UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

Rafael Victório Carvalho Guido

**Planejamento de inibidores da enzima gliceraldeído-3-fosfato desidrogenase de
Trypanosoma cruzi: biologia estrutural e química medicinal.**

São Carlos
2008

RAFAEL VICTÓRIO CARVALHO GUIDO

**Planejamento de inibidores da enzima gliceraldeído-3-fosfato desidrogenase de
Trypanosoma cruzi: biologia estrutural e química medicinal.**

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos, da Universidade de São Paulo, para a obtenção do título de Doutor em Ciências.

Área de concentração: Física Aplicada - opção Física Biomolecular.

Orientador: Prof. Dr. Glaucius Oliva

São Carlos
2008

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

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

Guido, Rafael Victório Carvalho

Planejamento de inibidores da enzima gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi*: biologia estrutural e química medicinal. / Rafael Victório Carvalho Guido; orientador Glaucius Oliva.-- São Carlos, 2008.

302 p.

Tese (Doutorado – Programa de Pós-Graduação em Física - Área de concentração: Física Aplicada - opção Física Biomolecular) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Planejamento de fármacos. 2. Química Medicinal. 3. Cristalografia de proteína. 4. Enzima. 5. Inibidor. I. Título.

FOLHA DE APROVAÇÃO

Rafael Victorio Carvalho Guido

Tese apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências. Área de Concentração: Física Aplicada – Opção: Física Biomolecular.

Aprovado em: 18/04/2008

Comissão Julgadora

Prof. Dr. Glaucius Oliva

Instituição: IFSC/USP

Assinatura

Profa. Dra. Elizabeth Igne ferreira

Instituição: FCF/USP

Assinatura

Profa. Dra. Chung Man Chin

Instituição: UNESP

Assinatura

Prof. Dr. Eliezer Jesus de Lacerda Barreiro

Instituição: UFRJ

Assinatura

Profa. Dra. Vanderlan da Silva Bolzani

Instituição: UNESP

Assinatura

AGRADECIMENTOS

À Deus por iluminar o meu caminho.

Aos meus pais Paschoal e Arlete, ao meu irmão Rodolfo, minha noiva Luciana, a minha tia Marlene, a minha avó Ely, meus sogros Eliseu e Mara por sempre me apoiarem e acreditarem em mim.

Ao Prof. Dr. Glaucius Oliva pela confiança, suporte, orientação, exemplo de liderança e amizade.

Ao Prof. Dr. Adriano D. Andricopulo pelas oportunidades, incentivo, amizade, e acima de tudo, pelo exemplo de compromisso e dedicação.

Ao Prof. Dr. Gerhard Klebe e seu grupo de pesquisa (AGKlebe) pela oportunidade dada e pela orientação durante o período de doutorado sanduíche na Alemanha.

À Profa. Dra. Chung Man Chin pela oportunidade inicial dada e por me ensinar os primeiros passos nos estudos em Química Medicinal.

Ao grande amigo Prof. Dr. Marcelo S. Castilho pela orientação no período inicial do projeto e pelas colaborações estabelecidas.

Aos meus grandes amigos Prof. Dr. Fábio F. Perazzo, Dr. Fábio C. Cruz e todos os varzeanos pelas longas discussões sobre ciência, incentivo e amizade.

Aos meus grandes amigos Napoleão F. Valadares, Dr. André L. Ambrósio, Dr. João R. C. Muniz, Ricardo P. Nicoluci, Ms. Matheus P. Postigo, Dra. Monique Mantovani, Dr. Ney R. Leite, Alécio A. P. Júnior e Dr. Wanius Garcia pelas longas e divertidas discussões sobre os mais diversos temas da vida.

Aos professores, funcionários e colegas do Centro de Biotecnologia Molecular Estrutural do Instituto de Física de São Carlos.

À FAPESP pelo apoio financeiro.

Aos meus pais Paschoal e Arlete e ao meu irmão Rodolfo por me apoiarem sempre.

RESUMO

A Doença de Chagas, causada pelo parasita *Trypanosoma cruzi*, atinge cerca de um quarto da população da América Latina. Os fármacos disponíveis para o tratamento desta doença são inapropriados, apresentam baixa eficácia e sérios efeitos colaterais que limitam o seu uso. Esse grave panorama torna urgente a descoberta de novos agentes quimioterápicos para o tratamento seguro e eficaz da doença. A via glicolítica é principal forma de obtenção de energia de tripanosomatídeos. Um alvo molecular atrativo desta via bioquímica que desempenha papel essencial no controle do fluxo glicolítico do *Trypanosoma cruzi*, a enzima gliceraldeído-3-fosfato desidrogenase (GAPDH), foi selecionada neste trabalho de Tese para estudos em biologia estrutural e química medicinal visando à identificação e planejamento de novos inibidores enzimáticos. Neste contexto, triagens biológicas resultaram na identificação de compostos de origem natural e sintética com atividade inibitória *in vitro* frente à GAPDH de *T. cruzi*, ampliando a diversidade química de moduladores seletivos deste alvo. Estudos cinéticos e estruturais demonstraram o comportamento não cooperativo entre os sítios ativos da enzima GAPDH de *T. cruzi* em relação à interação com o cofator NAD^+ , fornecendo importantes evidências mecanísticas e estruturais para uma melhor compreensão das bases moleculares envolvidas no processo de reconhecimento molecular. Os estudos das relações quantitativas entre a estrutura e atividade (QSAR 2D e QSAR 3D) resultaram na geração de modelos com elevada consistência estatística interna e externa, além de alto poder preditivo da propriedade-alvo. Além disso, estudos de modelagem molecular e de QSAR 3D revelaram aspectos estruturais relevantes para o planejamento de inibidores seletivos da enzima GAPDH de tripanosomatídeos. Por fim, uma estratégia de triagem virtual baseado na estrutura do receptor foi empregada para a identificação de novos inibidores da GAPDH de *T. cruzi*, consistindo, entre outros, na aplicação

de filtros hierárquicos sucessivos envolvendo restrições farmacofóricas e estudos de docagem molecular que resultaram na seleção de 35 candidatos a inibidores da enzima-alvo. Os trabalhos integrando estudos em química medicinal e biologia estrutural apresentados nessa Tese de Doutorado significam importantes contribuições no desenvolvimento de bases científicas sólidas para o planejamento de novos inibidores potentes e seletivos da enzima GAPDH de *T. cruzi*, um alvo molecular de alta prioridade em nosso grupo de pesquisa.

Palavras-chaves: Planejamento de Fármacos, Química Medicinal, Biologia Estrutural, Enzimas, Inibidores, Modelagem Molecular

ABSTRACT

Parasitic diseases are the foremost threat to human health and welfare around the world. Chagas' disease (also called American trypanosomiasis) is a tropical parasitic disease which occurs in Latin America, particularly in South America. The currently available drugs for this parasitic disease have severe limitations, including poor efficacy and high toxicity. The crucial dependence of trypanosomatids on glycolysis as a source of energy makes the glycolytic enzymes promising targets for drug design. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from *Trypanosoma cruzi*, a key enzyme in the glycolytic cascade, has been selected as an attractive drug target in this PhD Thesis work for studies in structural biology and medicinal chemistry for the identification and design of new enzyme inhibitors. In this context, compounds from both natural and synthetic sources with *in vitro* inhibitory activity against *T. cruzi* GAPDH were identified by screening assays, improving the chemical diversity of selective modulators of the target. Kinetic and structural studies have demonstrated the non-cooperative behavior between the *T. cruzi* active sites in the interaction with the NAD⁺ cofactor, shedding some light on the mechanistic and structural determinants underlying the biochemical recognition phenomenon. Quantitative structure–activity relationships (2D QSAR 2D and 3D QSAR) were successfully created, resulting in statistically significant models with good predictive ability for untested compounds. In addition, molecular modeling and 3D QSAR studies highlighted important structural aspects to assist the design of novel trypanosomatid GAPDH inhibitors. Finally, a structure-based virtual screening approach was employed for the identification of novel inhibitors of *T. cruzi* GAPDH, consisting of several consecutive hierarchical, fast pharmacophore matching and molecular docking, which afforded 35 inhibitor candidates for the target enzyme. The integration of structural biology and medicinal chemistry studies presented in this PhD Thesis are

important contributions in the development of strong scientific basis for the design of new selective and potent inhibitors of GAPDH from *T. cruzi*, a molecular target of highest priority in our research group.

Keywords: Drug Design, Medicinal Chemistry, Structural Biology, Enzymes, Inhibitors, Molecular Modeling

LISTA DE FIGURAS

Figura 1.1	Visão geral do processo de Pesquisa e Desenvolvimento de um novo fármaco.	35
Figura 1.2	Etapas detalhadas das Fases Pré-Clinica e Clínica realizadas durante o processo de desenvolvimento de novos fármacos.	36
Figura 1.3	Etapas de identificação, seleção e otimização de substâncias bioativas durante o processo de descoberta de novos fármacos (fase pré-clínica).....	37
Figura 1.4	Número de NCEs aprovadas no período 1996-2006 pela agência americana FDA.	40
Figura 1.5	Número de NCEs aprovadas no período 1945-2004 pela agência americana FDA.	41
Figura 1.6	(A) Incidência da doença de Chagas na América Latina. (B) Distribuição geográfica dos inseto-vetor causador da doença.....	45
Figura 1.7	Fármacos utilizados para o tratamento da doença de Chagas.....	46
Figura 1.8	Distribuição mundial da Leishmaniose.....	48
Figura 1.9	Fármacos utilizados no tratamento da leishmaniose.....	50
Figura 1.10	Distribuição das enzimas utilizadas como alvo molecular dos fármacos aprovados pelo FDA.....	52
Figura 1.11	Esquema do metabolismo da glicose e glicerol no glicossoma. As enzimas constituintes do glicossoma são: 1. HK, hexokinase; 2. PGI, fosfoglicose isomerase; 3. PFK, 6-fosfofruto-kinase; 4. ALDO, frutose-difosfato aldolase; 5. GDH, glicerol-3-fosfato desidrogenase; 6. GK, glicerol kinase; 7. TIM, triosefosfato isomerase; 8. GAPDH, gliceraldeído-3-fosfato desidrogenase; 9. PGK, 3-fosfoglicerato kinase.....	54

Figura 1.12	Estrutura quaternária da enzima GAPDH de <i>T. cruzi</i> . (B) Reação enzimática catalizada pela GAPDH.	55
Figura 3.1	Exemplos de inibidores do tipo reversível. Os esquemas para os diferentes tipos de inibição são apresentados na parte superior. Os gráficos de Lineweaver-Burk característicos dos diferentes tipos de inibição são apresentados na parte inferior.....	76
Figura 3.2	Gel de poliacrilamida 15% (SDS-PAGE) referente às etapas de purificação da GAPDH.	77
Figura 3.3	Mistura de componentes constituintes do óleo da castanha de caju.....	81
Figura 3.4	Curvas de concentração x respostas dos inibidor enzimático ácido anacárdico (1).....	82
Figura 3.5	Curvas de concentração x respostas dos inibidores enzimáticos 17 e 18	83
Figura 3.6	Curvas de concentração x respostas dos inibidores enzimáticos 23 , 28 , 30 e 31	87
Figura 4.1	Esquema dos modelos (A) MWC e (B) KNF propostos para explicar o fenômeno de cooperatividade entre as subunidades de proteínas oligoméricas..	92
Figura 4.2	Influência da concentração do cofator NAD^+ na reação catalisada pela enzima GAPDH de <i>T. cruzi</i>	96
Figura 4.3	Influência da concentração do substrato G3P na reação catalisada pela enzima GAPDH de <i>T. cruzi</i>	96

Figura 4.4	(A) Curva de ajuste dos dados cinéticos para a interação do NAD^+ em escala logarítmica para determinação do valor da constante de Hill (n_H). (B) Gráfico de Hill correspondente à saturação dos sítios da GAPDH com NAD^+	99
Figura 4.5	(A) Curva de ajuste dos dados cinéticos para a interação do G3P em escala logarítmica para determinação do valor da constante de Hill (n_H). (B) Gráfico de Hill correspondente à saturação dos sítios da GAPDH com G3P.....	99
Figura 4.6	Esquema dos métodos de difusão de vapor em gota suspensa (A) e gota assentada (B) usados para a obtenção de cristais de macromoléculas (M) com o auxílio de agentes precipitantes (AP).....	101
Figura 4.7	Mapa de densidade eletrônica $2F_{\text{obs}} - F_{\text{calc}}$ (contorno de 1.0σ) no sítio de ligação do NAD^+ nas subunidades A, B, C e D das estruturas apo (ciano), holo (magenta), 1NAD (cinza) e 2NAD (laranja). Em destaque os resíduos que interagem com o cofator através de ligação de hidrogênio. Os elementos de estrutura secundária estão representados como modelo de fitas e as moléculas de água como esferas.....	107
Figura 4.8	Diagrama de Ramachandran para as diferentes formas de GAPDH de <i>T. cruzi</i> .	109
Figura 4.9	(A) Sobreposição estrutural das formas apo (ciano) e holo (magenta) da GAPDH. As setas indicam a movimentação do domínio de ligação do NAD^+ que se fecha sobre o domínio catalítico durante a transição apo-holo. Ângulos de abertura das subunidades A da forma apo (B) e forma holo (C).....	113
Figura 4.10	Rotâmeros da Met39 na forma apo (ciano) e na forma holo (magenta). Alteração induzida pela ligação do NAD^+ durante a transição apo-holo.....	113

- Figura 4.11** Parte superior: modelo esquemático das estruturas **apo**, **1NAD**, **2NAD** e **holo** da GAPDH de *T. cruzi*, subunidades que adotam a conformação apo estão representadas como círculo e subunidades que adotam a conformação holo estão representadas como quadrado. Parte inferior: sobreposição dos C α das subunidades A, B, C e D da estrutura **1NAD** (cinza) e **2NAD** (laranja) nas subunidades correspondentes da estrutura **apo** (ciano) ou da estrutura **holo** (magenta). Os valores de rmsd da sobreposição das subunidades equivalentes estão mostrados na parte inferior da figura. Para facilitar a visualização, todas as subunidades estão apresentadas na mesma orientação. 115
- Figura 4.12** Organização espacial das quatro subunidades da GAPDH de *B. stearothermophilus* (esquerda) e *T. cruzi* (direita). As subunidades estão coloridas e indicadas por letras. As linhas tracejadas indicam os eixos de simetria. Os eixos-Q e -C são perpendiculares ao plano do papel..... 117
- Figura 4.13** Alinhamento seqüencial das GAPDHs de *T. cruzi*, *B. stearothermophilus*, *P. versicolor*, músculo de coelho e *E. coli*. Os números acima das seqüências correspondem à seqüência da GAPDH de *T. cruzi*. Resíduos conservados entre as diferentes espécies estão coloridos em cinza. As setas indicam resíduos que podem ser importantes para o comportamento não cooperativo observado na GAPDH de *T. cruzi*..... 118

Figura 4.14	Contatos estabelecidos pelo resíduo de Thr54 entre três subunidades diferentes da GAPDH de <i>T. cruzi</i> . A subunidade A está colorida em verde, a B em ciano, a C em magenta e a D em amarelo. Os resíduos que interagem com a Thr54 estão ilustrados em modelo de bastões, as interações do tipo van der Waals estão em ilustradas como pontos e as ligações de hidrogênio estão ilustradas como linhas tracejadas.....	119
Figura 4.15	Comparação entre os contatos inter-subunidades nas formas apo , 1NAD , 2NAD e holo . As subunidades A estão ilustradas em verde, subunidades B em ciano, subunidades C em magenta e subunidades D em amarelo. Resíduos que realizam interações com as subunidades vizinhas estão mostrados como bastões e as ligações de hidrogênio como linhas tracejadas. As estruturas foram sobrepostas nas subunidades correspondentes da forma apo e todo o tetramero foi utilizado no protocolo de sobreposição.....	120
Figura 5.1	Etapas envolvidas no processo de docagem molecular.....	129
Figura 5.2	Distribuição dos valores de pIC ₅₀ para o conjunto treinamento, conjunto teste e conjunto de dados.	135
Figura 5.3	Exemplos de impressão digital molecular e holograma molecular.....	145
Figura 5.4	Esquema ilustrativo do processo de geração dos hologramas moleculares.....	146
Figura 5.5	Etapas envolvidas na geração de um modelo de HQSAR.....	148
Figura 5.6	Gráfico dos valores experimentais <i>versus</i> preditos de pIC ₅₀ obtido com o melhor modelo de HQSAR (modelo 16 , $r^2_{\text{pred}} = 0,92$) para o conjunto treinamento (quadrados) e conjunto teste (triângulos) de inibidores da GAPDH de <i>L. mexicana</i>	154

Figura 5.7	Mapa de contribuição HQSAR para o inibidor mais potente do conjunto de dados (36).	155
Figura 5.8	Etapas envolvidas na construção de um modelo CoMFA.....	158
Figura 5.9	Curva de potenciais utilizadas nos métodos de QSAR 3D. O potencial de Lennard-Jones (verde) apresenta uma variação bastante abrupta próximo à superfície molecular, já o potencial de Coulomb (vermelho) apresenta uma variação mais suave, entretanto, ambos potenciais apresentam singularidades próximos as posições atômicas (nesta figura representado pela origem dos dados). Para evitar valores excessivamente elevados, a avaliação dos potenciais é restringida por um valor de exclusão arbitrário. A função tipo Guassiana (azul) utilizada no método CoMSIA apresenta um termo dependente da distância mais suave que as outras funções, evitando assim qualquer tipo de singularidade.....	160
Figura 5.10	Alinhamento seqüencial das GAPDHs de <i>L. mexicana</i> , <i>T. cruzi</i> , <i>T. brucei</i> , e <i>H. sapiens</i> . Os números acima das seqüências correspondem à seqüência da GAPDH de <i>L. mexicana</i> . Resíduos conservados entre as GAPDHs de diferentes organismos estão coloridos em cinza.....	163
Figura 5.11	Estrutura cristalográfica do inibidor NMDBA (verde) ligado no sítio ativo da GAPDH de <i>L. mexicana</i> (PDB 1I32) comparada com as conformações melhores pontuadas do próprio NMDBA (cinza) geradas pelos programas FLEXX (A) e GOLD (B) durante a etapa de validação do procedimento. Alinhamento molecular do conjunto de dados obtidos com o FLEXX (C) e GOLD (D) ilustradas no interior do sítio ativo da GAPDH de <i>L. mexicana</i>	165

Figura 5.12	Gráfico dos valores experimentais <i>versus</i> preditos de pIC_{50} obtido com os modelos de QSAR 3D para o conjunto treinamento (quadrados) e conjunto teste (triângulos) de inibidores da GAPDH de tripanosomatídeo.....	170
Figura 5.13	Mapas de contorno CoMFA obtidos com o modelo CoMFA _{FLEX} . Os inibidores 36 (azul claro) e 23 (laranja) estão representados como modelos de bastões e esferas no interior do sítio ativo da GAPDH de <i>L. mexicana</i> . O sítio ativo esta representado contornado pela superfície de Connolly. (A) Mapa de contorno estéreo está indicado em amarelo e verde (-0,005 e 0,035 Kcal/mol, respectivamente). (B) O mapa de contorno eletrostático está indicado em vermelho e azul (-0,05 e 0,05 Kcal/mol, respectivamente).....	172
Figura 5.14	Mapas de contorno CoMSIA obtidos com o modelo CoMSIA _{FLEX} . Os inibidores 36 (azul claro) e 23 (laranja) estão representados como modelos de bastões e esferas no interior do sítio ativo da GAPDH de <i>L. mexicana</i> . O sítio ativo esta representado contornado pela superfície de Connolly. Mapa de contorno estéreo está indicado em amarelo e verde (-0,0014 e 0,008 Kcal/mol, respectivamente). (B) Mapa de contorno eletrostáticos está indicado em vermelho e azul (-0,03 e 0,02 Kcal/mol, respectivamente). (C) Mapa de contorno hidrofóbico está indicado em branco e amarelo (-0,01 e 0,015 Kcal/mol, respectivamente). (D) Mapa de contorno doador de ligação de hidrogênio está indicado em roxo e ciano (-0,015 e 0,02 Kcal/mol, respectivamente) e o mapa de contorno aceptor de ligação de hidrogênio está indicado em vermelho e magenta (-0,07 e 0,015 Kcal/mol, respectivamente)....	175

Figura 5.15	Sobreposição estrutural dos Ca das GAPDHs de <i>L. mexicana</i> (ciano, PDB 1I32), <i>T. brucei</i> (amarelo, PDB 1GGA), <i>T. cruzi</i> (cinza, PDB 1ML3) e humana (magenta, PDB 1U8F) apresentados simultaneamente com os mapas de contorno estéreo (verde, favorável para grupos volumosos, 0,035 Kcal/mol) e eletrostático (vermelho, favorável para grupos eletronegativos, -0,05 Kcal/mol) derivados do modelo CoMFA _{FLEX} . As moléculas de NAD ⁺ e do inibidor 36 estão apresentadas como modelos de bastões e esferas e coloridas de acordo com a estrutura que estão ligadas.....	177
Figura 5.16	Sítio de ligação do substrato da enzima GAPDH de <i>T. cruzi</i> em complexo com um análogo reversível do substrato (S70) (PDB 1QXS). Os átomos de oxigênio do resíduo Asp210 (Leu194 na GAPDH humana) estão coloridos em vermelho e em destaque a cavidade próxima ao sítio ativo que pode ser explorada para a descoberta de novos inibidores.....	181
Figura 5.17	Esquema geral da análise de <i>Hot Spots</i>	185

Figura 5.18 Análise de *Hot Spots* do sítio de ligação do substrato da GAPDH de *T. cruzi* calculado pelo DrugScore (A, B e C) e pelo GRID (C, D e E). (A) Mapa de contorno estéreo (átomo de prova C.3 contornado com um nível de 83% em relação ao mínimo global). (B) Mapa de contorno para potenciais grupos doadores de ligação de hidrogênio na molécula do ligante (átomo de prova N.am, nível de contorno 80%). (C) Mapa de contorno para potenciais grupos aceptores de ligação de hidrogênio na molécula do ligante (átomo de prova O.2, nível de contorno 78%). (D) Mapa de contorno hidrofóbico (átomo de prova DRY, nível de contorno -0,55 Kcal/mol). (E) Mapa de contorno para potenciais grupos doadores de ligação de hidrogênio na molécula do ligante (átomo de prova NH de amida, nível de contorno -9,0 Kcal/mol). (F) Mapa de contorno para potenciais grupos aceptores de ligação de hidrogênio na molécula do ligante (átomo de prova O de carbonila, nível de contorno -7,5 Kcal/mol)..... 186

Figura 5.19 Modelo farmacofórico baseado na estrutura da proteína. 1) Um dos grupos aniônicos essenciais (ciano) devem ser posicionado no interior das esferas magenta. 2) Região no espaço onde um grupo doador de ligação de hidrogênio na molécula do ligante é capaz de interagir com O da cadeia lateral do Asp210. 3) Região no espaço onde um grupo aceptor de ligação de hidrogênio na molécula do ligante é capaz de interagir com o NH da cadeia principal da Lys209..... 188

Figura 5.20	Exemplos de modelos de interação para as soluções encontradas durante o estudo de VS. Os modos de ligação propostos satisfazem as restrições impostas pelo modelo farmacofórico incorporado durante o processo de docagem molecular dos compostos da base de dados.....	190
Figura 5.21	Modo de ligação de um dos compostos da base de dados. A molécula do ligante está representada em ciano, os resíduos de aminoácidos da proteína estão representados em verde e as ligações de hidrogênio estão ilustradas em amarelo. O sítio de ligação está representado contornado pela superfície de Connolly.	194
Figura 5.22	Modelo esquemático das etapas realizadas no estudo de VS com a GAPDH de <i>T. cruzi</i>	196
Figura 7.1	Inibidores da GAPDH de <i>T. cruzi</i> utilizados nos ensaios de co-cristalização.....	222
Figura 7.2	Mapa de densidade eletrônica 2Fobs-Fcalc referente ao sítio ativo da GAPDH ilustrando o resíduo cisteína covalentemente ligada ao inibidor IAA (CCS166) (contorno de 1.0 σ). Em destaque os principais resíduos, molécula de água e NAD ⁺ que interagem com o ligante. As ligações de hidrogênio estão indicadas como linhas tracejadas amarelas.....	223
Figura 7.3	Mecanismo de inativação da GAPDH de <i>T. cruzi</i> pelo inibidor IAA.....	224

LISTA DE TABELAS

Tabela 1.1	Fármacos para doenças negligenciadas introduzidos no mercado através de parcerias público-privadas.....	44
Tabela 1.2	Porcentagem de inibição das enzimas do glicossoma necessária para redução de 50% do fluxo glicolítico.....	54
Tabela 3.1	Compostos derivados de nucleosídeos testados <i>in vitro</i> contra a GAPDH....	83
Tabela 3.2	Série de bisfosfonatos como inibidores da enzima da GAPDH.....	86
Tabela 4.1	Parâmetros cinéticos calculados utilizando-se os métodos de regressão não-linear e Lineweaver-Burk.....	97
Tabela 4.2	Dados cristalográficos e estatísticas do refinamento das diferentes formas de GAPDH de <i>T. cruzi</i>	105
Tabela 4.3	Fatores de temperatura médio dos 359 resíduos das quatro formas da GAPDH de <i>T. cruzi</i>	109
Tabela 4.4	Comparação entre as subunidades do tetrâmero das diferentes formas de GAPDH de <i>T. cruzi</i> (valores de rmsd em Å). Os resíduos 3-356 foram utilizados no protocolo de sobreposição.....	111
Tabela 4.5	Comparação entre as subunidades das diferentes formas de GAPDH de <i>T. cruzi</i> (valores de rmsd em Å). A Tabela está colorida de acordo com o grau de similaridade, mostrando as subunidades similares em verde (< 0,61 Å) e as diferentes em vermelho (≥ 0,61 Å). Os resíduos 3-356 foram utilizados no protocolo de sobreposição.....	112
Tabela 5.1	Estruturas químicas e os valores associados de pIC ₅₀ do conjunto de dados de inibidores da GAPDH de <i>L. mexicana</i>	136

Tabela 5.2	Parâmetros de distinção de fragmentos no método HQSAR.....	147
Tabela 5.3	Parâmetros estatísticos das análises de HQSAR para várias combinações de distinção do fragmento utilizando o tamanho do fragmento padrão (4-7).....	150
Tabela 5.4	Influência do tamanho do fragmento nos parâmetros estatísticos dos de HQSAR.....	151
Tabela 5.5	Valor médio de q^2 e SDEP obtidos após 25 análises independentes do protocolo de validação cruzada LMO.....	152
Tabela 5.6	Valores experimentais, preditos e residuais de pIC_{50} calculados com o melhor modelo de HQSAR para o conjunto teste de 14 compostos inibidores da GAPDH de <i>L. mexicana</i>	153
Tabela 5.7	Parâmetros estatísticos das análises CoMFA e CoMSIA conduzidas nos alinhamentos FlexX e GOLD.....	167
Tabela 5.8	Parâmetros estatísticos obtidos durante os procedimentos de validação dos modelos CoMFA e CoMSIA.....	168
Tabela 5.9	Valores experimentais, preditos e residuais de pIC_{50} calculados com os modelos de QSAR 3D para o conjunto teste de 14 compostos inibidores da GAPDH de tripanosomatídeo.....	169
Tabela 5.10	Grupos funcionais freqüentemente encontrados interagindo com grupo guanidina de argininas em complexos cristalográficos.....	183
Tabela 5.11	Sub-conjunto de 10 moléculas representantes do compostos selecionados pelo VS para atividade biológica.....	197
Tabela 7.1	Parâmetros cristalográficos e estatísticos dos conjuntos dados coletados no LNLS.....	222

LISTA DE ABREVIATURAS E SIGLAS

2D	Bidimensional
3D	Tridimensional
A	Átomo
aa	Aminoácido
AChE	Enzima Acetilcolinesterase
ADME	Absorção, Distribuição, Metabolismo e Excreção
AP	Agente Precipitante
B	Ligação
BuChE	Enzima Butirilcolinesterase
C	Conectividade
CBME	Centro de Biotecnologia Molecular Estrutural
CETP	Proteína Colesteril Éster Transferase (<i>Cholesteryl Ester Transfer Protein</i>)
CG	Carbono Gama
Ch	Quiralidade
CoMFA	Análise Comparativa dos Campos Moleculares (<i>Comparative Molecular Field Analysis</i>)
CoMSIA	Análise Comparativa dos Índices de Similaridade Moleculares (<i>Comparative Molecular Similarity Indices Analysis</i>)
CP	Cruzaína
CRC	Algoritmo Cíclico de Verificação de Redundância (<i>Cyclic Reduncy Check</i>)
CSD	Banco de Dados Cristalográfico de Cambrigde (<i>Cambrigde Structural Database</i>)

DA	Doador e Aceptor
DNA	Ácido Desoxirribonucléico (<i>Deoxyribonucleic acid</i>)
E	Enzima Livre
EDTA	Ácido etilenodiaminotetracético
ES	Complexo Enzima-Substrato
FDA	Administração Federal de Alimentos e Medicamentos dos Estados Unidos (<i>Food and Drug Administration</i>)
FFCL-RP	Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
G3P	Gliceraldeído-3-Fosfato (<i>Glyceraldehyde 3-phosphate</i>)
GAPDH	Gliceraldeído-3-Fosfato Desidrogenase (<i>Glyceraldehyde 3-Phosphate Dehydrogenase</i>)
HDL	Lipoproteínas de Alta Densidade (<i>High-density lipoproteins</i>)
HEPES	4-(2-Hidroxietil) Piperazina ácido Etanosulfônico (<i>4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic acid</i>)
HQSAR	Holograma QSAR
HTS	Triagem em Larga Escala (<i>High-Throughput Screening</i>)
IAA	Ácido Iodo Acético (<i>Iodo Acetic Acid</i>)
IC ₅₀	Concentração Inibitória 50%
IFSC	Instituto de Física de São Carlos
IND	Investigação de Novo Fármaco (<i>Investigational New Drug</i>)
IQ	Instituto de Química
IUPAC	União Internacional de Química Pura e Aplicada (<i>International Union of Pure and Applied Chemistry</i>)

K_m	Constante de Michaelis-Menten
KNF	Modelo Koshland-Némethy-Filmer
LBDD	Planejamento Baseado na Estrutura do Ligante (<i>Ligand-Based Drug Design</i>)
LC	Leishmaniose Cutânea
LCD	Leishmaniose Cutânea Difusa
LDL	Lipoproteínas de Baixa Densidade (<i>Low-Density Lipoprotein</i>)
LMC	Leishmaniose Muco-Cutânea
LMO	Método Deixe Vários de Fora (<i>Leave-Many-Out method</i>)
LNLS	Laboratório Nacional de Luz Síncrotron
LOO	Método Deixe Um de Fora (<i>Leave-One-Out method</i>)
LQMC	Laboratório de Química Medicinal e Computacional
LTA	Leishmaniose Tegumentar Americana
LV	Leishmaniose Visceral
M	Macromolécula
MIF	Campo de Interação Molecular (<i>Molecular Interaction Field</i>)
MWC	Modelo Monod-Wyman-Changeux
MX-1	Linha de Cristalografia de Macromoléculas 1
NAD ⁺	Nicotinamida Adenina Dinucleotídeo - forma oxidada (<i>Nicotinamide Adenine Dinucleotide – oxidized form</i>)
NADH	Nicotinamida Adenina Dinucleotídeo - forma reduzida (<i>Nicotinamide Adenine Dinucleotide - reduced form</i>)
NCE	Nova Entidade Química (<i>New Chemical Entity</i>)
NCS	Simetria Não-Cristalográfica (<i>Non-Crystallographic Symmetry</i>)

NDA	Requisição de um Novo Fármaco (<i>New Drug Application</i>)
n_H	Constante de Hill
NIH	Instituto Nacional de Saúde dos Estados Unidos (<i>National Institute of Health</i>)
OMS	Organização Mundial da Saúde
P	Produto
P&D	Pesquisa e Desenvolvimento
PDB	Banco de Dados de Proteínas (<i>Protein Data Bank</i>)
pIC_{50}	Logaritmo da concentração inibitória 50%
PLS	Regressão Multivariada por Mínimos Quadrados Parciais (<i>Partial Least Square method</i>)
q^2	Coeficiente de correlação da validação cruzada
QSAR	Relações Quantitativas Estrutura-Atividade (<i>Quantitative Structure-Activity Relationships</i>)
r^2	Coeficiente de Correlação dos Dados
RD17	Cumarina Sintética (controle positivo dos ensaios cinéticos)
RMN	Ressonância Magnética Nuclear
RMSD	Raiz Quadrada do Desvio Médio (<i>Root Mean Square Deviation</i>)
RNA	Ácido Ribonucléico (<i>Ribonucleic Acid</i>)
SAR	Relações Estrutura-Atividade (<i>Structure-Activity Relationships</i>)
SBDD	Planejamento Baseado na Estrutura do Receptor (<i>Structure-Based Drug Design</i>)
SDEP	Desvio Padrão do Erro de Predição (<i>Standard Deviation Error of Prediction</i>)
SEE	Erro Padrão de Estimativa (<i>Standard Error of the Estimate</i>)

SLN	Linha de Notação do Sybyl (<i>Sybyl Line Notation</i>)
TEA	Trietanolamina
TR	Tripanotiona Redutase
UFBA	Universidade Federal da Bahia
UFSC	Universidade Federal de Santa Catarina
UFSCar	Universidade Federal de São Carlos
USP	Universidade de São Paulo
UV	Ultra-Violeta
V_0	Velocidade Inicial
$V_{\max}$	Velocidade Máxima
VS	Triagem Virtual

SUMÁRIO

CAPÍTULO 1	32
1 - INTRODUÇÃO	33
1.1 - Química medicinal e a descoberta de novos fármacos	33
1.1.2 - Fármacos e indústria farmacêutica	34
1.1.3 - Estratégias para o planejamento de candidatos a fármacos	42
1.2 - Doenças tropicais	42
1.2.1 - Doença de Chagas	44
1.2.2 - Leishmaniose	47
1.3 - Enzimas-alvo e quimioterapia	51
1.3.1 - Alvos terapêuticos para a Doença de Chagas	52
1.3.2 - A via glicolítica como alvo para o planejamento de fármacos antiparasitários	53
1.3.3 - Gliceraldeído-3-fosfato desidrogenase (GAPDH)	55
1.4 - Referências	57
CAPÍTULO 2	72
2 - OBJETIVOS	73
CAPÍTULO 3	74
3 - IDENTIFICAÇÃO DE NOVOS INIBIDORES DA ENZIMA GAPDH DE <i>T. cruzi</i>	75
3.1 - Inibidores enzimáticos	75
3.2 - Expressão e purificação da enzima GAPDH de <i>T. cruzi</i>	77
3.3 - Ensaios bioquímicos com a enzima GAPDH	78
3.4 - Inibidores da GAPDH de <i>T. cruzi</i>	80

3.4.1 - Produtos naturais como inibidores.....	80
3.4.2 – Compostos sintéticos como inibidores.....	82
<i>Derivados de nucleosídeos.....</i>	82
<i>Derivados bisfosfonatos.....</i>	84
3.5 – Referências.....	88
CAPÍTULO 4.....	90
4 - ESTUDOS CINÉTICOS E CRISTALOGRÁFICOS DA GAPDH DE <i>T. cruzi</i>	91
4.1 - Alosterismo e mudanças conformacionais.....	91
4.2 - Análise quantitativa da cooperatividade entre sítios de ligação.....	93
4.3 - A enzima GAPDH e a fenômeno de cooperatividade entre os sítios de ligação.....	94
4.4 - Estudos cinéticos da GAPDH de <i>T. cruzi</i>	95
4.4.1 - Determinação das constantes cinéticas da GAPDH de <i>T. cruzi</i>	95
4.4.2 - Determinação do comportamento cooperativo da GAPDH de <i>T. cruzi</i>	97
4.5 - Experimentos de cristalização e difração de raios X com a GAPDH de <i>T. cruzi</i>	100
4.5.1 - Cristalização.....	100
4.5.2 - Purificação e cristalização.....	102
4.5.3 - Coleta e processamento dos dados de difração de raios X.....	103
4.5.4 - Qualidade das estruturas cristalográficas.....	108
4.5.5 - Análise das estruturas de GAPDH de <i>T. cruzi</i>	110
<i>Comparação das formas apo e holo.....</i>	111
<i>Formas 1NAD e 2NAD.....</i>	114
<i>O mecanismo de cooperatividade e os contatos entre subunidades.....</i>	116
4.6 – Referências.....	121

CAPÍTULO 5.....	126
5 - ESTUDOS DE MODELAGEM MOLECULAR COM AS GAPDHs DE TRIPANOSOMATÍDEOS.....	127
5.1 - Química medicinal e as técnicas de modelagem molecular.....	127
5.1.1 - Planejamento baseado na estrutura do receptor.....	128
5.1.2 - Planejamento baseado na estrutura do ligante.....	131
5.2 - Relações quantitativas entre a estrutura-atividade.....	133
5.2.1 - Estudos de QSAR para os inibidores da GAPDH de <i>L. mexicana</i>	134
<i>Métodos computacionais</i>	134
<i>Conjunto de dados</i>	134
<i>Validação dos modelos de QSAR</i>	144
5.2.2 - Holograma QSAR (HQSAR).....	144
5.2.3 - Estudos de HQSAR para os inibidores de GAPDH de <i>L. mexicana</i>	149
<i>Análise de HQSAR</i>	149
5.2.4 - QSAR 3D.....	156
<i>Análise comparativa dos campos moleculares (CoMFA)</i>	156
<i>Análise comparativa dos índices de similaridades moleculares (CoMSIA)</i>	159
5.2.5 - Estudos de QSAR 3D (CoMFA e CoMSIA) com as GAPDHs de tripanosomatídeos...	162
<i>Estratégias de alinhamento molecular</i>	162
<i>Modelos CoMFA e CoMSIA</i>	166
<i>Análise dos mapas de contorno CoMFA</i>	171
<i>Análise dos mapas de contorno CoMSIA</i>	173
<i>Mapas de contorno de seletividade</i>	176

5.3 - Triagem virtual de compostos.....	178
5.3.1 - Triagem virtual de inibidores da GAPDH de <i>T. cruzi</i>	180
<i>Sítio de ligação</i>	180
<i>Base de dados de compostos</i>	182
<i>Filtragem preliminar por grupos funcionais</i>	182
<i>Modelo farmacofórico baseado na estrutura do alvo molecular</i>	184
<i>Incorporação do modelo farmacofórico no procedimento de docagem molecular</i>	188
<i>Pontuação por consenso</i>	191
<i>Análise visual dos resultados</i>	193
<i>Seleção dos compostos</i>	195
5.4 - Referências.....	198
CAPÍTULO 6	208
6 - COLABORAÇÕES.....	209
6.1 - Artigo de revisão.....	209
6.2 - Desenvolvimento de modelos de QSAR.....	210
6.2.1 - Estudos de QSAR 2D para uma série de inibidores da enzima butirilcolinesterase (BuChe).....	210
6.2.2 - Estudos de QSAR 2D para uma série de inibidores da proteína colesterol éster transferase (CETP)	211
6.2.3 - Estudos de QSAR para uma série de inibidores da enzima cruzaina de <i>T. cruzi</i>	212
6.3 - Desenvolvimento de biorreator enzimático utilizando a GAPDH de <i>T. cruzi</i>	212
6.4 - Capítulos de livros.....	214
6.4.1 - Construção de modelo farmacofórico baseado na estrutura do alvo receptor.....	214

6.5 - Referências.....	215
CAPÍTULO 7.....	217
7 - CONCLUSÕES E PERSPECTIVAS.....	218
7.1 - Conclusões.....	218
7.2 - Perspectivas.....	221
<i>Complexo cristalográfico GAPDH-IAA.....</i>	<i>221</i>
<i>Mecanismo de cooperatividade da GAPDH de T. cruzi.....</i>	<i>225</i>
<i>Triagem virtual de compostos.....</i>	<i>225</i>
<i>Artigos em preparação.....</i>	<i>226</i>
7.3 - Referências.....	228
APÊNDICES.....	229

Capítulo 1

Introdução

Nesse capítulo é realizada uma descrição breve da química medicinal e do seu papel no planejamento de novos fármacos, bem como de algumas etapas e técnicas envolvidas neste complexo paradigma. Uma revisão da literatura sobre as doenças negligenciadas também é apresentada, com atenção especial para a Doença de Chagas e a leishmaniose. Neste contexto, são discutidos dados epidemiológicos e o impacto médico-social destas parasitoses, além das formas atuais de prevenção e tratamento. A via glicolítica dos tripanosomatídeos é explorada, com destaque para o papel da enzima gliceraldeído-3-fosfato desidrogenase (GAPDH) como alvo biológico atrativo para o desenvolvimento de inibidores candidatos a novos agentes quimioterápicos contra as tripanosomíases.

1 - INTRODUÇÃO

1.1 - Química medicinal e a descoberta de novos fármacos

O processo de descoberta e desenvolvimento de novos fármacos é complexo, de elevado custo e tempo e fundamenta-se na integração de várias áreas estratégicas, tais como: inovação, conhecimento, tecnologia, gerenciamento e investimentos em Pesquisa e Desenvolvimento (P&D).¹ Os notáveis avanços da química e biologia, aliados a melhor compreensão de mecanismos fisiopatológicos e farmacológicos, tornaram possível o planejamento e a descoberta de fármacos capazes de representar inovações terapêuticas importantes para diversas doenças e disfunções orgânicas em humanos. Os avanços científicos e tecnológicos na área de P&D de fármacos levaram, de forma inegável, a melhorias significativas na qualidade de vida das diversas populações no mundo.²⁻⁴

Na estratégia de planejamento racional de fármacos, os estudos dos processos evolutivos de reconhecimento molecular em sistemas biológicos assumem grande importância, pois constituem as bases fundamentais para a potência, afinidade e seletividade dos fármacos por seus receptores-alvo. A química medicinal desempenha função central neste complexo paradigma de descoberta e desenvolvimento de novas moléculas com atividade biológica.⁵⁻⁷

A química medicinal possui forte caráter multidisciplinar, abrangendo diversas especialidades importantes, como a química orgânica, bioquímica, farmacologia, informática, biologia molecular e estrutural, entre outras. Segundo a União Internacional de Química Pura e Aplicada (IUPAC, do inglês *International Union of Pure and Applied Chemistry*), a química medicinal envolve a descoberta, desenvolvimento, identificação e interpretação do mecanismo de ação molecular de compostos biologicamente ativos. Além da descoberta de moléculas bioativas,

a química medicinal também estuda os fenômenos envolvidos no metabolismo e as relações entre a estrutura química e atividade.⁸ A química medicinal é uma ciência interdisciplinar fundamentada em química e que envolve aspectos importantes das ciências farmacêuticas, médicas, biológicas, físicas e computacionais.

1.1.2 - Fármacos e indústria farmacêutica

O setor de P&D de fármacos é extremamente competitivo e se caracteriza por elevados níveis de investimentos, que estão associados a riscos de proporcional magnitude. A idéia de que as maiores companhias farmacêuticas são as melhores em P&D tem se tornado um dogma mundial, já que os fármacos da categoria *blockbuster* (termo do inglês para designar os fármacos com vendas anuais superiores a US\$ 1 bilhão) são os principais responsáveis pelo crescimento da indústria farmacêutica. A capacidade da indústria em destinar recursos para P&D está diretamente relacionada à sua competência em gerar lucros através da venda de um conjunto atrativo de fármacos.⁹⁻¹¹

O complexo processo de descoberta e desenvolvimento de novas entidades químicas (NCEs, do inglês, *new chemical entities*) é longo e envolve altos investimentos.¹² Desde a concepção do projeto até a introdução de um único fármaco no mercado farmacêutico, são investidos de 12 a 15 anos em P&D, com custos totais estimados, no biênio 2004-2006, da ordem de US\$ 500-880 milhões, podendo em alguns casos alcançar cifras superiores a US\$ 1 bilhão (Figura 1.1).^{7,13-15}

Figura 1.1 - Visão geral do processo de P&D de um novo fármaco.¹⁴

Inovações terapêuticas de grande sucesso, entretanto, são capazes de gerar rendimentos significativos da ordem de US\$ bilhões anuais. Dentre os fármacos incluídos atualmente na categoria *blockbuster*, estão o hipolipemiante atorvastatina (Lípitor[®], Pfizer), que gerou extraordinários US\$ 12,7 bilhões com vendas somente no ano de 2007, o antitrombótico clopidogrel (Plavix[®], Sanofi Aventis), o antiulceroso esomeprazol (Nexium[®], AstraZeneca), o antipsicótico olanzapina (Zyprexa[®], Eli Lilly), entre outros.

O processo de descoberta e desenvolvimento de novos fármacos é dividido em duas grandes fases: (i) pré-clínica (descoberta ou pesquisa básica), e (ii) clínica (desenvolvimento), de acordo com a representação ilustrativa da Figura 1.2.^{7,16} Nos estágios iniciais da fase pré-clínica, as pesquisas se concentram na identificação e otimização de moléculas pequenas capazes de modular a atividade do alvo macromolecular eleito para o processo de planejamento (Figura 1.3). A validação do alvo molecular selecionado é fundamental, pois estabelece a sua relevância no processo fisiopatológico em estudo, além de evidenciar se a sua modulação seletiva é capaz de gerar a resposta farmacológica esperada no controle da doença ou desordem em humanos.^{8,17,18}

Figura 1.2 – Esquema da evolução das fases pré-clínica e clínica no processo de desenvolvimento e descoberta de novos fármacos.^{7,19}

As moléculas bioativas ou ligantes (do inglês, *hits*) tem sua origem a partir de produtos naturais ou através de síntese orgânica e coleções combinatórias (Figura 1.3). As moléculas bioativas podem ser identificadas através de triagens reais ou virtuais, ou ainda através de planejamento racional, mas em todos os casos as suas propriedades biológicas devem ser determinadas experimentalmente. Nesta fase inicial geralmente são identificadas moléculas com baixa afinidade, que necessitam ser otimizadas em relação a uma série de propriedades (e.g., potência, afinidade, seletividade, biodisponibilidade, toxidez). Os compostos bioativos com melhores propriedades são selecionados como compostos líderes (do inglês, *lead compounds*) para posterior otimização (Figura 1.3).^{17,20,21}

Com o auxílio de métodos em química medicinal é possível explorar o imenso espaço químico delineando o trabalho investigativo na identificação, seleção e otimização de moléculas capazes de interagir com alta afinidade e seletividade com o alvo molecular selecionado (e.g., enzima, receptor), o qual representa o espaço biológico do binômio químico-biológico em

questão. Várias estratégias em química medicinal podem ser empregadas para a investigação do espaço químico-biológico, como a organização de bases padrões de dados, a aplicação de filtros moleculares, o emprego de ensaios biológicos automatizados em larga escala (HTS, do inglês, *high-throughput screening*) e do ensaio virtual (VS, do inglês, *virtual screening*), além das técnicas de planejamento baseado na estrutura do receptor (SBDD, do inglês, *structure-based drug design*) e de planejamento baseado na estrutura do ligante (LBDD, do inglês, *ligand-based drug design*), e do estudo das relações quantitativas entre a estrutura e atividade (QSAR, do inglês, *quantitative structure-activity relationships*). Em todos estes casos, o gerenciamento qualificado da informação possibilita a geração de conhecimento apropriado a partir dos dados disponíveis.^{22,23}

Figura 1.3 - Etapas evolutivas de identificação, seleção e otimização de moléculas no processo de planejamento de fármacos (QSAR, estudo das relações quantitativas entre a estrutura e atividade; ADME/Tox, propriedades farmacocinéticas/tóxicas: absorção, distribuição, metabolismo e excreção).²⁴

Uma das etapas-chave da Figura 1.3 é a otimização de compostos líderes em relação às suas propriedades farmacodinâmicas e farmacocinéticas, com o objetivo da descoberta de NCEs com elevado potencial de desenvolvimento clínico. Variações estruturais são utilizadas para

adequar o perfil das moléculas às características necessárias ao uso terapêutico em humanos. Neste sentido, a química medicinal oferece várias estratégias capazes de gerar coleções dirigidas de compostos.^{25,26} Os estudos das relações entre a estrutura e atividade (SAR, do inglês, *structure-activity relationships*) e de QSAR são importantes para guiar a síntese de novas moléculas com propriedades otimizadas, minimizando o universo de compostos a serem considerados nos programas de triagem biológica. Idealmente, propriedades farmacodinâmicas como potência, afinidade e seletividade, além de propriedades farmacocinéticas de ADME (Absorção, Distribuição, Metabolismo e Excreção) devem ser consideradas conjuntamente, facilitando a eliminação de candidatos com propriedades inadequadas e reduzindo os custos do processo de P&D.^{27,28} Na fase pré-clínica de pesquisa básica é realizada uma série de testes *in vitro* e *in vivo* para avaliar o potencial dos candidatos a NCEs em relação a sua segurança e eficácia, visando a sua introdução em fases clínicas para testes em humanos.

Os métodos em quimio- e bioinformática estão integrados ao planejamento de fármacos. O seu emprego nos estudos em química medicinal vão desde a identificação, seleção e otimização de moléculas candidatas até a proposição de novas NCEs com elevado potencial terapêutico. A integração desses métodos ao trabalho químico e biológico experimental é um requerimento essencial para a geração de novas moléculas bioativas qualificadas, com propriedades múltiplas otimizadas.^{5,23,28-30}

Após a conclusão da fase pré-clínica, as NCEs selecionadas são submetidas à aprovação das agências regulatórias, como a Administração Federal de Alimentos e Medicamentos dos Estados Unidos (FDA, do inglês, *Food and Drug Administration*) para autorização da realização de testes em humanos através da licença para investigação de um novo candidato a fármaco (IND, do inglês, *investigational new drug*) (Figura 1.1). No Brasil, a responsável pelo controle das pesquisas clínicas e registro de fármacos é a Agência Nacional de Vigilância Sanitária

(ANVISA). Nas fases clínicas são realizados diversos estudos para avaliar a eficácia e a segurança das NCEs candidatas a fármacos (Figura 1.2).

