

UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA DE SÃO CARLOS

**MODELAGEM MOLECULAR DE UMA SÉRIE DE COMPOSTOS INIBIDORES DA
ENZIMA INTEGRASE DO VÍRUS HIV-1**

LUCIANA LUZIA DE CARVALHO

Tese apresentada ao Instituto de Química de São Carlos, da Universidade de São Paulo para obtenção do título de Doutor em Ciências (Físico-Química).

Orientador: Prof. Dr. Albérico B. F. da Silva

São Carlos

2011

À minha família, com amor.

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos a todos que, de forma direta ou indireta, influenciaram na realização deste trabalho:

Agradeço ao meu orientador, Prof. Dr. Albérico B. F da Silva, pela confiança, paciência, amizade e também por ter me ajudado na minha formação como pesquisadora.

Agradeço à Profa. Dra. Káthia M. Honório e ao Vinícius Maltarollo pelo auxílio na realização dos cálculos computacionais e discussão dos resultados.

Agradeço aos meus pais, Luiz T. V. de Carvalho e Gilda G. D. de Carvalho e à minha irmã Luana Carvalho por todo amor e apoio durante toda a minha vida.

A todos os meus colegas, funcionários e professores do Grupo de Química Estrutural pela amizade, apoio e vários favores já prestados;

A todos os meus queridos amigos que estiveram ao meu lado durante o doutorado: Alessandra, Andressa, Biankita, Camila, Dani, Fernando, Igor, Jana, Káthia, Karen, Layla, Maíra, Manu, Marcel, Paulinha e Shirley.

Agradeço ao CNPq pelo suporte financeiro;

Finalmente, faço um agradecimento especial ao meu marido Thiago Borges, não apenas por todas as formas de amor e carinho demonstradas, mas principalmente por fazer parte de minha vida.

RESUMO

Uma etapa essencial no ciclo de vida do vírus HIV é a integração do DNA viral no cromossomo hospedeiro. Essa etapa é catalisada pela enzima integrase (IN) de 32-kDa. HIV-1 IN é um importante e validado alvo, e as drogas que inibem seletivamente a enzima, quando utilizadas em combinação com os inibidores da transcriptase reversa (RT) e protease (PR), são consideradas altamente eficazes em suprimir a replicação viral. IN catalisa dois processos enzimáticos designados por 3' processamento e transferência de DNA. Agentes ativos contra integrase, inibindo a etapa de transferência da vertente já estão em fase clínica. O fármaco Raltegravir® é o primeiro nesta nova classe. Os ensaios clínicos no tratamento em novos pacientes têm uma atividade anti-retroviral potente e bem tolerado. Dada a sua potência, segurança e novo mecanismo de ação, os inibidores da integrase representam um importante avanço terapêutico contra o HIV-1. Na presente tese de doutorado, foram realizados estudos quimiométricos utilizando descritores teóricos e QSAR bi- (2D) e tridimensionais (3D) empregando, respectivamente, as técnicas holograma QSAR (HQSAR) e a análise comparativa dos campos moleculares (CoMFA), visando à geração de modelos preditivos para um conjunto de inibidores da integrase do vírus HIV-1. Modelos de QSAR com boa consistência interna, habilidade preditiva e estabilidade foram obtidos em todos os casos. Os modelos gerados, associados às informações obtidas pelos mapas de contribuição 2D e de contorno 3D, são guias químico-medicinais úteis no planejamento de novos inibidores mais potentes e seletivos da integrase do HIV-1.

ABSTRACT

An essential step in the HIV life cycle is integration of the viral DNA into the host chromosome. This step is catalyzed by a 32-kDa viral enzyme HIV integrase (IN). HIV-1 IN is an important and validated target, and the drugs that selectively inhibit this enzyme, when used in combination with reverse transcriptase (RT) and protease (PR) inhibitors, are believed to be highly effective in suppressing the viral replication. IN catalyzes two discrete enzymatic processes referred as 3' processing and DNA strand transfer. Agents active against HIV-1, which target the viral integrase by inhibiting the strand transfer step of integration, have now initialized the clinical trials. The Raltegravir® is the first drug in this new class. Clinical trials in treatment-experienced and in treatment-naive patients have shown that raltegravir-containing regimens have potent antiretroviral activity and are well tolerated. Given their potency, safety and novel mechanism of action, integrase inhibitors represent an important advance in HIV-1 therapy. In the present thesis, Bi- and Tridimensional Quantitative Structure-Activity Relationship (QSAR) studies were performed applying chemometric methods based on theoretical descriptors, Comparative Molecular Field Analysis (CoMFA) and Holograma QSAR (HQSAR) techniques, aiming to generate predictive models for a large set of HIV-1 IN inhibitors. QSAR models presenting good internal consistency, predictive power and stability were obtained in all cases. The final models along with the information resulted by 2D contribution and 3D contour maps should be useful in the design of new inhibitors with increased potency and selective within the chemical diversity of the data.

SUMÁRIO

1. INTRODUÇÃO	8
1.1. Síndrome da imunodeficiência adquirida e suas características.....	9
1.2. Ciclo de vida do vírus HIV-1.....	10
1.3. Terapia farmacológica.....	13
1.4. Estrutura da enzima integrase (IN) do vírus HIV-1 e seus domínios funcionais .	14
1.5. Mecanismo de integração	16
1.6. Inibidores da integrase do vírus HIV-1	18
1.7. Estrutura cristalográfica da enzima integrase com o inibidor 5CITEP	21
2. OBJETIVOS	23
2.1. Objetivos Específicos.....	23
3. METODOLOGIA.....	25
3.1. Métodos computacionais	25
3.1.1. Campo de Força em Mecânica Molecular	25
3.1.2. A Função da Energia Potencial	25
3.1.3. Otimização da Geometria Molecular.....	28
3.1.4. Mecânica Quântica	30
3.1.4.1. Teoria do Orbital Molecular	31
3.1.4.2. Métodos Semi-empíricos	35
3.1.4.3. Teoria do Funcional de Densidade.....	37
3.2. Química Medicinal e as Técnicas de QSAR	39
3.2.1. QSAR.....	40
3.2.1.1 Descrição dos parâmetros moleculares calculados.....	40
3.2.2. Docking Molecular	44
3.2.3. Análise Comparativa dos Campos Moleculares (CoMFA)	46
3.3.3. Holograma QSAR	49
4. ESTUDO QUIMIOMÉTRICO DOS COMPOSTOS 1,3 DICETOÁCIDOS E SEUS ANÁLOGOS 8-hidroxi -(1,6)-nafititidina-7-carboxiamida.....	54
4.1. Procedimento Computacional	54

4.2. Otimização das geometrias e cálculo dos descritores moleculares	58
4.3 Resultados e Discussão.....	61
4.3.1. Seleção de variáveis.....	61
4.3.2. Regressão PLS.....	62
4.3.3. Relação entre orbitais moleculares e valores logIC ₅₀	65
4.4. Conclusões	67
5. ESTUDO DAS RELAÇÕES QUANTITATIVAS ENTRE A ESTRUTURA E ATIVIDADE DOS INIBIDORES DA INTEGRASE CONTENDO O GRUPO N-Alquil-5- hidroxipirimidinona	69
5.1. Conjunto de dados.....	69
5.2. Docking Molecular	79
5.3. QSAR 3D	85
5.3.1. Procedimento computacional.....	85
5.3.2. Resultados e Discussão	87
5.4. HQSAR	92
5.4.1. Procedimento Computacional.....	93
5.4.2. Resultados e Discussão	95
5.5. Relação entre orbitais moleculares e valores logIC ₅₀	101
5.6. Conclusões	103
6. CONCLUSÕES FINAIS.....	105
REFERÊNCIAS.....	107

1. INTRODUÇÃO

A Química Computacional emergiu como uma área de pesquisa de grande interesse, tornando-se uma valiosa ferramenta para auxiliar nos estudos experimentais. A relação entre estudos computacionais com estudos experimentais é mutuamente benéfica. Técnicas computacionais são capazes de oferecer propriedades termodinâmicas e estruturais de sistemas moleculares, tornando possível o planejamento e a descoberta de novas moléculas bioativas (fármacos). Este avanço na Química Computacional foi impulsionado pelo grande e rápido desenvolvimento dos recursos computacionais, os quais diminuíram o tempo de processamento de dados, isto é, o tempo de cálculo, bem como o avanço nos programas disponíveis para cálculos dos parâmetros.¹

Neste sentido, a mecânica clássica e a química quântica têm auxiliado de modo significativo no estudo de moléculas com atividade biológica e de biomoléculas com milhares de átomos, como proteínas. Em particular, a bioinformática, anteriormente chamada de farmacologia quântica, vem sendo utilizada no cálculo da estrutura eletrônica de compostos no qual, juntamente com a termodinâmica estatística, com a integração da linguagem e de conceitos químicos, permitem a interpretação e a racionalização de propriedades macroscópicas com fundamentos em nível atômico-molecular.²

No presente trabalho, utilizaram-se diferentes técnicas computacionais com o intuito de relacionar propriedades químicas quânticas com a atividade de moléculas candidatas a novos fármacos. Para isso, foram abordados os métodos de quimiometria,

docking molecular, QSAR 2D e 3D para o estudo da relação estrutura atividade de uma séries de inibidores da enzima Integrase do vírus HIV-1.

1.1. Síndrome da imunodeficiência adquirida e suas características

Segundo o último relatório global da UNAIDS (Agência das Nações Unidas para o combate a AIDS), mais de 30 milhões de pessoas morreram desde que a epidemia começou há duas décadas (3,1 milhões só no ano passado), e aproximadamente 40 milhões de pessoas estão infectadas atualmente. Estima-se que em 2025 haverá 90 milhões apenas no continente africano. Em países pobres, 6 milhões de pessoas com HIV/AIDS precisam imediatamente de tratamento antiretroviral e mulheres e crianças fazem parte de uma grande proporção desses que precisam de cuidado, tratamento e suporte.³

Dada a sua rápida expansão e concentração nos países e comunidades mais pobres, a AIDS foi considerada como um problema de segurança mundial pelo Conselho de Segurança da ONU e os recursos econômicos internacionais destinados estão longe de satisfazer as demandas provenientes desta epidemia.

O sucesso dos coquetéis multidrogas, usando os inibidores de protease e transcriptase reversa, tem alterado o curso da AIDS pela supressão da replicação viral ao longo dos anos. Entretanto, problemas como toxicidade e resistências a essas drogas tornam imperativo o rápido desenvolvimento de novos fármacos anti-HIV³. A enzima HIV-1 integrase se tornou um alvo importante no desenvolvimento de fármacos anti-HIV devido ao seu papel fundamental exercido no ciclo de replicação viral.⁴⁻⁷

Entretanto, resultados do desenvolvimento de fármacos, baseados na estrutura do alvo, tornaram-se difíceis pela falta de informações estruturais, e especialmente no que se refere às propriedades eletrônicas que dão origem a atividade inibidora. Logo, tornou-se necessário a união e a racionalização de resultados experimentais com estudos teóricos no sentido de se obter relações preditivas que possam guiar a otimização do potencial de novos compostos inibidores, como fármacos anti-HIV.

1.2. Ciclo de vida do vírus HIV-1

A síndrome de imunodeficiência adquirida (AIDS, do inglês, *Acquired Immune Deficiency Syndrome*), é uma doença pandêmica que tem como agente etiológico o vírus HIV-1 (*Human Immunodeficiency Virus*), um retrovírus pertencente à subfamília Lentivirinae⁸. O vírus HIV possui duas formas geneticamente diferentes, mas relacionadas, o HIV-1 e o HIV-2. O HIV-1 é o tipo mais comum nas Américas e Europa, enquanto o HIV-2 é mais encontrado na África.⁹

Como a maioria dos retrovírus, o HIV é esférico, sendo envolvido por um envelope constituído por uma bicamada lipídica derivada da membrana celular de células do hospedeiro. O envelope é salpicado por duas glicoproteínas virais denominadas gp120 e gp41. No interior do capsídeo, está contido o material genético do vírus, que consiste de duas cópias de RNA, e as enzimas protease, transcriptase reversa (RT) e integrase (IN).

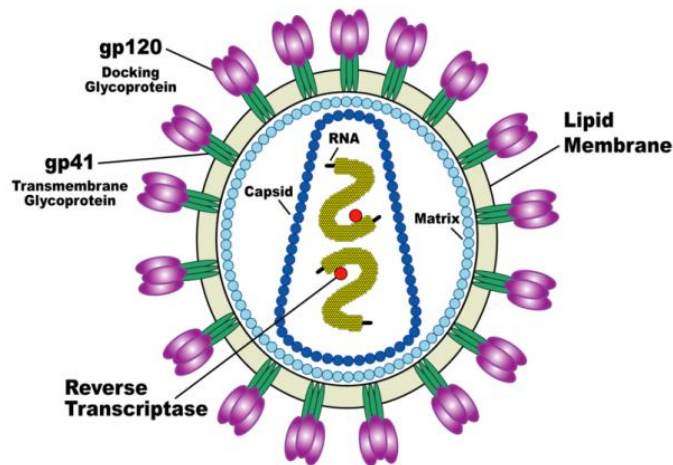


Figura 1. Vírus HIV.

O mapa genético do HIV consiste de três regiões principais: (1) região gag, que codifica as proteínas estruturais internas; (2) região pol (de poliproteína), que codifica as enzimas protease (PR), transcriptase reversa (RT) e integrase (IN) e (3) região env, responsável pela codificação das glicoproteínas do envoltório, gp120 (SU) e gp41 (TM), os quais fazem parte da formação do envelope viral.¹⁰

O alvo principal do vírus HIV, durante a infecção, é o linfócito TCD4+. A característica mais marcante da infecção é a depleção seletiva de linfócitos CD4+, porém macrófagos, monócitos, células de Langerhans são outros exemplos que também são susceptíveis à infecção por HIV¹¹. O início da infecção ocorre com a ligação do vírus à receptores específicos na superfície da célula alvo. Após a ligação com a membrana celular, a proteína gp120 dissocia-se da proteína gp41, que passa por modificações conformacionais, as quais promovem a fusão vírus-célula, permitindo, assim, a entrada do capsídeo na célula. Em seguida, há a liberação no citoplasma do conteúdo do capsídeo, o RNA genômico e enzimas virais, o que se faz necessário para

a etapa posterior, a transcrição reversa. A transcriptase reversa (RT) produz uma molécula de cDNA (DNA complementar) a partir da cópia de uma molécula de RNA.¹²

A integrase, a terceira enzima viral, catalisa duas etapas no ciclo de replicação viral, a clivagem endonucleolítica dos 3'-terminal do c-DNA viral e o a etapa de Transferência de Fita. *In vivo*, a integração ocorre dentro de um grande complexo de nucleoproteínas, chamado de complexo de pré-integração (PIC), ilustrado na Figura 2. Após a transcrição reversa no citoplasma, o cDNA viral está associada com a integrase no PIC até a translocação nuclear e integração dentro do cromossomo hospedeiro. A translocação do PIC para o núcleo é provavelmente alcançado através de interações com a rede de microtúbulos. Esse mecanismo de importação nuclear do PIC ainda não foi completamente elucidado¹³. Pelo ciclo de vida do vírus HIV, ilustrado na Figura 2, é possível verificar a formação do PIC e a sua entrada no núcleo na célula hospedeira. No PIC, um tetrâmero da integrase liga os dois LTR (do inglês, *Long Terminal Repeats*) do DNA viral ao DNA hospedeiro.

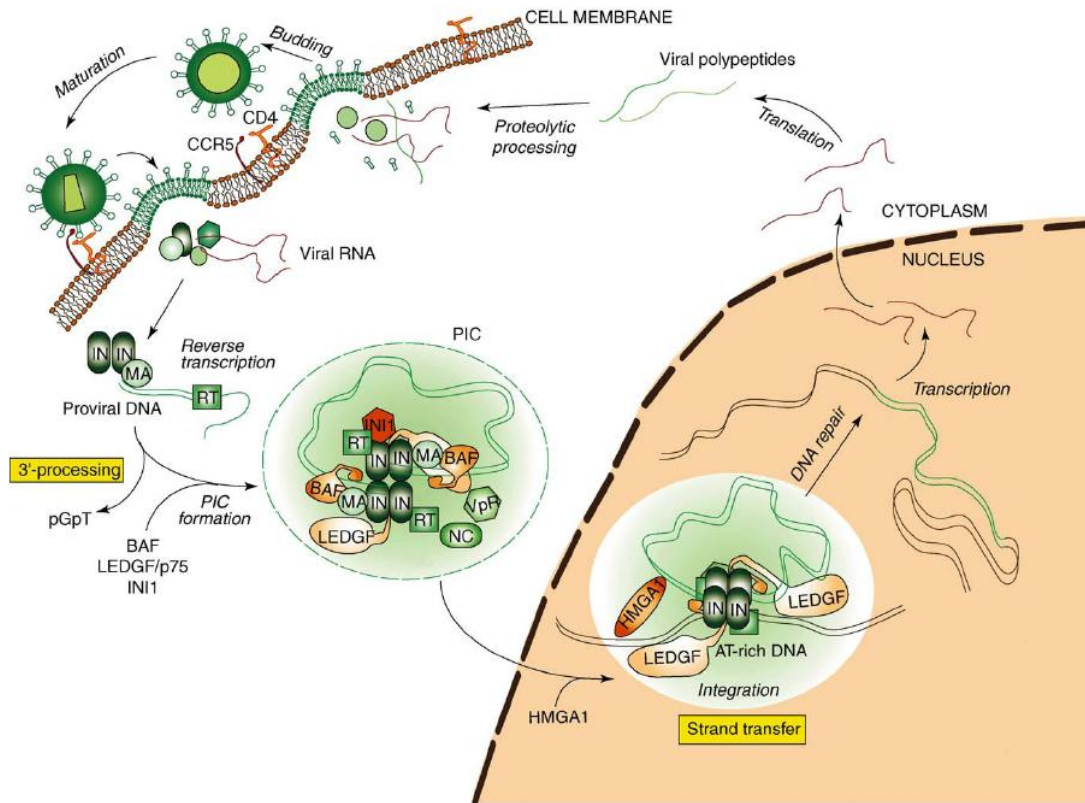


Figura 2. Representação do ciclo do vírus HIV.¹³

1.3. Terapia farmacológica

As terapias aprovadas pela agência norte americana FDA (*Food and Drug Administration*) atualmente consistem de inibidores para quatro etapas do ciclo do HIV-1: transcrição reversa, fusão, maturação proteolítica e integrase. A terapia tripla, comumente chamada de HAART (do inglês, *Highly Active Antiretroviral Therapy*) é um padrão para o tratamento e consiste de um inibidor para a protease (PI), ou um inibidor para a transcriptase reversa não nucleosídeo (NNRTI) em combinação com dois inibidores para a transcriptase reversa nucleosídeo (NRTI). Entretanto, esse tratamento não é bem tolerado pelos pacientes, requer muita disciplina e também é um tratamento

muito caro⁴. Por esse motivo, novos tratamentos estão sendo estudados visando minimizar esses problemas.

1.4. Estrutura da enzima integrase (IN) do vírus HIV-1 e seus domínios funcionais

A HIV-1 Integrase (IN) é uma enzima de 32 kDa e 288 aminoácidos que possui três domínios distintos¹⁴: (i) um domínio N-terminal (resíduos de aminoácidos 1-50), onde se encontram dois resíduos de histidina e dois de cisteína que promovem a ligação com Zn^{2+} ; (ii) o domínio central catalítico (resíduos de aminoácidos 50-212) que contém os aminoácidos ácidos altamente conservados Asp64, Asp116 e Glu52, onde se ligam 2 íons de Mg^{2+} , chamada de padrão D,D-35-E; (iii) um domínio C-terminal (aminoácidos 213-288) que contém resíduos de aminoácidos básicos e liga-se ao DNA. A estrutura cristalográfica de cada um desses domínios foi determinada isoladamente ou em forma de dímeros por difração de raios-X e ressonância magnética nuclear.¹⁴⁻¹⁶

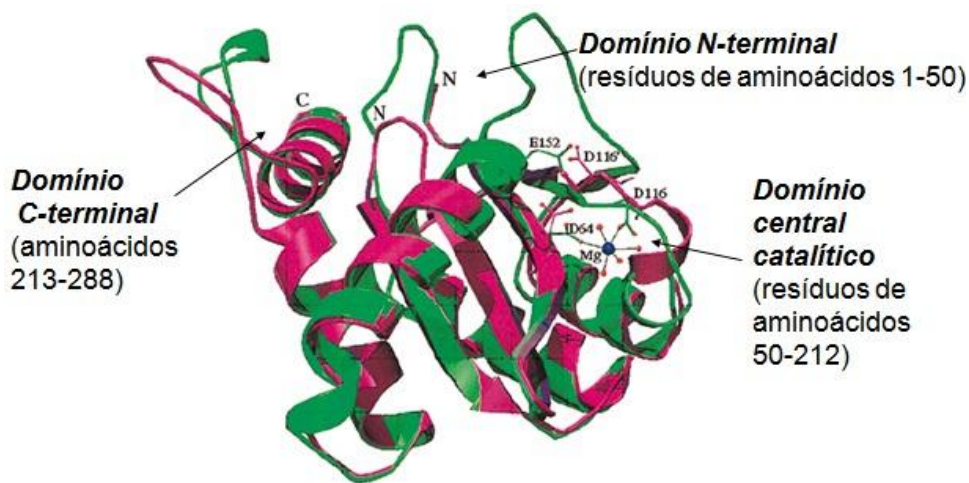


Figura 3. Domínios da enzima integrase do vírus HIV-1.

Análises mostram que o domínio central compreendido pelo padrão (D,D-35-E), D64, D116, E152 é crucial para a atividade da enzima¹⁵. Em uma estrutura de raios-X do domínio central, verificou-se que um íon Mg^{2+} é quelado pela Asp 64 e Asp 116 da cadeia lateral^{15,16}. Acredita-se que um segundo metal pode ser coordenado pelo glutamato E152 juntamente com os resíduos Asp 64 e Asp 116.

Tem sido proposto que um tetrâmero da enzima integrase é a forma requerida para a reação de integração. Porém, o arranjo cristalográfico dessa estrutura permanece desconhecido, uma vez que estudos de sua estrutura completa são dificultados pela propensão da enzima em formar grandes agregados sobre condições reacionais¹⁶. Vários modelos desse tetrâmero com as fitas do DNA viral e hospedeiro foram construídos por meio de modelagem por homologia, utilizando enzimas da mesma família como a Tn5 transposase^{17,18}. Através desses modelos é possível entender como a integrase integra o DNA viral no DNA hospedeiro, como ilustra a Figura 4.

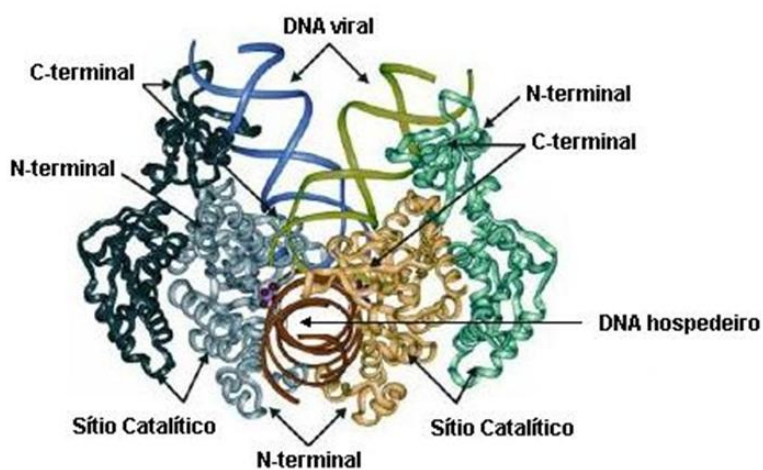


Figura 4. Forma ativa da enzima integrase proposta obtida por meio de modelagem por homologia.¹⁸

1.5. Mecanismo de integração

A enzima integrase catalisa duas etapas no ciclo de replicação viral. Primeiramente, ela catalisa a clivagem endonucleolítica do 3'-terminal do c-DNA viral. Em seguida, a integrase permanece ligada num complexo com o cDNA final, chamado de complexo de pré-integração (PICs), catalisando a inserção do cDNA viral no cromossomo hospedeiro. Este processo é chamado de Etapa de Transferência de Fita.^{6,16}

Na etapa de transferência de fita ocorre uma reação de transesterificação, em que o átomo de oxigênio 3'-terminal do DNA viral clivado ataca a ligação fosfodiéster do DNA hospedeiro através dos dois íons de Mg^{2+} . Acredita-se que essas reações sejam catalisadas em um único sítio ativo no núcleo da enzima. A Figura 5 ilustra essas duas etapas.

