holograma QSAR	78
3 QSAR 3D CoMFA	86
 PARTE B - Determinantes Químicos e Estruturais da Afinidade e Seletividade de Moduladores dos Subtipos do ER	 101
1 Conjuntos Padrões de Dados	102
2 Holograma QSAR	111
3 QSAR 3D CoMFA	115
4 GRID/PCA	123

CAPÍTULO 5	129
CONCLUSÕES E PERSPECTIVAS	
1 Conclusões	130
2 Perspectivas	131
CAPÍTULO 6	132
REFERÊNCIAS	
APÊNDICES	139

Capítulo 1

---

*I*NTRODUÇÃO

1 FÁRMACOS E INDÚSTRIA FARMACÊUTICA

O mercado farmacêutico é um setor extremamente competitivo e caracterizado por elevados níveis de investimentos associados a riscos de proporcional magnitude. A sobrevivência, desenvolvimento e crescimento contínuo da indústria farmacêutica requerem investimentos significativos em P&D (Pesquisa e Desenvolvimento). A ideia de que a maior companhia farmacêutica é a melhor em P&D tem se tornado um dogma para a indústria farmacêutica mundial, já que a descoberta de fármacos da categoria *blockbuster* (termo inglês para designar os fármacos com vendas anuais superiores a US\$ 1 bilhão) é a principal responsável pelo crescimento da indústria em nível mundial. Neste sentido, a capacidade da indústria em alocar recursos em P&D está diretamente ligada à sua capacidade em gerar lucros através da venda de um arsenal atrativo de fármacos.¹⁻³

O complexo processo de descoberta e desenvolvimento de novas entidades químicas (NCEs, do inglês, *New Chemical Entities*) é longo e envolve altos investimentos.⁴ Desde a concepção do projeto até a introdução de um único fármaco no mercado farmacêutico, são investidos de 12 a 15 anos em P&D, com custos totais da ordem de US\$ 500-880 milhões, podendo alcançar cifras superiores a US\$ 1 bilhão em alguns casos.⁵⁻⁸

Uma inovação terapêutica de alto sucesso, entretanto, pode gerar rendimentos significativos, da ordem de milhões a bilhões de dólares anuais. Dentre os fármacos incluídos na categoria *blockbuster* atualmente estão o hipolipemiante atorvastatina (Lípitor®, Pfizer), o antitrombótico clopidogrel (Plavix®, Sanofi Aventis), o antiulceroso esomeprazol (Nexium®, AstraZeneca) e o antipsicótico olanzapina (Zyprexa®, Eli Lilly), entre outros.

1.1 O Novo Paradigma

As últimas décadas foram marcadas por um aumento contínuo da expectativa de vida de diversas populações. Uma das razões fundamentais para tal foi o desenvolvimento de inovações terapêuticas associadas ao aumento da qualidade de vida e ao tratamento de doenças crônicas.⁹ No mesmo período, entretanto, o número de NCEs para uso farmacêutico,

submetidas e aprovadas pelas agências regulatórias, tem decrescido progressivamente, a despeito do aumento dos investimentos anuais da indústria farmacêutica em P&D, da disponibilidade de novas tecnologias, do seqüenciamento genômico e da revolução digital. Além disso, a competitividade do mercado farmacêutico vem se tornando cada vez mais acirrada, mesmo em face à enorme demanda existente.^{7,8,10-14}

Esse panorama, caracterizado pelo declínio da produtividade e aumento da competitividade, aponta para mudanças significativas no modelo tradicional de pesquisa, baseado principalmente na química e na biologia básica, e numa abordagem “triagem-erro”.^{1,15} De fato, a indústria farmacêutica vem passando por um momento crucial, marcado por grandes transformações guiadas pelo uso racional de inovações científicas e tecnológicas como ferramentas de pesquisa. As abordagens modernas estão fundamentadas em uma combinação de conhecimentos atuando em caráter multi- e interdisciplinar, onde a computação, a biotecnologia, a biologia e a química têm papel de destaque. Essa nova geração de modelos e conceitos reflete o papel central da química medicinal moderna nos processos inovativos da gênese planejada de fármacos.^{5,15-19} Nesse novo paradigma, o desenvolvimento de inovações terapêuticas requer um forte comprometimento com a inovação científica e tecnológica através de programas bem estabelecidos de P&D. O sucesso destas iniciativas, deverá resultar na introdução de fármacos pioneiros no mercado, ou ao menos em novas gerações de NCEs capazes de competir no mercado com fármacos já comercializados em suas respectivas classes terapêuticas.⁵

2 QUÍMICA MEDICINAL E A GÊNESE PLANEJADA DE FÁRMACOS

O processo de descoberta e desenvolvimento de novos fármacos é dividido em duas grandes fases (Figura 1): pré-clínica (ou descoberta) e (ii) clínica (ou desenvolvimento).^{5,20}

Figura 1. Etapas do processo de descoberta e desenvolvimento de novos fármacos.⁵

Nos estágios iniciais da fase pré-clínica, as pesquisas se concentram no estudo e caracterização de alvos macromoleculares com elevado potencial terapêutico e na busca por moléculas pequenas capazes de modular o alvo macromolecular selecionado (Figura 2). Uma vez identificado um alvo molecular promissor, é fundamental a sua validação, ou seja, é preciso verificar qual é a relevância deste alvo no processo fisiopatológico em questão.²¹⁻²³ É necessário investigar as vias bioquímicas envolvidas e estabelecer se a modulação seletiva da proteína-alvo, por moléculas pequenas, é capaz de gerar a resposta esperada no controle da doença ou desordem em humanos. Tais moléculas bioativas podem ser obtidas de fontes diversas, como produtos naturais, síntese química e coleções combinatórias, e devem ter sua propriedade biológica determinada experimentalmente. Neste estágio da pesquisa, buscam-se identificar moléculas que apresentem propriedades desejadas para um fármaco, predominantemente potência e seletividade, mas que ainda precisam ser otimizadas em relação a outras propriedades. Os compostos bioativos que se destacam nesta fase são selecionados como compostos-líderes (do inglês, *lead compounds*).²¹⁻²⁴

Figura 2. Etapas de identificação, seleção e otimização de substâncias bioativas durante o processo de descoberta de novos fármacos (fase pré-clínica).²⁴

No processo de identificação e seleção de compostos capazes de interagir com o alvo selecionado, é possível explorar o imenso espaço químico com o auxílio de técnicas em química medicinal, de forma a correlacionar a correta diversidade química com o espaço biológico em estudo. Entre essas técnicas citam-se a organização de bases padrões de dados, aplicação de filtros moleculares, e o emprego de métodos modernos como os ensaios biológicos automatizados em larga escala (HTS, do inglês, *High-Throughput Screening*) e o ensaio virtual (VS, do inglês, *Virtual Screening*), além do estudo das relações quantitativas entre a estrutura e atividade (QSAR, do inglês, *Quantitative Structure-Activity Relationships*). Em todos estes casos, o gerenciamento qualificado da informação possibilita a geração de conhecimento apropriado com base nos dados disponíveis.^{23,25}

A fase seguinte tem como objetivo a otimização dos compostos-líderes em relação às suas propriedades farmacodinâmicas e farmacocinéticas (Figura 2). Variações estruturais são utilizadas para adequar o perfil das moléculas às características necessárias à aplicação terapêutica. Neste sentido, a química medicinal moderna oferece várias estratégias capazes de gerar coleções dirigidas de compostos.²² Os estudos de Relação Estrutura-Atividade (SAR, do inglês, *Structure-Activity Relationships*) e QSAR são importantes para guiar a síntese de novas moléculas com propriedades otimizadas, minimizando o universo de compostos a serem considerados nos programas de triagens. Idealmente, parâmetros como potência, seletividade e afinidade, além de propriedades farmacocinéticas como ADME (Absorção, Distribuição, Metabolismo e Excreção), e certas características físico-químicas das substâncias, devem ser considerados conjuntamente, para eliminar candidatos com propriedades inadequadas nos

estágios iniciais, reduzindo os custos do processo de P&D.^{26,27} Em suma, na fase pré-clínica de pesquisa básica, são realizados testes para avaliar o potencial terapêutico de um composto e sua segurança em relação aos testes em humanos que serão realizados na fase clínica. Compostos que se destacam nessa fase, por apresentarem propriedades múltiplas otimizadas, são selecionados como NCEs.

Os métodos em quimio- e bionformática estão integrados ao planejamento de fármacos. O seu emprego nos estudos em química medicinal vão desde a identificação, seleção e otimização de moléculas candidatas até a proposição de novas NCEs com elevado potencial terapêutico. A integração desses métodos ao trabalho químico e biológico experimental é um requerimento essencial para a geração de novas moléculas bioativas qualificadas, com propriedades múltiplas otimizadas.^{17,25,27}

Após a conclusão da fase pré-clínica, as NCEs selecionadas são submetidas à aprovação das agências regulatórias, como a Administração Federal de Alimentos e Medicamentos dos Estados Unidos (FDA, do inglês, *Food and Drug Administration*) para autorização da realização de testes em seres humanos através da licença para investigação de um novo candidato a fármaco (IND, do inglês, *Investigational New Drug*). Nas fases clínicas são realizados diversos estudos para determinar a eficácia e a segurança das NCEs para uso clínico (Figura 1). Na fase clínica I são feitos testes de dosagem dos princípios ativos em um grupo de voluntários saudáveis e a avaliação preliminar dos níveis de toxidez. Na fase clínica II são realizados testes farmacológicos em um número reduzido de pacientes para a avaliação da dosagem, eficácia e segurança da composição. Na fase clínica III são realizados testes em um número significativo de pacientes para avaliação das propriedades farmacocinéticas, além de eficácia e segurança. Na fase clínica IV, é realizado o monitoramento pós-comercialização do novo fármaco.^{6,20}

2.1 Estratégias para o Planejamento de Candidatos a Fármacos

A estratégia de planejamento da arquitetura molecular de substâncias bioativas, candidatas a novos fármacos, fundamenta-se no conhecimento prévio do processo fisiopatológico envolvido e na seleção do alvo terapêutico adequado, como discutido anteriormente. Este alvo macromolecular (e.g., enzima, receptor biológico) pode ter a sua estrutura tridimensional conhecida ou não. O conhecimento da topografia molecular tridimensional do receptor, particularmente da cavidade de interação, permite o planejamento de inibidores enzimáticos ou agonistas/antagonistas de receptores, por processos de complementaridade molecular planejada. Neste processo, se planeja um candidato com propriedades estruturais adequadas ao processo de reconhecimento molecular pelo biorreceptor, reflexo de suas propriedades farmacodinâmicas. É o que se chama de planejamento baseado na estrutura do receptor (SBDD, do inglês, *Structure-Based Drug Design*). Graças a grande evolução das técnicas em biologia estrutural, como a cristalografia de raios-X, um considerável número de alvos moleculares tem suas estruturas tridimensionais disponíveis no banco de dados de proteínas (PDB, do inglês, *Protein Data Bank*)

Em contraste, a estrutura do alvo macromolecular eleito como alvo terapêutico pode não ser conhecida. Neste caso, o processo de planejamento se inicia pelo estudo das características multifuncionais dos ligantes conhecidos ou da micromolécula endógena envolvida na fisiopatologia do processo em estudo. O estudo pode, então, ser conduzido através do planejamento baseado na estrutura do ligante (LBDD, do inglês, *Ligand-Based Drug Design*). Em muitos casos, o uso combinado das duas estratégias pode ser útil no processo de planejamento, gerando conhecimentos novos através de uma sinergia entre as técnicas.^{22,28} Combinações dessas duas estratégias serão descritas na presente dissertação de mestrado.

3 CÂNCER DE MAMA

Câncer é a denominação para um grupo de doenças caracterizadas pelo crescimento descontrolado e migração de células anormais. Pode ser causado por fatores internos (mutações inerentes, hormônios, condições imunológicas) e externos (tabaco, substâncias químicas, radiação, agentes infecciosos), sendo que ambos os fatores causais podem atuar simultaneamente ou seqüencialmente para a iniciação ou promoção da carcinogênese.^{29,30}

De acordo com o NIH (*National Institute of Health*, dos Estados Unidos) aproximadamente 2,5 milhões de novos casos de câncer nos Estados Unidos serão diagnosticados em 2007.³⁰ No Brasil, as estimativas para o ano de 2006 apontaram mais de 470.000 novos casos, com maior incidência para os cânceres de mama e colo de útero no gênero feminino. Apenas para o câncer de mama, esperavam-se 48.000 novos casos, ou seja, 52 casos a cada 100.000 mulheres (Figura 3).³¹

Figura 3. Tipos de cânceres mais incidentes, estimados para 2006, exceto pele não melanoma, na população brasileira, nos gêneros masculino e feminino.³¹

Em todo o mundo, seguindo as mesmas tendências verificadas para o Brasil, o câncer de mama permanece o tipo de câncer mais freqüentemente diagnosticado em mulheres, e o

segundo em mortalidade, depois do câncer de pulmão. Apenas nos Estados Unidos, onde os dados estatísticos são disponíveis, são esperados mais de 180.000 novos casos de câncer de mama invasivo (178.000 em mulheres) em 2007, com cerca de 41.000 mortes provocadas por esse tipo de câncer (mais de 40.000 em mulheres). Sabe-se, entretanto, que, desde 1990, a taxa de mortalidade pelo câncer de mama na população feminina tem diminuído devido à detecção precoce e tratamento. A frequência de novos casos, entretanto, tem se mantido alta e constante.³⁰

3.1 Tratamento do Câncer de Mama

Considerando o tamanho e estágio do tumor, bem como a preferência do paciente, o tratamento pode envolver a remoção cirúrgica do tumor e suas margens (ou mastectomia), radioterapia, quimioterapia, terapia hormonal (antiestrógenos ou inibidores da aromatase), terapia biológica orientada (anticorpos monoclonais) ou combinações desses métodos. A maioria das opções é seguida por um tratamento concomitante ou de manutenção utilizando a terapia hormonal, com antiestrógenos.³⁰

A relação entre os hormônios ovarianos e o câncer de mama foi desvendada ao longo do século passado. Aproximadamente um terço das pacientes pré-menopausa com câncer de mama, submetidas à remoção dos ovários, no início do século, apresentavam regressão no tumor. Na década de 50, quando se comprovou a presença do receptor de estrógeno (ER) no útero, vagina e mamas, inferiu-se que esse receptor estava presente nos tumores das mamas, aumentando a probabilidade de resposta à terapia endócrina ablativa. Essa teoria se concretizou e levou ao desenvolvimento dos antiestrógenos, como uma alternativa simples à remoção cirúrgica dos ovários para o tratamento do câncer de mama.³²

Tamoxifeno (Nolvadex®, AstraZeneca), Toremifeno (Fareston®, Schering-Plough) e Fulvestranto (Faslodex®, AstraZeneca) são fármacos usados no tratamento do câncer de mama (Figura 4). Os dois primeiros são antiestrógenos não-esteróides de uso oral e o último é um antiestrógeno esteróide disponível na forma injetável. Enquanto o tamoxifeno (pró-fármaco do hidroxitamoxifeno) é aprovado para tratamento do câncer de mama em todos os estágios,

toremifeno (pró-fármaco do hidroxitoremifeno) e fulvestranto são geralmente utilizados em estágios mais avançados da doença (doença metastática), em mulheres pós-menopausa. O raloxifeno (Evista®, Eli Lilly) é um modulador não-esteróide também disponível comercialmente, entretanto é usado principalmente no tratamento da osteoporose.³²⁻³⁶

Figura 4. Moduladores do receptor de estrógeno comercialmente disponíveis.

O uso desses fármacos está, entretanto, associado a sérias reações indesejadas, incluindo aumento da frequência de eventos tromboembólicos, sangramento vaginal e alterações na parede do útero, com aumento da incidência de hiperplasia e tumor endometrial, além de resistência após uso prolongado, o que limita o tratamento do câncer de mama.³²⁻³⁶ Por essa razão, é necessário que se desenvolvam novos agentes terapêuticos para o emprego eficaz e seguro no tratamento do câncer de mama.

4 MENOPAUSA

As concentrações sanguíneas dos estrógenos sofrem flutuações durante a vida das mulheres. Até a puberdade, as concentrações do hormônio se mantêm estáveis e baixas, mas, após a puberdade, as concentrações sanguíneas aumentam até um máximo e a mulher atinge o ápice de sua capacidade reprodutiva. A partir dos 35 anos, em média, as concentrações de estradiol começam a diminuir. O processo de redução das concentrações hormonais não é linear e sofre certas variações. Aproximadamente aos 50 anos surgem os sintomas da pré-menopausa e cerca de um ano antes da última menstruação as concentrações hormonais decrescem bruscamente, o que caracteriza a menopausa.³⁷

A deficiência dos estrógenos durante a menopausa está inequivocamente associada a diversos sintomas vasomotores e genitais, incluindo ondas de calor, alterações do sono e humor, redução da densidade óssea, atrofia vaginal, inapetência sexual e incontinência urinária. Tais sintomas são experimentados por cerca de 75% das mulheres em algum momento da menopausa, comprometendo a sua qualidade de vida. Esses sintomas podem, entretanto, ser reduzidos pela reposição de hormônios durante e após a menopausa.³⁷⁻⁴⁰

Os hormônios esteróides são comumente prescritos para a terapia de reposição hormonal (HRT, do inglês, *hormone replacement therapy*) e estudos clínicos aleatórios placebo-controlados atestam que estão associados à redução dos sintomas vasomotores e genitais característicos da menopausa e à diminuição da morbidade e mortalidade cardiovascular e da osteoporose.³⁹⁻⁴³ Entretanto, os benefícios da reposição hormonal com esteróides são limitados pelos complexos efeitos colaterais originários principalmente de distúrbios da função endometrial, incluindo sangramentos, proliferação do endométrio e mesmo aumento da incidência dos cânceres uterino e de mama durante a terapia prolongada.^{40,44-47} Em razão desses efeitos indesejados, o interesse no planejamento de novos agentes terapêuticos capazes de reter os efeitos benéficos dos estrógenos, mas não apresentar as reações indesejadas nos tecidos reprodutivos e nas mamas tem tido um considerável interesse nos últimos anos.⁴⁸⁻⁵⁰

Estudos epidemiológicos sobre HRT e o consumo de isoflavonas, dentre as quais as principais são a genisteína e a daidzeína, indicam que esses produtos naturais estão relacionados a efeitos fisiológicos positivos no tratamento das desordens da menopausa, sem, entretanto, aumentar os riscos de eventos cardiovasculares ou de aparecimento dos cânceres de mama e endometrial (Figura 5).⁵¹⁻⁵⁴ Sabe-se, ainda, que essas moléculas interagem com o ER.^{52,55} Por essa razão, esses produtos naturais foram denominados fitoestrógenos. O padrão de reconhecimento molecular dessas moléculas bioativas representa esqueletos estruturais de grande utilidade para abordagens avançadas em química medicinal.

Figura 5. Estruturas dos fitoestrógenos mais importantes utilizados na terapia de reposição hormonal.

5 OS RECEPTORES DE ESTRÓGENO

Os estrógenos são hormônios endógenos de crucial importância e exercem profundos efeitos no desenvolvimento, na diferenciação e na homeostase metabólica de vários tecidos, incluindo os tratos reprodutivos, as mamas e o sistema esquelético (Figura 6).^{32, 56-58}

Figura 6. Localização dos tecidos-alvo dos estrógenos no corpo da mulher.

Os estrógenos naturais incluem o estradiol, a estrona e o estriol, e, dentre eles, o estradiol é o que apresenta maiores concentrações sanguíneas e maior potência nos vários tecidos (Figura 7).⁵⁸

Figura 7. Estrutura química do estradiol, principal estrógeno feminino.

Os efeitos fisiológicos dos estrógenos são mediados pelos dois subtipos do receptor de estrógeno em humanos, $hER\alpha$ e $hER\beta$, membros da superfamília dos receptores nucleares, que são fatores de transcrição intracelulares modulados por ligantes.⁵⁹⁻⁶² A ligação do estradiol (ligante natural) ou outros ligantes ao ER, dispara uma complexa rede de sinalização intracelular que leva ao recrutamento de complexos co-regulatórios e culmina com a transcrição ou repressão de genes-alvo específicos, relacionados à proliferação e diferenciação celular.^{32,56-58}

5.1 Domínios Estruturais dos Subtipos do Receptor de Estrógeno

O primeiro receptor de estrógeno humano (hER α) foi caracterizado bioquimicamente na década de 60 e subsequente clonado na década de 80. Apenas mais recentemente, em 1996, o segundo subtipo do receptor do estrógeno (hER β) foi descoberto.⁵⁹⁻⁶² Os hERs compartilham uma estrutura funcional conservada de seis domínios nomeados de A à F: uma região N-terminal variável com função de transativação (A/B), um domínio de ligação ao DNA bem conservado (DBD, do inglês, *DNA Binding Domain*, C), uma região gênica envolvida na dimerização e ligação de proteínas de choque térmico (D), o domínio de ligação ao ligante (LBD, do inglês, *Ligand Binding Domain*, E) e uma região carboxi-terminal que modula a ativação transcricional do ER (F). Os dois subtipos do receptor apresentam identidade seqüencial de 58% para os domínios de LBD e de 95% para DBD, enquanto regiões A/B, hinge e F não são conservadas (Figura 8).^{32,56}

Figura 8. Estruturas de domínios dos subtipos do receptor de estrógeno.

Os hERs apresentam duas regiões, conhecidas como funções de ativação (AF, do inglês, *Activation Function*), que contribuem para a atividade transcricional. Uma delas (AF-1) é localizada dentro da região A/B na porção N-terminal da proteína. A segunda região (AF-2) encontra-se no C-terminal da proteína. A região AF-2 interage com um grande número de ativadores de transcrição e esta interação pode ser modulada por ligantes do ER.^{32,56}

Várias estruturas cristalográficas do LBD (domínio E) de hER α e hER β em complexo com agonistas e antagonistas foram elucidadas. Para ambos os subtipos, a estrutura tridimensional do LBD é composta principalmente por α -hélices (Figura 9).³² A cavidade de interação com ligantes encontra-se em uma das extremidades da estrutura e é praticamente idêntica nos dois subtipos do ER, diferindo por apenas dois resíduos de aminoácidos, com

mudanças que são bastante conservativas (Leu384 de hER α é substituída por Met336 em hER β e Met421 de hER α é substituída por Ile373 em hER β).⁶²⁻⁶⁵ Essa é a razão pela qual o 17 β -estradiol apresenta afinidade similar para ambos os receptores.^{62,66}

Figura 9. Representação dos LBDs de hER α (azul) e hER β (magenta) em complexo com o estradiol (verde).

Os receptores de estrógeno estão amplamente distribuídos por todos os tecidos do corpo humano. Entretanto, existem certas diferenças na distribuição dos subtipos. Por exemplo, enquanto hER α é o principal receptor expresso no útero, ovários, mamas e fígado, hER β é o único receptor encontrado no trato gastrointestinal.^{62,67}

5.2 Mecanismo de Ação

O mecanismo de ação do estradiol envolve a difusão do ligante hidrofóbico através da membrana celular e sua ligação ao receptor no citoplasma. A ligação do estradiol ao ER promove mudanças conformacionais que provocam a dissociação de proteínas de choque térmico, migração ao núcleo, dimerização e ligação a seqüências específicas do DNA, estimulando a transcrição de genes responsivos (Figura 10). O controle da transcrição requer a

interação com um complexo de proteínas co-regulatórias (e.g., coativadores e co-repressores) que agem como intermediários de sinalização entre o ER e a maquinaria basal de transcrição.^{32,56-58}

Figura 10. Ativação do receptor de estrógeno mediada pelo ligante endógeno estradiol.

A estrutura cristalográfica do LBD de hER α em complexo com o estradiol demonstra que esse ligante de alta afinidade é posicionado na cavidade por complementaridade estereoquímica e eletrostática (Figura 11). A interação ocorre num bolsão hidrofóbico do LBD, onde o anel fenólico A do estradiol interage com os resíduos Glu353 e Arg394 e o anel D interage com His524 em hER α . Os resíduos são Glu305, Arg346 e His475, respectivamente, em hER β .^{32,68}

Figura 11. Ligação do estradiol no bolsão hidrofóbico de hER α .