De forma resumida, na fase clínica I são realizados testes de dosagem dos princípios ativos em um grupo de voluntários saudáveis (20-100 indivíduos) para avaliação da toxidez. Na fase clínica II iniciam os testes em pacientes que apresentam a doença ou desordem (100-500 indivíduos) para a avaliação da dosagem, eficácia e segurança da composição. Na fase clínica III são realizados testes farmacológicos em um número significativo de pacientes (1.000-5.000 indivíduos) distribuídos em regiões geográficas representativas, para uma melhor avaliação das propriedades farmacocinéticas, além de eficácia e segurança. Ao término da fase III, todos os dados clínicos obtidos são organizados e submetidos à agência regulatória (FDA) para solicitação da aprovação de um novo fármaco (NDA, do inglês, *new drug application*). Após a aprovação da NDA, o fármaco poderá ser lançado no mercado farmacêutico. Na fase clínica IV, é realizado o monitoramento pós-comercialização a longo prazo do novo fármaco.^{13,16}

Apesar da notável evolução das ciências químicas e biológicas, dos avanços significativos em P&D de fármacos e dos altos níveis de investimentos envolvidos no processo de descoberta, o número de NCEs aprovadas tem decrescido (Figura 1.4).³¹⁻³⁴

Esse panorama, caracterizado pelo declínio da produtividade e aumento da competitividade, aponta para mudanças significativas do modelo tradicional de pesquisa baseado principalmente em química e biologia básica e na abordagem “triagem-erro”, para um novo paradigma de planejamento racional “baseado no alvo terapêutico”. O desenvolvimento de inovações terapêuticas requer um forte comprometimento com a inovação científica e tecnológica através de programas bem estabelecidos de P&D de fármacos. De fato, a indústria farmacêutica vem passando por um momento crucial marcado por grandes transformações guiadas pelo uso

racional de inovações científicas e tecnológicas como ferramenta de pesquisa multidisciplinar avançada.³³⁻³⁵

Figura 1.4 - Número de NCEs aprovadas pelo FDA para uso em humanos no período entre 1996 e 2006.³³

Dentre as razões que contribuem para a diminuição do número de novos fármacos introduzidos no mercado, podem ser citadas cinco como as principais: (i) mecanismo: problemas no desenvolvimento de modelos animais e clínicos que descrevam apropriadamente o mecanismo fisiopatológico de doenças-alvo; (ii) composto líder: impossibilidade de otimização de múltiplas propriedades do composto líder; (iii) segurança do fármaco: os candidatos a fármaco apresentam propriedades tóxicas que não foram observadas previamente na fase pré-clínica; (iv) propriedades farmacocinéticas: identificação de propriedades de ADME inadequadas para uso clínico do candidato a fármaco, que não foram detectadas na fase pré-clínica; (v) eficácia: perda ou redução significativa da eficácia de NCEs nas triagens clínicas, devido a diversos fatores.³³

Embora as técnicas modernas envolvidas no novo paradigma de planejamento de fármacos ainda não foram capazes de satisfazer as rigorosas expectativas de produtividade da

indústria farmacêutica, alguns exemplos importantes ilustram o sucesso da aplicação da estratégia baseada no alvo terapêutico.³⁶⁻³⁹ Vale salientar que alguns dos métodos modernos de planejamento racional que foram introduzidos a partir do início da década de 90 estão em contínuo desenvolvimento devido aos marcantes avanços científicos e tecnológicos, de forma que o seu real potencial ainda não pode ser completamente avaliado.³³ Entretanto, uma análise comparativa no período de 1945 a 2004 revelou um crescimento constante do número de NCEs aprovadas em relação as últimas 4 décadas (Figura 1.5). Em particular, esse número é bastante expressivo na última década (1995-2004), período que coincide com o surgimento do emprego das técnicas modernas de planejamento racional (Figura 1.5).

Figura 1.5 - Número de NCEs aprovadas para uso clínico no período de 1945 a 2004.³⁵

1.1.3 - Estratégias de planejamento de fármacos

Estratégias de planejamento de candidatos a novos fármacos fundamentam-se no conhecimento prévio do processo fisiopatológico da doença e na seleção do alvo molecular adequado. O alvo selecionado (e.g., proteína, DNA ou RNA) pode ter a sua estrutura tridimensional conhecida ou não, fato que determina quais as estratégias que serão utilizadas no processo de planejamento.^{17,40} Os grandes avanços da genômica e proteômica, bem como a evolução de técnicas como a cristalografia de raios X e ressonância magnética nuclear (RMN), proporcionam um aumento significativo no número de alvos moleculares que possuem suas estruturas tridimensionais disponíveis no Banco de Dados de Proteínas (PDB, do inglês, *Protein Data Bank*).^{41,42} O conhecimento da estrutura do alvo macromolecular, seja este uma enzima ou receptor biológico, ou da estrutura de algum complexo ligante-receptor, permite o planejamento racional de inibidores enzimáticos, ou de agonistas ou antagonistas de receptores através de estratégia de SBDD.^{1,2,5}

Em contraste, quando a estrutura do alvo macromolecular eleito não é conhecida, aplicam-se técnicas de LBDD envolvendo o estudo de características multifuncionais de ligantes conhecidos.^{1,43,44} Em muitos casos, o uso integrado das técnicas de SBDD e LBDD pode gerar informações e novos conhecimentos úteis no planejamento de NCEs, explorando a sinergia e complementaridade inerente entre as técnicas.^{1,45}

1.2 - Doenças tropicais

As doenças endêmicas tropicais são consequências marcantes do subdesenvolvimento social, retratando com minúcia a situação de prevalência nos países pobres.⁴⁶ Neste grupo

encontram-se as doenças infecções causadas por protozoários (e.g., doença de Chagas, leishmaniose, doença do sono, malária), helmintos (e.g., filariase, ascaridíase, oncocercose, esquistossomose), bactérias (e.g., tuberculose, lepra, tracoma) e vírus (e.g., dengue e febre amarela).

As doenças tropicais afetam milhões de indivíduos em todo o mundo, entre homens, mulheres e crianças. Para a maioria destas doenças, que causam altos índices de morbidade e mortalidade, os tratamentos disponíveis são limitados e ineficazes, e o quadro é constantemente agravado pela emergência de cepas resistentes.^{47,48} O aumento considerável nos custos de desenvolvimento de novos fármacos, combinado com as perspectivas de retorno financeiro inadequado, tem como resultado a quase inexistência de investimentos em P&D na área de doenças tropicais por parte das companhias farmacêuticas.⁴⁹ Estima-se que do total investido em P&D de fármacos, somente 10% são destinados a doenças que representam cerca de 90% dos problemas globais de saúde humana.⁵⁰ Visando reverter esse panorama, iniciativas envolvendo parcerias efetivas na tríade governo-universidade-empresa vêm sendo estabelecidas.⁴⁹

O Programa Especial para Pesquisa e Treinamento em Doenças Tropicais (do inglês, *Special Programme for Research and Training in Tropical Diseases*)⁵¹ da Organização Mundial da Saúde (OMS), o Programa de Medicamentos para Malária (MMV, do inglês, *Medicines for Malaria Venture*),⁵² e a Iniciativa de Medicamentos para Doenças Negligenciadas (DNDi, do inglês, *Drugs for Neglected Diseases initiative*),⁵³ são exemplos de parcerias público-privadas que tem como objetivo criar, estimular e investir em P&D de fármacos contra uma variedade de doenças tropicais. Alguns exemplos de fármacos introduzidos no mercado através de esforços de parcerias público-privadas estão apresentados na Tabela 1.1.

Tabela 1.1 - Fármacos para doenças tropicais introduzidos no mercado através de parcerias público-privadas.⁵¹⁻⁵³

Fármaco	Ano de Registro	Indicação Clínica	Parceiro
Praziquantel	1980	Esquistossomose	Bayer
Mefloquina	1984	Malária	Hoffman La Roche, WRAIR
Ivermectina	1987	Oncocercose	Merck
Halofantrina	1988	Malária	Smith Kline Beecham, WRAIR
Eflornitina	1991	Doença do Sono	Marion Merrel Dow
Anfotericina B Liposomal	1994	Leishmaniose	NeXstar
Artemisinina	1997	Malária	Rhone Poulenc Rorer, Kunming
Artemisinina-lumefantrina	1999	Malária	Novartis
Artemotila	2000	Malária	Artecef, WRAIR
Miltefosina	2002	Leishmaniose	Zentaris, Indian CMR
Clorproguanil-dapsona	2003	Malária	GSK, DFID

1.2.1 - Doença de Chagas

A doença de Chagas, também conhecida como tripanossomíase americana, afeta 17 milhões de indivíduos em 18 países da América Latina, sendo que mais de 100 milhões encontram-se em áreas sob o risco de contrair a doença (Figura 1.6A). Anualmente, aproximadamente 200.000 indivíduos são infectados e 13.000 vão a óbito por complicações decorrentes da parasitose.⁵⁴⁻⁵⁶ A doença de Chagas é causada pela infecção com o protozoário *Trypanosoma cruzi*, o qual é principalmente transmitido pelo inseto-vetor hematófago conhecido vulgarmente como “barbeiro” ou “chupão” - da ordem *Hemiptera*; família *Reduviidae* - que habita as frestas das paredes de barro ou de madeira das chamadas cafuas. Outras formas de transmissão desta parasitose ocorrem por transfusão de sangue e transmissão congênita.^{51,54}

A prevenção é a estratégia mais recomendada para o controle desta parasitose, ou seja, com a eliminação do inseto vetor através de melhorias nas condições de habitação e higiene das populações afetadas. Bons resultados têm sido alcançados na redução da incidência da doença por esses mecanismos de controle.⁵⁶ Entretanto, existem vários reservatórios silvestres de *T. cruzi*, que inviabilizam a sua completa eliminação.⁵⁷ Dessa forma, o controle do inseto vetor requer ações constantes de saúde, seja através de aplicações de inseticidas nas regiões endêmicas, ou pela contínua educação da população sob risco de contaminação. Na prática, essas ações não são efetivas, especialmente em regiões com pequenos níveis de transmissão vetorial, como o caso do Brasil (Figura 1.6B). Outro fator relevante é a crescente descentralização do sistema de saúde em países em desenvolvimento, o que afeta sensivelmente os programas de controle do vetor.⁵⁸

Figura 1.6 - (A) Incidência da doença de Chagas na América Latina. **(B)** Distribuição geográfica do inseto vetor causador da doença.⁵⁹

Após quase 100 anos da descoberta da doença realizada pelo brilhante cientista brasileiro Carlos Chagas, e apesar do genoma⁶⁰ e proteoma⁶¹ do *T. cruzi* serem conhecidos, não existe um único fármaco seguro e eficaz para o tratamento quimioterápico da doença de Chagas. Atualmente, o tratamento é mais sintomático do que etiológico. Do ponto de vista sintomático, procura-se amenizar as diversas manifestações clínicas da doença (e.g., administração de diuréticos no caso de insuficiência cardíaca congestiva). Do ponto de vista etiológico, os fármacos disponíveis há mais de três décadas (Figura 1.7), nifurtimox (Lampit[®], Bayer) e benznidazol (Rochagan[®], Roche), são ativos somente na fase aguda da doença, sendo que o nifurtimox não é mais comercializado no Brasil e em outros países como a Argentina, Chile e Uruguai.^{62,63} Estes fármacos, além da baixa eficácia, apresentam graves efeitos adversos provenientes de sua administração, como hiporexia, perda de peso, náuseas, vômitos, alergia cutânea e neuropatia periférica.^{64,65}

Figura 1.7 - Fármacos utilizados no tratamento da doença de Chagas.

Diante do cenário de completa ausência de inovações terapêuticas, outros fármacos têm sido avaliados no da parasitose. Entre eles, destacam-se os antifúngicos cetoconazol e terbinafina os quais apresentam atividade tripanocida. Um possível mecanismo de ação para esses compostos está relacionado à inibição da biossíntese dos esteróis constituintes da membrana plasmática do parasita.⁶⁶⁻⁶⁹

Em relação à prevenção da doença de Chagas, protocolos de imunização utilizando vacinas de DNA e de proteínas isoladas do *T. cruzi* (e.g., genes que codificam a enzima trans-sialidase e proteínas de membrana) demonstraram potencial no controle da infecção.⁷⁰ Além disso, a imunização de cachorros com a forma epimastigota de *T. rangeli* induziu a formação de anticorpos contra *T. cruzi* e mostrou-se eficaz na redução da parasitemia.⁷¹ Apesar dos enormes esforços globais e de alguns avanços promissores, não há ainda uma perspectiva concreta para o desenvolvimento de uma vacina contra a doença de Chagas.⁷⁰⁻⁷⁶

1.2.2 - Leishmaniose

As leishmanioses formam um conjunto de doenças causadas por pelo menos 24 espécies de protozoários do gênero *Leishmania*, sendo transmitidas por cerca de 30 espécies de insetos vetores pertencentes aos gêneros *Phlebotomus* e *Lutzomyia*.⁷⁷⁻⁸⁰ A maioria das espécies de *Leishmania* causa graves doenças crônicas, manifestadas por lesões cutâneas ou viscerais. Anualmente, entre 1,5 e 2 milhões de adultos e crianças desenvolvem a doença sintomática, sendo o número de óbitos estimados na ordem de 70.000.⁷⁹ Aproximadamente 90% das infecções cutâneas ocorrem no Afeganistão, Paquistão, Síria, Arábia Saudita, Argélia, Irã, Brasil e Peru, enquanto que as infecções viscerais ocorrem em sua grande maioria na Índia, Sudão, Nepal, Bangladesh e Brasil (Figura 1.8).⁸¹⁻⁸⁵

Figura 1.8 - Distribuição geográfica da leishmaniose.⁷⁸

A leishmaniose pode ser classificada em quatro formas principais:⁸¹⁻⁸⁵ (i) leishmaniose cutânea (LC) causada por *L. major*, que é a forma mais comum da doença e caracteriza-se por lesões ulcerativas da derme com feridas comumente na face, braços e pernas; (ii) leishmaniose cutânea difusa (LCD) causada por *L. mexicana* e *L. amazonensis*, que é muito semelhante a LC, exceto pelo fato de que as lesões são encontradas espalhadas por todo o corpo; (iii) leishmaniose mucocutânea (LMC) causada por *L. braziliensis*, que é caracterizada pela disseminação dos parasitas via corrente sanguínea ou linfática causando ulcerações nos tecidos cartilagosos, principalmente na boca e nariz, produzindo lesões mutilantes. Essas três manifestações são agrupadas no continente americano em uma única doença denominada leishmaniose tegumentar americana (LTA), conhecida também como úlcera de Bauru, ferida brava, entre outras denominações regionais;^{77,86} (iv) leishmaniose visceral (LV), também conhecida como calazar, é a forma mais letal da doença, sendo causada por *L. donovani*. Esta forma acomete o sistema

retículo-endotelial, causando perda de peso, anemia e inchaço do baço e do fígado podendo levar a morte num período de dois a três anos.^{77,81,86}

O tratamento da leishmaniose, desde a década de 40, baseia-se no uso de antimoniais pentavalentes.⁸⁷⁻⁸⁹ Duas formulações de antimoniais pentavalentes são disponíveis atualmente, o estibogluconato de sódio (Pentostam[®], GlaxoSmithKline) e o antimoniato-N-metil glucamina (Glucantime[®], Aventis) (Figura 1.9), sendo este último distribuído gratuitamente no Brasil pelo Ministério da Saúde. Em virtude do tempo de meia-vida pequeno dos antimoniais na corrente sangüínea (cerca de 2 horas), doses diárias altas são requeridas por longos períodos de tempo (20 a 40 dias).⁹⁰ Além disso, doses maiores dos antimoniais acabam sendo empregadas devido ao aparecimento de resistência. Embora o mecanismo de ação não seja completamente elucidado, evidências demonstraram que estes fármacos atuam nas formas amastigotas do parasita, inibindo sua atividade glicolítica e a via oxidativa de ácidos graxos. Efeitos colaterais severos como trombose, toxidez hepática, hipersensibilidade e choque anafilático são associados ao tratamento quimioterápico.⁹¹

A pentamidina (Pentacarinat[®], Rhodia Farma) (Figura 1.9), tem sido utilizada desde a década de 50 como fármaco de segunda escolha contra a leishmaniose devido ao longo período de tratamento e aos graves efeitos colaterais, tais como: hipoglicemia, diabetes, taquicardia, entre outros.⁸⁸

A anfotericina B (Fungisone[®], Bristol Meyers Squibb) (Figura 1.9) é o agente antileishmaniose mais eficaz, atuando nas formas promastigotas e amastigotas do parasita através da ligação aos ésteres (ergosterol ou episterol) presentes na membrana plasmática do parasita. A administração do medicamento é feita por via intravenosa por períodos de até 4 horas. Diversos efeitos colaterais são causados pela anfotericina B, todos relacionados com a dose terapêutica administrada.^{88,89} Novas formulações da anfotericina B associadas à lipossomas, (AmBisome[®],

NeXstar Pharmaceuticals) e a dispersão coloidal (Amphocil[®], AstraZeneca), apresentam uma menor toxidez relativa e maior tempo de meia-vida no plasma. Entretanto, o custo elevado limita o uso sistemático destes fármacos no serviço de saúde pública.^{88,89} A miltefosina (Impavido[®], Zentaris), recentemente introduzida no mercado farmacêutico, foi o primeiro fármaco administrado por via oral para o tratamento da LV e LC.⁹² O desenvolvimento da miltefosina é um exemplo de serendipidade (do inglês, *serendipity*), já que este derivado fosfolipídico foi originalmente desenvolvido para o tratamento do câncer de pele. Devido a sua indicação para o tratamento tópico de lesões ulcerativas, a miltefosina foi testada contra a LC. Testes clínicos com administração por via oral demonstraram a alta eficácia do fármaco para o tratamento da LV e LC.^{92,93} O mecanismo de ação da miltefosina ainda é desconhecido, contudo, acredita-se que o fármaco possa estar envolvido em um ou mais dos seguintes mecanismos: (i) inibição do metabolismo de fosfolídeos; (ii) interferência na transdução de sinais através da membrana do parasita; e (iii) interferência na biossíntese de glicosilfosfatidilinositol.⁹²

Figura 1.9 - Fármacos utilizados no tratamento da leishmaniose.

Em relação à prevenção da leishmaniose, os métodos mais comuns são desinsetização das áreas de risco para a erradicação dos insetos vetores, uso de mosquiteiros impregnados de inseticidas e a aplicação de repelentes sobre a pele. O sacrifício de cães infectados também faz parte do controle da doença, uma vez que esses animais representam reservatórios da doença. Além disso, surge como esperança o desenvolvimento de vacinas utilizando proteínas recombinantes, parasitas vivos atenuados, vacinas de DNA e, mais recentemente, vacinas de antígenos derivados de glicoconjugados (lipofosfoglicano).^{76,94} Atualmente, a vacina Leishvacin[®] (Biobrás), constituída de uma mistura de antígenos provenientes da *L. amazonensis*, que encontra-se em fase clínica de desenvolvimento. Apesar da Leishvacin[®] ter demonstrado potencial imunogênico em voluntários saudáveis, sua eficácia ainda não está totalmente comprovada.^{95,96}

1.3 – Enzimas-alvo e quimioterapia

As enzimas são alvos biológicos extremamente importantes para o planejamento de novos fármacos, devido principalmente ao seu papel essencial em inúmeras vias bioquímicas associadas a desordens e doenças em humanos. Outros atrativos são a sua fácil obtenção, adequação para ensaios biológicos e versatilidade na aplicação de técnicas de SBDD e LBDD.⁹⁷⁻¹⁰⁰ Os inibidores enzimáticos do tipo reversível são os candidatos a fármacos mais estudados e podem ser classificados em competitivos, não-competitivos e incompetitivos. Existem ainda os inativadores enzimáticos irreversíveis, que atuam principalmente como agentes marcadores de afinidade (do inglês, *affinity labels*) ou como inativadores enzimáticos com base no mecanismo (do inglês, *mechanism-based inactivators*), levando a modificações covalentes da enzima.¹⁰¹

De acordo com os dados do FDA de 2005, das 1.278 NCEs registradas, 317 (~25%) interagem com 71 enzimas provenientes de cinco fontes distintas (Figura 1.10).⁹⁷

Figura 1.10 - Distribuição das enzimas utilizadas como alvos moleculares de fármacos aprovados pelo FDA de acordo com a sua fonte.⁹⁷

1.3.1 - Alvos terapêuticos para a doença de Chagas

O processo de planejamento de substâncias bioativas com propriedades antiparasitárias baseia-se, principalmente, na investigação e comparação de vias bioquímicas do parasita e do hospedeiro. Essa investigação tem como objetivo identificar possíveis alvos no parasita para modulação seletiva por pequenas moléculas.¹⁰²⁻¹⁰⁴ Neste contexto, várias enzimas envolvidas em rotas distintas do metabolismo do *T. cruzi* vêm sendo investigadas como alvos moleculares para o desenvolvimento de novos agentes tripanocidas. Dentre estas, destacam-se: (i) a cruzaina (CP), uma cisteína protease lisossomal que exerce função central para a nutrição do parasita, estando também envolvida em processos de penetração nas células do hospedeiro e em fases de diferenciação celular do ciclo de vida do parasita;¹⁰⁵⁻¹⁰⁸ (ii) a tripanotiona redutase (TR) e tioredoxina redutase (TRX), enzimas com função protetora do parasita contra agentes oxidantes, através da manutenção do meio redutor intracelular;^{109,110} (iii) alvos da via glicolítica, com

destaque especial para a gliceraldeído-3-fosfato desidrogenase (GAPDH),¹¹¹ enzima que exerce papel central na glicólise, fonte vital de energia para o *T. cruzi*.¹⁰⁴

1.3.2 – A via glicolítica como alvo para o planejamento de fármacos antiparasitários

O parasita *T. cruzi* apresenta uma organela especializada característica da família *Trypanosomatidae*, chamada glicossoma, que contém nove enzimas envolvidas nas reações da via glicolítica (Figura 1.11).^{102,104,112} Esta via bioquímica é a principal fonte de energia da forma tripomastigota de *T. brucei* e *T. cruzi*.^{113,114} Isso se deve ao fato de que o parasita possui mitocôndrias inativas durante determinadas fases de seu ciclo celular, o que impede o funcionamento de rotas metabólicas alternativas mais eficientes para a produção de ATP.^{102,104} Além disso, estudos realizados com a forma amastigota do *T. cruzi* demonstraram que esta forma também é amplamente dependente dessa via para obtenção de energia.^{115,116}

As enzimas envolvidas na via glicolítica de tripanosomatídeos se apresentam como alvos importantes para o planejamento de novos fármacos através de inibição seletiva.¹⁰⁴ De fato, uma avaliação do controle do fluxo glicolítico em tripanosomatídeos identificou as enzimas cuja inibição seletiva levaria a uma redução significativa da glicólise no parasita (Tabela 1.2).^{117,118}

Figura 1.11 - Esquema do metabolismo da glicose e glicerol no glicossoma. As enzimas constituintes do glicossoma são: **1.** HK, hexoquinase; **2.** PGI, fosfoglicose isomerase; **3.** PFK, 6-fosfofruto-quinase; **4.** ALDO, frutose-difosfato aldolase; **5.** GDH, glicerol-3-fosfato desidrogenase; **6.** GK, glicerol quinase; **7.** TIM, triosefosfato isomerase; **8.** GAPDH, gliceraldeído-3-fosfato desidrogenase; **9.** PGK, 3-fosfoglicerato quinase.¹¹²

Tabela 1.2 - Porcentagem de inibição das enzimas do glicossoma necessária para redução de 50% do fluxo glicolítico.

Enzima*	Inibição necessária (%)
Transporte de glicose	51
HK	93
PFK	93
ALDO	76
GAPDH	84
PGK	85
GDH	83

*HK, hexoquinase; PFK, 6-fosfofruto-quinase; ALDO, frutose-difosfato aldolase; GAPDH, gliceraldeído-3-fosfato desidrogenase; PGK, 3-fosfoglicerato quinase. GDH, glicerol-3-fosfato desidrogenase;

De acordo com os dados da Tabela 1.2, a enzima GAPDH exerce papel significativo no fluxo glicolítico, estabelecendo-se como alvo biológico de notável valor no planejamento de candidatos a fármacos contra as tripanosomíases.¹¹⁹⁻¹³⁰ Além disso, um estudo similar de regulação da glicólise em eritrócitos humanos indicou que a inibição de até 95% das enzimas ALDO, GAPDH e PGK não levaria ao aparecimento de efeitos adversos.¹¹⁷

1.3.3 - Gliceraldeído-3-fosfato desidrogenase

A enzima GAPDH (EC 1.2.1.12) catalisa a conversão do substrato gliceraldeído-3-fosfato (GAP) em 1,3-bisfosfoglicerato (1,3-BPG) na presença do cofator NAD^+ e fosfato inorgânico (Figura 1.12A e 1.12B). A GAPDH é uma enzima tetramérica composta de subunidades idênticas (359 resíduos) e massa molecular de 156 kDa. Dependendo de sua origem, a enzima pode apresentar comportamento cooperativo ou não-cooperativo em relação a interação com o NAD^+ .¹³¹⁻¹³⁷

Figura 1.12 - (A) Estrutura quaternária da enzima GAPDH de *T. cruzi*. **(B)** Reação enzimática catalisada pela GAPDH.

Além de seu papel na via glicolítica, a enzima GAPDH desempenha funções importantes em outros processos celulares. Classificada como uma proteína multifuncional, a GAPDH de mamíferos está envolvida em diversos processos celulares, como apoptose, transporte nuclear de RNA, replicação de DNA, reparo de DNA, atividade RNase, formação de microtúbulos e fusão de membrana nuclear.^{139,140} Evidências indicam que a GAPDH pode estar envolvida também em algumas doenças neurodegenerativas, (e.g., Alzheimer, Parkinson, Huntington), câncer de próstata e distúrbios metabólicos.¹⁴¹⁻¹⁴⁵

As enzimas GAPDH de *Plasmodium falciparum*,¹⁴⁶ *L. mexicana*,¹⁴⁷ *T. cruzi*¹¹¹ e *T. brucei*,¹⁴⁸ são consideradas alvos moleculares promissores para o desenvolvimento de novos agentes antiprotozoários, uma vez que apresentam diferenças estruturais significativas quando comparadas à enzima humana (~45 % de identidade sequencial), que podem ser investigadas para a inibição seletiva da enzima dos parasitas. A enzima GAPDH do *T. cruzi*, escolhida como alvo biológico nesta tese de doutorado, apresenta como vantagens: estar clonada em *E. coli*,¹⁴⁹ possuir protocolos padronizados de expressão e purificação, além de ter algumas propriedades bioquímicas e catalíticas caracterizadas.¹⁵⁰⁻¹⁵²

1.4 - Referências

1. GUIDO R. V. C.; OLIVA, G.; ANDRICOPULO, A. D. Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, v. 15, p. 37-46, 2008.
2. DAVIS, A. M.; TEAGUE, S. J.; KLEYWEGT, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed. Engl.*, v. 42, p. 2718-2736, 2003.
3. CAPRINO, L.; RUSSO, P. Developing a paradigm of drug innovation: an evaluation algorithm. *Drug Discov. Today*, v. 11, p. 999-1006, 2006.
4. YAGO, G.; AMRAM, M.; MAGULA, T. Financial innovations for funding early-stage. *Drug Discov. Dev.*, v. 12, p. 22-28, 2006.
5. KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.*, v. 3, p. 935-949, 2004.
6. KUHN, D. et al. From the similarity analysis of protein cavities to the functional classification of protein families using cavbase. *J. Mol. Biol.*, v. 359, p. 1023-1044, 2006.
7. LOMBARDINO, J. G.; LOWE, J. A. 3rd. The role of the medicinal chemist in drug discovery-then and now. *Nat. Rev. Drug Discov.*, v. 3, p. 853-862, 2004.
8. WERMUTH, C. G. *The Practice of Medicinal Chemistry*. London: Academic Press, 2003.
9. AGNEW, B. When pharma merges, R&D is the dowry. *Science*, v. 287, p. 1952-1953, 2000.
10. COHEN, F. J. Macro trends in pharmaceutical innovation. *Nat. Rev. Drug Discov.*, v. 4, p. 78-84, 2005.
11. COUZIN, J. Brains behind blockbuster. *Science*, v. 309, p. 728-730, 2005.

12. BERNDT, E. R. et al. Industry funding of the FDA: effects of PDUFA on approval times and withdrawal rates. *Nat. Rev. Drug Discov.*, v. 4, p. 545-554, 2005.
13. DIMASI, J. A.; HANSEN, R. W.; GRABOWSKI, H. G. The price of innovation: new estimates of drug development cost. *J. Health Econom.*, v. 22, p. 151-185, 2003.
14. UNITED STATES GOVERNMENT ACCOUNTABILITY OFFICE (2006). New drug development. Science, business, regulatory, and intellectual property issue cited as hampering drug development efforts. Disponível em: <<http://www.gao.gov/new.items/d0749.pdf>>. Acesso em: 14 Jan 2008.
15. DICKSON M., GAGNON J. P. Key factors in the rising cost of new drug discovery and development. *Nat. Rev. Drug. Discov.*, v. 5, p. 417-429, 2004.
16. MCGEE, P. Clinical Trials on the Move. *Drug Discov. Dev.*, v. 9, p. 16-22, 2006.
17. KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today*, v. 11, p. 580-594, 2006.
18. ANDERS, H. J.; VIELHAUER, V. Identifying and validating novel targets with in vivo disease models: guidelines for study design. *Drug Discov. Today*, v. 12, p. 446-451, 2007.
19. SALUM, L. B. *Estudos in silico no planejamento de candidatos a novos fármacos na terapia do câncer de mama e de reposição hormonal*. 2007. 140 f. Dissertação (Mestrado em Física Biomolecular) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2007.
20. BARREIRO, E. J.; FRAGA, C. A. M. *Química medicinal: as bases moleculares da ação dos fármacos*. Porto Alegre: Artmed Editora, 2001.
21. OPREA, T. I. et al. Lead-like, drug-like or "Pub-like": how different are they? *J. Comput. Aided Mol. Des.*, v. 21, p.113-119, 2007.
22. THOMAS, G. *Fundamentals of medicinal chemistry*. West Sussex: Wiley, 2003.

23. JÓNSDÓTTIR S. O.; JØRGENSEN, F. S.; BRUNAK, S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics*, v. 21, p. 2145-2160, 2005.
24. BELFIELD, G. P.; DELANEY, S. J. The impact of molecular biology on drug discovery. *Biochem. Soc. Trans.*, v. 34, p. 313-316, 2006.
25. KOGEJ, T. et al. Multifingerprint based similarity searches for targeted class compound selection. *J. Chem. Inf. Model.*, v. 46, p. 1201-1213, 2006.
26. BALAKIN, K. V. et al. Focused chemistry from annotated libraries. *Drug Discov. Today: Technol.*, v. 3, p. 397-403, 2006.
27. BUNIN, A. B. Increasing the efficiency of small molecule drug discovery. *Drug. Discov. Today*, v. 18, p. 823-826, 2003.
28. ANDRICOPULO, A. D.; MONTANARI, C. A. Structure-activity relationships for the design of Small-molecule inhibitors. *Mini Rev. Med. Chem.*, v. 5, p. 585-593, 2005.
29. HOLBROOK, J. D.; SANSEAU, P. Drug discovery and computational evolutionary analysis. *Drug Discov. Today*, v. 12, p. 826-832, 2007.
30. RUSSO, E. Chemistry plans a structural overhaul. *Nature*, v. 419, p. 4-7, 2002.
31. YOUNG, D. Drug innovation on the decline. *Am. J. Health Syst. Pharm.*, v. 64, p. 228-231, 2007.
32. FEDERSEL, H. J. In search of sustainability: process R&D in light of current pharmaceutical industry challenges. *Drug Discov. Today*, v. 11, p. 966-974, 2006.
33. BROWN, D. Unfinished business: target-based drug discovery. *Drug Discov. Today*, v. 12, p. 1007-1012, 2007.
34. SCHMID, E. F.; SMITH, D. A. Pharmaceutical R&D in the spotlight: why is there still unmet medical need? *Drug Discov. Today*, v. 12, p. 998-1006, 2007.

-
35. SCHMID, E. F.; SMITH, D. A. R&D technology investments: misguided and expensive or a better way to discover medicines? *Drug Discov. Today*, v. 11, p.775-784, 2006.
36. MITSUYA, H.; YARCHOAN, R.; BRODER, S. Molecular targets for AIDS therapy. *Science*, v. 249, p. 1533-1544, 1990.
37. THAIMATTAM, R. et al. Protein kinase inhibitors: structural insights into selectivity. *Curr. Pharm. Des.*, v. 13, p. 2751-2765, 2007.
38. VON ITZSTEIN, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, v. 363, p. 418-423, 1993.
39. IGLESIAS, P.; DÍEZ, J. J. New drugs for the treatment of hypercholesterolaemia. *Expert. Opin. Investig. Drugs.*, v. 12, p. 1777-1789, 2003.
40. PFEFFER, P.; GOHLKE, H. DrugScore^{RNA}-knowledge-based scoring function to predict RNA-ligand interactions. *J. Chem. Inf. Model.*, v. 47, p. 1868-1876, 2007.
41. BERMAN, H. M. et al. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, suppl, p. 957-959, 2000.
42. WESTBROOK, J. et al. The Protein Data Bank and structural genomics. *Nucleic Acids Res.*, v. 31, p. 489-491, 2003.
43. ECKERT, H.; BAJORATH, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today*, v. 12, p. 225-233, 2007.
44. WILLETT, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today*, v. 11, p. 1046-1053, 2006.
45. KUBINIY, H. *3D QSAR in drug design: theory, methods and applications*. Leiden:Kluwer Academic Publishers/Escom, 1993. v. 1.
46. BEYRER, C. et al. Neglected diseases, civil conflicts, and the right to health. *Lancet*, v. 370, p. 619-627, 2007.

47. TROUILLER, P. et al. Drug development for neglected diseases: a deficient market and a public-health policy failure. *Lancet*, v. 359, p. 2188-2194, 2002.
48. NWAKA, S.; HUDSON, A. Innovative lead discovery strategies for tropical diseases. *Nat. Rev. Drug Discov.*, v. 5, p. 941-955, 2006.
49. NWAKA, S.; RIDLEY, R. G. Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nat. Rev. Drug Discov.*, v. 2, p. 919-928, 2003.
50. WIDDUS, R. Public-private partnerships for health: their main targets, their diversity, and their future directions. *Bull. World Health Organization*, v. 79, p. 728-734, 2001.
51. WHO/TDR - World Health Organization - Special Programme for Research and Training in Tropical Diseases, 2008. Disponível em: <<http://www.who.int/tdr/index.html>>. Acessado em: 16 Jan 2008.
52. MMV - Medicines for Malaria Venture, 2008. Disponível em: <<http://www.mmv.org/>>. Acessado em: 16 Jan 2008.
53. DNDi - Drugs for Neglected Diseases initiative, 2008. Disponível em: <<http://www.dndi.org/>>. Acessado em: 16 Jan 2008.
54. WORLD HEALTH ORGANIZATION. *Control of Chagas disease: second report of the WHO expert committee*. Geneva: WHO, 2002. (Technical Report Series, 905).
55. MONCAYO, A. Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. *Mem. Inst. Oswaldo Cruz*, v. 98, p. 577-591, 2003.
56. MONCAYO, A.; ORTIZ YANINE, M. I. An update on Chagas disease (human American trypanosomiasis). *Ann. Trop. Med. Parasitol.*, v. 100, p. 663-677, 2006.
57. MOREL, C. M. Chagas disease, from discovery to control- and beyond: history, myths and lessons to take home. *Mem. Inst. Oswaldo Cruz*, v. 94, p. 3-16, 1999.

-
58. DIAS, J. C. P.; SCHOFIELD, C. J. The evolution of Chaga's Disease (American Trypanosomiasis) Control after 90 Years since Carlos Chagas Discovery. *Mem. Inst. Oswaldo Cruz*, v. 94, p. 103-121, 1999.
59. PAHO – Pan American Health Organization Chagas Disease (American trypanosomiasis) Disponível em: <<http://www.paho.org/english/ad/dpc/cd/chagas.htm>> Acessado em: 16 Jan 2008.
60. EL-SAYED; N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, v. 309, p. 409-415, 2005.
61. TARLETON, R. L. The *Trypanosoma cruzi* proteome. *Science*, v. 309, p. 473-476, 2005.
62. SCHOFIELD, C. J.; JANNIN, J.; SALVATELLA, R. The future of Chagas disease control. *Trends Parasitol.*, v. 22, p. 583-588, 2006.
63. FAIRLAMB, A.H. Future prospects for the chemotherapy of Chaga's disease. *Medicina-Buenos Aires*, v. 59, p. 179, 1999.
64. VERONESI, R. *Doenças infecciosas e parasitárias*, Rio de Janeiro: Guanabara Koogan, 1991.
65. CINQUE, G. M. et al. Structure-activity relationship of new growth inhibitors of *Trypanosoma cruzi*. *J. Med. Chem.*, v. 41, p. 1540-1554, 1998.
66. BUCKNER, F. S. et al. Potent anti-*Trypanosoma cruzi* activities of oxidosqualene cyclase inhibitors. *Antimicrob. Agents Chemother.*, v. 45, p. 1210-1215, 2001.
67. URBINA, J. A.; et al. Cure of short- and long-term experimental Chagas' disease using D0870. *Science*, v. 273, p. 969-971, 1996
68. NUSSBAUMER, P. et al. Synthesis and structure-activity relationships of sidechain-substituted analogs of the allylamine antimycotic Terbinafine lacking the central amino function. *J. Med. Chem.*, v. 38, p. 1831-1836, 1995.

-
69. LAZARDI, K.; URBINA, J. A.; DE SOUZA, W. Ultrastructural alterations induced by two ergosterol biosynthesis inhibitors, Ketoconazole and Terbinafine, on epimastigotes and amastigotes of *Trypanosoma (Schizotrypanum) cruzi*. *Antimicrob. Agents Chemother.*, v. 34, 2097-2105, 1990.
70. CHAMOND, N.; COATNOAN, N.; MINOPRIO, P. Immunotherapy of *Trypanosoma cruzi* infections. *Curr. Drug Targets Immune Endocr. Metabol. Disord.*, v. 2, p. 247-254, 2002.
71. BASSO, B. et al. Vaccination with *Trypanosoma rangeli* reduces the infectiousness of dogs experimentally infected with *Trypanosoma cruzi*. *Vaccine*, v. 25, p. 3855-3858, 2007.
72. ARAÚJO, F. G.; MOREIN, B. Immunization with *Trypanosoma cruzi* - epimastigote antigens incorporated into ISCOMS protects against lethal challenge in mice. *Infect. Immun.*, v. 59, p. 2909-2914, 1991.
73. BUA, J. *Trypanosoma cruzi* – Cellular and antibody response against the parasite in mice immunizes with a 19-amino acid synthetic peptide. *Exp. Parasitol.*, v. 72, p. 54-62, 1991.
74. LUHRS, K. A.; FOUTS, D. L.; Manning J. E. Immunization with recombinant paraflagellar rod protein induces protective immunity against *Trypanosoma cruzi* infection. *Vaccine*, v. 21, p. 3058–3069, 2003.
75. MUNIZ, J.; NÓBREGA, G.; CUNHA, M. Ensaios de vacinação preventiva e curativa nas infecções pelo *Scizotrypanum cruzi*. *Mem. Inst. Oswaldo Cruz*, v. 44, p. 529-541, 1946.
76. HOTEZ, P. J.; FERRIS, M. T. The antipoverty vaccines. *Vaccine*, v. 24, p. 5787-5799, 2006.
77. HERWALDT, B. L. Leishmaniasis. *Lancet*, v. 354, p. 1191-1199, 1999.
78. DESJEUX, P. Leishmaniasis. *Nat. Ver. Microbiol.*, v. 2, p. 692-693, 2004.
79. DESJEUX, P. Leishmaniasis: current situation and new perspectives. *Comp. Immunol. Microbiol. Infect. Dis.*, v. 27, p. 305-318, 2004.
80. DESJEUX, P. The increase in risk factors for leishmaniasis worldwide. *Trans. R. Soc. Trop. Med. Hyg.*, v. 95, p. 239-243, 2001.

-
81. MURRAY, H. W. et al. Advances in leishmaniasis. *Lancet*, v. 366, p. 1561-1577, 2005.
82. YAMEY, G.; TORRELE, E. The world's most neglected diseases. *BMJ*, v. 325, p. 176-177, 2002.
83. MAGILL, A. J. et al. Visceral infection caused by *Leishmania tropica* in veterans of Operation Desert Storm. *N. Engl. J. Med.*, v. 328, p. 1383-1387, 1993.
84. ALVAR, J. et al. *Leishmania* and human immunodeficiency virus coinfection: the first 10 years. *Clin. Microbiol. Rev.*, v. 10, p. 298-319, 1997.
85. DAVIDSON, R. N. AIDS and leishmaniasis. *Genitourin. Med.* v. 73, p. 237-239, 1997.
86. CROFT, S. L.; YARDLEY, V. Chemotherapy of leishmaniasis. *Curr. Pharm. Des.*, v. 8, p. 319-342, 2002.
87. DAVIES, C. R. et al. Leishmaniasis: new approaches to disease control. *BMJ*, v. 326, p. 77-82, 2003.
88. CROFT, S. L. Neglected diseases: progress in drug development. *Curr. Opin. Investig. Drugs*, v. 8, p. 103-104, 2007.
89. CROFT, S. L.; COOMBS, G. H. Leishmaniasis-current chemotherapy and recent advances in the search for novel drugs. *Trends Parasitol.*, v. 19, p. 502-508, 2003.
90. CROFT, S. L.; BARRETT, M. P.; URBINA, J. A. Chemotherapy of trypanosomiasis and leishmaniasis. *Trends Parasitol.*, v. 21, p. 508-512, 2005.
91. CROFT, S. L. Monitoring drug resistance in leishmaniasis. *Trop. Med. Int. Health*, v. 6, p. 899-905, 2001.
92. GEARY, T. G.; EDGAR, A. E. JENSEN, J. B. Leishmaniasis - current treatment. In: CAMPBELL, W.C. e REW. R.S. (eds.). *Chemotherapy of Parasitic Diseases*. New York: Plenum Press, 1989, p. 209-238.
93. SINDERMAN, H. et al. Miltefosine (Impavido): the first oral treatment against leishmaniasis. *Med. Microbiol. Immunol.*, v. 193, p. 173-180, 2004.

93. SINDERMANN, H.; ENGEL, J. Development of miltefosine as an oral treatment for leishmaniasis. *Trans. R. Soc. Trop. Med. Hyg.*, v. 100, p. S17-S20, 2006.
94. PINHEIRO, R. O. et al. Protection against cutaneous leishmaniasis by intranasal vaccination with lipophosphoglycan. *Vaccine*, v. 25, p. 2716-2722, 2007.
95. ARMIJOS, R. X. et al. Safety, immunogenicity, and efficacy of an autoclaved *Leishmania amazonensis* vaccine plus BCG adjuvant against NewWorld cutaneous leishmaniasis. *Vaccine*, v. 22, p. 1320-1326, 2004.
96. VELEZ, I. D. et al. Failure of a killed *Leishmania amazonensis* vaccine against American cutaneous leishmaniasis in Colombia. *Trans. R. Soc. Trop. Med. Hyg.*, v. 99, p. 593-598, 2005.
97. ROBERTSON, J. G. Mechanistic basis of enzyme-targeted drugs. *Biochemistry*, v. 44, p. 5561-5571, 2005.
98. LESPINET, O.; LABEDAN, B. Orphan enzymes could be an unexplored reservoir of new drug targets. *Drug Discov. Today*, v. 11, p. 300-305, 2006.
99. MESTRES, J. Representativity of target families in the Protein Data Bank: impact for family-directed structure-based drug discovery. *Drug Discov. Today*, v. 10, p. 1629-1637, 2005.
100. GUIDO R. V. C.; et al. Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*. *J. Enz. Inhib. Med. Chem.*, v. 23, 2008. In press
101. COPELAND, R. A. *Evaluation of enzyme inhibitors in drug discovery*. New Jersey:Wiley, 2005.
102. MICHELS P. A. M. Compartmentation of glycolysis in trypanosomes: a potential target for new trypanocidal drugs. *Biol. Cell.*, v. 64, p. 157-164, 1988.
103. MICHELS, P. A. M. et al. Metabolic functions of glycosomes in trypanosomatids. *Biochim. Biophys. Acta.*, v. 1763, p. 1463-1477, 2006.

-
104. VERLINDE, C. L. M. et al. Glycolysis as a target for the design of new anti trypanosome drugs. *Drug Resist. Updat.*, v. 4, p. 50-65, 2001.
105. CAZZULO J. J.; STOKA, V.; TURK, V. Cruzipain, the major cysteine proteinase from the protozoan parasite *Trypanosoma cruzi*. *Biol. Chem.*, v. 378, p. 1-10, 1997.
106. APARICIO, I. M.; SCHARFSTEIN, J.; LIMA, A. P. A new Cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. *Infect. Immun.*, v. 72, p. 5892-5902, 2004.
107. SOUTO-PADRÓN, T. et al. Cysteine proteinase in *Trypanosoma cruzi*: Immunocytochemical localization and involvement in parasite-host cell interaction. *J. Cell Sci.*, v. 96, p. 485-490, 1990.
108. FRANKE DE CAZZULO, B. M. et al. Effect of proteinase inhibitors on growth and differentiation of *Trypanosoma cruzi*. *FEMS Microbiol. Lett.*, v. 124, p. 81-86, 1994.
109. SCHMIDT, A.; KRAUTH-SIEGEL, R. L. Enzymes of the trypanothione metabolism as targets for antitrypanosomal drug development. *Curr. Topics Med. Chem.*, v. 2, p. 1239-1259, 2002.
110. KRAUTH-SIEGEL, R. L.; INHOFF, O. Parasite-specific trypanothione reductase as a drug target molecule. *Parasitol Res.*, v. 90, p. 77-85, 2003.
111. SOUZA, D. H. F. et al. *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design. *FEBS Lett.*, v. 424, p. 131-135, 1998.
112. MICHELS, P. A. M.; HANNAERT, V. The evolution of kinetoplastid glycosomes. *J. Bioenerg. Biomembr.*, v. 26, p. 213-219, 1994.
113. CLARKSON, A. B. JR.; BROHN, F. H. Trypanosomiasis: an approach to chemotherapy by the inhibition of carbohydrate catabolism. *Science*, v. 8, p. 204-206, 1976.
114. GUERRA, G. C.; QUIJADA, L.; CLAYTON, C. E. Compartmentation of enzymes in a microbody, the glycosome, is essential in *Trypanosome brucei*. *J. Cell Science*, v. 115, p. 2651-2658, 2002.

-
115. KENNEDY, K. J.; BRESSI, J. C.; GELB, M. H. A disubstituted NAD⁺ analogue is a nanomolar inhibitor of trypanosomal Glyceraldehyde-3-Phosphate Dehydrogenase. *Bioorg. Med. Chem. Lett.*, v.11, p. 95-98, 2001.
116. COURA, J. R.; CASTRO, S. L. A Critical Review on Chagas Disease Chemotherapy. *Mem. Inst. Oswaldo Cruz*, v. 97, p. 3-24, 2002.
117. BAKKER, B. M. et al. What controls Glycolysis in bloodstream form *Trypanosoma brucei*. *J. Biol. Chem.*, v. 274, p. 14551-14559, 1999.
118. BAKKER, B. M. et al. Metabolic control analysis of glycolysis in trypanosomes as an approach to improve selectivity and effectiveness of drugs. *Mol. Biochem. Parasitol.*, v. 106, p. 1-10, 2000.
119. OPPERDOES, F. R.; MICHELS, P. A. M. Enzymes of carbohydrate metabolism as potential drug targets. *Int. J. Parasitology*, v. 31, p. 482-490, 2001.
120. ARONOV, A. M. et al. Structure-based design of submicromolar, biologically active inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. *Proc. Natl. Acad. Sci. USA*, v. 96, p. 4273-4278, 1999.
121. ARONOV, A. M. et al. Selective tight binding inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase via structure-based design design. *J. Med. Chem.*, v. 41, p. 4790-4799, 1998.
122. VAN CALENBERGH, S. et al. Synthesis and structure-activity relationships of analogs of 2'-deoxy-2'-(3-methoxybenzamido)adenosine, a selective inhibitor of trypanosomal glycosomal glyceraldehyde-3-phosphate dehydrogenase. *J. Med. Chem.*, v. 38, p. 3838-3849, 1995.
123. VERLINDE, C. L. M. J. et al. Selective inhibition of trypanosomal glyceraldehyde-3-phosphate dehydrogenase by protein structure-based design: toward new drugs for the treatment of sleeping sickness. *J. Med. Chem.*, v. 37, p. 3605-3613, 1994.
124. BRESSI, J. C. et al. Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of *Trypanosomatidae* via structure-based design design. *J. Med. Chem.*, v. 44, p. 2080-2093, 2001.

-
125. GUIDO, R. V. C. et al. Structural basis for selective inhibition of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase: molecular docking and 3D QSAR studies. *J. Chem. Inf. Model.*, v. 48, 2008. In press
126. GUIDO, R. V. C. et al. Classical and hologram QSAR studies on a series of inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. *QSAR Comb. Sci.*, v. 27, 2008. In press
127. GALLO, M. B. C. et al. Enzymatic inhibitory activity and trypanocidal effects of extracts and compounds from *Siphoneugena densiflora* Berg and *Vitex polygama* Cham. *Z. Naturforsch. [C]*, v. 63, 2008. In press
128. LEITÃO, A. et al. Structure-activity relationships of novel inhibitors of glyceraldehyde-3-phosphate dehydrogenase. *Bioorg. Med. Chem. Lett.*, v. 14, p. 2199-2204, 2004.
129. MENEZES, I. R. et al. 3D QSAR studies on binding affinities of coumarin natural products for glycosomal GAPDH of *Trypanosoma cruzi*. *J. Comput. Aided Mol. Des.*, v. 17, p. 277-290, 2003.
130. DE MARCHI, A. A. et al. New 3-piperonylcoumarins as inhibitors of glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) from *Trypanosoma cruzi*. *Bioorg. Med. Chem.*, v. 12, p. 4823-4833, 2004.
131. NAGRADOVA, N. K. Study of the properties of phosphorylating D-glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry (Mosc.)*, v. 66, p. 1067-1076, 2001.
132. NAGRADOVA, N. K. et al. D-glyceraldehyde-3-phosphate dehydrogenase. Properties of the enzyme modified at arginine residues. *Appl. Biochem. Biotechnol.*, v. 61, p. 47-56, 1996.
133. EBY, D.; KIRTLY, M. E. Cooperativity and noncooperativity in the binding of NAD analogues to rabbit muscle glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*, v. 15, p. 2168-2171, 1976.
134. HENIS, Y. I.; LEVITZKI, A. Mechanism of negative cooperativity in glyceraldehyde-3-phosphate dehydrogenase deduced from ligand competition experiments. *Proc. Natl. Acad. Sci. USA*, v. 77, p. 5055-5059, 1980.