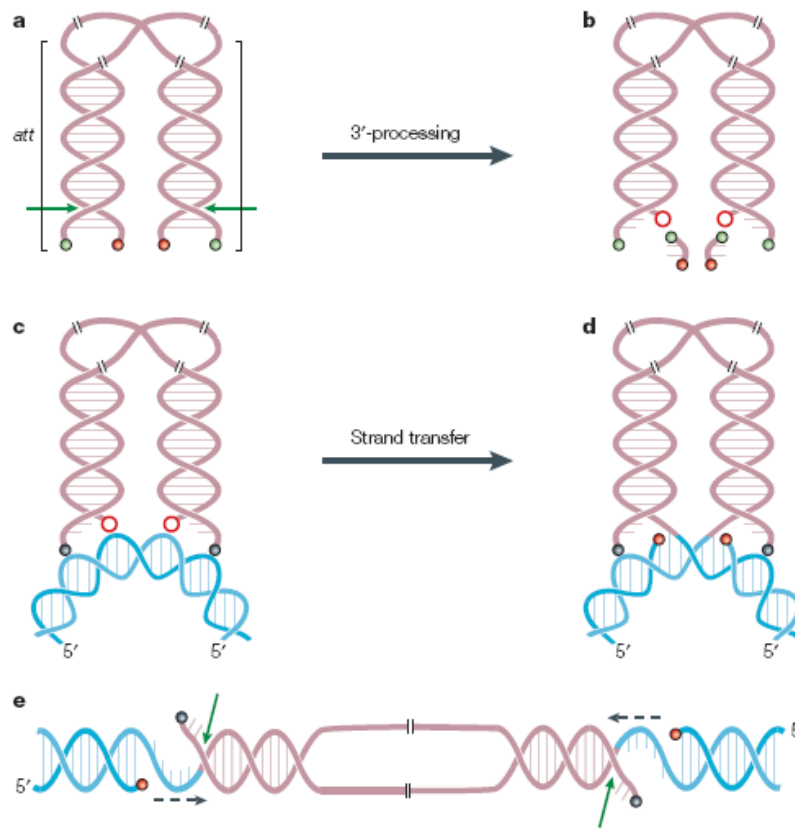


Figura 5. Reações catalisadas pela integrase (clivagem do 3'-terminal e transferência de fita)¹⁶. A figura ilustra os sítios de recombinação do DNA viral (*att*). (a) As flechas verdes mostram a região 3'-terminal onde ocorrerá a clivagem endonucleolítica; (b) O processo de clivagem gera 3'-hidroxilas reativas no final do DNA viral (círculos vermelhos); (c-d) A segunda reação catalisada pela integrase é a transferência de fita, que insere ambos os finais do DNA viral no cromossomo da célula hospedeira (DNA acceptor em azul). Cada 3'-hidroxila final do DNA viral ataca uma ligação fosfodiéster de cada fita do DNA acceptor. (e) DNA viral integrado no DNA hospedeiro.

A Figura 6 ilustra a reação de transesterificação catalisada pela enzima integrase no processo de transferência de fita.

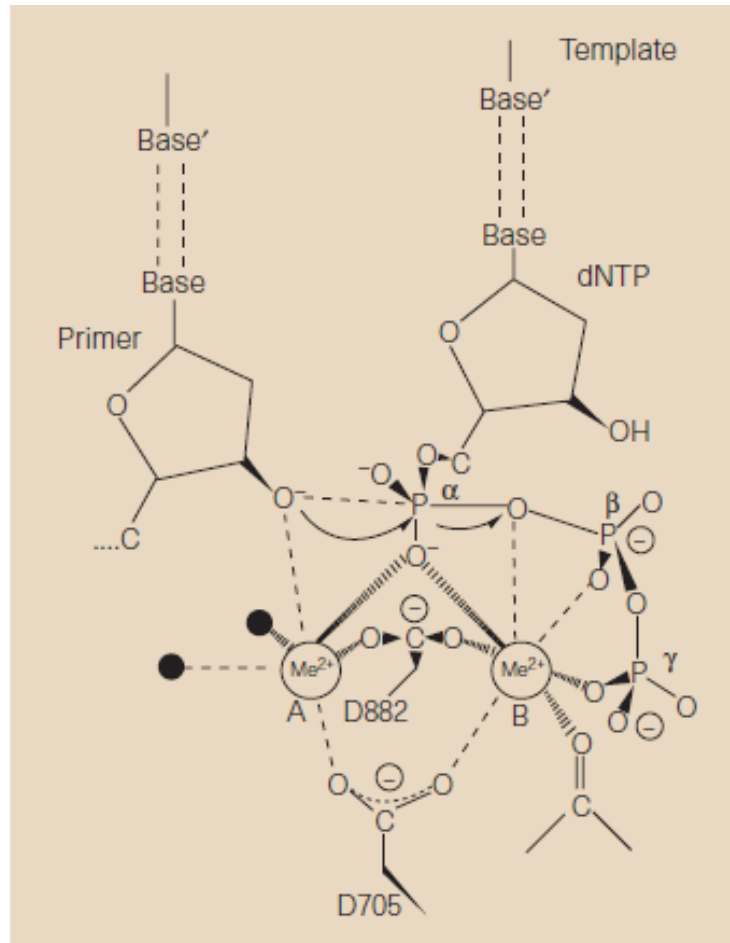


Figura 6. Reação de transesterificação catalisada pela enzima no processo de Transferência de Fita.

1.6. Inibidores da integrase do vírus HIV-1

Os primeiros inibidores seletivos da integrase HIV-1 foram sintetizados a partir dos derivados do 1,3 β -dicetoácido (DKAs). Historicamente, as duas primeiras moléculas de DKAs são relatadas como sendo 5CITEP, sintetizado pela empresa Shionogi & Co e L-731.988, pertencentes ao Grupo Merck e Co. Essa classe de inibidores é responsável pela inibição da etapa de transferência de fita. Neste processo, esses compostos ligam-se na interface do complexo integrase-DNA, diminuindo efetivamente a infecção pelo vírus HIV-1. Por apresentar alta seletividade e

inibição da atividade viral em escala nanomolar, esses compostos tornaram-se importantes agentes quimoterapêuticos¹⁹. A Figura 7 apresenta o mecanismo proposto para ação dos dicetoácidos e seus análogos.

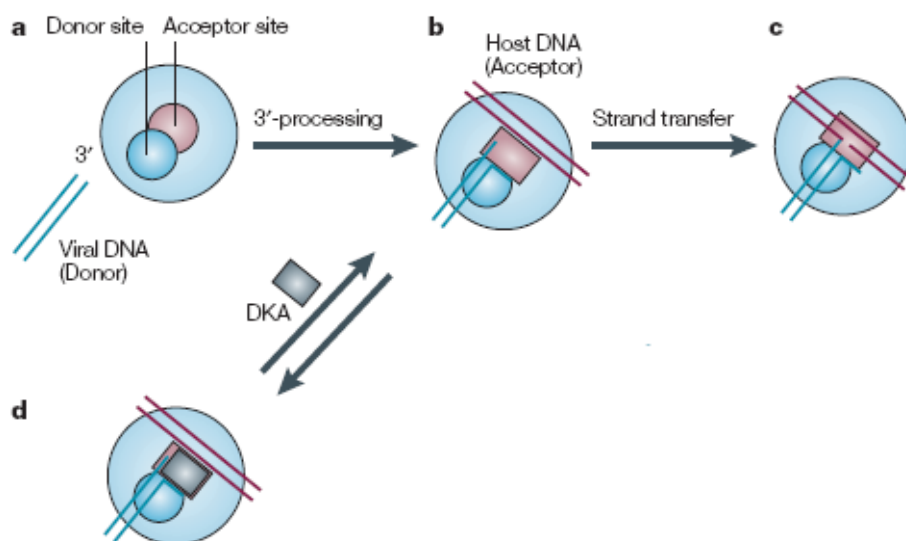


Figura 7. Mecanismo proposto para ação dos dicetoácidos¹⁶. A parte **a** mostra que a integrase possui dois sítios de ligação: um sítio doador para o DNA viral (círculo azul) e um sítio aceitador para o DNA hospedeiro (círculo vermelho). A parte **b** mostra os eventos após o processo de clivagem: o complexo DNA-integrase sofre uma mudança estrutural (retângulo vermelho) para que DNA hospedeiro se ligue ao complexo. Na parte **c**, em condições normais, a ligação do DNA hospedeiro no sítio aceitador conduz a transferência de fita. Na parte **d**, a molécula de dicetoácido (retângulo cinza) após o processo de clivagem, liga-se na interface do complexo DNA-integrase inibindo a transferência de fita.

A funcionalidade do grupo 1,3 β -dicetoácido é essencial para atividade inibitória. Sabe-se que a ligação e a atividade desses inibidores são dependentes do metal Mg^{2+} presente no sítio catalítico da enzima, o qual atua como uma ponte de ligação entre a tríade DDE, o cDNA viral e o DNA aceitador (DNA hospedeiro). Acredita-se que os dicetoácidos e seus derivados agem através do grupo carbonil e do oxigênio da hidroxila, quelando os dois íons de Mg^{2+} e estabilizando o complexo macromolecular integrase-DNA viral²⁰ (Figura 8). Por ter esse único mecanismo de ação, os dicetoácidos e seus análogos podem ser chamados de inibidores interfaciais²¹, pois

representam um dos poucos tipos de fármacos conhecidos que atuam em um intermediário de uma reação enzimática.

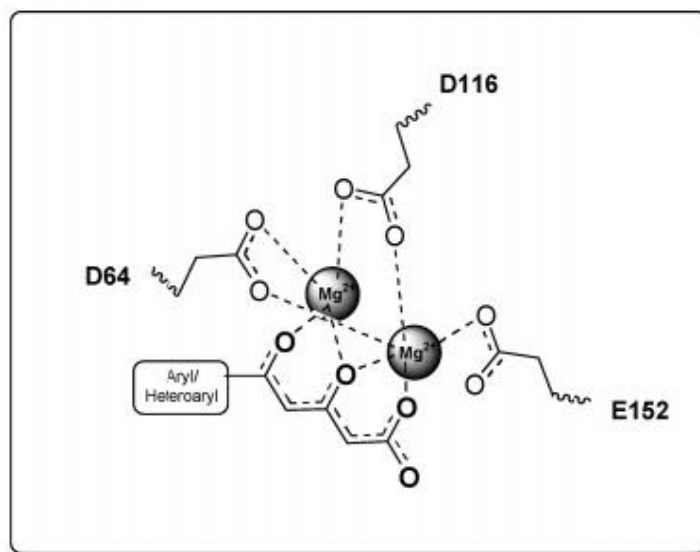


Figura 8. Mecanismo de quelação proposto para os β -dicetoácidos e análogos no sítio ativo da enzima impedindo a etapa de transferência de fita no ciclo de replicação viral²⁰.

Recentemente, foi sintetizado um conjunto de análogos mais potentes derivados dos dicetoácidos contendo os grupos 8-hidróxi-(1,6)-naftitidina-7-carboxiamida²²⁻²⁴ e *N*-Alquill-5-hidroxipirimidinona²⁵⁻²⁹. A presença desses dois grupos farmacofóricos permite que esses novos compostos ajam com o mesmo mecanismo de inibição dos dicetoácidos. A Figura 9 ilustra as estruturas de alguns fármacos contendo os grupos dicetoácido, naftiridina-carboxamida (L-870,810, um dos primeiros a entrar em análise clínica) e *N*-Alquill-5-hidroxipirimidinona (Raltegravir®, fármaco já aprovado pelo FDA em 2007), respectivamente.

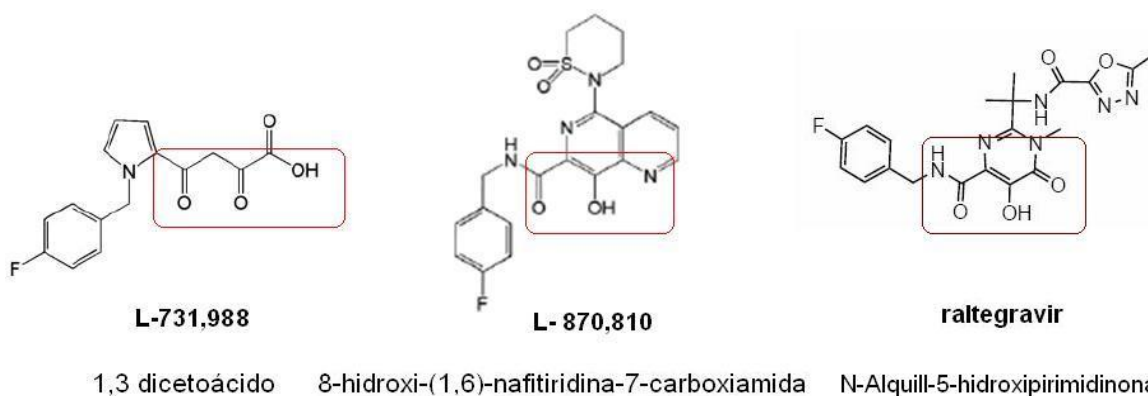


Figura 9. Farmacóforos responsáveis pela ligação com os íons Mg^{2+} no sítio ativo da enzima integrase.

Neste trabalho foram estudados 105 compostos derivados do grupo 1,3 β -dicetoácido¹⁹, 8-hidroxi-(1,6)-nafiridina-7-carboxiamida²²⁻²⁴ e N-Alquil-5-hidroxipirimidinona.²⁵⁻²⁹

1.7. Estrutura cristalográfica da enzima integrase com o inibidor 5CITEP

Atualmente, há somente uma estrutura cristalográfica da enzima integrase complexada com um inibidor no banco de dados PDB (do inglês, **Protein Data Bank**), sob o código 1QS4. Essa estrutura corresponde ao domínio core catalítico complexada com um átomo de Mg^{2+} e o composto dicetoácido 5CITEP.³⁰

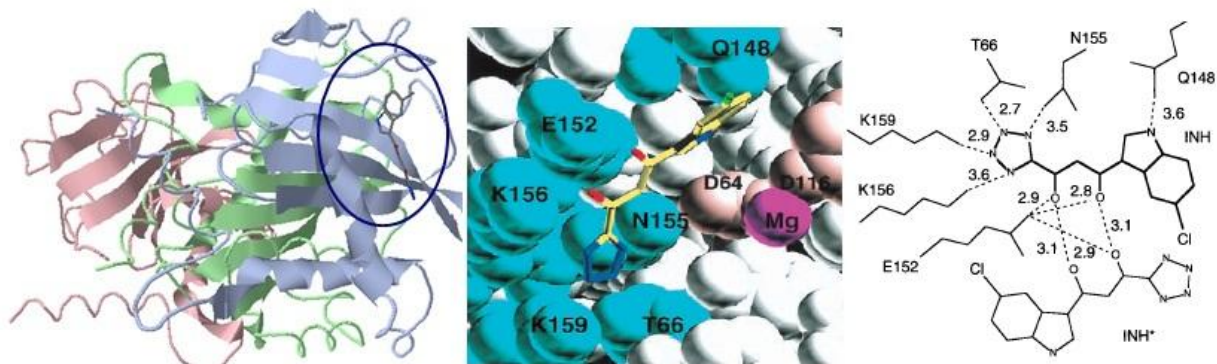


Figura 10. Estrutura cristalográfica da enzima integrase cristalizada com o inibidor *5CITEP*.³⁰

5CITEP é estabilizado no sítio ativo por ligações de hidrogênio com um dos resíduos catalíticos (E152), bem como com outros aminoácidos (T66, N155, Q148) e dois resíduos de lisina (K156 e K159), previamente relatado como sendo envolvidos na ligação com DNA³¹. Porém, na estrutura co-cristal *5CITEP*-IN não há um contato direto entre o composto e os outros dois resíduos catalíticos (D64 e D116), nem uma interação direta com o íon magnésio pode ser detectada. Apenas uma interação de van der Waals entre o anel indólico com uma das quatro moléculas de água de coordenação do íon magnésio pareceu ser plausível.

2. OBJETIVOS

Este trabalho tem como principal objetivo investigar teoricamente as interações observadas experimentalmente entre compostos derivados do grupo 1,3 dicetoácido, 8-*hidroxi-(1,6)-nafitiridina-7-carboxamida* e *N-Alquil-5-hidroxipirimidinona* com a enzima HIV-1 integrase, cuja habilidade está em inibir a integração do DNA viral no DNA das células hospedeiras a fim de possibilitar um maior entendimento das características importantes para tais interações.

2.1. Objetivos Específicos

- ❖ Partindo das estruturas otimizadas dos compostos, calcular propriedades eletrônicas, estéricas, hidrofóbicas (logP) e topológicas para determinar teoricamente quais as características que discriminam os compostos com maior e menor capacidade inibitória em relação à enzima integrase.
- ❖ Obter modelos teóricos da relação qualitativa (SAR) e quantitativa (QSAR) entre as concentrações inibitórias (IC₅₀), determinadas experimentalmente, e as propriedades calculadas que descrevem os aspectos eletrônicos, estéricos e topológicos dos compostos estudados
- ❖ Organizar um conjunto de dados padrão de inibidores da integrase do vírus HIV-1 para o uso da modelagem de QSAR.
- ❖ Obter informações estruturais da ligação desses inibidores com 2 íons Mg²⁺ no sítio ativo da enzima integrase por meio da técnica de *docking* molecular.

- ❖ Desenvolver modelos de QSAR 2D e 3D, empregando as técnicas de HQSAR e CoMFA, respectivamente, para um conjunto padrão de inibidores da integrase do vírus HIV-1.
- ❖ Avaliar a robustez estatística dos modelos gerados, o seu poder de correlação nos conjuntos de treinamento e predição externa nos conjunto teste.
- ❖ Obter informações a partir dos mapas de contribuição 2D gerados pelo método de HQSAR, e dos mapas de contorno 3D, gerados pelo método CoMFA, e utilizá-los como guias em futuros trabalhos em Química Medicinal no planejamento de novos inibidores da integrase do vírus HIV-1 mais potentes e seletivos.

3. METODOLOGIA

3.1. Métodos computacionais

3.1.1. Campo de Força em Mecânica Molecular

Campo de Força em mecânica molecular emprega aproximação empírica simples para descrever uma estrutura molecular. Um campo de força de mecânica molecular pode ser essencialmente dividido em duas partes: a função da energia potencial, um conjunto de equações que descrevem a energia de um sistema como uma função da posição nuclear; e um conjunto de parâmetros que descrevem as preferências e as características de um determinado tipo de átomo. O tipo de átomo define a hibridização e o ambiente local de um átomo. Cada campo de força terá os seus próprios parâmetros e tipos de átomo. O campo de força terá sido parametrizada a partir de dados experimentais ou, no caso de dificuldade para obtenção dos parâmetros, por exemplo, barreiras de torção, a partir de cálculos ab initio.

3.1.2. A Função da Energia Potencial

Campo de força em mecânica molecular reduz a quantidade de cálculos exigidos em relação à mecânica quântica, ignorando os movimentos eletrônicos do sistema. A aproximação de Born-Oppenheimer é utilizada para calcular a energia do sistema como

uma função das posições nucleares. A função da energia potencial é baseada em termos de energia de interação intra e intermoleculares:

$$V_{tot} = V_{ligação} + V_{angulos} + V_{diedros} + V_{improprios} + V_{vdW} + V_{elec} + V_{cross} \quad (1)$$

A função acima inclui estiramento da ligação, ângulo de flexão, barreiras de torção e interações não-ligadas. O termo de estiramento da ligação:

$$V_{ligação} = \sum_{ligação} \frac{K_b}{2} (r - r_0)^2 \quad (2)$$

é um potencial harmônico simples levando ao aumento da energia à medida que o comprimento de ligação r se estende desde o seu comprimento de referência r_0 . A constante de força K_b é geralmente alta, um reflexo do fato de que as ligações são fortes. O termo do ângulo de flexão utiliza um potencial harmônico similar:

$$V_{angulos} = \sum_{angulos} \frac{K_\theta}{2} (\theta - \theta_0)^2 \quad (3)$$

θ e θ_0 são o ângulo e o ângulo de referência, respectivamente. A constante de força do ângulo de flexão K_θ é baixa, pois é necessário menos energia para mudá-los a partir do ângulo de referência. As torções (ângulos diedros) são descritos por um período de expansão cosseno:

$$V_{diedro} = \sum_{diedro} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \quad (4)$$

onde ω é o ângulo de torção e n é a multiplicidade. γ é o fator de fase (ou ângulo), que determina onde o ângulo de torção passa por seu valor mínimo. V_n são constantes que determinam o tamanho da barreira de rotação.

Uma forma diferente, $V_{impr\u00f3rios}$, é necess\u00e1ria para os \u00e2ngulos de tor\u00e7\u00e3o impr\u00f3rios que exigem uma flex\u00e3o fora de plano para manter a planaridade. Essencialmente, eles s\u00e3o usados para manter os grupos como os an\u00e9is arom\u00e1ticos planares.

$$V_{impr\u00f3rios} = \sum_{impr\u00f3rios} K_{\varphi} (\varphi - \varphi_0)^2 \quad (5)$$

Os termos n\u00e3o-ligados incluem van der Waals e intera\u00e7\u00f5es eletrost\u00e1ticas. O termo van der Waals, V_{vdW} , entre os pares de \u00e1tomos \u00e9 aproximado pelo potencial 12-6 de Lennard-Jones:

$$V_{vdW} = \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (6)$$

onde r_{ij} \u00e9 a dist\u00e2ncia interat\u00f4mica, σ_{ij} \u00e9 o di\u00e2metro de colis\u00e3o (a dist\u00e2ncia em zero de energia) e ϵ_{ij} \u00e9 o potencial de profundidade. O termo repulsivo, $(\sigma_{ij}/r_{ij})^{12}$, prevalece a curtas dist\u00e2ncias e modela a repuls\u00e3o eletrost\u00e1tica internuclear e das for\u00e7as repulsivas de curto alcance. O termo atrativo, $(\sigma_{ij}/r_{ij})^6$, \u00e9 devido a for\u00e7as de van der Waals de dispers\u00e3o e prevalece a grandes dist\u00e2ncias.

As intera\u00e7\u00f5es eletrost\u00e1ticas entre os pares de \u00e1tomos V_{elec} s\u00e3o calculadas utilizando cargas parciais implementadas pela lei de Coulomb:

$$V_{elec} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (7)$$

onde os átomos i e j com cargas parciais q_i e q_j , são separados pela distância r_{ij} em um meio com constante dielétrica ϵ .