A alta flexibilidade da cavidade de interação do ER permite a interação do receptor com uma série de moléculas estruturalmente diversas. Além da afinidade pela cavidade, esses compostos podem apresentar diferentes atividades intrínsecas, ou seja, diferentes habilidades para produzir a alteração conformacional adequada do LBD e, conseqüentemente, o efeito biológico. Essas moléculas podem modular o receptor da mesma forma que o estradiol, ativando-o, e são, portanto, chamadas de agonistas do receptor. Entretanto, essas moléculas podem interagir com a cavidade de interação, mas não levar à transcrição dos genes-alvo. Nesse caso, essas moléculas irão competir com o estradiol pela cavidade de interação e impedir os seus efeitos e são, portanto, denominadas antagonistas. Assim, os agonistas do ER estabilizam a sua conformação ativa específica, ao passo que os antagonistas estabilizam conformações inativas.^{32,56,68-71}

5.3 Hélice 12

A ligação de coativadores a ambos os subtipos do ER é dependente das alterações conformacionais que ocorrem no LBD, provocadas pela interação com o estradiol. Na ausência do hormônio, a hélice 12 do LBD, apresenta uma conformação aberta. Entretanto, na presença do ligante, a hélice 12 é reposicionada de forma a interagir com o este, selando a entrada da cavidade de interação (Figura 12).

Figura 12. Mudanças conformacionais no LBD do ER provocadas pela interação com o estradiol. A. Na ausência do estradiol. B. Reposicionamento da hélice 12 na presença do estradiol.

As alterações conformacionais provocadas pelo reposicionamento da hélice 12 modificam o formato do LBD e criam uma superfície necessária para a interação com os coativadores. Portanto, o reposicionamento da hélice 12 é essencial para a interação do ER com coativadores e, consequentemente, para a transcrição gênica.^{32,68-70}

Estudos das estruturas tridimensionais de complexos dos receptores de estrógeno com diferentes ligantes, obtidas por difração de raios-X, demonstram que a posição da hélice 12 é bastante diferente em complexos com ligantes agonistas e com antagonistas. A posição da hélice é semelhante à observada para o estradiol quando a micromolécula age como agonista. Contudo, a hélice não se reposiciona de forma a selar a entrada da cavidade quando o ligante é um antagonista (Figura 13). A falha em assumir a conformação produtiva compromete a ligação do ER aos coativadores, alterando de maneira significativa a sua capacidade de ativar a transcrição. Assim, as alterações conformacionais na cavidade de interação, e particularmente na posição N-terminal da hélice 12, são os maiores determinantes das propriedades agonistas e antagonistas das substâncias capazes de modular o ER.^{32,68-70}

Figura 13. A. Dímero do LBD de hER α com estradiol (azul). Note conformação fechada da hélice 12 (magenta). B. Esta conformação é muito diferente da conformação aberta da mesma hélice em complexo com o antagonista raloxifeno (vermelho).

O entendimento do processo de reconhecimento molecular responsável pelos efeitos farmacológicos é, entretanto, muito mais complexo, já que a ação dos receptores de estrógeno é tripartida, envolvendo não só as características dos ligantes, mas também o subtipo do receptor e as proteínas co-regulatórias presentes no ambiente celular.^{50,72}

5.4 Interações Intermoleculares

A característica agonista ou antagonista dos diferentes compostos interagindo com o ER pode ser estudada à luz da teoria da perturbação macromolecular. Quando o receptor está no estado desligado, pode interagir com um agonista que produza uma perturbação conformacional específica e conseqüente ação estimuladora. Entretanto, a interação com um antagonista produz uma perturbação conformacional não-específica (Figura 14). A existência na micromolécula de uma cadeia lateral antiestrogênica corretamente posicionada é, então, essencial para o caráter antiestrogênico, e o comprimento da cadeia está relacionado à alteração conformacional (não específica) no ER. Assim, a cadeia lateral interage com uma “região antiestrogênica” no ER, impedindo sua ativação.^{32,73}

Figura 14. Modelo hipotético para a ligação estrogênica e antiestrogênica ao ER. O estradiol é ancorado no sítio fenólico (SF) com alta afinidade. O monohidroxitamoxifeno tem alta afinidade pelo sítio de ligação, mas possui uma cadeia lateral que interage com uma região antiestrogênica.

Os antiestrógenos raloxifeno e 4-hidroxitamoxifeno possuem complementaridade multifuncional adequada para interagir com a região hidrofóbica na cavidade de interação do LBD do ER, assim como o estradiol. Entretanto, a presença de uma “cadeia lateral antiestrogênica”, representada por um grupo aminoetoxifenila, previne a reorientação da hélice 12 e as alterações conformacionais necessárias para a interação dos coativadores e a formação do complexo de transcrição.

Os antagonistas de alta afinidade interagem, através de grupos hidroxila fenólicos, com os resíduos Glu353 e Arg394, em hER α , e se ancoram adequadamente na cavidade de interação, assim como o estradiol. Entretanto, a cadeia lateral, que é essencial para a atividade antiestrogênica, interage com um resíduo de Asp351 (Asp303 em hER β), através de uma interação eletrostática entre o aminoácido negativamente carregado e a cadeia lateral protonada (Figura 15 C e D). A distribuição de cargas ao redor do resíduo Asp351 (ou Asp303 em hER β) e a cadeia lateral volumosa previnem o reposicionamento adequado da hélice 12. Essa interpretação é consistente com os resultados de cristalografia de raios-X, já que a ligação do coativador na região AF-2 é bloqueada pela posição assumida pela hélice 12. O mesmo, entretanto, não ocorre para o agonista dietilbestrol, que apresenta posicionamento na cavidade semelhante ao estradiol (Figura 15 A). A genisteína é um agonista de hER β que também apresenta posicionamento semelhante ao estradiol na cavidade de interação (Figura 15 B).^{32,33,68-70}

6 OS RECEPTORES DE ESTRÓGENO COMO ALVOS MACROMOLECULARES PROMISSORES

Os estudos farmacológicos com o ER indicam algumas particularidades, já que *in vivo* muitos compostos possuem atividade pleiotrópica, ou seja, apresentam uma variedade de efeitos em diferentes tecidos. Considerando-se que são diversos os órgãos-alvo dos estrógenos, é perfeitamente explicável que as substâncias que atuam no ER também provoquem uma multiplicidade de efeitos farmacológicos.⁷² Além disso, alguns compostos moduladores do ER possuem ação agonista em alguns tecidos e antagonista em outros, e são, por isso, denominados moduladores seletivos. O termo Modulador Seletivo do Receptor de Estrógeno (SERM, do inglês, *Selective Estrogen Receptor Modulator*) foi criado pelos laboratórios Eli Lilly na década de 90 para descrever a multiplicidade de efeitos dessas moléculas, que interagem com o ER em diferentes sítios.

Apesar dos níveis de variação nas características agonistas e antagonistas em diferentes tecidos, sabe-se que todos esses compostos agem através dos ERs. Por exemplo, em mulheres, o raloxifeno apresenta forte atividade agonista nas células ósseas e no fígado, como indicado pela sua capacidade de manutenção da densidade óssea e dos perfis estrogênicos de lipídeos no sangue. Entretanto, esse fármaco apresenta uma pequena estimulação uterina e, ainda, ação antagonista nas mamas. A atividade agonista nas células ósseas explica a utilização do raloxifeno no tratamento da osteoporose. A mesma característica nas células do fígado explica o aumento da incidência de tromboembolismo associado ao uso desse fármaco. O tratamento com tamoxifeno, em mulheres com câncer de mama, revela que, além da atividade antagonista nas mamas, este fármaco apresenta ações semelhantes ao estrógeno na manutenção da densidade mineral óssea, nas concentrações sanguíneas de lipoproteínas, na elevada estimulação uterina e até nas mamas, quando ocorre resistência ao tratamento. A elevada estimulação uterina explica as graves reações adversas originárias de desordens da função endometrial, incluindo aumento da incidência de câncer uterino. Os antiestrógenos esteroidais, em contraste, como o fulvestranto, possuem um caráter antagonista em todos os tecidos-alvo dos estrógenos.^{32,33} Esse comportamento complexo pode ser analisado à luz de três mecanismos, em diferentes níveis: i. o efeito provocado pelo ligante; ii. o

efeito dependente do receptor; e, iii. o efeito dependente do sítio efetor. O primeiro efeito é dependente da capacidade dos ligantes em provocar o reposicionamento da hélice 12, como discutido anteriormente. No segundo mecanismo, tecidos distintos podem responder diferentemente aos mesmos moduladores porque apresentam um balanço diferente entre os subtipos do ER. Ensaio celulares demonstraram que ER β pode, inclusive, apresentar ações opostas ao ER α em alguns tecidos, como nas mamas, onde a expressão positiva de ER β está correlacionada com um prognóstico favorável no câncer desse tecido. Nos sistemas em que isso ocorre, as mesmas moléculas podem apresentar seletividade de ação como o resultado da interação diferencial com os diferentes subtipos do ER presentes no tecido-alvo.^{50,62,72} O efeito farmacológico provocado pelos moduladores do receptor de estrógeno é, ainda, dependente de um nível mais profundo de complexidade, associado aos sistemas efetores envolvidos nas diferentes respostas provocadas em diferentes tecidos.^{50,72}

Assim, os graves efeitos colaterais associados aos fármacos disponíveis hoje no mercado estão relacionados ao efeito da seletividade multidimensional característica das interações intermoleculares. Um SERM ideal deveria, portanto, apresentar diferentes comportamentos em tecidos distintos, como ação antagonista nos receptores presentes nos tecidos das mamas e do útero e concomitante ação agonista nos receptores presentes nas células ósseas.^{32,33}

Um aspecto fundamental no planejamento de novos moduladores do ER é explorar as propriedades da estrutura química dos compostos em combinação com a habilidade em induzir uma resposta farmacológica apropriada como consequência da sua interação com o receptor. Fazem-se, portanto, primordiais as ações de P&D de NCEs com reconhecimento molecular privilegiado em sítios de ação distintos, minimizando assim, de forma planejada, reações adversas indesejadas. Desta forma, novas estratégias podem ser exploradas para o desenvolvimento de moduladores que apresentem capacidade multifuncional apropriada nos diferentes tecidos.

7 QUÍMICA MEDICINAL E AS TÉCNICAS DE QSAR

Uma das técnicas muito empregadas em química medicinal no planejamento de NCEs candidatas a fármacos é o estudo das relações quantitativas entre a estrutura e atividade (QSAR, do inglês, *Quantitative Structure-Activity Relationships*). As técnicas de QSAR buscam identificar e quantificar as relações predominantes no complexo campo da modelagem, representadas pelas propriedades da estrutura química e atividade biológica correspondente, com o objetivo de encontrar moléculas com propriedades superiores em relação às aquelas apresentadas pelo conjunto de dados que deu origem ao processo de modelagem. Para que tais estudos sejam realizados de maneira padrão e comparativa, as moléculas do conjunto treinamento precisam ser quimicamente relacionadas (ocupar o mesmo espaço químico), exercer seus efeitos através de um mesmo receptor, na mesma cavidade de ligação e através do mesmo mecanismo de ação. A propriedade alvo, geralmente potência ou afinidade, deve ser padronizada e validada para ser útil no desenvolvimento de modelos quantitativos. Propriedades moleculares são calculadas e empregadas como descritores. Métodos estatísticos são usados para produzir uma relação que descreve as variações na propriedade alvo, em função dos descritores. Esta relação, denominada QSAR, pode ser utilizada em programas de planejamento para orientar a síntese de novas moléculas com propriedades otimizadas. O processo cíclico de formulação da QSAR, predição da propriedade alvo, síntese e ensaios biológicos, prossegue até que moléculas qualificadas, candidatas a NCEs, sejam selecionadas para desenvolvimento clínico.^{27,74,75}

As técnicas de QSAR 3D são muito exploradas em conjunto com técnicas de modelagem molecular quando estruturas tridimensionais dos alvos macromoleculares eleitos são disponíveis.⁷⁶ Uma vez que as características químicas e estruturais associadas às propriedades biológicas alvo sejam identificadas na modelagem de QSAR, estas podem ser comparadas com as informações obtidas no estudo das características topológicas tridimensionais dos receptores macromoleculares estudados. A integração dessas informações altamente qualificadas é uma estratégia importante em química medicinal, uma vez que as informações complementares geradas são úteis no planejamento de candidatos a NCEs mais

potentes e seletivas, como é o caso dos modulares do ER, alvo dos estudos dessa dissertação de mestrado.

8 HOLOGRAMA QSAR

O Holograma QSAR (HQSAR, do inglês, *Hologram Quantitative Structure-Activity Relationships*) é um método de QSAR bidimensional cuja premissa é a de que a estrutura molecular pode ser convertida em notações lineares, ou seja, seqüências numéricas que codificam a estrutura 2D das moléculas do conjunto de dados. Baseando-se no conceito central do QSAR de que a atividade biológica é função da estrutura molecular, é possível estabelecer relações quantitativas entre a atividade de uma série de moléculas a partir das informações codificadas nos descritores bidimensionais. O método se baseia na expansão do conceito de impressão digital molecular (do inglês, *fingerprint*), utilizando uma forma estendida (de impressão digital molecular) que é conhecida como holograma molecular, que se caracteriza por uma maior informação estrutural codificada em relação a impressão digital tradicional. Enquanto que no método tradicional apenas a notação binária (possui ou não o fragmento) é utilizada, o holograma molecular contém, além de todos os possíveis fragmentos moleculares possíveis, uma contagem do número de vezes que cada fragmento ocorre na molécula.⁷⁷⁻⁷⁹

Na técnica de HQSAR, os hologramas moleculares são originários da fragmentação molecular bidimensional. Todos os fragmentos moleculares possíveis (lineares, ramificados e sobrepostos) são obtidos e posteriormente distribuídos pelo holograma. Cada posição do holograma contém a frequência com que um determinado fragmento aparece na molécula, sendo que fragmentos iguais que ocorrem mais de uma vez são sempre alocados na mesma posição (Figura 16). Informações tridimensionais adicionais, como hibridização e quiralidade são também codificadas nos hologramas moleculares. Desta forma, os hologramas definem a dimensionalidade do espaço ocupado pelos descritores e os modelos de QSAR são, então, desenvolvidos empregando-se o método de regressões multivariadas pelos mínimos quadrados parciais (PLS, do inglês, *Partial Least Squares*) e técnicas de validação cruzada.

Figura 16. Esquema ilustrativo do processo de geração dos hologramas moleculares.

Durante o processo de modelagem de HQSAR, vários parâmetros relevantes para a obtenção dos hologramas, e, portanto, que afetam a qualidade dos modelos, podem ser definidos. Esses parâmetros são: i. comprimento do holograma; ii. tamanho dos fragmentos; e, iii. distinção dos fragmentos. O comprimento do holograma define o número de posições, no holograma, disponíveis para que diferentes fragmentos sejam incorporados. O tamanho dos fragmentos é definido pelo número máximo (N) e mínimo (M) de átomos, que é imposto aos fragmentos a serem gerados e os parâmetros de distinção de fragmentos definem as características moleculares que distinguem um fragmento do outro. Como, em geral, o comprimento do holograma é menor que o número total de fragmentos diferentes gerados, fragmentos distintos são alocados no mesmo holograma. A maneira como os fragmentos são distribuídos nas posições do holograma influencia diretamente a qualidade dos modelos.^{77,80} Várias combinações de distribuição de fragmentos nas posições do holograma molecular são testadas para impedir que haja a colisão de fragmentos, ou seja, que fragmentos distintos e importantes sejam agrupados numa mesma posição, através do algoritmo cíclico de verificação de redundância (CRC, do inglês, *cyclic redundancy check*).

Os parâmetros utilizados para distinguir os fragmentos são ajustáveis, e de fundamental importância, pois definem o tipo de informação 2D que será incluída na criação dos descritores. Na Tabela 1 estão apresentados os diferentes parâmetros de distinção de fragmentos considerados no método HQSAR. Combinações desses parâmetros são utilizadas para escolha da combinação que produz o resultado estatístico mais relevante.^{77,80}

Tabela 1. Parâmetros de distinção de fragmentos no método HQSAR.

Distinção de Fragmentos	Definição	Representação
Átomo (A)	Fragmentos são distinguidos com base nos tipos de átomos	CCCCCFN
Ligação (B)	Fragmentos são distinguidos com base nos tipos de ligação: simples, dupla, tripla e aromática	
Conectividade (C)	Fragmentos são diferenciados de acordo com a hibridização dos átomos contidos no fragmento	
Hidrogênio (H)	Fragmentos são diferenciados com base na presença e número de átomos de hidrogênio	
Quiralidade (Ch)	Fragmentos são distinguidos de acordo com a presença de centros estereogênicos	
Doador e acceptor	Fragmentos são diferenciados com base na presença de átomos aceptores ou doadores de ligações hidrogênio	

Os hologramas gerados são os descritores que são relacionados com a propriedade biológica alvo através do método PLS, para o desenvolvimento do modelo estatístico de QSAR (Figura 17). Utilizando esse método, padrões de fragmentos subestruturais relevantes para a propriedade alvo são identificados. Uma nova molécula ainda não testada, e, portanto, não incluída no conjunto de dados utilizado para a modelagem, pode, então, ter a sua propriedade biológica predita quando submetida ao modelo criado.^{79,81,82}

Figura 17. Etapas envolvidas na geração de um modelo de HQSAR.

Além de prever a propriedade alvo de moléculas ainda não testadas, o modelo gerado de HQSAR fornece importantes indicações sobre quais fragmentos moleculares são importantes para a propriedade estudada. Estas informações podem orientar a síntese de novas moléculas com propriedades otimizadas. No método HQSAR, esse objetivo é atingido pela geração de mapas de contribuição individuais 2D, que utilizam um código de cores para indicar a magnitude da contribuição de diferentes fragmentos moleculares para a propriedade em estudo.⁷⁹⁻⁸²

Esta técnica moderna, que faz parte de um leque de tecnologias disponíveis em química medicinal, é capaz de gerar modelos com alta capacidade de interpolação e extrapolação em diversos conjuntos de dados.

9 ANÁLISE COMPARATIVA DOS CAMPOS MOLECULARES

A análise comparativa dos campos moleculares (CoMFA, do inglês, *Comparative Molecular Field Analysis*) é um método de QSAR 3D bastante utilizado no processo de otimização estrutural de compostos líderes. O fundamento da técnica é que a atividade biológica está diretamente relacionada com as interações fármaco-receptor, ou seja, as moléculas bioativas geralmente produzem seus efeitos através de interações não-covalentes estereoquímicas e eletrostáticas com o alvo macromolecular. Como essas interações apresentam caráter tridimensional, este método considera as moléculas em três dimensões, utilizando campos moleculares de interação (MIFs, do inglês, *Molecular Interaction Fields*) como descritores.^{74,75,83,84} Os campos moleculares são obtidos pelo cálculo das energias de interação estereoquímicas e eletrostáticas entre as moléculas e sondas virtuais, que consistem de um determinado átomo apresentando uma carga de prova. As diferenças na propriedade alvo (variável dependente) são desta forma, descritas em função das diferenças nestes campos moleculares de interação.

Diferentemente do HQSAR, as conformações e as cargas parciais atômicas das moléculas alteram de forma significativa o cálculo dos descritores 3D, que são os campos moleculares estereoquímicos e eletrostáticos. Dessa forma, é necessário que sejam geradas conformações de baixa energia e que as cargas atômicas parciais sejam calculadas para as moléculas utilizadas na modelagem. Além disso, as conformações do conjunto de moléculas precisam ser submetidas a um alinhamento estrutural único de forma a apresentar mesma orientação no espaço cartesiano. Estas são as principais dificuldades inerentes a este método, quando comparado aos métodos de QSAR 2D.^{74,75,83,84}

Uma grade tridimensional reticulada suficientemente ampla é, então, posicionada ao redor das moléculas superpostas e uma distância de intersecção entre os retículos da grade é definida. Em cada um desses pontos de intersecção são calculadas as energias de interação estereoquímicas e eletrostáticas entre a sonda virtual e os átomos suficientemente próximos desse ponto. A sonda virtual pode ser definida, por exemplo, como um átomo de carbono com

hibridização sp^3 carregado positivamente. Esse processo é realizado até que todos os pontos de intersecção sejam percorridos pela sonda.

As energias de interação calculadas em cada ponto correspondem a uma tabela contendo centenas de colunas de descritores (variáveis independentes), que podem ser correlacionados com o valor da propriedade alvo (variável dependente) utilizando métodos de regressão multivariada. O método de regressão PLS é apropriado para esse propósito. Técnicas de validação cruzada são utilizadas para investigar a robustez interna do modelo obtido. O modelo estatístico gerado, dessa forma, correlaciona a propriedade alvo com os descritores e permite que novas moléculas ainda não testadas possam ter o valor da propriedade em estudo predito (Figura 18).^{83,84}

Figura 18. Etapas envolvidas na geração de um modelo de CoMFA.

Assim como no método HQSAR, o método CoMFA permite a avaliação visual das modificações estruturais relacionadas ao aumento do valor da propriedade biológica em estudo, através de poliedros especiais denominados mapas de contorno. Esses poliedros evidenciam regiões estereoquímicas e eletrostáticas, ao redor da molécula, que são posições de substituição com contribuições favoráveis ou desfavoráveis para a propriedade biológica em questão. O método CoMFA é uma das técnicas de QSAR mais empregadas pela indústria farmacêutica na construção de modelos preditivos para uma série de propriedades biológicas e moleculares.^{83,84}

10 O MÉTODO GRID/PCA

O método GRID é uma técnica computacional para determinação das regiões de interação energeticamente favoráveis em moléculas de estrutura tridimensional conhecida, podendo ser usado para o estudo de micro ou macromoléculas.^{76, 86}

Assim como no método CoMFA, a premissa básica é a de que as interações fármaco-receptor são extensivamente dirigidas por energias de interação não-covalentes, que apresentam caráter tridimensional. A diferença é que o estudo pode ser baseado na estrutura tridimensional da cavidade de interação do alvo macromolecular eleito. Os mapas tridimensionais são, nesse caso, obtidos pelo cálculo das energias de interação entre os átomos dos resíduos de aminoácidos na periferia da cavidade interação e as sondas químicas virtuais.^{76,85,86}

O mapa tridimensional pode ser visto como uma matriz contendo a medida das energias atrativas e repulsivas entre a molécula e a sonda. Cada vértice da grade reticulada representa um ponto a partir do qual as energias de interação com a sonda química são calculadas. O espaçamento da caixa pode ter a dimensão variada de 0,1 a 1 Å para que as energias de interação sejam exploradas adequadamente. Várias sondas, representando interações distintas, podem ser selecionadas, incluindo o grupo metila, o átomo de carbono aromático, os átomos de nitrogênio amino, amido e heterocíclicos, átomos de halogênio e enxofre, o carbono carbonílico, átomos de oxigênio de éter e de hidroxilas, o grupo carboxila,

água, metais e outros íons. As principais sondas virtuais estão definidas na Tabela 2. Regiões favoráveis para interação podem ser, assim, identificadas.^{85, 86}

Tabela 2. Principais sondas químicas utilizadas no método GRID.

Sondas Químicas	
OH2	Sonda água: é capaz de doar duas ligações de hidrogênio e receber duas. Simula o processo de solvatação e dessolvatação e explora interações eletrostáticas.
DRY	Sonda hidrofóbica: explora possíveis interações hidrofóbicas na superfície ou cavidade da molécula alvo. É a sonda água modificada, uma vez que é eletricamente neutra.

O método pode ser também utilizado para entender as diferenças estruturais relacionadas à seletividade, um aspecto fundamental no planejamento de fármacos.^{87,88} Para tal, as estruturas cristalográficas dos alvos macromoleculares envolvidos, desprovidas de qualquer ligante, devem ser preparadas e superpostas, e a grade reticulada deve ser definida de forma a envolver todos os resíduos da cavidade importantes para a interação em todas as macromoléculas estudadas. Em geral, mapas de interação hidrofóbicos (sonda DRY) e hidrofílicos (sonda OH2) são calculados. Diferentemente dos métodos de QSAR, em que uma correlação é feita entre os descritores e as variáveis dependentes (método supervisionado), no método GRID, o tratamento dos inúmeros descritores gerados é não-supervisionado, baseado apenas nas diferenças e similaridades entre as variáveis dependentes obtidas nos cálculos das energias de interação. Métodos como a análise das componentes principais (PCA, do inglês, *Principal Component Analysis*) são apropriados para a análise estatística dos campos (Figura 19).

Assim como nos métodos de QSAR descritos, o modelo GRID/PCA gerado permite a avaliação visual das regiões tridimensionais estreitamente relacionadas ao reconhecimento molecular e seletividade, através de mapas de contorno semelhantes aos gerados nos modelos de CoMFA.^{87,88}

Figura 19. Etapas envolvidas na geração de um modelo GRID/PCA.