-
135. LESLIE, A.G.; WONACOTT, A. J. Coenzyme binding in crystals of glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.*, v. 165, p. 375-391, 1983.
136. ROITEL, O. et al. P but not R-axis interface is involved in cooperative binding of NAD on tetrameric phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *J. Mol. Biol.*, v. 326, p. 1513-1522, 2003.
137. SHEN, Y. Q.; SONG, S. Y.; LIN, Z. J. Structures of D-glyceraldehyde-3-phosphate dehydrogenase complexed with coenzyme analogues. *Acta Crystallogr, D Biol. Crystallogr.*, v. 58, p. 1287-1297, 2002.
138. JEFFERY, C. J. Moonlighting proteins. *Trends Biochem. Sci.*, v. 24, p. 8-11, 1999.
139. SIROVER, M. A. New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells. *J. Cell. Biochem.*, v. 95, p. 45-52, 2005.
140. SIROVER, M. A. New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim. Biophys. Acta.*, v. 1432, p. 159-184, 1999.
141. TATTON, W. G. et al. Apoptosis in Parkinson's disease: signals for neuronal degradation. *Ann. Neurol.*, v. 53, p. S61-70, 2003.
142. TATTON, W. G. et al. Glyceraldehyde-3-phosphate dehydrogenase in neurodegeneration and apoptosis signaling. *J. Neural. Transm. Suppl.*, v. 60, p. 77-100, 2000.
143. WALDMEIER, P. C.; TATTON, W. G. Interrupting apoptosis in neurodegenerative disease: potential for effective therapy? *Drug Discov. Today*, v. 9, p. 210-218, 2004.
144. CARLILE, G. W. et al. Reduced apoptosis after nerve growth factor and serum withdrawal: conversion of tetrameric glyceraldehyde-3-phosphate dehydrogenase to a dimer. *Mol. Pharmacol.*, v. 57, p. 2-12, 2000.
145. BURKE, J. R. et al. Huntingtin and DRPLA proteins selectively interact with the enzyme GAPDH. *Nat. Med.*, v. 2, p. 347-350, 1996.

146. SATCHELL, J. F. et al. Structure of glyceraldehyde-3-phosphate dehydrogenase from *Plasmodium falciparum*. *Acta Crystallogr. D Biol. Crystallogr.*, v. 61, p. 1213-1221, 2005.
147. KIM, H. et al. Crystal Structure of glycosomal glyceraldehyde-3-Phosphate dehydrogenase from *Leishmania mexicana*: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry*, v. 34, p. 14975-14986, 1995.
148. VELLIEUX, F. M. et al. Structure of glycosomal glyceraldehyde-3-Phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue Data. *Proc. Natl. Acad. Sci. USA*, v. 90, p. 2355-2359, 1993.
149. HANNAERT, V.; OPPERDOES, F. R.; MICHELS, P. A. M. Glycosomal glyceraldehyde-3-phosphate dehydrogenase of *Trypanosoma brucei* and *Trypanosoma cruzi*: expression in *Escherichia coli*, purification, and characterization of the enzymes. *Protein Expr. Purif.*, v. 6, p. 244-250, 1995.
150. PAVÃO, F. et al. Structure of *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 Å resolution. *FEBS Lett.*, v. 520, p. 13-17, 2002.
151. LADAME, S. et al. Crystal structure of *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase complexed with an analogue of 1,3-bisphospho-d-glyceric acid. *Eur. J. Biochem.*, v. 270, p. 4574-4586, 2003.
152. CASTILHO, M. S. et al. Evidence for the two phosphate binding sites of an analogue of the thioacyl intermediate for the *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase-catalyzed reaction, from its crystal structure. *Biochemistry*, v. 42, p. 7143-7151, 2003.

Capítulo 2

Objetivos

Nesse capítulo são apresentados os objetivos fundamentais desta tese de doutorado, envolvendo os estudos em química medicinal e biologia estrutural no planejamento de inibidores da enzima GAPDH de tripanosomatídeos.

2 - OBJETIVOS

Os objetivos fundamentais desta tese de doutorado são:

- ✓ Realizar triagens bioquímicas de compostos de origem natural e sintética para a identificação de novos inibidores da GAPDH de *T. cruzi*.
- ✓ Determinar as constantes cinéticas do cofator e do substrato, além de avaliar o mecanismo de cooperatividade da GAPDH de *T. cruzi*.
- ✓ Desenvolver estudos estruturais para a elucidação das bases moleculares envolvidas nos fenômenos de cooperatividade entre as subunidades da enzima-alvo.
- ✓ Organizar um conjunto de dados padrão de inibidores da enzima GAPDH de tripanosomatídeos.
- ✓ Gerar modelos preditivos de QSAR empregando os métodos HQSAR (2D), CoMFA e CoMSIA (3D) para o conjunto padrão de inibidores.
- ✓ Realizar estudos de modelagem molecular e de QSAR 3D capazes de revelar aspectos estruturais relevantes para o planejamento de inibidores seletivos da enzima GAPDH de tripanosomatídeos.
- ✓ Executar uma estratégia completa de ensaio virtual baseado na estrutura do receptor para a identificação de candidatos e inibidores da enzima GAPDH de *T. cruzi*.

Capítulo 3

*Identificação de Novos Inibidores da Enzima GAPDH de *T. cruzi**

Nesse capítulo é realizada uma introdução sobre os principais tipos de inibidores enzimáticos existentes. Em seguida, há uma breve descrição do processo de obtenção da enzima GAPDH de *T. cruzi* (expressão e purificação) e de como são realizados os ensaios de triagem de compostos de origem natural e sintética. Por último, a potência dos inibidores identificados é determinada e os resultados são discutidos.

3 - IDENTIFICAÇÃO DE NOVOS INIBIDORES DA ENZIMA GAPDH DE *T. cruzi*

3.1 - Inibidores enzimáticos

Inibidores enzimáticos são substâncias químicas capazes de reduzir a velocidade de reações catalisadas por enzimas-alvo.¹ Nos estudos de cinética enzimática, o reagente é designado de substrato (S). O substrato se liga à enzima (E) para formar o chamado complexo enzima-substrato (E•S), o qual sofre subsequente transformação bioquímica seguida pela dissociação do produto (P) e liberação da enzima livre. Considerando um caso mais simples, a reação catalisada por uma enzima envolve a conversão de um único S em P. A formação do complexo E•S ocorre de forma rápida e reversível, envolvendo interações intermoleculares não-covalentes, sem o registro de modificações químicas. Os inibidores enzimáticos possuem notáveis aplicações terapêuticas, promovendo uma variedade de benfeitorias significativas na qualidade de vida de milhões de indivíduos no mundo todo.^{1,2}

Os inibidores enzimáticos atuam por mecanismos reversíveis ou irreversíveis, associando-se às enzimas através de ligações intermoleculares não-covalentes (E•I) ou covalentes (E-I), respectivamente. Os inibidores irreversíveis causam inativação enzimática, enquanto que os inibidores reversíveis podem se ligar as várias espécies da enzima em equilíbrio e não causam modificações na estrutura de enzima, sendo que a atividade enzimática pode ser recuperada com a remoção do inibidor.

Os inibidores enzimáticos do tipo reversível se ligam à enzima de forma rápida, não-covalente e não produzem mudanças na estrutura da enzima. Podem ser classificados em competitivos, não-competitivos e incompetitivos (Figura 3.1).

Figura 3.1 – Esquemas cinéticos de inibidores do tipo reversível. À esquerda, o esquema clássico de inibição competitiva, onde o inibidor se liga exclusivamente a enzima livre; na parte central, o esquema cinético para a inibição não-competitiva, onde o inibidor se liga a enzima livre e ao complexo enzima-substrato; enquanto que a direita tem-se um esquema similar para a inibição incompetitiva, onde o inibidor se liga exclusivamente ao complexo enzima-substrato. Os gráficos de duplo recíproco de Lineweaver-Burk característicos para os diferentes tipos de inibição são apresentados na parte inferior.^{1,2}

Os inibidores enzimáticos do tipo reversível competitivo se ligam à enzima de forma rápida, de tal forma que impedem a ligação do S à enzima livre. Por definição, este tipo de inibição requer a competição entre S e I pela ligação ao sítio ativo da enzima-alvo, sendo esta ligação mútua exclusiva (i.e., formação da espécie $E \cdot S$ ou $E \cdot I$).

Os inibidores do tipo não-competitivo não se ligam ao sítio ativo da enzima e não tem efeito na ligação do S. Os inibidores não-competitivos apresentam afinidade de ligação pela enzima livre (em outras regiões diferentes do sítio do S) e pelo complexo binário $E \cdot S$. Neste caso, S e I se ligam de forma reversível e independente em sítios diferentes da enzima. Os complexos resultantes $E \cdot I$ e $E \cdot S \cdot I$ são cataliticamente inativos.

Os inibidores do tipo incompetitivo se ligam exclusivamente ao complexo $E \cdot S$ e não possuem afinidade pela enzima livre. O complexo resultante $E \cdot S \cdot I$ é cataliticamente inativo. Este

mecanismo puro de inibição enzimática é o mais raro em estudos de planejamento de inibidores, embora existam na clínica médica exemplos de fármacos importantes que atuam por mecanismo incompetitivo (e.g., metotrexato, finasterida, camptotecina).¹

3.2 - Expressão e purificação da enzima GAPDH de *T. cruzi*

O gene que codifica a enzima GAPDH de *T. cruzi* foi clonado no vetor pET3a e introduzido em células de *E. coli* da linhagem BL21(DE3)gap⁻.³ No presente trabalho, apenas uma modificação foi introduzida no protocolo de purificação,⁴ já que a resina da coluna de troca catiônica Phospho Ultrogel AGR (IBF Biotechnics) não era mais produzida comercialmente. Por essa razão, a segunda coluna utilizada no protocolo de purificação foi substituída por uma coluna de celulose fosfato (Whatman). Essa coluna apresentou o mesmo perfil de eluição da coluna Phospho Ultrogel AGR e, portanto, não introduziu modificações significativas na pureza da enzima obtida (Figura 3.2).

Figura 3.2 - Gel de poliácridamida 15% (SDS-PAGE) referente às etapas de purificação da GAPDH.

Os ensaios de triagem bioquímica e de cristalização descritos nesta tese de doutorado foram realizados utilizando a enzima GAPDH produzida através deste processo.

3.3 - Ensaios bioquímicos com a enzima GAPDH

O processo de triagem bioquímica *in vitro* é de grande valor para a identificação de inibidores de enzimas-alvo em programas de planejamento de fármacos. Compostos promissores podem ser selecionados para posterior determinação de potência e afinidade.^{5,6} Em nossos estudos, a atividade enzimática foi monitorada através de ensaio espectrofotométrico de ultravioleta (UV), com base na reação bioquímica catalisada pela enzima GAPDH (Figura 1.12B). O meio reacional foi constituído de GAPDH de *T. cruzi* (10-20 nM), solução-tampão trietanolamina (100 mM), arseniato de sódio (30 mM), NAD⁺ (0,4 mM), β -mercaptoetanol (1 mM), G3P (0,6 mM) e EDTA (1 mM). As reações foram realizadas a 25°C em pH 7,4. Todos os reagentes comerciais utilizados foram de grau analítico (Sigma-Aldrich). Para a avaliação da atividade inibitória, os compostos em estudo foram adicionados ao meio reacional em concentrações crescentes. A reação enzimática foi iniciada com a adição da GAPDH ao meio reacional. O ensaio espectrofotométrico foi realizado monitorando-se a formação de NADH no comprimento de onda de 340 nm.

Alguns cuidados devem ser tomados antes da realização de triagens bioquímicas, para a validação das substâncias a serem consideradas no ensaio enzimático. A primeira etapa no pré-processo de triagem de inibidores consistiu na aquisição de um espectro de varredura da amostra a ser avaliada (candidato a inibidor), em nosso caso, na faixa entre 200 e 400 nm. Esta etapa é importante na identificação de possível interferência dos candidatos a inibidor na região do espectro de absorção em que se realizam as leituras do ensaio cinético. Na etapa seguinte, o

composto em estudo foi adicionado ao tampão na presença de NAD^+ e uma análise a 340 nm foi realizada. Este procedimento é de fundamental importância para avaliar se o composto *per si* é capaz de reduzir o NAD^+ a NADH, causando erros de leitura e interpretações equivocadas sobre a atuação dos candidatos a inibidores. Finalizados estes testes preliminares, que foram conduzidos rigorosamente para todas as amostras estudadas, os compostos puderam ser selecionados para os ensaios de inibição enzimática. Os inibidores são dissolvidos e adicionados à mistura reacional na concentração final de 10% DMSO. Nessa concentração, o solvente orgânico não afeta significativamente a atividade da enzima,⁷ como observado em nossos experimentos cinéticos.

No processo de triagem de substâncias químicas, um teste de inibição em dose (para mistura de substâncias) ou concentração (compostos puros) única é geralmente empregado na avaliação inicial. Compostos puros ou misturas que apresentam inibição maior do que 30% podem ser selecionados para novas avaliações e encaminhados para ensaios de determinação de potência. Em nossos estudos, a concentração única de 200 μM foi empregada nas triagens iniciais, sendo selecionados os compostos que apresentaram inibição substancial (> 30%). Para a determinação da potência através de valores de IC_{50} , que se referem à concentração de inibidor necessária para reduzir em 50% a atividade enzimática, foram utilizadas cinco ou seis concentrações de inibidor, capazes de reduzir a atividade enzimática entre 15-85%. Todas as medidas foram realizadas em triplicata. Em todos os testes de inibição, utilizou-se como controle negativo a mistura reacional acrescida de 10% de DMSO, e como controle positivo uma solução 200 μM de uma amostra denominada de inibidor padrão (uma cumarina sintética com IC_{50} de 64 μM). Os dados experimentais coletados foram analisados graficamente com o módulo de cinética enzimática disponível no programa SigmaPlot versão 7.0 e os valores de IC_{50} determinados através de análise de regressão não-linear de melhor ajuste.

3.4 - Inibidores da GAPDH de *T. cruzi*

Os compostos estudados nesta tese de doutorado são de origem natural ou sintética.

3.4.1 - Produtos naturais como inibidores

A natureza é uma fonte muito rica em moléculas de grande diversidade química e potencial biológico.⁸ Dados do FDA entre os anos de 1981 e 2006 indicam que 5% das NCEs aprovadas são produtos naturais. Entretanto, quando se considera o número de NCEs derivadas de produtos naturais ou planejadas a partir de produtos naturais, esse número sobe para 52%, demonstrando a enorme importância dos produtos naturais no processo de descoberta de fármacos.⁹

Com o objetivo de identificar novos compostos capazes de modular a atividade da enzima GAPDH de *T. cruzi*, extratos vegetais brutos, misturas fracionadas e compostos isolados da flora brasileira foram avaliados neste trabalho, seguindo atividades de triagem que rotineiramente são realizadas em nosso laboratório. O ensaio bioquímico utilizado para o processo de triagem foi padronizado e validado em nossos laboratórios. Entre os vários produtos naturais investigados, alguns extraídos da casca da castanha do caju merecem destaque.

A *Anacardium occidentale* L. é uma planta típica de campos e dunas da costa norte do Brasil e popularmente conhecida como cajueiro. O líquido proveniente da casca da castanha do caju apresenta misturas de lipídeos fenólicos de cadeias carbônicas longas saturadas e insaturadas (i.e., uma, duas ou três instaurações não conjugadas de configuração *cis*). O óleo natural é constituído majoritariamente de uma mistura de ácido anacárdico, cardol e cardanol (Figura 3.3).¹⁰ Em colaboração com o Prof. Dr. Paulo Cezar Vieira e a sua aluna de doutorado Richele Severino, do Laboratório de Produtos Naturais do Departamento de Química da Universidade

Federal de São Carlos (DQ-UFSCar), essas três misturas de compostos foram obtidas e tiveram a sua atividade inibitória avaliada frente à GAPDH de *T. cruzi*.

Figura 3.3 - Mistura de componentes constituintes do óleo da castanha de caju.¹⁰

Os resultados indicaram que a mistura de compostos correspondente à fração de ácido anacárdico (compostos 2, 3 e 4 da Figura 3.3; o composto 1 foi acrescido a essa figura apenas por conveniência) apresentou atividade inibitória promissora no teste de inibição em dose única. Diante disso, essa mistura foi submetida ao processo de separação para isolamento dos respectivos constituintes. Entretanto, devido às diversas dificuldades inerentes ao processo, não foi possível o isolamento dos componentes puros desta mistura. Uma das tentativas realizadas paralelamente, foi a reação de hidrogenação desta mistura de ácidos anacárdicos levando a obtenção do composto 1 puro (Figura 3.3). Esse composto se apresentou como um inibidor bastante promissor da enzima GAPDH de *T. cruzi*, com valor de IC_{50} de 28 μM (Figura 3.4). Devido a sua potência inibitória e a diversidade química nova para a enzima-alvo, esta molécula é um bom modelo para futuros estudos de modificação molecular e SAR.

Figura 3.4 - Curvas de concentração x respostas do inibidor enzimático ácido anacárdico saturado (1).

3.4.2 - Compostos sintéticos como inibidores

Paralelamente aos estudos de inibição com produtos naturais, vários compostos de origem sintética também foram avaliados frente à enzima GAPDH de *T. cruzi*. Entre eles destacam-se: (i) derivados de nucleosídeos; e (ii) derivados bisfosfonatos.

Derivados de nucleosídeos

Uma série de derivados nucleosídeos foram sintetizados no grupo do Prof. Dr. Marcus Mandolesi de Sá do Departamento de Química da Universidade Federal de Santa Catarina (UFSC) e enviados para avaliação bioquímica frente à GAPDH de *T. cruzi* (Tabela 3.1, Figura 3.5).¹¹ Esses compostos foram sintetizados com o objetivo de explorar as diferenças estruturais entre as enzimas GAPDH de tripanosomatídeos e humana, principalmente na vizinhança do sítio de ligação do NAD^+ .¹²

Tabela 3.1 - Compostos derivados de nucleosídeos testados *in vitro* contra a enzima GAPDH.

Código	X	R ₁	R ₂	% Inibição (200 µM)	IC ₅₀ (µM)
13	O	H	Cl	nd	nd
14	O	H	NO ₂	nd	nd
15	O	H	Br	26	> 250
16	O	H	F	nd	nd
17	O	OH	Cl	60	36 ± 6
18	O	OH	Br	74	25 ± 4
19	O	OH	F	0	--
20	S	NH ₂	Br	35	> 250

nd: não foi possível determinar a atividade inibitória devido ao composto absorver em 340 nm.

Figura 3.5 - Curvas de concentração x respostas dos inibidores enzimáticos 17 e 18.

A análise dos dados de inibição dos derivados de nucleosídeos mostra que o emprego da técnica de bioisosterismo^{13,14} foi importante para o aumento da potência inibitória desta classe de moléculas. A substituição bioisostérica clássica do grupamento –NH₂ (20) por –OH (18) na posição R₁ contribuiu consideravelmente para o aumento da potência do composto 18 (membro

mais potente desta pequena série, com $IC_{50} = 25 \mu M$). A diminuição do volume estereo no substituinte da posição R_2 não alterou significativamente a atividade inibitória (**17**, $IC_{50} = 36 \mu M$). Apesar de apresentarem potência de ordem micromolar típica para a enzima-alvo, esses compostos podem ser considerados apenas inibidores modestos, uma vez que inibidores nucleosídeos de ordem sub-micromolar já foram reportados anteriormente.¹⁵⁻²⁰

Uma possível explicação para a potência modesta destes inibidores seria o fato de que os grupos OH' do anel ribose estão protegidos. Um requerimento estrutural importante para o reconhecimento molecular desta classe de inibidores é a presença de grupos doadores de ligação de hidrogênio nestas posições do anel ribose para o estabelecimento de interação com o resíduo altamente conservado Asp38 presente no sítio de ligação do NAD^+ da GAPDH de tripanosomatídeos.¹² Diante disso, procedimentos para a obtenção de compostos com essas propriedades estão em andamento no laboratório do Prof. Dr. Marcus Mandolesi de Sá.

Derivados bisfosfonatos

A segunda classe de moléculas sintéticas estudada é composta de derivados bisfosfonatos. Esses compostos fazem parte de uma coleção combinatória disponibilizada pelo Prof. Dr. Eric Oldfield da Universidade de Urbana Champaign, Ilinoís, EUA.

A promissora atividade antiparasitária dos compostos bisfosfonatos foi demonstrada em ensaios com culturas de *T. brucei*, *T. cruzi*, *L. donovani*, *T. gondii*, e *P. falciparum*.²¹ O mecanismo de ação proposto para esses compostos envolve a inibição da farnesilpirofosfato sintase, enzima da via do mevalonato, essencial para a biosíntese de esteróis e poliisoprenóides (moléculas de vital importância para a sobrevivência dos parasitas). Devido ao fato da GAPDH

possuir sítios de ligação de grupos fosfatos, vários compostos desta classe foram selecionados para avaliação da atividade inibitória frente à enzima GAPDH de *T. cruzi*.²²

Dentre os mais de 100 compostos avaliados por triagem bioquímica (teste de inibição concentração única de 200 μ M), merecem destaque apenas aqueles apresentados na Tabela 3.2 (**23**, **28**, **30** e **31**). Esses compostos tiveram a potência inibitória determinada, conforme mostra os gráficos de concentração x resposta da Figura 3.6, e podem ser divididos em três subclasses com características estruturais importantes para a inibição da enzima GAPDH.

A primeira subclasse contém o composto mais potente da série (**23**, Tabela 3.2). Observa-se que a modificação molecular do grupo CH_2 no composto **21** ($\text{IC}_{50} > 250 \mu\text{M}$) por O no composto **23** ($\text{IC}_{50} = 57 \mu\text{M}$) alterou substancialmente a propriedade biológica. A substituição bioisostérica clássica do grupo divalente adjacente ao anel benzílico levou a um aumento expressivo da potência inibitória (**23**). Este dado sugere que o átomo de oxigênio da posição *R* do inibidor **23** poderia participar em uma ligação de hidrogênio com um grupo do sítio ativo da enzima de funcionalidade complementar. O composto mais ativo da segunda subclasse, **28** ($\text{IC}_{50} = 100 \mu\text{M}$), possui o substituinte naftil, sugerindo que há um limite para o aumento do volume estéreo tolerado pelo sítio de ligação. A terceira subclasse apresenta cadeias alifáticas de quatro (**31**) ou seis (**32**) grupos CH_2 na posição *R*. O composto **31**, com quatro grupos CH_2 foi o mais potente, com valor de $\text{IC}_{50} = 117 \mu\text{M}$.

Tabela 3.2 - Série de bisfosfonatos como inibidores da enzima GAPDH de *T. cruzi*.

Composto	Substituinte	% Inibição (200 μ M)	IC ₅₀ (μ M)
21	R = CH ₂ , n = 3	29	> 250
22	R = CH ₂ , n = 4	19	> 250
23	R = O, n = 3	90	57 $\pm$ 15

24	R = <i>p</i> -NH ₂ -benzil	Inativo	---
25	R = benzil	Inativo	---
26	R = <i>m</i> -F-benzil	Inativo	---
27	R = <i>m</i> -NO ₂ -benzil	Inativo	---
28	R = naftil	76	100 $\pm$ 20
29	R = <i>m</i> -NH ₂ -benzil	32	---
30	R = bifenil	35	460 $\pm$ 50

31	R = <i>n</i>-butil	74	117 $\pm$ 10
32	R = <i>n</i> -hexil	15	> 200

Figura 3.6 - Curvas de concentração x respostas dos inibidores enzimáticos **23**, **28**, **30** e **31**.

Todos os inibidores da GAPDH de *T. cruzi* identificados neste trabalho (compostos com IC_{50} determinada) foram submetidos a inúmeros ensaios de cristalização e difração de raios X. De fato, milhares de condições de cristalização foram avaliadas e os dados cristalográficos de centenas de cristais da enzima obtidos na presença dos inibidores foram coletados. Contudo, devido a baixa solubilidade dos inibidores em solução aquosa, não foi possível obter de complexos enzima-inibidor.

3.5 - Referências

1. COPELAND, R. A. *Enzymes – practical introduction to structure, mechanism and data analysis*, New York: Wiley VHC, 2000.
2. NELSON, D. N.; MICHAEL M. C. *Lehninger Principles of Biochemistry*, New York: Worth Publishers, 2004.
3. HANNAERT, V.; OPPERDOES, F. R.; MICHELS, P. A. M. Glycosomal glyceraldehyde-3-phosphate dehydrogenase of *Trypanosoma brucei* and *Trypanosoma cruzi*: expression in *Escherichia coli*, purification, and characterization of the enzymes. *Protein Expr. Purif.*, v. 6, p. 244-250, 1995.
4. SOUZA, D. H. F. et al. *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design. *FEBS Lett.*, v. 424, p. 131-135, 1998.
5. OPREA, T. I. et al. Lead-like, drug-like or "Pub-like": how different are they? *J. Comput. Aided Mol. Des.*, v. 21, p. 113-119, 2007.
6. OPREA, T. I. Current trends in lead discovery: are we looking for the appropriate properties? *J. Comput. Aided Mol. Des.*, v. 16, p. 325-334, 2002.
7. WIGGERS, H. J. et al. Effects of organic solvents on the enzyme activity of *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase in calorimetric assays. *Anal. Biochem.*, v. 370, p. 107-114, 2007.
8. HOSTETMANN, K.; QUERÓZ, E. F.; VIEIRA, P. C. *Princípios ativos de plantas superiores*. Série de textos da escola de verão em química. São Carlos: EdUFSCar, 2003.
9. NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.*, v. 70, p. 461-477, 2007.
10. PARAMASHIVAPPA, R. et al. Novel method for isolation of major phenolic constituents from cashew (*Anacardium occidentale* L.) nut shell liquid. *J. Agric. Food. Chem.*, v. 49, p. 2548-2551, 2001.
11. SA, M. M. et al. Synthesis of acylated nucleosides and ribonic-1,4-lactones as inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH), *Arkivoc*, p. 112-124, part 8, 2002.
12. SURESH, S. et al. Conformational changes in *Leishmania mexicana* glyceraldehyde-3-phosphate dehydrogenase induced by designed inhibitors. *J. Mol. Biol.*, v. 309, p. 423-435, 2001.

13. LIMA, L. M.; BARREIRO, E. J. Bioisosterism: a useful strategy for molecular modification and drug design. *Curr. Med. Chem.* v. 12, p. 23-49, 2005.
14. BARREIRO, E. J.; FRAGA, C. A. M. *Química medicinal: as bases moleculares da ação dos fármacos*. 2ª edição, Porto Alegre: Artmed Editora, 2008.
15. ARONOV, A. M. et al. Structure-based design of submicromolar, biologically active inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. *Proc. Natl. Acad. Sci. USA*, v. 96, p. 4273-4278, 1999.
16. ARONOV, A. M. et al. Selective tight binding Inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase via Structure-based drug design. *J. Med. Chem.*, v. 41, p. 4790-4799, 1998.
17. BRESSI, J. C. et al. Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of trypanosomatidae via structure-based drug design. *J. Med. Chem.*, v. 44, p. 2080-2093, 2001.
18. KENNEDY, K. J.; BRESSI, J. C.; GELB, M. H. A disubstituted NAD^+ analogue is a nanomolar inhibitor of trypanosomal glyceraldehyde-3-phosphate dehydrogenase. *Bioorg. Med. Chem. Lett.*, v.11, p. 95-98, 2001.
19. VAN CALENBERGH, S. et al. Synthesis and structure-activity relationships of analogs of 2'-deoxy-2'-(3-methoxybenzamido)adenosine, a selective inhibitor of trypanosomal glycosomal glyceraldehyde-3-phosphate dehydrogenase. *J. Med. Chem.*, v. 38, p. 3838-3849, 1995.
20. VERLINDE, C. L. M. J. et al. Selective inhibition of trypanosomal glyceraldehyde-3-phosphate dehydrogenase by protein structure-based design: toward new drugs for the treatment of sleeping sickness. *J. Med. Chem.*, v. 37, p. 3605-3613, 1994.
21. MARTIN, M. B. et al. Bisphosphonates inhibit the growth of *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania donovani*, *Toxoplasma gondii*, and *Plasmodium falciparum*: a potential route to chemotherapy. *J. Med. Chem.*, v. 44, p. 909-916, 2001.
22. CASTILHO, M. S. et al. Evidence for the two phosphate binding sites of an analogue of the thioacyl intermediate for the *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase-catalyzed reaction, from its crystal structure. *Biochemistry*, v. 42, p. 7143-7151, 2003.

Capítulo 4

Estudos Cinéticos e Cristalográficos da GAPDH de T. cruzi

Nesse capítulo é apresentada uma introdução sobre os fenômenos de alosterismo e cooperatividade de proteínas oligoméricas. Em seguida, são descritos os métodos utilizados e os resultados obtidos dos estudos de cinética e cooperatividade da enzima GAPDH de *T. cruzi*. Na seção seguinte, são apresentados os resultados dos estudos estruturais e a comparação entre quatro estruturas distintas da GAPDH que foram cristalizadas com quantidades crescentes de NAD^+ ligado. Os dados cristalográficos corroboraram os dados cinéticos, fornecendo informações estruturais importantes na compreensão das bases moleculares envolvidas no mecanismo de cooperatividade da GAPDH.

4 - ESTUDOS CINÉTICOS E CRISTALOGRAFICOS DA GAPDH DE *T. cruzi*

4.1 - Alostерismo e mudanças conformacionais

Para exercerem suas funções biológicas, algumas proteínas organizam-se em complexos oligoméricos. As subunidades formadoras do oligômero são em geral compostas de cópias da mesma (ou distinta) cadeia polipeptídica, podendo cada subunidade conter um sítio de ligação. No caso mais simples, os sítios ativos de cada subunidade funcionam independentemente, como se representassem unidades catalíticas separadas. Em outros casos, a interação de um ligante no sítio ativo de uma subunidade pode aumentar ou diminuir a afinidade do sítio de ligação em outra subunidade, dando origem ao fenômeno de cooperatividade, positiva ou negativa, respectivamente.¹

Para ocorrer uma interação cooperativa entre dois sítios distantes numa estrutura oligomérica, a interação do ligante em um sítio ativo deve induzir uma mudança conformacional na proteína que é transmitida, via cadeia peptídica, para o sítio ativo localizado na subunidade adjacente.^{1,2} O fenômeno de transmissão de informações via mudanças conformacionais em proteínas é conhecido como alosterismo. Portanto, uma proteína alostérica é aquela na qual a ligação de uma molécula no sítio de ligação produz alterações estruturais em sua estrutura terciária ou quaternária, induzindo mudanças nas propriedades de interação de outros sítios de ligação por uma segunda molécula (do mesmo ligante), no processo de cooperatividade.²

Os moduladores (ligantes) das proteínas alostéricas podem ser classificados em inibidores ou ativadores. Quando o modulador é o próprio ligante natural, tem-se o efeito alostérico de cooperatividade homotrópica. Por outro lado, se o modulador é um ligante diferente do ligante natural, o efeito alostérico é conhecido como cooperatividade heterotrópica.^{1,2}

Dois modelos teóricos foram propostos para explicar o comportamento cooperativo das proteínas. No primeiro modelo de Monod-Wyman-Changeux (MWC),³ assume-se que cada subunidade da proteína pode existir em no mínimo duas conformações. Essas conformações estão em equilíbrio dinâmico, podendo o ligante interagir com ambas as conformações, entretanto, podendo apresentar afinidade de ligação diferente. A interação sucessiva do ligante no estado de menor afinidade (mais estável na ausência do ligante) faz com que o equilíbrio seja deslocado para a conformação de maior afinidade (Figura 4.1A). Já no segundo modelo de Koshland-Némethy-Filmer (KNF),⁴ assume-se que a interação do ligante induz mudanças conformacionais na subunidade na qual se liga, influenciando a afinidade da proteína para a interação com a segunda molécula de ligante, e assim sucessivamente (Figura 4.1B). A Figura 4.1 ilustra o esquema simplificado dos modelos MWC e KNF.

Figura 4.1 - Esquema dos modelos (A) MWC e (B) KNF na descrição do fenômeno de cooperatividade entre subunidades de proteínas oligoméricas. L = ligante.

4.2 - Análise quantitativa da cooperatividade entre sítios de ligação

O grau de cooperatividade entre os sítios de ligação de uma proteína pode ser calculado utilizando-se a equação de Hill (Equação 4.6).⁵ O valor de n na equação de Hill corresponde ao grau de cooperatividade entre os sítios de ligação. A influência da cooperatividade entre os sítios de uma proteína pode ser quantificada de acordo com as Equações 4.1 e 4.2:

$$K = \frac{[E][L]^n}{[ES_n]} \quad (4.2)$$

Rearranjando a Equação 4.2 e substituindo-se o termo $[ES_n]$ e $[E]$ pelas Equações 4.3 e 4.4, respectivamente (sendo $[E]_0$ a quantidade de enzima total, A a fração de proteína ligada e $1 - A$ a fração de proteína livre):

$$A = \frac{[ES_n]}{[E]_0} \quad (4.3)$$

$$1 - A = \frac{[E]}{[E]_0} \quad (4.4)$$

Obtém-se a razão entre a quantidade de proteína ligada e proteína livre ou grau de saturação (Equação 4.5):

$$\frac{A}{1 - A} = \frac{[L]^n}{K} \quad (4.5)$$

Aplicando-se a função logarítmica em ambos os lados da Equação 4.5, tem-se:

$$\log \frac{A}{1 - A} = n \log [L] - \log K \quad (4.6)$$

A Equação 4.6 é conhecida como Equação de Hill e o valor da inclinação da reta é a constante de Hill (n_H), a qual indica o grau de cooperatividade entre os sítios de uma proteína. Assim, quando n_H é igual a 1, a interação entre o ligante e a proteína não é cooperativa, fenômeno que ocorre até mesmo em proteínas oligoméricas nas quais os sítios de ligação não se comunicam. Se o valor de n_H é maior do que 1, ocorre cooperatividade positiva entre os sítios de ligação, ou seja, a interação de um ligante com o primeiro sítio ativo favorece a interação do segundo ligante (maior afinidade), e assim sucessivamente. Finalmente, se o valor de n_H é menor do que 1, ocorre cooperatividade negativa entre os sítios de ligação, ou seja, a interação de um ligante com o primeiro sítio ativo dificulta a interação da segunda molécula de ligante (menor afinidade), e assim sucessivamente.^{1,2,6}

4.3 – A enzima GAPDH e a fenômeno de cooperatividade entre os sítios de ligação

A enzima GAPDH tem sido objeto importante de pesquisa na investigação do mecanismo de cooperatividade.⁷⁻¹⁴ Apesar de várias estruturas cristalográficas de GAPDH (e.g., complexos com ligantes naturais e sintéticos; formas apo e holo) serem conhecidas, as bases moleculares do comportamento cooperativo ainda não foram completamente elucidadas. A GAPDH apresenta comportamento cooperativo diferenciado em relação à interação com cofator dependendo do organismo. Por exemplo, uma cooperatividade positiva foi observada em GAPDH de leveduras,¹⁰ enquanto que uma cooperatividade negativa foi observada em GAPDH de músculo de coelho,¹⁵ esturjão,¹⁶ músculo da cauda de lagostas,^{17,18} *Escherichia coli*,¹⁹ e *Bacillus stearothermophilus*.²⁰

No presente trabalho, estudos cinéticos foram realizados com o objetivo de determinar o comportamento cooperativo da GAPDH de *T. cruzi*.

4.4 - Estudos cinéticos da GAPDH de *T. cruzi*

4.4.1 - Determinação das constantes cinéticas da GAPDH de *T. cruzi*

A enzima utilizada para os estudos cinéticos foi preparada de acordo com o procedimento descrito na seção 3.2. O ensaio cinético realizado para a determinação do comportamento cooperativo da GAPDH de *T. cruzi* é similar ao procedimento descrito na seção 3.3. Os parâmetros cinéticos da GAPDH, (i) K_M , que é a constante de Michaelis-Menten, definida como a concentração de substrato responsável por metade da velocidade máxima de reação; e (ii) V_{max} , que é a velocidade máxima de reação obtida em condições de saturação de substrato; foram obtidas para o cofator e para o substrato sob condições de estado estacionário (do inglês, *steady state*).

Nos ensaios de determinação dos parâmetros cinéticos para o cofator, foram utilizadas concentrações crescentes de NAD^+ no intervalo de 8-2.000 μM , mantendo-se constante a concentração do substrato (G3P, 400 μM). No caso do substrato, foram utilizadas concentrações crescentes de G3P no intervalo de 15-2.000 μM , mantendo-se constante a concentração do cofator (NAD^+ , 300 μM). Os dados experimentais coletados foram analisados com o módulo de cinética enzimática do programa SigmaPlot versão 7.0 e os parâmetros cinéticos calculados através de análise de regressão não-linear de melhor ajuste.²

A enzima GAPDH de *T. cruzi* segue a cinética de Michaelis-Menten, ou seja, em concentrações baixas de G3P e NAD^+ ($[S] \ll K_M$) a velocidade de reação (v_0) é diretamente proporcional às suas concentrações. Em concentrações altas de G3P e NAD^+ ($[S] \gg K_M$), v_0 se aproxima do seu valor máximo, sendo independente de suas concentrações (Figuras 4.2 e 4.3). Os valores dos parâmetros cinéticos para o cofator e o substrato são apresentados na Tabela 4.1.

Figura 4.2 - Influência da concentração do cofator NAD⁺ na reação catalisada pela enzima GAPDH de *T. cruzi*.

Figura 4.3 - Influência da concentração do substrato G3P na reação catalisada pela enzima GAPDH de *T. cruzi*.

Tabela 4.1 - Parâmetros cinéticos calculados utilizando-se o método de regressão não-linear.

Substrato	K_M (μM)	$V_{\max}$ ($\mu\text{M}/\text{min}$)	R^2
NAD⁺	232 ± 32	$0,12 \pm 0,01$	0,99
G3P	300 ± 45	$0,20 \pm 0,01$	0,99

O método utilizado para a determinação dos parâmetros cinéticos apresentou curvas de ajuste bastante significativas para os dados coletados experimentalmente, conforme indicado pelos valores de R^2 (Tabela 4.1). Além disso, os valores dos parâmetros cinéticos calculados para o NAD⁺ e G3P são similares aos reportados na literatura.^{21,22}

4.4.2 - Determinação do comportamento cooperativo da GAPDH de *T. cruzi*

Para a avaliação do comportamento cooperativo da GAPDH de *T. cruzi* foram empregados os valores de velocidade e K_M obtidos nos estudos cinéticos (seção 4.4). O grau de saturação da enzima (α) foi calculado de acordo com a equação 4.7 e o formalismo de Hill²³ foi utilizado para a determinação da cooperatividade entre os sítios da GAPDH em relação à interação com o NAD⁺ e o G3P (sendo L a concentração do ligante e n_H a constante de Hill, ou grau de cooperatividade).

$$\alpha = \frac{V_i - V_0}{V_{\max} - V_0} \quad (4.7)$$

$$\alpha = \frac{\frac{L^{n_H}}{K_m^{n_H}}}{1 + \frac{L^{n_H}}{K_m^{n_H}}} \quad (4.8)$$

Portanto, combinando-se as equações 4.7 e 4.8, obtém-se:

$$V_i = V_0 + (V_{\max} - V_0) \times \frac{\frac{L^{n_H}}{K_m^{n_H}}}{1 + \frac{L^{n_H}}{K_m^{n_H}}} \quad (4.9)$$

A substituição dos valores obtidos das constantes cinéticas do NAD^+ e G3P na equação 4.9 permitiu o cálculo da constante de Hill para a interação com o substrato e o cofator. As curvas de ajustes e os valores de n_H estão apresentados nas Figuras 4.4 e 4.5. Os valores da constante de Hill foram determinados através do método de regressão não-linear dos dados ajustados à equação 4.9 (Figuras 4.4A e 4.5A) e através do valor da inclinação da reta (Figuras 4.4B e 4.5B).

Os valores obtidos da constante de Hill ($n_H \sim 1$) indicam que a enzima GAPDH de *T. cruzi*, diferentemente das GAPDHs de outros organismos, apresenta comportamento não-cooperativo para a interação com o NAD^+ ou G3P. Resultado similar foi observado para o mutante tetramérico S48G da GAPDH de *B. stearothermophilus*.⁸ Entretanto, as bases moleculares envolvidas neste comportamento não são descritas. Diante disso, no presente trabalho, procurou-se obter estruturas cristalográficas da GAPDH de *T. cruzi* em complexo com diferentes quantidades de NAD^+ para a melhor compreensão dos determinantes estruturais envolvidos neste mecanismo bioquímico.

Figura 4.4 - (A) Curva de ajuste dos dados cinéticos para a interação do NAD^+ em escala logarítmica para determinação do valor da constante de Hill (n_H). (B) Gráfico de Hill correspondente à saturação dos sítios da GAPDH com NAD^+ .

Figura 4.5 - (A) Curva de ajuste dos dados cinéticos para a interação do G3P em escala logarítmica para determinação do valor da constante de Hill (n_H). (B) Gráfico de Hill correspondente à saturação dos sítios da GAPDH com G3P.

4.5 - Experimentos de cristalização e difração de raios X com a GAPDH de *T. cruzi*

4.5.1 - Cristalização

A cristalização de macromoléculas biológicas é uma técnica baseada na tentativa e erro por se tratar de um processo complexo multiparamétrico. Parâmetros de caráter físico (e.g., temperatura, pressão, superfície da molécula, tempo, etc.), químico (e.g., pH, agente precipitante, força iônica, grau de supersaturação, pureza da amostra, estado de agregação, ponto isoelétrico, estabilidade, etc.) interferem diretamente na formação de um cristal, de maneira que os diversos métodos utilizados tentam variar este espaço multiparamétrico com o objetivo de examinar os efeitos de combinações dessas variáveis. Esses métodos são geralmente aplicáveis à maioria das proteínas, DNAs, RNAs e complexos multimoleculares.²⁴

O processo de formação de cristais ocorre a partir de uma solução supersaturada de proteína, sendo a velocidade com que se atinge esse estado essencial para a formação de cristais, microcristais ou precipitado amorfo. Os cristais de proteínas crescem geralmente em condições toleráveis de temperatura, força iônica e com pequenas faixas de variação pH.²⁴

O preparo de cristais de proteínas de boa qualidade para difração de raios X requer uma quantidade considerável de proteína pura. Seja a proteína-alvo obtida diretamente de um organismo, ou produzida por expressão em sistema heterólogo, a mesma deve passar por um processo de purificação, visando a obtenção de uma amostra homogênea.²⁵

Existem diversos métodos de cristalização de proteínas, os quais são utilizados de acordo com o tipo da amostra. O método de difusão de vapor baseia-se no equilíbrio entre duas soluções através da fase de vapor num meio fechado. A solução menos concentrada perde seu solvente volátil até que os potenciais químicos das duas soluções se igualem. Para se poder controlar a

concentração final da solução de proteína, realiza-se o experimento de difusão de vapor com um volume pequeno de proteína contra um volume grande de solução precipitante. Assim, uma gota de solução da proteína a ser cristalizada é adicionada à solução tampão contendo agentes precipitantes e aditivos. Em seguida, esta gota é equilibrada contra um reservatório contendo solução de agentes precipitantes a uma concentração maior que a da gota com proteína.²⁵

O método de difusão de vapor pode ser conduzido de duas maneiras principais: (i) gota suspensa (do inglês, *hanging drop*), no qual a gota contendo a proteína de interesse é colocada sobre uma lamínula de vidro siliconizada e, posteriormente, vedada com o auxílio de graxa especial na parte superior do poço, de forma que a gota fique interna ao reservatório (Figura 4.6A); e (ii) gota assentada (do inglês *sitting drop*), no qual a solução contendo a proteína é colocada sobre um suporte fixado no centro do poço, o qual é posteriormente fechado com lamínula de vidro siliconizada e vedado com o auxílio de graxa especial (Figura 4.6B).

Figura 4.6 - Esquema representativo dos métodos de difusão de vapor em gota suspensa (A) e gota assentada (B) usados para a obtenção de cristais de macromoléculas com o auxílio de agentes precipitantes (M, macromolécula; AP, agente precipitante).

4.5.2 - Purificação e cristalização

A proteína GAPDH de *T. cruzi* foi purificada de acordo com a metodologia descrita na seção 3.2. Para a obtenção de cristais de GAPDH com diferentes quantidade de cofator nos sítios de ligação, primeiramente submeteu-se a enzima pura (com aproximadamente 60 mM de NAD^+ proveniente do processo de purificação) a diálise contra uma solução tampão 100 mM de TEA/HCl, pH 7,4, para a obtenção da enzima na forma apo (sem NAD^+). Posteriormente, adicionou-se a solução de proteína apo soluções de NAD^+ em concentrações crescentes, de maneira a se obter a relação estequiométrica GAPDH: NAD^+ de 1:1, 1:2, 1:3 e 1:4. Assim, os cristais obtidos deveriam conter, respectivamente, uma, duas, três e quatro moléculas de NAD^+ por unidade de tetrâmero. Os ensaios de cristalização foram realizados com solução de proteína pura a 7 mg/mL e solução de cristalização contendo solução tampão Bis-Tris, 0,1 M, pH 7,3 e agentes precipitantes acetato de cálcio, 0,1 M, e, polietileno glicol 8000, 18%. O método de gota suspensa foi empregado para a cristalização da proteína.

Quatro estruturas cristalográficas contendo quantidades crescentes de NAD^+ por tetrâmero foram obtidas. Uma estrutura na forma apo (**apo**, sem NAD^+ ligado), uma estrutura contendo apenas uma molécula de NAD^+ ligada ao tetrâmero (**1NAD**), uma estrutura contendo duas moléculas de NAD^+ por tetrâmero (**2NAD**) e uma estrutura na forma holo (**holo**, quatro moléculas de NAD^+ por tetrâmero). Somente duas estruturas cristalográficas da GAPDH com quantidades diferentes de NAD^+ são descritas na literatura, uma com a GAPDH de *B. stearothermophilus* (3 Å de resolução) contendo uma molécula de NAD^+ ligada por tetrâmero¹⁴ e outra com a GAPDH de músculo de coelho (2,4 Å de resolução) contendo duas moléculas de NAD^+ ligada por tetrâmero.²⁶ Portanto, essa é a primeira vez que se observa os estados intermediários de interação entre o cofator e a GAPDH de um mesmo organismo. A descrição

detalhada dos métodos de resolução das estruturas, análise estrutural e implicações para o comportamento não-cooperativo são discutidas nas próximas seções.

4.5.3 - Coleta e processamento dos dados de difração de raios X

Todas as coletas de dados das imagens de difração referente aos cristais obtidos foram realizadas à temperatura de 100 K com o auxílio de solução crio protetora (10-15% de glicerol adicionado à solução de cristalização). Os cristais foram rapidamente congelados num fluxo contínuo de nitrogênio a 100 K. Os cristais congelados foram armazenados em *deware* com nitrogênio líquido e transportados até o local da coleta de dados.

A coleta de dados das imagens de difração de raios X foi realizada na linha de Cristalografia de Macromoléculas (MX-1) no Laboratório Nacional de Luz de Síncrotron (LNLS).²⁷ O método utilizado para as coletas de dados foi de rotação-oscilação,²⁸ onde o cristal sofre pequenas e sucessivas oscilações em torno do eixo perpendicular à direção do feixe de raios X incidentes. A indexação e integração dos dados de difração cristalográficos, para todos os conjuntos, foram realizadas com o programa MOSFLM.²⁹ O escalonamento e a promediação dos dados foram realizados com o programa SCALA,³⁰ que inclui o protocolo TRUNCATE,³¹ ambos fazem parte do pacote de programas cristalográficos CCP4 (Collaborative Computational Project Number 4). O programa SCALA gera um arquivo final que contém os índices de Miller (*hkl*), os módulos dos fatores de estrutura, $|F(hkl)|$, e as respectivas incertezas ($\sigma(F)$) para cada reflexão, calculadas a partir das intensidades. Nesta etapa do processo, um conjunto de 5% do total de reflexões independentes foi separado para ser usado posteriormente como parâmetro de monitoramento da qualidade do refinamento (R_{free}).

O método de substituição molecular foi utilizado para a estimativa das fases iniciais. A estrutura da GAPDH de *T. cruzi* (PDB, 1K3T)³² sem as moléculas de água e cofator, na faixa de resolução de 15-3 Å foi utilizada como modelo de busca. O programa utilizado para realizar o método de substituição molecular foi o MOLREP,³³ que também faz parte do pacote de programas cristalográficos CCP4.

Após a obtenção das fases iniciais, deu-se início ao processo de refinamento. A primeira estratégia utilizada para o refinamento foi a realização de uma rotina de *simulated annealing*, implementado no pacote CNS,³⁴ onde foi simulado a elevação da temperatura até 10.000 K com passos sucessivos de resfriamento lento de 25 K. Essa metodologia foi empregada com o objetivo de superar mínimos locais que pudessem prejudicar o refinamento dos modelos. A análise preliminar dos mapas iniciais de densidade eletrônica ($2F_{obs}-F_{calc}$ e $F_{obs}-F_{calc}$) dos conjuntos de dados coletados indicava a obtenção de quatro estruturas distintas contendo quantidades crescentes de NAD⁺ ligados por tetrâmero. A reconstrução manual e o refinamento dos modelos macromoleculares no espaço real foram realizados com o programa COOT.³⁵ Os ciclos de refinamento no espaço recíproco foram realizados com o protocolo de refinamento implementado no programa CNS utilizando restrições cristalográficas e estereoquímicas. Em nenhum momento do processo de refinamento do modelo foi utilizada simetria não-cristalográfica (NCS). Ao longo do processo de refinamento, moléculas de água foram adicionadas automaticamente através do protocolo *water peak* também implementado no programa CNS, sendo posteriormente verificadas visualmente. O refinamento foi considerado terminado uma vez que novas alterações realizadas no modelo não resultavam mais na diminuição dos valores de R_{factor} e R_{free} .

A análise final dos modelos refinados revelou a obtenção de (i) uma estrutura a 2,15 Å de resolução, onde não foram observadas as densidades eletrônicas característica da molécula do NAD⁺ em nenhuma das subunidades (conformação **apo**); (ii) uma estrutura a 2,40 Å de

resolução, com apenas uma molécula de NAD^+ ligada por tetrâmero na subunidade D (conformação **1NAD**); (iii) uma estrutura a 2,23 Å de resolução, com duas moléculas de NAD^+ por unidade de tetrâmero, uma na subunidade A e outra na subunidade D (conformação **2NAD**); e (iv) uma estrutura a 2,15 Å de resolução, com os quatro sítios de ligação ocupados com NAD^+ (conformação **holo**). A Figura 4.7 apresenta os mapas de densidade eletrônica finais nos sítios de ligação das quatro estruturas obtidas com diferentes quantidades de NAD^+ . Os dados cristalográficos e as estatísticas dos refinamentos referentes aos conjuntos de dados coletados estão apresentados na Tabela 4.2.