Campo de Força de mecânica molecular freqüentemente inclui termos que lida com cross-independência de diferentes termos, por exemplo, ligação-ligação, ângulo-ligação, ângulo-ângulo, etc. Para os átomos no ângulo 1-2-3, o termo descreve a distorção da distância 1-3. O potencial de *Urey-Bradley*, V_{UB} é igual ao aumento da energia com a distância entre os átomos de 1,3, e S se desvia de sua referência S_0 . A constante de força K_{UB} e, portanto:

$$V_{UB} = K_{UB} (S - S_0)^2 \quad (8)$$

3.1.3. Otimização da Geometria Molecular

Conforme descrito na seção anterior, a energia potencial é uma função das coordenadas de um sistema. A variação da energia potencial com relação às coordenadas é chamada de superfície de energia potencial (PES). Para um sistema que contém N átomos a energia é uma função de $3N$ coordenadas cartesianas ou $3N-6$ coordenadas internas. Isso produz um número enorme de configurações possíveis, mesmo para uma molécula pequena. No entanto, um sistema molecular é apenas uma configuração estável quando está no mínimo da PES, do qual pode haver um número

muito grande em um sistema tão complexo como uma proteína. Para encontrar as geometrias de um sistema no mínimo da PES, um algoritmo de minimização é utilizado.

Para a localização do mínimo em uma PES, são utilizados métodos numéricos que mudam gradualmente as coordenadas no sentido de diminuir a energia até que um mínimo seja alcançado.

Os algoritmos de minimização mais populares encontram o mínimo utilizando derivadas. A primeira derivada indica a direção do mínimo e a segunda derivada indica a curvatura da função. A função é freqüentemente escrita como uma expansão em série de Taylor sobre o ponto x_0 :

$$V(x) = V(x_0) + (x - x_0)V'(x_0) + \frac{1}{2}(x - x_0)^2 V''(x_0) + \dots \quad (9)$$

Para funções multidimensionais, a variável x é substituída pelo vetor x e matrizes são usadas para descrever as derivadas. Para a energia potencial $V(x)$, existem $3N$ coordenadas cartesianas, portanto o vetor x possui $3N$ componentes e x_0 corresponde à configuração atual do sistema. Quando x_0 é um vetor, $V'(x_0)$ é um vetor $3N$, do qual cada elemento é uma derivada parcial de V , no que diz respeito à enésima coordenada. $V''(x_0)$ é uma matriz $3N \times 3N$ com os elementos de segunda derivada parcial de V com relação às coordenadas x_i e x_j . Isto é conhecido como a matriz Hessiana.

3.1.4. Mecânica Quântica

Para modelar reações químicas é geralmente necessária a utilização da mecânica quântica. A maioria dos métodos de mecânica quântica representa explicitamente os elétrons de um sistema que permite investigar reações químicas nas quais ligações são formadas e quebradas. Esta seção apresenta os métodos de mecânica quântica que foram utilizados neste estudo. A teoria da mecânica quântica é baseada em resolver a equação de Schrödinger, introduzida em 1926 pelo físico austríaco Erwin Schrödinger. A estrutura eletrônica e propriedades das moléculas podem ser obtidas através da fórmula independente do tempo:

$$\hat{H}\Psi = E\Psi \quad (10)$$

O Hamiltoniano \hat{H} é um operador de energia que descreve a energia cinética e potencial de uma partícula. Ψ é a função de onda, que caracteriza o movimento de partículas em um sistema, e E é a energia do sistema.

Exceto para os sistemas mais simples como, por exemplo, o átomo de hidrogênio, não é possível resolver analiticamente a equação de Schrödinger. Apenas soluções aproximadas podem ser obtidas. Essa aproximação é obtida por meio da aproximação de Born-Oppenheimer. A aproximação de Born-Oppenheimer separa os movimentos dos elétrons e dos núcleos. Isso se justifica pelo fato de que os núcleos têm massas muito maiores do que a dos elétrons. Daí os movimentos dos núcleos serem muito mais lentos enquanto que os elétrons são capazes de adaptar-se quase que instantaneamente a qualquer posição dos núcleos. Portanto, com essa

aproximação função de onda Ψ , é somente afetada por posições e não pelo movimento dos núcleos. A função de onda pode ser escrita como o produto de duas funções:

$$\Psi_{\text{tot}}(\text{nuc}, \text{ele}) = \Psi(\text{ele}) \Psi(\text{nuc}) \quad (11)$$

Ψ_{nuc} depende da posição dos núcleos e Ψ_{ele} sobre as posições dos elétrons, mas também é dependente das posições nucleares.

A equação de Schrödinger pode ser resolvida mediante a aplicação da aproximação Born-Oppenheimer e tratando separadamente os elétrons dos núcleos:

$$\hat{H}_{\text{ele}} \Psi_{\text{ele}} = E_{\text{ele}} \Psi_{\text{ele}} \quad (12)$$

Calculando a energia eletrônica, E_{ele} , é possível construir uma PES para diferentes configurações dos núcleos.

3.1.4.1. Teoria do Orbital Molecular

Resolver a função de onda para sistemas de muitos elétrons é ainda um assunto complicado. Teoria dos orbitais moleculares é uma abordagem que simplifica a determinação da função de onda. Ela aproxima a função de onda como o produto de uma série de funções de onda de um elétron ou orbitais:

$$\Psi_{\text{eletrônico}} = |\psi_1 \psi_2 \dots \psi_N| \quad (13)$$

Cada função de onda de um elétron é um orbital molecular que consiste de um termo espacial e de spin. No entanto, isso não satisfaz o princípio de exclusão de Pauli, que exige que a função de onda seja assimétrica (ou seja, o sinal da função de onda deve mudar se as coordenadas do spin espacial e de dois elétrons são alteradas). A solução para este problema é construir a função de onda a partir do Determinante de Slater. Se houver N elétrons em orbitais spin $x_1, x_2 \dots x_N$:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} x_1(1) & x_2(1) & \dots & x_N(1) \\ x_1(2) & x_2(2) & \dots & x_N(2) \\ \vdots & \vdots & & \vdots \\ x_1(N) & x_2(N) & \dots & x_N(N) \end{vmatrix} \quad (14)$$

Alterar quaisquer duas linhas do determinante é equivalente à troca de dois elétrons, que, em seguida, muda o sinal do determinante.

A próxima aproximação utilizada é a Combinação Linear de Orbitais Atômicos (LCAO). Ela representa cada orbital molecular como uma combinação linear de orbitais atômicos.

$$\psi_i = \sum_{\mu}^L c_{i\mu} \phi_{\mu} \quad (15)$$

onde o termo ψ_i é o orbital molecular contendo L orbitais atômicos ϕ_{μ} . Os orbitais atômicos são conhecidas como funções de base e as funções de base L definem o conjunto de bases. $c_{i\mu}$ é um coeficiente e determinando o conjunto de valores $c_{i\mu}$ para

cada MO obtém-se uma solução aproximada para a equação de Schrödinger. Isto se torna possível utilizando o método de Hartree-Fock.

O Método Hartree-Fock

Resolvendo a equação de Schrödinger desta forma se obtém uma resposta aproximada, sendo necessários valores de $c_{i\mu}$ que forneçam a melhor resposta. Para isso, este método utiliza o Princípio da Variação que afirma que a energia de uma função de onda aproximada não pode nunca ser inferior à energia exata. Dessa forma o objetivo do método Hartree-Fock (HF) é determinar os coeficientes que resultam na função de onda de menor energia. Isto é possível por meio do procedimento de Campo Auto-Consistente (SCF). Essencialmente, este trata o problema de muitos elétrons como um problema de um elétron. As distribuições eletrônicas em orbitais moleculares são tratadas independentemente uma da outra, enquanto que, na realidade, as posições de elétrons são dinamicamente correlacionadas. Suposições iniciais para $c_{i\mu}$ são feitas e uma nova função de onda é calculada. Isto é então repetido com os orbitais melhorados. A determinação dos coeficientes exige que as formas de todos os orbitais ocupados já sejam conhecidas. Portanto, o procedimento tem de ser continuado iterativamente até que os critérios de convergência sejam satisfeitos e como resultado obtém-se um conjunto de orbitais auto-consistentes.

Conjunto de Bases

Um aspecto importante dos cálculos de mecânica quântica para a estrutura eletrônica é a escolha de um conjunto de base. Como anteriormente detalhado, conjuntos de base são compostos por orbitais atômicos. Geralmente, os grandes conjuntos de base produzem resultados mais precisos, mas os cálculos serão mais caros computacionalmente. Conjunto de base pode ser mínimo ou estendido. Um conjunto de base mínimo contém apenas um número de funções necessárias para representar os orbitais em cada átomo. Conjuntos de base estendidas levam em conta vários recursos que melhoram a representação dos orbitais. Os conjuntos mais usados são *double zeta* e *split valence*. O conjunto de bases *double zeta* define o dobro do número de funções eletrônicas em relação ao conjunto de base mínima. O conjunto de base *split valence* define o dobro do número de funções usadas para representar os elétrons de valência, mas usa uma única função para as camadas internas. É racional que os orbitais do núcleo, diferentemente dos orbitais de valência, tenham pouco efeito sobre as propriedades químicas e apenas variem ligeiramente em todas as moléculas. Outras funções também podem ser incluídas. Funções de polarização levam em conta a polarização de orbitais atômicas, devido à presença de núcleos de outros átomos. Funções difusas representam a densidade de elétrons de uma distância longa a partir do núcleo. Isso é importante quando se tratam de ânions e espécies de único par.

Existem dois tipos comumente utilizados, o tipo de Slater e do tipo Gaussiana. As funções do tipo Slater levam a resultados muito precisos, mas são difíceis e os cálculos são demorados. A substituição dos orbitais do tipo Slater por orbitais do tipo Gaussiana resulta em uma equação mais fácil de resolver em relação ao custo de

precisão. No entanto, a precisão pode ser aumentada por meio de uma combinação linear de funções Gaussianas para representar cada orbital atômico. Verifica-se que pelo menos três gaussianas são necessárias para representar adequadamente um orbital do tipo Slater.^{32,33}

3.1.4.2. Métodos Semi-empíricos

A utilização de métodos *ab initio* Hartree-Fock para calcular a estrutura eletrônica de uma molécula pode fornecer resultados precisos. No entanto, eles são computacionalmente caros, um fator que aumenta consideravelmente com o tamanho da molécula. Uma maneira de reduzir o custo computacional é a utilização de métodos semi-empíricos. Estes são baseados na mesma teoria dos métodos *ab initio*.

A primeira simplificação feita por métodos semi-empíricos é considerar apenas os elétrons de valência de forma explícita.

Cálculos Hartree-Fock SCF envolvem demanda de tempo e manipulação de integrais. Métodos semi-empíricos usam integrais com base em dados experimentais, que ajudam a reduzir significativamente o custo computacional em comparação aos métodos *ab initio*. Estes dados reais também ajudam a compensar as aproximações utilizadas e incluir efeitos de correlação eletrônica. Na realidade, os movimentos dos elétrons são correlacionados e evitam um ao outro mais do que em métodos HF, isto leva à SCF, sendo maior em energia do que a verdadeira.

Outro método semi-empírico é o Austin Método 1 (AM1)³⁴. Há também o *modified neglect of diatomic overlap-Parametric Method Number 3* (PM3)³⁵. Ambos os

métodos baseiam-se na modificação do método modified neglect of diatomic overlap (MNDO) de Dewar e Thiel³⁶. Eles foram desenvolvidos para corrigir diversos problemas com o método MNDO, com base na tendência de superestimar a repulsão entre os átomos a distâncias muito curtas, igual à soma dos seus raios de van der Waals. O termo de repulsão núcleo-núcleo é modificado em AM1 e PM3 com funções Gaussianas. Funções Gaussianas atrativas foram usadas diretamente para superar a repulsão centrada nas regiões de grande repulsão. Funções Gaussianas repulsivas foram colocadas em pequenas separações internucleares. Como resultado das funções Gaussianas o número de parâmetros exigidos por átomo foram aumentadas.

Os métodos AM1 e PM3 são essencialmente os mesmos, mas foram parametrizadas usando métodos diferentes. Os parâmetros AM1 foram calculados com base em conhecimentos de química e da intuição³⁴. Os parâmetros PM3 foram desenvolvidos usando um processo automatizado desenvolvido por Stewart³⁵. Ambos os métodos são capazes de prever propriedades termodinâmicas e estruturais para um mesmo nível de precisão. No entanto, alguns problemas persistem incluindo a previsões erradas de geometrias e estabilidade de vários grupos funcionais. Particularmente exemplos relevantes incluem: interações como van der Waals e ligações de hidrogênio fracas, geometrias incorretas; barreiras rotacionais baixas para ligações com caráter de dupla ligação, especialmente na ligação C-N em amidas^{32, 33,36}. Apesar desses problemas, métodos semi-empíricos geram dados para geometria molecular e propriedades com gasto computacional significativamente inferior aos métodos *ab initio*.

3.1.4.3. Teoria do Funcional da Densidade

Teoria do Funcional da Densidade (DFT) tem atraído um grande interesse desde a década de 1980 sendo uma alternativa para os métodos *ab initio*. É baseado na idéia de que existe uma relação entre a energia total eletrônica e a densidade eletrônica em geral. O grande avanço veio com o artigo publicado por Hohenberg e Kohns em 1964, que mostrou a energia do estado fundamental de um sistema definida exclusivamente pela densidade de elétrons. Isso pode ser ilustrado através da definição E como um único funcional de $\rho(r)$, a densidade de elétrons. A função atua sobre uma variável para fornecer um número, já o funcional atua em uma função para fornecer o número. Em DFT o funcional de energia é a soma de dois termos:

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)] \quad (16)$$

$E[\rho(r)]$, que descreve a interação dos elétrons com o potencial externo $V_{ext}(r)$. O segundo termo, $F[\rho(r)]$, é a soma da energia cinética e eletrônica e a contribuição das interações intereletrônicas. A densidade eletrônica exata do estado fundamental corresponde ao mínimo de energia. O número de elétrons, N , é fixado:

$$N = \int \rho(r)dr \quad (17)$$

Isto permite uma abordagem semelhante à utilizada para encontrar funções de ondas com o método Hartree-Fock.

O método DFT foi desenvolvido por Kohn e Sham, em 1965, fornecendo uma solução prática para resolver o teorema de Hohenberg-Kohn. O termo $F[\rho(r)]$ na Eq 16 não é conhecido, mas pode ser aproximada em três termos:

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \quad (18)$$

Este descreve a energia cinética, $E_{KE}[\rho(r)]$, a energia coulômbica de elétron-elétron $E_H[\rho(r)]$, e as contribuições de troca e correlação, $E_{XC}[\rho(r)]$. A densidade, $\rho(r)$, foi escrita como a soma dos módulos dos quadrados de um conjunto de orbitais ortonormais de um elétron:

$$\rho(r) = \sum_{i=1}^N |\psi_i(r)|^2 \quad (19)$$

A introdução desses termos dá as equações de um elétron Kohn-Sham uma abordagem auto-consistente. Uma suposição inicial na densidade é feita e produz um conjunto de orbitais e uma melhor densidade. A densidade melhorada é usada para iniciar a próxima iteração e assim sucessivamente até a auto-consistência ser atingida.

Outras aproximações foram desenvolvidas para explicar as desvantagens na teoria que diz respeito ao potencial de correlação de troca. Uma forma de funcional explícita para esse potencial não foi encontrado. O método DFT que foi usado para correções neste trabalho foi o funcional B3LYP³⁷. Ele é baseado em uma classe de aproximação melhorada conhecido como o método Gradiente Corrigido que leva em conta o gradiente de densidade eletrônica para corrigir o termo de correlação de troca. O funcional B3LYP usa o gradiente de correção Becke para a troca e os gradientes de

correção de Lee, Yang e Parr para a correlação, os quais é parametrizado com dados experimentais. Ele também inclui um termo de troca Hartree-Fock o que faz dele um método híbrido Hartree-Fock/DFT.

Métodos DFT não são classificados como *ab initio* ou semi-empíricos, embora eles apresentem características de ambos. Eles podem fornecer descrições precisas da estrutura eletrônica de uma molécula. As vantagens da utilização de DFT sobre os métodos *ab initio* é que ele usa a densidade de elétrons, que ao contrário da função de onda é uma quantidade observável. Densidade de elétrons tem três coordenadas espaciais independentemente do número de elétrons, o que significa que é possível calcular as geometrias e as propriedades dos sistemas com centenas de átomos.

Neste trabalho, foi utilizado o método *ab initio* DFT no cálculo de otimização de geometrias, parâmetros energéticos, geométricos e eletrônicos do sistema investigado.

3.2. Química Medicinal e as Técnicas de QSAR

Uma das técnicas muito empregada em química medicinal no planejamento de novas moléculas candidatas a fármacos é o estudo das relações quantitativas entre a estrutura e atividade (QSAR, do inglês, *Quantitative Structure-Activity Relationships*).³⁸

As técnicas de QSAR são empregadas com a finalidade de se encontrar moléculas com propriedades superiores em relação àquelas apresentadas pelo conjunto de dados que deu origem ao processo de modelagem. A propriedade alvo, geralmente potência ou afinidade, deve ser padronizada e validada para ser útil em estudos de QSAR. Propriedades moleculares são calculadas e empregadas como

descritores. Métodos estatísticos são empregados para produzir uma relação que descreve as variações na propriedade alvo, em função dos descritores. Esta relação é denominada QSAR, a qual 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, prosseguem até que moléculas qualificadas candidatas a novos fármacos sejam selecionadas.

3.2.1. QSAR

Design de novos fármacos tem sido influenciado consideravelmente desde que o primeiro estudo quantitativo da relação estrutura-atividade (QSAR) foi publicado por Hansch e seus colaboradores em 1962³⁹. Este trabalho utilizava modelos matemáticos desenvolvidos a partir de métodos multivariados de análise de dados para correlacionar as alterações na atividade biológica com as mudanças na estrutura química para uma série de moléculas. Os métodos clássicos de QSAR descrevem as relações estrutura-atividade em termos de parâmetros físico-químicos, propriedades estéricas e de certas características estruturais, ou seja, obter os parâmetros responsáveis pela presença ou ausência de atividade inibitória nestes compostos.

3.2.1.1 Descrição dos parâmetros moleculares calculados

No presente estudo, para tentar obter as possíveis relações entre a atividade biológica dos compostos dicetoácidos e seus análogos, diversas propriedades foram

calculadas a partir dos métodos químico-quânticos e outras ferramentas computacionais.

Estas propriedades podem ser divididas em cinco grupos de parâmetros: eletrônicos, geométricos, constitucionais, topológico e lipofilia.

Parâmetros eletrônicos: calculados a partir de métodos químico-quânticos, estes relacionam a influência de certos grupos ou substituintes na densidade de distribuição eletrônica. Como por exemplo: energias dos orbitais de fronteira, cargas atômicas, potencial de ionização, energia eletrônica e polarizabilidade.

As energias dos orbitais de fronteira denominados HOMO (*High Occupied Molecular Orbital* – último orbital molecular ocupado) e LUMO (*Lower Unoccupied Molecular Orbital* – primeiro orbital molecular desocupado) envolvem o princípio de que os elétrons em uma molécula não estão localizados entre átomos, mas em orbitais moleculares deslocalizados, no qual um orbital molecular representa uma combinação linear dos orbitais atômicos. Os orbitais HOMO e LUMO são propriedades importantes, uma vez que medem a capacidade doadora e receptora de elétrons, respectivamente. Assim, quanto maior a energia do HOMO, maior será o caráter nucleofílico, ou seja, a capacidade doadora de elétrons, já que a energia necessária para retirar um elétron do HOMO será menor. Por outro lado, quanto mais baixo for o nível energético do LUMO, maior será o caráter eletrofílico ou afinidade eletrônica, pois os elétrons estarão em orbitais mais estáveis⁴⁰. A Figura 11 ilustra os orbitais de fronteira.

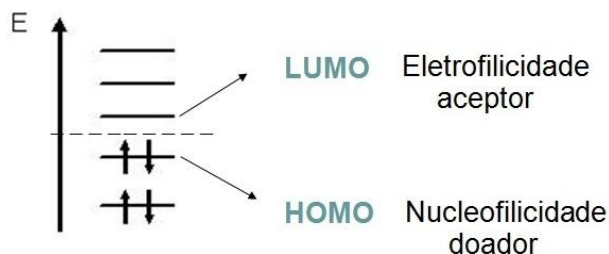


Figura 11. Orbitais de fronteira HOMO e LUMO.

O potencial de ionização está relacionado à força com que um elétron está ligado ao átomo. Assim, o primeiro potencial de ionização é a energia necessária para se retirar um elétron do átomo, e pode ser considerado como sendo análogo ao valor absoluto da energia do último orbital molecular ocupado correspondente (HOMO). Por outro lado, os valores de afinidade eletrônica são análogos aos dos correspondentes primeiros orbitais não ocupados (LUMO).

A polarizabilidade é uma medida da facilidade com que a nuvem eletrônica de um átomo pode ser distorcida, em função do efeito de dipolo de um campo elétrico externo, tendo desta forma uma maior dispersão de elétrons (maior deslocalização). Esta medida pode ser obtida através da diferença de energia (ΔE) HOMO-LUMO. Quanto maior for a diferença de energia HOMO-LUMO menor será a polarizabilidade.⁴¹

A energia eletrônica é calculada para a geometria de equilíbrio das moléculas, onde a configuração nuclear terá um valor mínimo de energia eletrônica.

Parâmetros geométricos: dependem do arranjo espacial dos átomos constituintes da molécula. Representam, por exemplo, informações sobre a superfície molecular, obtidas das áreas de van der Waals dos átomos, assim como sobre o volume molecular.⁴²

A superfície molecular total pode ser calculada somando-se todas as contribuições da superfície atômica, assim como o volume de um dado substituinte que também pode ser calculado.

O cálculo do volume molecular é feito pelo uso da técnica de integração, no qual se gera um conjunto de redes cúbicas tridimensionais. O centro da rede é posicionado em um átomo, e o diâmetro é a aresta do cubo. Todos os pontos da rede são testados, como se eles estivessem dentro do átomo, cuidando para evitar o problema da sobreposição atômica.

Parâmetros Constitucionais: são descritores calculados rápido e facilmente e estão relacionados aos elementos constituintes de sua estrutura. São exemplos a massa molecular, número total de átomos na molécula e número de átomos de identidade diferentes. Informações relacionadas a ligações são também consideradas, como números totais de ligações simples, duplas, triplas ou aromáticas, assim como número de anéis aromáticos.

Parâmetros Topológicos: tratam a estrutura da molécula como um objeto geométrico, tendo os átomos como vértices e as ligações covalentes como arestas. Com base nesta abordagem, muitos índices quantificando a conectividade molecular são definidos, como índice de Wiener que conta o número total de ligações em caminhos curtos entre todos os pares de átomos não-hidrogênio.

Parâmetros de Lipofilia: o logaritmo de coeficiente de partição ($\log P$), definido como sendo o logaritmo a razão entre a solubilidade de uma substância em um

solvente orgânico e a solubilidade da mesma em água. Este parâmetro está relacionado principalmente com a distribuição do fármaco no organismo.

Nesse trabalho foi realizado um estudo quimiométrico de 26 compostos pertencentes às classes 1,3 dicetoácido¹⁹ e 8-hidroxi-(1,6)-nafititidina-7-carboxiamida²²⁻²⁴, objetivando construir um modelo quantitativo da relação estrutura-atividade pela inibição da enzima integrase do vírus HIV-1. Os resultados deste estudo encontram-se no Capítulo 4 desta tese.

3.2.2. Docking Molecular

A compreensão dos mecanismos de reconhecimento molecular ligante-receptor vem sendo cada vez mais utilizada no planejamento racional de fármacos baseado na estrutura do receptor com redução do tempo e do custo envolvidos no processo de desenvolvimentos de novos medicamentos.