O método GRID/PCA apresenta grande utilidade no planejamento baseado na estrutura do receptor.⁷² Em conjunto com as técnicas de QSAR, pode fornecer informações complementares de grande valor no planejamento de moduladores mais potentes e seletivos.

Capítulo 2

OBJETIVOS

Os objetivos fundamentais dessa dissertação de mestrado são:

- Organizar conjuntos de dados padrão de moduladores dos subtipos alfa e beta do receptor de estrógeno.
- Desenvolver modelos preditivos de QSAR 2D e 3D, empregando as técnicas HQSAR e CoMFA, respectivamente, para os conjuntos padrão de dados selecionados
- Analisar os mapas de contribuição 2D e os mapas de contorno 3D e utilizá-los como guias em futuros trabalhos em Química Medicinal no planejamento de moduladores apresentando maior potência, afinidade e seletividade.
- Estudar a topografia molecular das cavidades de interação dos subtipos do receptor de estrógeno e desenvolver modelos para explicar as diferenças químicas entre os dois subtipos utilizando o método GRID/PCA.

Capítulo 3

PROCEDIMIENTO EXPERIMENTAL

1 MODELAGENS DE QSAR E GRID/PCA

Os estudos de modelagem molecular, análises estatísticas e geração de modelos de QSAR 2D, QSAR 3D e GRID/PCA, foram realizados em estações computacionais no Laboratório de Química Medicinal e Computacional (LQMC) do Centro de Biotecnologia Molecular Estrutural (CBME), no Instituto de Física de São Carlos da Universidade de São Paulo (IFSC/USP). Os trabalhos foram desenvolvidos em estações de trabalho Enterprise Linux utilizando as interfaces computacionais SYBYL 7.2 (Tripos, St. Louis, EUA), GOLD 3.0 (Cambridge Crystallographic Data Centre, Cambridge, UK), GRID 21 (Molecular Discovery Ltd., West Way House, Elms Parade, Oxford) e Pymol 0.99 (DeLano Scientific, San Carlos, USA).

2 CONJUNTOS DE DADOS

A identificação da diversidade química (esqueletos padrão) representada pelos compostos capazes de interagir com o ER envolveu uma busca intensiva na literatura e por patentes, empregando o “United States Patent and Trademark Office” (USPTO, <http://www.uspto.gov>). Após diversas pesquisas usando diferentes combinações de palavras-chave, um elevado número de patentes foi identificado, coletado e analisado individualmente. O processo de seleção das patentes de interesse resultou na identificação de um número reduzido de patentes relacionadas a substâncias capazes de interagir e modular a atividade do ER. O gerenciamento da informação obtida envolveu o agrupamento das patentes selecionadas em grupos de esqueletos químicos (classe estrutural) a que se referiam. Uma vez que o espaço químico–biológico de compostos capazes de se ligar ao receptor estrógeno foi identificado, moduladores representativos de cinco classes químicas foram identificados na literatura e bases de dados padrão foram organizadas para a realização do processo de modelagem de QSAR 2D e 3D.

Dois conjuntos padrões de dados foram organizados para moduladores de hER α pertencentes às classes das diidrobenezoxatiinas, flavanóides e diidrobenezoditiinas.⁸⁹⁻⁹⁹ O

conjunto de dados 1 consiste de 127 moduladores de hER α associados aos valores de afinidade de ligação (pIC₅₀) correspondentes, ao passo que o conjunto de dados 2 é um subconjunto (do conjunto de dados 1) contendo 69 moduladores de hER α associados aos respectivos valores de potência (inibição do crescimento de células MCF-7 *in vitro*). As estruturas químicas do conjunto de moduladores, associados aos valores correspondentes de afinidade e potência, são apresentadas na Tabela 3, no Capítulo 4 dessa dissertação.

As estruturas tridimensionais dos moduladores foram geradas utilizando-se as ferramentas disponíveis no programa computacional SYBYL 7.2, sendo posteriormente submetidas aos procedimentos de minimização de energia empregando o campo de forças Tripos. Moléculas apresentando a função amina foram modeladas no estado protonado, pois é esperado que estejam ionizadas em condições fisiológicas.

O conjunto de dados 3, empregado no desenvolvimento dos modelos de QSAR 2D e 3D, contém 81 moduladores das classes dos fenilnaftalenos e fenilquinolinas associados aos valores de afinidade de ligação (pIC₅₀) para os receptores hER α e hER β , como mostrado na Tabela 13, no Capítulo 4.^{100,101} As estruturas dos moduladores foram criadas utilizando-se as ferramentas disponíveis no programa computacional SYBYL 7.2, sendo posteriormente submetidas aos procedimentos de minimização de energia e cálculo de cargas atômicas parciais através do método AM1, no programa MOPAC 6.0.

Para todos os conjuntos de dados, as estruturas 3D geradas foram empregadas no processo de alinhamento estrutural, necessário para a realização do método CoMFA.

3 IDENTIFICAÇÃO DAS ESTRUTURAS TRIDIMENSIONAIS DE ERs

Aspectos estruturais do receptor alvo foram considerados na modelagem, tanto para o planejamento baseado na estrutura do receptor quanto nos estudos de QSAR 3D. Uma ampla pesquisa no banco de dados do PDB (<http://www.rcsb.org/pdb>) foi realizada para a identificação das estruturas cristalográficas de ERs humanos disponíveis. Este trabalho resultou em 31 estruturas cristalográficas referentes ao LBD do ER. Destas, 23 estruturas correspondem ao hER α e oito ao hER β . Trinta dessas estruturas apresentam-se em complexos com ligantes das mais diversas classes químicas, previamente identificadas em nossos estudos, enquanto que apenas uma não apresentava qualquer ligante em complexo. Três estruturas apresentavam mutações e treze estruturas foram cristalizadas com os respectivos coativadores ou co-repressores.

4 ALINHAMENTO ESTRUTURAL

O alinhamento estrutural tridimensional é uma das etapas cruciais na modelagem de QSAR 3D CoMFA, pois afeta os cálculos dos campos de interação moleculares e, por conseguinte, a qualidade dos resultados finais obtidos. Por isso, a escolha da estratégia de obtenção do alinhamento molecular é muito importante. A estratégia empregada foi o posicionamento das moléculas dos conjuntos de dados nas cavidades de interação dos subtipos do ER, de forma a encontrar conformações de interação no interior das cavidades. Para tal, foram utilizados os protocolos de acoplamento implementados no programa GOLD 3.0.^{102,103}

Para cada conjunto de dados, a estrutura cristalográfica adequada foi selecionada. Para o conjunto de dados 1, a estrutura de hER α (código PDB 1XP1) em complexo com um dos moduladores da base de dados, (2S,3R)-2-(4-{2-[(3R,4R)-3,4-dimetilpirrolidin-1-il]etoxi}fenil)-3-(4-hidroxifenil)-2,3-diidro-1,4-benzoxatíin-6-ol (composto 33, Tabela 3), foi escolhida. Para o conjunto de dados 3, as estruturas de códigos 1QKM (para hER β) e 1X7R (para hER α), em complexo com a genisteína, foram selecionadas, na indisponibilidade da estrutura em

complexo com um dos moduladores da base de dados. Outros critérios foram adicionalmente adotados na escolha das estruturas, como a conservação dos aminoácidos de ligação, a ausência de mutação e a resolução das estruturas.

O processo de preparo das estruturas selecionadas para a modelagem envolveu os seguintes procedimentos: (i) verificação da conservação da cavidade de interação; (ii) análise dos resíduos de aminoácidos principais; (iii) remoção dos ligantes na cavidade de interação; (iv) eliminação das moléculas de água; (v) adição de átomos de hidrogênio (utilizando o módulo Biopolymer, implementado na plataforma SYBYL 7.2); e (vi) verificação manual dos possíveis rotâmeros, dos estados de protonação e dos estados tautoméricos de resíduos de His, Gln e Asn (utilizando Pymol 0.99). O sítio de ligação foi centralizado no átomo C22 do modulador co-complexado para os procedimentos de acoplamento dos moduladores do conjunto de dados 1, e no átomo C7 da genisteína para as simulações de posicionamento dos moduladores do conjunto 3. Um raio de 10 Å foi considerado nas simulações. Os modelos de conformação bioativa obtidos foram exportados para o programa SYBYL 7.2 e utilizados para os estudos de QSAR 3D.

5 MODELOS DE HQSAR

O módulo de HQSAR da plataforma SYBYL 7.2 foi empregado na produção dos modelos de HQSAR para as séries de moduladores do receptor de estrógeno, apresentadas nas Tabelas 3 e 13. As estruturas das moléculas do conjunto de dados e os valores correspondentes das propriedades biológicas alvo, convertidas em unidades logarítmicas, pIC_{50} ($-\log IC_{50}$), foram transferidos para planilhas moleculares (MSS, do inglês, *molecular spreadsheet*), onde toda a informação para os processos de modelagem de QSAR ficou armazenada. A etapa seguinte consistiu na seleção de diversos parâmetros referentes à geração dos hologramas para as moléculas do conjunto treinamento. Estes parâmetros foram: comprimento do holograma, tamanho dos fragmentos e distinção dos fragmentos.

Os hologramas foram gerados usando diferentes combinações dos parâmetros de distinção: átomos (A), ligações (B), conectividade (C), átomos de hidrogênio (H) e átomos

doadores e aceptores de ligação de hidrogênio (DA). As análises de HQSAR foram executadas através de uma abordagem de triagem-erro, com uma série padrão de hologramas de diferentes comprimentos (entre 53 e 401), que forneceu a melhor consistência interna. Adicionalmente, a influência do tamanho do fragmento nos modelos obtidos também foi investigada, variando os números máximos e mínimos de átomos nos fragmentos (2-5, 3-6, 4-7, 5-8 e 6-9).⁷⁸⁻⁷⁰ Os padrões de frequência dos fragmentos nos hologramas gerados para os moduladores das séries de treinamento foram então relacionados aos parâmetros experimentais usando análise de PLS. Técnicas de validação cruzada foram aplicadas para determinar o número de componentes principais que resultavam em modelos mais consistentes. Os modelos finais foram utilizados no processo de validação externa dos conjuntos testes. Os resultados são apresentados no Capítulo 4 desta dissertação.

6 MODELOS CoMFA

O módulo CoMFA da plataforma SYBYL 7.2 foi empregado no desenvolvimento dos modelos de QSAR 3D para os mesmos conjuntos de dados de moduladores utilizados na modelagem de HQSAR. As moléculas alinhadas no espaço tridimensional foram utilizadas para o desenvolvimento dos modelos. Planilhas moleculares (MSS) foram criadas para a realização das análises de QSAR 3D, contendo toda a informação necessária em relação à estrutura e atividade das moléculas dos conjuntos de dados.

As moléculas dos conjuntos treinamento alinhadas foram inseridas numa grade tridimensional reticulada de dimensões suficientes, com espaçamento de 2,0 Å entre os vértices nas direções x, y e z. Os mapas estereoquímicos e eletrostáticos foram gerados em cada vértice utilizando o campo de forças Tripos, e como sonda virtual, um átomo de carbono de hibridização sp^3 carregado positivamente. As contribuições dos campos eletrostáticos e estereoquímicos foram calculadas de acordo com os potenciais de Coulomb e de Lennard-Jones, respectivamente, como implementado no método CoMFA.^{83,84} Os valores de corte para as energias de interação eletrostáticas e estereoquímicas foi de 30 Kcal/mol. Os descritores foram escalonados pelo método CoMFA padrão para que as energias de interação fossem

tratadas em mesma ordem de grandeza. O método PLS foi utilizado para correlacionar estatisticamente os valores das energias de interação em cada ponto com os valores das propriedades biológicas. Técnicas de validação cruzada foram aplicadas para determinar o número necessário de componentes principais. A possibilidade de geração de modelos aleatórios foi verificada pela permutação dos valores da propriedade biológica entre os compostos (do inglês, *progressive scrambling*).

Várias análises de QSAR foram realizadas para o refinamento dos modelos, introduzindo-se variações na geração dos descritores 3D e nas análises estatísticas. Os modelos gerados, com alta capacidade de correlação e predição, são apresentados no Capítulo 4 dessa dissertação.

7 MÉTODO GRID/PCA

Os cálculos das energias de interação molecular com as sondas químicas OH2 e DRY foram realizados utilizando o programa GRID 22.^{85,86} Estruturas cristalográficas dos subtipos alfa e beta do ER humano com ligantes diversos foram utilizadas para os estudos e estão apresentadas na Tabela 21, no Capítulo 4. As estruturas tridimensionais, desprovidas de qualquer ligante, foram verificadas no módulo Biopolymer, na plataforma SYBYL 7.2 e alinhadas pelos seus carbonos alfa. Íons de potássio foram adicionados como contraíons utilizando o módulo MINIM e FILMAP do programa GRID 22, para eliminar diferenças eletrostáticas relevantes entre as estruturas dos receptores. Uma grade tridimensional reticulada foi determinada de forma a incluir todas as posições importantes de interação ao redor da cavidade, resultando em uma grade de dimensões 16 x 12 x 15 Å. Um espaçamento de 0,5 Å entre os vértices da grade foi selecionado, nas direções x, y e z. A caracterização multivariada da cavidade de interação pelas energias de interação não-covalentes entre os resíduos de aminoácidos dos receptores e as sondas químicas OH2 e DRY foi realizada utilizando o campo de forças GRID 22. A diretiva MOVE do programa foi utilizada de forma a permitir que pares de elétrons não-ligantes, átomos de hidrogênio tautoméricos e os ângulos

de torção de grupos hidroxila sp^3 ou amina sp^2 se movessem em resposta às diferentes posições das sondas.^{85,86}

A avaliação estatística dos campos moleculares obtidos (descritores 3D) foi realizada empregando a análise fatorial implementada no programa SYBYL 7.2. Os campos de interação hidrofílicos e hidrofóbicos foram analisados em conjunto. Para tal, as variáveis dependentes foram escalonadas de forma que as diferentes magnitudes de energias de interação obtidas com as duas sondas diferentes fossem tratadas na mesma ordem de grandeza.^{87,88} O modelo gerado está apresentado no Capítulo 4 dessa dissertação.

Capítulo 4

RESULTADOS E DISCUSSÃO

Este capítulo foi dividido em duas partes para melhor entendimento do processo de construção e avaliação dos modelos de QSAR. Na Parte A é apresentado o planejamento de antagonistas do hER α de elevada afinidade e potência, com utilidade clínica no tratamento do câncer de mama. Na parte B, o foco é na otimização da afinidade e da seletividade de agonistas do hER β , úteis na terapia de reposição hormonal.

PARTE A

BASES QUÍMICAS E ESTRUTURAIS PARA A OTIMIZAÇÃO DA AFINIDADE E POTÊNCIA DE MODULADORES DO RECEPTOR DE ESTRÓGENO

1 CONJUNTOS PADRÕES DE DADOS

Os modelos de QSAR 3D CoMFA e QSAR 2D HQSAR foram desenvolvidos para uma série de 127 moduladores do hER α utilizando dois conjuntos de dados distintos⁸⁹⁻⁹⁹. O primeiro (conjunto de dados 1), contendo a série completa de 127 moduladores do ER, foi utilizado para criar os modelos de afinidade de ligação. A medida da afinidade é determinada pela capacidade do modulador em competir com o 17 β -estradiol. A concentração do modulador (em mol.L⁻¹) necessária para deslocar 50% do ligante natural marcado radioativamente determina o seu valor de IC₅₀. Esse é, portanto, um ensaio de competição com ³H 17 β -estradiol. Os valores de IC₅₀ variam de 0,30 a 1100 nM, um fator de aproximadamente 4.000 vezes.

O segundo conjunto de dados (conjunto de dados 2), contendo um subconjunto (do conjunto de dados 1) de 69 ligantes do ER, foi empregado para criar os modelos de potência. A medida da potência se refere à capacidade do modulador em inibir o crescimento *in vitro* de células de câncer de mama (MCF-7), na presença de 17 β -estradiol. Nesse ensaio celular, o biocaráter agonista ou antagonista do modulador pode ser determinado. A concentração do modulador necessária para reduzir em 50% a multiplicação das células, em relação ao padrão 17 β -estradiol, é a medida do valor de IC₅₀. Os valores de IC₅₀ para a série de 69 moduladores

variam de 0,03 a 560 nM, um fator de potência de aproximadamente 20.000 vezes. As estruturas químicas dos moduladores e os valores correspondentes da propriedade biológica são apresentados na Tabela 3.

Um ponto muito importante a ser destacado é que todos os valores de IC_{50} coletados neste trabalho foram determinados nas mesmas condições experimentais, o que é um fator essencial para o desenvolvimento de modelos confiáveis de QSAR.⁸⁹⁻⁹⁹ Os valores de IC_{50} das moléculas do conjunto de dados foram convertidos nos valores correspondentes na escala logarítmica de pIC_{50} ($-\log IC_{50}$) para o trabalho de modelagem de QSAR. Portanto, para uma análise comparativa, quanto maior o valor de pIC_{50} , maior a afinidade de ligação ou potência do composto da série.

As três principais classes químicas incluídas nos conjuntos de dados são as dos flavanóides, das diidrobenezoxatiinas e das diidrobenzoditiinas, e foram selecionadas cuidadosamente porque fornecem uma diversidade química apropriada para os estudos de QSAR (Figura 20).

Figura 20. Estrutura representativa da diversidade química dos conjuntos de dados.

Tabela 3. Estruturas químicas e valores associados de afinidade ligação e potência (pIC₅₀) do conjunto completo de moduladores de hERα

Código	Estrutura	Afinidade	Potência	Código	Estrutura	Afinidade	Potência
1		8,64	7,84	6		7,82	6,25
2		8,85	9,05	7		8,40	6,94
3		9,00	8,85	8		8,30	7,17
4		8,44	7,54	9		9,05	8,44
5		8,40	6,82	10		9,16	8,48

<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>	<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>
11		8,89	8,80	16		8,92	7,53
12		9,22	8,24	17		9,40	9,52
13		9,10	8,60	18		8,82	7,81
14		9,52	8,89	19		8,70	7,58
15		9,00	8,11	20		7,80	7,38

Código	Estrutura	Afinidade	Potência	Código	Estrutura	Afinidade	Potência
21		8,70	7,37	26		8,77	9,22
22		8,75	8,12	27		8,70	8,49
23		8,14	7,37	28		8,41	8,72
24		8,59	8,48	29		9,16	9,40
25		9,30	8,42	30		9,10	9,30

<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>	<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>
31		9,40	8,46	36		9,30	9,70
32		9,05	9,40	37		8,20	8,04
33		9,30	9,40	38		9,10	8,89
34		8,77	8,39	39		8,60	9,40
35		8,77	9,30	40		8,89	10,00

Código	Estrutura	Afinidade	Potência	Código	Estrutura	Afinidade	Potência
41		9,10	10,00	46		9,22	8,66
42		8,92	10,52	47		9,22	8,52
43		9,10	10,52	48		9,30	9,30
44		8,77	8,72	49		8,66	7,68
45		8,36	8,55	50		8,96	8,92

<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>	<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>
51		9,00	9,15	56		8,85	8,17
52		8,89	9,52	57		9,16	9,30
53		9,16	8,70	58		9,00	8,47
54		9,30	9,40	59		8,44	7,55
55		8,60	8,64	60		7,75	6,91

<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>	<i>Código</i>	<i>Estrutura</i>	<i>Afinidade</i>	<i>Potência</i>
61		7,68	6,95	66		9,05	8,80
62		9,16	8,59	67		9,52	8,42
63		8,66	7,95	68		9,16	9,40
64		8,52	9,05	69		8,92	8,51
65		9,00	9,00	70		7,85	

Código	Estrutura	Afinidade	Potência	Código	Estrutura	Afinidade	Potência
71		7,05		76		7,17	
72		7,41		77		6,19	
73		6,31		78		7,15	
74		6,26		79		7,24	
75		6,75		80		6,38	

Código	Estrutura	Afinidade	Potência	Código	Estrutura	Afinidade	Potência
81		6,03		86		5,96	
82		8,77		87		8,57	
83		8,21		88		6,31	
84		8,00		89		6,27	
85		5,98		90		8,06	

Código	Estrutura	Afinidade	Código	Estrutura	Afinidade
91		8,52	96		9,29
92		8,60	97		8,92
93		8,59	98		8,33
94		8,13	99		8,62
95		9,30	100		7,87

Código	Estrutura	Afinidade	Código	Estrutura	Afinidade
101		8,28	106		8,27
102		8,28	107		9,16
103		7,39	108		9,22
104		8,42	109		8,60
105		8,52	110		9,40

Código	Estrutura	Afinidade	Código	Estrutura	Afinidade
111		9,16	116		6,28
112		7,35	117		6,98
113		7,42	118		6,83
114		9,10	119		5,98
115		9,22	120		8,51

Código	Estrutura	Afinidade	Código	Estrutura	Afinidade
121		8,13	125		8,52
122		7,46	126		8,41
123		8,16	127		8,70
124		7,96			

As moléculas dos conjuntos de dados (afinidade e potência) foram divididas entre conjunto treinamento e conjunto teste. O conjunto treinamento é formado pelas moléculas que são utilizadas para a construção dos modelos de QSAR, ou seja, é a partir desta informação qualificada acerca da estrutura química e propriedade biológica que são gerados os descritores e são realizadas as análises estatísticas. O conjunto teste é utilizado no processo de validação externa, ou seja, na avaliação da capacidade do modelo em prever a propriedade alvo de novas moléculas que não foram incluídas durante o processo de modelagem no conjunto treinamento. Do total de 127 compostos do conjunto original, 28 (22%) foram incluídos no conjunto teste 1 enquanto que os outros 99 (78%) permaneceram no conjunto treinamento. Para o segundo conjunto de dados, 14 (20%) moduladores foram separados para a série teste, enquanto os outros 55 (80%) permaneceram no conjunto treinamento. Análises hierárquicas de agrupamentos (HCA, do inglês, *Hierarchical cluster analyses*) foram utilizadas para orientar a seleção apropriada dos conjuntos treinamento e teste para cada série. Em ambos os casos, os conjuntos foram selecionados de forma que moléculas estruturalmente diversas possuindo valores da propriedade biológica cobrindo toda a faixa da distribuição foram incluídas no conjunto teste (Figuras 21 e 22). O conjunto teste, em especial, apresenta uma distribuição representativa em comparação com o conjunto treinamento. Essa distribuição apropriada ao longo do espectro da propriedade biológica representa um aspecto muito importante na avaliação da consistência interna e externa dos modelos de QSAR.^{81,82} O método PLS foi utilizado em todas as análises de QSAR e a capacidade preditiva dos modelos foi avaliada de acordo com técnicas de validação cruzada e com as predições dos valores das propriedades da série teste.

Figura 21. Representação gráfica da distribuição dos valores de pIC_{50} para o conjunto de dados 1.

Figura 22. Representação gráfica da distribuição dos valores de pIC_{50} para o conjunto de dados 2.

2 HOLOGRAMA QSAR

Os modelos de HQSAR são particularmente sensíveis a variações dos parâmetros que especificam o tamanho e o tipo de fragmento codificado, além do comprimento do holograma molecular. Por isso, várias combinações desses parâmetros foram consideradas durante a otimização dos modelos de QSAR. A obtenção dos fragmentos moleculares foi realizada utilizando as seguintes distinções de fragmentos: átomos (A), ligações (B), conectividade, átomos de hidrogênio (H) e doadores e aceptores (DA). Foram empregadas 8 combinações desses parâmetros utilizando inicialmente o tamanho padrão de fragmento (4-7): A/B, A/B/C/H, A/B/C/H/DA, A/B/H, A/B/H/DA, A/B/DA, A/B/H/DA e A/B/C/DA. As análises foram realizadas considerando os 12 comprimentos de holograma disponíveis entre 53 e 401 (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401).⁷⁸⁻⁸⁰ Os resultados estatísticos a partir das análises de PLS para os conjuntos de dados 1 (afinidade) e 2 (potência) estão apresentados nas Tabelas 4 e 5, respectivamente.