Tabela 4.2 - Dados cristalográficos e estatísticas do refinamento das diferentes formas de GAPDH de *T. cruzi*.

	Apo	1NAD	2NAD	Holo
Dados cristalográficos				
Faixa de resolução (Å)	20,0 - 2,15 (2,25 - 2,15)*	20,0 - 2,40 (2,50 - 2,50)*	20,0 - 2,23 (2,35 - 2,23)*	20,0 - 2,15 (2,25 - 2,15)*
Grupo Espacial	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Parâmetros de rede (Å)	a=81,30	a=79,88	a=90,68	a=78,85
	b=83,47	b=83,17	b=83,63	b=84,93
	c=103,98	c=102,95	c=99,64	c=106,16
	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$
	$\beta=96,17^\circ$	$\beta=96,03^\circ$	$\beta=98,56^\circ$	$\beta=95,70^\circ$
Nº total de imagens	326	202	267	181
$\Delta\phi$ (°)	1	1	1	1
Nº total de reflexões	251.539 (34.376)*	198.013 (30.198)*	288.359 (40.142)*	230.044 (32.549)*
Nº de reflexões únicas	70.242 (9.959)*	52.568 (7.632)*	72.015 (10.417)*	74.208 (10.568)*
Completeza (%)	99,2 (99,2)*	99,8 (99,8)*	98,5 (99,3)*	97,7 (87,4)
R _{merge} (%)	7,0 (17,2)*	10,7 (35,2)*	7,5 (45,3)	8,9 (12,6)
$\langle I \rangle / \langle \sigma(I) \rangle$	8,8 (4,2)*	6,3 (2,2)*	11,2 (2,0)	7,4 (2,0)*
Multiplicidade	3,6 (3,5)*	3,8 (4,0)*	4,0 (3,9)*	3,1 (3,0)*
Refinamento				
Nº de aa / monômero	359	359	359	359
Moléculas de água	852	658	817	1091
R _{factor} (%)	19,6	22,0	20,4	17,9
R _{free} (%)	23,8	26,4	25,5	22,4
rmsd - ângulos de ligação (°)	1,1	0,925	1,5	2,1
rmsd - comprimento das ligações (Å)	0,006	0,006	0,007	0,016

*Valores obtidos para a camada mais externa de resolução.

continuação

Figura 4.7 - Mapa de densidade eletrônica $2F_{\text{obs}} - F_{\text{calc}}$ (contorno de 1.0σ) no sítio de ligação do NAD^+ nas subunidades A, B, C e D das estruturas **apo** (ciano), **holo** (magenta), **1NAD** (cinza) e **2NAD** (laranja). Em destaque os resíduos que interagem com o cofator através de ligação de hidrogênio. Os elementos de estrutura secundária estão representados como modelo de fitas e as moléculas de água como esferas.

4.5.4 - Qualidade das estruturas cristalográficas

Os mapas de densidade eletrônica $2F_{\text{obs}} - F_{\text{calc}}$ finais calculados para as quatro formas da GAPDH de *T. cruzi* apresentam boa qualidade, sendo possível observar com detalhes os átomos das cadeias principal e lateral dos aminoácidos. Somente seis resíduos (aa. 69-74) posicionados em uma alça flexível no N-terminal de cada estrutura apresentam densidade eletrônica descontínua. Essa alça está localizada na superfície da GAPDH e está exposta ao solvente. Apesar da descontinuidade da densidade eletrônica, a posição dos átomos da cadeia principal destes resíduos pode ser identificada.

A estereoquímica das estruturas é bastante satisfatória, com mais de 99% dos resíduos apresentando ângulos torsionais nas regiões favoráveis do diagrama de Ramachandran³⁶ (Figura 4.8). Somente a Val255 em todas as estruturas está em região desfavorável do diagrama. Este resíduo está localizado em uma alça flexível entre duas folhas- β consecutivas. Essa conformação incomum é observada em outras estruturas de GAPDH e esta relacionada com posicionamento correto do resíduo Cys166 em relação ao anel nicotinamida do cofator durante a catálise.^{32,37-42}

Os valores de temperatura médios (fator B) para os 359 resíduos de cada subunidade das quatro estruturas estão de acordo com a qualidade dos dados medidos (Tabela 4.3). A análise dos fatores de temperatura médios revelou que, para as estruturas na mesma faixa de resolução (2,15 Å), estes valores são significativamente menores para a forma **holo** quando comparado aos valores obtido para a forma **apo**. Resultado similar é observado nas estruturas **1NAD** e **2NAD**, ou seja, aquelas subunidades que contém a molécula do NAD^+ ligada apresentam fatores de temperatura ligeiramente menores. Esses dados sugerem que a interação com o cofator poderia estabilizar a estrutura da GAPDH. De fato, estudos espectroscópicos demonstraram que há uma diminuição significativa da dinâmica da GAPDH de músculo de coelho em função do aumento da quantidade de NAD^+ ligado.⁴³

Figura 4.8 - Diagrama de Ramachandran para as diferentes formas de GAPDH de *T. cruzi*.

Tabela 4.3 - Fatores de temperatura médio dos 359 resíduos das quatro formas da GAPDH de *T. cruzi*.

Estrutura	Resolução (Å)	Subunidade (Å)			
		A	B	C	D
apo	2,15	37,6	27,8	26,9	29,8
1NAD	2,40	28,3	31,0	34,4	26,8
2NAD	2,23	28,7	32,8	31,0	29,6
holo	2,15	18,6	22,1	23,8	18,1

4.5.5 - Análise das estruturas de GAPDH de *T. cruzi*

As quatro formas da GAPDH pertencem ao sistema cristalino monoclinico, grupo espacial $P2_1$ com um tetrâmero por unidade assimétrica. As subunidades das estruturas da GAPDH estão relacionadas entre si por um eixo de simetria não cristalográfico de ordem 222. Cada subunidade é composta por dois domínios: domínio de ligação do NAD^+ (resíduos 1-150 e 331-359) e um domínio catalítico (resíduos 151-330). As moléculas de NAD^+ nas quatro formas diferentes de GAPDH de *T. cruzi* adotam uma conformação similar àquelas obtidas em GAPDH de outros organismos (Figura 4.7).^{18,44-46}

Os valores de rmsd (raiz quadrada do desvio médio, do inglês, *root mean square deviation*) da comparação entre as subunidades que compõem o tetrâmero de cada estrutura cristalográfica estão apresentados na Tabela 4.4. A análise dos dados da Tabela 4.4 revelou que as subunidades que compõem o tetrâmero das estruturas **apo** e **holo** são bastante similares entre si (comparação **1** e **2**). Em contraste, as comparações entre as subunidades pertencentes à estrutura **1NAD** diferem entre si quando comparadas com o monômero que apresenta a molécula do NAD^+ ligada (comparação **3**, A/D; B/D e C/D). Comportamento similar é observado para a estrutura **2NAD**, ou seja, as comparações entre as subunidades que possuem o NAD^+ ligado com aquelas que não possuem a molécula do cofator apresentam valores de rmsd significativamente maiores (comparação **4**, A/B; A/C; B/D e C/D).

Tabela 4.4 - Comparação entre as subunidades do tetrâmero das diferentes formas de GAPDH de *T. cruzi* (valores de rmsd em Å). Os resíduos 3-356 foram utilizados no protocolo de sobreposição.

Comparação	Subunidades					
	A / B	A / C	A / D	B / C	B / D	C / D
(1) apo / apo	0,32	0,50	0,34	0,38	0,31	0,50
(2) holo / holo	0,24	0,23	0,24	0,16	0,20	0,23
(3) 1NAD / 1NAD	0,51	0,52	0,67	0,50	0,74	0,67
(4) 2NAD / 2NAD	0,75	0,72	0,33	0,43	0,66	0,63

Comparação das formas apo e holo

As estruturas das formas **apo** e **holo** da GAPDH de *T. cruzi* foram determinadas na mesma resolução (2,15 Å). A sobreposição dos C α dos tetrâmeros resultou em um valor médio de rmsd de 0,70 Å (Tabela 4.5), enquanto que o valor de rmsd para a sobreposição os domínios catalíticos é de 0,54 Å e para o domínio de ligação do NAD⁺ de 1,14 Å. Estes dados indicam a ocorrência de uma mudança conformacional na estrutura da enzima devido à movimentação do domínio de ligação do NAD⁺. A mudança conformacional induzida pela ligação do NAD⁺ foi primeiramente observada na GAPDH de *B. stearothermophilus*^{14,38} e posteriormente na GAPDH de *P. versicolor*.¹⁸

Baseado nos resultados obtidos, o menor valor de rmsd dos C α para a transição da forma **apo** para a forma **holo** é de 0,61 Å (comparação 1, Tabela 4.5). Diante disso, com a finalidade de comparar as diferentes formas da GAPDH, esse número foi definido como valor de referência para a comparação e classificação das subunidades das demais formas. Assim, subunidades equivalentes que apresentem valor de rmsd menor do que o valor de referência são consideradas

subunidades similares ($\text{rmsd} < 0,61 \text{ \AA}$), da mesma forma, subunidades equivalentes que apresentem valor de rmsd maior ou igual ao valor de referência são consideradas diferentes ($\geq 0,61 \text{ \AA}$). A Tabela 4.5 está colorida de acordo com o critério de similaridade descrito acima, mostrando as subunidades similares em verde e as diferentes em vermelho.

Tabela 4.5 - Comparação entre as subunidades das diferentes formas de GAPDH de *T. cruzi* (valores de rmsd em $\text{\AA}$).

Comparação*	Subunidades			
	A / A	B / B	C / C	D / D
(1) apo / holo	0,73	0,65	0,61	0,84
(2) apo / 1NAD	0,44	0,36	0,39	0,80
(3) apo / 2NAD	0,79	0,36	0,26	0,78
(4) holo / 1NAD	0,69	0,78	0,66	0,44
(5) holo / 2NAD	0,37	0,61	0,62	0,40
(6) 1nad / 2NAD	0,72	0,45	0,38	0,34

*A Tabela está colorida de acordo com o grau de similaridade, mostrando as subunidades similares em verde ($< 0,61 \text{ \AA}$) e as diferentes em vermelho ($\geq 0,61 \text{ \AA}$). Os resíduos 3-356 foram utilizados no protocolo de sobreposição.

A mudança conformacional induzida pela ligação do NAD^+ na GAPDH de *T. cruzi* pode ser descrita como uma rotação de corpo rígido do domínio de ligação do NAD^+ em relação ao domínio catalítico. Durante a transição da forma **apo** para a forma **holo** o domínio de ligação do NAD^+ sofre uma rotação de aproximadamente 6° em relação ao domínio catalítico. Essa mudança conformacional promove um fechamento da subunidade (Figura 4.9). A função biológica desse fechamento seria proteger o sítio ativo e a molécula do NAD^+ dos efeitos do solvente favorecendo assim o processo de catálise enzimática.^{14,38}

Figura 4.9 - (A) Sobreposição estrutural das formas **apo** (ciano) e **holo** (magenta) da GAPDH. As setas indicam a movimentação do domínio de ligação do NAD⁺ que se fecha sobre o domínio catalítico durante a transição apo-holo. Ângulos de abertura das subunidades A da forma **apo** (B) e forma **holo** (C).

A mudança conformacional induzida pela ligação do NAD⁺ além de promover o fechamento do sítio ativo da enzima, faz com que a Met39 adote um rotâmero estrutural diferente daquele observado na estrutura **apo**. Este rotâmero da Met39, na estrutura **holo**, auxilia na formação de um bolsão hidrofóbico no sítio de ligação da porção adenina do NAD⁺, o qual favorece a estabilização da molécula do cofator através da formação de interações hidrofóbica entre o Se da Met39 e o anel adenina da molécula do cofator (Figura 4.10).

Figura 4.10 - Rotâmeros da Met39 na forma **apo** (ciano) e na forma **holo** (magenta). Alteração induzida pela ligação do NAD⁺ durante a transição apo-holo.

Formas 1NAD e 2NAD

Os dados apresentados na Tabela 4.5 mostram que as subunidades A, B e C da forma **1NAD** adotam conformação similar às subunidades equivalentes da forma **apo**, apresentando valores de rmsd de 0,44; 0,36 e 0,39 Å; respectivamente (comparação **2**, **apo** / **1NAD**), enquanto a subunidade D, que contém a molécula do NAD^+ ligada, é similar a subunidade D da forma **holo**, apresentando valor de rmsd de 0,44 Å (comparação **4**, **holo** / **1NAD**). Resultados semelhantes são observados para a forma **2NAD**, a qual possui as subunidades B e C com conformação similar à forma **apo**, apresentando valores de rmsd de 0,36 e 0,26 Å, respectivamente (comparação **3**, **apo** / **2NAD**) e as outras duas subunidades, A e D, que contém a molécula de NAD^+ ligada, adotam conformação similar a forma **holo**, apresentando valores de rmsd de 0,37 e 0,40 Å para as subunidades A e D, respectivamente (comparação **5**, **holo** / **2NAD**). Baseado nestes resultados a estrutura **1NAD** pode ser descrita como um tetrâmero constituído de três subunidades na conformação apo, e a quarta subunidade, que apresenta o NAD^+ ligado no sítio ativo, na conformação holo. Enquanto a estrutura **2NAD** pode ser descrita como um tetrâmero constituído de duas subunidades na conformação apo e duas subunidades na conformação holo. Esses resultados podem ser visualizados na Figura 4.11, que apresenta um modelo esquemático referentes às estruturas cristalográficas obtidas e as sobreposições estruturais das subunidades equivalentes das estruturas **1NAD** e **2NAD** nas respectivas subunidades das estruturas **apo** e **holo**.

Figura 4.11 – Parte superior: modelo esquemático das estruturas **apo**, **1NAD**, **2NAD** e **holo** da GAPDH de *T. cruzi*, subunidades que adotam a conformação apo estão representadas como círculo e subunidades que adotam a conformação holo estão representadas como quadrado. Parte inferior: sobreposição dos Ca das subunidades A, B, C e D da estrutura **1NAD** (cinza) e **2NAD** (laranja) nas subunidades correspondentes da estrutura **apo** (ciano) ou da estrutura **holo** (magenta). Os valores de rmsd da sobreposição das subunidades equivalentes estão mostrados na parte inferior da figura. Para facilitar a visualização, todas as subunidades estão apresentadas na mesma orientação.

O mecanismo de cooperatividade e os contatos entre subunidades

O mecanismo de cooperatividade da GAPDH foi extensamente investigado com o objetivo de elucidar a relação entre a interação com NAD^+ e a transmissão das mudanças conformacionais entre as subunidades. Muitos dos conceitos hoje compreendidos sobre a cooperatividade em macromoléculas biológicas são resultados dos estudos realizados com GAPDH.^{8-14,47} Um dos primeiros estudos relacionados ao mecanismo de cooperatividade da GAPDH revelou que moléculas análogas ao NAD^+ interagem com a enzima de modo não-cooperativo.⁴⁸ Neste estudo foi demonstrado a importância do anel piridina na molécula do NAD^+ para a cooperatividade. Em estudos paralelos, foi determinada que a GAPDH de leveduras apresentava cooperatividade positiva,¹⁰ enquanto a enzima de músculo de coelho,¹⁵ esturção,¹⁶ lagosta,¹⁷ *B. stearothermophilus*,⁴⁹ e mais recentemente *E. coli*⁹ apresentavam cooperatividade negativa.

Estudos de mutações sítio-dirigidas dissociaram o tetrâmero da GAPDH de *B. stearothermophilus* em dímeros e caracterizaram os resíduos de aminoácidos que seriam importantes para a estabilização da estrutura quaternária e cooperatividade.^{7,8} Os critérios de seleção dos aminoácidos para mutagênese foram baseados no grau de conservação dos resíduos entre enzimas GAPDH de diferente espécies e na localização destes na estrutura da enzima. Vários mutantes foram obtidos, os quais apresentaram diferentes estados de oligomerização e cooperatividade distinta em relação à interação com o NAD^+ . Os mutantes tetraméricos, Y46G, E276G, D186G e D282G exibiram cooperatividade negativa, similar à enzima selvagem, enquanto, os mutantes diméricos D186G/E276G, Y46G/S52G e Y46G/S48G exibiram cooperatividade positiva. Diante desses resultados, foi concluído que o dímero O-P (análogo as subunidade A-B na GAPDH de *T. cruzi*) seria o dímero responsável pela cooperatividade, e que a

interface formada pelo eixo-*P* seria essencial para a transmissão das mudanças conformacionais induzidas pela interação com o NAD⁺ (Figura 4.12).

Um estudo similar de mutagênese sugeriu que alterações na região 39-52 na estrutura da GAPDH de *E. coli* (correspondente à região 44-57 na estrutura de *T. cruzi*) poderiam explicar a função desta interface no mecanismo de cooperatividade.⁹ Baseado nesta informação, estudos de mutação sítio-dirigida foram realizados com a GAPDH de *B. stearothermophilus* os quais revelaram o papel significativo do resíduo Ser48 (Thr54 na estrutura da GAPDH de *T. cruzi*) para o mecanismo de cooperatividade. A GAPDH mutante tetramérica S48G não apresentou cooperatividade entre os sítios, entretanto, os autores não puderam explicar as bases moleculares do comportamento exibido.⁸

Figura 4.12 - Organização espacial das quatro subunidades da GAPDH de *B. stearothermophilus* (painel esquerdo) e *T. cruzi* (painel direito). As subunidades estão coloridas e indicadas por letras. As linhas tracejadas indicam os eixos de simetria. Os eixos-*Q* e -*C* são perpendiculares ao plano do papel.

O alinhamento sequencial da GAPDH de *T. cruzi* com enzimas GAPDH de outros organismos que apresentam cooperatividade negativa revelou várias regiões altamente conservadas. Entretanto algumas mutações pontuais chamam a atenção (Figura 4.13). Por exemplo, a GAPDH de *T. cruzi* apresenta na região inter-subunidades um resíduo Thr na posição 54, enquanto que as demais enzimas GAPDH que possuem cooperatividade negativa apresentam o resíduo Ser altamente conservado e importante para a cooperatividade. Além disso, outras mutações não conservativas localizadas na mesma região inter-subunidades podem ser observadas, por exemplo, a Val219 (Glu na estrutura de *B. stearothermophilus* e Gln nas outras GAPDHs) e Ala299 (Arg na estrutura de *B. stearothermophilus*, Cys em *P. versicolor* e Thr em *E. coli*).

Figura 4.13 - Alinhamento sequencial das GAPDHs de *T. cruzi*, *B. stearothermophilus*, *P. versicolor*, músculo de coelho e *E. coli*. Os números acima das seqüências correspondem à seqüência da GAPDH de *T. cruzi*. Resíduos conservados entre as diferentes espécies estão coloridos em cinza. As setas indicam resíduos que podem ser importantes para o comportamento não cooperativo observado na GAPDH de *T. cruzi*.

capaz de realizar interações adicionais entre as subunidades, tornando as subunidades adjacentes menos susceptíveis as mudanças conformacionais induzidas pela interação com o NAD^+ e consequentemente contribuindo para o mecanismo não cooperativo observado. Além disso, a análise estrutural das diferentes formas de GAPDH de *T. cruzi* revelou que, apesar do movimento de 6º domínio do cofator em função da interação com o NAD^+ , este não é observado nas regiões que formam os contatos entre as subunidades de nenhuma forma obtida, dado que está de acordo com o comportamento não cooperativo da GAPDH de *T. cruzi* (Figura 4.15).

Figura 4.15 - Comparação entre os contatos inter-subunidades nas formas **apo**, **1NAD**, **2NAD** e **holo**. As subunidades A estão ilustradas em verde, subunidades B em ciano, subunidades C em magenta e subunidades D em amarelo. Resíduos que realizam interações com as subunidades vizinhas estão mostrados como bastões e as ligações de hidrogênio como linhas tracejadas. As estruturas foram sobrepostas nas subunidades correspondentes da forma **apo** e todo o tetrâmero foi utilizado no protocolo de sobreposição.

4.6 - Referências

1. COPELAND, R. A. *Enzymes – practical introduction to structure, mechanism and data analysis*, New York: Wiley VHC, 2000.
2. NELSON, D. N.; MICHAEL M. C. *Lehninger principles of biochemistry*, New York: Worth Publishers, 2004.
3. MONOD, J.; WYMAN, J.; CHANGEUX, J. P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, v. 12, p. 88-118, 1965.
4. KOSHLAND, D. E.; NÉMETHY, G.; FILMER, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, v. 5, p. 365-385, 1966.
5. FERSHT, A. *Structure and mechanism in protein science – A guide to enzyme catalysis and protein folding*. USA: W. H: Freeman, 2002.
6. MEIGHEN, E. A. *Enzyme and their inhibition – drug development*. Florida: CRC Press, 2005.
7. ROITEL, O. et al. P but not R-axis interface is involved in cooperative binding of NAD on tetrameric phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *J. Mol. Biol.*, v. 326, p. 1513-1522, 2003.
8. ROITEL, O.; SERGIENKO, E.; BRANLANT, G. Dimers generated from tetrameric phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* are inactive but exhibit cooperativity in NAD binding. *Biochemistry*, v. 38, p. 16084-16091, 1999.
9. DUEE, E. et al. Comparison of the structures of wild-type and a N313T mutant of *Escherichia coli* glyceraldehyde 3-phosphate dehydrogenases: implication for NAD binding and cooperativity. *J. Mol. Biol.*, v. 257, p. 814-838, 1996.
10. KIRSCHNER, K. Co-operative binding of nicotinamide-adenine dinucleotide to yeast glyceraldehyde-3-phosphate dehydrogenase. II. Stopped-flow studies at pH 8–5 and 40 degrees C. *J. Mol. Biol.*, v. 58, p. 51-68, 1971.

11. COOK, R. A.; KOSHLAND, D. E., JR. Positive and negative cooperativity in yeast glyceraldehydes 3-phosphate dehydrogenase. *Biochemistry*, v. 9, p. 3337-3342, 1970.
12. ALLEN, G.; HARRIS, J. I. The binding of nicotinamide-adenine dinucleotide to glyceraldehydes 3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *Biochem. J.*, v. 151, p. 747-749, 1975.
13. GENNIS, L. S. Negative homotropic cooperativity and affinity heterogeneity: preparation of yeast glyceraldehyde-3-phosphate dehydrogenase with maximal affinity homogeneity. *Proc. Natl Acad. Sci. USA*, v. 73, p. 3928-3932, 1976.
14. LESLIE, A. G. W.; WONACOTT, A. J. Structural evidence for ligand-induced sequential conformational changes in glyceraldehyde 3-phosphate dehydrogenase. *J. Mol. Biol.*, v. 178, p. 743-772, 1984.
15. CONWAY, A.; KOSHLAND, D. E. JR. Negative cooperativity in enzyme action. The binding of diphosphopyridine nucleotide to glyceraldehyde 3-phosphate dehydrogenase. *Biochemistry*, v. 7, p. 4011-4023, 1968.
16. SEYDOUX, F. et al. Preparation and active-site specific properties of sturgeon muscle glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*, v. 12, p. 4290-4300, 1973.
17. DE VIJLDER, J. J.; BOERS, W.; SLATER, E. C. Binding and properties of NAD⁺ in glyceraldehydephosphate dehydrogenase from lobster-tail muscle. *Biochim. Biophys. Acta.*, v. 191, p. 214-220, 1969.
18. SHEN, Y. Q. et al. Structure of apo-glyceraldehyde-3-phosphate dehydrogenase from *Palinurus versicolor*. *J. Struct. Biol.*, v. 130, p. 1-9, 2000.
19. CORBIER, C. et al. The nicotinamide subsite of glyceraldehyde-3-phosphate dehydrogenase studied by site-directed mutagenesis. *Biochimie*, v. 72, p. 545-554, 1990.
20. CLERMONT, S. et al. Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase: role of the acidic residue in the fingerprint region of the nucleotide binding fold. *Biochemistry*, v. 32, p. 10178-10184, 1993.

21. VIGNAIS, M. L. et al. Circular permutation within the coenzyme binding domain of the tetrameric glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *Protein Sci.* 1995, v. 5, p. 994-1000, 1995.
22. HANNAERT, V. et al Purification and characterization of the native and the recombinant *Leishmania mexicana* glycosomal glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* v. 225, p. 143-149, 1994.
23. CANTOR, C. R.; SCHIMMEL, P. R. *Biophysical chemistry part III: The Behavior of Biological Macromolecules*, New York: W. H. Freeman and Co., 1980.
24. DRENTH, J. *Principles of X-ray crystallography*, Heidelberg: Springer, 1999.
25. MCREE, D. E. *Practical protein crystallography*, San Diego: Academic Press, 1999.
26. COWAN-JACOB, S. W. et al. Structure of rabbit-muscle glyceraldehyde-3-phosphate dehydrogenase *Acta Crystallogr. D Biol. Crystallogr.*, v. 9, p. 2218-2227, 2003.
27. POLIKARPOV, I. et al. Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory, *J. Synchrotron Radiat.*, v. 5, p. 72-76, 1998.
28. ARNDT, U. W. et al. Design and Use of a Camera for Low-Angle X-ray Diffraction Experiments with Synchrotron Radiation., *J. Phys. Sci. Instr.*, v. 10, p. 1035-1044, 1977.
29. LESLIE, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.*, v. 62, p. 48-57, 2006.
30. EVANS, P. R. Proceedings of the CCP4 Study Weekend. Recent Advances In: WILSON, K. S.; DAVIES, G.; ASHTON, A. W. e BAILEY S. (eds.). *Phasing*. Warrington: Daresbury Laboratory, 1997, p. 97-102.
31. FRENCH, G. S.; WILSON, K. S. On the treatment of negative intensity observations. *Acta Crystallogr. A*, v. 34, p. 517-525, 1978.

32. PAVÃO, F. et al. Structure of *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 Å resolution. *FEBS Lett.*, v. 520, p. 13-17, 2002.
33. VAGIN, A.; TEPLYCOV, A. MOLREP: An automated program for molecular replacement, *J. Appl. Cryst.*, v. 30, p. 1002-1025, 1997.
34. BRÜNGER, A. T. et al. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination., *Acta Crystallogr. D Biol. Crystallogr.* v. 54, p. 905-921, 1998.
35. EMSLEY, P.; COWTAN, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr. D*, v. 60, p. 2126-2132, 2004.
36. RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKLARAN, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* v. 7, p. 95-99, 1963.
37. WATSON, H. C., DUEE, E., MERCER, W. D. Low resolution structure of glyceraldehyde 3-phosphate dehydrogenase. *Nat. New Biol.*, v. 240, p. 130-133, 1972.
38. SKARZYŃSKI, T.; WONACOTT, A. J. Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *J. Mol. Biol.*, v. 203, p. 1097-1118, 1988.
39. VELLIEUX, F. M. D. et al. Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. *Proc. Natl. Acad. Sci. USA*, v. 90, p. 2355-2359, 1993.
40. KIM, H. et al. Crystal structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Leishmania mexicana*: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry*, v. 34, p. 14975-14986, 1995.
41. CASTILHO, M. S. et al. Evidence for the two phosphate binding sites of an analogue of the thioacyl intermediate for the *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase-catalyzed reaction, from its crystal structure. *Biochemistry*, v. 42, p. 7143-7151, 2003.

42. LADAME, S. et al. Crystal structure of *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase complexed with an analogue of 1,3-bisphospho-d-glyceric acid. *Eur. J. Biochem.*, v. 270, p. 4574-4586, 2003.
43. WILLIAMS, D. H.; ZHOU, M.; STEPHENS, E. Ligand binding energy and enzyme efficiency from reductions in protein dynamics. *J. Mol. Biol.*, v. 355, p. 760-767, 2006.
44. SKARZYNSKI, T.; MOODY, P. C.; WONACOTT, A. J. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.*, v. 193, p. 171-87, 1987.
45. YUN, M. et al. Structural analysis of glyceraldehydes-3-phosphate dehydrogenase from *Escherichia coli*: direct evidence of substrate binding and cofactor-induced conformational changes. *Biochemistry*, v. 39, p. 10702-10710, 2000.
46. LORENTZEN, E. et al. Structural Basis of allosteric regulation and substrate specificity of the non-phosphorylating glyceraldehydes-3-Phosphate dehydrogenase from *Thermoproteus tenax*. *J. Mol. Biol.*, v. 341, 815-828, 2004.
47. TRAUT, T. W. Dissociation of enzyme oligomers: a mechanism for allosteric regulation. *Crit. Rev. Biochem. Mol. Biol.*, v. 29, p. 125-163, 1994.
48. EBY, D.; KIRTLEY, M. E. Interaction of nicotinamide-adenine dinucleotide and its analogs with glyceraldehyde 3-phosphate dehydrogenase. *Biochemistry*, v. 10, p. 2677-2682, 1971.
49. LESLIE, A. G.; WONACOTT, A. J. Coenzyme binding in crystals of glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.*, v. 165, p. 375-391, 1983.

Capítulo 5

Estudos de Modelagem Molecular com a GAPDH de Tripanosomatídeos

Esse capítulo inicia com uma breve introdução sobre a modelagem molecular e os métodos computacionais em Química Medicinal, enfatizando algumas estratégias empregadas no planejamento de fármacos. Em seguida, há uma descrição detalhada das etapas envolvidas na construção de uma base de dados de inibidores da GAPDH de tripanosomatídeos e dos métodos de QSAR 2D (HQSAR) e QSAR 3D (CoMFA e CoMSIA) empregados na geração de modelos preditivos. Na última seção é apresentado o emprego de uma estratégia de ensaio virtual para a identificação de novos ligantes da GAPDH de *T. cruzi*.

5 - ESTUDOS DE MODELAGEM MOLECULAR COM AS ENZIMAS GAPDH DE TRIPANOSOMATÍDEOS

5.1 - Química medicinal e as técnicas de modelagem molecular

Os projetos de descoberta e planejamento de novos fármacos são guiados pela inovação e uma combinação de métodos experimentais e computacionais.¹ Os avanços nas ciências biológicas contribuíram significativamente para a identificação e validação de novos alvos biológicos de interesse terapêutico. Além disso, iniciativas como os programas genoma e proteoma de vários organismos têm fornecido dados importantes para o detalhamento das bases moleculares responsáveis pela estrutura e função de biomoléculas. Simultaneamente, o aprimoramento das técnicas de determinação estrutural e análise de moléculas, como a cristalografia de raios X, ressonância magnética nuclear (RMN) e a calorimetria, têm contribuído substancialmente para a melhor compreensão dos componentes energéticos e espaciais que compõe as interações entre fármacos e receptores.¹⁻³

Métodos computacionais são ferramentas amplamente empregadas nos projetos de descoberta e desenvolvimento de novos fármacos. Estes métodos são utilizados desde a identificação de moléculas com atividade biológica até a otimização das propriedades químicas e farmacêuticas de compostos líderes. Devido à rápida evolução na área de *hardware*, *software* e métodos, os processos computacionais anteriormente calculados exclusivamente em estações de trabalho de alto desempenho são hoje em dia realizados em microcomputadores numa escala de tempo bastante razoável. Diante deste cenário, a integração entre técnicas experimentais e computacionais assume função central no processo de planejamento de novos fármacos.^{1,2}

Um dos maiores desafios da química medicinal é o de aumentar a taxa de sucesso na identificação de novas NCEs provenientes de coleções de compostos reais ou virtuais.^{1,4} Neste contexto, duas estratégias de grande impacto destacam-se, o SBDD e o LBDD.

5.1.1 - Planejamento baseado na estrutura do receptor

O SBDD é uma técnica baseada no conhecimento do arranjo topológico de alvos moleculares, logo, utiliza como pré-requisito a informação 3D detalhada da macromolécula em estudo. Essa informação é geralmente adquirida através da análise de estruturas obtidas por cristalografia de raios X, estudos de RMN ou modelagem por homologia.^{1,4} Estas estruturas são depositadas em bases de dados públicas onde podem ser acessadas livremente. As principais bases de dados são: PDB,⁵ InterPro,⁶⁻⁸ ExPASy,⁹ e Relibase.¹⁰⁻¹²

A docagem molecular (do inglês, *molecular docking*) é uma das principais estratégias de SBDD empregada na descoberta de novas moléculas bioativas. Esta técnica consiste na predição da conformação bioativa de uma pequena molécula (ligante) no sítio de ligação de uma macromolécula (e.g., enzima, receptor, DNA ou RNA) seguido de avaliação (pontuação) e classificação do modo de ligação proposto.^{1,2} A Figura 5.1 apresenta um esquema geral do processo de docagem molecular. Uma estratégia bastante difundida que emprega os métodos de docagem molecular é a triagem virtual. Nesta aplicação, os programas de docagem molecular são utilizados para identificar e classificar as conformações bioativas de moléculas pertencentes a bases de dados podendo conter centenas de milhares de compostos disponíveis para aquisição comercial ou síntese.²

Figura 5.1 - Etapas envolvidas no processo de docagem molecular.

A docagem molecular foi introduzida no início da década de 80 e tem contribuído significativamente tanto na fase de descoberta de novas moléculas bioativas, quanto na fase de otimização de compostos líderes em relação a uma série de propriedades, como potência, afinidade, propriedades farmacocinéticas, entre outras.^{2,13}

Um dos primeiros grandes desafios em SBDD é a seleção do programa de docagem molecular. Os métodos disponíveis são baseados em conceitos que diferem ligeiramente entre si, fazendo com que um programa em particular seja mais adequado para um determinado projeto. É estimado que existam aproximadamente 30 programas de docagem molecular disponíveis,¹⁴ entre os quais destacam-se o DOCK,¹⁵ GOLD,¹⁶ FlexX,¹⁷ GLIDE,^{18,19} AUTODOCK,²⁰ e SURFLEX-DOCK.²¹

O processo de docagem molecular pode ser dividido em duas etapas complexas: modelagem do modo de ligação (predição da conformação bioativa) e predição da afinidade do ligante pelo sítio. A predição do modo de ligação de moléculas no sítio de ligação de receptores-alvo é considerada a etapa mais simples e robusta do processo, com considerável nível sucesso alcançado.⁴ Para realizar essa primeira etapa, os programas utilizam diferentes métodos de amostragem que atribuem flexibilidade à molécula do ligante. Os métodos de amostragem podem ser classificados em três categorias principais: (i) **Métodos sistemáticos** (e.g., Construção

Incremental, Amostragem Conformacional); (ii) **Métodos aleatórios** ou **Estocásticos** (e.g., Método de Monte Carlo, Algoritmo Genético); e (iii) **Métodos de simulação** (e.g., Dinâmica Molecular e Minimização de Energia). A etapa de predição da afinidade do ligante pelo sítio de ligação é realizada baseada em funções de pontuação desenvolvidas para avaliar e classificar o modo de ligação proposto na etapa anterior. As funções de pontuação atualmente utilizadas podem ser agrupadas em três categorias básicas:²² (i) **Campos de força**, geralmente baseados em métodos de mecânica molecular, quantificam a energia proveniente das interações entre o ligante e receptor e a energia interna do ligante (e.g., G-score e D-score, implementados no pacote de programas SYBYL;²³ Amber,²⁴ implementado no programa AUTODOCK;²⁰ GOLDscore implementado no programa GOLD);¹⁶ (ii) **Empíricas**, funções baseadas no ajuste teórico de dados experimentais através do emprego de equações (e.g., regressão linear) (e.g., LigScore,²⁵ PLP,²⁵ LUDI,²⁵ F-Score,²³ ChemScore²³ e X-score²⁶); (iii) **Funções baseadas no conhecimento**, essas funções utilizam dados estatísticos dos potenciais de pares atômicos de interação, os quais são derivados de conjuntos de dados proveniente de complexos cristalográficos entre proteínas e ligantes (e.g., PMF²⁵ e DrugScore²⁷).

No processo de reconhecimento molecular e afinidade a energia livre de ligação (Energia livre de Gibbs) é determinada pela combinação das contribuições entálpicas e entrópicas.⁴ Esse conceito é implementado nas funções de pontuação de maneira aproximada, uma vez que o componente entálpico está razoavelmente caracterizado e descrito, entretanto, as contribuições entrópicas são difíceis de descrever, pois ainda não são totalmente compreendidas e dependem de um número significativo de fatores, como, por exemplo, a área de superfície hidrofóbica, a liberação de moléculas de água ligadas à cavidade de ligação (dessolvatação) e a imobilização das ligações rotacionáveis na molécula do ligante.⁴

Um estudo realizado com 100 complexos cristalográficos entre proteína-ligante avaliou a capacidade preditiva de 11 funções de pontuação através da correlação entre as energias de ligação preditas e os dados experimentais dos ligantes dos complexos. O estudo revelou que somente as funções X-Score, PLP, DrugScore, e G-Score apresentaram um coeficiente de correlação maior do que 0,50.²² Esses resultados indicaram que a predição exata da energia livre de ligação ainda necessita de uma melhor descrição dos componentes determinantes da afinidade de ligação. Portanto, as funções de pontuação implementadas nos programas de docagem molecular são funções simplificadas e aproximadas. Entretanto, apesar de explicarem parcialmente o complexo processo de interação macromolécula-ligante, são amplamente empregadas para a identificação e seleção de novos compostos com atividade biológica.^{1,2,4}

As seções 5.2.5 e 5.3 apresentam com mais detalhes os procedimentos e métodos de docagem molecular utilizados como ferramentas importantes para a obtenção do alinhamento estrutural utilizado nos estudos de QSAR 3D e para os estudos de triagem virtual de compostos inibidores da GAPDH de *T. cruzi*, respectivamente.

5.1.2 - Planejamento baseado na estrutura do ligante

Métodos em química medicinal com base na estrutura do ligante são muito empregados na identificação de novas moléculas bioativas e na otimização de compostos líderes.²⁸ A descoberta de uma grande quantidade de compostos biologicamente ativos foi possível devido ao uso de métodos de LBDD que não exigem informação sobre a topologia do alvo molecular.²⁹

As principais estratégias de LBDD são baseadas na análise comparativa de moléculas bioativas e podem ser classificadas em quatro categorias principais: (i) **Buscas por similaridade**, utilizam principalmente métodos de impressão digital molecular (do inglês, *molecular*

fingerprint) para identificar e classificar compostos como similares ou diferentes de um ligante bioativo.³⁰ Esses métodos são baseados no princípio das propriedades similares, ou seja, moléculas estruturalmente relacionadas podem apresentar atividade biológica similar; (ii) **Farmacóforos**, comumente definido como um arranjo tridimensional de propriedades moleculares que determinam um conjunto mínimo de condições necessárias para o reconhecimento molecular. Esses métodos envolvem a identificação de padrões farmacofóricos comuns entre um conjunto de moléculas bioativas e a utilização deste padrão (modelo farmacofórico) para buscas de novas moléculas que satisfaçam os requisitos impostos pelo modelo;³¹ (iii) **Inteligência artificial**, esta técnica baseia-se em regras classificatórias derivadas a partir de um conjunto de moléculas ativas e inativas utilizado para construir e treinar modelos computacionais (do inglês, *machine learning*). Posteriormente, o modelo desenvolvido é aplicado para a identificação e seleção de novos compostos com atividade biológica;³² (iv) **Métodos de QSAR**, buscam identificar e quantificar as relações predominantes entre a estrutura química e atividade biológica de um conjunto treinamento de moléculas quimicamente relacionadas através do uso de descritores químicos.³³

No presente trabalho, os métodos de QSAR 2D e QSAR 3D foram amplamente empregados para o estudo das bases moleculares determinantes da potência e seletividade dos inibidores das enzimas GAPDH de tripanosomatídeos. Os detalhes destes estudos são apresentados na seção 5.2.2 e 5.2.3.

5.2 - Relações quantitativas entre a estrutura-atividade

O QSAR é uma das técnicas mais amplamente empregadas em química medicinal para o planejamento de NCEs candidatas a fármacos. Os métodos de QSAR têm como objetivo o planejamento de moléculas com propriedades superiores em relação àquelas apresentadas pelo conjunto de dados que deu origem ao processo de modelagem. Para que tais estudos sejam realizados de maneira padrão e comparativa, as moléculas do conjunto treinamento precisam ser quimicamente relacionadas (ocupar o mesmo espaço químico), exercer seus efeitos em um mesmo alvo terapêutico, na mesma cavidade de ligação e através do mesmo mecanismo de ação.³⁴⁻³⁷ A propriedade alvo, geralmente potência ou afinidade, deve ser padronizada e validada para ser útil no desenvolvimento de modelos quantitativos. Propriedades moleculares são calculadas e empregadas como descritores. Métodos estatísticos são usados para produzir uma relação que descreve as variações na propriedade alvo, em função dos descritores moleculares. O processo cíclico de formulação de QSAR, predição da propriedade alvo, síntese e ensaios biológicos, prossegue até que novas moléculas candidatas a NCEs sejam selecionadas com elevado potencial de desenvolvimento clínico.³⁸⁻⁴⁰

As técnicas de QSAR são muito exploradas em conjunto com técnicas de SBDD quando estruturas 3D dos alvos macromoleculares eleitos são disponíveis.^{41,42} Uma vez que as características químicas e estruturais associadas às propriedades biológicas alvo sejam identificadas na modelagem de QSAR, estas podem ser comparadas com as informações obtidas no estudo das características topológicas 3D dos receptores macromoleculares estudados. A integração dessas informações altamente qualificadas é uma estratégia importante em química medicinal, uma vez que as informações complementares geradas são úteis no planejamento de candidatos a NCEs mais potentes e seletivas.^{37,43}

5.2.1 - Estudos de QSAR para os inibidores da GAPDH de *L. mexicana*

Métodos computacionais

Os estudos de modelagem molecular, análises estatísticas e construção de modelos de QSAR 2D e QSAR 3D foram realizados em estações computacionais no Laboratório de Química Medicinal e Computacional (LQMC) do Centro de Biotecnologia Molecular Estrutural (CBME), no Instituto de Física de São Carlos da Universidade de São Paulo (IFSC/USP). Os trabalhos foram desenvolvidos em estações de trabalho Enterprise Linux utilizando as interfaces computacionais SYBYL 7.3 (Tripos, St. Louis, USA), GOLD 3.1 (Cambridge Crystallographic Data Centre, Cambridge, UK), FlexX 2.1 (BioSolveIT GmbH, Sankt Augustin, Germany) e Pymol 0.99 (DeLano Scientific, San Carlos, USA).

Conjunto de dados

As análises de QSAR realizadas no presente trabalho foram conduzidas em um conjunto de dados contendo 70 inibidores da GAPDH de *L. mexicana* selecionados de cinco publicações reportadas pelo mesmo laboratório.⁴⁴⁻⁴⁸ A propriedade biológica avaliada dos compostos do conjunto de dados foi a potência inibitória, representada pelo valor de IC₅₀. Os dados de inibição utilizados foram determinados sob as mesmas condições experimentais, requisito essencial para os estudos de QSAR.³⁴⁻³⁷

Os compostos investigados no presente trabalho apresentam diversidade estrutural moderada. O conjunto de dados consiste de derivados de adenosina substituídos nas posições N6 e C9 do anel adenina que exploram diferenças estruturais significativas observadas no sítio de ligação do NAD⁺ da GAPDH de parasitas.⁴⁴

A geração de modelos estatísticos consistentes depende da constituição dos conjuntos treinamento e teste.³⁵⁻³⁷ Assim, o conjunto de dados original foi dividido em conjunto treinamento (compostos **1-56**, Tabela 5.1) e conjunto teste (composto **57-70**, Tabela 5.1) na razão de 5:1 (20%). Os compostos foram cuidadosamente selecionados para garantir uma distribuição adequada das moléculas com diferentes valores de potência em ambos os conjuntos (Figura 5.2). Para tanto, foi conduzido um estudo de análise de agrupamento (do inglês, *cluster*) com o programa TSAR 3D versão 3.3 (Accelrys, San Diego, USA) para confirmar que a composição dos conjuntos treinamento e teste eram representativos de todo o conjunto de dados. Os valores de IC_{50} para os compostos dos conjuntos treinamento e teste foram convertidos em pIC_{50} ($-\log IC_{50}$) e usados como variável dependente nas análises de QSAR. Os valores de pIC_{50} estão aceitavelmente distribuídos ao longo de um intervalo de potência de três ordens e meia de magnitude (Figura 5.2).

Figura 5.2 - Distribuição dos valores de pIC_{50} para o conjunto treinamento, conjunto teste e conjunto de dados.

Tabela 5.1 - Estruturas químicas e os valores associados de pIC₅₀ do conjunto de dados de inibidores da GAPDH de *L. mexicana*.

Conjunto Treinamento					
Cpd	Estrutura	pIC ₅₀	Cpd	Estrutura	pIC ₅₀
1		3,30	2		2,52
3		3,15	4		2,40
5		2,80	6		2,22
7		3,40	8		3,60
9		3,12			

Continua

Continuação

Conjunto Treinamento

Cpd	R ¹	R ²	R ³	pIC ₅₀
10	CH ₃	H	H	2,22
11	H	H		2,40
12	H		H	2,43
13	H		H	3,15
14	H		H	3,15
15	H		H	2,74
16	H		H	3,44
17	H		H	3,47
18	H		H	3,62

Continua

Continuação

Conjunto Treinamento

Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
19	H	H	H	H	H	H	2,48
20	H	H	H	H	C ₆ H ₅	H	3,70
21	H	H	H	OCH ₂ CH ₃	H	H	3,40
22	H	H	H	OCH ₃	OH	H	3,30
23	H	Br	H	H	H	H	2,52
24	H	H	H	OH	OH	OH	3,60
25	H	H	H	OH	H	OH	3,19
26		H	H	OCH ₃	H	OCH ₃	4,70
27		H	H	OCH ₃	H	H	4,80
28		H	H	OCH ₃	H	OCH ₃	4,60
29		H	H	OCH ₃	H	H	4,60
30		H	H	OCH ₃	H	OCH ₃	4,60

Continua

Continuação

Conjunto Treinamento

Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
31		H	H	OCH ₃	H	H	4,60
32		H	H	OCH ₃	H	H	5,26
33		H	H	OCH ₃	H	H	3,82
34		H	H	OCH ₃	H	H	4,10
35		H	H	Cl	H	H	5,00
36		H	H	OCH ₃	H	OCH ₃	5,70
37		H	H	Cl	H	Cl	4,92
38		H	H	OAc	H	OAc	5,00
39		H	H	OH	H	OH	5,30
40		H	H	H	OAc	H	4,00

Continua

Continuação

Conjunto Treinamento

Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
41		H	H	H	OH	H	4,10
42		H	H	OAc	OAc	H	4,60
43		H	H	OH	OCH ₃	H	4,43
44		H	H	OCH ₃	OH	H	5,00
45		H	H	N(CH ₃) ₂	H	H	4,60
46		H	H	OCH ₃	H	OCH ₃	5,70
47		H	H	Cl	H	Cl	4,60
48		H	H	OCH ₃	H	OCH ₃	4,22
49		H	H	OCH ₃	H	H	5,22

Continua

Continuação

Conjunto Treinamento

Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
50		H	H	Cl	H	H	5,40
51		H	H	OCH ₃	H	OCH ₃	5,70
52		H	H	Cl	H	Cl	4,60
53		H	H	OCH ₃	H	H	5,70
54		H	H	Cl	H	Cl	4,60
55		H	H	OCH ₃	H	H	5,30
56		H	H	OCH ₃	H	H	4,08

Continua

Continuação

Conjunto Teste								
Cpd	Estrutura	pIC ₅₀	Cpd	Estrutura	IC ₅₀	Cpd	Estrutura	pIC ₅₀
57		2,62	58		3,52	59		3,30
60		3,22	61		3,07	62		3,82
63		5,30	64		5,40	65		5,70

Continua

Conclusão

Conjunto Teste								
Cpd	Estrutura	pIC ₅₀	Cpd	Estrutura	pIC ₅₀	Cpd	Estrutura	pIC ₅₀
66		4,30	67		4,60	68		4,00
69		4,74	70		5,00			

As estruturas 3D dos inibidores foram construídas utilizando os parâmetros geométricos padrão do módulo SKETCHER implementado no pacote de programas SYBYL 7.3. A conformação otimizada de cada molécula do conjunto de dados foi obtida através do processo de minimização de energia empregando o campo de força Tripos⁴⁹ e o algoritmo Powell⁵⁰ de gradientes conjugados com critério de convergência de 0,05 kcal/mol.Å. As cargas atômicas parciais foram calculadas pelo método Gasteiger-Hückel⁵¹ também implementado no SYBYL 7.3.

Validação dos modelos de QSAR

Todos os modelos de QSAR gerados no presente trabalho foram investigados utilizando-se o método de validação cruzada r^2 (q^2) baseado nas análises de regressões multivariadas pelo método de mínimos quadrados parciais (PLS, do inglês, *partial least squares*). O protocolo de validação cruzada LOO (deixe um de fora, do inglês, *leave-one-out*) implementado no programa SAMPLS⁵² foi aplicado para a determinação do número ótimo de componentes principais que geravam os modelos com maior capacidade de predição. O protocolo LMO (deixe muitos de fora, do inglês, *leave-many-out*) com dez (LMO₁₀) e cinco (LMO₅) grupos, aleatoriamente selecionados, foi aplicado como um teste mais rigoroso de validação cruzada para avaliar a estabilidade e significância estatística dos modelos. Cada análise LMO foi realizada independentemente e repetida 25 vezes. Os valores médios de q^2 e do desvio padrão do erro de predição (SDEP) são reportados para cada verificação. O método de validação externa, que avalia a real capacidade de predição dos modelos, foi realizado com um conjunto teste de 14 compostos que não foram considerados durante as etapas de construção dos modelos de QSAR.

5.2.2 - Holograma QSAR

O holograma QSAR (HQSAR, do inglês, *hologram quantitative structure-activity relationships*) é um método de QSAR bidimensional que converte a estrutura molecular em notações lineares, ou seja, seqüências numéricas que codificam a estrutura 2D das moléculas do conjunto de dados. Baseando-se no conceito central do QSAR de que a atividade biológica é função da estrutura molecular, é possível estabelecer relações quantitativas entre a atividade de uma série de moléculas a partir das informações codificadas nos descritores bidimensionais. O

método é fundamentado na expansão do conceito de impressão digital molecular, utilizando uma forma estendida de impressão digital molecular conhecida como holograma molecular. O holograma molecular é caracterizado por codificar uma quantidade maior de informação estrutural em relação à impressão digital tradicional. Enquanto que o método tradicional utiliza apenas uma notação binária (possui ou não o fragmento), o holograma molecular contém, além de todos os possíveis fragmentos moleculares, uma contagem do número de vezes que cada fragmento ocorre na molécula.⁵³⁻⁵⁵ A Figura 5.3 apresenta um exemplo de impressão digital molecular e holograma molecular.