Dentre as várias técnicas computacionais disponíveis para o planejamento de fármacos, é interessante a utilização da estrutura do receptor biológico quando esta se encontra disponível, modelando a estrutura química do ligante e relacionando-a com a atividade biológica. Neste caso, o *docking* é uma das técnicas mais utilizadas atualmente tanto nas universidades como nas indústrias farmacêuticas⁴³, a qual consiste na utilização da estrutura do receptor como molde para o desenvolvimento de novos ligantes, dissimilares daqueles já conhecidos, através de uma estratégia de busca (conformação e orientação) e uma função de escore, no sítio ativo da enzima, estimando-se a afinidade de ligação entre ligante e receptor.⁴⁴

Embora a introdução de flexibilidade do receptor seja desejável para uma representação mais realista do sistema, a maioria dos programas hoje disponíveis para estudo de *docking* considera a estrutura do receptor como sendo rígida e apenas o espaço conformacional do ligante como flexível.⁴⁵

Neste trabalho, utilizou-se o programa GOLD⁴⁶, para o estudo de *docking* molecular, o qual utiliza o Algoritmo Genético (GA) na determinação da melhor conformação do ligante no sítio ativo da enzima⁴⁷. Neste programa, o ligante é totalmente flexível, enquanto que na biomacromolécula, isto é, na proteína, apenas as cadeias laterais são flexíveis. Resultados de *docking* disponíveis na literatura obtidos por esse programa reproduziram, de maneira satisfatória, a afinidade experimental para 68% das moléculas estudadas⁴⁸, mostrando-se útil e confiável na avaliação de novas moléculas no sítio ativo de seus receptores biológicos. Esta predição é obtida através da utilização das funções de escore as quais são baseadas em campos de força e incluem termos de ligação de hidrogênio, potencial de dispersão intermolecular 4-8 e um potencial 8-12 para a energia interna do ligante.⁴⁹

Nesta tese, a docagem foi empregada no alinhamento estrutural de uma série de moléculas do conjunto de dados (apresentados na Tabela 7, item 5.1 Capítulo 5) no sítio ativo da enzima integrase do vírus HIV-1 utilizando-se o programa GOLD 5.0[®]. Primeiramente, foi realizada uma pesquisa no banco de dados PDB para a seleção de uma estrutura cristalográfica da HIV-1 IN. Foi selecionada uma estrutura da enzima em complexo com o dicetoácido 5CITEP e um átomo de Mg²⁺ (código PDB: 1QS4), com resolução de 2.1Å.³⁰

O processo de preparo da estrutura selecionada (alvo) para a modelagem envolveu o seguinte procedimento: (i) verificação da conservação do sítio ativo; (ii) análise dos aminoácidos principais; (iii) remoção do ligante do sítio ativo da enzima; (iv) eliminação das moléculas de água; e, (v) adição de átomos de hidrogênio.

Depois de finalizados os processos de preparação do alvo e conjunto de dados, foram estabelecidos os parâmetros de docagem. A preparação do conjunto de dados selecionados após as pesquisas na literatura²⁵⁻²⁹ foi feita utilizando-se o programa SYBYL 8.1^{® 50}. As estruturas dos inibidores foram criadas utilizando-se as ferramentas disponíveis no SYBYL, sendo posteriormente submetidas aos procedimentos de minimização de energia e cálculo de cargas atômicas parciais através do método AM1, no programa MOPAC implementado no SYBYL. Os resultados da docagem molecular para um conjunto de **inibidores contendo o grupo N-Alquil-5-hidroxipirimidinona** são apresentados no item 5.2 do Capítulo 5 desta tese.

3.2.3. Análise Comparativa dos Campos Moleculares (CoMFA)

O método da análise comparativa dos campos moleculares (CoMFA, do inglês, *Comparative Molecular Field Analysis*) é a técnica de QSAR 3D mais utilizada. O conceito não qual se baseia, é o de que a atividade biológica está diretamente relacionada com as interações fármaco-receptor. Estas interações apresentam caráter tridimensional, portanto, este método considera as moléculas em três dimensões, utilizando campos moleculares de interação como variáveis independentes. Estes campos de interação são obtidos através do cálculo da energia de interação entre as

moléculas e sondas virtuais, que consistem em determinado tipo de átomo. As diferenças na atividade biológica (variável dependente) são desta forma, descritas em função das diferenças nestes campos de interação. O método CoMFA emprega predominantemente os campos de interações eletrostáticas e interações estereoquímicas, baseando-se na hipótese de as interações fármaco-receptor são extensivamente dirigidas por estas forças de natureza não covalente.⁵¹

O conjunto de molécula precisa ser submetido a um alinhamento estrutural, ou seja, as moléculas precisam apresentar conformações semelhantes e mesma orientação no espaço cartesiano. Esta é a principal dificuldade inerente a este método, quando comparado aos métodos de QSAR 2D, como o HQSAR. Obviamente, cargas atômicas devem ser incluídas nas moléculas, desde que interações eletrostáticas serão calculadas⁵². As moléculas alinhadas e superpostas são inseridas em uma grade tridimensional reticulada. Em cada ponto de interação desta grade, são calculadas as energias de interação estereoquímica e eletrostática entre uma sonda virtual, que pode ser, por exemplo, um átomo de carbono, e os átomos próximos a este ponto. Este processo é realizado até que todos os pontos sejam escaneados pela sonda. A partir destes cálculos, é gerada uma tabela contendo em umas as colunas, o valor da atividade biológica (variável dependente), e nas outras colunas os descritores moleculares (variáveis independentes). Cada ponto da grade é representado por duas colunas, uma delas contendo o valor da energia de interação eletrostática e a outra o valor para a interação estereoquímica⁵³. A partir desta tabela, o método de regressão PLS, gera um modelo estatístico correlacionado a propriedade alvo com os descritores. Este modelo permite que novas moléculas que ainda não foram sintetizadas ou

testadas, possam ter o valor da propriedade biológica em questão predita^{51, 53}. Este processo é esquematizado na Figura 12.

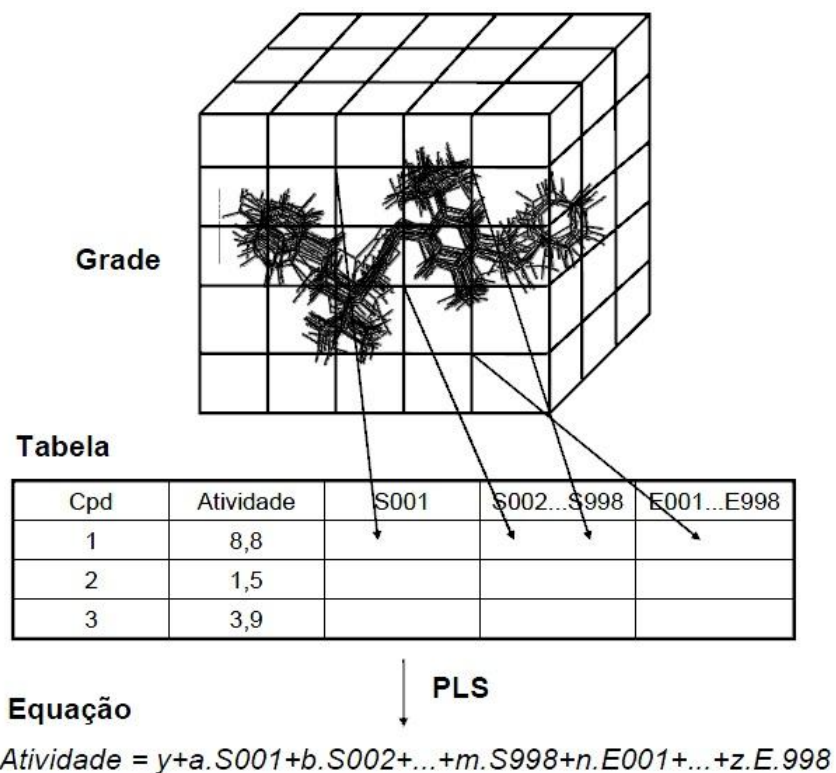


Figura 12. Geração de um modelo CoMFA.

O método CoMFA permite a avaliação visual de modificações na estrutura das moléculas que levem a um aumento da atividade biológica. Para tanto, poliedros que apresentam um código de cores, denominados mapas de contorno, são constituídos em torno das moléculas.

A metodologia 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.

O módulo CoMFA da plataforma SYBYL 8.1[®] foi empregado no desenvolvimento dos modelos de QSAR 3D. No método CoMFA, se faz necessário que as moléculas estejam alinhadas no espaço tridimensional (descrito anteriormente no item 3.2 deste capítulo). A atribuição de cargas atômicas parciais foi necessária para os cálculos das interações eletrostáticas e geração dos campos correspondentes através do método semi-empírico AM1.

Uma planilha molecular (MSS, do inglês, *molecular spreadsheet*) foi criada para a realização das análises de QSAR 3D, contendo toda a informação necessária em relação à estrutura e atividade das moléculas do conjunto de dados. A seguir, foram especificados os parâmetros iniciais referentes à geração dos descritores e posteriormente foram realizadas várias análises de QSAR, introduzindo-se variações na geração dos descritores e nas análises estatísticas com a finalidade de refinar os modelos e produzir um modelos CoMFA com alta capacidade de correlação e predição para o conjunto de dados gerados. Os resultados são apresentados no item 5.3 Capítulo 5 desta tese.

3.3.3. Holograma QSAR

O método holograma QSAR (HQSAR, do inglês, *Hologram Quantitative Structure-Activity Relationships*), baseia-se no conceito de impressão digital molecular (do inglês, *fingerprint*). A premissa da HQSAR é que a estrutura molecular pode ser convertida em seqüências numéricas que codificam a estrutura bidimensional das moléculas de um conjunto de dados. Baseando-se no conceito central da QSAR, de que a atividade biológica é função da estrutura molecular, é possível fazer predições a

respeito da atividade de uma molécula a partir das informações codificadas nos hologramas moleculares.^{54,55}

Os hologramas moleculares empregados na técnica de HQSAR são uma forma diferente de impressão digital. O holograma é gerado a partir da fragmentação da estrutura bidimensional das moléculas, onde todos os fragmentos possíveis são obtidos e posteriormente distribuídos pelo holograma molecular. Cada posição do holograma contém o número de fragmentos alocados nesta determinada posição, sendo que fragmentos iguais que ocorrem mais de uma vez são sempre colocados na mesma posição. Desta forma, o holograma contém informações a respeito dos tipos e da quantidade de fragmentos moleculares. Isto distingue o holograma da impressão digital tradicional, onde a notação é simplesmente binária, contendo apenas características 0 e 1.⁵⁶

Durante o processo de modelagem de HQSAR, há a possibilidade de se definir vários parâmetros relevantes na obtenção dos hologramas, como o comprimento do holograma, o tamanho dos fragmentos e os parâmetros utilizados para distinguir diferentes fragmentos. O tamanho dos fragmentos é definido pelo número máximo (N) e mínimo (M) de átomos que cada fragmento gerado poderá conter. Os parâmetros apresentados na Tabela 1 podem ser utilizados para definir as diferenças ente os fragmentos⁵⁶. Há a liberdade de se gerar modelos selecionando todos os parâmetros, ou combinações de apenas alguns parâmetros. Escolhendo ao final aquela combinação que produz o resultado estatístico mais relevante.

Tabela 1. Parâmetros de distinção de fragmentos no método HQSAR

Parametro de Distinção de fragmento	Definição
(A) Átomos	Permite que os fragmentos sejam distinguidos com base nos tipos de átomos.
(B) Ligações	Permite que fragmentos sejam distinguidos com base nos tipos de ligações: simples, duplas, triplas e aromáticas
(C) Conectividade	Permite que os fragmentos sejam diferenciados baseando-se no estado de hibridização dos átomos contidos no fragmento.
(H) Hidrogênio	A distinção dos fragmentos é feita com base na presença e número de átomos de hidrogênio.
(Ch) Quiralidade	Permite que os fragmentos sejam distinguidos com base na presença de átomos quirais.
(DA) Doador e Receptor	Esta opção permite que os fragmentos sejam distinguidos com base na presença de átomos aceptores ou doadores de ligações hidrogênio.

A partir da geração dos hologramas moleculares, o método de regressão por mínimos quadrados parciais (PLS, do inglês, *partial least squares*) correlaciona a atividade biológica com a informação contida nos hologramas, gerando um modelo estatístico de QSAR. Uma nova molécula que ainda não foi testada pode então ter sua atividade predita quando submetida a este modelo⁵⁵. O processo é ilustrado na Figura 13.

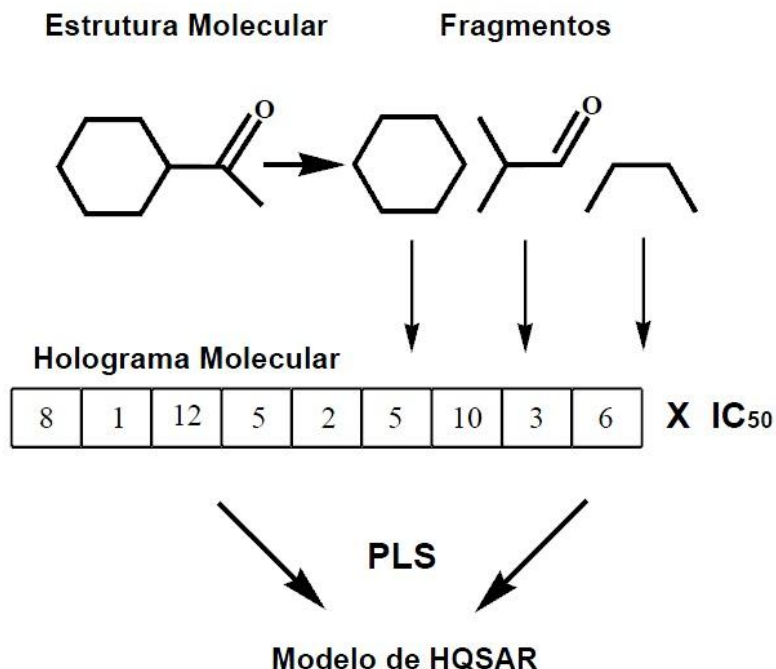


Figura 13. Geração de um modelo HQSAR.

Um importante papel de um modelo de QSAR, além de prever a atividade de moléculas não testadas, é fornecer indicações sobre quais fragmentos moleculares são importantes para a atividade biológica. Estas informações podem orientar a síntese de novas moléculas com propriedades otimizadas. No método de HQSAR, este objetivo é atingido através da construção de um mapa de contribuição 2D, que utiliza um código de cores para indicar a contribuição dos fragmentos moleculares para atividade das moléculas⁵⁶. Esta metodologia, que atualmente faz parte de um leque de tecnologias disponíveis em química medicinal, é capaz de gerar modelos de QSAR com alta capacidade de interpolação e extrapolação em diversos conjuntos de dados.

Para o mesmo conjunto de dados da HIV-1 IN utilizado na modelagem de CoMFA foi empregado o módulo HQSAR da plataforma SYBYL. As estruturas das

moléculas do conjunto de dados e os valores correspondentes de pIC_{50} ($-\log IC_{50}$) foram transferidos para uma planilha molecular (MSS), onde fica armazenada toda a informação para o processo de modelagem de QSAR. A próxima etapa consistiu na seleção de diversos parâmetros referentes à geração dos hologramas para as moléculas do conjunto de treinamento. Várias análises de QSAR, empregando inúmeras combinações de parâmetros, foram realizadas com o objetivo de otimizar os modelos até produzir um modelo final de HQSAR com elevada consistência interna. Esse modelo foi então utilizado no processo de validação externa de um conjunto teste. Os resultados são apresentados no item 5.4 do Capítulo 5.

Todos os trabalhos de cálculos quânticos, análise estatística e produção do modelo QSAR foram realizados em estações computacionais no Grupo de Química Estrutural, no Instituto de Química de São Carlos da Universidade de São Paulo. Já os cálculos de *docking* molecular, CoMFA e HQSAR foram realizados na Escola de Artes, Ciências e Humanidades da Universidade de São Paulo.

4. ESTUDO QUIMIOMÉTRICO DOS COMPOSTOS 1,3 DICETOÁCIDOS E SEUS ANÁLOGOS 8-hidroxi -(1,6)-nafititidina-7-carboxiamida

Neste capítulo, serão discutidos os resultados do estudo quimiométrico de 26 compostos pertencentes às classes 1,3 dicetoácido e 8-hidroxi -(1,6)-nafititidina-7-carboxiamida, objetivando construir um modelo quantitativo da relação estrutura-atividade pela inibição da enzima integrase do vírus HIV-1. Foi calculado um grande número de descritores eletrônicos, estruturais e topológicos para serem utilizados em análises realizadas com o método quimiométrico PLS (Mínimos Quadrados Parciais), com o intuito de identificar descritores que pudessem ser relacionados com as atividades biológicas apresentadas, de acordo com o procedimento descrito a seguir.

4.1. Procedimento Computacional

Dos compostos contendo o grupo 1,3 dicetoácido ¹⁹ e o grupo 8-hidroxi -(1,6)-nafititidina-7-carboxiamida ²²⁻²⁴ selecionados na literatura, foram escolhidos 20 compostos para construir o conjunto de treinamento (mostrados na Tabela 2) e 6 compostos para fazerem parte do conjunto teste, para a validação externa do modelo obtido. (ver Tabela 3). A escolha dos compostos foi feita de maneira que os dois conjuntos apresentassem diversidade estrutural e uma boa distribuição da propriedade biológica (valores de pIC₅₀ variando entre 5,12 e 8,15). Quanto maior o valor de pIC₅₀ maior a atividade. O conjunto de compostos foi dividido em duas classes: compostos com valores de pIC₅₀ > 7.4 foram considerados como compostos com maior afinidade

em se ligar com os átomos de Mg^{2+} na interface da enzima e aqueles com $pIC_{50} < 7.4$ foram considerados compostos com menor afinidade em se ligar com os átomos de Mg^{2+} na interface da enzima. A propriedade utilizada, IC_{50} refere-se à concentração de inibidor necessária para diminuir em 50% a atividade enzimática. Todos os valores de IC_{50} foram determinados nas mesmas condições experimentais, o que é um fator essencial para o desenvolvimento de modelos confiáveis de QSAR.

A estrutura representativa da diversidade química é apresentada na Figura 16. A unidade farmacofórica é constituída pelos fragmentos 1,3 dicetoácido e o grupo 8-hidroxi-(1,6)-nafititidina-7-carboxiamida, sendo fundamental no processo de reconhecimento molecular desta série de inibidores pela a HIV-1 IN. A conservação da unidade farmacofórica é requerida para a manutenção da potência inibitória, uma vez que a diversidade estrutural desta série é conferida pelos variados substituintes presentes nos fragmentos A e B da Figura 14.

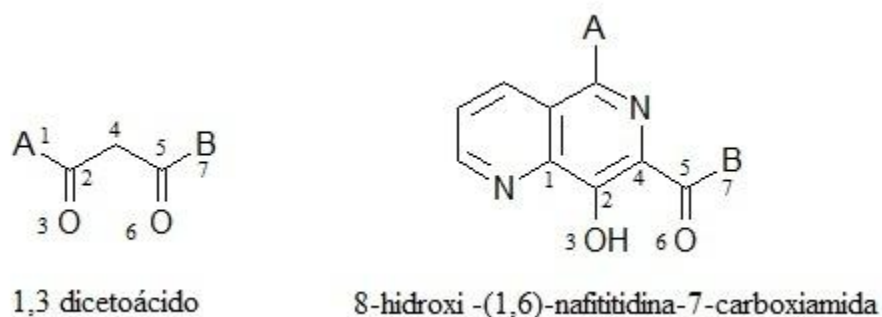


Figura 14. Estrutura representativa da diversidade química do conjunto de dados

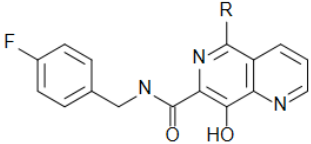
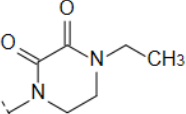
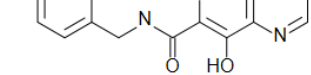
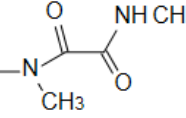

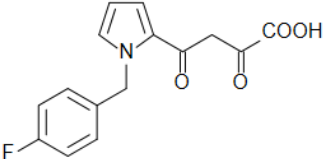
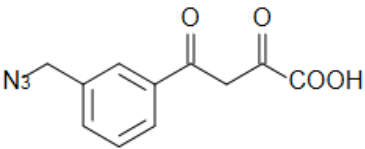
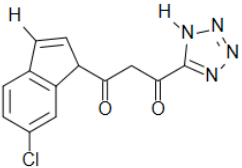
Tabela 2. Estruturas químicas e valores de pIC_{50} dos compostos do conjunto de treinamento.

Composto	Estrutura geral	R	R'	pIC_{50} (nM)
1			-	8.15
2		H	H	7.95
3		Me	H	7.95
4		-	-	7.88
5		H	H	7.88
6		-	-	7.69
7			-	7.69
8			-	7.60
9			-	7.60

Tabela 2. continuação

Composto	Estrutura geral	R	R'	pIC ₅₀ (nM)
10		Me	Me	7.56
11			-	7.52
12		Me	Me	7.45
13		-	-	7.39
14			-	7.39
15			-	7.39
16			-	7.30
17		-	-	6.43
18			-	6.32
19		-	-	6.30
20		-	-	5.12

Tabela 3. Estruturas e valores de pIC_{50} dos compostos de conjunto teste.

Composto	Estrutura geral	R	R'	pIC_{50}
21			-	8.00
22			-	7.82
23		H	-	7.48
24		-	-	7.30
25		-	-	6.39
26		-	-	5.39

4.2. Otimização das geometrias e cálculo dos descritores moleculares

A obtenção de parâmetros moleculares utilizados no estudo da relação estrutura-atividade é uma etapa importante do trabalho, pois estes parâmetros representam as propriedades físico-químicas dos compostos em estudo. Sendo assim, é necessário analisar cuidadosamente a importância de cada parâmetro para a atividade biológica em questão, dada a grande complexidade encontrada no estudo das relações entre estrutura química e atividade biológica.

Para realizar os cálculos das propriedades moleculares que descrevem os compostos, é necessário conhecer a estrutura mais adequada do ponto de vista

energético. Para esse fim, foi empregado inicialmente um procedimento de otimização das geometrias dos compostos, que consiste em encontrar a geometria de equilíbrio correspondente ao arranjo atômico que minimiza a energia total das moléculas a serem estudadas. Embora essa geometria não corresponda à conformação bioativa, que é aquela associada à atividade biológica, este é um procedimento freqüentemente adotado em estudos da relação estrutura-atividade por ser menos dispendioso computacionalmente, e foi empregada no presente trabalho, dada a quantidade de compostos a serem estudados e o nível de teoria utilizado (DFT, *Density Functional Theory* – Teoria do Funcional da Densidade). A partir da estrutura de menor energia para cada composto, foram calculadas as propriedades geométricas e eletrônicas dos compostos, como energia eletrônicas, polarizabilidade, dipolo, energia dos orbitais de fronteira HOMO-2, HOMO-1, HOMO, LUMO, LUMO+1 e LUMO+2, cargas atômicas, etc. Todos esses cálculos foram feitos empregando o método *ab initio* da Teoria do Funcional de Densidade com o funcional de troca-correlação B3LYP. O B3LYP é um método com custo computacional similar ao Hartre-Fock e com resultados entre o método de perturbação de segunda (MP2) e terceira (MP3) ordens de Møller- Pleset⁵⁷. Já o método DFT foi utilizado, pois leva a excelentes resultados para a análise das energias^{58,59}. O conjunto de base empregado foi o 6-31G**, implementado no programa Gaussian 03⁶⁰. Estimados a partir de cálculos químico-quânticos, os parâmetros eletrônicos descrevem não somente as propriedades eletrônicas das moléculas, mas também a influência eletrônica de certos grupos ou substituintes na interação com o receptor biológico.

As cargas atômicas foram calculadas usando o método CHELPG do programa GAUSSIAN 03. Esse método permite avaliar o potencial eletrostático da molécula a partir de cargas atômicas pontuais de modo que represente o melhor potencial molecular para um conjunto de pontos definidos em volta da molécula.⁶¹

Com o programa ChemPlus⁶², foram calculadas algumas propriedades hidrofóbicas e estéricas, tais como; área superficial, volume, volume dos substituintes, massa molecular, refratividade molar e polarizabilidade.