Como pode ser observado para o modelo de afinidade, os melhores resultados estatísticos foram obtidos utilizando A/B/C/DA como informação para distinção de fragmentos ($q^2 = 0,71$ e $r^2 = 0,92$, com SEE de 0,27). Para o modelo de potência, entretanto, os melhores resultados estatísticos foram obtidos utilizando A/B/DA como informação para distinção dos fragmentos ($q^2 = 0,74$ e $r^2 = 0,92$, com SEE de 0,28). Em ambos os casos, os melhores modelos de HQSAR possuem o parâmetro DA (doador e aceptor). Esses resultados estão em concordância com estudos anteriores, os quais indicam a importância de ligações hidrogênio com resíduos específicos da estrutura da proteína para a atividade biológica.^{32,33,104,105}

Tabela 4. Análises de HQSAR para o modelo de afinidade de ligação para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7).

Distinção de Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
A/B	0,70	0,89	0,32	199	6
A/B/C/H	0,63	0,78	0,44	83	6
A/B/C/H/DA	0,61	0,85	0,37	353	7
A/B/H	0,60	0,75	0,48	401	5
A/B/H/DA	0,57	0,77	0,45	307	6
A/B/C/DA	0,71	0,92	0,27	257	7
A/B/DA	0,69	0,91	0,28	257	7
A/B/H/DA	0,570	0,774	0,453	307	6

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes. Distinção de fragmento: A, átomo; B, ligação; C, conectividade; H, hidrogênio; DA, doador e aceptor.

Tabela 5. Análises de HQSAR para o modelo de potência para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7).

Distinção de Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
A/B	0,72	0,93	0,26	353	7
A/B/C/H	0,69	0,89	0,33	53	7
A/B/C/H/DA	0,71	0,82	0,40	53	4
A/B/H	0,71	0,91	0,29	199	7
A/B/H/DA	0,68	0,89	0,32	97	7
A/B/C/DA	0,62	0,90	0,31	151	6
A/B/DA	0,74	0,92	0,28	97	7
A/B/H/DA	0,68	0,89	0,33	97	7

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes. Distinção de fragmento: A, átomo; B, ligação; C, conectividade; H, hidrogênio; DA, doador e aceptor.

A influência da variação do tamanho dos fragmentos nos parâmetros estatísticos foi investigada para os melhores modelos de HQSAR gerados previamente. Esses parâmetros podem ser modificados de maneira a incluir fragmentos maiores ou menores. Os resultados desta análise estão mostrados nas Tabelas 6 e 7.

Tabela 6. Análise de HQSAR para o modelo de afinidade de ligação em função da variação do tamanho dos fragmentos (distinção de fragmentos: A/B/C/DA).

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
2–5	0,71	0,89	0,32	257	7
3–6	0,71	0,89	0,32	257	7
4–7	0,71	0,92	0,27	257	7
5–8	0,69	0,85	0,33	61	7
6–9	0,71	0,88	0,32	353	7

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes.

Tabela 7. Análise de HQSAR para o modelo de potência em função da variação do tamanho dos fragmentos (distinção de fragmentos: A/B/DA).

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
2–5	0,72	0,91	0,30	401	7
3–6	0,72	0,92	0,28	199	7
4–7	0,74	0,92	0,28	97	7
5–8	0,72	0,91	0,29	59	7
6–9	0,67	0,94	0,25	353	7

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes.

Os resultados estatísticos para as análises individuais dos diferentes tamanhos dos fragmentos avaliados (2–5, 3–6, 4–7, 5–8 e 6–9) não levaram a diferenças significativas e não resultaram em melhorias nos parâmetros estatísticos nestas séries comparativas.

Os modelos finais de HQSAR gerados com os conjuntos treinamento possuem a validação cruzada, obtida pelo método “deixe um fora” (LOO, do inglês, *Leave-One-Out*), como

indicativo maior de sua robustez. Para melhor avaliar a consistência interna desses modelos, foi empregado o método “deixe muitos fora” (LMO, do inglês, *Leave-Many-Out*). Neste processo, grupos de 10 moléculas foram aleatoriamente retirados do conjunto de dados 1, e grupos de 7 moléculas do conjunto de dados 2, e os resultados não indicaram uma influência significativa nos parâmetros estatísticos. Esses parâmetros indicam a boa capacidade de correlação dos dados nos conjuntos treinamento e a elevada consistência interna dos modelos. Entretanto, apenas a validação cruzada não pode ser considerada suficiente, pois bons coeficientes de correlação associados a uma elevada consistência interna muitas vezes não resultam em modelos com elevada capacidade de prever a propriedade alvo para novos compostos.¹⁰⁶

Para que um modelo de QSAR apresente utilidade na otimização das propriedades de novas substâncias bioativas, é necessário que o mesmo seja validado externamente quanto a sua real capacidade preditiva. O teste mais importante de um modelo de QSAR é justamente a sua capacidade de prever o valor da propriedade alvo para compostos não incluídos no conjunto treinamento. Uma vez que a estrutura química codificada nos hologramas moleculares está diretamente relacionada ao valor da propriedade biológica (e.g., afinidade de ligação, potência) nos conjuntos de treinamento, os modelos de HQSAR são também capazes de prever o valor da propriedade de novos moduladores estruturalmente relacionados a partir dessa informação. Assim, o poder preditivo para os melhores modelos de afinidade e potência foi avaliado pela predição dos valores de pIC_{50} para as moléculas do conjunto teste, que foram completamente excluídas durante a derivação do modelo. Portanto, o conjunto teste é apropriado para avaliar o poder preditivo do modelo para novas moléculas dentro da mesma diversidade química. Os resultados do processo de validação externa são apresentados nas Tabelas 8 e 9. As pequenas diferenças entre os valores experimentais e preditos (resíduos das Tabelas 8 e 9) indicam a boa capacidade preditiva dos modelos HQSAR gerados. Os valores preditos se aproximam dos experimentais, desviando em não mais que 0,37 unidades logarítmicas em ambos os modelos.

Tabela 8. Valores preditos e experimentais de afinidade de ligação (pIC_{50}), associados aos resíduos, para os 28 compostos do conjunto teste.

Código	pIC_{50}	pIC_{50}	Resíduo
	Experimental	Predito	
43	9,10	9,05	0.05
44	8,77	8,62	0.15
45	8,36	8,13	0.23
46	9,22	9,25	-0.03
47	9,22	9,14	0.08
48	9,30	9,06	0.24
49	8,66	8,42	0.24
50	8,96	8,90	0.06
51	9,00	9,05	-0.05
52	8,89	8,84	0.05
53	9,16	9,11	0.05
54	9,30	9,11	0.19
55	8,60	8,59	0.01
56	8,85	9,04	-0.19
57	9,16	9,09	0.07
106	8,27	8,36	-0.09
116	6,28	6,65	-0.37
117	6,98	6,85	0.13
118	6,83	6,72	0.11
119	5,98	6,35	-0.37
120	8,51	8,34	0.17
121	8,13	8,43	-0.3
122	7,46	7,33	0.13
123	8,16	8,15	0.01
124	7,96	8,31	-0.35
125	8,52	8,60	-0.08
126	8,41	8,54	-0.13
127	8,70	8,51	0.19

Tabela 9. Valores preditos e experimentais de potência (pIC_{50}), associados aos resíduos, para os 14 compostos do conjunto teste.

Código	pIC_{50}	pIC_{50}	Resíduo
	Experimental	Predito	
56	8,17	8,33	-0.16
57	9,30	9,13	0.17
58	8,47	8,83	-0.36
59	7,55	7,44	0.11
60	6,91	6,94	-0.03
61	6,95	6,70	0.25
62	8,59	8,71	-0.12
63	7,95	8,07	-0.12
64	9,05	8,94	0.11
65	9,00	8,94	0.06
66	8,80	8,79	0.01
67	8,42	8,66	-0.24
68	9,40	9,43	-0.03
69	8,51	8,73	-0.22

A análise gráfica dos valores experimentais e preditos para os conjuntos treinamento e teste, para os dois modelos, confirma a boa concordância desses valores (Figuras 23 e 24). Como os resíduos são modestos, pode-se inferir que o modelo final gerado apresenta significativa consistência interna e externa para a diversidade estrutural estudada, possuindo comprovada capacidade preditiva.

Figura 23. Gráfico de HQSAR dos valores experimentais e preditos para as 127 moléculas do conjunto de dados (modelo para afinidade).

Figura 24. Gráfico de HQSAR dos valores experimentais e preditos para as 69 moléculas do conjunto de dados (modelo para potência).

Além de prever o valor das propriedades afinidade e potência de novas moléculas não incluídas no conjunto de treinamento, os modelos de HQSAR fornecem importantes indicadores sobre os diversos fragmentos moleculares que estão diretamente relacionados com a propriedade biológica em estudo. Estas informações estruturais diretamente associadas com as contribuições em afinidade e potência nos modelos gerados, são úteis no planejamento de novas moléculas com propriedades otimizadas.^{78,79,81,82} As informações são codificadas nos chamados mapas de contribuições de HQSAR, que determinam as contribuições individuais de diferentes átomos em uma determinada molécula com relação à propriedade alvo. O módulo de HQSAR implementado na plataforma SYBYL 7.2 utiliza um código de cores para atribuir as diferentes magnitudes de contribuição de cada átomo para a propriedade alvo, refletindo a contribuição dos diferentes fragmentos da molécula. Cores na região do vermelho do espectro (vermelho, laranja avermelhado e laranja) refletem contribuições desfavoráveis ou negativas. Cores na região do verde (amarelo, azul esverdeado e verde) refletem contribuições favoráveis ou positivas. Átomos com contribuições intermediárias são representados em branco. Os fragmentos mais importantes para os compostos **67** e **43**, os derivados apresentando a maior afinidade e a maior potência, respectivamente, podem ser visualizados em seus mapas de contribuições individuais (Figura 25).

O fragmento molecular representado pelo substituinte *para*-fenol é fortemente associado aos parâmetros farmacológicos potência e afinidade desses compostos. O mesmo pode ser sugerido para o nitrogênio protonado nos anéis pirrolidina e para os dois carbonos ligados a esse átomo de nitrogênio. Esses resultados concordam com as interações descritas para o ER, e serão discutidas no próximo tópico. Adicionalmente, para quaisquer moléculas dos conjuntos de dados, regiões com contribuições negativas ou intermediárias podem ser identificadas como alvos potenciais de modificação para a síntese e os estudos de SAR.

Figura 25. Mapas de contribuição de HQSAR. A. Modulador com maior potência (**43**). B. modulador com maior afinidade (**67**).

3 QSAR 3D CoMFA

No método CoMFA, são estabelecidas as relações entre os valores da propriedade biológica e as propriedades estereoquímicas e eletrostáticas de uma série de compostos. Essa estratégia se baseia na premissa de que as alterações na afinidade de ligação estão relacionadas com as propriedades moleculares representadas pelos campos moleculares.^{83,84} O alinhamento é uma variável crucial e afeta de forma direta os resultados das análises 3D. O alinhamento molecular tridimensional deve representar as conformações de interação adotadas pelas moléculas do conjunto treinamento no interior da cavidade de interação do receptor alvo. Para o modulador **33** (Tabela 3) em complexo com a estrutura utilizada nos estudos de acoplamento (código PDB: 1XP1), as interações no interior da cavidade são semelhantes às descritas para o raloxifeno (Figura 26).

Figura 26. Complementaridade multifuncional entre o modulador **33** e hER α (código PDB: 1XP1).

No programa GOLD 3.0, utilizado nos estudos de docagem molecular na cavidade de interação do hER α para a obtenção dos modelos de conformações bioativas, uma série de parâmetros podem ser ajustados pelo usuário. Dentre os parâmetros mais importantes se destacam a quantidade de posições de acoplamento produzida para cada ligante, a definição da região utilizada para o posicionamento das diferentes moléculas e a função de pontuação utilizada pelo programa. A região a ser considerada para o acoplamento das moléculas (uma esfera) pode ser definida a partir de um ponto com coordenadas cartesianas definidas, ou de átomo de um ligante ou, ainda, de um resíduo de aminoácido da proteína. Uma vez definida a região da cavidade a ser considerada no processo de acoplamento, as moléculas do conjunto de dados só poderão ser alocadas neste espaço disponível no interior da esfera.¹⁰³ Para o processo de acoplamento dos moduladores do conjunto de dados 1, a cavidade de interação foi definida com um raio de 10 Å ao redor das coordenadas cartesianas do átomo central (C22) do ligante. Apenas as três melhores posições para cada modulador foram consideradas e as moléculas foram classificadas de acordo com a função *GoldScore* (padrão do programa).

Outro parâmetro importante se refere às restrições que podem ser impostas no processo de acoplamento, podendo favorecer certas interações que devem ocorrer entre a proteína e os ligantes, como por exemplo, ligações hidrogênio e interações hidrofóbicas. Para

evitar o posicionamento invertido dos sistemas de anéis benzoxatiina, benzoditiina e flavanóide no interior do sítio hidrofóbico do estradiol, as interações de grupos hidroxila ligados a esses anéis, direta ou indiretamente, foram fixadas em relação aos resíduos de aminoácidos Glu353, Arg394 e His524, de acordo com o observado para a hipótese farmacofórica (Figura 26). De acordo com as restrições, o sistema de anéis fundidos é posicionado do mesmo lado que os resíduos de Arg394 e Glu353.

Após o processo de alinhamento, foram geradas 381 conformações, sendo três para cada molécula do conjunto de dados. Cada uma das conformações foi avaliada quanto à satisfação dos requisitos da hipótese farmacofórica (Figura 26). Para as moléculas contendo cadeias laterais β -etilamina ligadas por quaisquer grupos espaçadores ao sistema de anéis benzoxatiina, benzoditiina ou flavanóide, a hipótese farmacofórica só foi satisfeita (conformação adequada só foi verificada) quando os nitrogênios amino desses grupos foram modelados em sua forma protonada.

A geometria relativa das conformações obtidas mais bem classificadas foi também analisada, contudo, nenhum problema foi detectado. O processo resultou, inclusive, no posicionamento adequado do modulador **33** do conjunto de dados, com as principais interações intermoleculares semelhantes às aquelas observadas no co-complexo. Uma única conformação foi selecionada para cada molécula (Figura 27).

Figura 27. Alinhamento molecular estrutural do conjunto de dados.

O alinhamento molecular foi então submetido às análises de QSAR 3D CoMFA. Os modelos CoMFA foram desenvolvidos com base nas mesmas combinações de conjuntos treinamento e teste utilizadas na modelagem de HQSAR (item 2 deste capítulo).

Utilizando os parâmetros especificados no item 6 do capítulo 3 (campo de forças Tripos, campos eletrostáticos e estereoquímicos, valores de corte eletrostáticos e estereoquímicos de 30 Kcal/mol, grade com espaçamento de 2 Å entre os pontos e átomo de carbono sp^3 como sonda, com carga +1), os descritores foram obtidos. O método de validação cruzada LOO para um número máximo de 8 componentes foi utilizado para estabelecer o número ótimo de componentes principais para o modelo de QSAR, com os valores das energias de interação escalonados pelo método CoMFA padrão. Modelos com parâmetros estatísticos insatisfatórios foram inicialmente obtidos, como apresentados na Tabela 10. Na tentativa de obter modelos com maior capacidade preditiva, foi empregado o recurso de focagem da melhor região (do inglês, QSAR CoMFA Region Focus).

Tabela 10. Resultados dos modelos iniciais CoMFA construídos.

Conjunto de Dados	Num	q^2	N	r^2	SEE	F	Fração	
							S	E
1	127	0,51	8	0,86	0,30	250,7	0,58	0,42
2	69	0,53	8	0,88	0,25	256,4	0,51	0,49

Conjunto de dados 1, afinidade; conjunto de dados 2, potência; Num, número de compostos do conjunto de dados original, q^2 : coeficiente de correlação com validação cruzada (LOO); N: número ótimo de componentes; r^2 : coeficiente de correlação; SEE: erro padrão de estimativa; F: valor do teste F; S: contribuição estereoquímica; E: contribuição eletrostática.

Cada análise de PLS associa uma série de parâmetros a cada vértice de cada campo utilizado. O escalonamento CoMFA padrão atribui, inicialmente, o mesmo peso para todos esses parâmetros. Assim, muitos descritores gerados em vários pontos da grade reticulada não contribuem de forma positiva para o modelo e são considerados pontos de redundância ou baixa variância. Como muitos desses pontos são incluídos na análise, geralmente os modelos gerados inicialmente possuem menor qualidade. O recurso da focagem da melhor região consiste em atribuir um peso maior para aqueles pontos da grade que contribuem de maneira mais significativa para o modelo e menor peso para pontos que apresentem redundância ou baixa variância. Vários parâmetros podem ser selecionados para evidenciar os campos apropriados, incluindo o método pelo qual os campos terão seus pesos diferenciados e a magnitude dessa diferenciação (de 0,3 a 1,5 geralmente). Um desses métodos é a diferenciação dos campos pelo coeficiente de desvio padrão (SDC, do inglês, Standard Deviation Coefficient), que a partir de uma análise previamente realizada, atribui o peso que determinado ponto terá em uma nova análise, multiplicando a variação dos valores dos campos de interação deste ponto pelo coeficiente que foi atribuído a este ponto pela análise de PLS anterior. Além disso, o espaçamento da grade reticulada pode ser modificado na nova região criada. O espaçamento padrão inicial é de 2 Å e pode ser modificado para valores menores, com aumento do número de descritores gerados, podendo aumentar o ruído em alguns casos.¹⁰⁷

Os modelos pouco satisfatórios obtidos inicialmente, foram submetidos aos processos de focagem da melhor região e demonstraram parâmetros estatísticos significativos quando os descritores importantes foram potencializados pelo SDC com peso de 0,8. A diminuição do espaçamento entre pontos da grade não resultou em melhora alguma na consistência interna dos modelos. Os resultados estão mostrados na Tabela 11. A estratégia de focagem da melhor região não apenas aumenta os valores dos parâmetros estatísticos durante o processo de geração do modelo, mas também resulta no refinamento dos mapas de contorno 3D. A possibilidade de geração de modelos aleatórios foi verificada pela permutação dos valores da propriedade biológica entre os compostos e a estabilidade dos modelos foi confirmada.

Tabela 11. Resultados dos modelos finais CoMFA construídos.

Conjunto de Dados	Num	q^2	N	r^2	SEE	F	Fração	
							S	E
1	127	0,79	7	0,93	0,26	161	0,74	0,26
2	69	0,72	7	0,94	0,24	105	0,66	0,34

Conjunto de dados 1, afinidade; conjunto de dados 2, potência; Num, número de compostos do conjunto de dados original, q^2 : coeficiente de correlação com validação cruzada (LOO); N: número ótimo de componentes; r^2 : coeficiente de correlação; SEE: erro padrão de estimativa; F: valor do teste F; S: contribuição estereoquímica; E: contribuição eletrostática.

Os modelos finais gerados foram submetidos à validação externa utilizando o conjunto teste, cujos compostos não haviam sido incluídos no processo de geração dos modelos. Os compostos do conjunto teste foram submetidos à minimização de energia e ao alinhamento molecular de forma idêntica às moléculas pertencentes ao conjunto treinamento. Os resultados do processo de validação externa são apresentados nas Tabelas 12 e 13, para afinidade de ligação e potência, respectivamente. Como pode ser observado, os valores de pIC_{50} preditos se aproximam dos experimentais, desviando não mais que 0,42 unidades logarítmicas. A pequena diferença entre os valores experimentais e preditos indica uma elevada consistência externa do modelo, com boa capacidade de predição. Os modelos podem, então, ser usados para o planejamento de novos moduladores com propriedades farmacológicas otimizadas.

Tabela 12. Valores preditos e experimentais de afinidade de ligação (pIC_{50}), associados aos resíduos, para os 28 compostos do conjunto teste.

Código	pIC_{50}	pIC_{50}	Resíduo
	Experimental	Predito	
43	9,10	9,21	-0,11
44	8,77	8,60	0,17
45	8,36	8,46	-0,1
46	9,22	9,56	-0,34
47	9,22	9,24	-0,02
48	9,30	9,41	-0,11
49	8,66	8,65	0,01
50	8,96	8,64	0,32
51	9,00	8,83	0,17
52	8,89	8,90	-0,01
53	9,16	9,01	0,15
54	9,30	9,16	0,14
55	8,60	8,86	-0,26
56	8,85	9,14	-0,29
57	9,16	9,19	-0,03
106	8,27	8,59	-0,32
116	6,28	5,97	0,31
117	6,98	6,77	0,21
118	6,83	6,99	-0,16
119	5,98	5,94	0,04
120	8,51	8,93	-0,42
121	8,13	8,07	0,06
122	7,46	7,29	0,17
123	8,16	8,55	-0,39
124	7,96	8,00	-0,04
125	8,52	8,64	-0,12
126	8,41	8,66	-0,25
127	8,70	8,88	-0,18

Tabela 13. Valores preditos e experimentais de potência (pIC_{50}), associados aos resíduos, para os 14 compostos do conjunto teste.

Código	pIC_{50} Experimental	pIC_{50} Predito	Resíduo
56	8,17	8,32	-0,15
57	9,30	9,05	0,25
58	8,47	8,41	0,06
59	7,55	7,64	-0,09
60	6,91	6,89	0,02
61	6,95	6,94	0,01
62	8,59	8,50	0,09
63	7,95	7,91	0,04
64	9,05	8,88	0,17
65	9,00	8,84	0,16
66	8,80	8,91	-0,11
67	8,42	8,64	-0,22
68	9,40	9,43	-0,03
69	8,51	8,36	0,15

A análise gráfica dos valores experimentais e preditos para os conjuntos treinamento e teste dos dois modelos confirma a boa concordância desses valores (Figuras 28 e 29). O processo de validação externa comprova que este modelo é capaz de prever a propriedade biológica de novas moléculas dentro desta classe química, que é representada pela diversidade contida no espaço químico explorado.

Figura 28. Gráfico CoMFA de afinidade de ligação dos valores experimentais e preditos para as 127 moléculas do conjunto de dados.

Figura 29. Gráfico CoMFA de potência dos valores experimentais e preditos para as 69 moléculas do conjunto de dados.

Como pode ser verificado na Tabela 11, o descritor de campos estereoquímicos para o conjunto de dados 1 explica 74% da variância total dos dados, enquanto o descritor de campos eletrostáticos é responsável pelos 26% adicionais. Para o conjunto de dados 2, os campos estereoquímicos e eletrostáticos explicam 66% e 34%, respectivamente. Como esperado, os descritores de campos estereoquímicos foram mais importantes para a explicação da variância total nos modelos de afinidade e potência, uma vez a cavidade de interação com ligantes do ER apresenta características hidrofóbicas. Assim, é de se esperar que essas características sejam importantes na interação de ligantes seletivos, como confirmado pelo esqueleto estrutural com características hidrofóbicas do estradiol, o ligante natural endógeno.

Os mapas de contorno 3D CoMFA permitem a visualização dos campos estereoquímicos e eletrostáticos que apresentam maior importância para explicar os diferentes valores das propriedades (afinidade e potência) apresentados pelas moléculas do conjunto de dados. Os mapas gerados a partir das análises de PLS mostram onde a variabilidade nos campos das moléculas explica as diferenças nas propriedades alvo (Figura 30). O entendimento destes campos é muito útil nos trabalhos em química medicinal, com ênfase nos estudos de SAR e síntese planejada.

As representações dos campos 3D com cores verde e amarelo referem-se a características estereoquímicas, ao passo que os campos 3D com cores azul e vermelho se referem a características eletrostáticas. Poliedros verdes em torno de determinadas regiões da molécula, sugerem substituições com grupos mais volumosos, ao passo que poliedros amarelos sugerem grupos menos volumosos. A presença de poliedros azuis em determinada região da molécula sugere substituições por grupos com maior carga positiva, ao passo que poliedros vermelhos sugerem grupos mais eletronegativos.^{83,84,107}

Figura 30. Mapas de contorno 3D CoMFA eletrostáticos e estereoquímicos (contorno 80:20 padrão). A. modulador de maior afinidade (**67**). B. modulador mais potente (**43**).

Não há diferenças significativas entre os mapas de contorno eletrostáticos para os modelos de potência e afinidade. Para ambos os modelos, o átomo de nitrogênio protonado do grupo pirrolidina está relacionado ao aumento da afinidade e da potência. Os descritores eletrostáticos nessa posição, entretanto, são muito mais importantes para diferenciar os compostos em relação à potência, uma vez que a medida da variabilidade nos campos das moléculas que explica as diferenças dos valores de potência é maior do que a medida para afinidade.