Impressões digital molecular	0	0	1	1	0	0	0	1	1	1	1	0	0	0	0
Holograma molecular	0	0	6	18	0	0	0	12	5	14	42	0	0	0	0

Figura 5.3 - Exemplos de impressão digital molecular e de holograma molecular.

Na técnica de HQSAR, os hologramas moleculares são originários da fragmentação molecular bidimensional. Todos os fragmentos moleculares possíveis (lineares, ramificados e sobrepostos) são obtidos e posteriormente distribuídos pelo holograma. Cada posição do holograma contém a frequência com que um determinado fragmento aparece na molécula, sendo que fragmentos iguais que ocorrem mais de uma vez são sempre alocados na mesma posição. No exemplo apresentado na Figura 5.3 o holograma molecular que representa uma estrutura química arbitrária contém 97 fragmentos (6+18+12+5+14+42). Informações tridimensionais adicionais, como hibridização e quiralidade são também codificadas nos hologramas moleculares. Desta forma, os hologramas definem a dimensionalidade do espaço ocupado pelos descritores e os

modelos de QSAR são desenvolvidos empregando-se o método PLS e técnicas de validação cruzada. O processo de obtenção do holograma molecular esta resumido na Figura 5.4.

Figura 5.4 - Esquema ilustrativo do processo de geração dos hologramas moleculares.

Durante o processo de modelagem de HQSAR, alguns parâmetros relevantes para a obtenção dos hologramas afetam a qualidade dos modelos. Esses parâmetros são: (i) comprimento do holograma; (ii) tamanho de fragmentos; e, (iii) distinção de fragmentos. O comprimento do holograma define o número de posições no holograma disponível para que diferentes fragmentos sejam incorporados. O tamanho de fragmentos é definido pelo número máximo e mínimo de átomos que é imposto aos fragmentos a serem gerados, enquanto o parâmetro de distinção de fragmentos define as características moleculares que distinguem um fragmento do outro. O modo como os fragmentos são distribuídos nas posições do holograma influencia diretamente a qualidade dos modelos.^{53,56} O algoritmo cíclico de verificação de redundância (CRC, do inglês, *cyclic redundancy check*) avalia várias combinações de distribuição

de fragmentos nas diferentes posições do holograma molecular para impedir que haja a colisão de fragmentos, ou seja, essa estratégia evita que fragmentos distintos e importantes sejam agrupados numa mesma posição no holograma.

Os parâmetros utilizados para distinguir os fragmentos são variáveis ajustáveis e definem o tipo de informação 2D que será incluída na criação dos descritores. Na Tabela 5.2 estão apresentados os parâmetros de distinção de fragmentos considerados no método HQSAR. Combinações desses parâmetros são utilizadas para determinar a melhor correlação entre as estruturas e a propriedade-alvo.^{53,56}

Tabela 5.2 - Parâmetros de distinção de fragmentos no método HQSAR.

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

O desenvolvimento do modelo estatístico de HQSAR utiliza os hologramas gerados como descritores que são relacionados com a propriedade-alvo através do método PLS (Figura 5.5). Uma vez determinado, o modelo de HQSAR pode ser utilizado para a predição da propriedade-alvo de novas moléculas.^{35,36,57,58}

Figura 5.5 - Etapas envolvidas na geração de um modelo de HQSAR.

Além de prever a propriedade-alvo de moléculas ainda não testadas, o método HQSAR fornece importantes informações sobre os fragmentos estruturais associados à propriedade estudada. Estas informações são obtidas através da análise dos mapas de contribuição e podem orientar a síntese de novas moléculas com propriedades otimizadas. Esses mapas utilizam um código de cores para indicar a magnitude da contribuição de diferentes fragmentos moleculares.⁵³ Esta técnica de QSAR é capaz de gerar modelos com alta capacidade de interpolação e extrapolação em diversos conjuntos de dados.

5.2.3 - Estudos de HQSAR para os Inibidores de GAPDH de *L. mexicana*

Análise de HQSAR

A geração dos hologramas moleculares foi realizada utilizando-se várias combinações de distinção de fragmentos: AB, ABC, ABCH, ABCHCh, ABCHChDA, ABH, ABCCCh, ABDA, ABCDA, ABHDA, ABCHDA, e ABHChDA. O tamanho padrão de fragmentos (4-7) foi empregado para a obtenção dos modelos iniciais de HQSAR. As análises de HQSAR foram realizadas com os 12 padrões de comprimento do holograma (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401). Os resultados estatísticos das análises de PLS para os 56 inibidores do conjunto treinamento utilizando as diversas combinações de distinção de fragmentos estão apresentados na Tabela 5.3.

Os valores elevados de q^2 e r^2 revelaram que modelos estatísticos de qualidade razoável foram obtidos utilizando-se diferentes combinações de distinção de fragmentos. Os melhores resultados estatísticos foram obtidos para os modelos **6**, **10** e **12**, derivados com os parâmetros de distinção de fragmentos A/B/H, A/B/H/DA, e A/B/H/Ch/DA, respectivamente (Tabela 5.3).

Tabela 5.3 - Parâmetros estatísticos das análises de HQSAR para várias combinações de distinção de fragmentos utilizando o tamanho padrão de fragmentos (4-7).

Modelo	Distinção de Fragmentos	q^2	SDEP	r^2	SEE	HL	PC
1	A/B	0,831	0,439	0,885	0,355	257	3
2	A/B/C	0,835	0,436	0,888	0,350	353	3
3	A/B/C/H	0,836	0,429	0,884	0,362	307	3
4	A/B/C/H/Ch	0,835	0,432	0,887	0,357	353	3
5	A/B/C/H/Ch/DA	0,823	0,447	0,886	0,359	353	3
6	A/B/H	0,851	0,409	0,895	0,343	151	3
7	A/B/C/Ch	0,832	0,437	0,884	0,356	353	3
8	A/B/DA	0,834	0,435	0,894	0,345	97	3
9	A/B/C/DA	0,827	0,442	0,886	0,359	199	3
10	A/B/H/DA	0,848	0,414	0,896	0,342	151	3
11	A/B/C/H/DA	0,821	0,449	0,876	0,374	151	3
12	A/B/H/Ch/DA	0,855	0,405	0,903	0,331	307	3

Distinção de fragmentos: A, átomo; B, ligação; C, conectividade; H, hidrogênio; Ch, quiralidade; DA, doador e aceptor. q^2 , coeficiente de correlação da validação cruzada; SDEP, desvio padrão do erro de predição; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento do holograma; PC, número ótimo de componentes principais.

A influência do tamanho de fragmentos foi posteriormente avaliada para os três melhores modelos obtidos na etapa anterior (modelos **6**, **10** e **12**) empregando os 12 padrões de comprimento do holograma. Os resultados estatísticos apresentados na Tabela 5.4 demonstram que a variação do tamanho do holograma apresenta pouca influência para a geração de modelos de HQSAR para essa classe de compostos. Modelos com parâmetros estatísticos ligeiramente melhores foram obtidos com a variação do tamanho de fragmentos (modelos **16**, **22** e **28** obtidos com o tamanho de fragmento 5-8).

Tabela 5.4 - Influência do tamanho de fragmentos nos parâmetros estatísticos dos modelos de HQSAR.

Distinção de Fragmentos	Modelo	Tamanho de Fragmentos	q^2	SDEP	r^2	SEE	HL	PC
A/B/H	13	2-5	0,837	0,428	0,880	0,368	61	3
	14	3-6	0,845	0,418	0,888	0,355	257	3
	15	4-7	0,851	0,409	0,895	0,343	151	3
	16	5-8	0,864	0,391	0,905	0,326	59	3
	17	6-9	0,840	0,422	0,892	0,348	353	3
	18	7-10	0,835	0,431	0,883	0,362	353	3
A/B/H/DA	19	2-5	0,835	0,431	0,887	0,357	71	3
	20	3-6	0,837	0,428	0,889	0,354	257	3
	21	4-7	0,848	0,414	0,896	0,342	151	3
	22	5-8	0,857	0,399	0,895	0,344	307	3
	23	6-9	0,836	0,429	0,888	0,355	97	3
	24	7-10	0,838	0,427	0,884	0,360	257	3
A/B/H/Ch/DA	25	2-5	0,839	0,426	0,889	0,354	71	3
	26	3-6	0,843	0,420	0,894	0,345	97	3
	27	4-7	0,855	0,405	0,903	0,331	307	3
	28	5-8	0,860	0,397	0,897	0,340	97	3
	29	6-9	0,821	0,449	0,870	0,383	71	3
	30	7-10	0,808	0,465	0,858	0,400	71	3

Com o objetivo de avaliar a significância estatística e a estabilidade dos modelos gerados, um protocolo mais consistente de validação cruzada (LMO) foi aplicado nos três melhores modelos de HQSAR. Os valores médios de q^2 e SDEP obtidos no protocolo de validação cruzada LMO₁₀ e LMO₅, que seleciona aleatoriamente 10% e 20% dos compostos do conjunto treinamento, respectivamente, estão reportados na Tabela 5.5. Os dados da Tabela 5.5 são bastante similares àqueles previamente obtidos pelas análises LOO e confirmam a estabilidade e robustez dos modelos.

Tabela 5.5 - Valor médio de q^2 e SDEP obtidos após 25 análises independentes do protocolo de validação cruzada LMO.

Distinção de Fragmentos	Modelo	LMO ₁₀ ^a		LMO ₅ ^b	
		q^2	SDEP	q^2	SDEP
A/B/H	16	0,861	0,396	0,858	0,399
A/B/H/DA	22	0,839	0,425	0,840	0,424
A/B/H/Ch/DA	28	0,838	0,427	0,841	0,422

^aValor médio de q^2 e SDEP obtidos após 25 análises independentes do protocolo de validação cruzada com 10% dos compostos do conjunto treinamento. ^bValor médio de q^2 e SDEP obtidos após 25 análises independentes do protocolo de validação cruzada com 20% dos compostos do conjunto treinamento.

A capacidade de predição dos três melhores modelos foi avaliada utilizando-se um conjunto teste de moléculas que não fizeram parte da etapa de construção dos modelos (compostos **57-70**, Tabela 5.1). Com o objetivo de selecionar o modelo com maior poder preditivo, os modelos **16**, **22** e **28** foram utilizados para predizer a atividade inibitória dos compostos do conjunto teste. O melhor modelo foi selecionado com base no valor de r^2_{pred} (r^2 -preditivo), que indica a capacidade de predição do modelo para moléculas não pertencentes ao conjunto treinamento. O valor de r^2_{pred} igual a 1 indica o modelo ideal, portanto, quanto mais próximo de 1, maior a capacidade preditiva do modelo construído. Os valores calculados de r^2_{pred} para os modelos **16**, **22** e **28** foram 0,92; 0,83 e 0,85; respectivamente. Conseqüentemente, o modelo **16** foi selecionado como o melhor modelo de HQSAR para o conjunto de inibidores da GAPDH de *L. mexicana*. A Figura 5.6 apresenta o gráfico dos resultados experimentais *versus* preditos para os conjuntos treinamento e teste. Os resultados da validação externa do modelo **16** estão apresentados na Tabela 5.6. Os dados da Tabela 5.6 demonstram que os valores preditos de pIC₅₀ dos compostos do conjunto teste estão próximos dos valores experimentais, uma vez que

não desviavam mais que 0,66 unidade logarítmica. Esses dados indicam que o modelo de HQSAR construído possui alta capacidade de predição da potência inibitória de novas moléculas.³⁶

Tabela 5.6 - Valores experimentais, preditos e residuais de pIC₅₀ calculados com o melhor modelo de HQSAR para o conjunto teste de 14 inibidores da GAPDH de *L. mexicana*.

Conjunto Teste			
Cpd	Experimental	HQSAR	
		Predito	Resíduo ^a
57	2,62	2,63	-0,01
58	3,52	3,35	0,17
59	3,30	3,32	-0,02
60	3,22	3,27	-0,05
61	3,07	3,17	-0,10
62	3,82	3,71	0,11
63	5,30	5,31	-0,01
64	5,40	5,12	0,28
65	5,70	5,75	-0,05
66	4,30	4,91	-0,61
67	4,60	4,87	-0,27
68	4,00	4,66	-0,66
69	4,74	4,72	0,02
70	5,00	4,99	0,01

^aDiferença entre o valor de pIC₅₀ experimental e predito.

Figura 5.6 - Gráfico dos valores experimentais *versus* preditos de pIC₅₀ obtido com o melhor modelo de HQSAR (modelo **16**, $r^2_{\text{pred}} = 0,92$) para o conjunto treinamento (quadrados) e conjunto teste (triângulos) de inibidores da GAPDH de *L. mexicana*.

Além de prever a propriedade biológica de novos compostos, os modelos de HQSAR fornecem informações importantes sobre a relação entre os fragmentos gerados e a atividade biológica.^{35,36,57-59} Os modelos de HQSAR podem ser graficamente representados na forma de mapas de contribuição com esquema de cores de acordo com a contribuição de cada fragmento para a atividade biológica. As cores no espectro vermelho (e.g., vermelho ao laranja) representam as contribuições negativas, enquanto que as cores no espectro verde (e.g., amarelo ao verde) representam as contribuições positivas para a propriedade biológica investigada. Átomos com contribuições intermediárias ou neutras são marcados em branco.

Os fragmentos mais importantes para a atividade biológica do composto **36** (uma das moléculas mais potentes do conjunto de dados, com IC₅₀ = 2 μ M) estão ilustrados na Figura 5.7.

Figura 5.7 - Mapa de contribuição HQSAR para o inibidor mais potente do conjunto de dados (**36**).

O mapa de contribuição indica que o anel benzamida substituído contribui favoravelmente para a atividade inibitória do composto. Essas observações estão de acordo com resultados previamente descritos na literatura⁴⁵ e são corroboradas pelo fato de que compostos do conjunto de dados com substituintes quinolina, bifenila na mesma posição (**1** e **2**, respectivamente) ou aqueles que não possuem o substituinte benzamida (**8-18**), apresentam diminuição substancial da potência inibitória. Além disso, os mapas de contribuição para os compostos **19-56** sugerem que padrões de substituição diferentes no anel benzamida são toleráveis pelo receptor e, simultaneamente, indicam contribuição favorável dos substituintes na posição *meta* do anel. Essa observação é corroborada pelos dados de SAR.⁴⁸ Por exemplo, o composto **36** ($IC_{50} = 2 \mu M$), com substituintes *m*-metoxila no anel benzamida apresenta atividade inibitória da ordem de 50 e 40 vezes maior do que os compostos *para*-substituídos **40** ($IC_{50} = 100 \mu M$) e **41** ($IC_{50} = 80 \mu M$), respectivamente.

Os resultados de HQSAR com os inibidores da GAPDH de *L. mexicana* foram combinados com a análise de QSAR clássico para o mesmo conjunto de dados, realizado em colaboração com o Prof. Dr. Marcelo S. Castilho da Faculdade de Farmácia da Universidade Federal da Bahia (UFBA). Uma vez que ambas as técnicas são complementares, a análise integrada dos modelos de QSAR 2D forneceu informações relevantes para a melhor compreensão dos dados de SAR, revelando a função essencial dos componentes estéreos e eletrostáticos para a atividade inibitória dos compostos do conjunto de dados.³⁶ Esses resultados foram organizados em um manuscrito intitulado *Classical and Hologram QSAR Studies on a Series of Inhibitors of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase* e submetido para o periódico científico *QSAR & Combinatorial Science*, o qual foi aceito para publicação (Apêndice A).

5.2.4 - QSAR 3D

Análise comparativa dos campos moleculares

A análise comparativa dos campos moleculares (CoMFA, do inglês, *comparative molecular field analysis*) é um método de QSAR 3D amplamente utilizado no processo de otimização estrutural de compostos líderes. A técnica baseia-se na premissa de que a atividade biológica está diretamente relacionada com as interações fármaco e receptor, ou seja, as moléculas bioativas geralmente produzem seus efeitos através de interações (não-covalentes) estereoquímicas e eletrostáticas com o alvo macromolecular. Como essas interações apresentam caráter tridimensional, este método considera as moléculas em três dimensões, utilizando campos moleculares de interação (MIFs, do inglês, *molecular interaction fields*) como descritores. Os campos moleculares são obtidos através do cálculo das energias de interação estereoquímicas e

eletrostáticas entre as moléculas do conjunto de dados e sondas virtuais. As sondas virtuais consistem de um determinado átomo apresentando uma carga de prova. As diferenças na propriedade-alvo (variável dependente) são, desta forma, descritas em função das diferenças nestes campos moleculares de interação.^{41,60-62}

Diferentemente do método HQSAR, as conformações e as cargas parciais atômicas das moléculas em estudo alteram de forma significativa o cálculo dos campos moleculares estereoquímicos e eletrostáticos (descritores 3D). Dessa forma, é necessário que sejam geradas conformações próximas as bioativas e que as cargas atômicas parciais sejam apropriadamente calculadas para as moléculas utilizadas na modelagem. Além disso, o método requer que as conformações das moléculas do conjunto de dados apresentem a mesma orientação no espaço cartesiano (alinhamento estrutural 3D). Estas são as principais observações que diferem o método CoMFA dos métodos de QSAR 2D.⁶⁰⁻⁶²

Uma vez gerado o alinhamento estrutural dos compostos em estudo, uma grade tridimensional reticulada suficientemente ampla para envolver todas as moléculas do conjunto de dados é posicionada e uma distância de intersecção entre os retículos da grade é definida (e.g., 1-2 Å). Em cada um desses pontos de intersecção são calculadas as energias de interação estereoquímicas e eletrostáticas entre a sonda virtual e os átomos próximos desses pontos. A sonda molecular pode ser definida, por exemplo, como um átomo virtual de carbono sp^3 carregado positivamente. Esse processo é realizado até que todos os pontos de intersecção sejam mapeados pela sonda.^{41,60}

As energias de interação calculadas em cada ponto correspondem a uma tabela contendo centenas de colunas de descritores (variáveis independentes), que são correlacionados com o valor da propriedade alvo (variável dependente) utilizando métodos de regressão multivariada. O método de regressão PLS é apropriado para esse propósito. Técnicas de validação cruzada são

utilizadas para investigar a robustez interna do modelo obtido. O modelo estatístico gerado correlaciona a propriedade alvo com os descritores e permite que novas moléculas ainda não testadas possam ter o valor da propriedade em estudo predito (Figura 5.8).^{41,60-62}

Figura 5.8 - Etapas envolvidas na construção de um modelo CoMFA.

Assim como no método HQSAR, o método CoMFA permite a avaliação visual das modificações estruturais relacionadas ao aumento do valor da propriedade-alvo através da análise de mapas de contorno. Esses mapas evidenciam regiões estereoquímicas e eletrostáticas ao redor das moléculas do conjunto de dados com contribuições favoráveis ou desfavoráveis para a propriedade-alvo (potência inibitória).

Análise comparativa dos índices de similaridades moleculares

Nos estudos comparativos de QSAR 3D os valores da propriedade-alvo podem ser razoavelmente descritos pela somatória dos diversos descritores moleculares.^{4,60} Conforme descrito anteriormente, no método CoMFA mudanças graduais nas propriedades de interações são mapeadas através da medida da energia potencial em pontos determinados ao redor das moléculas dos ligantes alinhados. A contribuição devido às forças de dispersão entre moléculas é descrita pelo potencial de Lennard-Jones e as propriedades eletrostáticas são caracterizadas pelo potencial de Coulomb (Figura 5.9). Diante disso, os campos moleculares calculados pelo método CoMFA apresentam algumas particularidades, como, por exemplo, o potencial de Lennard-Jones sofre uma variação extremamente abrupta próximo à superfície de van der Waals. Conseqüentemente, o potencial energético nos pontos próximos a superfície molecular sofrem alterações drásticas. Devido a este fato, pequenas alterações na orientação espacial ou pequenas diferenças conformacionais das moléculas em estudo (e.g., diferenças no alinhamento molecular) podem resultar em grandes variações nos descritores moleculares. Esta particularidade faz com que o método CoMFA seja altamente dependente do alinhamento molecular utilizado.⁶⁰ Além disso, tanto o potencial de Lennard-Jones quanto o potencial de Coulomb apresentam singularidades próximos às posições atômicas.⁶³

Figura 5.9 - Curva de potenciais utilizadas nos métodos de QSAR 3D.⁶⁰ O potencial de Lennard-Jones (verde) apresenta uma variação bastante abrupta próximo à superfície molecular, já o potencial de Coulomb (vermelho) apresenta uma variação mais suave, entretanto, ambos potenciais apresentam singularidades próximos as posições atômicas (nesta figura representado pela origem dos dados). Para evitar valores excessivamente elevados, a avaliação dos potenciais é restringida por um valor de exclusão arbitrário. A função tipo Gaussiana (azul) utilizada no método CoMSIA apresenta um termo dependente da distância mais suave que as outras funções, evitando assim qualquer tipo de singularidade.

Para evitar valores extremamente elevados de energia, a determinação dos potenciais é normalmente restrita as regiões ao redor das moléculas em estudo, empregando-se para tanto, um valor arbitrário de exclusão (do inglês, *cutoff*). Devido às diferenças na inclinação dos potenciais de Lennard-Jones e Coulomb, os valores de exclusão são excedidos para os diferentes termos em distâncias diferentes das moléculas em estudo, isso faz com que outros ajustes arbitrários sejam empregados para adequar simultaneamente os campos estérico e eletrostático, o que pode levar a perda de informação de um dos campos.⁶⁰

Os mapas de contorno gerados pelo método CoMFA são ferramentas extremamente úteis para a avaliação das contribuições espaciais dos diferentes campos moleculares. Entretanto, devido à inclinação acentuada dos potenciais próximos à superfície molecular e a necessidade do

emprego de um valor arbitrário de exclusão, esses mapas de contorno são frequentemente descontínuos, fato que pode comprometer sua interpretação.^{60,63}

Para contornar as particularidades inerentes ao método CoMFA uma técnica alternativa de análise de campos moleculares foi desenvolvida. O método de análise comparativa dos índices de similaridades moleculares (CoMSIA, do inglês, *Comparative Molecular Similarity Indices Analysis*)^{63,64} emprega uma estratégia diferente para calcular os descritores moleculares. Uma função tipo Gaussiana (sem singularidades) foi selecionada para descrever os termos energéticos em função da distância entre a sonda molecular e os átomos das moléculas em estudo. O emprego de uma função tipo Gaussiana define uma função mais “suave” para o cálculo do termo dependente da distância quando comparado ao potencial de Lennard-Jones e ao potencial de Coulomb (Figura 5.9). Dessa forma, no método CoMSIA não há necessidade do emprego de um valor arbitrário de exclusão, fazendo com que os índices de similaridades possam ser calculados em todos os pontos de interação e não somente ao redor das moléculas, como no método CoMFA. Os valores dos índices de similaridades, obtidos através do mapeamento das moléculas alinhadas no interior da grade tridimensional reticulada, não representam uma medida direta de similaridade entre as moléculas. Na realidade, esses valores são medidas indiretas da similaridade de cada uma das moléculas do conjunto de dados com a sonda molecular.^{63,64} Os índices de similaridade então obtidos são avaliados através da análise de PLS de modo similar ao método CoMFA.

Em princípio, qualquer propriedade físico-química relevante para a atividade biológica poderia ser considerada para calcular os campos referentes aos índices de similaridade. Entretanto, as propriedades estereoquímicas, eletrostáticas, hidrofóbicas e de ligação de hidrogênio (doador e aceptor) são as comumente empregadas, pois assume-se que este conjunto

de propriedades agrega as contribuições mais importantes responsáveis pelo reconhecimento molecular e a afinidade de ligação de uma molécula pequena e seu alvo biológico.^{63,64}

Enquanto os mapas de contorno gerados pelo método CoMFA evidenciam as regiões no espaço onde as moléculas alinhadas poderiam interagir favorável ou desfavoravelmente com um possível ambiente, os mapas CoMSIA apresentam as regiões na molécula do ligante onde a presença de um determinado grupo químico (e implicitamente sua propriedade físico-química) é favorável ou desfavorável para a atividade. Baseado na análise visual dos mapas CoMSIA é possível avaliar se todas as características importantes para a atividade biológica estão presentes nas moléculas dos ligantes em estudo.^{60,63,64}

A aplicação do método CoMSIA e CoMFA para um mesmo conjunto de dados resulta geralmente em modelos de significância estatística bastante similares, entretanto, as vantagens da utilização do método CoMSIA em relação ao método CoMFA estão relacionadas a: (i) menor sensibilidade em relação ao alinhamento molecular (função tipo Guassiana); (ii) possibilidade de incorporação de campos moleculares adicionais (campos hidrofóbicos e de ligação de hidrogênio); e (iii) mapas de contorno mais simples e de fácil interpretação.

5.2.5 - Estudos de QSAR 3D de enzimas GAPDH de tripanosomatídeos

Estratégias de alinhamento molecular

As enzimas GAPDH de *T. cruzi*,⁶⁵ *T. brucei*,⁶⁶ e *L. mexicana*⁶⁷ possuem alta identidade sequencial (~90 %) enquanto que apresentam diferenças estruturais significativas quando comparadas à enzima humana (~45 % de identidade sequencial), podendo, desta forma, serem exploradas para a inibição seletiva.³⁷ O alinhamento sequencial das GAPDHs revelou vários

resíduos que são altamente conservados entre as estruturas dos parasitas e que diferem substancialmente daqueles observados na enzima humana (Figura 5.10). Assim, é provável que estudos de SAR, QSAR e técnicas de SBDD possam ser aplicadas para o desenvolvimento de inibidores potentes e seletivos das enzimas GAPDH de tripanosomatídeos utilizando como base os dados de inibição e a estrutura da enzima de um dos parasitas.³⁷

Figura 5.10 - Alinhamento seqüencial das enzimas GAPDH de *L. mexicana*, *T. cruzi*, *T. brucei*, e *H. sapiens*. Os números das seqüências correspondem à seqüência da GAPDH de *L. mexicana*. Resíduos conservados entre as GAPDHs dos diferentes organismos estão coloridos em cinza.

A determinação do alinhamento molecular é uma etapa crucial para os estudos de QSAR 3D.^{68,69} Diante disso, estudos de docagem molecular foram conduzidos utilizando-se os programas GOLD 3.0¹⁶ e FlexX 2.1¹⁷, para gerar os alinhamentos moleculares. As coordenadas cristalográficas da GAPDH de *L. mexicana* em complexo com o inibidor N⁶-(1-metil-nafil)-2'-deoxi-2'-(3,5-dimetoxi benzamida) adenosina (PDB, 1I32)⁷⁰ (NMDBA, composto **36**, Tabela 5.1) ligado no sítio ativo foi utilizada como alvo molecular. Os átomos de hidrogênio foram

adicionados na geometria padrão utilizando o módulo BIOPOLYMER implementado no SYBYL 7.3. Os resíduos de histidinas, glutaminas e asparaginas presentes na cavidade de ligação foram manualmente checados quanto à orientação, estado de protonação e tautômeros mais prováveis com o programa PYMOL 0.99. O sítio ativo para estudos de docagem molecular foi definido incorporando todos os resíduos de aminoácidos presentes no interior de uma esfera de 8 Å de raio, posicionada ao redor da molécula do ligante NMDBA. Os parâmetros padrões dos programas de docagem molecular foram utilizados para obtenção dos alinhamentos moleculares.

As coordenadas cristalográficas do inibidor NMDBA⁷⁰ foram utilizadas para validar os resultados da docagem molecular. Os programas FlexX e GOLD posicionaram a molécula do inibidor numa conformação bastante próxima àquela observada experimentalmente (Figura 5.11A e 5.11B, respectivamente). O valor de rmsd referente à sobreposição de todos os átomos do inibidor NMDBA entre a conformação cristalográfica e a conformação melhor pontuada pelos programas FlexX e GOLD foi de 1.43 Å e 1.39 Å, respectivamente. Baseado nestes resultados, os alinhamentos moleculares foram obtidos aplicando-se o mesmo protocolo padrão de docagem molecular, e somente as conformações melhores pontuadas para cada inibidor foram consideradas para o desenvolvimento de modelos de QSAR 3D. Os alinhamentos moleculares utilizados nos estudos de modelagem estão apresentados na Figura 5.11C e 5.11D. As conformações melhores pontuadas estão razoavelmente alinhadas no sítio de ligação da GAPDH de *L. mexicana*, confirmando a robustez dos procedimentos de docagem molecular utilizados.^{22,71,72}

Figura 5.11 - Estrutura cristalográfica do inibidor NMDBA (verde) ligado no sítio ativo da GAPDH de *L. mexicana* (PDB, 1I32)⁷⁰ comparada com as conformações melhores pontuadas do próprio NMDBA (cinza) geradas pelos programas FlexX (A) e GOLD (B) durante a etapa de validação do procedimento. Alinhamento molecular do conjunto de dados obtidos com o FlexX (C) e GOLD (D) ilustradas no interior do sítio ativo da GAPDH de *L. mexicana*.

Modelos CoMFA e CoMSIA

Com o objetivo de investigar as propriedades químicas e estruturais relacionadas à potência inibitória do conjunto de dados de inibidores da GAPDH de tripanosomatídeos, os alinhamentos moleculares obtidos foram quantitativamente analisados utilizando-se os métodos CoMFA e CoMSIA. Os dois alinhamentos geraram quatro modelos de QSAR, a saber: CoMFA_{FlexX}, CoMFA_{GOLD}, CoMSIA_{FlexX} e CoMSIA_{GOLD}.

Os modelos CoMFA e CoMSIA foram criados com base nos valores calculados de energias de interação entre sondas moleculares apropriadas e o átomos dos ligantes alinhados no interior de uma grade tridimensional reticulada, com uma distância de 2 Å entre as intersecções dos retículos. A grade tridimensional envolveu todos os compostos do conjunto de dados alinhados com uma margem de segurança de 4 Å nos vértices nas direções x, y e z. Um carbono sp^3 com carga +1 foi utilizado para calcular os campos CoMFA estéreo e eletrostático de acordo com os parâmetros padrão do SYBYL (campo de força Tripos, constante dielétrica 1/r e valor de exclusão 30 kcal/mol). Os campos CoMSIA foram calculados para as propriedades estereo, eletrostáticas, hidrofóbicas e ligação de hidrogênio empregando-se sonda molecular com carga +1, raio atômico 1, hidrofóbica, propriedade de ligação de hidrogênio +1, e fator de atenuação da função Gaussiana dependente da distância de 0,3. Os valores de pIC_{50} foram utilizados como variável dependente e correlacionados com os campos moleculares (descritores) através do método de validação cruzada LOO. Os resultados estatísticos das análises estão apresentados na Tabela 5.7.

Tabela 5.7 - Parâmetros estatísticos das análises CoMFA e CoMSIA conduzidas nos alinhamentos FlexX e GOLD.

Parâmetros Estatísticos	FlexX		GOLD	
	CoMFA	CoMSIA	CoMFA	CoMSIA
q^2_{LOO}	0,83	0,72	0,82	0,79
r^2	0,93	0,92	0,93	0,93
SEE	0,27	0,30	0,29	0,29
F	259	153	218	167
PC	3	4	3	4
Fração (%)				
S	44	18	48	15
E	56	26	52	27
H	-	26	-	26
doador-H	-	18	-	21
aceptor-H	-	12	-	11

q^2_{LOO} , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; F , valor do teste- F ; PC, número ótimo de componentes principais; S, contribuição dos campos estéreos, E, contribuição dos campos eletrostáticos; H, contribuição dos campos hidrofóbicos.

Os alinhamentos FlexX e GOLD geraram modelos CoMFA estatisticamente significativos, conforme indicado pelos valores dos coeficientes r^2 e q^2 ($\text{CoMFA}_{\text{FlexX}}$, $r^2 = 0,93$ e $q^2_{\text{LOO}} = 0,83$; $\text{CoMFA}_{\text{GOLD}}$, $r^2 = 0,93$ e $q^2_{\text{LOO}} = 0,82$). Os modelos CoMSIA gerados para o mesmo conjunto treinamento também apresentam parâmetros estatísticos razoáveis ($\text{CoMSIA}_{\text{FlexX}}$, $r^2 = 0,92$ e $q^2_{\text{LOO}} = 0,72$; $\text{CoMSIA}_{\text{GOLD}}$, $r^2 = 0,93$ e $q^2_{\text{LOO}} = 0,79$). Apesar dos modelos CoMSIA apresentarem resultados estatísticos ligeiramente menores, os modelos construídos são estatisticamente relevantes. Resultados similares são observados na literatura para outros conjuntos de dados.^{63,64,73-75} Os resultados da Tabela 5.7 indicam que os modelos de QSAR 3D não foram substancialmente afetados pelos diferentes alinhamentos. Este resultado pode estar relacionado ao fato de que as moléculas do conjunto treinamento agregam informação estrutural suficiente para o reconhecimento molecular privilegiado pelo sítio de ligação da GAPDH de *L. mexicana*, gerando assim, campos moleculares similares que explicam a variância dos modelos de QSAR 3D.

As avaliações de significância estatística e estabilidade dos modelos construídos através do protocolo LMO (LMO₁₀ e LMO₅) confirmaram a estabilidade e robustez dos modelos de QSAR 3D, uma vez que os resultados de LMO são similares àqueles previamente obtidos pela análise LOO (Tabela 5.8).

Tabela 5.8 - Parâmetros estatísticos obtidos durante os procedimentos de validação dos modelos CoMFA e CoMSIA.

Parâmetros Estatísticos	FlexX		GOLD	
	CoMFA	CoMSIA	CoMFA	CoMSIA
$q^2_{\text{LMO}_{10}}$ ^a	0,82 ± 0,01	0,72 ± 0,02	0,82 ± 0,01	0,79 ± 0,01
$q^2_{\text{LMO}_5}$ ^b	0,82 ± 0,01	0,70 ± 0,03	0,82 ± 0,01	0,78 ± 0,02
r^2_{boot} ^c	0,95 ± 0,01	0,95 ± 0,01	0,95 ± 0,01	0,95 ± 0,01
SEE _{boot} ^c	0,24 ± 0,11	0,25 ± 0,11	0,23 ± 0,11	0,23 ± 0,11
r^2_{pred} ^d	0,86	0,83	0,80	0,78

^aValor médio de q^2 e SDEP obtidos após 25 análises independentes do protocolo de validação cruzada com 10% dos compostos do conjunto treinamento. ^bValor médio de q^2 e SDEP obtidos após 25 análises independentes do protocolo de validação cruzada com 20% dos compostos do conjunto treinamento. ^cValor médio de correlação da reamostragem (r^2_{boot}) e erro padrão de correlação da reamostragem (SEE_{boot}) obtidos na análise de reamostragem após 100 simulações de amostragem. ^dValor de r^2 -preditivo dos modelos de QSAR para os compostos do conjunto teste (validação externa).

A possibilidade de geração de modelos aleatórios foi verificada pela permutação dos valores da propriedade biológica entre os compostos (do inglês, *progressive scrambling*).⁷⁶ Os resultados indicaram que os modelos são consistentes e a estabilidade dos modelos foi confirmada. Para validação da precisão dos modelos, a análise estatística de reamostragem (do inglês, *bootstrapping*)^{77,78} foi empregada para a determinação do coeficiente de correlação da reamostragem (r^2_{boot}) e o erro padrão de correlação da reamostragem (SEE_{boot}) para cada modelo. O valor médio de r^2_{boot} de 0,95 obtido para os modelos CoMFA e CoMSIA indica que os modelos são bastante precisos (Tabela 5.8). Os resultados dos testes de validação interna indicam que os modelos de QSAR construídos são bastante estáveis e possuem alta significância estatística. A validação externa dos modelos de QSAR 3D foi realizada com o conjunto teste de 14 compostos

que foram completamente excluídos das etapas de construção dos modelos. Os resultados da validação externa estão apresentados na Tabela 5.9 e na Figura 5.12.

Tabela 5.9 - Valores experimentais, preditos e residuais de pIC_{50} calculados com os modelos de QSAR 3D para o conjunto teste de 14 compostos inibidores da GAPDH de tripanosomatídeo.

Cpd	Exp ^a	FlexX				GOLD			
		CoMFA		CoMSIA		CoMFA		CoMSIA	
		Predito	Resíduo ^b	Predito	Resíduo ^b	Predito	Resíduo ^b	Predito	Resíduo ^b
57	2,62	3,18	-0,56	3,31	-0,69	3,11	-0,49	2,89	-0,27
58	3,52	3,45	0,08	3,65	-0,13	3,42	0,10	3,26	0,26
59	3,30	3,30	0,00	3,17	0,13	2,68	0,62	2,80	0,50
60	3,22	2,95	0,27	3,13	0,09	3,38	-0,16	3,17	0,05
61	3,07	3,17	-0,10	3,29	-0,22	3,17	-0,10	3,11	-0,04
62	3,82	3,60	0,22	3,27	0,55	3,08	0,74	3,28	0,54
63	5,30	4,94	0,36	5,25	0,05	4,86	0,44	5,20	0,10
64	5,40	4,93	0,47	4,93	0,47	4,88	0,52	4,63	0,77
65	5,70	5,66	0,04	5,42	0,28	5,53	0,17	5,30	0,40
66	4,30	4,62	-0,32	4,75	-0,45	4,96	-0,66	5,05	-0,75
67	4,60	5,07	-0,47	4,91	-0,31	4,66	-0,06	5,04	-0,44
68	4,00	4,56	-0,56	4,45	-0,45	4,36	-0,36	4,40	-0,40
69	4,74	4,46	0,28	4,27	0,47	4,84	-0,10	5,12	-0,38
70	5,00	4,57	0,43	4,44	0,56	4,55	0,46	4,46	0,54

^aValores experimentais de pIC_{50} ; ^bDiferença entre o valor de pIC_{50} experimental e predito.

Valores de r^2_{pred} de 0,86; 0,83; 0,80 e 0,78 foram obtidos para os modelos CoMFA_{FlexX}, CoMSIA_{FlexX}, CoMFA_{GOLD} e CoMSIA_{GOLD}, respectivamente, indicando concordância significativa entre os dados experimentais e preditos (Tabela 5.8). A qualidade estatística, estabilidade e substancial capacidade preditiva dos modelos construídos indicam que estes podem ser empregados para o planejamento de novos inibidores das enzimas GAPDH de tripanosomatídeos.

Figura 5.12 - Gráfico dos valores experimentais *versus* preditos de pIC_{50} obtido com os modelos de QSAR 3D para o conjunto treinamento (quadrados) e conjunto teste (triângulos) de inibidores da GAPDH de tripanosomatídeo.

O conhecimento do padrão de interação entre enzima e inibidor é um requisito importante para o planejamento de novas moléculas mais potentes.¹ Assim, além de prever satisfatoriamente o valor da atividade de interesse, os modelos de QSAR 3D também fornecem informações gráficas extremamente úteis relacionadas com a propriedade-alvo. Umas das

características mais exploradas dos métodos de QSAR 3D é a capacidade de visualizar as regiões no espaço responsáveis pelo aumento ou diminuição da variável dependente (e.g., pIC_{50} , pK_i , pEC_{50}). Diante disso, considerando-se que os dois alinhamentos geraram modelos estatisticamente similares, a discussão dos mapas de contorno CoMFA e CoMSIA será realizada baseada nos modelos derivados do alinhamento FlexX, uma vez que estes modelos apresentaram resultados estatísticos ligeiramente melhores e maior capacidade preditiva.³⁷

Análise dos mapas de contorno CoMFA

O mapa de contorno estéreo CoMFA está apresentado na Figura 5.13A. O contorno verde ao redor do substituinte naftila do inibidor **36** ($pIC_{50} = 5,70$) indica que grupos volumosos nesta região são favoráveis para o aumento da potência inibitória, enquanto que uma região desfavorável para a presença de grupos volumosos (contorno amarelo) é observada próxima a posição 8 do anel adenina, indicando que substituintes desta natureza devem ser evitados nesta posição. O modo de ligação proposto para o composto **23** ($pIC_{50} = 2,52$) orienta o grupo bromo nesta região desfavorável. Assim, a ausência de grupos volumosos próximos as regiões favoráveis combinado com a presença de substituintes na região desfavorável poderiam explicar a baixa potência exibida pelo composto **23**. O mapa CoMFA de contorno eletrostático está apresentado na Figura 5.13B. O mapa ilustra três regiões favoráveis para a presença de grupos eletropositivos (contorno azul). Duas dessas regiões estão posicionadas em locais onde o sítio de ligação está exposto ao solvente, uma próxima a posição 2 do anel naftila e outra próxima à hidroxila 6' do anel ribose. A terceira região favorável encontra-se adjacente ao átomo de enxofre da cadeia lateral da Met39, indicando que a presença de grupos eletropositivos nesta região poderia aumentar a potência inibitória do composto. Além disso, duas regiões favoráveis para grupos

eletronegativos são observadas na Figura 5.13B, uma próxima a posição N6 do anel adenina que está exposta ao solvente, e outra localizada adjacente ao grupo hidroxila da cadeia lateral da Ser40. O oxigênio do substituinte 3-metoxila do potente composto **36** está orientado nesta região favorável para grupos eletronegativos. O mapa de contorno eletrostático concorda com os dados de SAR, uma vez que a ausência de grupos eletronegativos no anel benzamida (compostos **19** e **23**, Tabela 5.1) diminui a potência inibitória desta classe de inibidores da GAPDH de tripanosomatídeos.^{45,46}

Figura 5.13 - Mapas de contorno CoMFA obtidos com o modelo CoMFA_{FlexX}. Os inibidores **36** (azul claro) e **23** (laranja) estão representados como modelos de bastões e esferas no interior do sítio ativo da GAPDH de *L. mexicana*. O sítio ativo está representado contornado pela superfície de Connolly. (A) Mapa de contorno estérico está indicado em amarelo e verde (-0,005 e 0,035 kcal/mol, respectivamente). (B) O mapa de contorno eletrostático está indicado em vermelho e azul (-0,05 e 0,05 kcal/mol, respectivamente).

Análise dos mapas de contorno CoMSIA

O ambiente químico do sítio de ligação da GAPDH foi mapeado por cinco sondas moleculares implementadas no método CoMSIA (estéreo, eletrostático, hidrofóbico, doador e acceptor de ligação de hidrogênio).³⁷ A Figura 5.14A mostra o mapa de contorno estéreo. O mapa apresenta contornos similares àqueles previamente observados no mapa CoMFA e identifica duas regiões favoráveis para a presença de grupos volumosos (contorno verde). A primeira região está localizada entre as cadeias laterais dos resíduos Met39 e Arg92, enquanto a segunda encontra-se na porção inferior da cavidade de ligação, no então denominado “bolsão de seletividade”,⁴⁴ entre as cadeias laterais da Met39 e Val206. Compostos contendo grupos volumosos em ambas as regiões favoráveis apresentam maior potência inibitória (compostos **26-55**, **63-70**, Tabela 5.1).⁴⁶⁻⁴⁸ O mapa de contorno estéreo desfavorável também indica o substituinte 8-bromo do composto **23** como sendo prejudicial para a potência inibitória.

O mapa de contorno eletrostático ilustrado na Figura 5.14B indica que inibidores contendo grupos eletronegativos nas regiões contornadas em vermelho ou grupos eletropositivos nas regiões contornadas em azul poderiam apresentar maior potência inibitória. O substituinte 3-metoxila do composto **36** orienta o átomo de oxigênio próximo a região favorável para grupos eletronegativos enquanto o átomo de carbono está localizado próximo a região favorável para grupos eletropositivos.

Os contornos amarelos e brancos observados na Figura 5.14C indicam regiões no sítio de ligação favoráveis para a presença de grupos hidrofóbicos e hidrofílicos, respectivamente. Duas regiões favoráveis para grupos hidrofóbicos são indicadas, uma localizada próxima aos átomos de carbono das cadeias laterais da Met39 e Arg92, e outra localizada no “bolsão de seletividade” entre as cadeias laterais da Met39 e Val206. Os substituintes naftila e 3-metoxila do composto **36**

estão orientados nestas regiões favoráveis para grupos hidrofóbicos. Essas regiões se sobrepõem às regiões favoráveis identificadas no mapa de contorno estéreo. Assim, a análise combinada dos mapas CoMSIA sugere que inibidores derivados de adenosina apresentando substituintes volumosos e com caráter hidrofóbico nas posição N6 da adenina e C3' da ribose seriam moléculas potentes inibidoras da GAPDH de tripanosomatídeos.

Os mapas de contorno indicando as regiões favoráveis e desfavoráveis para ligação de hidrogênio estão apresentados na Figura 5.14D. O contorno em ciano próximo ao grupo carboxila da cadeia lateral do resíduo Asp38 indica que a presença de grupos doadores de ligação de hidrogênio próximo a esta região seria favorável para a potência inibitória. De fato, o grupo carboxila do Asp38 representa um ponto importante de interação para o reconhecimento molecular entre as enzimas GAPDH de tripanosomatídeos.⁷⁰ Um contorno magenta, favorável para grupos aceptores de ligação de hidrogênio pode ser observado próximo à cadeia principal do resíduo Gln91, sugerindo esta região como sendo um ponto atrativo para ligação de hidrogênio.

Figura 5.14 - Mapas de contorno CoMSIA obtidos com o modelo CoMSIA_{FlexX}. Os inibidores **36** (azul claro) e **23** (laranja) estão representados como modelos de bastões e esferas no interior do sítio ativo da GAPDH de *L. mexicana*. O sítio ativo está representado contornado pela superfície de Connolly. Mapa de contorno estérico está indicado em amarelo e verde (-0,0014 e 0,008 kcal/mol, respectivamente). **(B)** Mapa de contorno eletrostáticos está indicado em vermelho e azul (-0,03 e 0,02 kcal/mol, respectivamente). **(C)** Mapa de contorno hidrofóbico está indicado em branco e amarelo (-0,01 e 0,015 kcal/mol, respectivamente). **(D)** Mapa de contorno doador de ligação de hidrogênio está indicado em roxo e ciano (-0,015 e 0,02 kcal/mol, respectivamente) e o mapa de contorno aceptor de ligação de hidrogênio está indicado em vermelho e magenta (-0,07 e 0,015 kcal/mol, respectivamente).

Mapas de contorno de seletividade

O planejamento de novos inibidores baseado na estrutura da GAPDH explora as diferenças estruturais entre as enzimas dos parasitas e a humana.⁷⁹ Baseado neste conceito, são descritos na literatura inibidores derivados de adenosina seletivos para a GAPDH de *L. mexicana*, *T. brucei* e *T. cruzi*.⁴⁴⁻⁴⁸ Com o objetivo de obter informações significantes para a melhor compreensão das bases moleculares envolvidas na inibição seletiva dessa classe de moléculas, foi realizado um estudo detalhado dos campos moleculares derivados das análises de QSAR 3D. Para tanto, uma análise criteriosa dos mapas de contorno estéreo e eletrostático obtidos com o modelo CoMFA_{FlexX} (melhor modelo construído) foi realizada levando-se em consideração a organização estrutural e o ambiente químico das cavidades de ligação das GAPDHs de *L. mexicana*, *T. brucei*, *T. cruzi* e humana.

Segundo o modelo CoMFA construído, os campos moleculares eletrostático explicam 56% da variância dos dados enquanto os campos estéreos explicam os 44% restantes (Tabela 5.7). A região favorável para a presença de grupos eletronegativos (contorno vermelho, Figura 5.15) é complementar ao ambiente químico observado nos sítios de ligação das GAPDHs dos parasitas, uma vez que o mapa de contorno eletrostático está localizado próximo ao grupo NH₂ da cadeia lateral dos resíduos Asn (Asn40 em *T. cruzi* e Asn39 em *T. brucei*) e próxima ao grupo OH da cadeia lateral da Ser40 na GAPDH de *L. mexicana*. Entretanto, essa região favorável para grupos eletronegativos está próxima ao grupo carbonila da cadeia principal da Phe37 na GAPDH de humano, sugerindo que substituintes eletronegativos na posição *para* do anel benzamida poderiam apresentar maior seletividade para as enzimas dos parasitas. O mapa de contorno estéreo (verde, Figura 5.15) também apresenta boa complementaridade com os sítios de ligação das GAPDHs de tripanosomatídeos e indica que substituintes volumosos seriam capazes de

interagir na região entre as cadeias laterais da Arg92 e Met39 (Met38 em *T. brucei*). Essa região na GAPDH humana apresenta impedimento estérico em função da cadeia lateral volumosa do resíduo Phe37. Portanto, compostos com substituintes volumosos no anel naftila também são interessantes para a obtenção de compostos com maior seletividade para as enzimas dos parasitas.³⁷

Figura 5.15 - Sobreposição estrutural dos α das GAPDHs de *L. mexicana* (ciano, PDB, 1I32), *T. brucei* (amarelo, PDB, 1GGA), *T. cruzi* (cinza, PDB, 1ML3) e humana (magenta, PDB, 1U8F) apresentados simultaneamente com os mapas de contorno estérico (verde, favorável para grupos volumosos, 0,035 kcal/mol) e eletrostático (vermelho, favorável para grupos eletronegativos, -0,05 kcal/mol) derivados do modelo CoMFA_{FlexX}. As moléculas de NAD⁺ e do inibidor **36** estão apresentadas como modelos de bastões e esferas e coloridas de acordo com a estrutura que estão ligadas.

Os resultados dos estudos de QSAR 3D para as enzimas GAPDH de tripanosomatídeos foram organizados em um manuscrito intitulado *Structural Basis for Selective Inhibition of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase: Molecular Docking and 3D QSAR Studies* e submetidos para o periódico científico *Journal of Chemical Information and Modeling*, o qual foi aceito para publicação (Apêndice B).

5.3 - Triagem Virtual de Compostos

Uma estratégia promissora para a identificação de novos ligantes surgiu no final da década de 80 quando foram combinados os métodos de ensaios biológicos automatizados em larga escala (HTS, do inglês, *high-throughput screening*) e a química combinatória. A expectativa de obtenção de um número elevado de novos compostos líderes era bastante grande, pois se acreditava que a integração dessas técnicas resultaria num aumento substancial no número de novos fármacos lançados no mercado. Entretanto, a expectativa inicial não foi contemplada da forma esperada como pode ser observado nos baixos índices de novos ligantes descobertos e nos altos custos envolvidos no processo.^{4,80} Este cenário propiciou o desenvolvimento de novas técnicas, como a triagem virtual de compostos (VS, ensaio virtual). Esta técnica, fundamentada em métodos computacionais, surgiu como estratégia alternativa para a descoberta de novos compostos com potencial terapêutico.