Além dos parâmetros eletrônicos, diversos parâmetros topológicos foram calculados pelo programa DRAGON 2.1⁶³. Estes parâmetros caracterizam atributos estruturais das moléculas (índices de forma molecular e valores de equivalência topológica), caracterizando átomos e grupos no esqueleto molecular. Parâmetros estéricos e hidrofóbicos, como área superficial, volume molecular, massa e ângulos diedros das moléculas e logP, também foram calculados, pois são fatores importantes na interação de um ligante com seu respectivo receptor.

O peso de Fisher foi calculado com o objetivo de selecionar as variáveis com maior poder de discriminação dentro deste conjunto de moléculas. Quanto maior o seu valor, maior será a contribuição de uma dada propriedade, pois maior será a separação do valor das médias das duas classes em questão⁶⁴. O Peso de Fisher, W_{1-2} , para o i -ésimo descritor e amostras pertencentes às classes 1 e 2 é calculado pela Eq. 20:

$$W_{1-2}(i) = \frac{[\bar{X}_i(1) - \bar{X}_i(2)]^2}{S_i^2(1) + S_i^2(2)} \quad (20)$$

onde X_i os valores médios para amostras de cada classe e S_i^2 são os valores das variâncias de cada classe.

Para a análise estatística, o modelo de regressão dos mínimos quadrados parciais (do inglês PLS, *Partial Least Square*) tem como objetivo principal correlacionar uma variável dependente através de uma relação quantitativa; neste caso, o valor da atividade biológica com as variáveis independentes - propriedades calculadas para a série de compostos. Este método está baseado na combinação linear de todas as variáveis originais, e ortogonais das demais componentes principais⁶⁵. A análise quimiométrica PLS foi realizada com o programa Pirouette 3.11⁶⁶. O objetivo desta etapa do trabalho foi tentar prever quais propriedades moleculares têm relevância na atividade biológica e verificar a possibilidade de predição de resultados futuros baseados nestas informações.

4.3 Resultados e Discussão

4.3.1. Seleção de variáveis

Como foi calculado um extenso conjunto de propriedades e descritores, faz-se necessária uma seleção dos mesmos através da redução deste conjunto, com o intuito de eliminar as variáveis que não estejam diretamente relacionadas à atividade biológica dos compostos.

A partir das propriedades calculadas para os 26 compostos, efetuou-se o cálculo dos pesos de Fisher dessas variáveis, para verificar quais teriam o maior potencial discriminante. Aquelas com maior peso foram selecionadas para serem utilizadas nas análises estatísticas. De 1150 variáveis foram eliminadas as variáveis

apresentando altas correlações entre si e 16 variáveis com pesos de Fisher acima de 2,5 foram selecionadas para a geração do modelo PLS.

4.3.2. Regressão PLS

O método estatístico PLS foi empregado neste trabalho para modelar a propriedade experimental de interesse (variável dependente pIC_{50}), Y , com os possíveis descritores X (variáveis independentes), representando as estruturas químicas dos compostos em estudo. Primeiro, um modelo foi constituído e refinado com base no conjunto de treinamento e, em seguida, foi usado para fazer previsões sobre os compostos do conjunto de teste (validação externa do modelo). O melhor modelo é escolhido com base na soma dos quadrados dos erros de previsão (PRESS) obtidos no procedimento de validação cruzada do modelo. O número ótimo de componentes PLS é aquele que minimiza o valor de PRESS. Se o modelo apresenta boa qualidade, o que é verificado pelos valores dos coeficientes de correlação r^2 e q^2 e também pelos resíduos de previsão pode ser utilizado para fazer previsões da propriedade biológica de compostos desconhecidos que sejam estruturalmente similares àqueles presentes no conjunto de treinamento.

O melhor modelo encontrado foi construído com os descritores energia do orbital LUMO (E_{LUMO}), polarizabilidade, carga no átomo 1 (Q_1) e distância de ligação entre os átomos 2 e 3 (r_2). O número ótimo de componentes PLS foi escolhido analisando-se o valor de PRESS. Pela Tabela 4, verifica-se que o número de componentes PLS correspondente ao menor valor de PRESS e ao maior valor de q^2 é igual a 6, sendo portanto o número ótimo de componentes PLS para o conjunto de dados. Este número

resultou em um modelo com coeficientes de regressão $r^2 = 0,86$ e $q^2 = 0,65$, o que indica que um bom modelo foi obtido.

Tabela 4. Porcentagem de variância, PRESS, r^2 e q^2

Componente PLS	Variância Cumulativa%	PRESS	q^2	r^2
1	48	3,66	0,599	0,804
2	56	4,07	0,595	0,840
3	63	4,22	0,569	0,859
4	78	4,70	0,660	0,907
5	87	4,29	0,562	0,848
6	88	3,90	0,652	0,863
7	95	4,28	0,495	0,759
8	100	4,81	0,589	0,799
9	100	4,52	0,615	0,825

A Figura 14 mostra a reta de regressão para os valores de pIC_{50} para o conjunto de treinamento. A maioria dos resíduos de predição é menor do que 0,30. A Figura 15 mostra também a reta de regressão para os compostos do conjunto de teste usados na validação externa no modelo PLS. Os resultados da predição para este conjunto são mostrados na Tabela 5, onde é possível observar que os erros de predição são todos inferiores a 0, indicando que um modelo com bom poder preditivo foi obtido.

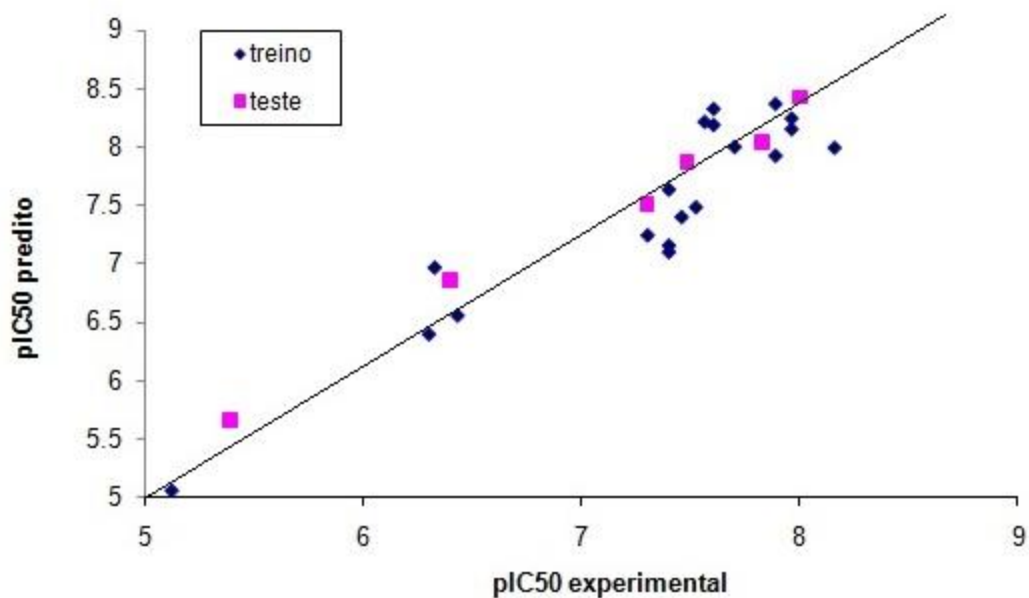


Figura 15. Valores de pIC_{50} preditos *versus* experimentais para a regressão PLS.

Tabela 5. Resíduos de predição para o conjunto teste

Composto	$\log IC_{50}$ experimental	$\log IC_{50}$ predito	Resíduo
21	8,00	8,43	-0,42
22	7,82	8,04	-0,22
23	7,48	7,86	-0,39
24	7,30	7,51	-0,21
25	6,40	6,86	-0,46
26	5,40	5,66	-0,26

O melhor modelo de regressão obtido está representado na Eq. 21.

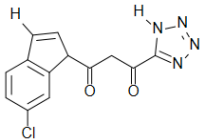
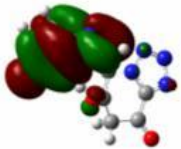
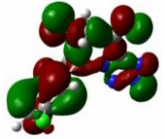
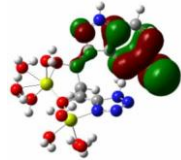
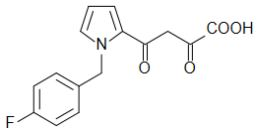
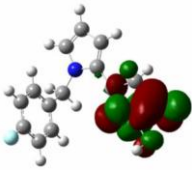
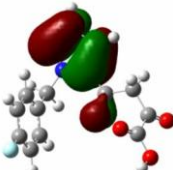
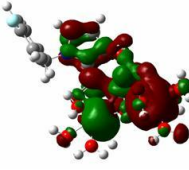
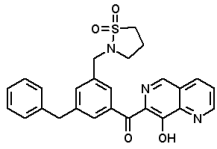
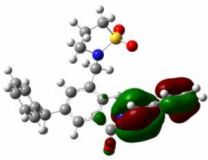
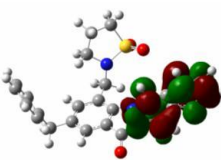
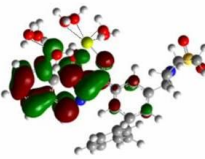
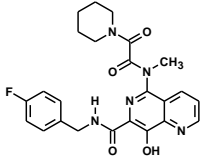
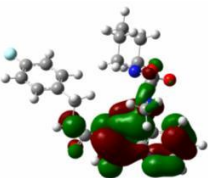
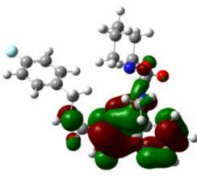
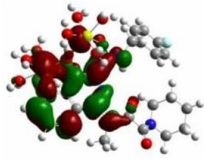
$$\log IC_{50} = -0.5047(E_{lumo}) - 0,7851(polar) + 0,00768(Q_1) + 0,3106(r_2) \quad (21)$$

A partir dessa equação pode-se observar que a E_{LUMO} e a polarizabilidade contribuem negativamente, enquanto que a carga no átomo 1 e a distância r_2 entre os átomos 2 e 3 contribuem positivamente. Deste modo é possível dizer que a contribuição das quatro variáveis (multiplicadas por seus coeficientes na equação) determina a presença ou ausência de atividade. Portanto, para que um inibidor da enzima integrase seja ativo, Q_1 e r_2 precisam apresentar valores baixos enquanto E_{LUMO} e a polarizabilidade valores altos.

4.3.3. Relação entre orbitais moleculares e valores $\log IC_{50}$

Como discutido no Capítulo 1, a atividade desses inibidores são dependentes dos íons Mg^{2+} presentes no sítio catalítico da enzima, o qual atua como uma ponte de ligação entre a tríade catalítica DDE, o cDNA viral e o DNA hospedeiro. Com o objetivo de melhor entender a atividade dessa classe de moléculas foram calculados mapas de HOMO e LUMO, com o método *ab initio* DFT B3LYP 6-31G**, para algumas moléculas mais ativas e outra menos ativas sem complexar com os íons e complexadas com os íons Mg^{2+} ($[L-2Mg(H_2O)_4]^{4+}$). Pela Tabela 6, é possível observar que para as moléculas mais ativas (moléculas contendo o grupo 8-hidroxi -(1,6)-nafititidina-7-carboxiamida) o HOMO encontra-se na região farmacofórica responsáveis pela formação do complexo com os íons Mg^{2+} . O que fica claro, pois para a ligação com os cátions é necessária uma densidade eletrônica maior nos átomos de oxigênio envolvidos.

Tabela 6. Orbitais moleculares para os compostos mais ativos e menos ativos

Molécula	pIC50 (nM)	Homo	Lumo	Lumo do complexo ligante Mg ²⁺
	5,39			
	7,30			
	7,88			
	8,15			

Avaliando as estruturas, verifica-se que os anéis do grupo 8-hidroxi -(1,6)-nafititidina-7-carboxiamida contribuem com caráter nucleofílico para os átomos de oxigênio, o que já não acontece no grupo 1,3 dicetoácido visto a ausência de anéis aromáticos. Como conseqüência, as moléculas de naitiridinas são mais ativas que as

moléculas de dicetoácidos. Após a complexação com os íons de Mg^{2+} , torna-se agora necessário que o LUMO esteja sobre os átomos de magnésio, uma vez que estes irão interagir com uma região nucleofílica do sitio catalítico da enzima. Para as moléculas mais ativas (naftiridinas), observa-se que o LUMO está principalmente sobre os átomos de magnésio, ao passo que para as moléculas menos ativas (dicetoácidos), novamente, o orbital encontra-se diluído ou concentrado em uma região diferente dos íons de magnésio.

4.4. Conclusões

Neste capítulo, foram apresentados modelo SAR e QSAR confiáveis e preditivos, obtidos com o método quimiométrico PLS. O modelo apresentou consistência interna e foi validado com um conjunto de teste. Os descritores selecionados, que contêm informações úteis sobre vários aspectos da estrutura molecular, mostraram boa correlação com o modo de inibição da enzima integrase do vírus HIV-1. As análises de QSAR e o estudo teórico com o método DFT indicaram que, para as moléculas mais ativas, os orbitais moleculares HOMO e LUMO estão localizados principalmente nos farmacóforos β -dicetoácidos e 8-hidroxi-(1,6)-naftiridina-7-carboxamida. Logo há uma importante contribuição das cargas negativas sobre os átomos de oxigênio que estão envolvidos na coordenação com os cátions de Mg^{2+} . Este efeito do orbital produz um elevado caráter nucleofílico sobre esses sítios ativos, necessário para deslocar duas moléculas de água a partir de $[Mg(H_2O)_6]^{2+}$. Finalmente, o LUMO nas moléculas mais ativas em $[L-2Mg(H_2O)_4]^{4+}$ está localizado na

região dos íons Mg^{2+} , possibilitando a ligação com os átomos de oxigênio dos resíduos aspartato (D64 e D116) e glutamato (E152) no sítio ativo da enzima.

5. ESTUDO DAS RELAÇÕES QUANTITATIVAS ENTRE A ESTRUTURA E ATIVIDADE DOS INIBIDORES DA INTEGRASE CONTENDO O GRUPO **N-Alquil-5-hidroxipirimidinona**

Em outubro de 2007 foi aprovado pelo FDA o primeiro medicamento pertencente à classe dos inibidores da integrase (inibindo a etapa de transferência de fita, INIs), conhecido comercialmente como ISENTRESS™. Essa decisão do FDA foi baseada nos resultados da análise clínica realizada durante 24 semanas em que ISENTRESS™ foi combinado à terapia padrão (OBT) em pacientes no tratamento de teste, providenciando reduções na carga de HIV RNA viral e no aumento da contagem de células CD4²⁹. **O ISENTRESS™ possui como princípio ativo o *raltegravir*, um composto contendo o grupo *N-Alquil-5-hidroxipirimidinona*, sintetizado a partir de dicetoácidos e nafiridinas. Com o objetivo de obter um modelo quantitativo mais condizente com a realidade atual, optou-se por estudar agora as moléculas do grupo *N-Alquil-5-hidroxipirimidinona* por meio de **Docking Molecular, QSAR 3D e HQSAR.****

5.1. Conjunto de dados

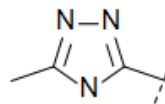
Os inibidores da HIV-1 IN empregados no desenvolvimento dos modelos de QSAR foram selecionados por meio de pesquisas na literatura²⁵⁻²⁹. O conjunto de dados dos inibidores da HIV-1 IN empregado no desenvolvimento dos modelos de *Docking Molecular*, QSAR 2D e 3D possui 79 moléculas associadas aos respectivos

valores de potência (**IC₅₀**). As estruturas químicas e os valores da atividade biológica (IC₅₀) são apresentados na Tabela 7.

Tabela 7. Conjunto padrão de dados de inibidores da HIV-1 IN.

Compostos do conjunto de treinamento				
Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
1			Me	8.5
2				8.7
3			Me	7.7
4			Me	7.9
5			H	8
6			H	8.4
7			Me	7.8
8			-	7.8
9			-	8.1
10			-	8.0
11			-	8.2
12			-	7.8

13

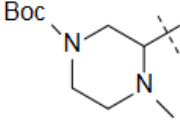
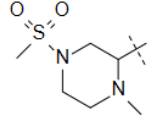
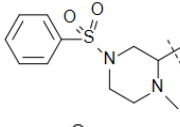
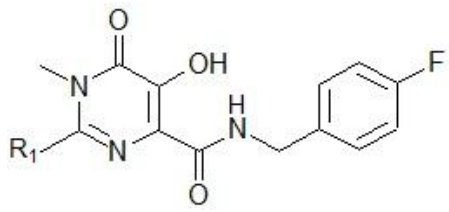
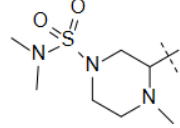
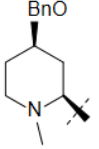
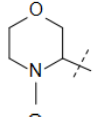
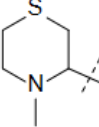
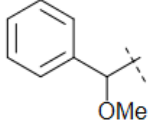
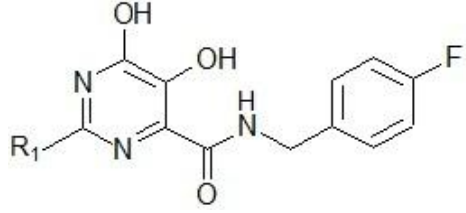
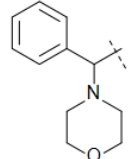
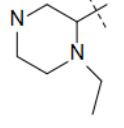
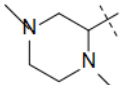


-

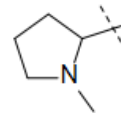
8.4

Tabela 7. Continuação

Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
14			-	7.2
15			-	6.1
16			-	6.7
17			-	7.0
18			-	6.7
19			-	7.2
20			-	7.7
21			-	7.0
22			-	6.7
23			-	6.8

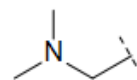
24			-	7.4
Tabela 7. Continuação				
Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
25			-	7.1
26			-	8.2
27			-	8.1
28			-	7.2
29			-	7.2
30			-	7.1
31			-	7.1
32			-	7.3
33			-	7.0
34			-	7.0

35



- 6.6

36



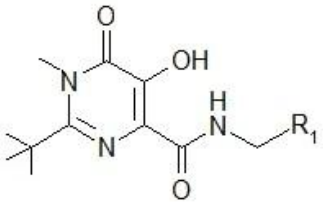
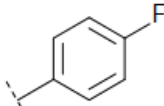
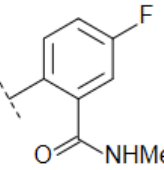
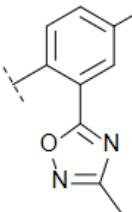
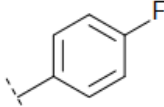
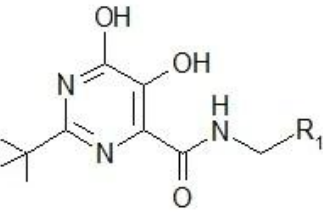
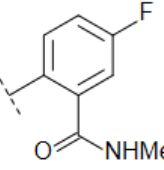
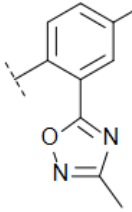
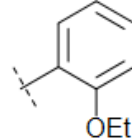
- 6.7

Tabela 7. Continuação

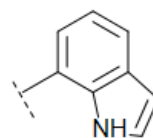
Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
37			-	5.7
38			-	7.3
39			-	7.2
40			-	8.1
41			-	8.0
42			-	7.3
43			-	7.5
44		Pr	-	7.0
45		H	-	7.2
46		H	H	7.2
47		2-F	H	6.7
48		3-F	H	7.2
49		3-Cl	H	7.2
50		3-Br	H	7.7
51		3-OMe	H	6.8

52		4-F	3-F	7.6
53		4-F	3-Me	7.1
54		4-F	3-Cl	7.5
55		4-F	3-Br	7.3

Tabela 7. Continuação

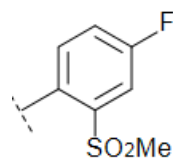
Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
56			-	8.0
57			-	7.7
58			-	7.7
59			-	8.0
60			-	7.7
61			-	7.7
62				8.0

63



7.2

64

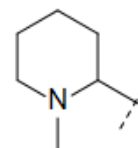


8.0

Tabela 7. Continuação

Compostos do conjunto teste				
Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
65				Me 6.6
66				Me 8.1
67				H 7.8
68				H 7.7
69				- 7.7
70				- 8.2
71				- 7.7
72				- 7.3

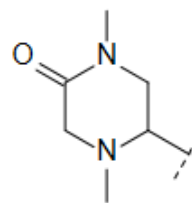
73



-

6.4

74



-

7.3

Tabela 7. Continuação

Comp.	Estrutura Geral	R ₁	R ₂	pIC ₅₀
75				- 6.7
76				- 8
77				- 7.4
78				- 7.0
79				- 7.1

A propriedade utilizada, IC₅₀ refere-se à concentração de inibidor necessária para diminuir em 50% a atividade enzimática. Os valores de IC₅₀ variam de 2 a 2000nM, um fator de potência de aproximadamente 1000 vezes. Todos os valores de IC₅₀ foram determinados nas mesmas condições experimentais, o que é um fator essencial para o desenvolvimento de modelos confiáveis de QSAR.^{67,68}

O conjunto de dados apresenta boa diversidade química. A estrutura representativa da diversidade química é apresentada na Figura 16. A unidade

farmacofórica é constituída pelo fragmento *N-Alquil-5-hidroxipirimidinona*, sendo fundamental no processo de reconhecimento molecular desta série de inibidores pela a HIV-1 IN. A conservação da unidade farmacofórica é requerida para a manutenção da potência inibitória, uma vez que a diversidade estrutural desta série é conferida pelos variados substituintes presentes nos fragmentos **A** e **B** da Figura 16.

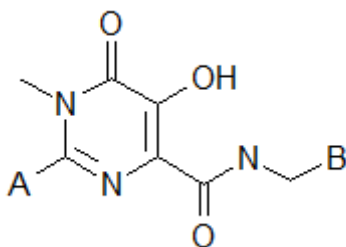


Figura 16. Estrutura representativa da diversidade química do conjunto de dados.

Como pode ser observado na Tabela 7, as moléculas do conjunto de dados foram divididas em dois conjuntos, denominados conjunto treinamento e conjunto teste. O conjunto treinamento é formado pelas moléculas que são empregadas na construção dos modelos de QSAR, ou seja, com as informações sobre a estrutura e atividade biológica das moléculas são gerados os descritores e realizadas as análises estatísticas. O conjunto teste é empregado na avaliação externa dos modelos em relação a sua capacidade de predizer a propriedade alvo de novas moléculas que não foram incluídas durante o processo de modelagem no conjunto de treinamento.

Do total de 79 inibidores do conjunto de dados, 15 (20%) foram separados para o conjunto teste, sendo que os 64 restantes (80%) constituem o conjunto treinamento. O conjunto treinamento é formado pelos inibidores de **1** a **64** e, enquanto que o conjunto teste é formado pelos inibidores de **65** a **79**.

Para melhor compreensão da distribuição dos valores de atividade, os valores de IC_{50} foram convertidos nos valores correspondentes na escala logarítmica de pIC_{50} ($-\log IC_{50}$). Portanto, para uma análise comparativa, quanto maior o valor de pIC_{50} mais potente será o composto. A distribuição dos valores de pIC_{50} é mostrada no histograma da Figura 17.