Os mapas de contorno estereoquímicos são também semelhantes para os dois modelos. Substituintes com menor impedimento estérico ligados ao anel benzoxatiina são positivamente relacionados a ambos os parâmetros biológicos. Além disso, a presença de grupos substituintes volumosos ligados ao átomo enxofre do mesmo anel está relacionado à diminuição da potência e da afinidade. A presença de substituintes nas posições *para* ou *meta* no anel fenila é considerada uma contribuição positiva para afinidade, mas não para potência, como indicado pelo mapa de contorno verde ao redor do grupo fenila. Ainda de acordo com os mapas 3D estereoquímicos, há espaço adicional para grupos mais volumosos ligados ao anel

pirrolidina, os quais poderiam levar ao aumento da potência e afinidade. Essas posições de substituição são especialmente atrativas para modificação molecular e estudos posteriores de SAR. Os modelos de QSAR gerados são compatíveis com o ambiente químico tridimensional na cavidade de interação do ER (Figura 31).

Figura 31. Representação esquemática do modo de ligação do modulador apresentando mais alta afinidade no interior da estrutura cristalográfica de hER α .

A importância dos substituintes nos anéis benzoxatiina e fenila corrobora com as interações estabelecidas com os resíduos His524, Arg394 e Glu353, no interior da cavidade de interação do hER α (Figuras 31 e 32).^{32,33,104,105} Substituintes volumosos ligados ao átomo de enxofre do anel benzoxatiina (ex. modulador **86**) estão relacionados à diminuição da afinidade em razão do impedimento estérico provocado pelas cadeias laterais dos resíduos Phe404 e Leu 346 (Figura 33).

Figura 32. Mapa de contorno estereoquímico (modulador **67** na cavidade de interação do hER α).

Figura 33. Impedimento estérico ao posicionamento de grupos volumosos ligados ao átomo de enxofre do anel benzoxatiina (modulador **86** na cavidade de interação do hER α).

Os mapas tridimensionais também indicam que o átomo de nitrogênio pirrolidínico é importante para a característica antagonista, o que está de acordo às estruturas cristalográficas do receptor em complexo com moduladores antagonistas. Além disso, alguns estudos indicam que as interações entre o N da heterociclo e o resíduo Asp351 são responsáveis pela atividade antiestrogênica (Figura 34).¹⁰⁹⁻¹¹²

Figura 34. Mapa de contorno eletrostático (modulador **67** na cavidade de interação do hERα).

Em concordância com os estudos de HQSAR, a importância dos substituintes nos anéis fenol e benzoxatiina, e do nitrogênio pirrolidínico foi verificada. Além disso, as posições de acoplamento, e os mapas de contorno eletrostáticos e de contribuição 2D, demonstram que, para uma série homóloga na cadeia lateral *aminoalquinoxifenila*, 2 é o número ótimo de carbonos espaçadores entre o grupo fenol e o átomo nitrogênio pirrolidínico.⁹⁶

Esses resultados podem, ainda, ser interpretados de acordo com os conceitos de afinidade e potência para receptores nucleares. Enquanto a afinidade de ligação é dependente apenas da capacidade do modulador de competir com o estradiol pela cavidade de interação do hERα, a potência está relacionada com a capacidade do modulador de inibir a multiplicação das células de câncer de mama. A afinidade de ligação é função do tipo, do número e da

orientação das diversas interações intermoleculares na cavidade de interação e da perda de entropia conformacional por ligações rotacionáveis na molécula, ao assumir a conformação preferencial de interação.¹¹³ Os resultados das análises 3D, para o modelo de afinidade, sugerem que moléculas possuindo grupos de volume adequado, em posições determinadas nos anéis fenol e benzoxatiina, possuirão maior afinidade de ligação. Além disso, a existência de uma interação eletrostática, com o nitrogênio pirrolidínico protonado, também favorece o aumento da afinidade.

A potência, entretanto, é também função da capacidade do modulador em impedir as alterações conformacionais que levem à transcrição de genes relacionados à multiplicação celular. De acordo com as análises 3D para o modelo de potência, os campos eletrostáticos localizados na região do anel pirrolidínico, na cadeia lateral antiestrogênica, são os principais responsáveis pelas diferenças nos valores das potências dos antagonistas.^{32,109-112} Os descritores eletrostáticos nessa posição são, inclusive, muito mais importantes para diferenciar os compostos em relação à potência do que em relação à afinidade.

Os modelos podem ser utilizados no planejamento baseado na estrutura do receptor e são úteis no planejamento de novos antagonistas estruturalmente relacionados.¹¹⁴

PARTE B

DETERMINANTES QUÍMICOS E ESTRUTURAIS DA AFINIDADE E SELETIVIDADE DE MODULADORES DOS SUBTIPOS DO ER

Os dois subtipos do ER são originados de genes separados e apresentam diferentes perfis farmacológicos, padrões de expressão, distribuição tecidual e funções biológicas.^{62,63,66,67,118} Ao longo dos últimos anos, vários experimentos têm sido realizados na tentativa de definir suas funções. Enquanto a importância de hER α nos tecidos reprodutivos e seu impacto na etiologia do câncer de mama é bem estabelecida, a função de hER β permanece incerta.^{34-36,119}

Como comentado no Capítulo I dessa dissertação, estudos epidemiológicos indicam que o consumo de isoflavonas está associado a efeitos positivos nas desordens da menopausa. Alguns estudos demonstram que esses efeitos são devidos à interação seletiva desses ligantes de origem natural com hER β , e também ao recrutamento de proteínas co-regulatórias que disparam mecanismo transcricionais mediados por esse subtipo do ER.^{55,65,119} Essas evidências indicam que o hER β é um alvo atrativo para o desenvolvimento de agonistas seletivos, os quais poderiam manter os efeitos benéficos dos estrógenos, sem entretanto aumentar o risco de complicações nos tecidos reprodutivos e nas mamas, já que hER β não é o receptor principal expresso nesses locais.^{48-50,62,66,120}

A cavidade de interação com ligantes é muito similar nos dois subtipos do ER, diferindo por apenas dois resíduos de aminoácidos, com mudanças que são bastante conservativas: Leu384 de hER α é substituída por Met336 em hER β e Met421 de hER α é substituída por Ile373 em hER β (Figura 35).⁶²⁻⁶⁵ Entretanto, o surgimento de moduladores subtipo-seletivos indica que essa variação sutil entre os receptores hER α e hER β é suficiente para propiciar oportunidades de planejamento de novos ligantes subtipo-seletivos.^{119,121,122}

Figura 35. Diferenças nas cavidades de interação com ligantes de hER α (verde) e hER β (magenta), em complexos com a genisteína.

1 CONJUNTOS PADRÕES DE DADOS

O conjunto de dados empregado no desenvolvimento dos modelos de QSAR 2D e 3D possui 81 moduladores para os quais as medidas de afinidade de ligação (IC_{50}) foram coletadas para hER α e hER β .^{100,101} A diversidade química do conjunto de dados é representada por fenilnaftalenos e fenilquinolinas, os quais foram planejados para mimetizar o esqueleto estrutural da genisteína, um modulador hER β -seletivo (40 vezes), de forma a gerar ligantes com maiores afinidade e seletividade.¹⁰⁰ As estruturas químicas dos moduladores e os valores das propriedades biológicas associadas estão apresentados na Tabela 14.

Os valores de afinidade de ligação variam de 0,001 a 5 μ M para hER β (um fator de 4.500), e de 0,01 a 6,6 μ M para hER α (um fator de 660), e foram determinados nas mesmas condições experimentais. Esse conjunto de dados foi selecionado porque fornece uma diversidade estrutural apropriada para a modelagem de QSAR e porque os valores experimentais foram determinados de forma padrão para ambos os subtipos do hER.

Tabela 14. Estruturas químicas e valores correspondentes de afinidade de ligação para hER α e hER β .

Código	Estrutura	β pIC ₅₀	α pIC ₅₀	Código	Estrutura	β pIC ₅₀	α pIC ₅₀
<i>Conjunto treinamento</i>							
1		8,96	7,40	9		8,66	7,35
2		8,96	7,14	10		8,64	7,52
3		8,92	7,14	11		8,64	6,98
4		8,85	8,00	12		8,62	6,63
5		8,82	7,68	13		8,60	6,95
6		8,80	7,66	14		8,60	7,04
7		8,70	7,08	15		8,57	6,68
8		8,68	7,02	16		8,55	7,57

Código	Estrutura	βpIC_{50}	αpIC_{50}	Código	Estrutura	βpIC_{50}	αpIC_{50}
17		8,48	7,44	25		8,28	6,61
18		8,47	7,46	26		8,26	6,74
19		8,47	6,55	27		8,22	6,39
20		8,40	6,84	28		8,16	6,73
21		8,37	6,67	29		8,15	7,11
22		8,35	6,94	30		8,14	6,28
23		8,34	6,67	31		8,08	7,04
24		8,30	7,04	32		8,07	6,93

Código	Estrutura	βpIC_{50}	αpIC_{50}	Código	Estrutura	βpIC_{50}	αpIC_{50}
33		8,00	7,40	40		7,77	6,55
34		8,00	6,46	41		7,64	5,98
35		7,97	6,28	42		7,52	6,20
36		7,96	6,52	43		7,49	6,45
37		7,92	6,88	44		7,44	6,11
38		7,89	6,58	45		7,36	6,41
39		7,79	6,68	46		7,28	6,20

Código	Estrutura	βpIC_{50}	αpIC_{50}	Código	Estrutura	βpIC_{50}	αpIC_{50}
47		7,24	6,26	54		6,70	6,15
48		7,22	6,30	55		6,69	5,84
49		7,12	5,82	56		6,68	5,74
50		7,10	5,76	57		6,66	5,50
51		7,06	5,94	58		6,61	6,20
52		6,94	6,05	59		6,43	5,87
53		6,77	5,75	60		6,25	5,58

Código	Estrutura	β plC ₅₀	α plC ₅₀	Código	Estrutura	β plC ₅₀	α plC ₅₀
61		6,13	5,91	64		5,70	5,30
62		6,12	5,32	65		5,30	5,30
63		5,92	5,18				

Código	Estrutura	β plC ₅₀	α plC ₅₀	Código	Estrutura	β plC ₅₀	α plC ₅₀
<i>Conjunto teste</i>							
66		8,92	7,24	74		8,22	6,96
67		8,70	7,62	75		8,20	6,61
68		8,64	7,99	76		7,96	6,97
69		8,60	6,95	77		7,89	7,40
70		8,52	7,01	78		7,57	6,76
71		8,47	6,64	79		7,55	6,34
72		8,36	6,60	80		7,03	5,47
73		8,25	6,51	81		6,28	5,47

Como a geração de modelos confiáveis é dependente da criação de conjuntos treinamento e teste com apropriada diversidade e distribuição da propriedade alvo, a divisão dos moduladores do conjunto de dados original foi orientada por HCA. Do total de 81 moduladores do conjunto original, 65 (80%) foram selecionados no conjunto treinamento para a construção do modelo, enquanto os outros 16 (20%) foram incluídos no conjunto teste para a validação externa. Os conjuntos de treinamento e teste foram selecionados de forma que moléculas estruturalmente diversas, possuindo valores das propriedades biológicas em toda a faixa da distribuição, fossem incluídas em ambos os conjuntos. Os mesmos conjuntos treinamento e teste foram empregados para todas as análises, para que os resultados pudessem ser comparados (Figuras 36 e 37). Os valores de afinidade de ligação foram convertidos nos respectivos valores de pIC_{50} e usados como variáveis dependentes nas investigações. O método PLS foi utilizado em todas as análises e a capacidade preditiva dos modelos foi avaliada pelos valores de q^2 e pelas predições do conjunto teste.

Figura 36. Representação gráfica da distribuição dos valores de pIC_{50} para hER β .

Figura 37. Representação gráfica da distribuição dos valores de pIC_{50} para hER α .

2 HOLOGRAMA QSAR

A geração de fragmentos moleculares foi realizada utilizando várias combinações dos parâmetros de distinção, para o tamanho de fragmento padrão (4-7): A/B, A/B/C, A/B/C/DA, A/B/C/H/DA e A/B/H/DA. As análises de HQSAR foram realizadas considerando os 12 comprimentos de holograma disponíveis entre 53 e 401. Os resultados estatísticos a partir das análises de PLS para os 65 compostos do conjunto treinamento utilizando várias combinações de distinção de fragmento estão apresentados nas Tabelas 15 e 16.

Tabela 15. Modelo de HQSAR de $hER\beta$ para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmento (4-7).

Distinção de Fragmentos	Parâmetros Estatísticos				
	$hER\beta$				
	q^2	r^2	SEE	HL	N
A/B	0,62	0,91	0,31	151	8
A/B/C	0,53	0,89	0,32	151	7
A/B/C/DA	0,57	0,94	0,24	353	7
A/B/C/H/DA	0,67	0,93	0,26	353	6
A/B/H/DA	0,73	0,91	0,28	199	6

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes. Distinção de fragmento: A, átomo; B, ligação; C, conectividade; H, hidrogênio; DA, doador eceptor.

Tabela 16. Modelo de HQSAR de $hER\alpha$ para várias combinações de distinção fragmentos utilizando o tamanho padrão de fragmento (4-7).

Distinção de Fragmentos	Parâmetros Estatísticos				
	$hER\alpha$				
	q^2	r^2	SEE	HL	N
A/B	0,59	0,86	0,27	83	7
A/B/C	0,58	0,88	0,24	151	6
A/B/C/DA	0,59	0,88	0,24	353	5
A/B/C/H/DA	0,67	0,89	0,23	307	5
A/B/H/DA	0,71	0,90	0,23	199	6

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes. Distinção de fragmento: A, átomo; B, ligação; C, conectividade; H, hidrogênio; DA, doador eceptor.

Os melhores resultados estatísticos para os dois modelos gerados foram obtidos utilizando A/B/H/DA como informação para distinção de fragmentos (hER β : $q^2 = 0,73$ e $r^2 = 0,91$, com SEE de 0,28; e, hER α : $q^2 = 0,71$ e $r^2 = 0,90$, com SEE de 0,23). Os resultados estatísticos para as análises individuais dos diferentes tamanhos de fragmentos avaliados (2–5, 3–6, 4–7, 5–8 e 6–9) não levaram a diferenças significativas e não resultaram em melhorias nos parâmetros estatísticos nestas séries comparativas (resultados não mostrados). Vale ser ressaltado que nos melhores modelos de HQSAR, o parâmetro DA foi utilizado, fato que está em concordância com estudos anteriores que indicam a importância de interações intermoleculares não-ligantes com resíduos específicos da estrutura da proteína para a atividade biológica.^{32,33,104,105,114}

O poder preditivo para os modelos para hER β e hER α foi verificado pela predição dos valores de pIC₅₀ para as moléculas da série teste, que foram completamente excluídas durante a derivação do modelo. Os resultados do processo de validação externa são apresentados nas Tabelas 17 e 18.

Tabela 17. Valores preditos e experimentais de afinidade de ligação (β pIC₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.

Código	β pIC ₅₀ Experimental	β pIC ₅₀ Predito	Resíduo
66	6,28	6,27	0,01
67	7,03	6,95	0,08
68	7,55	7,28	0,27
69	7,57	7,03	0,54
70	7,87	8,02	-0,15
71	7,96	8,02	-0,06
72	8,20	8,08	0,12
73	8,22	8,12	0,1
74	8,25	8,47	-0,22
75	8,36	8,30	0,06
76	8,47	8,25	0,22
77	8,52	8,78	-0,26
78	8,60	8,76	-0,16
79	8,64	8,51	0,13
80	8,70	8,43	0,27
81	8,92	8,40	0,52
Genisteína	8,01	7,72	0,19

Tabela 18. Valores preditos e experimentais de afinidade de ligação (α pIC₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.

Código	α pIC ₅₀ Experimental	α pIC ₅₀ Predito	Resíduo
66	5,47	5,89	-0,42
67	5,47	5,89	-0,42
68	6,34	5,96	0,38
69	6,76	6,39	0,37
70	7,40	7,45	-0,05
71	6,97	6,88	0,09
72	6,61	6,61	0
73	6,96	6,80	0,16
74	6,51	6,90	-0,39
75	6,60	6,53	0,07
76	6,64	6,65	-0,01
77	7,00	7,23	-0,23
78	6,95	6,73	0,22
79	7,99	7,61	0,38
80	7,62	7,39	0,23
81	7,41	7,30	0,11
Genisteína	6,40	6,36	0,04

As pequenas diferenças entre os valores experimentais e preditos indicam a boa capacidade preditiva dos modelos de HQSAR gerados. Os valores preditos se aproximam dos experimentais, desviando em não mais que 0,55 unidades logarítmicas em ambos os modelos. Além de possuírem capacidade de prever a afinidade de ligação desta série de compostos, os modelos apresentam também substancial capacidade de prever a seletividade dos compostos. A genisteína, por exemplo, apresenta valores preditos e experimentais muito próximos (resíduos modestos) para ambos $hER\beta$ e $hER\alpha$, comprovando a capacidade do modelo no estudo da seletividade.

A análise gráfica dos valores experimentais e preditos para os conjuntos treinamento e teste (para os dois modelos) confirma a boa concordância desses valores (Figuras 38 e 39). Como os resíduos são baixos, os modelos de HQSAR obtidos são altamente confiáveis e podem ser utilizados para prever os valores de afinidade de ligação para novos compostos estruturalmente relacionados.

Figura 38. Gráfico HQSAR dos valores experimentais e preditos para as 81 moléculas do conjunto de dados (modelo para $hER\beta$).

Figura 39. Gráfico HQSAR dos valores experimentais e preditos para as 81 moléculas do conjunto de dados (modelo para hER α).

A máxima subestrutura comum apresentada por todos os moduladores do conjunto de dados está representada em azul claro nos mapas de contribuição de HQSAR (Figura 40, modulador **2** de maior afinidade para o receptor hER β). Os mapas 2D indicam que o substituinte hidroxila fenólico na posição 2 (no anel naftaleno) está fortemente relacionado à afinidade desse composto para ambos os subtipos do ER. Essas observações estão de acordo com estudos anteriores, os quais indicam a importância de substituintes volumosos e eletronegativos nessa posição para ambos os subtipos do ER e também corroboram com as interações descritas para hER α (His524, Arg394 e Glu353) e hER β (His 475, Arg346 e Glu305).^{32,100,114-117} Além disso, átomos de hidrogênio são os substituintes preferidos nas posições *orto* à hidroxila do anel fenila. A contribuição do átomo de hidrogênio na posição 5 (no anel naftaleno) evidencia a preferência pelo anel naftaleno, em detrimento do anel quinolina. De fato, no conjunto de dados, os fenilnaftalenos possuem maior afinidade que as fenilquinolinas (ex. moduladores 10 e 23). Para quaisquer moléculas dos conjuntos de dados, regiões com contribuições negativas ou intermediárias podem ser identificadas como alvos potenciais de modificação para a síntese e os estudos de SAR.¹²³

Figura 40. Mapa de contribuição de HQSAR para o modulador 1 A) modelo para hER β e B) modelo para hER α .

Embora os mapas de contribuição sejam capazes de evidenciar contribuições importantes relacionadas à afinidade de ligação, eles não fornecem informações adicionais em relação à seletividade dentro desta série de moduladores. Isso se deve muito provavelmente à predominância de requerimentos tridimensionais nas interações entre os ligantes e os subtipos do ER, que não são suficientemente caracterizados pela análise de fragmentos moleculares 2D.

3 QSAR 3D CoMFA

Os protocolos de acoplamento implementados no programa GOLD 3.0 foram empregados para identificar possíveis conformações de interação dos moduladores na cavidade dos ERs. Os moduladores desse conjunto de dados foram planejados a partir do esqueleto estrutural da genisteína, um agonista modestamente hER β -seletivo (40 vezes).^{100,101} Portanto, na ausência de estruturas cristalográficas em complexos com moduladores do conjunto de dados, estruturas de hER α e hER β em complexos com a genisteína foram selecionadas para os procedimentos de acoplamento (códigos PDB 1QKM para hER β e 1X7R for hER α). Para os cálculos, o ligante genisteína e as moléculas de água foram removidos das estruturas, sendo os átomos de hidrogênio, em contrapartida, adicionados. Para ambos os subtipos, as cavidades de interação foram centralizadas no átomo C7 da genisteína e um raio de 10 Å foi considerado nos procedimentos de posicionamento dos moduladores do conjunto

de dados. Nenhuma restrição foi imposta e os modelos de conformações bioativas obtidos foram utilizados para o alinhamento estrutural do conjunto de dados. Apenas as conformações mais bem pontuadas foram consideradas para os estudos de QSAR 3D.

Os alinhamentos 3D obtidos pelo acoplamento nos dois subtipos do ER são diferentes para o mesmo conjunto de moduladores (Figura 41). Embora os resíduos de aminoácidos importantes para o ancoramento correto dos moduladores nas cavidades de interação sejam conservados nos dois subtipos, as substituições de resíduos de aminoácidos nas cavidades de interação dos dois subtipos do ER provocam diferenças nas interações com os moduladores, o que resulta em diferentes modelos de conformações bioativas. Uma vez que os cálculos dos campos moleculares de interação (eletrostáticos e estereoquímicos) dependem do alinhamento estrutural do conjunto de dados, os diferentes posicionamentos tridimensionais dessas moléculas explicam os diferentes valores de afinidade de ligação para os subtipos do ER, que conseqüentemente, levam a seletividade. Portanto, para os estudos de seletividade, a obtenção de alinhamentos estruturais adequados, considerando-se os mesmos moduladores do conjunto treinamento, é crucial para a posterior comparação dos modelos 3D.

Figura 41. Diferentes tipos de alinhamento estrutural do conjunto de dados (hERα à esquerda e hERβ à direita).

Os modelos CoMFA inicialmente obtidos foram otimizados pelo procedimento de focagem da melhor região, resultando no aumento da consistência interna dos modelos e no refinamento dos mapas de contorno 3D. A estabilidade dos modelos foi testada pelo método da

permutação dos valores da propriedade biológica entre os compostos e a sua consistência interna confirmada. Parâmetros estatísticos significativos foram obtidos para os modelos otimizados, como apresentados na Tabela 19.

Tabela 19. Resultados dos modelos finais CoMFA construídos.

Modelo	Subtipo	q^2	N	r^2	SEE	F	Fração	
							S	E
1	hER α	0,76	8	0,94	0,17	118,9	0,49	0,51
2	hER β	0,73	8	0,96	0,19	184,9	0,41	0,59

q^2 : coeficiente de correlação com validação cruzada (LOO); N: número ótimo de componentes; r^2 : coeficiente de correlação; SEE: erro padrão de estimativa; F: valor do teste F; S: contribuição estereoquímica; E: contribuição eletrostática.

Os resultados da validação externa dos modelos, utilizando os 16 moduladores do conjunto teste e a genisteína, são apresentados nas Tabelas 20 e 21.

Tabela 20. Valores preditos e experimentais de afinidade de ligação (α plC₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.

Código	α plC ₅₀ Experimental	α plC ₅₀ Predito	Resíduo
66	5,47	6,12	-0,65
67	5,47	5,65	-0,18
68	6,34	6,39	-0,05
69	6,76	6,78	-0,02
70	7,40	7,19	0,21
71	6,97	7,1	-0,13
72	6,61	6,84	-0,23
73	6,96	6,6	0,36
74	6,51	6,65	-0,14
75	6,60	6,31	0,29
76	6,64	6,52	0,12
77	7,00	7,46	-0,46
78	6,95	6,99	-0,04
79	7,99	7,8	0,19
80	7,62	7,2	0,42
81	7,23	6,73	0,50
Genisteína	6,40	6,55	-0,15

Tabela 21. Valores preditos e experimentais de afinidade de ligação (β plC₅₀), associados aos resíduos, para os 16 compostos do conjunto teste.

Código	β plC ₅₀ Experimental	β plC ₅₀ Predito	Resíduo
66	6,28	6,32	-0,04
67	7,03	7,43	-0,40
68	7,55	7,18	0,37
69	7,57	7,86	-0,29
70	7,87	8,03	-0,16
71	7,96	8,25	-0,29
72	8,20	8,10	0,10
73	8,22	8,23	-0,01
74	8,25	8,16	0,09
75	8,36	8,15	0,21
76	8,47	8,94	-0,47
77	8,52	8,63	-0,11
78	8,60	8,81	-0,21
79	8,64	8,60	0,04
80	8,70	8,95	-0,25
81	8,92	8,64	0,28
Genisteína	8,01	7,77	0,24

A análise gráfica dos valores experimentais e preditos para os conjuntos treinamento e teste confirma a robustez dos modelos de QSAR 3D construídos (Figuras 42 e 43). Os valores de pIC_{50} são bem preditos, desviando não mais que 0,65 unidades logarítmicas.