VS é uma estratégia que envolve a análise de grandes bases de dados virtuais de compostos através de métodos computacionais com o objetivo de identificar um pequeno subgrupo de compostos que podem ser adquiridos através de representantes comerciais ou de síntese orgânica para avaliação da atividade biológica.⁸¹ Diferentemente do HTS, que necessita de um investimento substancial e disponibilidade física de coleções de moléculas, no VS os compostos em estudo não precisam, num primeiro momento, necessariamente existir. Além disso, esta técnica não requer a utilização de tecnologias sofisticadas de automação e robótica como no HTS. Por outro lado, o VS exige como pré-requisito conhecimento dos componentes espaciais e energéticos responsáveis pela interação entre um o ligante e o alvo molecular em estudo. Assim, tanto a estrutura tridimensional do alvo macromolecular como a estrutura dos

ligantes devem ser previamente conhecidas e estarem disponíveis para a realização do ensaio virtual.^{1,4}

Com o advento dos projetos genoma e proteoma o conhecimento da estrutura tridimensional de uma grande variedade de macromoléculas torna-se cada vez mais disponíveis. Grande parte deste conhecimento é armazenado em base de dados de acesso público como o PDB onde encontra-se atualmente depositadas as coordenadas tridimensionais de ~50.000 macromoléculas, muitas delas consideradas potenciais alvos terapêuticos.

O VS é uma estratégia tipicamente baseada no conhecimento (do inglês, *knowledge-based*). A primeira etapa deste processo consiste na obtenção da estrutura 3D do alvo terapêutico em potencial. Subseqüentemente, a estrutura resolvida é utilizada como alvo molecular para a descoberta de novos ligantes. As moléculas dos potenciais ligantes são depositadas em base de dados de compostos as quais são empregadas em estudos de docagem molecular para a construção de modelos de interação entre os ligantes e o alvo macromolecular. Um pré-requisito para o sucesso desse método é um conhecimento detalhado das propriedades estruturais do alvo molecular e as características energéticas que determinam o processo de reconhecimento molecular da molécula do ligante.⁸²

A etapa crítica do um ensaio virtual é a seleção de ferramentas computacionais robustas capazes de gerar resultados confiáveis.^{4,80} Portanto, a escolha do método de docagem molecular mais apropriado é fundamental para obtenção de modos de interação plausíveis e estimativas coerentes da afinidade de ligação das moléculas em estudo.¹ Exemplos de desenvolvimento e emprego de estratégias de VS direcionadas e apropriadas para cada caso resultaram na identificação de novas moléculas com capacidade inibitória de ordem nano e submicromolar para enzimas como a anidrase carbônica II humana,⁸⁰ aldose redutase humana⁸² e a tRNA-guanina transglicosilase de *Shigellae*.^{83,84}

5.3.1 - Triagem virtual de inibidores da GAPDH de *T. cruzi*

Essa parte do trabalho foi realizada durante o período de doutorado-sanduíche no Institut für Pharmazeutische-Chemie, Philipps Universität - Marburg, Alemanha, sob a orientação do Prof. Dr. Gerhard Klebe.

A estratégia de VS utilizada consiste na aplicação de vários filtros hierárquicos consecutivos direcionados para as características do sítio ativo da GAPDH de *T. cruzi*. Em seguida, os compostos selecionados são orientados na cavidade de ligação através de docagem molecular e como última etapa são aplicados os critérios de seleção para aquisição das moléculas.

Sítio de ligação

Até o presente momento há depositado no PDB três estruturas da enzima GAPDH de *T. cruzi* em complexo com diferentes inibidores. Um complexo com o derivado de produto natural chalepina (PDB, 1K3T)⁸⁵, um complexo com um análogo irreversível do substrato G3P (PDB, 1ML3)⁸⁶ e um complexo com um análogo reversível do substrato G3P (PDB, 1QXS)⁸⁷. A análise desses complexos cristalográficos revelou características estruturais importantes para a seleção da estrutura mais adequada para o ensaio virtual.

A enzima GAPDH utiliza para a catálise bioquímica uma molécula de substrato e uma molécula de cofator conforme descrito na seção 1.3.2. Dessa forma, essa enzima apresenta dois sítios distintos de ligação, um para a molécula do substrato e outro para a molécula do cofator. As diferenças estruturais entre o sítio de ligação do cofator da enzima GAPDH de *L. mexicana* e seu homólogo humano foram exploradas e inibidores potentes e seletivos derivados da adenosina foram obtidos.⁴⁴⁻⁴⁸

A análise dos complexos cristalográficos da GAPDH de *T. cruzi* revelou um alto grau de conservação da estrutura terciária do domínio de ligação do substrato. A sobreposição dos Ca destas estruturas apresenta valores de rmsd entre 0,2-0,4 Å, indicando um pequeno rearranjo estrutural do sítio de ligação do substrato entre os diferentes complexos da GAPDH de *T. cruzi*. Já a análise sequencial das GAPDHs de *T. cruzi* e humana mostra que o resíduo Asp210 da enzima do parasita é substituída por um resíduo Leu194 na enzima humana (Figura 5.10). No presente estudo, essa diferença estrutural é explorada para a descoberta de novos inibidores da GAPDH de *T. cruzi*. Além disso, a análise estrutural da enzima do parasita revelou a presença de uma cavidade adjacente ao resíduo Asp210 que poderia ser preenchida por grupos volumosos da molécula do ligante (Figura 5.16).

Figura 5.16 - Sítio de ligação do substrato da enzima GAPDH de *T. cruzi* em complexo com um análogo reversível do substrato (S70) (PDB, 1QXS). Os átomos de oxigênio do resíduo Asp210 (Leu194 na GAPDH humana) estão coloridos em vermelho e em destaque a cavidade próxima ao sítio ativo que pode ser explorada para a descoberta de novos inibidores.

Base de dados de compostos

A escolha da base de dados para a realização do VS levou em consideração a facilidade para aquisição das moléculas pertencentes à base de dados. Diante disso, a base de dados de compostos selecionada foi o subgrupo ZINC *drug-like*.⁸⁸ Esse sub-grupo contém ~2.000.000 de compostos que satisfazem os critérios indicados para biodisponibilidade oral (≤ 5 doadores de ligação de hidrogênio, ≤ 10 aceptores de ligação de hidrogênio, massa molecular ≤ 500 e coeficiente de partição $\log P \leq 5$).⁸⁹ Além disso, essa base de dados fornece suporte para múltiplos estados de protonação, formas tautoméricas, e estereoisômeros.

Filtragem preliminar por grupos funcionais

A análise do sítio de ligação do substrato na GAPDH de *T. cruzi* e o conhecimento do mecanismo bioquímico evidenciaram a importância do resíduo Arg249 para a atividade catalítica da enzima.⁸⁶ Localizado na porção central do sítio de ligação do substrato, o grupo guanidina da Arg249, que em pH fisiológico encontra-se na forma protonada, poderia ser explorado como grupo essencial para a interação com moléculas de inibidores que apresentassem grupos aniônicos. Dessa forma, foi realizada uma busca por grupos químicos carregados negativamente mais freqüentemente encontrados interagindo com grupos guanidinas de argininas com o auxílio do programa ISOSTAR.⁹⁰

ISOSTAR consiste em uma ferramenta para análise de base de dados de interações intermoleculares que utiliza informações provenientes dos bancos de dados CSD (*Cambridge Structural Database*) e PDB. Para a seleção dos grupos químicos carregados negativamente que interagem com grupo guanidina de argininas foi utilizado o banco de dados ISOSTAR de

complexos cristalográficos depositados no PDB. A Tabela 5.10 apresenta a estrutura química e respectivos códigos SLN (*Sybyl Line Notation*) dos grupos funcionais selecionados de acordo com os critérios estabelecidos.

Tabela 5.10 - Grupos funcionais freqüentemente encontrados interagindo com grupo guanidina de argininas em complexos cristalográficos.

Grupo funcional	Código SLN
C(=O)O ⁻
S(=O)(O ⁻)O ⁻
P(=O)(O ⁻)O ⁻
P(O ⁻)O ⁻

Uma vez determinado os grupos funcionais mais favoráveis para interação com o resíduo Arg249 (Tabela 5.10), o código SLN destes grupos foi utilizado na primeira etapa do processo de filtragem da base de dados de compostos. Uma busca 2D com o programa UNITY²³ foi realizada para a seleção de moléculas que apresentavam os grupos funcionais pré-selecionados.

Ao final da aplicação do filtro molecular 2D a base de dados inicialmente com ~2.000.000 moléculas foi reduzida para ~79.000 moléculas que possuíam os grupos aniônicos de interesse.

Modelo farmacofórico baseado na estrutura do alvo molecular

A etapa seguinte da filtragem hierárquica envolveu o desenvolvimento de um modelo farmacofórico onde as moléculas presentes na base de dados deveriam satisfazer as restrições impostas pelo modelo. Esta etapa do estudo baseia-se no princípio que o microambiente do sítio de ligação determina as características principais que os ligantes em potencial devem apresentar para o reconhecimento molecular e assim exercer sua função biológica ou terapêutica.⁹¹

A análise de *Hot Spots* foi empregada para o mapeamento das regiões do sítio de ligação mais favoráveis para a interação com determinados grupos funcionais. Este mapeamento é realizado com o auxílio de sondas moleculares que apresentam propriedades físico-químicas comumente observadas em moléculas de ligantes.⁹² Dois métodos distintos foram empregados para realizar essa análise, DrugScore²⁷ e GRID⁹³. O primeiro método emprega uma abordagem baseada no conhecimento derivado das interações entre pares de potenciais de ligantes e proteínas observadas em complexos cristalográficos, enquanto o segundo utiliza uma técnica baseada em campos de força para a avaliação das energias de interação.

Na análise de *Hot Spots* uma grade reticulada é colocada e posicionada ao redor do sítio de ligação da macromolécula em estudo. Em seguida as energias de interação entre os resíduos presentes no sítio de ligação e as diferentes sondas moleculares são avaliadas nos diversos pontos da grade reticulada. As sondas moleculares, representadas por átomos ou grupos químicos, simulam as características físico-químicas dos substituintes encontrados nas moléculas bioativas. A Figura 5.17 apresenta um esquema geral das etapas envolvidas na análise de *Hot Spots*. Uma das principais características deste método é a possibilidade de identificação das regiões do sítio de ligação mais favoráveis para a interação (*Hot Spots*) com um determinado tipo de átomo através da visualização de mapas de contorno em diferentes níveis de energia.⁹²

Figura 5.17 - Esquema geral da análise de *Hot Spots*.

A análise de *Hot Spots* foi realizada no sítio de ligação da GAPDH de *T. cruzi* em complexo com um análogo reversível do substrato que apresenta grupos carregados negativamente (i.e., fosfatos e fosfonato) interagindo com o grupo guanidina da Arg249 (PDB, 1QXS).⁸⁶ Para a análise com o método GRID foram considerados as sondas moleculares: NH de amida (doador de ligação de hidrogênio) O de carbonila (aceptor de ligação de hidrogênio) e a sonda DRY para determinação de interações hidrofóbicas. A análise realizada com o método DrugScore utilizou como sondas moleculares: N.am (nitrogênio de grupos amida, doador de ligação de hidrogênio), O.2 (oxigênio sp^2 , acceptor ligação de hidrogênio) e C.3 (carbono sp^3 , determinação dos campos estéreo). A Figura 5.18 ilustra os campos de interação obtidos através das análises de *Hot Spots* no sítio de ligação da GAPDH de *T. cruzi*.

Figura 5.18 - Análise de *Hot Spots* do sítio de ligação do substrato da GAPDH de *T. cruzi* calculado pelo DrugScore (A, B e C) e pelo GRID (C, D e E). (A) Mapa de contorno estéreo (átomo de prova C.3 contornado com um nível de 83% em relação ao mínimo global). (B) Mapa de contorno para potenciais grupos doadores de ligação de hidrogênio na molécula do ligante (átomo de prova N.am, nível de contorno 80%). (C) Mapa de contorno para potenciais grupos aceptores de ligação de hidrogênio na molécula do ligante (átomo de prova O.2, nível de contorno 78%). (D) Mapa de contorno hidrofóbico (átomo de prova DRY, nível de contorno - 0,55 kcal/mol). (E) Mapa de contorno para potenciais grupos doadores de ligação de hidrogênio na molécula do ligante (átomo de prova NH de amida, nível de contorno -9,0 kcal/mol). (F) Mapa de contorno para potenciais grupos aceptores de ligação de hidrogênio na molécula do ligante (átomo de prova O de carbonila, nível de contorno -7,5 kcal/mol).

A combinação das informações obtidas através das diferentes análises de *Hot Spots* e o conhecimento do modo de interação entre proteína e ligantes conhecidos (complexos cristalográficos) determina um conjunto de características físico-químicas essenciais que devem estar presentes na molécula do ligante em potencial, resultando assim numa hipótese farmacofórica.^{92,94}

Para a construção do modelo farmacofórico utilizado foram geradas duas esferas de 2,5 Å de raio ao redor dos grupos fosfatos pertencentes ao análogo de substrato **S70** presente no sítio de ligação da GAPDH de *T. cruzi* (PDB, 1QXS). As esferas foram posicionadas de maneira a restringir no espaço aquelas regiões favoráveis para a interação com grupos aniônicos (sítios Pi e

Ps, sítio de interação do fosfato inorgânico e fosfato do substrato, respectivamente).⁸⁶ Além disso, a análise de *Hot Spots* revelou: i. que a cavidade adjacente ao resíduo Asp210 possui um volume adequado para ser explorado por grupos volumosos na molécula do ligante (Figura 5.18A); ii. uma região favorável para interação com um grupo doador de ligação de hidrogênio próximo a cadeia lateral do Asp210 (Figura 5.18B e 5.18E); iii. uma região favorável para a interação com um grupo aceptor de ligação de hidrogênio próximo ao NH da cadeia principal da Lys209 (Figura 5.18C e 5.18F). Dessa maneira o modelo farmacofórico construído baseado na estrutura do alvo receptor e posteriormente utilizado como filtro molecular 3D apresenta as seguintes características:

1. Duas esferas (volumes de exclusão) posicionadas delimitando as regiões favoráveis para interação com grupos aniônicos;
2. Uma restrição de interação (superfície de interação) capaz de aceitar ligação de hidrogênio (oxigênio da cadeia lateral do Asp210);
3. Uma restrição de interação (superfície de interação) capaz de doar ligação de hidrogênio (NH da cadeia principal da Lys209).

A Figura 5.19 ilustra as características do modelo farmacofórico derivado da análise estrutural e de *Hot Spots* utilizado como filtro molecular 3D.

Figura 5.19 - Modelo farmacofórico baseado na estrutura da proteína. **1)** Um dos grupos aniônicos essenciais (ciano) devem ser posicionado no interior das esferas magenta. **2)** Região no espaço onde um grupo doador de ligação de hidrogênio na molécula do ligante é capaz de interagir com O da cadeia lateral do Asp210. **3)** Região no espaço onde um grupo aceptor de ligação de hidrogênio na molécula do ligante é capaz de interagir com o NH da cadeia principal da Lys209.

Incorporação do modelo farmacofórico no procedimento de docagem molecular

Para a incorporação do modelo farmacofórico como um filtro 3D foi utilizado o módulo FlexX-Pharm⁹⁵ do programa de docagem molecular FlexX. FlexX-Pharm, permite a incorporação de informações sobre o modo de interação entre proteína e ligante durante o processo de docagem molecular. Essa informação é inserida no programa como um conjunto de restrições de interação e volumes de exclusão (e.g., modelo farmacofórico). Uma vez incorporado as restrições farmacofóricas no módulo FlexX-Pharm somente são produzidas soluções que satisfazem as restrições impostas pelo modelo. Além de apresentar uma função de busca robusta, outra característica importante deste módulo é a velocidade de obtenção dos resultados, tornando-o

uma ferramenta atrativa para estudos de VS. Primeiramente, o módulo avalia a capacidade das moléculas presente na base de dados de satisfazerem as restrições impostas pelo modelo farmacofórico, dessa maneira, aquelas moléculas que não se ajustam ao modelo são eliminadas antes do processo de docagem molecular. Isso faz com que haja um ganho considerável na velocidade de obtenção de soluções e na qualidade dos resultados obtidos.⁹⁵ Essas características associada à geração de modos de interações confiáveis fazem com que a utilização deste módulo seja cada vez mais considerada para estudos de VS.⁹⁶⁻¹⁰¹

O módulo FlexX-Pharm oferece ao usuário a possibilidade de definir as restrições farmacofóricas de duas maneiras diferentes: restrições de interações e restrições espaciais. No primeiro caso, pode-se determinar superfícies de interações no sítio ativo que serão utilizadas na interação com o ligante. Para o segundo caso, volumes de exclusão são determinados no interior do sítio ativo os quais deverão apresentar um determinado tipo de átomo ou grupo funcional. Além disso, o módulo permite ao usuário classificar as restrições como essenciais ou opcionais. No presente estudo os seguintes parâmetros de restrições foram utilizados:

- **Restrição de interação:** superfícies de interações correspondentes à capacidade de aceitar ou doar ligação de hidrogênio foram geradas para os átomos de oxigênios da cadeia lateral do Asp210 e do NH da cadeia principal da Lys209, respectivamente (Figura 5.19). A formação destas duas ligações de hidrogênio foi considerada essencial durante o processo de docagem molecular.
- **Restrição espacial:** volumes de exclusão foram gerados no interior do sítio ativo correspondentes às regiões no espaço capazes de interagir com o grupo guanidina da Arg249. No interior destas esferas foi requisitado o posicionamento de um dos grupos aniônicos essenciais (COO^- , SO_3^{-2} , PO_3^{-2} e PO_2^{-2}). A presença de um destes grupos

aniônicos no interior de uma ou outra esfera foi considerada essencial durante o processo de docagem molecular (Figura 5.19).

Dois exemplos de modelos de interação propostos são apresentados na Figura 5.20. Esses modelos satisfazem as restrições farmacofóricas incorporadas no módulo FlexX-Pharm.

Figura 5.20 - Exemplos de modelos de interação para as soluções encontradas durante o estudo de VS. Os modos de ligação propostos satisfazem as restrições impostas pelo modelo farmacofórico incorporado durante o processo de docagem molecular dos compostos da base de dados.

Após o processo de filtragem utilizando o modelo farmacofórico restaram 2.211 moléculas na base de dados de compostos. Esse número de moléculas representa ~0,11 % da base de dados inicial. O modo de ligação proposto das moléculas que satisfazem as restrições farmacofóricas foram posteriormente submetidas ao processo de minimização de energia no interior do sítio de ligação com o programa GOLD. Esse procedimento visou a eliminação daquelas moléculas que apesar de satisfazerem as exigências farmacofóricas apresentavam conformações de alta energia. Para tanto, o valor de rmsd entre as conformações propostas pelo FlexX (restrições farmacofóricas) e GOLD (minimizadas sem restrições) foi utilizado. Aquelas

moléculas que apresentaram valor de $\text{rmsd} < 0,6 \text{ \AA}$ foram selecionadas para a etapa posterior de avaliação. Esse critério selecionou 1.127 moléculas.

Pontuação por consenso

A função de pontuação implementada nos programas de docagem molecular tem como objetivo identificar moléculas bioativas. Para tanto, esta função pontua e classifica as soluções encontradas durante a etapa de amostragem.¹⁰² Entretanto, a limitação para determinação da energia livre de ligação decorrente da falta de conhecimento específico sobre o processo de dessolvatação, constantes dielétricas do microambiente do sítio de ligação, ausência de um termo que defina corretamente a entropia do sistema, aliados a necessidade de se gerar e pontuar soluções em uma escala de tempo aceitável, faz com que as funções de pontuação não sejam robustas o suficiente para a determinação precisa da afinidade de ligação.¹⁰³ Diante disso, os programas de docagem molecular abordam esse problema de maneiras diferentes, adaptando a função de pontuação a uma determinada característica de interesse. Para tanto, os métodos de pontuação apresentam, por exemplo, esquemas diferentes de classificação de tipo de átomos e utilização de métodos diversos para a determinação das cargas parciais.¹⁰³

Para diminuir o problema decorrente das limitações atualmente encontradas pelas funções de pontuação, a pontuação por consenso (do inglês, *consensus scoring*) vem sendo utilizada com sucesso para a determinação de moléculas bioativas.^{81,102,103}

A pontuação por consenso utiliza a combinação de diferentes funções de pontuação para auxiliar na seleção de compostos para aquisição e teste biológico.¹⁰⁴ Essa estratégia envolve a pontuação do modo de ligação proposto dos compostos com diferentes métodos (e.g., 2-5 métodos) e a combinação dos N primeiros compostos classificados por cada função.

No presente estudo, o modo de pontuação do programa DrugScore foi empregado como função de pontuação adicional. DrugScore é uma função de pontuação baseada no conhecimento de complexos entre proteínas e ligantes. Dado um complexo ligante-proteína, o programa avalia as interações não covalentes entre as moléculas. A função de pontuação DrugScore^{PDB} foi desenvolvida através da conversão das informações estruturais de 1.376 complexos cristalográficos presentes no PDB em potenciais de pares dependentes da distância.²⁷ Além do modo DrugScore^{PDB}, o programa oferece a possibilidade de utilização do modo DrugScore^{CSD}, o qual assume que as interações não covalentes provenientes do empacotamento cristalino de pequenas moléculas orgânicas são governadas pelos mesmos princípios físicos das interações entre proteínas e ligantes. Dessa forma, a base de dados cristalográficos CSD proporciona uma fonte bastante rica de informações sobre as geometrias de interações entre átomos observadas em 28.642 cristais de pequenas moléculas.¹⁰⁵

Diante disso, foram utilizados, além da função de pontuação do FlexX, os modos DrugScore^{PDB}, DrugScore^{CSD}, e a função de pontuação GOLDScore, implementada no programa GOLD para realizar a pontuação por consenso das soluções obtidas no estudo de VS.

O modo mais simples de combinar diferentes funções de pontuação baseia-se na somatória das pontuações individuais. Entretanto, apesar dos valores obtidos pelas diferentes funções de pontuação serem uma medida da energia de interação, estes valores não estão em uma mesma escala dimensional, assim, uma estratégia mais robusta de normalização deve ser utilizada para realizar a pontuação por consenso.¹⁰³ A estratégia utilizada no presente estudo baseou-se na somatória das posições de classificação das moléculas obtidas pelas diferentes funções, ou seja, a molécula com o modo de interação melhor classificado apresentou menor valor na somatória das classificações. Por exemplo, a molécula melhor pontuada pela pontuação por consenso foi classificada nas posições 3, 10, 5 e 25 de acordo com as funções de pontuação FlexX,

DrugScore^{PDB}, DrugScore^{CSD} e GOLDScore, respectivamente, logo a somatória das posições de classificação na pontuação por consenso foi de 43 (3+10+5+25). Este procedimento de pontuação por consenso foi empregado para todas as 1.127 moléculas que passaram pelos filtros moleculares 3D e as 500 moléculas melhores moléculas classificadas foram selecionadas para análise visual.

Análise visual dos resultados

Para auxiliar na análise visual dos resultados foi empregado o programa DrugScore^{ONLINE}.¹⁰⁶ Este programa possibilita ao usuário visualizar a contribuição de cada átomo, tanto do ligante como da proteína, para a pontuação do modo de interação proposto de acordo com a função DrugScore^{PDB} ou DrugScore^{CSD}. Essa ferramenta computacional facilita a compreensão das diferentes geometrias das moléculas, efeito de modificações moleculares e importância da interação com determinadas regiões do sítio ativo.

Na Figura 5.21 estão ilustradas as contribuições de cada átomo para o modo de interação proposto para um ligante da base de dados após a etapa de minimização no sítio de ligação com o programa GOLD.

Figura 5.21 - Modo de ligação de um dos compostos da base de dados. A molécula do ligante está representada em ciano, os resíduos de aminoácidos da proteína estão representados em verde e as ligações de hidrogênio estão ilustradas em amarelo. O sítio de ligação está representado contornado pela superfície de Connolly.

As interações atômicas favoráveis estão ilustradas pelas esferas azuis, enquanto as interações desfavoráveis estão ilustradas pelas esferas vermelhas. O tamanho da esfera é proporcional aos valores de contribuições por átomo, ou seja, uma esfera azul de raio maior indica uma interação mais favorável que uma esfera azul de raio menor, enquanto uma esfera vermelha com raio maior indica uma interação mais desfavorável que uma esfera vermelha de raio menor.

O complexo ilustrado na Figura 5.21 apresenta o substituinte carboxílico (grupo essencial) interagindo com o grupo guanidina da Arg249, fazendo com que esta interação contribua positivamente na pontuação desta molécula. Na outra extremidade da molécula, a carbonila do grupo benzamida interage com o NH da Lys209 contribuindo substancialmente para a pontuação final da molécula (esfera azul ao redor dos átomos de oxigênio da carbonila). Já o

substituinte sulfonilamino orienta o grupo sulfonila para o solvente sugerindo uma pequena contribuição desfavorável para a pontuação final. Entretanto, devido ao fato que o DrugScore não leva em consideração as moléculas de água que estariam interagindo com os oxigênios do grupo sulfonila o modo de ligação proposto para esta molécula pode ser considerado bastante razoável.

Observações como estas detalhadas acima, foram utilizadas durante a análise visual dos 500 melhores compostos classificados pelo método de pontuação por consenso.

Seleção dos compostos

Após a análise visual dos 500 melhores compostos segundo a pontuação por consenso, foram selecionados 108 compostos que apresentavam modo de ligação razoável de acordo com os critérios: complementaridade geral da conformação proposta e interação dos grupos aniônicos (grupos essenciais) presente nas moléculas dos ligantes com a Arg249.

O último critério adotado de seleção dos compostos para aquisição foi novamente baseado no valor de rmsd entre o modo de ligação proposto pelos programas FlexX e GOLD, entretanto, um valor de rmsd mais rigoroso ($\leq 0,25$ Å) foi adotado para a seleção final das moléculas. Esse procedimento resultou em 58 moléculas que foram verificadas quanto à disponibilidade e preço para aquisição. A seleção final resultou em 35 compostos, pertencentes às bases de dados ChemBrigde e da InterbioScreen. As moléculas selecionadas apresentam as seguintes características: tamanho adequado para futuras otimizações, as massas moleculares variam entre 280 e 410, solubilidade razoável em solução aquosa, o Log *P* varia entre 0,3 e 3,58 e números de ligações rotacionáveis variam entre 4 e 8, evitando assim os efeitos entrópicos desfavoráveis devido a imobilização de conformações. Um subconjunto de 10 moléculas representativo da diversidade química e estrutural das moléculas selecionadas está ilustrado na Tabela 5.11.

O pedido de compra das moléculas selecionadas foi realizado e no presente momento estas encontram-se em fase de entrega. A Figura 5.22 ilustra de maneira resumida as etapas e procedimentos utilizados para a seleção de compostos.

Figura 5.22 - Modelo esquemático das etapas realizadas no estudo de VS com a GAPDH de *T. cruzi*.

Tabela 5.11 - Sub-conjunto de 10 moléculas representantes do compostos selecionados pelo VS para atividade biológica.

Estrutura	Log <i>P</i>	Massa molecular	código	Base de dados
1,39	319,32	5914938	ChemBrigde
2,65	337,31	7357604	ChemBrigde
1,03	301,33	7488857	ChemBrigde
2,16	342,38	7685630	ChemBrigde
3,58	385,35	5276965	ChemBrigde
1,02	302,31	9021138	ChemBrigde
3,2	410,87	116699	InterbioScreen
0,3	320,34	8748190	InterbioScreen
3,09	332,36	862838	InterbioScreen
0,9	280,28	561612	InterbioScreen

5.4 - Referências

1. GUIDO R. V. C.; OLIVA, G.; ANDRICOPULO, A. D. Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, v. 15, p. 37-46, 2008.
2. KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.*, v. 3, p. 935-949, 2004.
3. DAVIS, A. M.; TEAGUE, S. J.; KLEYWEGT, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed. Engl.*, v. 42, p. 2718-2736, 2003.
4. KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today.*, v. 11, p. 580-594, 2006.
5. BERMAN, H. M. et al. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, suppl, p. 957-959, 2000.
6. ZDOBNOV, E. M.; APWEILER, R. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, v. 17, p. 847-848, 2001.
7. MULDER, N. J. et al. New developments in the InterPro database. *Nucleic Acids Res.*, v. 35, p. D224-D228, 2007.
8. MULDER, N. J. et al. InterPro, progress and status in 2005. *Nucleic Acids Res.*, v. 33, p. D201-D205, 2005.
9. GASTEIGER, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, v. 31, p. 3784-3788, 2003.
10. HENDLICH, M. et al. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.*, v. 326, p. 607-620, 2003.
11. GÜNTHER, J. et al. Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.*, v. 326, p. 621-636, 2003.

12. BERGNER, A. et al. Use of relibase for retrieving complex 3D interaction patterns including crystallographic packing effects. *Biopolymers*, v. 61, p. 99-110, 2002.
13. KUNTZ, I. D. et al. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.*, v. 161, p- 269-288, 1982.
14. LEACH, A. R.; SHOICHET, B. K.; PEISHOFF, C. E. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* v. 49, p. 5851-5855, 2006.
15. SHOICHET, B.K.; KUNTZ, I.D. Protein docking and complementarity. *J. Mol. Biol.*, v. 22, p. 327-346, 1991.
16. JONES, G. et al. Development and validation of a genetic algorithm for flexible docking., *J. Mol. Biol.*, v. 267, p. 727-748, 1997.
17. RAREY, M. et al. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, v. 261, p. 470-489, 1996.
18. FRIESNER, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, v. 47, p. 1739-1749, 2004.
19. HALGREN, T. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.*, v. 47, p. 1750-1759, 2004.
20. MORRIS, G. M. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, v. 19, p. 1639-1662, 1998.
21. JAIN, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.*, v. 46, p. 499-511, 2003.
22. WANG, R.; LU, Y.; WANG, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.*, v. 46, p. 2287-2303, 2003.
23. TRIPOS – Sybyl, 2008. Disponível em: <<http://www.tripos.com/>>. Acessado em: 12 Fev 2008.

24. WEINER, S. J. et al. An all-atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, v. 7, p. 230-252, 1986.
25. ACCELERYS - Cerius², 2008. Disponível em: <http://www.accelrys.com/products/cerius2/>>. Acessado em: 12 Fev 2008.
26. WANG, R.; LAI, L.; WANG, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.*, v. 16, p. 11-26, 2002.
27. GOHLKE, H.; HENDLICH, M.; KLEBE, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, v. 295, p. 337-356, 2000.
28. LENGAUER, T. et al. Novel technologies for virtual screening. *Drug Discov. Today*, v. 9, p. 27-34, 2004.
29. ECKERT, H.; BAJORATH, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today*, v. 12, p. 225-233, 2007.
30. WILLETT, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today*, v. 11, p. 1046-1053, 2006.
31. MASON, J. S.; GOOD, A. C.; MARTIN, E. J. 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.*, v. 7, p. 567-597, 2001.
32. WARMUTH, M. K. et al. Active learning with support vector machines in the drug discovery process. *J. Chem. Inf. Comput. Sci.*, v. 43, p. 667-673, 2003.
33. PATRICK, G. L. *An Introduction to Medicinal Chemistry*. New York: Oxford University Press, 2001.
34. FARUTIN, V. et al. Structure-activity relationships for a class of inhibitors of purine nucleoside phosphorylase. *J. Med. Chem.*, v. 42, p. 2422-2431, 1999.

35. GUIDO R. V. C.; et al. Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*. *J. Enz. Inhib. Med. Chem.*, v. 23, 2008. In press
36. GUIDO, R. V. C. et al. Classical and hologram QSAR studies on a series of inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. *QSAR Comb. Sci.*, v. 27, 2008. In press
37. GUIDO, R. V. C. et al. Structural basis for selective inhibition of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase: molecular docking and 3D QSAR studies. *J. Chem. Inf. Model.*, v. 48, 2008. In press
38. ANDRICOPULO, A. D.; MONTANARI, C. A. Structure-activity relationships for the design of Small-molecule inhibitors. *Mini Rev. Med. Chem.*, v. 5, p.585-593, 2005.
39. KUBINYI, H. *QSAR: Hansch analysis and related approaches*. New York:VCH, 1993.
40. HANSCH, C.; LEO, A. *Exploring QSAR: fundamentals and applications in chemistry and biology*. Washington: American Chemical Society, 1995.
41. KUBINYI, H. *3D QSAR in drug design – theory, methods and applications*. Leiden: Escom, 1993.
42. VON ITZSTEIN, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, v. 363, p. 418-423, 1993.
43. SIPPL, W. Development of biologically active compounds by combining 3D QSAR and structure-based design methods. *J. Comput. Aided Mol. Des.* v. 16, p. 825-830, 2002.
44. VERLINDE, C. L. M. J. et al. Selective inhibition of trypanosomal glyceraldehyde-3-phosphate dehydrogenase by protein structure-based design: toward new drugs for the treatment of sleeping sickness. *J. Med. Chem.*, v. 37, p. 3605-3613, 1994.
45. VAN CALENBERGH, S. et al. Synthesis and structure-activity relationships of analogs of 2'-deoxy-2'-(3-methoxybenzamido) adenosine, a selective inhibitor of trypanosomal glycosomal glyceraldehyde-3-phosphate dehydrogenase. *J. Med. Chem.*, v. 38, p. 3838-3849, 1995.

46. ARONOV, A. M. et al. Selective tight binding inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase via structure-based drug design. *J. Med. Chem.*, v. 41, p. 4790-4799, 1998.
47. ARONOV, A. M. et al. Structure-based design of submicromolar, biologically active inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. *Proc. Natl. Acad. Sci. USA*, v. 96, p. 4273-4278, 1999.
48. BRESSI, J. C. et al. Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of trypanosomatidae via structure-based drug design. *J. Med. Chem.*, v. 44, p. 2080-2093, 2001.
49. CLARK, M.; CRAMER, R. D., III; VAN OPDENBOSCH, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.*, v. 10, p. 982-1012, 1989.
50. POWELL, M. J. D. Restart Procedures for the Conjugate Gradient Method. *Math. Program.*, v. 12, p. 241-254, 1977.
51. GASTEIGER, J.; MARSILI, M. Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron*, v. 36, p. 3219-3228, 1980.
52. BUSH, B. L.; NACHBAR, R. B. JR. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided. Mol. Des.*, v. 7, p. 587-619, 1993.
53. LOWIS, D. R. *HQSAR: a new, highly predictive QSAR technique*. USA: Tripos, 1997. (Technical Notes, v. 1).
54. TONG, W. et al. Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.*, v. 38, p. 669-677, 1998.
55. HQSARTM *Manual*. St. Louis:Tripos Inc., 2006.
56. SEEL, M.; TURNER, D.B.; WILLETT, P. Effect of parameter variations on the effectiveness of HQSAR analyses. *QSAR*, v. 18, p. 245-252, 1999.

57. CASTILHO, M. S.; GUIDO, R. V. C.; ANDRICOPULO, A. D. 2D Quantitative structure-activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. *Bioorg. Med. Chem.*, v. 15, p. 6242-6252, 2007.
58. CASTILHO, M. S.; GUIDO, R. V. C.; ANDRICOPULO, A. D. Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors *Lett. Drug Des. Discov.*, v. 4, p. 106-113, 2007.
59. HONÓRIO, K. M.; GARRATT, R. C.; ANDRICOPULO, A. D. Hologram quantitative structure-activity relationships for a series of farnesoid X receptor activators. *Bioorg. Med. Chem. Lett.*, v. 15, p. 3119-3125, 2005.
60. KUBINYI, H. Comparative molecular field analysis. In: *Encyclopedia of computational chemistry*. USA: John Wiley & Sons, Ltd., 1998, v. 1, p. 1-12.
61. CRAMER, R. D. et al. 1. Effect of shape on binding of steroids to carrier proteins *J. Am. Chem. Soc.*, v. 110, p. 5959-5867, 1988.
62. CRAMER, R. D. III; PATTERSON, D. E.; BUNCE, J.D. Recent Advances in Comparative Molecular Field Analysis (CoMFA). *Prog. Clin. Biol. Res.*, v. 291, p. 161-165, 1989.
63. KLEBE, G.; ABRAHAM, U.; MIETZNER, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.*, v. 37, p. 4130-4146, 1994.
64. KLEBE, G.; ABRAHAM, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.*, v. 13, p. 1-10, 1999.
65. SOUZA, D. H. F. et al. *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design. *FEBS Lett.*, v. 424, p. 131-135, 1998.
66. VELLIEUX, F. M. et al. Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. *Proc. Natl. Acad. Sci. USA*, v. 90, p. 2355-2359, 1993.

67. KIM, H. et al. Crystal structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Leishmania mexicana*: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry*, v. 34, p. 14975-14986, 1995.
68. SALUM, L. B., POLIKARPOV, I.; ANDRICOPULO, A. D. Structural and chemical Basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. *J. Mol. Graph. Model.*, v. 26, p. 434-442, 2007.
69. HONORIO, K. M. et al. 3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators. *J. Mol. Graph. Model.*, v. 25, p. 921-927, 2007.
70. SURESH, S. et al. Conformational changes in *Leishmania mexicana* glyceraldehyde-3-phosphate dehydrogenase induced by designed inhibitors. *J. Mol. Biol.*, v. 309, p. 423-435, 2001.
71. BISSANTZ, C.; FOLKERS, G.; ROGNAN, D. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.*, v. 43, p. 4759-4767, 2000.
72. WARREN, G. L. et al. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, v. 49, p. 5912-5931, 2006.
73. DODDAREDDY, M. R. et al. 3D QSAR studies on T-type calcium channel blockers using CoMFA and CoMSIA. *Bioorg. Med. Chem.*, v. 12, p. 1613-1621, 2004.
74. AMIN, E. A.; WELSH, W. J. Highly predictive CoMFA and CoMSIA models for two series of stromelysin-1 (MMP-3) inhibitors elucidate S1' and S1-S2' binding modes. *J. Chem. Inf. Model.*, v. 46, p. 1775-1783, 2006.
75. ZHU, L. P.; YE, D. Y.; TANG, Y. Structure-based 3D-QSAR studies on thiazoles as 5-HT3 receptor antagonists. *J. Mol. Model.*, v. 13, p. 121-131, 2007.
76. CLARK, R. D.; FOX, P. C. Statistical variation in progressive scrambling. *J. Comput.-Aided Mol. Des.*, v. 17, p. 1-14, 2004.

77. CRAMER, R. D. III et al. Cross-Validating, Bootstrapping, and Partial Least-Squares Compared with Multiple-Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.*, v. 7, p. 18-25, 1988.
78. WEHRENS, R.; VAN DER LINDEN, W. E. Bootstrapping Principal Component Regression Models. *J. Chemom.*, v. 11, p. 157-171, 1997.
79. VERLINDE, C. L. M. et al. Glycolysis as a target for the design of new anti trypanosome drugs. *Drug Resist. Updat.*, v. 4, p. 50-65, 2001.
80. GRÜNEBERG, S.; STUBBS, M. T.; KLEBE, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J. Med. Chem.* v. 45, p. 3588-35602, 2002.
81. LYNE, P. D. et al. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J. Med. Chem.*, v. 47, p. 1962-1968, 2004.
82. KRAEMER, O. et al. Virtual screening for inhibitors of human aldose reductase. *Proteins*, v. 55, p. 814-823, 2004.
83. BRENN, R. et al. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.*, v. 46, p. 1133-1143, 2003.
84. BRENN, R. et al. Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening. *J. Mol. Biol.*, v. 338, p. 55-75, 2004.
85. PAVÃO, F. et al. Structure of *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 Å resolution. *FEBS Lett.*, v. 520, p. 13-17, 2002.
86. CASTILHO, M. S. et al. Evidence for the two phosphate binding sites of an analogue of the thioacyl intermediate for the *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase-catalyzed reaction, from its crystal structure. *Biochemistry*, v. 42, p. 7143-7151, 2003.

87. LADAME, S. et al. Crystal structure of *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase complexed with an analogue of 1,3-bisphospho-d-glyceric acid. *Eur. J. Biochem.*, v. 270, p. 4574-4586, 2003.
88. IRWIN, J. J.; SHOICHET, B. K. ZINC - a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* v. 45, p. 177-182, 2005.
89. LIPINSKI, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods*, v. 44, p. 235-249, 2000.
90. BRUNO, I. J. et al. Isostar: A library of information about non-bonded interactions., *J. Comput.-Aided Mol. Des.*, v. 11, p. 525-537, 1997.
91. SOTRIFFER, C.; KLEBE, G. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. *Farmaco*, v. 57, p. 243-251, 2002.
92. BRENK, R.; KLEBE G. Hot Spot analysis of protein-binding sites as a prerequisite for structure-based virtual screening and lead optimization. In: LANGER, T.; HOFFMANN R. D. (eds.). *Pharmacophores and Pharmacophore Searches*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA, 2006, p. 171-192.
93. GOODFORD, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.*, v. 28, p. 849-857, 1985.
94. ALVAREZ J.C., SHOICHET B. *Virtual screening in drug discovery*. Florida: CRC Press, 2005.
95. HINDLE, S. A. et al. Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.*, v. 2, p. 129-149, 2002.
96. POLGAR, T. et al. Comparative virtual and experimental high-throughput screening for glycogen synthase kinase-3 β inhibitors. *J. Med. Chem.*, v. 48, p. 7946-7959, 2005.

97. POLGAR, T.; KESERU, G. M. Virtual screening for beta-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.*, v. 11, p. 3749-3755, 2005.
98. SHAIKH, M. S.; MITTAL, A.; BHARATAM, P. V. Design of fructose-2,6-bisphosphatase inhibitors: A novel virtual screening approach. *J. Mol. Graph. Model.*, v. 26, p. 900-906, 2008.
99. CHEN, J. Z.; WANG, J.; XIE, X. Q. GPCR structure-based virtual screening approach for CB2 antagonist search. *J. Chem. Inf. Model.*, v. 47, p. 1626-1637, 2007.
100. KIM, R.; SKOLNICK, J. Assessment of programs for ligand binding affinity prediction. *J. Comput. Chem.*, v. 29, 2008. In press
101. CHENNA, B. C. et al. Identification of novel inhibitors of bacterial surface enzyme *Staphylococcus aureus* Sortase A. *Bioorg. Med. Chem. Lett.*, v. 18, p. 380-385, 2008.
102. CHARIFSON, P. S. et al. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.*, v. 42, p. 5100-5109, 1999.
103. CLARK, R. D. et al. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* v. 20, p. 281-295, 2002.
104. FEHER, M. Consensus scoring for protein-ligand interactions. *Drug Discov. Today*, v. 11, p. 421-428, 2006.
105. VELEC, H. F. G.; GOHLKE, H.; KLEBE, G. DrugScore^{CSD}-Knowledge-based scoring function derived from Small Molecule Crystal data with superior recognition rate of near native ligand poses and better affinity prediction. *J. Med. Chem.*, v. 48, p. 6296-6303, 2005.
106. DRUGSCORE^{ONLINE} - Scoring Functions for Protein-Ligand Complexes. Disponível em : <<http://www.agklebe.de/>>. Acessado em 12 Fev 2008.

6 - COLABORAÇÕES

No decorrer do doutoramento, houve ótimas oportunidades para o estabelecimento de colaborações produtivas em outros projetos atualmente em andamento no LQMC/CBME, que resultaram em trabalhos importantes que são descritos a seguir.

6.1 - Artigo de revisão

Um artigo de revisão foi elaborado em colaboração com o Prof. Dr. Adriano D. Andricopulo do Instituto de Física de São Carlos da Universidade de São Paulo (IFSC-USP). O artigo mostra a integração de técnicas modernas de planejamento de fármacos. Neste contexto, são discutidos os fundamentos das técnicas de triagem virtual de compostos baseados tanto na estrutura do ligante quanto na estrutura do alvo receptor. O artigo é finalizado com exemplos de aplicações resultantes da combinação destes métodos com outras técnicas modernas de planejamento de fármacos, como o HTS e QSAR.¹

Essas observações foram organizadas em um manuscrito intitulado *Virtual screening and its integration with modern drug design technologies*, de autoria de Rafael V. C. Guido, Glaucius Oliva e Adriano D. Andricopulo, e publicado no *Current Medicinal Chemistry*, v. 15, p. 37-46, 2008 (Apêndice C).

6.2 - Desenvolvimento de modelos de QSAR

6.2.1 - Estudos de QSAR 2D para uma série de inibidores da enzima butirilcolinesterase (BuChe)

Este trabalho foi realizado em colaboração com o Prof. Dr. Marcelo S. Castilho da Faculdade de Farmácia da Universidade Federal da Bahia (UFBA). O mal de Alzheimer é uma das principais doenças neurodegenerativas e afeta principalmente pessoas acima de 65 anos de idade. Um dos principais tratamento é baseado na utilização de inibidores da enzima acetilcolinesterase (AChE). Estudos recentes mostraram que a enzima butirilacolinesterase (BuChe) é considerado um alvo potencial para o tratamento do mal de Alzheimer e que a inibição seletiva desta enzima poderia ser útil na terapia da doença.^{4,5} Diante disso, estudos de QSAR 2D foram conduzidos numa série de inibidores potentes da enzima BuChe. As técnicas de QSAR clássico e HQSAR foram empregadas num conjunto treino de 40 inibidores derivados de tacrina. Os modelos de QSAR clássico mostraram moderada correlação dos dados ($r^2 = 0,84$; $q^2 = 0,75$), entretanto, o modelo não apresentou capacidade preditiva significativa para os 10 compostos do conjunto teste. Já o melhor modelo de HQSAR ($r^2 = 0,93$; $q^2 = 0,72$) mostrou substancial capacidade preditiva apresentando valores residuais (diferença entre os dados experimentais e preditos) $< 0,8$ unidade logarítmica. Portanto, o modelo final de HQSAR é útil para o desenvolvimento e predição da potência de novas moléculas dessa classe de inibidores da enzima BuChe.⁶

Esses resultados foram organizados em um manuscrito intitulado *Classical and Hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors*, de autoria de Marcelo S. Castilho, Rafael V. C. Guido e Adriano D. Andricopulo e publicado no periódico científico *Letters in Drug Design & Discovery*, v. 4, p. 106-113, 2007 (Apêndice D).

6.2.2 - Estudos de QSAR 2D para uma série de inibidores da proteína colesterol éster transferase (CETP)

Este trabalho também foi realizado em colaboração com o Prof. Dr. Marcelo S. Castilho da Faculdade de Farmácia da Universidade Federal da Bahia (UFBA). As doenças cardiovasculares contribuem significativamente para morbidez e mortalidade da população mundial. Dentre elas, merece destaque as doenças coronarianas, que são caracterizadas como distúrbios vasculares que comprometem a circulação nas artérias coronarianas e, conseqüentemente, afetam a irrigação do miocárdio. Este tipo de distúrbio caracteriza-se pelo estreitamento progressivo, agudo ou crônico, devido ao depósito de substâncias gordurosas na parede dessas artérias.⁷ Atualmente, um dos principais tratamentos das doenças cardiovasculares é baseado na redução dos níveis plasmáticos das lipoproteínas de baixa densidade (LDL), entretanto, acredita-se que o novo paradigma no tratamento será baseado no aumento dos níveis plasmáticos das lipoproteínas de alta densidade (HDL).⁸ Uma das estratégias mais promissoras para o aumento do HDL é a inibição da proteína colesterol éster transferase (CETP). Diante disso, no trabalho realizado em colaboração, modelos de QSAR clássico e HQSAR foram construídos baseados num conjunto 85 moléculas derivadas do trifluor-3-amino-2-propanol, potentes inibidores de CETP. Modelos robustos e com razoável capacidade preditiva foram obtidos. A análise combinada dos resultados de QSAR permitiu identificar características estruturais importantes para o desenvolvimento de novos inibidores de CETP.⁹

Esses resultados foram organizados em um manuscrito intitulado *2D Quantitative structure–activity relationship studies on a series of cholesteryl ester transfer protein inhibitors*, de autoria de autoria de Marcelo S. Castilho, Rafael V. C. Guido e Adriano D. Andricopulo e publicado no periódico científico *Bioorganic & Medicinal Chemistry*, v. 15, p. 6242-6252, 2007. (Apêndice E).

6.2.3 - Estudos de QSAR para uma série de inibidores da enzima cruzaina de *T. cruzi*

Este trabalho foi realizado em colaboração com a Profa. Dra. Elizabeth I. Ferreira e com o doutorando Gustavo H. G. Trossini da Faculdade de Ciências Farmacêuticas da Universidade de São Paulo (FCF-USP). Com o objetivo de estudar as bases químicas e estruturais da inibição da enzima cruzaina de *T. cruzi*, alvo promissor para o desenvolvimento de novos agentes tripanocidas,¹⁰ estudos de QSAR 2D e 3D foram conduzidos numa série de 55 potentes inibidores derivados (tio)semicarbazonas. Os modelos de QSAR clássico e HQSAR apresentaram parâmetros estatísticos significantes (QSAR clássico, $r^2 = 0,83$; $q^2 = 0,72$; HQSAR, $r^2 = 0,96$; $q^2 = 0,75$). A validação externa dos modelos com um conjunto teste de 10 inibidores mostrou que os valores de potência predita dos compostos estavam de acordo com os dados experimentais (QSAR clássico, $r^2_{\text{pred}} = 0,91$; HQSAR, $r^2_{\text{pred}} = 0,95$).¹¹

Esses resultados foram organizados em um manuscrito intitulado *Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from Trypanosoma cruzi*, de autoria de Rafael V. C. Guido, Gustavo H. G. Trossini, Marcelo S. Castilho, Glaucius Oliva, Elizabeth I. Ferreira e Adriano D. Andricopulo e publicado no periódico científico *Journal of Enzyme Inhibition and Medicinal Chemistry*, DOI: 10.1080/14756360701810322 (Apêndice F).

6.3 - Desenvolvimento de biorreator enzimático utilizando a GAPDH de *T. cruzi*

Este trabalho foi realizado em colaboração com a Profa. Dra. Quézia B. Cass do Instituto de Química da Universidade Federal de São Carlos (IQ-UFSCar) e com a Profa. Dra. Carmen L.

Cardoso do Departamento de Química da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCL-RP - USP).

Tendo em vista a necessidade de desenvolvimento de métodos rápidos e eficazes para a identificação de novos compostos inibidores da GAPDH de *T. cruzi*, a enzima do parasita foi alvo de estudos de imobilização em coluna cromatográfica para o desenvolvimento de biorreatores enzimáticos. A estratégia de se construir biorreatores com enzimas imobilizadas é atrativa, pois apresenta algumas vantagens quando comparada ao ensaio realizado *in vitro*, como, por exemplo, o aumento da estabilidade, pequenas quantidades de amostra e a possibilidade de implementação de ensaios de triagem de alta eficiência.¹² Diante disso, foram realizados estudos para avaliar o melhor método de imobilização da GAPDH. Diversas variáveis como solução tampão, pH e temperatura foram investigados para a determinação das condições ideais de preparação do bioreator. A melhor condição para imobilização foi obtida com 20 mM de solução tampão HEPES em pH 8,2 e 25°C. Após a determinação das condições ideais de preparação, o biorreator foi utilizado para a determinação das constantes cinéticas da GAPDH imobilizada. Os estudos cinéticos mostraram que a enzima imobilizada permanece ativa com os valores calculados das constantes cinéticas entre a enzima imobilizada e livre diferindo ligeiramente. Sendo assim, o biorreator pode ser usado para a identificação de novos inibidores da GAPDH de *T. cruzi*.¹³

Esses resultados foram organizados em um manuscrito intitulado *The development of an immobilized enzyme reactor containing glyceraldehyde-3-phosphate dehydrogenase from Trypanosoma cruzi: the effect of species' specific differences on the immobilization* de autoria de Carmen L. Cardoso, Marcela C. de Moraes, Rafael V. C. Guido, Glaucius Oliva, Adriano D. Andricopulo, Irving W. Wainer e Quézia B. Cass, e publicado no periódico científico *The Analyst*, v. 133, p. 93-99, 2008 (Apêndice G).