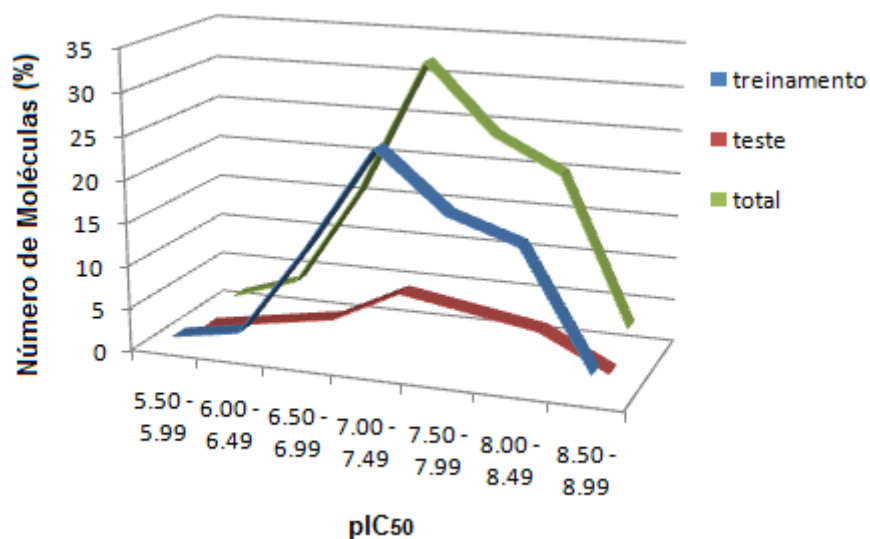


Figura 17. Representação gráfica da distribuição dos valores de pIC_{50} .

É importante ressaltar a distribuição representativa do conjunto teste em comparação com o conjunto treinamento. Essa distribuição apropriada em relação à propriedade biológica representa um fator importantíssimo na avaliação da consistência interna e externa dos modelos de QSAR. Portanto, o conjunto padrão de dados organizados para os inibidores da HIV-1 IN é apropriado para o desenvolvimento dos modelos preditivos de QSAR 2D e 3D propostos nesta tese.

As estruturas dos inibidores foram criadas empregando-se o módulo CONCORD⁶⁹ disponível no programa computacional SYBYL 8.1, sendo posteriormente submetidas aos procedimentos de minimização de energia e cálculo de cargas atômicas parciais através do método semi-empírico AM1, no programa MOPAC implementado no programa SYBYL 8.1. As estruturas 3D geradas foram empregadas no processo de alinhamento estrutural através de *docking* molecular, necessário para a realização do método CoMFA. Todos os cálculos e visualizações foram realizadas no pacote SYBYL 8.1, em plataforma Linux.

5.2. Docking Molecular

O programa GOLD 5.0 oferece uma série de parâmetros que podem ser controlados pelo usuário. Dentre os parâmetros mais importantes se destacam: (i) a quantidade de soluções de docagem (acoplamentos) que deve ser produzida para cada ligante; e, (ii) a definição da área do sítio ativo para o acoplamento das diferentes moléculas. Essa área, que consiste de uma esfera, pode ser definida a partir de um ponto com coordenadas cartesianas definidas, ou de um átomo de um ligante ou ainda de um resíduo de aminoácido de proteína. Uma vez definido o sítio ativo, todas as moléculas (com todos os seus átomos constituintes) só poderão ser alocadas no espaço disponível dentro desta esfera. Outro parâmetro se refere às restrições que podem ser impostas. Estas restrições podem especificar interações que devem ocorrer entre a proteína e os ligantes, como por exemplo, ligações hidrogênio e interações

hidrofóbicas. A Figura 18 apresenta um esquema geral das interações entre os inibidores do conjunto de dados e HIV-1 IN. Estas interações foram consideradas no processo de docagem molecular.

Primeiramente, foi realizada uma pesquisa no banco de dados PDB (do inglês, *Protein Data Bank* - <http://www.rcsb.org/pdb/home/home.do>) para a seleção de uma estrutura cristalográfica da HIV-1 IN. Foi selecionada a única estrutura em complexo com um inibidor existente (código PDB: 1QS4), com resolução 2.1Å. Esta se caracteriza por ser o domínio core catalítico complexada com um átomo de magnésio e o inibidor dicetoácido 5CITEP.³⁰

O processo de preparo da estrutura selecionada (alvo) para a modelagem envolveu o seguinte procedimento: (i) verificação da conservação do sítio ativo; (ii) análise dos aminoácidos principais; (iii) remoção do ligante do sítio ativo; (iv) adição de um segundo íon de Mg^{2+} entre os resíduos D64 e E152; (v) eliminação das moléculas de água; e (vi) adição dos átomos de hidrogênio.

Depois de finalizados os processos de preparação do alvo e do conjunto de dados, foram estabelecidos os parâmetros de docagem. O programa GOLD foi utilizado para a docagem dos inibidores no sítio ativo da enzima.

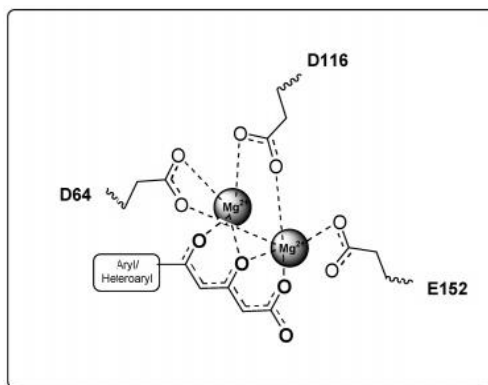


Figura 18. Esquema geral das interações entre os inibidores do conjunto de dados e a HIV-1 IN.

Com base na análise estrutural das interações intermoleculares responsáveis pelo processo de reconhecimento molecular, algumas características foram estabelecidas no processo de docagem. O sítio ativo foi definido com um raio de 5 Å ao redor do ligante 5CITEP. Em relação às restrições mais importantes, foi definido que: cada íon de Mg^{2+} seria quelado pelos oxigênios do grupo *hidroxipirimidinona* e pelos oxigênios dos resíduos D116, D64 e E152. Foi estabelecido ainda que para cada ligante o algoritmo genético (GA) foi reproduzido 30 vezes empregando desempenho máximo (rotação de todas as ligações rotacionáveis, exploração de todos os ângulos e diedros possíveis de ligação do ligante).

Depois de realizado o processo de alinhamento, foram geradas 10 soluções (conformações) para cada molécula do conjunto de dados. Cada conformação foi avaliada quanto à satisfação da hipótese farmacofórica apresentada na Figura 18. Desta maneira, verificou-se se os grupos químicos dos inibidores foram alocados de forma correta no sítio da proteína. Também foi avaliado o grau de similaridade entre as conformações das diferentes moléculas e o seu acoplamento relativo dentro do sítio de ligação. Baseando-se nestes critérios, foi escolhida uma única conformação para cada molécula. Deve ser enfatizado que diversos outros alinhamentos também foram testados antes da escolha final deste alinhamento gerado com o programa GOLD.

A partir dos resultados de *docking* molecular observou-se que todos os compostos apresentaram um modo de ligação interfacial similar em que o grupo farmacofórico *hidroxipirimidinona* quela os íons de Mg^{2+} juntamente com a tríade

catalítica DDE, como discutido na literatura^{16,19-21}. A Figura 19 ilustra esse modo de ligação.

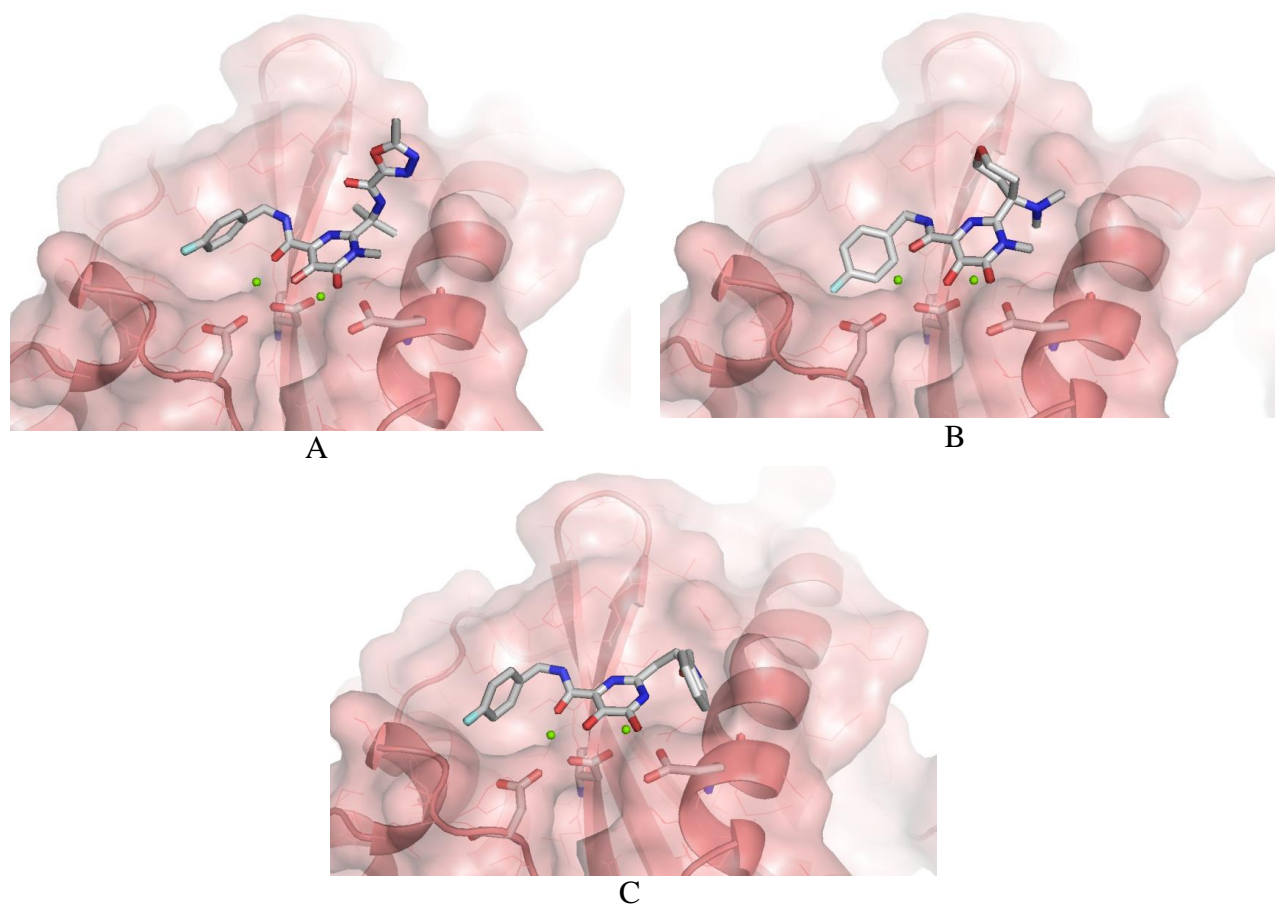


Figura 19. Docking molecular do fármaco raltegravir 12 (A), molécula mais ativa 2 (B) e menos ativa 37(C) no sítio ativo da enzima integrase. Os dois pontos verdes correspondem aos dois íons de Mg²⁺ quelados pelos oxigênios da tríade catalítica DDE e pelos oxigênios do domínio quelante do inibidor.

Na Figura 20 estão ilustrados as distâncias entre os íons de Mg²⁺ e o fármaco raltegravir, a molécula mais ativa e menos ativa respectivamente.

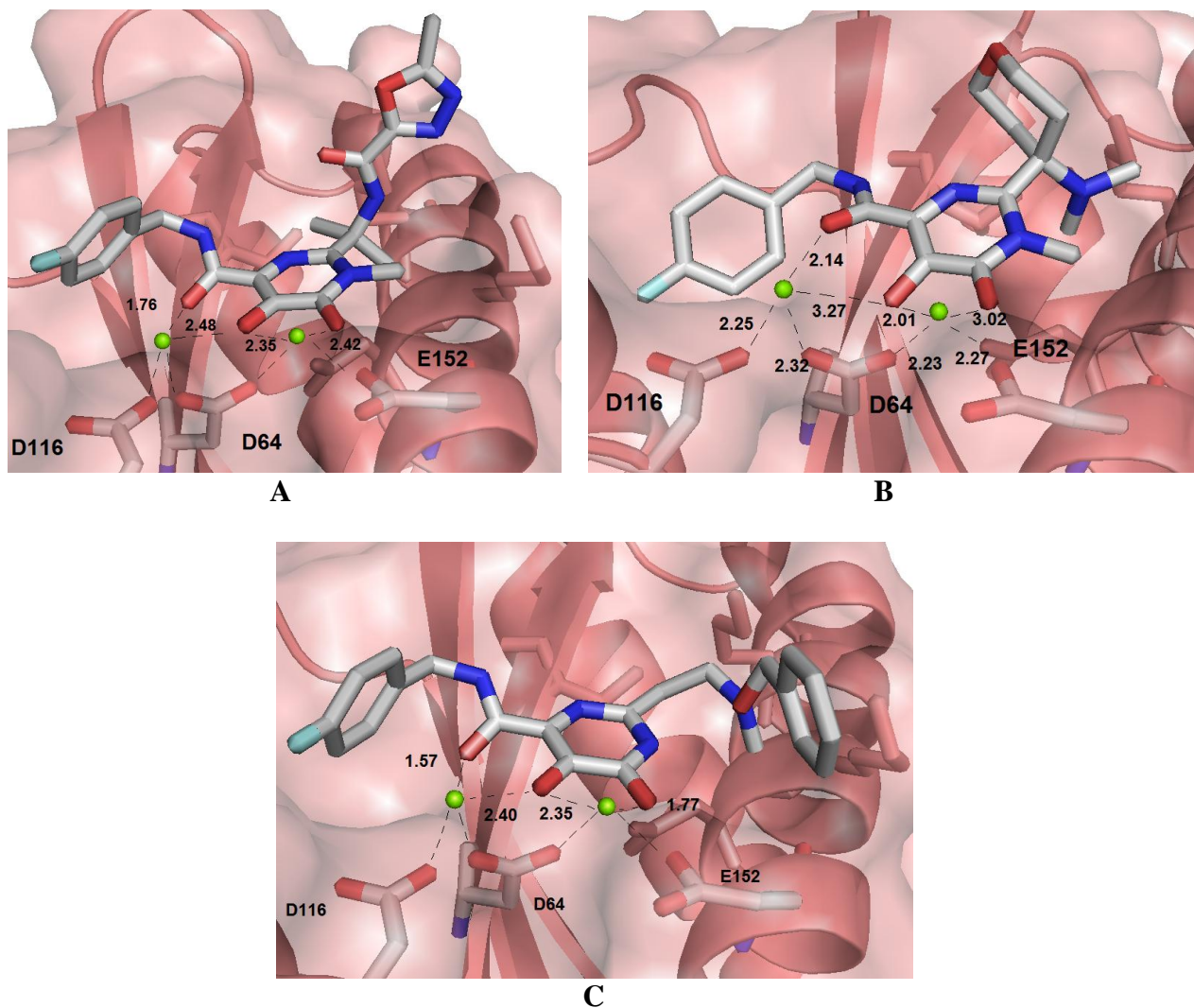


Figura 20. Distâncias entre os íons de Mg²⁺ e o fármaco raltegravir (A), a molécula mais ativa (B) e menos ativa (C) e a tríade catalítica na inibição do processo de transferência de fita.

Além das interações com os íons Mg²⁺, há também interações com resíduos importantes no processo de integração como o treonina T66 e os dois resíduos de lisina K156 e K159, previamente relatados como sendo envolvidos na ligação com a fita de DNA viral no processo de transferência de fita catalisada pela integrase³¹. A Figura 21 ilustra essas interações para o fármaco raltegravir, molécula **2** e molécula **37**.

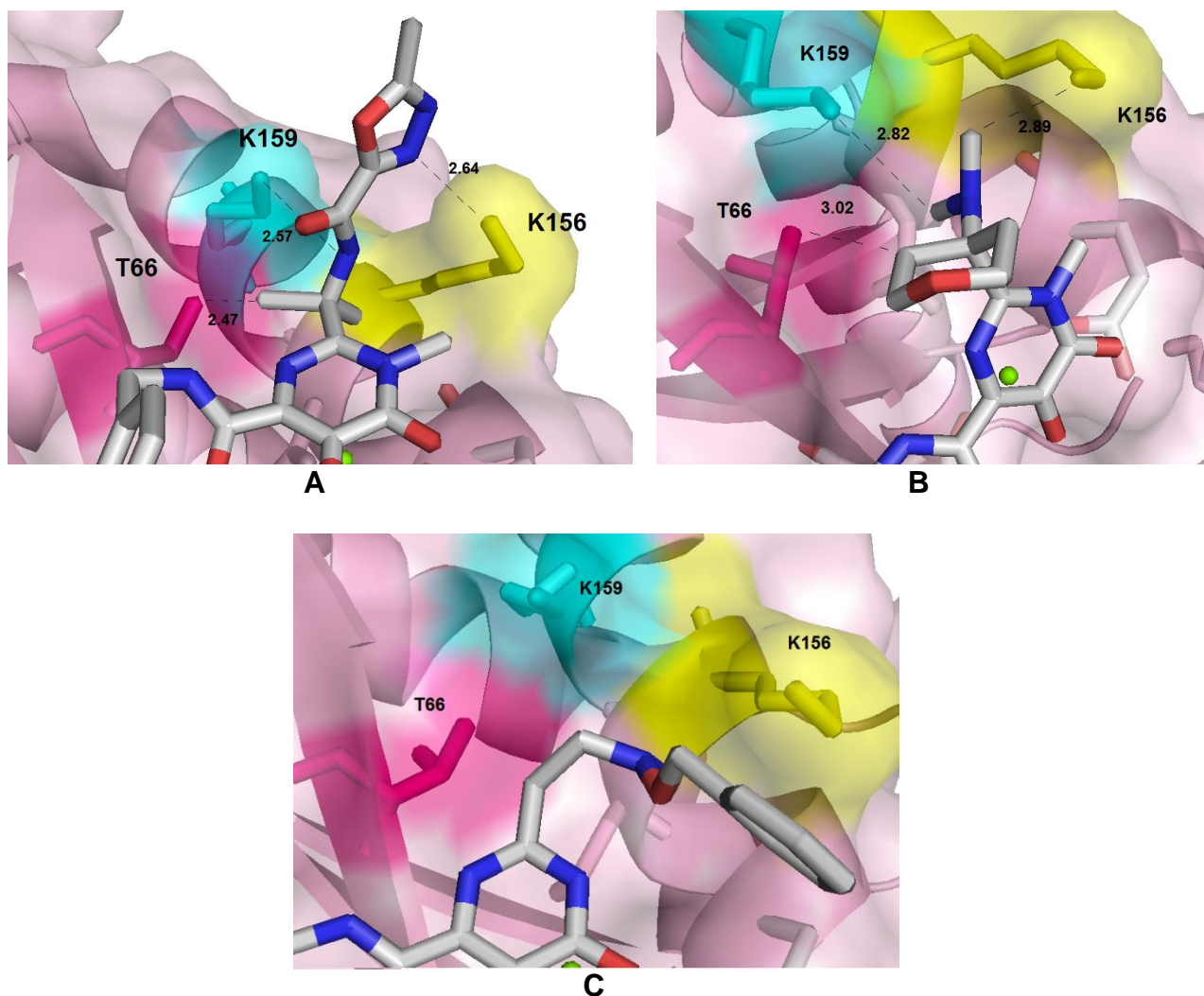


Figura 21 – Interações do fragmento A do fármaco raltegravir (A), molécula 2 (B) e molécula 37 (C) com os resíduos T66, K159 e K156. Esses resíduos se ligam com o DNA viral no processo de transferência de fita.

Na Figura 21-A e 21-B é possível verificar as interações do fragmento A do fármaco raltegravir e da molécula 2 com os resíduos da enzima HIV-1 IN. Há ligações de hidrogênio entre o oxigênio do resíduo T66 e o hidrogênio de um dos grupos metila, do oxigênio do resíduo K159 e o hidrogênio do grupo carboxamida e entre o resíduo K156 e o anel oxadiazol. Estas interações estão de acordo com o modelo proposto por Goldgur *et al.* (1999)³⁰. É possível observar que as distâncias entre as ligações hidrogênio para o fármaco raltegravir são

menores em relação às distâncias da molécula mais ativa 2. Já na Figura 21-C verifica-se que para o fármaco menos ativo não há interações do fragmento A com esses resíduos.

5.3. QSAR 3D

Os métodos de QSAR 3D baseiam-se no caráter tridimensional das interações entre um ligante e seu receptor biológico e na utilização de compos moleculares de interação com os descritores. As diferenças na variável dependente (propriedade biológica) são descritas em função das diferenças nestes campos de interação. Neste trabalho, foi empregado o método CoMFA, que é um dos mais utilizados no planejamento de fármacos e foi brevemente descrito no Capítulo 3. Estudos comparativos de análise molecular de campo foram realizados com o objetivo de melhor compreender e explorar as contribuições eletrostáticas e estéricas dos inibidores da integrase do vírus HIV em estudo para a sua atividade biológica.

5.3.1. Procedimento computacional

A etapa crucial na geração do modelo CoMFA é o alinhamento molecular 3D das moléculas do conjunto de dados. O alinhamento 3D usado neste trabalho é com base em conformações moleculares obtidas a partir de um processo em que uma conformação única otimizada de cada molécula no conjunto de dados foi minimizada energeticamente, utilizando o átomo parcial centrado nos cálculos de carga

implementados no programa MOPAC. Esta escolha para o alinhamento molecular pode ser justificada porque há uma estrutura cristalográfica determinada para a HIV-1 IN.

A estratégia empregada para a geração do alinhamento 3D baseado na estrutura do receptor foi a docagem (acoplamento) das moléculas do conjunto de dados no sítio ativo da enzima, utilizando-se o programa GOLD 5.0 previamente detalhada no item anterior.

Após o alinhamento de todas as conformações moleculares, o próximo passo foi o cálculo das propriedades estéricas e eletrostáticas de acordo com os potenciais Lennard-Jones e os de Coulomb, respectivamente. Para isso, o conjunto de 79 moléculas alinhadas foram inseridas em uma grade tridimensional com espaçamento de 2.0Å entre os vértices nas direções x, y e z. Os mapas estéricos e eletrostáticos foram gerados utilizando o método AM1 e um carbono sp^3 positivamente carregado como uma sonda química. O valor padrão de 30 kcal/ mol foi definido como o corte de energia máxima estérica e eletrostática. Para que as energias de interação fossem tratadas como tendo a mesma ordem de grandeza, os foram escalonados pelo método CoMFA padrão.

Uma vez gerados os descritores, foram especificados os seguintes parâmetros para a análise estatística com o método PLS para a obtenção do número ótimo de componentes para o modelo de QSAR: método de validação com “deixe um de fora” (LOO, do inglês, *Leave-one-Out*), para um número máximo de 6 componentes, empregando o escalonamento CoMFA padrão (*standard*).

Com o objetivo de obter um modelo com a máxima capacidade preditiva possível, foi utilizado ao recurso de *focagem da melhor região*, que consiste em atribuir

um peso maior para os pontos da grade que contribuem de maneira mais significativa para o modelo e um menor peso para pontos que apresentam redundância ou baixa variância dos dados. Uma maneira de se evidenciar os pontos de maior contribuição é fazer a diferenciação dos campos pelo coeficiente de desvio padrão (SDC). A partir da análise realizada previamente, atribui-se o peso que determinado ponto terá em uma nova análise pela multiplicação da variação dos valores dos campos de interação deste ponto pelo coeficiente que foi atribuído a este ponto pela análise anterior.

Finalmente, o método CoMFA permite avaliar as posições de substituição que fazem contribuições favoráveis ou desfavoráveis para a propriedade biológica em estudo, por meio de poliedros denominados *mapas de contorno*, que evidenciam regiões estereoquímicas e eletrostáticas ao redor das moléculas responsáveis pelas diferenças nos valores da propriedade biológica.

5.3.2. Resultados e Discussão

As 79 moléculas alinhadas para o processo de modelagem de QSAR 3D estão representadas na Figura 22. As características observadas neste modelo estão de acordo com a hipótese de ligação dos dicetoácidos e seus análogos em forma de complexo com os 2 átomos de Mg^{2+} e a tríade catalítica DDE.