Figura 42. Gráfico CoMFA do modelo de hER β dos valores experimentais e preditos para as 81 moléculas do conjunto de dados.

Figura 43. Gráfico CoMFA do modelo de hER α dos valores experimentais e preditos para as 81 moléculas do conjunto de dados.

Os descritores estereoquímicos para o modelo 1 (hER α) explicam 49% da variância total dos dados, enquanto os descritores eletrostáticos são responsáveis por 51%. Para o modelo 2 (hER β) os campos estereoquímicos compreendem 41% e os campos eletrostáticos são responsáveis por 59% da variância total. Esses resultados indicam que os campos eletrostáticos são mais importantes para diferenciar a afinidade de ligação dos moduladores para hER β do que para hER α . Os mapas de contorno 3D eletrostáticos e estereoquímicos para os modelos desenvolvidos para hER α e hER β indicam importantes características (semelhanças e diferenças), como mostrado para o modulador **1**, apresentando maior afinidade para hER β (Figura 44).

Figura 44. Mapas de contorno eletrostático (A e C) e estereoquímico (B e D) gerados nos modelos para hER α (A e B) e hER β (C e D).

A presença de substituintes nas posições 2, 4, 5 e 7 (no anel naftaleno) está associada ao aumento da afinidade para ambos os subtipos do ER. Essas são regiões em que modificações estruturais possivelmente levariam a otimização da afinidade, sem apresentar sensíveis diferenças na seletividade dos subtipos do ER. A importância dos substituintes na posição 2 é compatível com o ambiente químico tridimensional da cavidade de interação do hER α e com os resultados anteriores, que demonstram a importância das ligações de hidrogênio estabelecidas com os resíduos de His524 e His475 nos receptores alfa e beta, respectivamente.^{32,104,105,114-117}

A presença de grupos eletronegativos na posição *para* do anel fenila (posição 4') está associada ao aumento da afinidade de ligação para hER β . Esse resultado é também compatível com o ambiente tridimensional de hER β e com os resultados anteriores, que demonstram a importância das ligações de hidrogênio estabelecidas com os resíduos de Glu305 e Arg346.¹¹⁵⁻¹¹⁷ Na posição 3', no anel fenila, substituintes eletronegativos estão associados ao aumento da afinidade para hER β , mas não para hER α . A presença de substituintes eletronegativos nas posições 1 e 8 (representados pelo átomos de flúor e cloro no modulador **1**) está associada ao aumento da afinidade para os receptores hER α e hER β (Figura 44A e 44C). Para hER α , o grupo substituinte nessas posições deve ser pouco volumoso, como mostrado pelo mapa de contorno estereoquímico (Figura 44B). Esse resultado também é compatível com o ambiente tridimensional do hER α , uma vez que o posicionamento das cadeias laterais volumosas dos resíduos de Met421 e Leu384 provocam o impedimento estérico ao posicionamento de substituintes volumosos, principalmente na posição 1 (Figura 45A). Entretanto, o posicionamento das cadeias laterais dos resíduos de Ile373 e Met336 no hER β é capaz de acomodar substituintes mais volumosos na posição 1, o que é confirmado pelo mapa de contorno estereoquímico (Figuras 44D e 45B). Substituintes eletronegativos volumosos poderiam, portanto, levar ao aumento da afinidade e seletividade para hER β .

Figura 45. Impedimento estérico ao posicionamento de grupos volumosos em hER α (A). O mesmo não ocorre para hER β .

As posições 1 e 8 encontram-se, na realidade, rodeadas pelos resíduos de aminoácidos que são diferentes nos dois subtipos do receptor. Isso sugere que essas posições de substituição estão relacionadas diretamente com a seletividade (Figuras 45 e 46).

Figura 46. Cadeias laterais de resíduos substituídos em hER α (azul) e hER β (rosa) rodeando as posições 1 e 8 de substituição no anel naftaleno.

A importância dos substituintes eletronegativos, principalmente na posição 1, também pode ser explicada pelas interações com o átomo de enxofre do resíduo de Met336. A interação favorável entre a cadeia lateral do resíduo de metionina e os grupos eletronegativos, apresentando, ambos, alta densidade eletrônica, pode ser explicada levando em consideração a geometria da interação. O átomo de enxofre do resíduo de metionina se aproxima pelo lado deficiente de elétrons nos grupos polarizáveis, de tal forma que os orbitais moleculares doadores no átomo de enxofre interagem com os orbitais aceptores no grupo substituinte (Figura 47). O mesmo não ocorre, entretanto, para hER α , em que os posicionamentos das cadeias laterais dos resíduos de Met421 e Leu384, além de não permitirem tal interação, impedem a acomodação de grupos volumosos.^{64,124}

Figura 47. Posicionamento do modulador **2** na cavidade de interação de hER β .

Portanto, os modelos 3D construídos concordam com o ambiente tridimensional das cavidades de interação dos dois subtipos do hER, e evidenciam principalmente a posição 1 como uma região particularmente interessante para o planejamento de moduladores β -seletivos. Além disso, substituintes eletronegativos na posição 3', no anel fenila, também poderiam estar relacionados à otimização da afinidade apenas para hER β .

4 GRID/PCA

O estudo baseado nas estruturas 3D dos receptores visa identificar as regiões 3D que diferenciam os subtipos alfa e beta. Para isso, é necessário que as variações nos campos moleculares de interação sejam devidas apenas às diferenças entre hER α e hER β . Como as cavidades de interação dos hERs são bastantes flexíveis e se alteram em resposta a diferentes ligantes, estruturas dos LBDs dos dois subtipos do hER, em complexos com os mesmos ligantes (genisteína e WAY-244), foram selecionadas, como mostra a Tabela 22. As estruturas dos ligantes foram removidas, pois somente as interações na cavidade do alvo macromolecular foram estudadas.

Tabela 22. Características dos receptores de estrogênio humano utilizados nos estudos.

Estrutura (código PDB)	Subtipo	Ligante
		Genisteína
1X7R	hER α	
		Genisteína
1QKM	hER β	
		WAY-244
1X7E	hER α	
		WAY-244
1X78	hER β	
		ERB-041
1X7B	hER β	

Todas as estruturas foram alinhadas pelos seus carbonos-alfa em relação à estrutura de código 1QKM e apresentaram um RMSD menor do que 1,3. As energias de interação das sondas químicas OH2 e DRY com os resíduos de aminoácidos no interior da grade reticulada tridimensional, na cavidade de interação dos subtipos de ER, foram calculadas separadamente para cada sonda (Figura 48). A cada vértice da grade foram associadas energias de interação estereoquímicas e eletrostáticas, para cada sonda virtual. A energia total de interação em cada ponto é uma soma dessas energias componentes.

Figura 48. Esquema ilustrativo da definição das dimensões da grade reticulada tridimensional utilizada para os cálculos dos campos moleculares de interação.

As duas sondas foram escolhidas para mimetizar as interações intermoleculares dos grupos substituintes dos moduladores, já discutidas anteriormente, com os resíduos de aminoácidos no interior da grade 3D. A sonda OH2 representa as interações hidrofílicas e a possibilidade de formação de ligações de hidrogênio. A sonda DRY é uma sonda que representa interações hidrofóbicas e a formação de ligações de hidrogênio é detrimental para a estabilidade da interação intermolecular.⁸⁵ No programa SYBYL 7.2, os descritores gerados utilizando as duas sondas químicas foram tratados em conjunto. Para isso, os valores das energias hidrofílicas e hidrofóbicas foram escalonados. A análise quimiométrica resultou em duas componentes principais com 67% de descrição da variância total do sistema (PC1, 41%;

PC2, 26%). A primeira componente principal distinguiu os subtipos do ER, separando as estruturas tridimensionais em dois grupos (Figura 49).

Figura 49. Resultado do modelo de PCA para os subtipos do hER. As estruturas cristalográficas do hER β , 1QKM, 1X7B e 1X78 (rosa) foram posicionadas em valores positivos de PC1. Estruturas cristalográficas do hER α , 1X7R e 1X7E, foram posicionadas em valores negativos de PC1 (azul). Estruturas em complexo com a genisteína, 1QKM e 1X7R, foram posicionadas em valores positivos de PC2, enquanto que as estruturas em complexo com WAY-244, 1X7E e 1X78, foram posicionadas em valores negativos de PC2.

A primeira componente principal é responsável pela maior porcentagem de descrição da variância total do sistema estudado (41%), sendo também a componente responsável por discernir entre os subtipos do hER. Esse importante resultado sugere que as interações responsáveis pela seletividade podem ser identificadas e exploradas no planejamento de ligantes seletivos.^{87,88} Os descritores que apresentam maior importância para a diferenciação dos dois subtipos do hER podem ser convertidos em mapas de contorno, que são as regiões tridimensionais evidenciadas pelo modelo estatístico como sendo as responsáveis pelo processo de reconhecimento molecular seletivo.^{87,88} Apenas os descritores com alto poder de

discriminação foram selecionados para as análises dos mapas de contorno 3D. As regiões relacionadas às interações importantes para hER β foram coloridas em rosa e as regiões importantes para hER α em azul (Figura 50). A região rodeada pelas cadeias laterais dos resíduos substituídos nos dois subtipos do hER (hER β Met336/Leu384 hER α , hER β Ile373/Met421 hER α) é importante para diferenciá-los. Grupos volumosos e capazes de interagir eletrostaticamente com Leu336 podem estar associados à seletividade para hER β . Esse resultado é consistente com o posicionamento do grupo acetileno mostrado para ERB041, um notável modulador β -seletivo, e, em menor grau, para o posicionamento do substituinte hidroxila fenólico da genisteína, que é menos volumoso (Figura 50A e 50B). Além disso, grupos volumosos e eletronegativos próximos das regiões ocupadas pela posição *orto* dos anéis fenila (como em ERB041), parecem ser substituintes importantes para a seletividade- β .

Esses resultados são semelhantes aos obtidos para os mapas de contorno 3D CoMFA, no estudo dos fenilnaftalenos e fenilquinolinas. Para o modulador **2**, por exemplo, que possui a mesma afinidade para hER β que o modulador **1**, mas é mais seletivo, as razões para a maior seletividade podem ser explicadas pelos resultados de ambos os estudos (Figura 51). A carbonila, na posição 1 do modulador **2**, é mais volumosa e mais polarizável que o átomo de cloro, na mesma posição do modulador **1**. Além disso, o modulador **2** possui um átomo de flúor substituinte na posição 3', no anel fenila (Figura 51). Como a cavidade de interação de hER β é menor que de hER α (390 Å *versus* 490 Å), o posicionamento de substituintes eletronegativos capazes de interagir com os resíduos de Arg394 (ou Arg346) e Glu353 (ou Glu305) também pode influenciar a seletividade (Figura 50).

Figura 50. Mapas de contorno obtidos dos pesos da PC1 para as sondas A. DRY e B. OH2. No centro da cavidade está mostrado o ligante ERB041, que foi retirado do sítio durante a análise.

Figura 51. Mapas de contorno hidrofóbicos com altos pesos para PC1. Regiões importantes para a seletividade-hER β são mostradas em rosa, e para hER α em azul. No centro da cavidade está mostrado o modulador 2.

A segunda componente foi capaz de discernir as alterações conformacionais provocadas pelos diferentes ligantes (genisteína e WAY-244). Os descritores apresentando maior contribuição para essa PC representam interações não-seletivas e, portanto, não serão apresentados.

A informação de natureza complementar obtida a partir dos métodos baseados na estrutura dos ligantes e na estrutura dos subtipos do ER sugere que, a despeito da pequena variação entre as cavidades de interação com ligantes desses receptores, é possível o planejamento de ligantes seletivos, de uma maneira racional, com o emprego de métodos avançados em Química Medicinal.

Capítulo 5

CONCLUSÕES E PERSPECTIVAS

1 CONCLUSÕES

- ✓ A diversidade química e a distribuição dos valores da propriedade farmacológica (afinidade e potência) nos conjuntos de dados organizados neste trabalho são apropriadas, proporcionando considerável cobertura do espaço químico-biológico explorado.
- ✓ O uso dos conjuntos teste foi de extrema importância no processo de validação externa dos modelos de QSAR.
- ✓ Os modelos finais de HQSAR e CoMFA para os moduladores do hER α dos conjuntos de dados 1 e 2, e para os moduladores de hER α e hER β do conjunto de dados 3, apresentam, como indicado pelos bons coeficientes de correlação e capacidade preditiva, elevada consistência interna e externa.
- ✓ Os diferentes alinhamentos do conjunto de moduladores nas cavidades de hER α e hER β , e os modelos de QSAR CoMFA gerados sugerem que os requerimentos tridimensionais nas interações entre os ligantes e os subtipos do ER são predominantes em relação a seletividade.
- ✓ Os resultados dos estudos tridimensionais baseados na estrutura dos moduladores e dos subtipos do ER estão em concordância e evidenciam as regiões particularmente interessantes para modificações estruturais dirigidas ao planejamento de moduladores β -seletivos.
- ✓ Em todos os modelos, os mapas de contribuição 2D e de contorno 3D são compatíveis com os ambientes químicos tridimensionais das cavidades do hERs e podem ser

utilizados no planejamento de NCEs apresentando potência, afinidade e seletividade otimizadas.

- ✓ O uso de métodos baseados na estrutura do receptor em conjunto com o emprego de técnicas de QSAR 3D (CoMFA) e QSAR 2D (HQSAR) confirma, uma vez mais, a validade da integração de estratégias procurando um sinergismo útil no planejamento de novos candidatos a fármacos com propriedades otimizadas. Os modelos finais são úteis para guiar futuros trabalhos em química medicinal no planejamento de novos moduladores do ER.

2 PERSPECTIVAS

Esse trabalho foi parte de um projeto de pesquisa multidisciplinar bastante amplo, envolvendo um convênio intitulado “Descoberta de ligantes do receptor de estrógeno candidatos a novos fármacos no tratamento da osteoporose, câncer de mama e reposição hormonal (Convênio FINEP Ref. 3039/04)”, entre o Instituto Internacional de Pesquisas Farmacêuticas-EMS Sigma Pharma e o CBME-IFSC. Os modelos de QSAR construídos foram utilizados na avaliação e otimização de novos candidatos a moduladores seletivos dos ERs. Métodos avançados em Química Medicinal, como o ensaio virtual e buscas 2D e 3D em bases de dados de compostos, foram também utilizados. Alguns compostos foram selecionados, comprados e testados. Esses resultados estão sob proteção de propriedade intelectual, e não foram, portanto, apresentados nessa dissertação.

Capítulo 6

REFERÊNCIAS

- 1 AGNEW, B. When pharma merges, R&D is the dowry. *Science*, v. 287, p.1952-1953,2000.
- 2 COHEN, F. J. Macro trends in pharmaceutical innovation. *Nat. Rev. Drug Discov.*, **2005**, 4, 78-84.
- 3 COUZIN, J. Brains behind blockbuster. *Science*, v. 309,p.728-730,2005.
- 4 BERNDT, E. R. et al. Industry fencing of FDA: Effects of PDUFA on approval times and withdrawal rates. *Nat. Re. Drug Discov.*, v.4, p.545-554,2005.
- 5 LOMBARDINO, J. G.; LOWE III, J. A. The role of the medicinal chemist in drug discovery – then and now. *Nat. Rev. Drug. Discov.*,v. 3,p. 853-862.,2004.
- 6 DIMASI, J. A.; HANSEN, R. W.; GRABOWSKI, H. G. The price of innovation: new estimates of drug development cost. *J. Health Econom.*, v. 22, p.151-185,2003.
- 7 UNITED STATES GOVERNMENT ACCOUNTABILITY OFFICE (2006). New drug development. science, business, regulatory, and intellectual property issue cited as hampering drug development efforts.Disponível em:< <http://www.gao.gov/new.items/d0749.pdf>>. Acesso em: 15 Jul 2007.
- 8 DICKSON M., GAGNON J. P. Key factors in the rising cost of new drug discovery and development. *Nat. Rev. Drug. Discov.*, v.5, p.417-429,2004.
- 9 DREWS, J. Basic science and pharmaceutical innovation. *Nat. Biotechnol.*,v.17, p.406,1999.
- 10 FDA. Food and Drug Administration (2004). Innovation or stagnation. challenge and opportunity on the critical path to new medical products. Disponível em:<<http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html>> .Acesso em: 15 Jul 2007.
- 11 EMEA, European Agency for the Evaluation of Medicinal Products (2006) Annual report of the European Medicines Agency 2005.Disponível em:<<http://www.emea.europa.eu/pdfs/general/direct/emeaar/AnnualReport2005.pdf>> .Acesso em: 15 Jul 2007.
- 12 CAPRINO, L.; RUSSO, P. Developing a paradigm of drug innovation: an evaluation algorithm. *Drug Discov. Today*, v.11,p.999-1006,2006.
- 13 YOUNG, D. Drug innovation on the decline. *Am. J. Health Syst. Pharm.*, v.64,p. 228-231,2007.
- 14 FEDERSEL, H. J. In search of sustainability: process R&D in light of current pharmaceutical industry challenges. *Drug Discov. Today*, v.11,p.966-974,2006.
- 15 THE ECONOMIST (2006). Dispensed with yet another boss of a big drugs company loses his job. Disponível em: <www.electronicconomist.com/research>.Acesso em: 15 Jul 2007.
- 16 UPPENBRINK, J.; MERVIS, J. An information revolution. *Science*, v. 287, p.1951,2000.
- 17 RUSSO, E. Chemistry plans a structural overhaul. *Nature*, v.419, p.4-7,2002.
- 18 PAMMOLLI, F.; RICCABONI, M. Market structure and drug innovation. *Health Aff (Millwood)*, v.23, p.48-50,2004.
- 19 DREWS, J. Drug discovery: a historical perspective. *Science*, v.287, p.1960-1964,2000.
- 20 MCGEE, P. Clinical Trials on the Move. *Drug Discov. Dev.*,v. 9, p.16-22, 2006.
- 21 WERMUTH, C. G. The practice of medicinal chemistry. London: Academic Press, 2003.
- 22 BARREIRO, E. J.; FRAGA, C. A. M. Química medicinal: as bases moleculares da ação dos fármacos. Porto Alegre:Artmed Editora, 2001.
- 23 THOMAS, G. Fundamentals of medicinal chemistry. Chicheste: Wiley, 2003

- 24 WESS, G.; URMANN, M.; SICKENBERGER, B. Medicinal chemistry: challenges and opportunities. *Angew Chem. Int. Ed. Engl.*, v.40, p.341-3350, 2001.
- 25 JÓNSDÓTTIR S. O.; JØRGENSEN, F. S.; BRUNAK, S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics*, v.21, p.2145-2160, 2005.
- 26 BUNIN, A. B. Increasing the efficiency of small molecule drug discovery. *Drug. Discov. Today*, v. 18, p.823-826, 2003.
- 27 ANDRICOPULO, A. D.; MONTANARI, C. A. Structure-activity relationships for the design of Small-molecule inhibitors. *Mini Rev. Med. Chem.*, v. 5, p.585-593, 2005.
- 28 KUBINIY, H. 3D QSAR in drug design: theory, methods and applications. Leiden: Kluwer Academic Publishers/Escom, 1993. v.1.
- 29 NCI – NATIONAL CANCER INSTITUTE. U.S. National Institutes of Health. (2007). Comprehensive cancer information .Disponível em: < <http://www.cancer.gov/>> Acesso em: 15 Jul 2007.
- 30 ACS – AMERICAN CANCER SOCIETY (2007). Cancer facts & figures 2007. Disponível em: <<http://www.cancer.org> > .Acesso em: 15 Jul 2007.
- 31 INSTITUTO NACIONAL DO CÂNCER, INCA (2007). Estimativa 2006 – Incidência de câncer no Brasil. Disponível em: <<http://www.inca.gov.br/>>. Acesso em: 15 Jul 2007.
- 32 JORDAN, V.C. Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 1. receptor interactions. *J. Med. Chem.*, v. 46, p.883-908, 2003.
- 33 JORDAN, V.C. Antiestrogens and selective estrogen receptor modulators as multifunctional medicines: 2. clinical considerations and new agents. *J. Med. Chem.*, v. 46, p.1081-1111, 2003.
- 34 LEGHA, S.S.; CARTER, S.K. Antiestrogens in the treatment of breast cancer. *Cancer Treat. Rev.*, v.3, p.205-216, 1976.
- 35 LERNER, L.J.; JORDAN, V.C. The development of antiestrogens for the treatment of breast cancer. *Cancer Res.*, v.50, p.4177-4189, 1990.
- 36 ARIAZI, E.A. et al. Estrogen receptors as therapeutic targets in breast cancer. *Curr. Top. Med. Chem.*, v.6, p.181-202, 2006.
- 37 PANSINI, F. et al. The menopausal transition: a dynamic approach to the pathogenesis of neurovegetative complaints. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, v.57, p.103-109, 1994.
- 38 KHAN, S. A. et al. Climacteric symptoms in healthy middle-aged women. *Br. J. Clin. Pract.*, v.48, p.240-242, 1994.
- 39 PALACIOS, S. Current perspectives on the benefits of HRT in menopausal women. *Maturitas*, p.S1-S13, 1999.
- 40 BARRETT-CONNOR, E.; GRADY, D.; STEFANICK, M. L. The rise and fall of menopausal hormone therapy. *Annu. Rev. Public Health*, v.26, p.115-40, 2005.
- 41 SHERWIN, B. B. Hormones, mood, and cognitive functioning in postmenopausal women. *Obstet. Gynecol.*, v.87, p.20S-26S, 1996.
- 42 ROSANO, G. M.; VITALE, C.; FINI, M. Hormone replacement therapy and cardioprotection: what is good and what is bad for the cardiovascular system? *Ann. N. Y. Acad. Sci.*, v.1092, p.341-348, 2006.
- 43 VESTERGAARD, P.; REJNMARK, L.; MOSEKILDE, L. Fracture reducing potential of hormone

- replacement therapy on a population level. *Maturitas*, v.54,p.285-293,2006.
- 44 SHANG, Y. Molecular mechanisms of oestrogen and SERMs in endometrial carcinogenesis. *Nat. Rev. Cancer*, v.6, p.60-368,2006.
- 45 FERNANDEZ, E. et al. Hormone replacement therapy and cancer risk: a systematic analysis from a network of case-control studies. *Int. J. Cancer*, v.105, p.408-412,2003.
- 46 YEH, I. T. Postmenopausal hormone replacement therapy: endometrial and breast effects. *Adv. Anat. Pathol.*, v.14, p.17-24,2007.
- 47 KENEMANS, P. et al. Perspectives in hormone replacement therapy. *Maturitas*, v.15, p.S41-S48,2001.
- 48 NILSSON S.; KOEHLER, K. F. Oestrogen receptors and selective oestrogen receptor modulators: molecular and cellular pharmacology. *Basic Clin. Pharmacol. Toxicol.*, v.96, p.5-25,2005.
- 49 SAJI, S.; HIROSE, M.; TOI, M. Clinical significance of estrogen receptor β in breast cancer. *Cancer Chemother. Pharmacol.*, v.56, p.21-26,2005.
- 50 KATZENELLENBOGEN, B. S. et al. Molecular mechanisms of estrogen action: selective ligands and receptor pharmacology. *J. Steroid. Biochem. Mol. Biol.*, v.74, p.279-285,2000.
- 51 ADLERCREUTZ, H. Phyto-oestrogens and cancer. *Lancet Oncol.*, v.6, p.364-373,2002.
- 52 MCCARTY, M. F. Isoflavones made simple - genistein's agonist activity for the beta-type estrogen receptor mediates their health benefits. *Med. Hypotheses*, v.66, p.1093-11,2006.
- 53 ADLERCREUTZ, H.; HEINONEN, S. M.; PENALVO-GARCIA, J. Phytoestrogens, cancer and coronary heart disease. *Biofactors*, v.22, p.229-36,2004.
- 54 BECK, V.; ROHR, U.; JUNGBAUER, A. Phytoestrogens derived from red clover: an alternative to estrogen replacement therapy? *J. Steroid. Biochem. Mol. Biol.*, v.94, p.499- 518,2005.
- 55 KUIPER, G. et al. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor β . *Endocrinology*, v.139, p.4252-4263,1998.
- 56 NILSSON, S.; GUSTAFSSON, J.-Å. Biological role of estrogen and estrogen receptors. *Crit. Rev. Biochem. Mol. Biol.*, v.37, p.1-28,2002.
- 57 OSBORNE, C. K.; SCHIFF, R. Estrogen-receptor biology: continuing progress and therapeutic implications. *J. Clin. Oncol.*, v.23, p.1616-1622, 2005.
- 58 GRUBER, C.J. et al. Production and actions of estrogens. *N. Engl. J. Med.*, v.346, p.340-352,2002.
- 59 GREEN, S. et al. Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature*, v.320, p.134-139,1986.
- 60 GREENE, G. L. et al. Sequence and expression of human estrogen receptor complementary DNA. *Science*, v.231, p.1150-1154,1986.
- 61 KUIPER, G.G. et al. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc. Natl. Acad. Sci. U.S.A.*, v.93, p.5925-5930,1986.
- 62 MOSSELMAN, S.; POLMAN, J.; DIJKEMA, R. ER β : Identification and characterization of a novel human estrogen receptor. *FEBS Lett.*, v.392, p.49-53,1996.
- 63 KUIPER, G. G. et al. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology*, v.138, p.863-870,1997.
- 64 SUN, J. et al. Molecular basis for the subtype discrimination of the estrogen receptor-beta-selective ligand, diarylpropionitrile. *Mol. Endocrinol.*, v.17, p.247-258,2003.