6.4 - Capítulo de livro

6.4.1 - Construção de modelo farmacofórico baseado na estrutura do alvo receptor

Um capítulo de livro foi elaborado abordando os fundamentos das técnicas modernas de construção de modelos farmacofóricos baseado na estrutura do alvo receptor. O capítulo inicia-se definindo o conceito de farmacóforos e como os métodos de identificação e construção de modelos farmacofóricos 2D e 3D evoluíram. Em seguida, o leitor é apresentado as técnicas e fundamentos das estratégias modernas de identificação de farmacóforos e como estas são utilizadas para a obtenção de modelos farmacofóricos baseados na estrutura do alvo receptor (e.g., análise de *Hot Spots*, método FLAP e o método *LigandScout*). A segunda parte do capítulo discute os métodos baseado em exemplos recentes de aplicação das técnicas discutidas nas seções anteriores.²

Esse capítulo intitulado *Structure-Based Pharmacophore Strategies in Drug Discovery*, de autoria de Rafael V. C. Guido, Marcelo S. Castilho, Glaucius Oliva e Adriano D. Andricopulo será publicado no livro *Current Methods in Medicinal Chemistry and Biological Physics*, v. II, 2008, p. 1-39, Research Signpost, editado por Carlos H. T. P. Silva e Carlton Taft (em fase final de produção).

6.5 - Referências

1. GUIDO R. V. C.; OLIVA, G.; ANDRICOPULO, A. D. Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, v. 15, p. 37-46, 2008.
2. GUIDO, R. V. C. et al. Structure-based pharmacophore strategies in drug discovery. In: SILVA, C. H. T. P.; TAFT, C. A. (eds.) *Current Methods in Medicinal Chemistry and Biological Physics*, India: Research Signpost, 2008, v. II, p. 1-39. (In press)
3. GUIDO, R. V. C.; CASTILHO M. S.; ANDRICOPULO A. D. Planejamento de Fármacos. In: FERREIRA E. I. (ed). *Química Medicinal*, Brasil, 2008, p. 1-45. (No prelo)
4. ROSINI, M.; et al. Rational approach to discover multipotent anti-Alzheimer drugs. *J. Med. Chem.*, v. 48, p. 360-363, 2005.
5. LISTON, D. R. et al. Pharmacology of selective acetylcholinesterase inhibitors: implications for use in Alzheimer's disease. *Eur. J. Pharmacol.*, v. 486, p. 9-17, 2004.
6. CASTILHO, M. S.; GUIDO, R. V. C.; ANDRICOPULO, A. D. Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors *Lett. Drug Des. Discov.*, v. 4, p. 106-113, 2007.
7. ABILDSTROM, S. Z. et al. Trends in incidence and case fatality rates of acute myocardial infarction in Denmark and Sweden. *Heart*, v. 89, p. 507-511, 2003.
8. MILANI, R. V.; LAVIE, C. J. Cholesteryl ester transfer protein inhibition: the next frontier in combating coronary artery disease? *J. Am. Coll. Cardiol.*, v. 48, p. 1791-1792, 2006.
9. CASTILHO, M. S.; GUIDO, R. V. C.; ANDRICOPULO, A. D. 2D Quantitative structure-activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. *Bioorg. Med. Chem.*, v. 15, p. 6242-6252, 2007.

10. MCKERROW, J. H.; ENGEL, J. C; CAFFREY, C. R. Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg. Med. Chem.*, v. 7, p. 639-644, 1999.
11. GUIDO R. V. C.; et al. Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*. *J. Enz. Inhib. Med. Chem.*, v. 23, 2008. In press
12. GIRELLI, A. M.; MATTEI, E. Application of immobilized enzyme reactor in on-line high performance liquid chromatography: a review. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, v. 819, p. 3-16, 2005.
13. CARDOSO, C. L. et al. The development of an immobilized enzyme reactor containing glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma cruzi*: the effect of species' specific differences on the immobilization. *Analyst*, v. 133, p. 93-99, 2008.

Capítulo 7

Conclusões e Perspectivas

Nesse capítulo são apresentadas as conclusões desta tese de doutorado que envolveu os estudos em química medicinal e biologia estrutural no planejamento de inibidores da enzima GAPDH de tripanosomatídeos. Além disso, são apresentadas algumas perspectivas de trabalhos que serão realizados para a finalização plena de estudos iniciados nesta tese.

7 - CONCLUSÕES E PERSPECTIVAS

7.1 - Conclusões

Nesta tese de doutorado foram apresentados os resultados obtidos no decorrer de cinco anos de trabalho. Várias técnicas e estratégias foram empregadas para investigar as propriedades estruturais e biológicas da enzima GAPDH de *T. cruzi*.

Diversos ensaios de triagem bioquímica *in vitro* de compostos de origem natural e sintética foram realizados. Neste contexto, foram identificadas novas classes estruturais que servem de modelos para futuros estudos de SAR. Ensaios de cristalização e de difração de raios X foram conduzidos com a GAPDH e os inibidores identificados, entretanto, apesar dos cristais de GAPDH crescerem na presença de soluções de inibidores, as análises dos mapas de densidade eletrônica dos conjuntos de dados coletados não revelaram a formação dos complexos enzima-inibidor esperados.

As propriedades cinéticas da GAPDH foram determinadas e os parâmetros cinéticos reportados. Os resultados mostraram de forma pioneira que a enzima GAPDH de *T. cruzi* não apresenta comportamento cooperativo entre as subunidades da enzima em relação à interação com o substrato ou o cofator. Estudos estruturais realizados forneceram dados cristalográficos inéditos da GAPDH, evidenciando quatro estados conformacionais diferentes da enzima. A comparação entre as estruturas revelou uma mudança conformacional induzida pela ligação da molécula do NAD⁺. A análise dos estados intermediários de ligação com o NAD⁺ mostrou que a interação com a molécula do cofator em uma subunidade não determina mudanças conformacionais significativas nas demais subunidades. Dessa forma, os dados estruturais corroboram os dados cinéticos, uma vez que a interação da molécula do cofator em um sítio de

ligação não altera a organização estrutural e, portanto, a afinidade dos sítios nas subunidades adjacentes. A análise estrutural comparativa dentre a GAPDH de *T. cruzi* e outras enzimas GAPDH de organismos que apresentam cooperatividade negativa sugere que a presença de um resíduo mutado na GAPDH de *T. cruzi*, numa posição chave na interface entre três subunidades, poderia contribuir significativamente com interações químicas adicionais para o comportamento não cooperativo.

Estudos de modelagem molecular com enzimas GAPDH de tripanosomatídeos ajudaram a elucidar as bases predominantes no processo de reconhecimento molecular, afinidade e seletividade dos inibidores investigados. Os modelos de QSAR baseados nos alinhamentos obtidos por técnicas de SBDD podem ser úteis no planejamento de NCEs apresentando maior potência e seletividade. Além disso, esses estudos evidenciam como estas técnicas em química medicinal podem ser empregadas de modo combinado para se extrair informações detalhadas e qualificadas para o planejamento de novas substâncias bioativas. Os resultados dos estudos de QSAR 2D e QSAR 3D estão publicados (versão on-line) nos periódicos *QSAR & Combinatorial Science* DOI: 10.1080/14756360701810322 (Apêndice A) e *Journal of Chemical Information and Modeling* DOI: 10.1021/ci700453j (Apêndice B), respectivamente.

Durante o período de doutorado-sanduiche na Philipps-Universität-Marburg sob a orientação do Prof. Dr. Gerhard Klebe, foi possível aprender diversas técnicas e métodos úteis para o planejamento de inibidores enzimáticos que foram aplicados nesta tese de doutoramento. Entre eles, destaca-se o estudo de triagem virtual de compostos, que consistiu na aplicação de filtros hierárquicos sucessivos estabelecidos de acordo com as características e propriedades do sítio de ligação da enzima-alvo. O método de docagem molecular incorporou informações de um modelo farmacofórico construído baseado na estrutura do receptor-alvo e os compostos melhores classificados pela estratégia de pontuação por consenso foram visualmente analisados, sendo que

um conjunto de 35 moléculas foi selecionado para aquisição e futura determinação da atividade biológica.

Trabalhos em colaboração foram realizados aplicando-se técnicas aprendidas durante o período de doutoramento. Foi elaborado um artigo de revisão discutindo a integração de técnicas modernas de planejamento de fármacos, o qual está publicado no periódico *Current Medicinal Chemistry* (Apêndice C). Além disso, um capítulo de livro foi escrito apresentando e discutindo os fundamentos das técnicas modernas de construção de modelos farmacofóricos baseado na estrutura do receptor-alvo. Este capítulo fará parte do livro *Current Methods in Medicinal Chemistry and Biological Physics*, v. II, que encontra-se em fase final de publicação. Três trabalhos distintos envolvendo os estudos das relações entre a estrutura e atividade foram elaborados: (i) estudos de QSAR 2D para uma série de inibidores da enzima butirilcolinesterase (BuChe), publicado no periódico *Letters in Drug Design & Discovery* (Apêndice D); (ii) estudos de QSAR 2D para uma série de inibidores da proteína colesterol éster transferase (CETP), publicado no periódico *Bioorganic & Medicinal Chemistry* (Apêndice E); e (iii) estudos de QSAR 2D para uma série de inibidores da enzima cruzaina de *T. cruzi*, publicado no periódico *Journal of Enzyme Inhibition and Medicinal Chemistry* (Apêndice F). Por fim, foi elaborado um trabalho descrevendo os resultados dos estudos de imobilização em coluna cromatográfica da GAPDH de *T. cruzi* para o desenvolvimento de biorreatores enzimáticos, o qual está publicado no periódico *The Analyst* (Apêndice G).

7.2 - Perspectivas

Esta tese de doutorado faz parte de uma das principais linhas de pesquisa estabelecidas no CBME, que visa o desenvolvimento de novos agentes antiparasitários. Diante disso, a identificação e otimização de novos inibidores da enzima GAPDH de *T. cruzi* continua sendo um dos principais objetivos desta pesquisa. Este projeto será continuado pela doutoranda Tatiane L. Balliano (aluna de doutorado do IFSC em Física Aplicada, opção Física Biomolecular), já responsável pelos estudos cinéticos, cristalográficos e de modelagem molecular com a enzima GAPDH de *T. cruzi*.

Complexo cristalográfico GAPDH-IAA

Estudos de co-cristalização da enzima na presença de inibidores de origem natural e sintética são realizados rotineiramente nos laboratórios do CBME. Em colaboração com a doutoranda Tatiane L. Balliano, novos inibidores de origem sintética foram identificados e submetidos a ensaio de co-cristalização com a GAPDH (Figura 7.1). Os dados cristalográficos foram coletados na linha MX-1 do LNLS. Dentre eles, o cristal de GAPDH obtido na presença de solução contendo 5 mM do inibidor **IAA** foi o único que apresentou densidade eletrônica que pudesse ser atribuída à molécula do respectivo inibidor. Os dados cristalográficos e parâmetros estatísticos dos conjuntos de dados coletados estão detalhados na Tabela 7.1.

Figura 7.1 - Inibidores da GAPDH de *T. cruzi* utilizados nos ensaios de co-cristalização.

Tabela 7.1 - Parâmetros cristalográficos e estatísticos dos conjuntos dados coletados no LNLS.

	PR-1	PR-2	PR-3	PR-4	IAA
Faixa de resolução (Å)	20,0 - 2,15	20,0 - 2,15	20,0 - 2,40	20,0 - 2,10	20,0 - 2,30
Parâmetros de rede (Å)	a= 81,30	a=78,85	a=79,88	a=81,45	a=79,98
	b=83,47	b=84,93	b=83,17	b=83,79	b=83,60
	c=103,98	c=106,16	c=102,95	c=102,87	c=101,95
	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$	$\alpha=\gamma=90^\circ$
	$\beta=96,17^\circ$	$\beta=95,70^\circ$	$\beta=96,03^\circ$	$\beta=96,44^\circ$	$\beta=95,93^\circ$
Nº total de imagens	406	186	212	210	210
$\Delta\phi$ (°)	1	1	1	1	1
Nº total de reflexões	251.466 (35.574)*	232.271 (33.748)*	199.759 (28.989)*	200.042 (25.611)*	200.472 (27.936)*
Nº de reflexões únicas	70.242 (10.168)*	74.208 (11.325)*	52.568 (7.283)*	58.836 (8.235)*	60.749 (9.383)*
Completeza (%)*	99,2 (99,2)*	97,7 (87,4)*	99,8 (99,8)*	98,4 (97,8)*	96,4 (95,8)*
R_{merge} (%)**	7,0 (17,2)	8,9 (12,6)	10,7 (35,2)	8,2 (48,2)	10,2 (40,2)
$\langle I \rangle / \sigma(I)$	8,8 (4,2)	7,4 (2,0)	6,3 (2,2)	6,3 (2,4)	7,3 (2,2)
Multiplicidade	3,6 (3,5)	3,1 (3,0)	3,8 (4,0)	3,4 (3,1)	3,3 (3,0)

*Valores obtidos para a camada mais externa de resolução.

A análise dos mapas iniciais de densidade eletrônica ($2F_{obs}-F_{calc}$ e $F_{obs}-F_{calc}$) revelou uma densidade eletrônica contínua muito próxima a cisteína catalítica (Cys166), sugerindo que o inibidor **IAA** estaria ligado covalentemente ao átomo de enxofre na cadeia lateral deste resíduo. Vários ciclos de refinamento foram realizados e a análise do modelo final do complexo GAPDH-**IAA** explica, indubitavelmente, a densidade eletrônica observada no sítio ativo (Figura 7.2). A estrutura final refinada apresenta um tetrâmero de GAPDH na unidade assimétrica, 923 moléculas de água, quatro moléculas de NAD^+ , uma molécula do inibidor **IAA** covalentemente ligada a Cys166 do monômero C, resolução de 2,3 Å e valores de $R_{\text{factor}}=0,19$ e $R_{\text{free}}=0,24$.

Figura 7.2 - Mapa de densidade eletrônica $2F_{obs}-F_{calc}$ referente ao sítio ativo da GAPDH ilustrando o resíduo cisteína covalentemente ligada ao inibidor **IAA** (CCS166) (contorno de 1.0 σ). Em destaque os principais resíduos, molécula de água e NAD^+ que interagem com o ligante. As ligações de hidrogênio estão indicadas como linhas tracejadas amarelas.

O inibidor **IAA** é uma molécula conhecida por sua capacidade de inibir especificamente a GAPDH.^{1,2} O mecanismo de ação desse composto foi caracterizado como uma inibição irreversível ocasionada pela formação de uma ligação covalente entre **IAA** e a cisteína catalítica (Figura 7.3). Essa ligação impediria a formação do complexo catalítico enzima/ NAD^+ e a transferência de cargas entre a cisteína e o NAD^+ que ocorre durante a etapa de desidrogenação do substrato.³⁻⁵

Figura 7.3 - Mecanismo de inativação da GAPDH de *T. cruzi* pelo inibidor **IAA**.

Neste estudo, pode-se observar que o inibidor **IAA** está ligado covalentemente a cisteína catalítica, entretanto, diferentemente do mecanismo previamente proposto,³⁻⁵ o cofator está simultaneamente presente no sítio ativo da GAPDH. De fato, a molécula do NAD^+ auxilia na estabilização do modo de ligação do inibidor **IAA**, uma vez que os elétrons- π do anel nicotinamida interagem com os elétrons- π do grupo carboxílico do **IAA**, através de uma interação do tipo “ π -stacking”, e pela formação de uma ligação de hidrogênio entre o oxigênio carboxílico do **IAA** e o grupo 3-OH’ do anel ribose intermediada por uma molécula de água (W324) (Figura 7.2). Assim, os resultados obtidos no presente estudo sugerem que o mecanismo pelo qual **IAA** inibe a GAPDH estaria relacionado com a etapa de formação do intermediário tiohemiacetal, ou seja, a cisteína catalítica covalentemente ligada ao **IAA** (CCS166) estaria impossibilitada de realizar o ataque nucleofílico ao grupo carbonila do substrato.^{6,7}

Um manuscrito descrevendo esses resultados está sendo elaborado e será submetido para o periódico científico *Chemical Communications*.

Mecanismo de cooperatividade da GAPDH de T. cruzi

Planeja-se conduzir experimentos de mutação sítio-dirigida para obtenção de enzimas GAPDH mutantes. As mutações sugeridas T54G, T54S, T54A possuem modificações estruturais significativas em posições chaves da interface de dimerização que exploram diferentemente as interações inter-subunidades que podem estar relacionadas com o comportamento não cooperativo. Além disso, serão realizados estudos cinéticos e cristalográficos com as enzimas mutantes para comprovação da hipótese sugerida de que o resíduo Thr54 exerce função central para o comportamento não cooperativo observado. Esses resultados serão combinados aos resultados já obtidos para a conclusão do trabalho sobre o mecanismo de cooperatividade da GAPDH de *T. cruzi*.

Um manuscrito descrevendo todas as etapas deste estudo esta em fase de elaboração e será submetido para publicação em periódico especializado a ser definido.

Triagem virtual de compostos

Serão realizados ensaios cinéticos *in vitro* para a avaliação da atividade inibitória das 35 moléculas identificadas nos estudos de triagem virtual. Se obtivermos sucesso no processo de triagem, os novos inibidores da GAPDH de *T. cruzi* terão a sua potência e mecanismo de ação determinados, e serão posteriormente submetidos a experimentos de cristalização, difração de raios X e modelagem molecular. Os resultados do trabalho de triagem virtual associados aos dos ensaios de inibição (se forem positivos) deverão ser incorporados em um novo manuscrito a ser submetido para publicação em periódico especializado.

Artigos em preparação

Atualmente encontram-se em fase final de redação e revisão outros cinco artigos e um capítulo de livro:

- ✓ Um artigo em colaboração com a Profa. Dra. Elizabeth I. Ferreira e o doutorando Gustavo H. G. Trossini (FCF-USP), entre outros autores, descrevendo os resultados de QSAR 3D CoMFA e CoMSIA para o conjunto de dados de inibidores da enzima cruzaina de *T. cruzi*.
- ✓ Um artigo em colaboração com o Prof. Dr. Adriano D. Andricopulo e a doutoranda Livia B. Salum (IFSC), entre outros autores, descrevendo os resultados de QSAR e docagem molecular com um conjunto de dados de 47 derivados do discodermolídeo, potentes agentes estabilizadores de microtúbulos e potencialmente úteis para a terapia do câncer .
- ✓ Um artigo em colaboração com as Profa. Dra. Quézia B. Cass (IQ-UFSCar) e com a Profa. Dra. Carmen L. Cardoso (FFCL-RP - USP), entre outros autores, discutindo as bases estruturais relacionadas as diferenças observadas nas constantes cinéticas entre a GAPDH de *T. cruzi* imobilizada em coluna cromatográfica e livre em solução.
- ✓ Um artigo em colaboração com a Dra. Monique Mantovani (Eurofarma) e o Prof. Dr. Otavio H. Thiemann (IFSC), entre outros autores, descrevendo os resultados da caracterização molecular, estrutural e cinética da enzima adenilosuccinato liase de *L. major* (ADSL, EC 4.3.2.2).
- ✓ Um artigo em colaboração com o Prof. Dr. Paulo Cezar Vieira e a doutoranda Richele Severino (IQ-UFSCar), entre outros autores, descrevendo os resultados de identificação, determinação da atividade biológica e modelagem molecular de produtos naturais como inibidores da enzima catepsina V humana (Cath-V, EC 3.4.22.43).

- ✓ Um capítulo de livro intitulado *Planejamento de Fármacos*, de autoria de Rafael V. C. Guido, Marcelo S. Castilho, Glaucius Oliva e Adriano D. Andricopulo que será publicado no livro *Química Farmacêutica*, editado por Eliezer J. L. Barreiro e Elizabeth I. Ferreira.

7.3 - Referências

1. THOMPSON, J. Lactose metabolism in *Streptococcus lactis*: phosphorylation of galactose and glucose moieties in vivo. *J. Bacteriol.*, v. 140, p. 774-785, 1979.
2. POOLMAN, B. et al. Control of glycolysis by glyceraldehyde-3-phosphate dehydrogenase in *Streptococcus cremoris* and *Streptococcus lactis*. *J. Bacteriol.*, v. 169, p. 5887-5890, 1987.
3. EVEN, S. et al. Pyruvate metabolism in *Lactococcus lactis* is dependent upon glyceraldehyde-3-phosphate dehydrogenase activity. *Metab. Eng.*, v. 1, p. 198-205, 1999.
4. NAGRADOVA, N. K. et al. D-Glyceraldehyde-3-phosphate dehydrogenase. Properties of the enzyme modified at arginine residues. *Appl. Biochem. Biotechnol.*, v. 61, p. 47-56, 1996.
5. AMELUNXEN, R. E.; CARR, D. O. Glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle. *Methods Enzymol.*, v. 41, p. 264-267, 1975.
6. NAGRADOVA, N. K. Study of the properties of phosphorylating D-glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry (Mosc.)*, v. 66, p. 1067-1076, 2001.
7. CASTILHO, M. S. et al. Evidence for the two phosphate binding sites of an analogue of the thioacyl intermediate for the *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase-catalyzed reaction, from its crystal structure. *Biochemistry*, v. 42, p. 7143-7151, 2003.

Apêndices

Classical and Hologram QSAR Studies on a Series of Inhibitors of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase

Rafael V. C. Guido^a, Marcelo S. Castilho^b, Sabrina G. R. Mota^b, Glaucius Oliva^a and Adriano D. Andricopulo^{a*}

^a Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970 São Carlos, SP, Brazil, E-mail: aandrico@ifsc.usp.br

^b Laboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-290 Salvador, BA, Brazil

Keywords: Enzyme inhibitors, GAPDH, QSAR, Tropical diseases, Trypanosomatids

Received: September 26, 2007; Accepted: February 02, 2008

DOI: 10.1002/qsar.200710139

APÊNDICE A

Abstract

Leishmaniasis and trypanosomiasis are major causes of morbidity and mortality in both tropical and subtropical regions of the world. The current available drugs are limited, ineffective, and require long treatment regimens. Due to the high dependence of trypanosomatids on glycolysis as a source of energy, some glycolytic enzymes have been identified as attractive targets for drug design. In the present work, classical Two-Dimensional Quantitative Structure–Activity Relationships (2D QSAR) and Hologram QSAR (HQSAR) studies were performed on a series of adenosine derivatives as inhibitors of *Leishmania mexicana* Glyceraldehyde-3-Phosphate Dehydrogenase (LmGAPDH). Significant correlation coefficients (classical QSAR, $r^2=0.83$ and $q^2=0.81$; HQSAR, $r^2=0.91$ and $q^2=0.86$) were obtained for the 56 training set compounds, indicating the potential of the models for untested compounds. The models were then externally validated using a test set of 14 structurally related compounds and the predicted values were in good agreement with the experimental results (classical QSAR, $r^2_{\text{pred}}=0.94$; HQSAR, $r^2_{\text{pred}}=0.92$).

1 Introduction

Parasitic diseases are the foremost threat to human health and welfare around the world. In tropical and subtropical regions of the world, the consequences of parasitic infection are devastating both in terms of human morbidity and mortality, reaching dangerous levels which threaten future social stability and economic development of these areas. The currently available drugs for most of the parasitic diseases are inadequate, poorly tolerated, and ineffective [1–3]. A group of tropical diseases, including leishmaniasis and trypanosomiasis, is neglected and lie outside of the world pharmaceutical market. On one hand, it is estimated that 12 million people are currently affected by leishmaniasis in its different forms, with about 350 million living in risk areas. The most severe form, visceral leishmaniasis or

kala azar, is a serious disease that affects 500 000 people resulting in the death of over 65 000 every year [4]. On the other hand, Chagas' disease (also called American trypanosomiasis) is a tropical parasitic disease which occurs in Latin America, particularly in South America, where approximately 16–18 million people are infected causing over 45 000 deaths each year [5, 6].

Since safe, effective, and affordable chemotherapeutic agents against trypanosomatids are urgently needed, the identification of new macromolecular targets and small-molecule modulators is of utmost importance. The bloodstream form of trypanosomatids has no functional tricarboxylic acid cycle and is highly dependent on glycolysis for ATP production [7, 8]. The crucial dependence on glycolysis as a source of energy makes the glycolytic parasite enzymes promising targets for drug design. Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH, Enzyme Commission 1.2.1.12) is a key enzyme in the glycolytic cascade, which catalyzes the reversible oxidative phosphorylation of the substrate Glyceraldehyde-3-Phosphate (GAP) to 1,3-diphosphoglycerate in the presence of the cofactor Nicotinamide Adenine Dinucleotide (NAD^+) and inor-

ganic phosphate. The glycolytic GAPDH enzymes from the pathogenic protozoan parasites *Trypanosoma brucei* [9], *T. cruzi* [10], and *Leishmania mexicana* [11] are quite similar to each other (about 90% sequence identity) and have important structural differences when compared to their human counterpart [12] (about 45% sequence identity). Comparisons between parasite and human GAPDH crystal structures revealed significant differences in the neighboring of the NAD⁺ (adenosine moiety) binding site, which have been explored for the development of new selective inhibitors of the trypanosomatid enzymes [13–17].

Quantitative Structure–Activity Relationships (QSAR) have been successfully employed as a valuable tool to assist the design of compounds of pharmaceutical interest [18–20]. As part of a research program aimed at the design of selective GAPDH inhibitors [21–24], and in order to evaluate the quantitative relationships of a large series of adenosine derivatives, we have created predictive 2D QSAR models employing both classical and hologram QSAR (HQSAR) methods [25–27].

2 Experimental Section

2.1 Data Sets

The data set used for the QSAR studies contains 70 inhibitors of GAPDH (adenosine derivatives), which were selected from the literature [13–17]. The structures and corresponding pIC₅₀ (–log IC₅₀; where IC₅₀ is the concentration required for 50% inhibition of GAPDH) values for the whole set of inhibitors are included in Table 1. The complete data set was divided into training (compounds 1–56, Table 1) and test (compounds 57–70, Table 1) sets in the ratio of 5:1 (20%). A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) using the complete linkage clustering method (Euclidean distances) with no data standardization. The cluster analysis grouped the compounds into two major clusters, with a relatively well-balanced distribution of training and test set compounds. Considering that the structural diversity of the data set is mostly represented by adenine analogues of moderate structural diversity, the composition of the QSAR training and test sets is necessarily representative of the whole data set in terms of chemical similarity.

2.2 Descriptors Calculation and Selection

Classical 2D QSAR studies require the calculation of a variety of molecular descriptors that are used as independent variables in QSAR modeling. In the present work, these studies were carried out using the DRAGON 5.4 (Taletto SRL, Milan, Italy), BUILDQSAR [28], and PIROUETTE 3.11 software (Infometrix, Washington, USA). The 2D

molecular descriptors, including topological descriptors, connectivity indices, 2D autocorrelation descriptors, Burden eigenvalues indices, among others, were computed using the software DRAGON 5.4 and used as independent variables in the QSAR analyses. Approximately 930 molecular descriptors were subjected to the following selection criteria. Descriptors with constant values or found to have poor correlation to biological property were discarded ($r^2 < 0.10$). This strategy afforded 800 descriptors. Then, the BUILDQSAR software was employed to systematically search for models of up to 4 variables that give rise to MLR models with $r^2 > 0.70$. All descriptors present in the MLR models were pooled together, autoscaled, and used for the Partial Least Square (PLS) analysis performed with the PIROUETTE software.

2.3 HQSAR Analysis

The HQSAR modeling analyses, calculations, and visualizations were performed using the SYBYL 7.3 package (Tripos Inc., St. Louis, USA) running on Red Hat Linux 7.3 workstations, as previously described [25, 27]. This method requires selecting values for parameters that specify the length of the hologram as well as the size and type of the fragments that are to be encoded [25, 27]. The generation of the molecular holograms was carried out using several combinations of the following fragment distinction: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). The fragment size default (4–7) was used to generate the QSAR models. HQSAR analysis was performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. The influence of fragment size, which controls the minimum and maximum length of fragments to be included in the hologram, was further investigated by using 6 distinct fragment sizes over the 12 default series of hologram lengths values ranging from 53 to 401 bins. The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using the PLS analysis.

2.4 QSAR Model Validation

All QSAR models were investigated using full cross-validated r^2 (q^2) PLS. Leave-One-Out (LOO) cross-validation was applied to determine the number of components that yield optimally predictive models. Leave-Many-Out (LMO) with 10 (LMO₁₀) and 5 (LMO₅) randomly selected groups was used as a more rigorous test to assess model stability and statistical significance. Each random cross-validation run was repeated 25 times to obtain mean values for q^2 and the corresponding Standard Error of Prediction (SDEP). External validation was performed with a test set of 14 compounds which were not considered for QSAR model generation.

Table 1. Chemical structures and corresponding pIC₅₀ values for a series of inhibitors of *Lm*GAPDH.

Training set compounds

Compound	Structure	pIC ₅₀	Compound	Structure	pIC ₅₀
1		3.30	2		2.52
3		3.15	4		2.40
5		2.80	6		2.22
7		3.40	8		3.60
9		3.12			

Table 1. (continued)

Training set compounds

Compound	R^1	R^2	R^3	pIC ₅₀
10	CH ₃	H	H	2.22
11	H	H		2.40
12	H		H	2.43
13	H		H	3.15
14	H		H	3.15
15	H		H	2.74
16	H		H	3.44
17	H		H	3.47
18	H		H	3.62

Table 1. (cont.)

Training set compounds

Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
19	H	H	H	H	H	H	2.48
20	H	H	H	H	C ₆ H ₅	H	3.70
21	H	H	H	OCH ₂ CH ₃	H	H	3.40
22	H	H	H	OCH ₃	OH	H	3.30
23	H	Br	H	H	H	H	2.52
24	H	H	H	OH	OH	OH	3.60
25	H	H	H	OH	H	OH	3.19
26		H	H	OCH ₃	H	OCH ₃	4.70
27		H	H	OCH ₃	H	H	4.80
28		H	H	OCH ₃	H	OCH ₃	4.60
29		H	H	OCH ₃	H	H	4.60
30		H	H	OCH ₃	H	OCH ₃	4.60

Training set compounds

Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
31		H	H	OCH ₃	H	H	4.60
32		H	H	OCH ₃	H	H	5.26

Table 1. (cont.)

Training set compounds							
Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
33		H	H	OCH ₃	H	H	3.82
34		H	H	OCH ₃	H	H	4.10
35		H	H	Cl	H	H	5.00
36		H	H	OCH ₃	H	OCH ₃	5.70
37		H	H	Cl	H	Cl	4.92
38		H	H	OAc	H	OAc	5.00
39		H	H	OH	H	OH	5.30
40		H	H	H	OAc	H	4.00
Training set compounds							
Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
41		H	H	H	OH	H	4.10
42		H	H	OAc	OAc	H	4.60

Table 1. (cont.)

Training set compounds							
Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
43		H	H	OH	OCH ₃	H	4.43
44		H	H	OCH ₃	OH	H	5.00
45		H	H	N(CH ₃) ₂	H	H	4.60
46		H	H	OCH ₃	H	OCH ₃	5.70
47		H	H	Cl	H	Cl	4.60
48		H	H	OCH ₃	H	OCH ₃	4.22
49		H	H	OCH ₃	H	H	5.22
50		H	H	Cl	H	H	5.40
Training set compounds							
Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
51		H	H	OCH ₃	H	OCH ₃	5.70
52		H	H	Cl	H	Cl	4.60
53		H	H	OCH ₃	H	H	5.70

Table 1. (cont.)

Training set compounds

Compound	R^1	R^2	R^3	R^4	R^5	R^6	pIC ₅₀
54		H	H	Cl	H	Cl	4.60
55		H	H	OCH ₃	H	H	5.30
56		H	H	OCH ₃	H	H	4.08

Compound	Structure	pIC ₅₀	Compound	Structure	pIC ₅₀	Compound	Structure	pIC ₅₀
57		2.62	58		3.52	59		3.30
60		3.22	61		3.07	62		3.82

Table . (cont.)

CompoundStructure	pIC ₅₀ CompoundStructure	pIC ₅₀ CompoundStructure	pIC ₅₀ CompoundStructure
63	5.30	64	5.40
		65	5.70
CompoundStructure	pIC ₅₀ CompoundStructure	pIC ₅₀ CompoundStructure	pIC ₅₀ CompoundStructure
66	4.30	67	4.60
		68	4.00
69	4.74	70	5.00

3 Results and Discussion

3.1 Chemical and Biological Data

Classical QSAR and HQSAR models were derived for a series of 70 inhibitors of trypanosomatid GAPDHs listed in Table 1 [13–17]. The family of adenosine derivatives examined has moderate structural diversity. The bulk of the structural diversity lies in the nature of the substituent linked to the N⁶- and 9-position of the adenine system. The *in vitro* IC₅₀ values of *Lm*GAPDH employed in this work were measured under the same experimental conditions [13–17], a fundamental requirement for QSAR studies [29, 30]. The generation of consistent statistical models is dependent on the adequacy of the training and test sets. Therefore, from the original data set, 56 inhibitors (compounds 1–56, Table 1) were selected as members of the training set for model generation, whereas the other 14 inhibitors (compounds 57–70, Table 1) were held as members of the test set for external validation. A statistical cluster analysis confirmed that structurally diverse molecules possessing activities of wide range were included in both training and test sets. The pIC₅₀ values used as dependent variables in the QSAR analyses span about three and a half orders of magnitude and are acceptably distributed across the range of values. Thus, the data set is appropriate for the purpose of QSAR model development.

3.2 Classical QSAR Analysis

The strategy used in this work with the program DRAGON 5.4 resulted in the selection of 12 descriptors (Table 2) that were then explored using more robust statistical methods such as Principal Component Analysis (PCA) and PLS, as implemented in the PIROUETTE software.

The PCA results indicated that three principal components accounted for 95% of total variance, while other additional components did not significantly contribute to these models. The first PC broadly accounts for potency,

while PC2 roughly describes molecular steric properties. The less potent inhibitors have negative PC1 values whereas the most potent ones display positive values. These preliminary results prompted us to use the selected descriptors for QSAR modeling studies. PLS QSAR models were created using the LOO cross-validation procedure, and the results are shown in Figure 1. As it can be seen, the best statistical model has substantial correlation within the training set as shown by the high values of r^2 (0.83) and q^2 (0.81). The use of the LMO cross-validation protocols did not lead to considerable quantitative changes in the parameters of the final models (LMO₅ q^2 =0.81, and LMO₁₀ q^2 =0.80), thus confirming the stability and statistical significance of the model generated with LOO.

Figure 1. Plot of predicted vs. experimental values of pIC₅₀ for the training (open squares) and test (black triangle) set GAPDH inhibitors for the best classical QSAR model.

Table 2. Descriptors selected for QSAR model development^a.

Symbol	Definition and description
Har2	Square reciprocal distance sum index
dZ	Pogliani index
SRW01	Self-returning walk count of order 01
RHyDp	Reciprocal hyper-distance-path index
HaR	Harary H index
CID	Randic ID number
Pol	Polarity number
EPS0	Edge connectivity index of order 0
EEig02d	Eigenvalue 02 from edge adjacency matrix weighted by dipole moments
O-058	Atom centered fragments corresponding to carbonyl
O-060	Atom centered fragment corresponding to Al–O–Ar/Ar–O–Ar–R–O–R/R–O–C=X
TPSA(NO)	Topological polar surface area using Nitrogen and oxygen polar contributions

^aDescriptors considered in the final model are highlighted in Table, and were selected according to the regression vector score.

Table 3. Experimental and predicted activities (pIC_{50}) with residual values for the 14 test set compounds.

Com- pound	Experi- mental	Classical QSAR		HQSAR	
		Predicted	Residual ^a	Predicted	Residual ^a
Test set					
57	2.62	2.51	0.11	2.63	−0.01
58	3.52	3.26	0.26	3.35	0.17
59	3.30	3.33	−0.03	3.32	−0.02
60	3.22	3.42	−0.20	3.27	−0.05
61	3.07	2.80	0.27	3.17	−0.10
62	3.82	3.72	0.10	3.71	0.11
63	5.30	4.88	0.42	5.31	−0.01
64	5.40	5.36	0.04	5.12	0.28
65	5.70	5.96	−0.26	5.75	−0.05
66	4.30	4.50	−0.20	4.91	−0.61
67	4.60	4.27	0.33	4.87	−0.27
68	4.00	4.27	−0.27	4.66	−0.66
69	4.74	4.43	0.31	4.72	0.02
70	5.00	5.12	−0.12	4.99	0.01

^aThe difference between experimental and predicted values.

Although cross-validation protocols give a suitable representation of the internal consistency and predictive power of QSAR models, they may lead to very optimistic statistics [27–29]. The predictive power of the best QSAR model derived using the 56 training set molecules (compounds 1–56, Table 1) was assessed by predicting the biological property for 14 test set compounds (57–70, Table 1). The results of the external validation process are listed in Table 3. The predicted values fall close to the experimental pIC_{50} values, not deviating by more than 0.42 log units. No outliers were detected in this series of GAPDH inhibitors. The good agreement between actual and predicted pIC_{50} values for the test set compounds suggests that the constructed model is reliable and has predictive power for new inhibitors within this structural class.

In addition to their ability to predict the biological property of interest, QSAR models should also provide useful chemical insights for drug design. In this context, Figure 2 illustrates the influence of the descriptors over the regression vector, which can be regarded as a weighted sum of the loadings included in the final model. This information was used to select the six most relevant descriptors during the development of the QSAR models, highlighted in Table 2.

It can be noted that Har2 and EEig2d represent, respectively, the most important positive and negative contributions to the final QSAR model. Descriptors with small coefficients did not significantly affect the QSAR model. Har2 [31, 32] is a topological descriptor derived from distance matrix which accounts for through-bond interactions in molecular systems. Due to its reciprocal nature, central atoms have larger influence than the external ones on the final summed value. This is clearly observed in our data set once the structurally related and nearly equipotent compounds 46, 51, and 53 show different Har2 values (192,

Figure 2. Descriptors score to regression vector.

201, and 207, respectively). This result indicates that structural modifications that do not significantly alter the main scaffold have minor effects on the inhibitory potency. Additional insight into the influence of molecular shape on the inhibitory potency comes from the interpretation of the EEig02 values [33, 34]. While the most potent inhibitors have the same EEig02d index (3.49), decreases in the inhibitory potency were observed for larger or smaller values. Although the classical QSAR analysis has extracted valuable SAR data, the topological descriptors (*e.g.*, Har2, EEig02d) provide only information about the character of the whole molecule. Accordingly, the information collected does not differentiate major and minor accounts from individual molecular fragments to the inhibitory potency. In the present work, in order to gain further insight into the fragment-based structure–activity relationships for this series of GAPDH inhibitors, and also for the better understanding of the synergies and challenges between these approaches, we have employed the HQSAR method [25, 27].

3.3 HQSAR Analysis

The specialized fragment fingerprint (called molecular hologram) consists of structural fragments of a molecule that are counted in bins of a fixed length array. The bins represent all of the unique fragments included within a particular molecule and are assigned by a cyclic redundancy check (CRC) algorithm. The bin occupancies of the molecular hologram are structural descriptors (independent variables) encoding compositional and topological molecular information. Accordingly, the hologram includes information on the quantity and type of each fragment in each molecule. HQSAR encodes all possible molecular fragments (*i.e.*, linear, branched, and overlapping) without requiring any explicit 3D information of the ligands (*e.g.*, determination of 3D structure, putative binding conformations, and molecular alignment). The generation of molecular holograms was carried out using the fragment size de-

Table 4. HQSAR analysis for various fragment distinctions on the key statistical parameters using fragment size default (4–7).

Model	Fragment distinction	q^2	SDEP	r^2	SEE	HL	PC
1	A/B	0.831	0.439	0.885	0.355	257	3
2	A/B/C	0.835	0.436	0.888	0.350	353	3
3	A/B/C/H	0.836	0.429	0.884	0.362	307	3
4	A/B/C/H/Ch	0.835	0.432	0.887	0.357	353	3
5	A/B/C/H/Ch/DA	0.823	0.447	0.886	0.359	353	3
6	A/B/H	0.851	0.409	0.895	0.343	151	3
7	A/B/C/Ch	0.832	0.437	0.884	0.356	353	3
8	A/B/DA	0.834	0.435	0.894	0.345	97	3
9	A/B/C/DA	0.827	0.442	0.886	0.359	199	3
10	A/B/H/DA	0.848	0.414	0.896	0.342	151	3
11	A/B/C/H/DA	0.821	0.449	0.876	0.374	151	3
12	A/B/H/Ch/DA	0.855	0.405	0.903	0.331	307	3

Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor. q^2 , cross-validated correlation coefficient; SDEP, standard error of prediction; r^2 , noncross-validated correlation coefficient; SEE, noncross-validated standard error; HL, hologram length; PC, optimal number of components.

fault (4–7), the 12 default series of hologram lengths (53–401), and the following combinations of fragment distinction: AB, ABC, ABCH, ABCHCh, ABCHChDA, ABH, ABCCh, ABDA, ABCDA, ABHDA, ABCHDA, and ABHChDA. The statistical results from the PLS analyses for the 56 training set inhibitors using several fragment distinction combinations are summarized in Table 4.

Statistical models of very good quality were obtained using different combinations of fragment distinction as revealed by their high q^2 and r^2 values. The best statistical results were obtained for models **6**, **10**, and **12** (Table 4), which were derived using A/B/H, A/B/H/DA, and A/B/H/Ch/DA, respectively. The influence of different fragment sizes, which control the minimum and maximum length of

fragments to be included in the hologram, was further investigated for the three best HQSAR models as shown in Table 5. As it can be seen, the variation of the fragment size led to the generation of slightly better HQSAR models (models **16**, **22**, and **28**, generated with the fragment size 5–8).

In order to investigate the statistical significance and stability of the models, a more consistent cross-validation test was carried out for the three best HQSAR models using 10 and 5 groups (LMO₁₀ and LMO₅, respectively). Each model was evaluated 25 times by measuring its accuracy in predicting the activity of the remaining about 10 and 20% dataset compounds. The results of the LMO analyses are presented in Table 6, confirming the stability and robustness of the models generated with LOO.

The predictive ability of the models was assessed using the same test set compounds employed in the classical 2D QSAR studies (compounds **57–70**, Table 1). In order to select the best predictive statistical model, values of r^2_{pred} (predictive- r^2) were calculated for the three models **16**, **22**,

Table 6. Mean values of q^2 and SDEP obtained after 25 independent runs of LMO cross-validation protocol.

Fragment distinction	Model	LMO ₁₀ ^a		LMO ₅ ^b	
		q^2	SDEP	q^2	SDEP
A/B/H	16	0.861	0.396	0.858	0.399
A/B/H/DA	22	0.839	0.425	0.840	0.424
A/B/H/Ch/DA	28	0.838	0.427	0.841	0.422

Fragment distinction: A, atoms; B, bonds; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

a Mean value of q^2 and SDEP obtained after 25 independent runs of cross-validation protocol with 10% of compounds left out.

b Mean value of q^2 and SDEP obtained after 25 independent runs of cross-validation protocol with 20% of compounds left out.

Table 5. HQSAR analysis for the influence of different fragment sizes on the statistical parameters.

Fragment distinction	Model	Fragment size	q^2	SDEP	r^2	SEE	HL	PC
A/B/H	13	2–5	0.837	0.428	0.880	0.368	61	3
	14	3–6	0.845	0.418	0.888	0.355	257	3
	15	4–7	0.851	0.409	0.895	0.343	151	3
	16	5–8	0.864	0.391	0.905	0.326	59	3
	17	6–9	0.840	0.422	0.892	0.348	353	3
A/B/H/DA	18	7–10	0.835	0.431	0.883	0.362	353	3
	19	2–5	0.835	0.431	0.887	0.357	71	3
	20	3–6	0.837	0.428	0.889	0.354	257	3
	21	4–7	0.848	0.414	0.896	0.342	151	3
	22	5–8	0.857	0.399	0.895	0.344	307	3
A/B/H/Ch/DA	23	6–9	0.836	0.429	0.888	0.355	97	3
	24	7–10	0.838	0.427	0.884	0.360	257	3
	25	2–5	0.839	0.426	0.889	0.354	71	3
	26	3–6	0.843	0.420	0.894	0.345	97	3
	27	4–7	0.855	0.405	0.903	0.331	307	3
	28	5–8	0.860	0.397	0.897	0.340	97	3
	29	6–9	0.821	0.449	0.870	0.383	71	3
	30	7–10	0.808	0.465	0.858	0.400	71	3

Figure 3. Plot of predicted vs. experimental values of pIC_{50} for the training (open squares) and test (black triangle) set GAPDH inhibitors for the best HQSAR model.

and **28**, which presented r^2_{pred} of 0.92, 0.83, and 0.85, respectively. The results of the external validation for the best predictive model (**16**, $r^2_{pred} = 0.92$) are shown in Table 3, and the graphic results for the experimental *versus* predicted for both training and test sets are shown in Figure 3. As can be seen, the test set compounds are well predicted within the range of values, not deviating by more than 0.66 log units. Therefore, the model presents good correlative and predictive abilities.

Besides predicting the property value of interest of untested molecules, HQSAR models should also provide hints about the relationships of different molecular fragments to biological activity [25, 27]. HQSAR models can be graphically represented in the form of contribution maps where the color of each molecular fragment reflects the contribution of an atom or a small number of atoms to the activity of the molecule under study. The colors at the red end of the spectrum (e.g., red, red–orange, and orange) reflect poor (or negative) contributions, while colors at the green end (e.g., yellow, green–blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are colored white. For example, the most important fragments of compound **36** (one of the most potent inhibitor of the data set) are shown in Figure 4. The HQSAR contribution map shows that the substituted benzamide ring has a favorable contribution to the inhibitory activity. This observation is in good agreement with previously reported results [14], and it is further supported by the fact that when the benzamide ring is replaced by the quinoline ring, biphenyl rings (**1**, **2**) or is absent (**8**–**18**), it causes a significant decrease in potency.

Figure 4. HQSAR contribution map for the GAPDH inhibitor **36**.

Furthermore, the contribution maps of compounds (**19**–**56**) indicate that a different substitution pattern around the benzamide moiety is tolerable by the receptor (maps not shown) as previously described [17], and simultaneously, suggest a positive influence of *meta* substituents for the inhibitory potency. This can be seen, for instance, in the analysis of compound **36** ($IC_{50} = 2 \mu M$) for which the incorporation of a methoxy substituent at the *meta*-position of the benzamide ring results in 50- and 40-fold increase in potency when compared to the *para*-substituted compounds **40** ($IC_{50} = 100 \mu M$) and **41** ($IC_{50} = 80 \mu M$), respectively.

HQSAR is an important drug design technique which encodes useful information that combines the fragment-based and topological description of molecular structures. Nevertheless, for a more robust QSAR analysis, it is important to incorporate additional detailed description of molecular properties that are relevant for activity (e.g., physicochemical parameters). For this reason, the integration of classical and HQSAR methods are complementary in nature, especially because vital components such as steric or electrostatic play a major role towards the inhibitory potency of series of compounds. In this context, it can be demonstrated that compound **36** has higher Har2 index (59.51) than compounds **40** (50.86) and **41** (52.81), whereas the EEig02d value is optimal for all the three compounds (3.49). These results are confirmed by the HQSAR contribution maps, and also suggest that *meta* substituted benzamide adenosine analogs have a steric component that is favorable for potency and hence show increased inhibitory potency, confirming previously reported findings [35].

4 Conclusion

The classical and hologram QSAR models described in this work exhibit both good internal and external consistency, with substantial predictive power. It is worth noting that we have employed the same combination of training and test sets for both QSAR analyses. The highly consistent results confirmed that investigations can be carried out concomitantly to search for synergies between 2D QSAR technologies, revealing a potential of shedding some light on the effects of the substitution pattern over molecular fragments within the data set compounds. Finally, as the models reported herein have been internally and externally validated employing different 2D QSAR approaches, they appear to be valuable tools for ligand-based enzyme inhibitor design.

Acknowledgements

We gratefully acknowledge financial support from the State of São Paulo Research Foundation (FAPESP), the State of Bahia Research Foundation (FAPESB), and the National Council for Scientific and Technological Development (CNPq), Brazil.