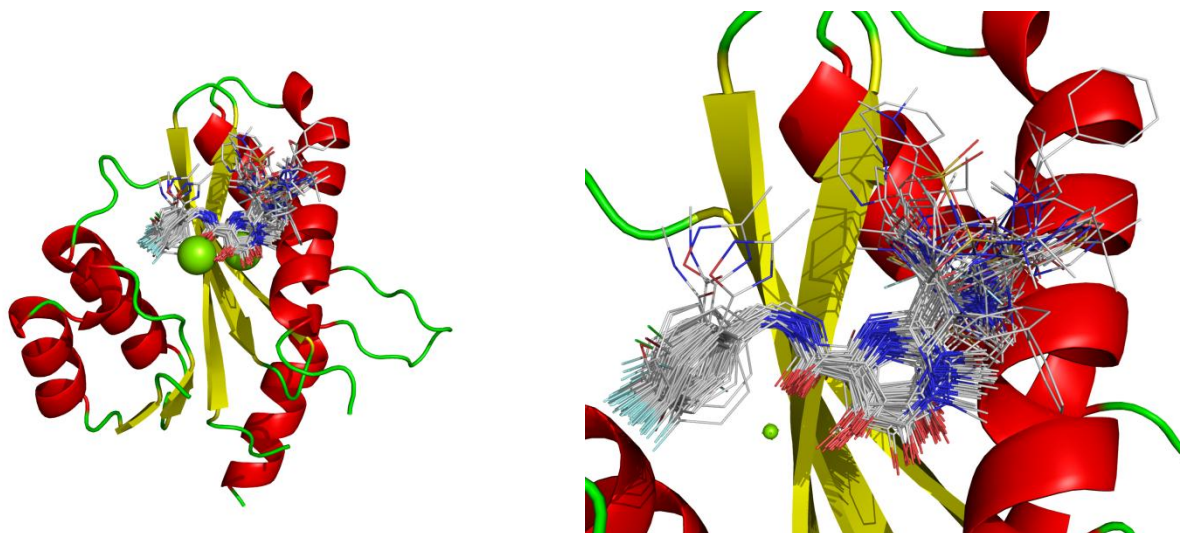


Figura 22. Alinhamento molecular estrutural do conjunto de dados.

Como o primeiro modelo CoMFA obtido apresentou parâmetros estatísticos insatisfatórios, foi empregado o recurso de focagem de melhor região, sendo possível obter um aumento na consistência interna e um refinamento dos mapas de contorno. Os melhores resultados foram obtidos quando os descritores relevantes foram potencializados pelo SDC com peso $w = 1,0$, como ilustrado na Tabela 8.

Tabela 8. Parâmetros estatísticos nas análises de PLS para o modelo CoMFA

Modelo	Parâmetros estatísticos			Porcentagens de contribuição	
	q^2	r^2	N	estérica	eletrostática
sem focagem	0,401	0,947	4		
W = 1.0	0,719	0,904	6	49%	51%

Com o intuito de avaliar o poder preditivo deste modelo, foram feitas as predições dos compostos do conjunto de teste, que foram submetidos aos mesmos procedimentos de minimização de energia, cálculo de cargas parciais e alinhamento

molecular tais como os compostos do conjunto de treinamento. Na Tabela 9 são mostrados os resultados das previsões para o conjunto teste, que indicam a boa capacidade preditiva do modelo obtido, uma vez que os resíduos de predição são bastantes baixos.

Tabela 9. Resíduos de predição para os compostos do conjunto de teste

Composto	pIC₅₀ Experimental	pIC₅₀ Predito	Resíduo
65	6,64	7,43	- 0,79
66	7,82	8,04	-0,22
67	7,70	7,87	-0,17
68	7,70	7,81	-0,11
69	8,15	7,86	0,29
70	7,28	7,65	-0,37
71	7,30	7,81	-0,51
72	8,00	7,19	0,81
73	7,39	7,21	0,18
74	7,05	7,03	0,02
75	7,10	7,65	-0,55
76	6,35	6,78	-0,43
77	8,10	7,55	0,55
78	7,70	7,66	0,04
79	6,70	6,67	0,03

A diferença entre os valores experimentais e preditos da Tabela 9 indica uma elevada consistência externa do modelo, com boa capacidade de predição. Dentre a 15 moléculas do conjunto teste, 8 apresentam resíduos abaixo de 0,2 unidades logarítmicas, 6 ficaram abaixo de 0,4 e apenas 1 acima de 0,4. O gráfico dos valores de afinidade pela HIV-1 IN experimentais *versus* preditos para os conjuntos de treinamento e teste é ilustrado na Figura 23, onde pode ser observada a boa concordância entre estes dados.

O processo de validação externa comprova que este modelo é capaz de prever a potência biológica de outras moléculas dentro desta classe química, representada pela diversidade contida dentro do espaço químico explorado.

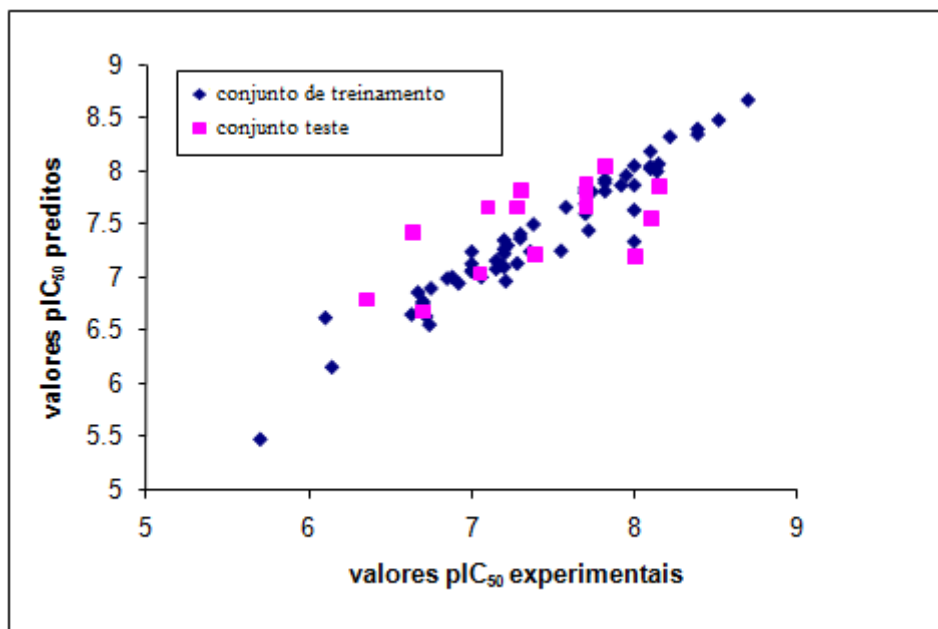


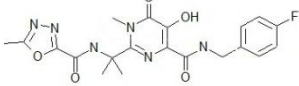
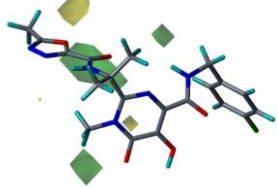
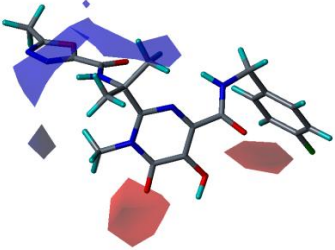
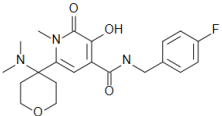
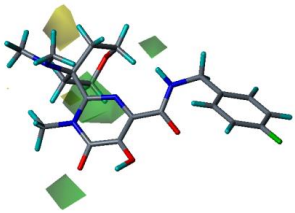
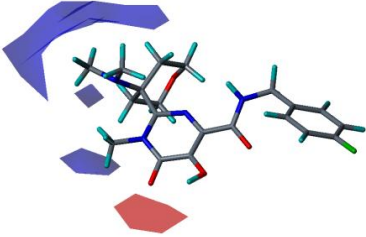
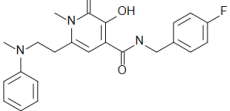
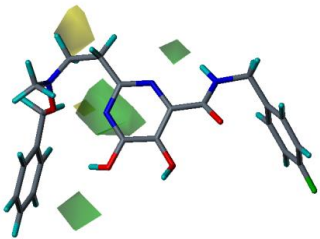
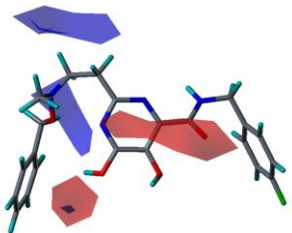
Figura 23. Valores de pIC_{50} preditos versus experimentais para o conjunto de treinamento e para o conjunto de teste resultantes do modelo CoMFA.

Os mapas de contorno 3D CoMFA permitem a visualização dos campos estéricos e eletrostáticos que apresentam maior importância para explicar as diferenças nos valores de afinidade da série de compostos em estudo, sendo úteis no trabalho químico medicinal, com ênfase nos estudos de SAR e síntese planejada.

Os mapas dos campos 3D nas cores verde e amarelo representam a característica estereoquímica: poliedros verdes sugerem substituições por grupos mais volumosos, enquanto poliedros amarelos sugerem grupos menos volumosos. Por

outro lado, os mapas azuis e vermelhos se referem às características eletrostáticas: poliedros azuis sugerem substituições por grupos positivos e os vermelhos sugerem substituições por grupos eletronegativos. O mapa de contorno é mostrado, de forma ilustrativa para o fármaco raltegravir (**12**), para a molécula mais potente (**2**), e menos potente (**37**) do conjunto de dados, na Tabela 10.

Tabela 10. Mapas de contorno 3D CoMFA estéricos e eletrostáticos para o fármaco raltegravir, para o composto com maior afinidade e menor afinidade.

Composto	Mapa de contorno estérico	Mapa de contorno eletrostático	pIC ₅₀ (nM)
 Raltegravir 12			7,8
 2			8,7
 37			5,7

As informações mais relevantes obtidas pelos mapas de contorno para o fragmento **A** sugerem, pela presença de poliedros verdes, a introdução de grupos mais

volumosos, e pela presença de poliedros azuis, grupos mais positivos. Esta informação é coerente com os estudos de docagem realizados para a obtenção do alinhamento estrutural. O melhor posicionamento das moléculas no sítio ativo da enzima permite que grupos sejam adicionados nestas posições. Moléculas como o fármaco raltegravir possuem um fragmento **A** maior com mais grupos positivos como o anel oxadiazol e o grupo carboxamida, ambos com átomos de nitrogênio, em relação aos inibidores menos ativos. A presença desses grupos é necessária para as interações com os resíduos polares K156, K159 e T66 através de ligações hidrogênio.

Já a presença de poliedros vermelhos na região quelante indica que altas densidades eletrônicas podem aumentar as afinidades dos ligantes. Substituintes eletronegativos nessa posição exercerão um efeito maior sobre a afinidade com os íons de Mg^{2+} .

5.4. HQSAR

A metodologia HQSAR (do inglês, *Hologram Quantitative Structure-Activity*), tem como objetivo buscar parâmetros bidimensionais e encontrar informações necessárias para a síntese de novos ligantes. Com a metodologia HQSAR é possível quantificar as variações nos hologramas moleculares de uma série de moléculas com o emprego do método PLS para a correlação estatística, explicando assim as diferenças na atividade biológica do conjunto estudado. Em geral, os resultados produzidos apresentam qualidade comparável a das técnicas tridimensionais, como CoMFA, com a vantagem

de eliminar a necessidade de estabelecer uma conformação ótima e um alinhamento estrutural das moléculas da série.⁷⁰

5.4.1. Procedimento Computacional

Assim como no modelo apresentado anteriormente, foi utilizado o conjunto treinamento com 64 compostos e o conjunto teste com 15 compostos. Esses dados para a modelagem de HQSAR são apresentados na Tabela 7.

Na metodologia HQSAR, cada molécula do conjunto de dados é dividida em uma série de fragmentos estruturais únicos, que são arranjados para formar um holograma molecular. Diferentemente de outros métodos baseados em *molecular fingerprints*, o método HQSAR considera todos os fragmentos moleculares possíveis, assim como informação tridimensional sobre hibridização e quiralidade das moléculas.

A análise HQSAR envolve três passos principais: a geração de fragmentos sub-estruturais para cada uma das moléculas do conjunto de treinamento; a transformação desses fragmentos em hologramas; e a correlação destes com os dados biológicos disponíveis. De acordo com este protocolo, uma molécula é descrita como uma série de números ou “bins” (holograma molecular). Os bins representam cada um dos fragmentos incluídos em uma molécula em particular e são atribuídos por um algoritmo de redundância cíclica (do inglês, CRC – *cyclic redundancy check*).

Para a geração dos modelos, primeiramente foram especificados os seguintes parâmetros relativos à obtenção dos descritores:

i. Distinção de fragmentos: todos os fragmentos estruturais lineares, de ramificação e de sobreposição foram utilizados para gerar os descritores HQSAR. Os fragmentos estruturais de cada molécula original consistiam de números de átomos mínimo e máximo previamente definidos, que são chamados de parâmetros de *tamanho de fragmento*. A informação em cada fragmento foi definida por parâmetros de *distinção de fragmentos*, incluindo A (átomos), B (ligações), C (conectividade), H (átomos de hidrogênio), Ch (quiralidade) e DA (doadores e aceptores de hidrogênio). Estas combinações são mostradas a seguir: A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/B/C/H/Ch/DA, A/B/H, A/B/C/Ch, A/B/DA, A/B/H/DA, A/B/C/Ch/DA, A/B/C/H/DA E A/B/H/Ch/DA. Os fragmentos gerados foram então divididos em um arranjo de comprimento fixo para produzir um holograma molecular. Este comprimento fixo foi definido como parâmetro *de comprimento de holograma*.

ii. Comprimento do holograma: as análises foram realizadas utilizando-se os seguintes comprimentos de holograma: 53, 69, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401. Esses são números primos para minimizar a possibilidade da chamada “*fragment collision*” (quando os fragmentos mais importantes para a atividade biológica são atribuídos ao mesmo *bin* no holograma, sem discriminação).

iii. Tamanho dos fragmentos: inicialmente foi utilizado o tamanho de fragmento 4-7.

A natureza particular dos fragmentos sub-estruturais gerados pelo HQSAR e, conseqüentemente, a informação contida nos hologramas moleculares resultantes é alterada pelo ajuste desses parâmetros. Finalmente, a correlação dos descritores obtidos dessa forma com a propriedade biológica foi feita empregando o método PLS.

5.4.2. Resultados e Discussão

Diversas análises estatísticas de HQSAR foram conduzidas com base nos parâmetros selecionados. Ao final do processo de modelagem, foram obtidos os resultados apresentados na Tabela 11.

Tabela 11. Resultados da análise HQSAR utilizando vários tipos de fragmentos, com o tamanho de fragmento padrão 4 -7

Modelo	Tipo de Fragmento	q^2	SEP	r^2	SEE	HL	N
1	A/B	0,482	0,461	0,737	0,329	353	3
2	A/B/C	0,464	0,469	0,722	0,388	353	3
3	A/B/C/H	0,510	0,461	0,894	0,214	257	6
4	A/B/C/H/Ch	0,547	0,443	0,857	0,233	97	6
5	A/B/C/H/Ch/DA	0,643	0,349	0,908	0,199	61	6
6	A/B/H	0,469	0,480	0,847	0,258	61	6
7	A/B/C/Ch	0,505	0,451	0,740	0,327	353	3
8	A/B/DA	0,527	0,449	0,823	0,274	83	5
9	A/B/C/DA	0,504	0,464	0,895	0,214	151	6
10	A/B/H/DA	0,561	0,433	0,834	0,266	83	5
11	A/B/C/Ch/DA	0,554	0,440	0,882	0,227	307	6
12	A/B/C/H/DA	0,516	0,450	0,752	0,322	71	4
13	A/B/H/Ch/DA	0,600	0,416	0,870	0,238	83	6

q^2 , coeficiente de correlação da validação cruzada; SEP, erro padrão da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão de calibração; HL, comprimento do holograma; N, número ótimo de componentes PLS. Tipos de Fragmentos: A, átomos; B, ligações; C, conexões; H, átomos de hidrogênio; Ch, quiralidade; DA, doador e receptor.

Como pode ser observado, o modelo que produziu os melhores resultados estatísticos ($q^2 = 0,643$ e $r^2 = 0,908$) foi aquele com a combinação de distinção de fragmentos A/B/C/H/Ch/DA, apresentando comprimento no holograma 61 e número ótimo de componentes de PLS igual a 6. Este modelo foi selecionado para investigação da influência da variação do tamanho dos fragmentos sobre os parâmetros estatísticos. Desta forma, foram realizadas análises individuais empregando tamanhos de fragmentos distintos: 2-5, 3-6, 5-8, 6-9, 7-10 e 8-11, representados no intervalo pelo número mínimo e máximo de átomos. A Tabela 12 ilustra os resultados obtidos neste processo.

Tabela 12. Resultados da análise HQSAR variando o tamanho do fragmento, com a combinação A/B/C/H/Ch/DA

Tamanho de fragmentos	q^2	SEP	r^2	S	HL	N
1-4	0,390	0,514	0,842	0,262	401	6
2-5	0,506	0,463	0,895	0,214	151	6
3-6	0,563	0,435	0,896	0,212	401	6
4-7	0,643	0,394	0,908	0,199	61	6
5-8	0,559	0,433	0,855	0,248	199	5
6-9	0,586	0,426	0,916	0,190	307	6
7-10	0,577	0,428	0,912	0,196	151	6

Nesta série comparativa, há uma diferença significativa entre os fragmentos estudados, sendo o modelo gerado com o tamanho do fragmento 4-7 para a combinação A/B/C/H/Ch/DA, o melhor resultado estatístico ($q^2 = 0,643$ e $r^2 = 0,908$). O modelo final de HQSAR gerado com o conjunto treinamento possui a validação cruzada

como indicativo de robustez. Este parâmetro demonstra a capacidade de correlação dos dados e a consistência interna do modelo. Contudo, apenas esta validação não pode ser considerada suficiente, pois bons coeficientes de correlação associados a uma elevada consistência interna muitas vezes não resultam em modelos com capacidade de prever a propriedade alvo para novos compostos.

Para que o modelo de QSAR apresente utilidade no planejamento de fármacos, é necessário que o mesmo passe por um processo de validação externa para uma avaliação da sua real capacidade preditiva. O processo de validação externa do modelo de HQSAR mais robusto foi feito com o emprego do conjunto teste apresentado na Tabela 7. As moléculas contidas no conjunto teste não foram utilizadas na 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 na Tabela 13.

As diferenças modestas entre os valores experimentais e preditos (resíduos da Tabela 13) indicam a boa capacidade preditiva do modelo gerado.

Tabela 13. Resíduos de predição para os compostos do conjunto teste

Composto	pIC₅₀ Experimental	pIC₅₀ Predito	Resíduo
65	6,64	7,129	-0,489
66	7,82	8,133	-0,313
67	7,70	7,905	-0,205
68	7,70	8,041	-0,341
69	8,15	8,010	0,140
70	7,28	7,026	0,254
71	7,30	6,967	0,333
72	8,00	7,213	0,787
73	7,39	6,447	0,943
74	7,05	7,681	-0,631
75	7,10	7,722	-0,622
76	6,35	7,246	-0,896
77	8,10	7,959	0,141
78	7,70	7,563	0,137
79	6,70	7,209	-0,509

A Figura 24 apresenta o gráfico dos valores experimentais contra os valores preditos para os conjuntos treinamento e teste. A análise gráfica mostra a boa concordância dos valores preditos e experimentais. O modelo final gerado apresenta, portanto, significativa consistência interna e externa para a diversidade estrutural estudada, possuindo comprovada capacidade preditiva.

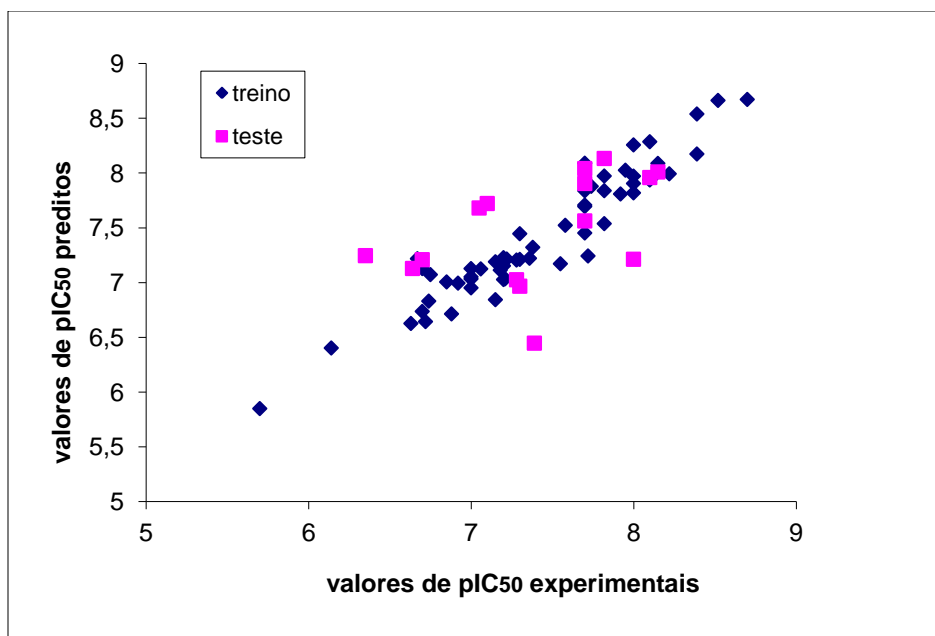


Figura 24. Gráfico HQSAR dos valores experimentais e preditos para as 79 moléculas do conjunto de dados.

Além de prever a potência de novas moléculas não incluídas no conjunto de treinamento, um modelo de HQSAR fornece importantes indicativos sobre os diversos fragmentos moleculares que estão relacionados com a atividade biológica. Esta informação é codificada em um mapa de contribuição, que apresenta cada átomo de uma determinada molécula e a sua contribuição para a atividade biológica. Neste mapa, um código de cores é atribuído, refletindo a contribuição individual de cada átomo e dos diferentes fragmentos da molécula. Cores na região do vermelho no 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 coloridos em branco. Na Figura 25 é mostrado o mapa de contribuição para o inibidor mais potente (**2**) do conjunto de dados.

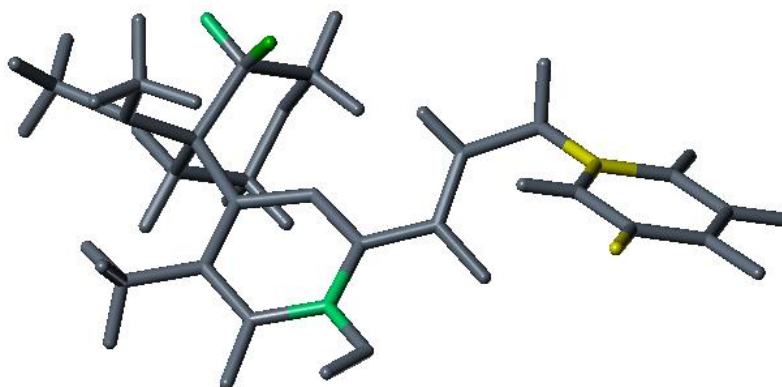


Figura 25. Mapa de contribuição de HQSAR para o inibidor **2**.

Como pode ser observado, o mapa de contribuição fornece informações importantes, que associadas ao conhecimento em química sintética, podem ser úteis no planejamento da síntese de novas moléculas com propriedades otimizadas. Por exemplo, os átomos em verde e amarelo são fragmentos que contribuem de forma positiva para a atividade biológica desta molécula.