- 65 MILLER, C. P.; COLLINI, M. D.; HARRIS, H. A. Constrained phytoestrogens and analogues as ERbeta selective ligands. *Bioorg. Med. Chem. Lett.*, v.13, p.2399-2403,2003.
- 66 COUSE, J. F. et al. Tissue distribution and quantitative analysis of estrogen receptor-alpha (ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic acid in the wild-type and ERalpha-knockout mouse. *Endocrinology*, v.138, p.4613-4621,1997
- 67 GUSTAFSSON, J-A. Estrogen receptor beta – a new dimension in estrogen mechanism of action. *J. Endocrinol.*, v.163, p.379-383,1999.
- 68 BRZOZOWSKI, A. M. et al. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature*, v.389, p.753-758,1997.
- 69 SHIAU, A. K. et al. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*, v.95, p.927-937,1988.
- 70 MACGREGOR, J.I.; JORDAN, V.C. Basic guide to the mechanisms of antiestrogen action. *Pharmacol. Rev.*, v.50, p.151-196,1998.
- 71 GRONEMEYER, H.; GUSTAFSSON, J-Å.; LAUDET, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.*, v.3, p.950-964,2004.
- 72 KATZENELLENBOGEN, J.A.; O'MALLEY, B.W.; KATZENELLENBOGEN, B.S. Tripartite steroid hormone receptor pharmacology: interaction with multiple effector sites as a basis for the cell- and promoter-specific action of these hormones. *Mol. Endocrinol.*, v.10, p.119-131,1996.
- 73 LIEBERMAN, M. E.; GORSKI, J.; JORDAN, V. C. An estrogen receptor model to describe the regulation of prolactin synthesis by antiestrogens in vitro. *J. Biol. Chem.*, v.258, p.4741-4745,1983.
- 74 KUBINYI, H. QSAR: hansch analysis and related approaches. New York:VCH, 1993.
- 75 HANSCH, C.; LEO, Albert. *Exploring QSAR: fundamentals and applications in chemistry and biology*. Washington: American Chemical Society,1995.
- 76 VON ITZSTEIN, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, v. 363,p.418-423,1993.
- 77 LOWIS, D. R. HQSAR: a new, highly predictive QSAR technique. *Tripos Technical Notes*, v.1, 1997.
- 78 HQSAR^{TM} Manual. St. Louis:Tripos Inc., 2006.
- 79 TONG, W. et al. Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.*, v.38, p.669-677,1998.
- 80 SEEL, M.; TURNER, D.B.; WILLETT, P. Effect of parameter variations on the effectiveness of HQSAR analyses. *QSAR*, v.18,p.245-25,1999.
- 81 CASTILHO, M.S. et al. Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg. Med. Chem.*, v.14, p.516-527,2006.
- 82 HONORIO, K.M.; GARRATT, R.C.; ANDRICOPULO, A.D. Hologram quantitative structure-activity relationships for a series of farnesoid X receptor activators. *Bioorg. Med. Chem. Lett.*, v.15,p.3119-3125,2005.
- 83 CRAMER, RD. III , PATTERSON, D.E., BUNCE, J.D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.*, v.291, p.161-165,1989.
- 84 KUBINYI, H. Comparative molecular field analysis. IN: Encyclopedia of computational chemistry, 1998. v.1.

- 85 GRID Manual. Oxford: Molecular Discovery Ltd.,c.2004.
- 86 GOODFORD, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, v.28, p.849-857,1985
- 87 KASTENHOLZ, M. A. et al. GRID/CPCA: a new computational tool to design selective ligands *J. Med. Chem.*, v.43, p.3033-3044,2000.
- 88 TERP, G. E. et al. Structural differences of matrix metalloproteinases with potential implications for inhibitor selectivity examined by the GRID/CPCA approach. *J. Med. Chem.*, v.45,p.2675 – 2684,2002.
- 89 CHEN, H.Y. et al. Estrogen receptor ligands. Part 1: the discovery of flavanoids with subtype selectivity. *Bioorg. Med. Chem. Lett.*, v.14, p.1417-1421,2004.
- 90 KIM, S. et al. Estrogen receptor ligands II: discovery of benzoxathiins as potent, selective estrogen receptor α modulators. *J. Med. Chem.*, v.47, p.2171-2175,2004.
- 91 CHEN, H. Yet al. Estrogen receptor ligands. Part 3: the SAR of dihydrobenzoxathiin SERMs. *Bioorg. Med. Chem. Lett.*, v.14, p. 2551-2554,2004.
- 92 KIM, S. et al. Estrogen receptor ligands. Part 4: the SAR of the syn-dihydrobenzoxathiin SERAMs. *Bioorg. Med. Chem. Lett.*, v.14,p. 2741-2745,2004.
- 93 TAN, Q. et al. Estrogen receptor ligands. Part 5: the SAR of dihydrobenzoxathiins containing modified basic side chains. *Bioorg. Med. Chem. Lett.*, v.14, p.3747-3751,2004.
- 94 TAN, Qet al. Estrogen receptor ligands. Part 6: synthesis and binding affinity of dihydrobenzodithiins. *Bioorg. Med. Chem. Lett.*, v.14, p.3753-3755,2004.
- 95 BLIZZARD, T. A. et al. Estrogen receptor ligands. Part 7: dihydrobenzoxathiin SERAMs with bicyclic amine side chains. *Bioorg. Med. Chem. Lett.*, v.14, p.3861-3864,2004.
- 96 BLIZZARD, T. A. et al. Estrogen receptor ligands. Part 8: dihydrobenzoxathiin SERAMs with heteroatom-substituted side chains. *Bioorg. Med. Chem. Lett.*, v.14, p.3865-3868,2004.
- 97 BLIZZARD, T. A. et al. Estrogen receptor ligands. Part 9: dihydrobenzoxathiin SERAMs with alkyl substituted pyrrolidine side chains and linkers. *Bioorg. Med. Chem. Lett.*, v.15, p.107-113,2005.
- 98 TAN, Q. et al. Estrogen receptor ligands. Part 10: chromanes,2005 old scaffolds for new SERAMs. *Bioorg. Med. Chem. Lett.*, v.15, p.1675-1681,2005.
- 99 BLIZZARD, T. A. et al. Estrogen receptor ligands. Part 13: dihydrobenzoxathiin SERAMs with an optimized antagonist side chain. *Bioorg. Med. Chem. Lett.*, v.15, p.3912-3916,2005.
- 100 MEWSHAW, R. E. et al. ERbeta ligands:exploiting two binding orientations of the 2-phenylnaphthalene scaffold to achieve ERbeta selectivity. *J. Med. Chem.*, v.48, p.3953-3979,2005.
- 101 VU, A. Tet al. ERbeta ligands. Part 4: synthesis and structure-activity relationships of a series of 2-phenylquinoline derivatives. *Bioorg Med Chem Lett.*, v.15, p.4520-4525,2005.
- 102 JONES, G. et al. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, v.267, p.727-748,1997.
- 103 VERDONK, M. L. et al. Improved protein-ligand docking using GOLD. *Proteins*,v.52, p.609-623,2003.
- 104 GAO, H. et al. Comparative QSAR analysis of estrogen receptor ligands. *Chem. Rev.*, v.99, p.723-744,1999.
- 105 MINUTOLO, F. et al.Synthesis, binding affinity, and transcriptional activity of hydroxy- and methoxy-substituted 3,4-diarylsalicylaldoximes on estrogen receptors alpha and beta. *Bioorg. Med.*

- Chem.*, v.11, p.1247-1257,2003.
- 106 GOLBRAIKH, A.; TROPSHA, A. Beware of q²! *J. Mol. Graph. Model.*, v.20, p.269-276,2002
  - 107 CoMFA MANUAL. St. Louis:Tripos Inc., 2006.
  - 108 GOLD 3.0 Manual 3.0 . Cambridge:Cambridge Crystallographic Data Centre,s.d.
  - 109 LEVENSON, A. S.; JORDAN, V. C. The key to the antiestrogenic mechanism of raloxifene is amino acid 351 (aspartate) in the estrogen receptor. *Cancer Res.*, v.58, p.1872 – 1875,1998.
  - 110 KIM, J. H. et al. Role of aspartate 351 in transactivation and active conformation of estrogen receptor alpha. *J. Mol. Endocrinol.*, v.35, p.449-464,2005.
  - 111 ZHAO, C.; ABRAMS, J.; SKAFAR, D. F. Targeted mutation of key residues at the start of helix 12 in the hERalpha ligand-binding domain identifies the role of hydrogen-bonding and hydrophobic interactions in the activity of the protein. *J. Steroid Biochem. Mol. Biol.*, v.98, p.1-11,2006.
  - 112 DAYAN, G. et al. Tamoxifen and raloxifene differ in their functional interactions with aspartate 351 of estrogen receptor alpha. *Mol. Pharmacol.*, v.70, p.579-588,2006.
  - 113 KUBINYI, H. QSAR and 3D QSAR in drug design. Part 2: applications and problems. *Drug Discov. Today*, v.12, p.538–546,1997.
  - 114 SALUM, L. B.; POLIKARPOV, I.; ANDRICOPULO, A. D. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. *J. Mol. Graph. Model.*, v.27, p. 434-442,2007.
  - 115 DEMYTTEAERE-KOVATCHEVA, A. et al. Identification of the structural requirements of the receptor-binding affinity of diphenolic azoles to estrogen receptors α and β by three-dimensional quantitative structure-activity relationship and structure-activity relationship analysis. *J. Med. Chem.*, v.48, p.7628-7636,2005.
  - 116 TONG, W. et al. QSAR models for binding of estrogenic compounds to estrogen receptor alpha and beta subtypes. *Endocrinology*, v.138, p.4022-4025,1997.
  - 117 ZHU, B. T. et al. Quantitative structure-activity relationship of various endogenous estrogen metabolites for human estrogen receptor α and β subtypes: insights into the structural determinants favoring a differential subtype binding. *Endocrinology*, v.147, p.4132-4150,2006.
  - 118 HARRIS, H. A. et al. The ligand binding profiles of estrogen receptors alpha and beta are species dependent. *Steroids*, v.67, p.379-384,2002.
  - 119 HARRIS, H. A. et al. Evaluation of an estrogen receptor-beta agonist in animal models of human disease. *Endocrinology*, v.144,p.4241-4219,2003.
  - 120 NILSSON, S.; KUIPER, G.; GUSTAFSSON, J.-A.; ER β : a novel estrogen receptor offers the potential for new drug development. *Trends Endocrinol. Met.*, v.9, p.387-395,1998.
  - 121 MALAMAS, M. S. et al. Design and synthesis of aryl diphenolic azoles as potent and selective estrogen receptor-beta ligands. *J. Med. Chem.*,v.47,p.5021-5040,2004.
  - 122 GÜNGÖR, T. et al. Synthesis and characterization of 3-arylquinazolinone and 3-arylquinazolinethione derivatives as selective estrogen receptor beta modulators. *J. Med. Chem.*, v.49, p.2440-2455,2006.
  - 123 SALUM, L. B.; POLIKARPOV, I.; ANDRICOPULO, A. D. Quantitative structure-activity relationships for a series of selective estrogen receptor-beta modulators . *SAR QSAR Environ. Res.* **2007**.No prelo.
  - 124 HSIEH, R. W. et al. Identification of ligands with bicyclic scaffolds provides insights into mechanisms of estrogen receptor subtype selectivity. *J. Biol. Chem.*, v.281, p.17909-17919,2006.

A PÊNDICES

Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies

Lívia de B. Salum, Igor Polikarpov, Adriano D. Andricopulo*

Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-Carlense 400, 13560-970 São Carlos, SP, Brazil

Received 15 December 2006; received in revised form 31 January 2007; accepted 2 February 2007

Available online 8 February 2007

Abstract

The estrogen receptor (ER) is an important drug target for the development of novel therapeutic agents for the treatment of breast cancer. Progress towards the design of more potent and selective ER modulators requires the optimization of multiple ligand-receptor interactions. Comparative molecular field analyses (CoMFA) and hologram quantitative structure–activity relationships (HQSAR) were conducted on a large set of ER α modulators. Two training sets containing either 127 or 69 compounds were used to generate QSAR models for *in vitro* binding affinity and potency, respectively. Significant correlation coefficients (affinity models, CoMFA, $r^2 = 0.93$ and $q^2 = 0.79$; HQSAR, $r^2 = 0.92$ and $q^2 = 0.71$; potency models, CoMFA, $r^2 = 0.94$ and $q^2 = 0.72$; HQSAR, $r^2 = 0.92$ and $q^2 = 0.74$) were obtained, indicating the potential of the models for untested compounds. The generated models were validated using external test sets, and the predicted values were in good agreement with the experimental results. The final QSAR models as well as the information gathered from 3D contour maps should be useful for the design of novel ER α modulators having improved affinity and potency.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Estrogen receptor; Drug design; Breast cancer; CoMFA; HQSAR

1. Introduction

Estrogens are crucial endogenous hormones that modulate the development and homeostasis of a wide range of female target tissues, including reproductive tracts, breast and skeletal system [1–3]. Their physiological effects are mediated by the estrogen receptor (ER) subtypes, ER α and ER β , which are members of the nuclear receptor superfamily of ligand-modulated transcriptional factors [4–6]. The binding of the natural ligand estradiol or other ligands to ER triggers complex signaling networks, leading to the recruitment of coregulatory complexes and to the transcription or repression of specific genes.

The high flexibility of the ER ligand-binding cavity allows the binding of a series of structurally diverse small molecules, which can exhibit agonist and antagonist effects, depending on

intrinsic cellular differences in the recruitment of coregulators [1,2,7–11].

The ligand-induced conformational changes in the ER binding site and the receptor structure in general, particularly at the N-terminal helix 12 position and its conformation, are major determinants of either agonist or antagonist properties. The understanding of the molecular recognition process responsible for the pharmacological effects of ER is very complex and has been extensively discussed in the literature [9,12,13].

Given the established role of ER α on the etiology and pathophysiology of breast cancer, the second leading cause of cancer death in women, selective ER modulation is an important strategy in the treatment of this disease [14–16]. The ER modulators tamoxifen (generic) and fulvestrant (Faslodex®) are well-known drugs employed in cancer therapy. Since resistance and serious side effects have been reported, including endometrial changes, hot flashes and irregular menstrual periods, there is an increasing clinical need for new therapeutic agents [2,8,9,16–20].

* Corresponding author. Tel.: +55 16 3373 8095; fax: +55 16 3373 9881.

E-mail address: aandrico@ifsc.usp.br (A.D. Andricopulo).

The key issue in the design of new selective ER modulators is to explore the properties of the chemical structure in combination with its ability of inducing a pharmacological response as a consequence of receptor-binding. As part of our ongoing research program aimed at discovering new selective and potent ER antagonists, and in order to investigate the quantitative structure–activity relationships (QSAR) of a large series of ER α ligands [21–31], we have employed the comparative molecular field analysis (CoMFA) and the hologram QSAR (HQSAR) methods to generate predictive 3D and 2D QSAR models, respectively [32,33]. The identification of key structural features responsible for binding affinity and potency should be useful for the design of novel modulators having promise of utility in clinical medicine.

2. Methodology

2.1. Data sets

The data set of 127 ER α modulators used in the QSAR analyses was selected from the literature [21–31]. Table 1 shows the general chemical structures of the series of ER α modulators employed in the QSAR studies. The complete version of Table 1, including all chemical structures and corresponding biological data, is provided as supplementary information, as well as the chemical structures in SMILES format for the complete data set.

2.2. Computational approach

QSAR modeling analyses, calculations, and visualizations for CoMFA and HQSAR were performed using the SYBYL 7.2 package (Tripos Inc., St. Louis, USA) running on Red Hat Enterprise Linux workstations. The 3D structures of the modulators were constructed using standard geometric parameters of molecular modeling software package SYBYL 7.2. Each single optimized conformation of each molecule in the data set was energetically minimized employing Tripos force field and Gasteiger–Huckel charges. Ligands presenting tertiary amine function were modeled in the protonated state, since under physiological conditions these ligands are expected to be in this form.

2.3. Molecular modeling and structural alignment

Docking protocol as implemented in GOLD 3.0 (Cambridge Crystallographic Data Centre, Cambridge, UK) was employed to search the possible binding conformations of ligands into the ER α ligand-binding site. Crystallographic data for ER α in complex with (2*S*,3*R*)-2-(4-{2-[(3*R*,4*R*)-3,4-dimethylpyrrolidin-1-yl]ethoxy}phenyl)-3-(4-hydroxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol (compound **33**, Table 1) used in simulations were retrieved from the Protein Data Bank (PDB code: 1XP1). All docking simulations were performed with the X-ray coordinates of the estrogen receptor structure to which the ligand and water molecules were removed. Hydrogen atoms were added in standard geometry using the Biopolymer module

Table 1

Representative chemical structures and associated binding affinity values (pIC₅₀) of the series of ER α modulators employed in the QSAR studies

Compound	pIC ₅₀	Compound	pIC ₅₀
	6.31		7.54
	7.05		8.13
	7.85		8.40
	5.98		5.96

as implemented is Sybyl 7.2. His, Gln, and Asn residues were manually checked for possible flipped orientation, protonation, and tautomeric states with Pymol 0.99 (DeLano Scientific, San Carlos, USA) side-chain wizard script. The binding site was centered on C22 atom of the (2*S*,3*R*)-2-(4-{2-[(3*R*,4*R*)-3,4-dimethylpyrrolidin-1-yl]ethoxy}phenyl)-3-(4-hydroxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol and a radius of 10 Å was considered for the docking procedure. Default parameters and GOLDScore function were employed in all runs, and only the best ranked conformation of each modulator was considered for 3D QSAR studies [34,35,36]. The aligned data set is depicted in Fig. 1.

2.4. 3D QSAR: CoMFA models

To better understand and explore the contributions of electrostatic and steric fields in the binding affinity and potency of the ER α modulators, and in order to build predictive 3D

Fig. 1. Three-dimensional data set alignment for conformations generated by GOLD 2.1.

QSAR models, CoMFA studies were performed based on the molecular alignment described. The contributions of steric and electrostatic fields were calculated according to Lennard–Jones and Coulomb potentials, respectively [32]. The aligned training set of molecules was placed in a 3D grid box such that the entire set was included in it. CoMFA steric and electrostatic fields were generated at each grid point with Tripos force field using a sp^3 carbon atom probe carrying a +1 net charge. The CoMFA grid spacing of 2.0 Å in the x , y , and z directions, and the grid region were generated by the CoMFA routine to encompass all molecules with an extension of 2.0 Å in each direction. CoMFA region focusing method was applied to increase the resolution of CoMFA models. The default value of 30 kcal mol⁻¹ was set as the maximum steric and electrostatic energy cutoff. All models were investigated using full cross-validated r^2 (q^2) partial least squares (PLS) leave-one-out (LOO) method, with CoMFA standard options for scaling of variables. Progressive scrambling method was applied to determine the sensitivity of the QSAR models to chance correlations.

2.5. 2D QSAR: HQSAR Models

Since the integrated information obtained from 2D and 3D methods is recognized as a valuable strategy in drug design [34,37,38], we have explored the 2D molecular features related to the biological properties of this series of ER ligands. Predictive 2D QSAR models were constructed using the HQSAR technique, which requires only 2D structures and biological activity values as input [33,34,39,40]. In this method, each molecule in the data set is broken down into several unique structural fragments, which are arranged to form a molecular hologram. HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length,

fragment size, and fragment distinction. Several combinations of fragment distinction were considered during the QSAR modeling runs. Holograms were generated using distinct combinations of atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA) as fragment distinctions. The HQSAR analysis was performed by screening the 12 default series of hologram length values ranging from 53 to 401 bins using distinct fragment sizes. The patterns of fragment counts from the training set modulators were then related to the experimental biological parameters using PLS analysis. LOO cross-validation was applied to determine the number of components that yield optimally predictive model.

3. Results and discussion

3.1. Data set characterization

3D CoMFA and 2D HQSAR models were derived for 127 ER α ligands using two distinct data sets (Table 1). The first data set (data set 1) containing the whole set of 127 ER ligands was used to create the binding affinity (as measured by IC₅₀ values) models. From the original data set, 99 compounds were selected as members of the training set for model construction while the other 28 compounds were held out as members of the test set for external validation. The second data set (data set 2) containing a subset of 69 ER ligands was employed to create the potency (*in vitro* inhibition of MCF-7 cell growth) models. Data set 2 was divided as follows: 55 compounds in the training set whereas the other 14 in the test set (Table 1).

The three major chemical classes included in the data sets are flavanoids, dihydrobenzoxathiins and dihydrobenzodithiins. These chemical classes were selected both because they provide an appropriate structural diversity for QSAR modeling and because the availability of high quality biological data. Hierarchical cluster analysis performed with Tsrar 3D (Accelrys, San Diego, USA) was used as previously described to guide an appropriate compound selection. In both cases, training and test sets were selected in such a way that structurally diverse molecules possessing activities of a wide range were included in both sets [34,39]. Thus, the data sets are appropriate for the purposes of QSAR model development. The PLS (partial least squares) method was used for all 3D and 2D QSAR analyses. The predictive ability of the models was assessed by their q^2 .

3.2. CoMFA models

In CoMFA, relationships for a set of compounds are established between the biological activities and their steric and electrostatic properties. This approach is based on the assumption that changes in ligand-binding affinities are related to changes in molecular properties represented by molecular fields. The alignment rule is a crucial variable, affecting the outcome of the 3D statistical analysis. An ideal alignment should represent the ligand-binding conformations adopted in the receptor binding site. After the alignment process, each of the employed conformations was inspected individually and no conformational problem was revealed. The molecular

Table 2
COMFA results

Data set	Num	q^2	N	r^2	S.E.E.	F	Fraction	
							S	E
1	127	0.79	7	0.93	0.26	161	0.74	0.26
2	69	0.72	7	0.94	0.24	105	0.66	0.34

Data set 1, binding affinity; data set 2, potency; Num, number of compounds used in the modelling; q^2 , leave-one-out (LOO) cross-validated correlation coefficient; N , optimum number of components; r^2 , non-cross-validated correlation coefficient; S.E.E., standard error of estimate; F , F -test value; S , steric field; E , electrostatic field.

alignment shown in Fig. 1 was submitted to the CoMFA analyses, and the statistical results are presented in Table 2. Significant correlation coefficients were obtained. As it can be seen, a cross-validated correlation coefficient q^2 of 0.79 and a

conventional non-cross-validated correlation coefficient r^2 of 0.93 with a standard error of estimate (S.E.E.) of 0.26 were obtained for the data set 1 (binding affinity), while a q^2 of 0.72 and an r^2 of 0.94 with a S.E.E. of 0.24 were obtained for the data set 2 (potency). In both cases, no significant differences were obtained using the LMO (leave-many-out) method. Progressive scrambling of the data sets was carried out to check for possible chance correlations and test the stability of the models. The results of progressive scrambling further confirmed consistency of the models as defined by the critical slope, and optimum statistics for cSDEP and Q^{*2} obtained at the end of different runs. The region focusing was weighted by $\text{StDev} \times \text{Coefficient}$ values ranging from 0.3 to 1.5, and grid spacing ranging from 1.0 to 4.0. This strategy not only increased q^2 values during the process of model generation, but also resulted in the refinement of 3D contour maps.

Fig. 2. Data set 1 CoMFA steric and electrostatic contour maps for the derivative with highest affinity, **67** (A). Corresponding maps for the data set 2 and the most potent compound **43** (B). Greater values of the affinity and potency are correlated with: more bulk near green; less bulk near yellow and more positive charge near blue.

Fig. 3. Representation of the binding mode of the highest affinity modulator (**67**, colored in pink) into crystal structure of ER α . In blue, some important residues in the ligand-binding site.

CoMFA steric field descriptor for data set 1 explains 74% of the total variance, while the electrostatic descriptor accounts for the rest 26%. For data set 2, steric and electrostatic fields explain 66 and 34%, respectively. As expected, the steric field descriptors were more important to explain the total variance in both affinity and potency models [41–44], since the ER ligand-binding cavity possess hydrophobic features which must be fitted correctly by selective ligands, as confirmed by the structural hydrophobic core of estradiol, its natural endogenous ligand.

The data set 1 CoMFA steric and electrostatic fields contour maps for compound **67** (highest affinity of the series) are shown in Fig. 2A, whereas the corresponding maps for the data set 2 and compound **43** (the most potent of the series) are shown in Fig. 2B as PLS StDev*Coefficient plots. A comparison between affinity and potency electrostatic contour maps reveals that there is no significant difference, since the protonated nitrogen atom in the pyrrolidine ring is related to increasing affinity and potency. CoMFA steric fields are also similar in both 3D contour maps. Less steric bulky substituents attached to the benzoxathiin ring are positively related to the biological parameters [45,46]. The presence of either *para*- or *meta*-substituents is considered a positive contribution to affinity, but not to potency, as indicated by the green region in the affinity map around the phenyl ring. There is still room for bulkier groups in the pyrrolidine ring, which would increase both affinity and potency as suggested by the 3D contour maps. These chemical connections are especially attractive for molecular modification and further SAR studies aimed at increasing affinity and potency.

The 3D QSAR models generated are compatible with the 3D protein environment in the ER binding site, as shown in Fig. 3. The importance of the substituents in the benzoxathiin and phenyl rings corroborates with the estrogen receptor interaction sites HIS524, ARG394 and GLU353, as previously reported [9,17,41,47]. The maps indicate that the nitrogen pyrrolidine ring is important for the antagonist biocharacter. This parameter was similar to some crystallographic structures co-complexed with antagonist ligands and some studies, which designate the interactions between N of the heterocycle and ASP351 to be responsible for the antiestrogenic activity [48–51]. Our results also agree with previous studies, which reported that the optimal length of the link is represented by two carbons [28]. These results suggest that these models can be used in structure-based drug design and should be useful for the design of novel structurally related ER antagonists.

The predictive ability of the CoMFA models was assessed by predicting biological activities of external test sets. Prior to prediction, the test set compounds were processed identically to the training set compounds, as previously described. The external validation process can be considered the most valuable validation method, as these compounds were completely excluded during the training of the models. The results are listed in Tables 3 and 4, and the graphic results simultaneously displayed in Fig. 4. The predicted values fall close to the experimental pIC₅₀ values, not deviating by more than 0.42 log units. No outliers were detected in this series of ER ligands. The good agreement between actual and predicted pIC₅₀ values for test set compounds in both 3D QSAR models suggests that the constructed models are reliable and can be used for the design of modulators with improved properties.

Table 3

Experimental and predicted activities (pIC₅₀) for the ligand-binding assay with residual values for 28 test compounds

Test set compounds	Experimental	Predicted CoMFA	CoMFA residuals	Predicted HQSAR	HQSAR residuals
43	9.10	9.21	−0.11	9.05	0.05
44	8.77	8.60	0.17	8.62	0.15
45	8.36	8.46	−0.1	8.13	0.23
46	9.22	9.56	−0.34	9.25	−0.03
47	9.22	9.24	−0.02	9.14	0.08
48	9.30	9.41	−0.11	9.06	0.24
49	8.66	8.65	0.01	8.42	0.24
50	8.96	8.64	0.32	8.90	0.06
51	9.00	8.83	0.17	9.05	−0.05
52	8.89	8.90	−0.01	8.84	0.05
53	9.16	9.01	0.15	9.11	0.05
54	9.30	9.16	0.14	9.11	0.19
55	8.60	8.86	−0.26	8.59	0.01
56	8.85	9.14	−0.29	9.04	−0.19
57	9.16	9.19	−0.03	9.09	0.07
116	6.28	5.97	0.31	6.65	−0.37
117	6.98	6.77	0.21	6.85	0.13
118	6.83	6.99	−0.16	6.72	0.11
119	5.98	5.94	0.04	6.35	−0.37
120	8.51	8.93	−0.42	8.34	0.17
121	8.13	8.07	0.06	8.43	−0.3
122	7.46	7.29	0.17	7.33	0.13
123	8.16	8.55	−0.39	8.15	0.01
124	7.96	8.00	−0.04	8.31	−0.35
125	8.52	8.64	−0.12	8.60	−0.08
126	8.41	8.66	−0.25	8.54	−0.13
127	8.70	8.88	−0.18	8.51	0.19
128	8.27	8.59	−0.32	8.36	−0.09

CoMFA and HQSAR models.

3.3. HQSAR models

HQSAR investigations require the variation of parameters that specify the length, size and type of fragment that are to be encoded in the hologram. In the present paper, several combinations of these parameters were considered during the optimization of the models as previously described [34,39]. The generation of molecular fragments was initially carried out using the fragment size default (4–7) and several fragment distinctions, over the 12 default series of hologram lengths. The

statistical results from the PLS analyses for affinity (data set 1) and potency (data set 2) are presented in Tables 5 and 6, respectively. As it can be seen, the best statistical results concerning the affinity model were obtained using A/B/C/DA as distinction information ($q^2 = 0.71$, and $r^2 = 0.92$, with a S.E.E. of 0.27). On the other hand, the best statistical results for the potency model were obtained using A/B/DA as distinction information ($q^2 = 0.74$, and $r^2 = 0.92$, with a S.E.E. of 0.28).

The influence of different fragment sizes in the statistical parameters was further investigated for the best HQSAR

Table 4

Experimental and predicted activities (pIC₅₀) for the inhibition of MCF-7 cell growth assay with residual values for 14 test compounds

Test set compounds	Experimental	Predicted CoMFA	CoMFA residuals	Predicted HQSAR	HQSAR residuals
56	8.17	8.32	−0.15	8.33	−0.16
57	9.30	9.05	0.25	9.13	0.17
58	8.47	8.41	0.06	8.83	−0.36
59	7.55	7.64	−0.09	7.44	0.11
60	6.91	6.89	0.02	6.94	−0.03
61	6.95	6.94	0.01	6.70	0.25
62	8.59	8.50	0.09	8.71	−0.12
63	7.95	7.91	0.04	8.07	−0.12
64	9.05	8.88	0.17	8.94	0.11
65	9.00	8.84	0.16	8.94	0.06
66	8.80	8.91	−0.11	8.79	0.01
67	8.42	8.64	−0.22	8.66	−0.24
68	9.40	9.43	−0.03	9.43	−0.03
69	8.51	8.36	0.15	8.73	−0.22

CoMFA and HQSAR models.

Fig. 4. Plot of predicted values of pIC_{50} vs. the corresponding experimental values for the training (black squares) and test (gray squares) set compounds for affinity (A) and potency models (B).

Table 5
HQSAR analysis for various fragment distinction on the key statistical parameters using fragment size (4–7)

Fragment distinction	Statistical parameters				
	q^2	r^2	S.E.E.	HL	N
A/B	0.70	0.89	0.32	199	6
A/B/C/H	0.63	0.78	0.44	83	6
A/B/C/H/DA	0.61	0.85	0.37	353	7
A/B/H	0.60	0.752	0.48	401	5
A/B/H/DA	0.57	0.77	0.45	307	6
A/B/C/DA	0.71	0.92	0.27	257	7

Model HQSAR for ligand-binding assay. q^2 , cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; S.E.E., non-cross-validated standard error; HL, hologram length; N , optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; DA, donor and acceptor.

Table 6

HQSAR analysis for various fragment distinction on the key statistical parameters using fragment size (4–7)

Fragment distinction	Statistical parameters				
	q^2	r^2	S.E.E.	HL	N
A/B	0.72	0.93	0.26	353	7
A/B/C/H	0.69	0.89	0.33	53	7
A/B/C/H/DA	0.71	0.82	0.40	53	4
A/B/H	0.71	0.91	0.29	199	7
A/B/H/DA	0.62	0.90	0.31	151	6
A/B/DA	0.74	0.92	0.28	97	7
A/B/H/DA	0.68	0.89	0.33	97	7

Model HQSAR for inhibition of MCF-7 cell growth assay. q^2 , cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; S.E.E., non-cross-validated standard error; HL, hologram length; N , optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; DA, donor and acceptor.

Fig. 5. Plot of predicted values of pIC_{50} vs. the corresponding experimental values for the training (black squares) and test (gray squares) set compounds for affinity (A) and potency models (B).

models generated previously. Fragment size parameters control the minimum and maximum lengths of fragments to be included in the hologram fingerprint. The statistical results for the different fragment sizes evaluated (2–5, 3–6, 4–7, 5–8 and 6–9) show no improvements in the HQSAR models. A measure of internal consistency is available in the form of q^2 . However, the most important test of a QSAR model is its ability to predict the property value for new compounds, not included in the training set. As the structure encoded in a 2D fingerprint is directly related to the biological activity of molecules within training set, HQSAR models are also able to predict the activity of new structurally related modulators from its fingerprint. Thus, the predictive power of the best affinity and potency HQSAR models derived using the training sets was assessed by predicting pIC_{50} values for test sets. The results of the external validation are listed in Tables 3 and 4. Again, the good agreement between experimental and predicted pIC_{50} values for the test set compounds indicates the robustness of the HQSAR models. The predicted values fall close to the experimental pIC_{50} values, deviating by no more than 0.37 log units in both models. The graphic results for the experimental versus predicted pIC_{50} values of both training and test sets are displayed in Fig. 5. From the low residual values, it can be seen that the affinity and potency HQSAR models obtained are highly reliable and can be used to predict the biological activity of novel compounds within these structural class. There are no outliers in both sets.

4. Conclusions

The QSAR models described herein exhibited both good internal and external consistency, showing substantial predictive power. The use of 3D structural information is useful in drug design and the CoMFA models generated in our studies are clearly stable and robust. Moreover, the 3D QSAR models are compatible with the 3D protein environment in the ER binding site, and should be useful for the design of novel structurally related ER α modulators. On the other hand, a powerful strategy is to study the integration of 2D and 3D methods in the development of predictive QSAR models. The HQSAR technique requires only 2D structures and biological activity as input, removing the necessity of conformer generation and molecular alignment which makes use of QSAR non-trivial in a number of settings. It is worth noting that we have employed the same training and test sets for all 3D and 2D QSAR analyses. The highly consistent results confirmed that investigations can be carried out concomitantly to search for synergies between 3D and 2D QSAR technologies. These important tools should be useful for the design of novel structurally related ER α antagonists with improved affinity and potency.

Acknowledgments

We gratefully acknowledge financial support from the Brazilian grant agencies FINEP (Research and Projects Financing), CNPq (The National Council for Scientific and

Technological Development) and FAPESP (The State of São Paulo Research Foundation).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.02.001.

References

- [1] S. Nilsson, J.-Å. Gustafsson, Biological role of estrogen and estrogen receptors, *Crit. Rev. Biochem. Mol. Biol.* 37 (2002) 1–28.
- [2] C.K. Osborne, R. Schiff, Estrogen-receptor biology: continuing progress and therapeutic implications, *J. Clin. Oncol.* 23 (2005) 1616–1622.
- [3] C.J. Gruber, W. Tschugguel, C. Schneeberger, J.C. Huber, Production and actions of estrogens, *N. Engl. J. Med.* 346 (2002) 340–352.
- [4] S. Green, P. Walter, V. Kumar, A. Krust, J.M. Bornert, P. Argos, P. Chambon, Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A, *Nature* 320 (1986) 134–139.
- [5] S. Mosselman, J. Polman, R. Dijkema, ER β : identification and characterization of a novel human estrogen receptor, *FEBS Lett.* 392 (1996) 49–53.
- [6] J.-Å. Gustafsson, What pharmacologists can learn from recent advances in oestrogen signaling, *Trends Pharmacol. Sci.* 24 (2003) 479–485.
- [7] H. Gronemeyer, J.-Å. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily, *Nat. Rev. Drug Discov.* 3 (2004) 950–964.
- [8] J.I. MacGregor, V.C. Jordan, Basic guide to the mechanisms of antiestrogen action, *Pharmacol. Rev.* 50 (1998) 151–196.
- [9] V.C. Jordan, Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 1. Receptor interactions, *J. Med. Chem.* 46 (2003) 883–908.
- [10] A.M. Brzozowski, A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.-Å. Gustafsson, M. Carlquist, Molecular basis of agonism and antagonism in the oestrogen receptor, *Nature* 389 (1997) 753–758.
- [11] B.S. Katzenellenbogen, B. Bhardwaj, H. Fang, B.A. Ince, F. Pakdel, J.C. Reese, D.J. Schodin, C.K. Wrenn, Hormone binding and transcription activation by estrogen receptors: analyses using mammalian and yeast systems, *J. Steroid. Biochem. Mol. Biol.* 47 (1993) 39–48.
- [12] J.A. Katzenellenbogen, B.W. O'Malley, B.S. Katzenellenbogen, Tripartite steroid hormone receptor pharmacology: interaction with multiple effector sites as a basis for the cell- and promoter-specific action of these hormones, *Mol. Endocrinol.* 10 (1996) 119–131.
- [13] M. Dutertre, C.L. Smith, Molecular mechanisms of selective estrogen receptor modulators (SERM) action, *J. Pharmacol. Exp. Ther.* 295 (2000) 431–437.
- [14] S.S. Legha, S.K. Carter, Antiestrogens in the treatment of breast cancer, *Cancer Treat. Rev.* 3 (1976) 205–216.
- [15] L.J. Lerner, V.C. Jordan, The development of antiestrogens for the treatment of breast cancer, *Cancer Res.* 50 (1990) 4177–4189.
- [16] E.A. Ariazi, J.L. Ariazi, F. Cordera, V.C. Jordan, Estrogen receptors as therapeutic targets in breast cancer, *Curr. Top. Med. Chem.* 6 (2006) 181–202.
- [17] V.C. Jordan, Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 2. Clinical considerations and new agents, *J. Med. Chem.* 46 (2003) 1081–1111.
- [18] H.J. Tagnon, Antiestrogens in treatment of breast cancer, *Cancer* 39 (1977) 2959–2964.
- [19] A.E. Wakeling, M. Dukes, J. Bowler, A potent specific pure antiestrogen with clinical potential, *Cancer Res.* 51 (1991) 3867–3873.
- [20] Y. Shang, Molecular mechanisms of oestrogen and SERMs in endometrial carcinogenesis, *Nat. Rev. Cancer* 6 (2006) 360–368.
- [21] H.Y. Chen, K.D. Dykstra, E.T. Birzin, K. Frisch, W. Chan, Y.T. Yang, R.T. Mosley, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 1. The discovery of flavanoids with subtype selectivity, *Bioorg. Med. Chem. Lett.* 14 (2004) 1417–1421.

- [22] S. Kim, J.Y. Wu, E.T. Birzin, K. Frisch, W. Chan, L.Y. Pai, Y.T. Yang, R.T. Mosley, P.M. Fitzgerald, N. Sharma, J. Dahllund, A.G. Thorsell, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen Receptor Ligands. II. Discovery of benzoxathiins as potent, selective estrogen receptor α modulators, *J. Med. Chem.* 47 (2004) 2171–2175.
- [23] H.Y. Chen, S. Kim, J.Y. Wu, E.T. Birzin, W. Chan, Y.T. Yang, J. Dahllund, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 3. The SAR of dihydrobenzoxathiin SERMs, *Bioorg. Med. Chem. Lett.* 14 (2004) 2551–2554.
- [24] S. Kim, J. Wu, H.Y. Chen, E.T. Birzin, W. Chan, Y.T. Yang, L. Colwell, S. Li, J. Dahllund, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 4. The SAR of the syn-dihydrobenzoxathiin SERAMs, *Bioorg. Med. Chem. Lett.* 14 (2004) 2741–2745.
- [25] Q. Tan, E.T. Birzin, W. Chan, Y. Tien Yang, L.Y. Pai, E.C. Hayes, C.A. DaSilva, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 5. The SAR of dihydrobenzoxathiins containing modified basic side chains, *Bioorg. Med. Chem. Lett.* 14 (2004) 3747–3751.
- [26] Q. Tan, E.T. Birzin, W. Chan, Y.T. Yang, L.Y. Pai, E.C. Hayes, C.A. DaSilva, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 6. Synthesis and binding affinity of dihydrobenzodithiins, *Bioorg. Med. Chem. Lett.* 14 (2004) 3753–3755.
- [27] T.A. Blizzard, F. DiNinno, J.D. Morgan II, H.Y. Chen, J.Y. Wu, C. Gude, S. Kim, W. Chan, E.T. Birzin, Y. Tien Yang, L.Y. Pai, Z. Zhang, E.C. Hayes, C.A. DaSilva, W. Tang, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 7. Dihydrobenzoxathiin SERAMs with bicyclic amine side chains, *Bioorg. Med. Chem. Lett.* 14 (2004) 3664–3861.
- [28] T.A. Blizzard, F. DiNinno, J.D. Morgan II, J.Y. Wu, H.Y. Chen, S. Kim, W. Chan, E.T. Birzin, Y.T. Yang, L.Y. Pai, Z. Zhang, E.C. Hayes, C.A. DaSilva, W. Tang, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 8. Dihydrobenzoxathiin SERAMs with heteroatom-substituted side chains, *Bioorg. Med. Chem. Lett.* 14 (2004) 3865–3868.
- [29] T.A. Blizzard, F. Dininno, J.D. Morgan II, H.Y. Chen, J.Y. Wu, S. Kim, W. Chan, E.T. Birzin, Y.T. Yang, L.Y. Pai, P.M. Fitzgerald, N. Sharma, Y. Li, Z. Zhang, E.C. Hayes, C.A. DaSilva, W. Tang, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 9. Dihydrobenzoxathiin SERAMs with alkyl substituted pyrrolidine side chains and linkers, *Bioorg. Med. Chem. Lett.* 15 (2005) 107–113.
- [30] Q. Tan, T.A. Blizzard, J.D. Morgan II, E.T. Birzin, W. Chan, Y.T. Yang, L.Y. Pai, E.C. Hayes, C.A. DaSilva, S. Warrier, J. Yudkovitz, H.A. Wilkinson, N. Sharma, P.M. Fitzgerald, S. Li, L. Colwell, J.E. Fisher, S. Adamski, A.A. Reszka, D. Kimmel, F. DiNinno, S.P. Rohrer, L.P. Freedman, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 10. Chromanes: old scaffolds for new SERAMs, *Bioorg. Med. Chem. Lett.* 15 (2005) 1675–1681.
- [31] T.A. Blizzard, F. DiNinno, H.Y. Chen, S. Kim, J.Y. Wu, W. Chan, E.T. Birzin, Y.T. Yang, L.Y. Pai, E.C. Hayes, C.A. DaSilva, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 13. Dihydrobenzoxathiin SERAMs with an optimized antagonist side chain, *Bioorg. Med. Chem. Lett.* 15 (2005) 3912–3916.
- [32] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Recent advances in comparative molecular field analysis (CoMFA), *Prog. Clin. Biol. Res.* 291 (1989) 161–165.
- [33] W. Tong, D.R. Lowis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, D.M. Sheehan, Evaluation of quantitative structure–activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor, *J. Chem. Inf. Comput. Sci.* 38 (1998) 669–677.
- [34] M.S. Castilho, M.P. Postigo, C.B. de Paula, C.A. Montanari, G. Oliva, A.D. Andricopulo, Two- and three-dimensional quantitative structure–activity relationships for a series of purine nucleoside phosphorylase inhibitors, *Bioorg. Med. Chem.* 14 (2006) 516–527.
- [35] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R.D. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* 267 (1997) 727–748.
- [36] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved protein–ligand docking using GOLD, *Proteins* 52 (2003) 609–623.
- [37] W. Zhu, G. Chen, L. Hu, X. Luo, C. Gui, C. Luo, C.M. Puah, K. Chen, H. Jiang, QSAR analyses on ginkgolides and their analogues using CoMFA, CoMSIA, and HQSAR, *Bioorg. Med. Chem.* 13 (2005) 313–322.
- [38] M.A. Avery, M. Alvim-Gaston, C.R. Rodrigues, E.J. Barreiro, F.E. Cohen, Y.A. Sabnis, J.R. Woolfrey, Structure–activity relationships of the anti-malarial agent artemisinin. 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies, *J. Med. Chem.* 45 (2002) 292–303.
- [39] K.M. Honorio, R.C. Garratt, A.D. Andricopulo, Hologram quantitative structure–activity relationships for a series of farnesoid $\times$ receptor activators, *Bioorg. Med. Chem. Lett.* 15 (2005) 3119–3125.
- [40] M.R. Doddareddy, Y.J. Lee, Y.S. Cho, K.I. Choi, H.Y. Koh, A.N. Pae, Hologram quantitative structure activity relationship studies on 5-HT₆ antagonists, *Bioorg. Med. Chem.* 12 (2004) 3815–3824.
- [41] H. Gao, J.A. Katzenellenbogen, R. Garg, C. Hansch, Comparative QSAR analysis of estrogen receptor ligands, *Chem. Rev.* 99 (1999) 723–744.
- [42] J.H. Poupaert, D.M. Lambert, J. Vamecq, Y.J. Abul-Hajj, Molecular modeling studies on 11 β -aminoethoxyphenyl and 7 α -aminoethoxyphenyl estradiols. Evidence suggesting a common hydrophobic pocket in estrogen receptor, *Bioorg. Med. Chem. Lett.* 5 (1995) 839–842.
- [43] H. Fang, W. Tong, L.M. Shi, R. Blair, R. Perkins, W. Branham, B.S. Hass, Q. Xie, S.L. Dial, C.L. Moland, D.M. Sheehan, Structure–activity relationships for a large diverse set of natural, synthetic, and environmental estrogens, *Chem. Res. Toxicol.* 14 (2001) 280–294.
- [44] C.L. Waller, A comparative QSAR study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds, *J. Chem. Inf. Comput. Sci.* 44 (2004) 758–765.
- [45] C.L. Waller, T.I. Oprea, K. Chae, H.K. Park, K.S. Korach, S.C. Laws, T.E. Wiese, W.R. Kelce, L.E. Gray Jr., Ligand-based identification of environmental estrogens, *Chem. Res. Toxicol.* 9 (1996) 1240–1248.
- [46] I.R. Menezes, A. Leitao, C.A. Montanari, Three-dimensional models of non-steroidal ligands: a comparative molecular field analysis, *Steroids* 71 (2006) 417–428.
- [47] F. Minutolo, M. Antonello, S. Bertini, S. Rapposelli, A. Rossello, S. Sheng, K.E. Carlson, J.A. Katzenellenbogen, M. Macchia, Synthesis, binding affinity, and transcriptional activity of hydroxy- and methoxy-substituted 3,4-diarylsalicylaldoximes on estrogen receptors α and β , *Bioorg. Med. Chem.* 11 (2003) 1247–1257.
- [48] A.S. Levenson, V.C. Jordan, The key to the antiestrogenic mechanism of raloxifene is amino acid 351 (aspartate) in the estrogen receptor, *Cancer Res.* 58 (1998) 1872–1875.
- [49] J.H. Kim, M.H. Lee, B.J. Kim, J.H. Kim, S.J. Han, H.Y. Kim, M.R. Stallcup, Role of aspartate 351 in transactivation and active conformation of estrogen receptor α , *J. Mol. Endocrinol.* 35 (2005) 449–464.
- [50] C. Zhao, J. Abrams, D.F. Skafar, Targeted mutation of key residues at the start of helix 12 in the hER α ligand-binding domain identifies the role of hydrogen-bonding and hydrophobic interactions in the activity of the protein, *J. Steroid Biochem. Mol. Biol.* 98 (2006) 1–11.
- [51] G. Dayan, M. Lupien, A. Auger, S.I. Anghel, W. Rocha, S. Croisietiere, J.A. Katzenellenbogen, S. Mader, Tamoxifen and raloxifene differ in their functional interactions with aspartate 351 of estrogen receptor α , *Mol. Pharmacol.* 70 (2006) 579–588.