Figura 26. Mapa de contribuição de HQSAR para o fármaco raltegravir (**12**).

O mesmo pode ser observado para o fármaco raltegravir mostrado na Figura 26. O átomo de hidrogênio em verde no grupo metila presente no fragmento **A** faz ligação hidrogênio com o oxigênio do resíduo T66.



Figura 27. Mapa de contribuição de QSAR para o inibidor menos ativo (**37**).

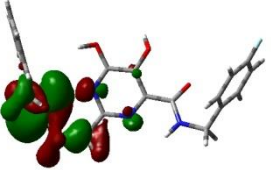
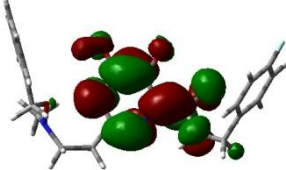
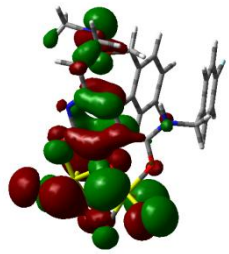
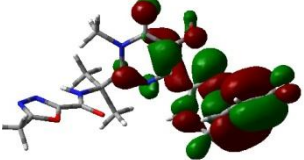
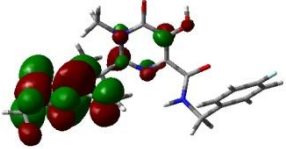
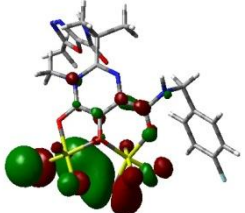
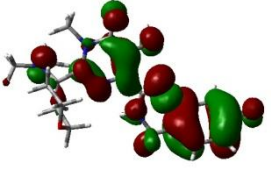
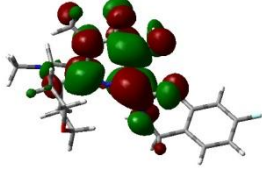
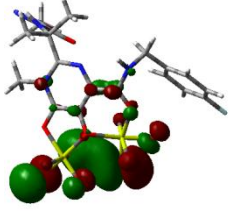
Já para o composto menos ativo (**37**), Figura 27, os átomos em cor vermelha no fragmento **A** e no anel do domínio quelante contribuem de forma negativa para a atividade biológica deste composto.

5.5. Relação entre orbitais moleculares e valores $\log IC_{50}$

Como demonstrado no item 4.3.3 do capítulo 4, a posição do HOMO nas moléculas antes da complexação com os íons de Mg^{2+} e a posição do LUMO após a complexação são muito importantes para a atividade dessas moléculas. Por esse motivo, novamente foi realizado o estudo da relação entre os orbitais moleculares para as moléculas *hidroxipirimidinonas*. A Tabela 14 apresenta esses resultados. Para o

fármaco raltegravir, assim como para as moléculas naftiridinas, o HOMO encontra-se principalmente sobre os átomos de oxigênio do domínio quelante.

Tabela 14. HOMO e LUMO para as moléculas mais ativa, menos ativa e para o fármaco raltegravir.

Composto	pIC ₅₀ (nM)	HOMO	LUMO	LUMO
37	5.7			
12 raltegravir	7.8			
2	8.7			

O mesmo acontece para a molécula mais ativa (composto **2**). Já para a molécula menos ativa (composto **37**), o HOMO encontra-se fora da região farmacofórica. Comparando agora o LUMO após a complexação, como já esperado, o mesmo encontra-se sobre os íons de magnésio tanto para o fármaco como para a molécula mais ativa, ao passo que este orbital está diluído sobre boa parte da molécula menos ativa. Outro ponto importante que deve ser ressaltado sobre as atividades das moléculas diz respeito à vizinhança dos átomos de oxigênio, que se ligam aos íons de magnésio. Para os compostos de *hidroxipirimidinonas*, os átomos de oxigênios,

correspondentes aos grupos carbonila e hidroxila, estão ligados a um anel aromático com caráter nucleofílico maior que os átomos de oxigênio dos compostos naftiridinas e dicetoácidos.

5.6. Conclusões

Neste capítulo, foram apresentados modelos de *docking* molecular, CoMFA e HQSAR aplicados com sucesso para o conjunto de inibidores da IN. Pela análise do *docking* molecular foi possível verificar o modo de ligação interfacial desses ligantes no sítio da enzima através da quelação dos 2 íons de Mg^{2+} . Além disso, foi possível observar que o fármaco raltegravir e as moléculas mais potentes apresentaram um melhor posicionamento do fragmento A em uma das cavidades do sítio ativo, devido às ligações de hidrogênio com os resíduos T66, K159 e K156. O modelo de QSAR 3D resultante foi capaz de fornecer uma correlação significativa entre os campos estéricos e eletrostáticos com os resultados obtidos pela docagem. Os mapas de contorno indicaram que na região quelante são necessários grupos mais eletronegativos, e que no fragmento A substituições por grupos mais volumosos e mais positivos contribuem com as interações com os resíduos polares T66, K159 e K156. As informações obtidas com os mapas de contribuição HQSAR indicam a importância de substituintes no fragmento A para o planejamento de novos compostos inibidores da IN.

O estudo teórico com o método DFT indicou que, para as moléculas mais ativas, os orbitais moleculares HOMO estão localizados principalmente no domínio quelante. Logo há uma importante contribuição das cargas negativas sobre os átomos

de oxigênio que estão envolvidos na coordenação com os cátions de Mg^{2+} . Este efeito do orbital produz um elevado caráter nucleofílico sobre esses sítios ativos, necessário para deslocar duas moléculas de água a partir de $[Mg(H_2O)_6]^{2+}$. Finalmente, o LUMO nas moléculas mais ativas em $[L-2Mg(H_2O)_4]^{4+}$ está localizado na região dos íons Mg^{2+} , possibilitando a ligação com os átomos de oxigênio dos resíduos aspartato (D64 e D116) e glutamato (E152) no sítio ativo da enzima.

6. CONCLUSÕES FINAIS

A enzima integrase do vírus HIV-1 exerce papel fundamental no ciclo de replicação viral. Os inibidores da integrase representam o maior avanço nas pesquisas sobre AIDS, mostrando potentes efeitos antiretrovirais em análises clínicas avançadas. Embora se tenham feito estudos ao longo de décadas nesse campo, várias questões com relação às interações dessa enzima com seus inibidores permanecem sem resposta, como os aminoácidos envolvidos na ligação, possíveis interações dos metais com o DNA viral, resistência a mutações e conformações assumidas pelos inibidores em complexo com a enzima. A elucidação destas questões é crucial, dada a exigência em restringir a IN de catalisar a inserção do DNA viral no genoma humano, levando a latência retroviral e persistência durante a terapia.

Nesse contexto, o presente trabalho teve o intuito de contribuir para o entendimento das características dos compostos derivados dos β -dicetoácidos responsáveis pelas interações com a enzima integrase. Os principais resultados dos estudos de *docking* e QSAR indicam algumas características como sendo importantes para altas afinidades. A presença de grupos volumosos e mais positivos no fragmento **A** permitiria ligações hidrogênio com os resíduos K156, K159 e T66, conforme indicado pelos mapas de contribuição gerados pelo método CoMFA e estudos de *docking* molecular. Sabe-se que esses resíduos estão envolvidos nas interações com a fita de DNA viral. Outro dado importante é a eletronegatividade da região quelante. Com os mapas de HOMO obtidos por cálculos quânticos e com os mapas de contribuição

eletrostática gerados pelo método CoMFA verificou-se que compostos mais ativos possuem maior caráter nucleofílico necessário para a ligação com os íons de Mg^{2+} .

Assim podemos afirmar que os métodos escolhidos foram empregados com sucesso na caracterização dos requisitos estruturais necessários para a interação no sítio ativo da enzima integrase do vírus HIV-1. A aplicação de estratégias integradas em química quântica medicinal é importante no planejamento de fármacos. O uso de métodos baseados na estrutura do receptor em conjunto com o emprego de técnicas de *Docking* Molecular, QSAR 3D (CoMFA) e QSAR 2D (HQSAR), confirma que é válido associar estratégias procurando uma sinergia útil no planejamento de candidatos a novos fármacos. Os modelos finais de QSAR com boa consistência interna, habilidade preditiva e estabilidade obtidos devem ser úteis para guiar futuros trabalhos de química medicinal no planejamento de novos inibidores HIV-1 IN.

REFERÊNCIAS

1. BULTINCK, P.; TOLLENAERE, J. P.; WINTER, H. **Computational medicinal chemistry for drug discovery**. New York: Marcel Dekker, 2003. 794 p.
2. EDGEcombe, K. E.; WEAVER, D. F.; SMITH JR, V. H. **Electronic structure analysis of compounds of interest in drug design: mono- and dicarboxylated pyridines**. *Canadian Journal of Chemistry*, v. 72, p. 1388-1403 1994.
3. Joint United Nations Programme on HIV/AIDS. <<http://www.unaids.org>>. Acesso em: 15 de maio 2011.
4. COHEN, J. Confronting the limits of success. *Therapies Science*, v. 296, p. 2320-2324, 2002.
5. FESEN, M. R.; KOHN, K. W.; LETEURTRE, F.; POMMIER, Y. Inhibitors of human immunodeficiency virus integrase. **Proceedings of the National Academy of Sciences**, v. 90, p. 2399-2403, 1993.
6. HEUER, T. S.; BROWN, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. **Biochemistry**, v. 36, p. 10655-10665, 1997.
7. MIKOL, V.; CLÉMENT-MELLA, C.; MAIGNAN, S.; GUILLOTEUAG, J. P.; ZHOU-LIU, Q. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. **Proceedings of the National Academy of Sciences**, v. 97, n. 15, p. 8233-8238, 2000.
8. GALLO, R. C.; MONTAGNEIR, L. Explanations of the origin(s) of the HIV and the etiology of AIDS. **Scientific American**, v. 41, p. 259, 1988.
9. TIEMESSEN, C. T.; HUNT, G. M.; PAPATHANASOPOULOS, M. A. Evolution and diversity of HIV-1 in africa. **Virus Genes**, v. 26, p. 151-163, 2003.
10. FRANKEL, A. D.; YOUNG, J. A. T.; **Annu. Rev. Biochem**, v.1, p. 67, 1998.
11. VAISHNAV, Y. N.; WONG-STAAAL, F. **Annu. Rev. Biochem**, v. 60, p. 577, 1991.
12. DEEN, K. C.; MCDUGAL, J. S.; INACKER, R.; FULENA-WASSERMAN, G.; ARTHOS, J. **Nature**, v. 82, p.331, 1988
13. POMMIER, Y.; MADDALI, K.; METIFIOT, M.; MARCHAND, C. HIV-1 IN inhibitors: 2010 update and perspectives. **Current Topics in Medicinal Chemistry**. V. 9, p. 1016-1037, 2009

14. CRAIGIE, R.; WANG J.; LING H.; YANG W. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. **The EMBO Journal**, v. 20, n. 24, p. 7333-7343, 2001.
15. MIKOL, V.; CLÉMENT-MELLA, C.; MAIGNAN, S.; GUILLOTEAU, J P., ZHOU-LIU, Q. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. **Proceedings of the National Academy of Sciences**, v. 97, n. 15, p. 8233-8238, 2000.
16. POMMIER, Y.; JOHNSON A., A.; MARCHAND, C. Integrase inhibitors to treat HIV/AIDS. **Nature**, v. 4, p. 236-248, 2005.
17. DAVIES, D. R. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. **Science**, v. 289, p. 77-85, 2000.
18. WIELENS, J.; CROSBY, I. T., CHALMERS, D. K. A three-dimensional model of the human immunodeficiency virus type 1 integration complex. **Journal of Computer-Aided Molecular Design**, v. 19, p. 301-317, 2005.
19. HAZUDA, D. J. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. **Science**, v. 287, p. 646-650, 2000.
20. GOLDGUR, Y.; DYDA, F.; HICKMAN, A. B.; JENKINS, T. M.; CRAIGIE, R.; DAVIES, D. R. Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. **Proceedings of the National Academy of Sciences**, v. 95, p. 9150-9154, 1998.
21. POMMIER, Y.; CHERFILS, J. Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. [Trends in Pharmacological Sciences](#), v. 26, p.138-145, 2005.
22. ZHUANG, L.; WAI, J. S.; VACCA, J. P. Design and synthesis of 8-Hydroxy-[1,6]Naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells. **Journal of Medicinal Chemistry**, v. 46, p. 453-456, 2003.
23. EMBREY, M. W.; HAZUDA, D. J.; ZHUANG, L.; VACCA, J. P. A series of 5-(5,6)-dihydrouracil substituted 8-hydroxy-[1,6]naphthyridine-7-carboxylic acid 4-fluorobenzamide inhibitors of HIV-1 integrase and viral replication in cells. [Bioorganic & Medicinal Chemistry Letters](#), v. 15, p. 4550-4554, 2005.
24. GUARE, J. P.; WAI, J. S.; GOMEZ, R. P.; ANTHONY, N. J.; JOLLY, S. M.; VACCA, J. P.; HAZUDA, J. D. A series of 5-aminosubstituted 4-fluorobenzyl-8-hydroxy-[1,6]naphthyridine-7-carboxamide HIV-1 integrase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v.16, p. 2900-2904, 2006.

25. SUMMA, V.; PETROCCHI, A.; MATASSA, V. G.; GARDELLI, C.; MURAGLIA, E.; ROWLEY, M.; PAZ, O. G.; LAUFER, R.; MONTEAGUDO, E.; PACE, P. 4,5-Dihydroxypyrimidine Carboxamides and *N*-Alkyl-5-hydroxypyrimidinone carboxamides are potent, selective HIV integrase inhibitors with good pharmacokinetic profiles in preclinical species. **Journal of Medicinal Chemistry**, v. 49, p. 6646-6649, 2006.
26. PACE, P.; GOMEZ, R. P.; ANTHONY, N. J.; JOLLY, S. M.; VACCA, J. P.; HAZUDA, J. D. Dihydroxypyrimidine-4-carboxamides as novel potent and selective HIV integrase inhibitors. **Journal of Medicinal Chemistry**, v. 50, p. 2225-2239, 2007.
27. GARDELLI, C.; GOMEZ, R. P.; ANTHONY, N. J.; JOLLY, S. M.; VACCA, J. P.; HAZUDA, J. D. Discovery and synthesis of HIV integrase inhibitors: development of potent and orally bioavailable *N*-Methyl pyrimidones. **Journal of Medicinal Chemistry**, v. 50, n. 20, p. 4953-4975, 2007.
28. SUMMA, V. *et al.* Development of 2-*t*-butyl-*N*-methyl pyrimidones as potent inhibitors of HIV integrase. **Bioorganic & Medicinal Chemistry Letters**, 2008.
29. SUMMA, V.; EMBREY, M. W.; HAZUDA, D. J.; ZHUANG, L.; VACCA, J. P. Discovery of raltegravir, a potent, selective orally bioavailable HIV-1 inhibitor for the treatment of HIV-AIDS infection. **Journal of Medicinal Chemistry**, v. 51, p. 5843-5855, 2008.
30. GOLDBERG, Y.; CRAIGIE, R.; COHEN, G. H.; FUJIWARA, T.; YOSHINAGA, T.; FUJISHITA, T.; SUGIMOTO, H.; ENDO, T.; MURAI, H.; DAVIES, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. **Proceedings of the National Academy of Sciences**, v. 96, p. 13040–13043, 1999.
31. ESPOSITO, D.; CRAIGIE, R. Sequence selectivity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interactions. **EMBO JOURNAL**, v. 17, p. 5832–5843, 1998.
32. JENSEN, F. **Introduction to Computational Chemistry**. New York: John Wiley 1999.
33. LEACH, A. R. **Molecular modeling. Principles and applications**. 2 ed. Personal Education, 2001
34. DEWAR, M. J. S.; ZOEBISH, E. G.; HEALY, E. F.; STEWART, J. J. P. The development and use of quantum-mechanical molecular-models. 76. AM1 – a new general-purpose quantum-mechanical molecular-model. **Journal of the American Chemical Society**, v. 107, p. 3909, 1985.
35. STEWART, J. J. P. Optimization of parameters for semiempirical methods. 1. method. **Journal of Computational chemistry**, v. 10, p. 209-220, 1989.

36. DEWAR, M. J. S., THIEL, W. **Journal of the American Chemical Society**. V. 99, p. 4899-4907, 1977.
37. BECKE, A. D. **J. Chem. Phys.** v. 98, p.5648, 1993
38. KUBINYI, H. **QSAR: Hansh analysis and related approaches**. New York: VCH, 1993. 240p.
39. HANSCH, C.; MALONEY, P.P.; FUJITA, T.; MUIR R.M. **Nature**, v. 194, p. 178, 1962.
40. HONÓRIO, K. M.; DA SILVA, A. B. F. An AM1 study on the electron-donating and electron-accepting character of biomolecules. **International Journal of Quantum Chem.**, v. 95, p. 126-132, 2003.
41. ATKINS, P. W. **Quanta**. 2. ed. New York: Oxford University Press, 1991. 434p
42. HIGO, J.; GO, N. Algorithm for rapid calculation of excluded volume of large molecules. **Journal of Computational Chemistry**, v.10, p. 367-379, 1989
43. MAGALHÃES, C. S.; BARBOSA, H. J. C.; DARDENNE, L. E. Métodos de docking receptor-ligante para o desenho racional de compostos bioativos. In: MORGON, N. H.; COUTINHO, K. (Ed.). **Métodos de química teórica e modelagem molecular**. São Paulo: Livraria da USP, 2007. p. 489-531.
44. KUKOL, A. (Ed.). **Molecular modeling of proteins**. Totowa: Humana Press, 2008. 275 p. (Methods in Molecular Biology, v. 443).
45. KITCHEN, D. B.; DECOREZ, H.; FURR, J. R.; BAJORATH, J. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature Reviews in Drug Discovery**, v. 3 p. 935-948, 2004.
46. GOLD 5.1:Cambridge, UK, 2004. [Programa de Computador]
47. JONES, G.; WILLETT, P.; GLEN, R. C.; LEACH, A. R.; TAYLOR, R. Development and validation of a genetic algorithm for flexible docking. **Journal of Molecular Biology**, v. 267, p. 727-748, 1997.
48. NISSINK, J. W. M.; MURRAY, C. W.; HARTSHORN, M. J.; VERDONK, M. L.; COLE, J. C.; TAYLOR, R. A new test set for validating predictions of protein ligand interactions. **Proteins**, v. 49, p. 457-471, 2002.
49. SOUZA, S. F.; FERNANDES, P. A.; RAMOS, M. J. Protein-ligand docking: current status and future challenges. **Proteins: Structure, Function, and Bioinformatics**, v. 65, p. 15-26, 2006.

50. TRIPOS INC. **SYBYL 8.1** St Louis [Programa de Computador]
51. WATERBEEMD, H. **Advanced computer assisted techniques in drug discovery**. New York: VCH, 1994.
52. QSAR and Advanced CoMFA manual. Tripos Inc., St. Louis, 2005.
53. KUBINIY, H. **3D QSAR in drug design: theory, methods and applications**, Leiden: Klumer Academic, 1993.
54. DODDAREDDY, M.; LEE, Y. Hologram quantitative structure activity relationship studies on 5-HT₆ antagonists. **Bioorganic Medicinal Chemistry**, v. 12, p. 3815-3824, 2004.
55. LOWIS, D. R. HQSAR: a new, highly predictive QSAR technique. Tripos Technical Notes, 1, 1997.
56. HQSARTM Manual, Tripos Inc, St. Lois, 2006.
57. CAMARGO, A. J. **Estudo químico-quântico ab initio e semi-empírico de compostos inorgânicos com possíveis aplicações tecnológicas**. 2001. 195f. Tese (Doutorado) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2001.
58. EL-AZHARY, A. A.; SUTTER, H. U. Comparison between optimezed geometries and vibrational frequencies calculated by the DFT methods. **Journal of Physical Chemistry**, v.100, p. 15056, 1996.
59. TURECEK, F. proton affinity of dimethyl sulfoxide and relative stabilities of C₂H₆OS molecules and C₂H₇OS⁺ ions. A comparative G2(MP2) ab initio and density functional theory study. **Journal of Physical Chemistry A**. v. 1012, n. 34, p. 4703-4713, 1998.
60. FRISCH, M. J.; TRUCKS, G. W.; SCHLEGEL, H. B.; SCUSERIA, G. E.; ROBB, M. A.; CHEESEMAN, J. R.; MONTGOMERY, J. A.; VREVEN, J. T.; KUDIN, K. N.; BURANT, J. C.; MILLAM, J. M.; LYENGAR, S. S.; TOMASI, J.; BARONE, V.; MENNUCCI, B.; COSSI, M.; SCALMANI, G.; REGA, N.; PETERSSON, G. A.; NAKATSUJI, H.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; KLENE, M.; LI, X.; KNOX, J. E.; HRATCHIAN, H. P.; CROSS, J. B.; ADAMO, C.; JARAMILLO, J.; GOMPERTS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; AYALA, P. Y.; MOROKUMA, K.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; ZAKRZEWSKI, V. G.; DAPPRICH, S.; DANIELS, A. D.; STRAIN, M. C.; FARKAS, O.; MALICK, D. K.; RABUCK, A. D.; RAGHAVACHARI, K.; FORESMAN, J. B.; ORTIZ, J. V.; CUI, Q.; G. BABOUL, A.; CLIFFORD, S.; CIOSLOWSKI, J.; STEFANOV, B. B.; LIU, G.; LIASHENKO, A.;

PISKORZ, P.; KOMAROMI, I.; MARTIN, R. L.; FOX, D. J.; KEITH, T.; AL-LAHAM, M. A.; PENG, C. Y.; NANAYAKKARA, A.; CHALLACOMBE, M.; GILL, P. M. W.; JOHNSON, B.; CHEN, W.; WONG, M. W.; GONZALEZ, C.; AND POPLE, J. A. Gaussian 03. Pittsburgh PA: GAUSSIAN, 2003. [Programa de Computador]

61. BRENEMAN, C. M.; WIBERG, K. B. Determining atom-centered monopole from molecular electrostatic potentials – the need for high sampling density in formamide conformational analysis. **Journal of Computational Chemistry**, v.11, n. 3, p. 361-373, 1990.

62. HYPERCUBE, **Chemplus**: Release 4.5 for Windows. Ontario, 1995. [Programa de Computador]

63. TODESCHINI, R.; CONSONNI, V.; PAVAN, M. **Dragon 2.1**. Milano, 2002. [Programa de Computador]

64. BRUNS, R. E.; FAIGLE, J. F. G. Quimiometria. **Química Nova**, v. 8, p. 84-99, 1985.

65. WATERBEEMD, H., ed. **Chemometrics Methods in Molecular Design**. New York, VCH, 1995. V. 2, p. 346.

66. INFOMETRIX INC. **Pirouette 3.11**. Woodinville, 2002. [Programa de Computador]

67. BOYER, F.; PRASAD, J. 5,6-Dihydropyran-2-ones possessing various sulfonyl functionalities: potent nonpeptidic inhibitors of HIV protease. **Journal of Medicinal Chemistry**, v. 43, p. 843-848, 2000.

68. HAGEN, S.; DOMOAGALA, J. 4-Hydroxy-5,6-dihydropyrones as Inhibitors of HIV protease: the effect of heterocyclic substituents at C-6 on antiviral potency and pharmacokinetic Parameters. **Journal of Medicinal Chemistry**, v. 44, p. 2319-2332, 2001.

69. CLARK, M.; CRAMER III, R. D.; VAN OPDENBOSH, N. Validation of the general purpose tripos 5.2 force field. **Journal of Computational Chemistry**, v. 10, p. 982-1012, 1989.

70. HONÓRIO, K.; ANDRICOPULO, A. Hologram quantitative structure-activity relationships for a series of Farnesoid X Receptor Activators. **Bioorganic and Medicinal Chemistry**, v. 15, p. 3119-3125, 2005.