References

- [1] B. M. Watkins, *Trends Parasitol.* **2003**, *19*, 477–478.
- [2] B. D. Nicholson, J. D. Walley, D. S. Baguley, *Trop. Med. Int. Health* **2006**, *11*, 1339–1340.
- [3] J. H. McKerrow, C. Caffrey, B. Kelly, P. Loke, M. Sajid, *Annu. Rev. Pathol. Mech. Dis.* **2006**, *1*, 497–536.
- [4] S. L. Croft, G. H. Coombs, *Trends Parasitol.* **2003**, *19*, 502–508.
- [5] World Health Organization, *Tech. Rep. Ser.* **2002**, *905*, 1–109.
- [6] C. E. Sanz-Rodriguez, J. L. Concepcion, S. Pekerar, E. Oldfield, J. A. Urbina, *J. Biol. Chem.* **2007**, *282*, 12377–12387.
- [7] K. J. Kennedy, J. C. Bressi, M. H. Gelb, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 95–98.
- [8] G. C. Guerra, L. Quijada, C. E. Clayton, *J. Cell Sci.* **2002**, *115*, 2651–2658.
- [9] F. M. Vellieux, J. Hajdu, C. L. M. J. Verlinde, H. Groendijk, R. J. Read, T. J. Greenhough, J. W. Campbell, K. H. Kalk, J. A. Littlechild, H. C. Watson, W. G. J. Hol, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2355–2359.
- [10] D. H. F. Souza, R. C. Garratt, A. P. Araujo, B. G. Guimaraes, W. D. Jesus, P. A. M. Michels, V. Hannaert, G. Oliva, *FEBS Lett.* **1998**, *424*, 131–135.
- [11] H. Kim, I. K. Feil, C. L. M. J. Verlinde, P. H. Petra, W. G. Hol, *Biochemistry* **1995**, *34*, 14975–14986.
- [12] J. L. Jenkins, J. J. Tanner, *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 290–301.
- [13] C. L. M. J. Verlinde, M. Callens, S. Van Calenbergh, A. Van Aerschot, P. Herdewijn, V. Hannaert, P. A. M. Michels, F. R. Opperdoes, W. G. Hol, *J. Med. Chem.* **1994**, *37*, 3605–3613.
- [14] S. Van Calenbergh, C. L. M. J. Verlinde, J. Soenens, A. De Bruyn, M. Callens, N. M. Blaton, O. M. Peeters, J. Rozen-skik, W. G. Hol, P. Herdewijn, *J. Med. Chem.* **1995**, *38*, 3838–3849.
- [15] A. M. Aronov, C. L. M. J. Verlinde, W. G. Hol, M. H. Gelb, *J. Med. Chem.* **1998**, *41*, 4790–4799.
- [16] A. M. Aronov, S. Suresh, F. S. Buckner, W. C. Van Voorhis, C. L. M. J. Verlinde, F. R. Opperdoes, W. G. Hol, M. H. Gelb, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4273–4278.
- [17] J. C. Bressi, C. L. M. J. Verlinde, A. M. Aronov, M. L. Shaw, S. S. Shin, L. N. Nguyen, S. Suresh, F. S. Buckner, W. C. Van Voorhis, I. D. Kuntz, W. G. Hol, M. H. Gelb, *J. Med. Chem.* **2001**, *44*, 2080–2093.
- [18] V. Farutin, L. Masterson, A. D. Andricopulo, J. Cheng, R. Riley, R. Hakimi, J. Frazer, E. H. Cordes, *J. Med. Chem.* **1999**, *42*, 2422–2431.
- [19] A. D. Andricopulo, R. A. Yunes, *Chem. Pharm. Bull.* **2001**, *49*, 10–17.
- [20] A. D. Andricopulo, A. Willian, R. Correa, A. R. S. Santos, R. J. Nunes, R. A. Yunes, V. Cechinel, *Pharmazie* **1998**, *53*, 493–498.
- [21] M. S. Castilho, F. Pavao, G. Oliva, S. Ladame, M. Willson, J. Perie, *Biochemistry* **2003**, *42*, 7143–7151.
- [22] S. Ladame, M. S. Castilho, C. H. Silva, C. Denier, V. Hannaert, J. Perie, G. Oliva, M. Willson, *Eur. J. Biochem.* **2003**, *270*, 4574–4586.
- [23] A. Leita, A. D. Andricopulo, G. Oliva, M. T. Pupo, A. A. de Marchi, P. C. Vieira, M. F. G. F. da Silva, V. F. Ferreira, M. C. B. V. de Souza, M. M. Sa, V. R. S. Moraes, C. A. Montanari, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2199–2204.
- [24] A. A. de Marchi, M. S. Castilho, P. G. Nascimento, F. C. Archanjo, G. del Ponte, G. Oliva, M. T. Pupo, *Bioorg. Med. Chem.* **2004**, *12*, 4823–4833.
- [25] M. S. Castilho, R. V. C. Guido, A. D. Andricopulo, *Bioorg. Med. Chem.* **2007**, *15*, 6242–6252.
- [26] M. S. Castilho, M. P. Postigo, C. B. de Paula, C. A. Montanari, G. Oliva, A. D. Andricopulo, *Bioorg. Med. Chem.* **2006**, *14*, 516–527.
- [27] K. M. Honorio, R. C. Garratt, A. D. Andricopulo, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119–3125.
- [28] D. B. Oliveira, A. C. Gaudio, *Quant. Struct. Act. Relat.* **2003**, *19*, 599–601.
- [29] K. M. Honorio, R. C. Garratt, I. Polikarpov, A. D. Andricopulo, *J. Mol. Graph. Model.* **2007**, *25*, 921–927.
- [30] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
- [31] M. Randic, A. F. Kleiner, L. M. DeAlba, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277–286.
- [32] M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 292–299.
- [33] E. Estrada, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
- [34] E. Estrada, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701–707.
- [35] S. Suresh, J. C. Bressi, K. J. Kennedy, C. L. M. J. Verlinde, M. H. Gelb, W. G. Hol, *J. Mol. Biol.* **2001**, *309*, 423–435.

Structural Basis for Selective Inhibition of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase: Molecular Docking and 3D QSAR Studies

Rafael V. C. Guido, Glaucius Oliva, Carlos A. Montanari, and Adriano D. Andricopulo*

Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil, and Grupo de Química Medicinal de Produtos Naturais, Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13566-970, São Carlos-SP, Brazil

Received December 6, 2007

APÊNDICE B

The glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is as an attractive target for the development of novel antitrypanosomatid agents. In the present work, comparative molecular field analysis and comparative molecular similarity index analysis were conducted on a large series of selective inhibitors of trypanosomatid GAPDH. Four statistically significant models were obtained ($r^2 > 0.90$ and $q^2 > 0.70$), indicating their predictive ability for untested compounds. The models were then used to predict the potency of an external test set, and the predicted values were in good agreement with the experimental results. Molecular modeling studies provided further insight into the structural basis for selective inhibition of trypanosomatid GAPDH.

1. INTRODUCTION

Parasitic diseases are a major global cause of illness, morbidity, long-term disability, and death, with severe medical and psychological consequences for millions of men, women, and children.^{1–3} According to the World Health Organization, in both tropical and subtropical regions worldwide, leishmaniasis and trypanosomiasis (sleeping sickness and Chagas' disease) are serious and life-threatening diseases, affecting about 30 million people, with nearly another 400 million living in high-risk areas.^{4–6} In spite of the alarming health, economic, and social consequences of these parasitic infections, the limited existing drug therapy (e.g., antimonials, arsenicals, and nitroheterocyclic drugs) suffers from a combination of drawbacks including poor efficacy, resistance, and serious side effects. Therefore, there is an urgent need for new drugs that can overcome resistance and are safe and effective for use in humans.

The identification of novel molecular targets as well as safe and affordable drug candidates is of utmost importance in drug discovery.^{7–11} The glycolytic pathway of trypanosomatids has been investigated as a promising biochemical target for drug intervention.¹² The bloodstream form of trypanosomatids has no functional tricarboxylic acid cycle and is highly dependent on glycolysis for ATP production.^{13,14} This vital dependence on glycolysis as a source of energy makes the glycolytic enzymes attractive targets for drug design. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH; EC 1.2.1.12), one of the key enzymes in the glycolytic cascade, is a homotetramer that catalyzes the reversible NAD^+ -dependent oxidative phosphorylation of glyceraldehydes-3-phosphate to 1,3-diphosphoglycerate in the presence of inorganic phosphate. The glycolytic GAPDH enzyme from the pathogenic parasites *Trypanosoma cruzi*,¹⁵ *Trypanosoma brucei*,¹⁶ and

*Leishmania mexicana*¹⁷ are closely related (about 90% sequence identity) and possess important structural differences for drug design when compared to the homologue protein of the mammalian host (about 45% sequence identity).^{18,19}

Structure- and ligand-based approaches have become vital components of modern drug design.²⁰ Quantitative structure–activity relationships (QSARs) have been successfully used as valuable tools to assist the design of small-molecule inhibitors or receptor ligands having promise of utility in clinical medicine.^{21–27} In the present work, comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) have been performed to investigate the QSAR of a large series of adenosine analogs, which explore significant structural differences between trypanosomatids and human GAPDHs, leading to a selective inhibition of the parasite enzymes. Molecular modeling studies were also employed to study the molecular basis underlying the selective inhibition of trypanosomatid GAPDH.

2. RESULTS AND DISCUSSION

2.1. Chemical and Biological Data. The data set employed for the 3D QSAR CoMFA and CoMSIA analyses contains 70 adenosine analogs which were collected from five publications reported by one laboratory.^{28–32} The *in vitro* IC_{50} values employed in this work were measured under the same experimental conditions, a fundamental requirement for QSAR studies.^{24–26} Importantly, the reported statistical error limits have been calculated and add up to 10%. The family of GAPDH enzyme inhibitors under investigation has moderate structural diversity. The adenosine analogs bear bulk substituents linked to the N^6 and 9 positions of the adenine system, exploring significant differences observed in the NAD^+ -binding-site vicinity of the parasite enzymes.

* Corresponding author. Tel.: +55 16 3373-8095. Fax: +55 16 3373-9881. E-mail: aandrico@ifsc.usp.br.

The structures and corresponding pIC_{50} values ($-\log \text{IC}_{50}$, where the IC_{50} is the concentration required for 50% inhibition of GAPDH) for the whole set of inhibitors are included in Table 1. The pIC_{50} values span about 3.5 orders of magnitude and are acceptably distributed across the range of values (Figure 1). The generation of consistent statistical models depends on the composition of the training and test sets. From the original data set, 56 inhibitors (compounds 1–56, Table 1) were selected as members of the training set for model generation, whereas the other 14 inhibitors (compounds 57–70, Table 1) were selected as members of the test set for external model validation. The compounds of both training and test sets were carefully selected in order to ensure appropriate property coverage on the entire range of pIC_{50} values. A statistical cluster analysis confirmed that the composition of both training and test sets is representative of the whole data set. The partial least squares (PLS) and leave-one-out (LOO) cross-validation methods were used for all 3D QSAR analyses.

2.2. Molecular Alignment Strategies. Due to the high dependency of trypanosomatids on glycolysis for survival, the energy production could be blocked through the action of a potent competitive inhibitor of an enzyme of the glycolytic pathway.³³ Glycolytic enzymes are potential targets for antitrypanosomatid drug design provided that reasonable parasite versus host selectivity can be obtained.^{28,33} It is important to note that the trypanosomatid GAPDHs possess low sequence identity in comparison with their mammalian host, as shown in Figure 2. The sequence alignments reveal that there are several highly conserved residues among the parasite enzymes that substantially differ from those of the human homologue. It is likely that SAR, QSAR, and structure-based drug design techniques would be useful for the design of GAPDH inhibitors that selectively target the trypanosomatid enzymes.²⁸

The determination of spatial molecular alignments is a crucial step in 3D QSAR studies, since the analyses are highly dependent on the quality of the alignments.^{22,23} Thus, molecular docking studies were carried out using the programs FlexX³⁴ and GOLD (Genetic Optimisation for Ligand Docking)^{35,36} in order to generate specialized molecular alignments for selective QSAR model building. The crystallographic coordinates of the inhibitor N^6 -(1-naphthalene methyl)-2'-deoxy-2'-(3,5-dimethoxy benzamido) (NMDBA, compound 36, Table 1) bound to the *L. mexicana* GAPDH binding site were employed for validating the docking studies.³⁷ Both FlexX and GOLD successfully docked the NMDBA inhibitor in good agreement with the experimentally determined X-ray crystal structure, placing the adenine moiety, as well as the N^6 and 9 substituents close to the bioactive conformation (Figure 3A and B). The heavy atom root-mean-square deviation for the best docking solution in comparison with the crystallographic conformation was 1.43 and 1.39 Å, according to FlexX and GOLD, respectively. On the basis of that, the molecular alignments were performed by applying the implemented scoring functions followed by the selection of the top-ranked docking solutions generated for each ligand of the data set. The final FlexX and GOLD alignments, respectively, are depicted in Figure 3C and D. As can be seen, the top-ranked conformations are reasonably well predicted within the *L. mexicana* GAPDH active site, confirming the robustness of the applied docking procedures.^{38–40}

2.3. 3D QSAR Analysis. **2.3.1. CoMFA and CoMSIA Models.** In order to investigate the structural and chemical features related to the biological activity of this series of trypanosomatid GAPDH inhibitors, the molecular alignments derived from both FlexX and GOLD docking simulations were quantitatively analyzed using the CoMFA^{41,42} and CoMSIA^{43,44} methods. These 3D QSAR techniques are based on the assumption that changes in binding affinities of ligands are related to changes in molecular properties represented by molecular fields. In the present study, the two molecular alignments were submitted to CoMFA and CoMSIA analyses, generating four QSAR models referred to as CoMFA_{FlexX}, CoMFA_{GOLD}, CoMSIA_{FlexX}, and CoMSIA_{GOLD}. The statistical results of the QSAR analyses are presented in Table 2.

Both 3D structural alignments yielded significant statistical CoMFA models, as indicated by the conventional non-cross-validated correlation (r^2) and cross-validated correlation (q^2_{LOO}) coefficients (CoMFA_{FlexX}, $r^2 = 0.93$ and $q^2_{\text{LOO}} = 0.83$; CoMFA_{GOLD}, $r^2 = 0.93$ and $q^2_{\text{LOO}} = 0.82$). The corresponding CoMSIA models derived for the same training set are also of good statistical quality (CoMSIA_{FlexX}, $r^2 = 0.92$ and $q^2_{\text{LOO}} = 0.72$; CoMSIA_{GOLD}, $r^2 = 0.93$ and $q^2_{\text{LOO}} = 0.79$). Although the CoMFA models had slightly better statistical parameters in comparison with the corresponding CoMSIA models, in accordance with similar results reported in the literature for other data sets,^{43–47} the results shown in Table 2 indicate that the models were not substantially affected by the different alignment protocols (FlexX and GOLD). This may be due to the fact that the training set molecules incorporate sufficient structural information to allow privileged molecular recognition within the binding site of *L. mexicana* GAPDH, resulting in similar field descriptors that explain the variance of the 3D QSAR models.

2.3.2. QSAR Model Validation. Leave-many-out cross-validation (q^2_{LMO}) employing 10 or 5 groups (LMO₁₀ or LMO₅, respectively) was performed as a more rigorous test for the stability and statistical significance of the models. Accordingly, the data set was divided into 10 and 5 randomly selected groups of equal size, and subsequently, each group was left out once during the cross-validation process. Each model was evaluated 25 times by measuring its accuracy in predicting the activity of the remaining 10% and 20% data set compounds (Table 3). The results confirm the stability and reliability of the 3D QSAR models, considering that the statistical values obtained for the LMO analyses are comparable to those of the LOO analysis.

Progressive scrambling of the data set was applied to determine the sensitivity of the QSAR models to chance correlations, and to test the stability of the models.²³ The results of the progressive scrambling further confirmed consistency of the models as defined by the critical slope, and optimum statistics for cross-validated standard deviation of error prediction and Q^2 obtained at the end of several different runs. Finally, to validate the models' accuracy, the bootstrapping function was used to determine the error on the r^2 (r^2_{boot}) and the standard error of estimate (SEE_{boot}) of each model. An average r^2_{boot} value of 0.95 (SEE_{boot} of about 0.25) was obtained for both CoMFA and CoMSIA models. These results indicate that the QSAR models generated in this study are very stable and possess substantial statistical significance and high internal predictive ability.

Training Set Compounds

Cpd	R ¹	R ²	R ³	pIC ₅₀
10	CH ₃	H	H	2.22
11	H	H		2.40
12	H		H	2.43
13	H		H	3.15
14	H		H	3.15
15	H		H	2.74
16	H		H	3.44
17	H		H	3.47
18	H		H	3.62

Training Set Compounds

Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
19	H	H	H	H	H	H	2.48
20	H	H	H	H	C ₆ H ₅	H	3.70
21	H	H	H	OCH ₂ CH ₃	H	H	3.40
22	H	H	H	OCH ₃	OH	H	3.30
23	H	Br	H	H	H	H	2.52
24	H	H	H	OH	OH	OH	3.60
25	H	H	H	OH	H	OH	3.19
26		H	H	OCH ₃	H	OCH ₃	4.70
27		H	H	OCH ₃	H	H	4.80
28		H	H	OCH ₃	H	OCH ₃	4.60
29		H	H	OCH ₃	H	H	4.60
30		H	H	OCH ₃	H	OCH ₃	4.60

Table 1. Continued

Training Set Compounds							
Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
31		H	H	OCH ₃	H	H	4.60
32		H	H	OCH ₃	H	H	5.26
33		H	H	OCH ₃	H	H	3.82
34		H	H	OCH ₃	H	H	4.10
35		H	H	Cl	H	H	5.00
36		H	H	OCH ₃	H	OCH ₃	5.70
37		H	H	Cl	H	Cl	4.92
38		H	H	OAc	H	OAc	5.00
39		H	H	OH	H	OH	5.30
40		H	H	H	OAc	H	4.00
41		H	H	H	OH	H	4.10
42		H	H	OAc	OAc	H	4.60
43		H	H	OH	OCH ₃	H	4.43
44		H	H	OCH ₃	OH	H	5.00
45		H	H	N(CH ₃) ₂	H	H	4.60
46		H	H	OCH ₃	H	OCH ₃	5.70
47		H	H	Cl	H	Cl	4.60
48		H	H	OCH ₃	H	OCH ₃	4.22
49		H	H	OCH ₃	H	H	5.22
50		H	H	Cl	H	H	5.40

Training Set Compounds							
Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
51		H	H	OCH ₃	H	OCH ₃	5.70
52		H	H	Cl	H	Cl	4.60
53		H	H	OCH ₃	H	H	5.70
54		H	H	Cl	H	Cl	4.60
55		H	H	OCH ₃	H	H	5.30
56		H	H	OCH ₃	H	H	4.08

Test Set Compounds								
Cpd	Structure	pIC ₅₀	Cpd	Structure	IC ₅₀	Cpd	Structure	pIC ₅₀
57		2.62	58		3.52	59		3.30
60		3.22	61		3.07	62		3.82
63		5.30	64		5.40	65		5.70
66		4.30	67		4.60	68		4.00
69		4.74	70		5.00

Figure 1. Distribution of pIC_{50} values for the training set, test set, and complete data set.

Figure 2. Sequence alignments of *L. mexicana*, *T. cruzi*, *T. brucei*, and *H. sapiens* GAPDHs. Numbers above the sequences refer to the *L. mexicana* GAPDH sequence. Residues conserved among the different species are highlighted in gray.

Although the q^2_{LOO} and q^2_{LMO} procedures may give a suitable representation of the predictive power of the models for untested GAPDH inhibitors, the external validation process can be considered the most valuable validation method.^{21–23} Therefore, the predictive ability of the models was assessed by predicting the potency of an external test set of 14 compounds (57–70, Table 1), which were completely excluded from model generation. On the basis of the appropriate representation of chemical diversity and distribution of the biological property (Figure 1), the test set meets the requirements for the purpose of external model validation.^{21–27} The results of the external validation are listed in Table 4, and the graphic results are simultaneously displayed in Figure 4.

Predictive r^2 values of 0.86, 0.83, 0.80, and 0.78 were obtained for CoMFA_{FlexX}, CoMSIA_{FlexX}, CoMFA_{GOLD}, and CoMSIA_{GOLD}, respectively (Table 3). The good agreement

between experimental and predicted pIC_{50} values for the test set compounds indicates the reliability of the QSAR models, as expressed by the high predictive r^2 values. This is probably the best statistical parameter to assess the external predictive power of QSAR models.^{48–51} Hence, in addition to the good statistical quality and internal predictive ability, the models show excellent predictive power for novel GAPDH inhibitors within this structural diversity.

2.3.3. 3D Contour Maps. The understanding of protein–ligand interactions is essential for the design of new inhibitors with improved potency.²⁰ Besides predicting the property value of interest for untested molecules, 3D QSAR models should also provide insights into molecular properties directly related to biological activity. One prominent feature of the 3D QSAR methods employed in this study is that they allow easy visualization of regions in space responsible for increases or decreases in the values of a particular type of

Figure 3. X-ray crystal structure 1I32 of the NMDBA inhibitor (green) bound to *L. mexicana* GAPDH and the best docking solution of NMBDA (gray) generated by FlexX (A) and GOLD (B). FlexX (C) and GOLD (D) data set molecular alignments are shown inside the Connolly surface of the adenine portion of the NAD⁺ binding pocket.

Table 2. CoMFA and CoMSIA Statistical Results for the FlexX and GOLD 3D Alignments^a

statistical parameters	FlexX		GOLD	
	CoMFA	CoMSIA	CoMFA	CoMSIA
q^2_{LOO}	0.83	0.72	0.82	0.79
r^2	0.93	0.92	0.93	0.93
SEE	0.27	0.30	0.29	0.29
F	259	153	218	167
N	3	4	3	4
fraction (%)				
S	44	18	48	15
E	56	26	52	27
H		26		26
H donor		18		21
H acceptor		12		11

^a q^2 , leave-one-out (LOO) cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; N , optimum number of components; SEE, standard error of estimate; F , F test value; **S**, steric field; **E**, electrostatic field; **H**, hydrophobic field.

dependent variable (e.g., pIC₅₀). Considering that both alignment protocols generated models of similar statistical quality, the discussion about the CoMFA and CoMSIA contour maps will be restricted to the models derived with the FlexX alignment, because they show slightly better statistical parameters and predictive power. The CoMFA (steric and electrostatic) and CoMSIA (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) con-

Table 3. Statistical Data for Validation Procedure of CoMFA and CoMSIA Models^a

statistical parameters	FlexX		GOLD	
	CoMFA	CoMSIA	CoMFA	CoMSIA
q^2 LMO ₁₀	0.82 ± 0.01	0.72 ± 0.02	0.82 ± 0.01	0.79 ± 0.01
q^2 LMO ₅	0.82 ± 0.01	0.70 ± 0.03	0.82 ± 0.01	0.78 ± 0.02
q^2_{boot}	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01
SEE _{boot}	0.24 ± 0.11	0.25 ± 0.11	0.23 ± 0.11	0.23 ± 0.11
predictive r^2	0.86	0.83	0.80	0.78

^a LMO, mean and standard deviation values of q^2 obtained after 25 independent runs of cross-validation protocols. Bootstrapping, mean, and standard deviation values of r^2 and SEE obtained after 100 simulation samplings.

tour maps are displayed as PLS standard deviation × coefficient maps in Figures 5 and 6, respectively.

2.3.3.1. CoMFA Contour Maps. The CoMFA steric contour map is shown in Figure 5A. The green contour surrounding the naphthyl substituent of the inhibitor **36** (pIC₅₀ = 5.70) indicates that bulky substituents would be favorable in this region. Conversely, unfavorable steric regions located near the 8 position of the adenine moiety suggest that bulky substituents at this position are detrimental for potency. In fact, in the proposed binding mode of compound **23** (pIC₅₀ = 2.52), the bromine substituent is oriented toward this unfavorable region. Thus, the absence of bulky groups in the sterically favorable region combined with the presence of a steric unfavorable substituent in the

Table 4. Experimental and Predicted Activities (pIC_{50}) with Residual Values for the Test Set Compounds

cpd	exp ^a	FlexX				GOLD			
		CoMFA		CoMSIA		CoMFA		CoMSIA	
		pred ^b	res ^c	pred ^b	res ^c	pred ^b	res ^c	pred ^b	res ^c
57	2.62	3.18	−0.56	3.31	−0.69	3.11	−0.49	2.89	−0.27
58	3.52	3.45	0.08	3.65	−0.13	3.42	0.10	3.26	0.26
59	3.30	3.30	0.00	3.17	0.13	2.68	0.62	2.80	0.50
60	3.22	2.95	0.27	3.13	0.09	3.38	−0.16	3.17	0.05
61	3.07	3.17	−0.10	3.29	−0.22	3.17	−0.10	3.11	−0.04
62	3.82	3.60	0.22	3.27	0.55	3.08	0.74	3.28	0.54
63	5.30	4.94	0.36	5.25	0.05	4.86	0.44	5.20	0.10
64	5.40	4.93	0.47	4.93	0.47	4.88	0.52	4.63	0.77
65	5.70	5.66	0.04	5.42	0.28	5.53	0.17	5.30	0.40
66	4.30	4.62	−0.32	4.75	−0.45	4.96	−0.66	5.05	−0.75
67	4.60	5.07	−0.47	4.91	−0.31	4.66	−0.06	5.04	−0.44
68	4.00	4.56	−0.56	4.45	−0.45	4.36	−0.36	4.40	−0.40
69	4.74	4.46	0.28	4.27	0.47	4.84	−0.10	5.12	−0.38
70	5.00	4.57	0.43	4.44	0.56	4.55	0.46	4.46	0.54

^a Experimental value. ^b Predicted value. ^c Residual: the difference between experimental and predicted values.

Figure 4. Plot of predicted values of pIC_{50} versus the corresponding experimental values for the 56 training set inhibitors (open square) and 14 test set inhibitors (solid triangles).

Figure 5. 3D contour maps (standard deviation $\times$ coefficient) for CoMFA_{FlexX}. The inhibitors **36** (in light blue) and **23** (in orange) are represented by ball-and-stick models in the active site of *L. mexicana* GAPDH. The protein Connolly surface is indicated. The hydrogen atoms were omitted for clarity. (A) CoMFA steric contour maps are shown in yellow and green (-0.005 and 0.035 kcal/mol, respectively); (B) CoMFA electrostatic contour maps are shown in red and blue (-0.05 and 0.05 kcal/mol, respectively).

Figure 6. 3D contour maps (standard deviation $\times$ coefficient) for CoMSIA_{FlexX}. The GAPDH inhibitors model within the *Lm*-GAPDH binding site. The inhibitors **36** (in light blue) and **23** (in orange) are represented by ball-and-stick models in the active site of *L. mexicana* GAPDH. The protein-solvent-accessible surface is indicated as a solid surface. The hydrogen atoms were omitted for clarity. (A) CoMSIA steric contour maps are shown in yellow and green (-0.0014 and 0.008 kcal/mol, respectively); (B) CoMSIA electrostatic contour maps are shown in red and blue (-0.03 and 0.02 kcal/mol, respectively). (C) CoMSIA hydrophobic contour maps are shown in white and yellow (-0.01 and 0.015 kcal/mol, respectively); (D) CoMSIA hydrogen-bond donor and acceptor contour maps. H-donor contour fields are shown in purple and cyan (-0.015 and 0.020 kcal/mol, respectively). H-acceptor contour fields are shown in red and magenta (-0.07 and 0.015 kcal/mol, respectively).

adenine ring could explain the poor inhibitory activity exhibited by compound **23** (Figure 5A). The CoMFA electrostatic contour plots are shown in Figure 5B. The map highlights three favorable regions for electropositive groups (blue contour). Two of them are solvent-exposed, one being near the 2 position of the naphthyl ring and the other near the 6'-hydroxyl of the ribose ring. The third favorable region can be found surrounding the sulfur in the side chain of Met39, indicating that electropositive groups at this position may enhance the inhibitory potency. Furthermore, there are two favorable regions for negatively charged groups: one,

solvent exposed, near the N⁶-position of the adenine moiety, and the other, located near to the side-chain hydroxyl group of Ser40. The 3-methoxy benzamide substituent of **36** is close to this electronegatively favorable region. The contour map also agrees with previous SAR studies, which indicated that the absence of electronegative groups in the benzamide ring (compounds **19** and **23**) is remarkably unfavorable for inhibitory potency within this series of trypanosomatid GAPDH inhibitors.^{28,29}

2.3.3.2. CoMSIA Contour Maps. The chemical environment of the enzyme binding site is represented by five

Figure 7. 3D contour maps (standard deviation $\times$ coefficient) for the CoMFA_{FlexX} model showed within the active site of the superposed crystallographic structures of trypanosomatid GAPDHs. The GAPDH structures are shown as *L. mexicana*, in cyan (PDB: 1I32); *T. brucei* in yellow (PDB: 1GGA); *T. cruzi* in gray (PDB: 1ML3); and human, in magenta (PDB: 1U8F). The inhibitor **36** (cyan) and the cofactor NAD⁺ are illustrated as ball-and-stick models. The hydrogen atoms were omitted for clarity. The CoMFA steric favorable map is shown in green, whereas the favorable CoMFA electronegative contour map is shown in red (+0.035 and -0.05 kcal/mol, respectively).

molecular fields in the CoMSIA method (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor). Figure 6A shows the CoMSIA steric contour plots, which agree with the contour plots of the corresponding CoMFA map, highlighting two important steric regions (green isopleths). The first one is positioned between the side chains of Met39 and Arg92, while the second one lies in the lower portion of the binding pocket, at the so-called “selective cleft”, between the chains of Met39 and Val206. Ligands possessing bulky groups in these regions have enhanced inhibitory potency (compounds **26–55**, **63–70**).^{30–32} The CoMSIA unfavorable steric map also highlights the 8-bromine substituent from compound **23** as detrimental for activity.

The CoMSIA electrostatic contour map of Figure 6B shows that inhibitors having either negatively charged groups in the red regions or positively charged groups in the blue regions would substantially enhance inhibitory potency toward the parasite enzymes. As previously noted for the CoMFA electrostatic contour map of Figure 5A, compound **36** orients the electronegative oxygen of the 3-methoxy benzamide group in a favorable region for negative charges. In addition, the methyl of the 3-methoxy group is located in the favorable positively charged lower portion of the binding site.

The yellow and white isopleths in Figure 6C indicate that hydrophobic and hydrophilic regions correlate with inhibitory potency. There are two hydrophobically favorable regions (yellow contours); one encompasses the N⁶ substituent of the adenine system, while the other surrounds the 5-methoxyl of the benzamide moiety. The former region partially overlaps with the steric volume favoring bulk groups, thus, suggesting that ligands with bulk and hydrophobic substituents at this site would be beneficial for enhanced potency. This corroborates with the chemical environment of the binding site and agrees with experimental results.^{28–32} The other region favoring hydrophobic contacts is located near the side chain of Met39. In addition, the contour map displays two unfavorable hydrophobic regions. One region can be

observed around the bromine substituent of compound **23**, while the other partly overlaps with the favored partially negatively charged region, suggesting that electron-withdrawing groups with hydrophilic properties would be beneficial at this region.

The CoMSIA hydrogen-bond contour map is illustrated in Figure 6D, being discriminated for donor (cyan and purple) or acceptor (magenta and red) regions. The presence of cyan regions near the carboxyl group in the side chain of Asp38 denotes the regions where hydrogen-bond acceptors might exist in the receptor, and thus a hydrogen-bond-donating substituent could be favored. In fact, the carboxyl group of the Asp38 is an important interaction point for molecular recognition toward the trypanosomatid GAPDHs.³⁷ A magenta isopleth, favorable for hydrogen-acceptor groups, is in good agreement with the binding site environment, suggesting the main-chain carbonyl group of Gln91 as a reasonable site for hydrogen-bond formation.

2.3.4. 3D QSAR Selectivity Map. The structure-based GAPDH inhibitor design can take advantage of the substantial structural differences between trypanosomatid and human enzymes. Selective inhibitors of *L. mexicana*, *T. brucei*, and *T. cruzi* GAPDH enzymes were obtained on the basis of adenosine analogs bearing functional groups that bind in the intersubunit (selectivity cleft) of the NAD⁺-binding site.^{28–32,37} Due to the fact that the human enzyme possesses a more narrow “selectivity cleft”, these inhibitors selectively bind to the trypanosomatid GAPDHs. In order to obtain further insights into the molecular basis underlying the selective inhibition of trypanosomatid GAPDH enzymes, a detailed analysis of the most relevant 3D QSAR fields was carried out through the superposition of the binding sites of *T. cruzi*,¹⁵ *T. brucei*,¹⁶ *L. mexicana*,¹⁷ and human¹⁸ GAPDHs (Figure 7). For this study, we have employed the corresponding steric and electrostatic fields of the CoMFA_{FlexX} model (the best QSAR model generated), which have significant complementarity to the GAPDH binding site.

As can be seen, the CoMFA electrostatic field descriptor explained 56% of the variance, while the proportion of the steric descriptor accounted for the remaining 44% (Table 2). According to the aforementioned CoMFA and CoMSIA results, the biological potency of the inhibitors is directly affected by the electrostatic fields. The electronegatively favorable region (red contour) located within the superposed binding sites is in accordance with the chemical environment of the GAPDH parasite enzymes. The map is near the NH_2 group in the side chain of the Asn residues (Asn40 in *T. cruzi* and Asn39 in *T. brucei*), and also near the OH group in the side chain of the Ser40 residue in *L. mexicana* GAPDH. Conversely, this electronegatively favorable region near the carbonyl group in the main chain of Phe37 of the human homologue suggests that para-substituted benzamides with partially negative charged groups may be exploited to enhance both potency and selectivity toward the trypanosomatid enzymes. The CoMFA steric field is also in good agreement with the binding site of the parasite enzymes, indicating that bulky substituents are able to fit between the side chains of the Arg92 and Met39 (Met38 in *T. brucei*) residues. In contrast, this region is sterically hindered in the human enzyme by a bulkier side chain of the Phe37 residue.

3. CONCLUSION

In this study, CoMFA and CoMSIA QSAR models were developed on the basis of the docked conformations of 70 adenosine analogs as trypanosomatid GAPDH inhibitors. The 3D QSAR models described possess high internal and external consistency. The good correlation between experimental and predicted pIC_{50} values for the test set compounds further indicated the robustness of the QSAR models. The CoMFA and CoMSIA contour maps, representing the effects of the steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor molecular fields, are in good agreement with the chemical environment of the GAPDH binding site and, therefore, could be useful to guide further structural modification as well as the structure-based design of new selective trypanosomatid-selective inhibitors of GAPDH. The integration of molecular docking and 3D QSAR is a useful tool for the generation of predictive models and for the identification of key structural features responsible for selectivity and biological potency.

4. MATERIALS AND METHODS

4.1. Data Set. The data set used for the QSAR studies contains 70 inhibitors of GAPDH, which were selected from the literature.^{28–32} The structures and corresponding pIC_{50} values for the complete set of inhibitors are included in Table 1. The complete data set was divided into training (compounds 1–56, Table 1) and test (compounds 57–70, Table 1) sets in the ratio of 5:1 (20%), and the pIC_{50} values were used as dependent variables in the QSAR analyses.

4.2. Computational Approach. The QSAR modeling analyses, calculations, and visualizations for CoMFA and CoMSIA were performed using the SYBYL 7.3 package (Tripos Inc., St. Louis, MO) running on Red Hat Enterprise Linux workstations. The 3D structures of the inhibitors were constructed using standard geometric parameters of the molecular modeling software package SYBYL 7.3. Each single-optimized conformation of each molecule in the data set was

energetically minimized employing the Tripos force field⁵² and the Powell conjugate gradient algorithm⁵³ with a convergence criterion of 0.05 kcal/mol·Å. Partial atomic charges were calculated by the Gasteiger–Hückel method in the SYBYL 7.3 program.⁵⁴ Molecular modeling studies were carried out with FlexX (Tripos Inc., St. Louis, MO) and GOLD 2.1 (Cambridge Crystallographic Data Centre, Cambridge, U.K.). With the default parameters of FlexX and GOLD, the docking procedures were repeated 30 times for each data set compound. The implemented FlexX and GOLDScore scoring functions were employed to select the representative conformation for each compound. The FlexX and GOLD top-ranked poses were then used to produce the statistically significant alignments. A statistical cluster analysis was carried out with Tsar 3D, version 3.3 (Accelrys, San Diego, CA), using the complete linkage clustering method (Euclidean distances) with no data standardization.

4.3. Molecular Modeling and Structural Alignment.

The data set molecular alignment protocols applied were based on the top-ranked conformation selected by the FlexX and GOLD 2.1 docking programs. The crystallographic data of *L. mexicana* GAPDH in complex with NMDBA (compound 51, Table 1) were used as a molecular target for all docking simulations (PDB code: 1I32).³⁷ Hydrogen atoms were added in standard geometry using the Biopolymer module as implemented in SYBYL 7.3. The histidine, glutamine, and asparagine residues in the active binding site were manually checked for possible flipped orientation, protonation, and tautomeric states with a Pymol 0.99 (DeLano Scientific, San Carlos, CA) side-chain wizard script. The active site was defined incorporating all amino acid residues within a radius sphere of 8.0 Å centered on the bound ligand. In both docking approaches, default parameters were employed, and only the top-ranked conformation of each inhibitor was considered for 3D QSAR studies.

4.4. 3D QSAR: CoMFA and CoMSIA Analyses. The CoMFA and CoMSIA analyses were carried out on the basis of the interaction energies between a suitable probe atom and the aligned ligand atoms calculated within a grid box of 2 Å spacing. The grid box embedded all ligands with a margin of at least 4 Å in each direction. A positively charged sp^3 carbon probe atom was employed to calculate steric and electrostatic CoMFA fields using SYBYL standard parameters (TRIPOS standard field, dielectric constant 1/ r , cutoff 30 kcal/mol). CoMSIA fields were computed for steric, electrostatic, hydrophobic, and hydrogen bonding properties, using a probe of charge +1, a radius of 1, hydrophobicity and hydrogen bonding properties of +1, and an attenuation factor of 0.3 for the Gaussian distance-dependent function. According to each alignment, the pIC_{50} values were correlated with the field descriptors using SAMPLS⁵⁵ in a LOO cross-validation analysis. After determining the optimal number of components, a PLS analysis was performed without cross-validation, applying no column filtering. A progressive scrambling method⁵⁶ was employed to determine the sensitivity of the QSAR models to chance correlations. This validation technique addresses the overly optimistic cross-validation or response randomization results for redundant data sets. In this approach, small random perturbations are introduced into a data set. This causes the nominal predictivity of unstable models to fall off rapidly, whereas robust models are relatively stable. LMO with 10 (LMO₁₀)

and 5 (LMO₅) randomly selected groups were used as a more rigorous test to assess model stability and statistical significance. Each random cross-validation run was repeated 25 times to obtain mean values for q^2 and the corresponding SDEP. The bootstrapping procedure was used to validate each model. In this procedure, n random selections of the original set of n objects were performed several times (100 times was used to produce consistent statistical information) to simulate different sampling from a larger set of objects. In each individual run, some objects may not be included in the PLS analysis (same method to determine the q^2), whereas other objects can be included more than once. Confidence intervals for each term were estimated through this procedure, providing an independent measure for the stability of the PLS model.^{57,58} The CoMFA and CoMSIA contour maps were generated by interpolation of the pairwise products between the PLS coefficients and standard deviations of the corresponding descriptors values. External model validation was performed with a test set of 14 compounds, which were not considered for QSAR model generation. After generation of the PLS training set models, the dependent variables (pIC₅₀) were predicted for the test set compounds, allowing predictive- r^2 values to be determined for the individual 3D QSAR models.

ACKNOWLEDGMENT

We gratefully acknowledge financial support from the State of São Paulo Research Foundation (FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo) and the National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Pesquisa e Desenvolvimento), Brazil. We also thank Prof. Gerhard Klebe and his research group (AGKlebe) from the University of Marburg—Germany, for important scientific support and very stimulating discussions.

REFERENCES AND NOTES

- (1) Watkins, B. M. Drugs for the Control of Parasitic Diseases: Current Status and Development. *Trends Parasitol.* **2003**, *19*, 477–478.
- (2) Nicholson, B. D.; Walley, J. D.; Baguley, D. S. Leishmaniasis, Chagas disease and Human African Trypanosomiasis revisited: Disease Control Priorities in Developing Countries. *Trop. Med. Int. Health* **2006**, *11*, 1339–1340.
- (3) McKerrow, J. H.; Caffrey, C.; Kelly, B.; Loke, P.; Sajid, M. Proteases in Parasitic Diseases. *Annu. Rev. Pathol. Mech. Dis.* **2006**, *1*, 497–536.
- (4) Croft, S. L.; Coombs, G. H. Leishmaniasis - Current Chemotherapy and Recent Advances in the Search for Novel Drugs. *Trends Parasitol.* **2003**, *19*, 502–508.
- (5) *Control of Chagas Disease*; World Health Organization, Tech. Rep. Ser.: Geneva, Switzerland, 2002; Vol. 905, pp 1–109.
- (6) Sanz-Rodriguez, C. E.; Concepcion, J. L.; Pekerar, S.; Oldfield, E.; Urbina, J. A. Bisphosphonates as Inhibitors of *Trypanosoma cruzi* Hexokinase: Kinetic and Metabolic Studies. *J. Biol. Chem.* **2007**, *282*, 12377–12387.
- (7) Andricopulo, A. D.; Montanari, C. A. Structure-Activity Relationships for the Design of Small-Molecule Inhibitors. *Mini-Rev. Med. Chem.* **2005**, *5*, 585–593.
- (8) Marton, M. J.; DeRisi, J. L.; Bennett, H. A.; Iyer, V. R.; Meyer, M. R.; Roberts, C. J.; Stoughton, R.; Burchard, J.; Slade, D.; Dai, H.; Bassett, D. E., Jr.; Hartwell, L. H.; Brown, P. O.; Friend, S. H. Drug Target Validation and Identification of Secondary Drug Target Effects Using DNA Microarrays. *Nat. Med.* **1998**, *4*, 1293–1301.
- (9) Brinkman, R. R.; Dube, M. P.; Rouleau, G. A.; Orr, A. C.; Samuels, M. E. Human Monogenic Disorders - a Source of Novel Drug Targets. *Nat. Rev. Genet.* **2006**, *7*, 249–260.
- (10) Wadman, M. Drive for Drugs Leads to Baby Clinical Trials. *Nature* **2006**, *440*, 406–407.
- (11) Reichert, J. M. Trends in Development and Approval Times for New Therapeutics in the United States. *Nat. Rev. Drug Discovery* **2003**, *2*, 695–702.
- (12) Verlinde, C. L. M. J.; Hannaert, V.; Blonski, C.; Willson, M.; Perie, J. J.; Fothergill-Gilmore, L. A.; Opperdoes, F. R.; Gelb, M. H.; Hol, W. G.; Michels, P. A. M. Glycolysis as a Target for the Design of New Anti-Trypanosome Drugs. *Drug Resist. Updates* **2001**, *4*, 50–65.
- (13) Bakker, B. M.; Mensonides, F. I.; Teusink, B.; van Hoek, P.; Michels, P. A. M.; Westerhoff, H. V. Compartmentation Protects Trypanosomes from the Dangerous Design of Glycolysis. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2087–2092.
- (14) Michels, P. A. M.; Hannaert, V.; Bringaud, F. Metabolic Aspects of Glycosomes in Trypanosomatidae - New Data and Views. *Parasitol. Today* **2000**, *16*, 482–489.
- (15) Souza, D. H. F.; Garratt, R. C.; Araujo, A. P.; Guimaraes, B. G.; Jesus, W. D.; Michels, P. A. M.; Hannaert, V.; Oliva, G. *Trypanosoma cruzi* Glycosomal Glyceraldehyde-3-Phosphate Dehydrogenase: Structure, Catalytic Mechanism and Targeted Inhibitor Design. *FEBS Lett.* **1998**, *424*, 131–135.
- (16) Vellieux, F. M.; Hajdu, J.; Verlinde, C. L. M. J.; Groendijk, H.; Read, R. J.; Greenhough, T. J.; Campbell, J. W.; Kalk, K. H.; Littlechild, J. A.; Watson, H. C.; Hol, W. G. J. Structure of Glycosomal Glyceraldehyde-3-phosphate Dehydrogenase from *Trypanosoma brucei* Determined from Laue Data. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2355–2359.
- (17) Kim, H.; Feil, I. K.; Verlinde, C. L. M. J.; Petra, P. H.; Hol, W. G. Crystal Structure of Glycosomal Glyceraldehyde-3-Phosphate Dehydrogenase from *Leishmania mexicana*: Implications for Structure-Based Drug Design and a New Position for the Inorganic Phosphate Binding Site. *Biochemistry* **1995**, *34*, 14975–14986.
- (18) Jenkins, J. L.; Tanner, J. J. High-Resolution Structure of Human D-Glyceraldehyde-3-Phosphate Dehydrogenase. *Acta Crystallogr., Sect. D* **2006**, *62*, 290–301.
- (19) Ismail, S. A.; Park, H. W. Structural Analysis of Human Liver Glyceraldehyde-3-Phosphate Dehydrogenase. *Acta Crystallogr., Sect. D* **2005**, *61*, 1508–1513.
- (20) Guido, R. V. C.; Oliva, G.; Andricopulo, A. D. Virtual Screening and Its Integration with Modern Drug Design Technologies. *Curr. Med. Chem.* **2008**, *15*, 37–46.
- (21) Castilho, M. S.; Guido, R. V. C.; Andricopulo, A. D. 2D Quantitative Structure-Activity Relationship Studies on a Series of Cholesteryl Ester Transfer Protein Inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 6242–6252.
- (22) Salum, L. B.; Polikarpov, I.; Andricopulo, A. D. Structural and Chemical Basis for Enhanced Affinity and Potency for a Large Series of Estrogen Receptor Ligands: 2D and 3D QSAR Studies. *J. Mol. Graphics Modell.* **2007**, *26*, 434–442.
- (23) Honorio, K. M.; Garratt, R. C.; Polikarpov, I.; Andricopulo, A. D. 3D QSAR Comparative Molecular Field Analysis on Nonsteroidal Farnesoid X Receptor Activators. *J. Mol. Graphics Modell.* **2007**, *25*, 921–927.
- (24) Farutin, V.; Masterson, L.; Andricopulo, A. D.; Cheng, J.; Riley, R.; Hakimi, R.; Frazer, J.; Cordes, E. H. Structure-Activity Relationships for a Class of Inhibitors of Purine Nucleoside Phosphorylase. *J. Med. Chem.* **1999**, *42*, 2422–2431.
- (25) Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. Hologram Quantitative Structure-Activity Relationships for a Series of Farnesoid X Receptor Activators. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119–3125.
- (26) Guido, R. V. C.; Trossini, G. H. G.; Castilho, M. S.; Oliva, G.; Ferreira, E. I.; Andricopulo, A. D. Structure-Activity Relationships for a Class of Selective Inhibitors of the Major Cysteine Protease from *Trypanosoma cruzi*. *J. Enzyme Inhib. Med. Chem.* **2008**. DOI: 10.1080/14756360701810322.
- (27) Castilho, M. S.; Postigo, M. P.; de Paula, C. B.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D. Two- and Three-Dimensional Quantitative Structure-Activity Relationships for a Series of Purine Nucleoside Phosphorylase Inhibitors. *Bioorg. Med. Chem.* **2006**, *14*, 516–527.
- (28) Aronov, A. M.; Suresh, S.; Buckner, F. S.; Van Voorhis, W. C.; Verlinde, C. L. M. J.; Opperdoes, F. R.; Hol, W. G.; Gelb, M. H. Structure-Based Design of Submicromolar, Biologically Active Inhibitors of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4273–4278.
- (29) Van Calenbergh, S.; Verlinde, C. L. M. J.; Soenens, J.; De Bruyn, A.; Callens, M.; Blaton, N. M.; Peeters, O. M.; Rozenskik, J.; Hol, W. G.; Herdewijn, P. Synthesis and Structure-Activity Relationships of Analogs of 2'-deoxy-2'-(3-methoxybenzamido)adenosine, a Selective Inhibitor of Trypanosomal Glycosomal Glyceraldehyde-3-Phosphate Dehydrogenase. *J. Med. Chem.* **1995**, *38*, 3838–3849.
- (30) Aronov, A. M.; Verlinde, C. L. M. J.; Hol, W. G.; Gelb, M. H. Selective Tight Binding Inhibitors of Trypanosomal Glyceraldehyde-3-Phosphate Dehydrogenase Via Structure-Based Drug Design. *J. Med. Chem.* **1998**, *41*, 4790–4799.

- (31) Verlinde, C. L. M. J.; Callens, M.; Van Calenbergh, S.; Van Aerschot, A.; Herdewijn, P.; Hannaert, V.; Michels, P. A. M.; Opperdoes, F. R.; Hol, W. G. Selective Inhibition of Trypanosomal Glyceraldehyde-3-Phosphate Dehydrogenase by Protein Structure-Based Design: Toward New Drugs for the Treatment of Sleeping Sickness. *J. Med. Chem.* **1994**, *37*, 3605–3613.
- (32) Bressi, J. C.; Verlinde, C. L. M. J.; Aronov, A. M.; Shaw, M. L.; Shin, S. S.; Nguyen, L. N.; Suresh, S.; Buckner, F. S.; Van Voorhis, W. C.; Kuntz, I. D.; Hol, W. G.; Gelb, M. H. Adenosine Analogues as Selective Inhibitors of Glyceraldehyde-3-Phosphate Dehydrogenase of Trypanosomatidae Via Structure-Based Drug Design. *J. Med. Chem.* **2001**, *44*, 2080–2093.
- (33) Bakker, B. M.; Michels, P. A. M.; Opperdoes, F. R.; Westerhoff, H. V. What Controls Glycolysis in Bloodstream form *Trypanosoma brucei*. *J. Biol. Chem.* **1999**, *274*, 14551–14559.
- (34) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (35) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. D. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (36) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003**, *52*, 609–623.
- (37) Suresh, S.; Bressi, J. C.; Kennedy, K. J.; Verlinde, C. L. M. J.; Gelb, M. H.; Hol, W. G. Conformational Changes in *Leishmania mexicana* Glyceraldehyde-3-Phosphate Dehydrogenase Induced by Designed Inhibitors. *J. Mol. Biol.* **2001**, *309*, 423–435.
- (38) Bissantz, C.; Folkers, G.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767.
- (39) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
- (40) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (41) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5867.
- (42) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Recent Advances in Comparative Molecular Field Analysis (CoMFA). *Prog. Clin. Biol. Res.* **1989**, *291*, 161–165.
- (43) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- (44) Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to Study Hydrogen-Bonding Properties and to Score Combinatorial Libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1–10.
- (45) Doddareddy, M. R.; Jung, H. K.; Cha, J. H.; Cho, Y. S.; Koh, H. Y.; Chang, M. H.; Pae, A. N. 3D QSAR Studies on T-type Calcium Channel Blockers Using CoMFA and CoMSIA. *Bioorg. Med. Chem.* **2004**, *12*, 1613–1621.
- (46) Amin, E. A.; Welsh, W. J. Highly Predictive CoMFA and CoMSIA Models for Two Series of Stromelysin-1 (MMP-3) Inhibitors Elucidate S1' and S1-S2' Binding Modes. *J. Chem. Inf. Model.* **2006**, *46*, 1775–1783.
- (47) Zhu, L. P.; Ye, D. Y.; Tang, Y. Structure-Based 3D-QSAR Studies on Thiazoles as 5-HT₃ Receptor Antagonists. *J. Mol. Model.* **2007**, *13*, 121–131.
- (48) Oprea, T. I.; Waller, C. L. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley: New York, 1997; Vol. 11, Chapter 3, pp. 127–182.
- (49) Golbraikh, A.; Tropsha, A. Beware of q²! *J. Mol. Graphics Model.* **2002**, *20*, 269–276.
- (50) Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational Selection of Training and Test Sets for the Development of Validated QSAR Models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241–253.
- (51) Doweyko, A. M. 3D-QSAR Illusions. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 587–596.
- (52) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (53) Powell, M. J. D. Restart Procedures for the Conjugate Gradient Method. *Math. Program.* **1977**, *12*, 241–254.
- (54) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- (55) Bush, B. L.; Nachbar, R. B., Jr. Sample-Distance Partial Least Squares: PLS Optimized For Many Variables, with Application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587.
- (56) Clark, R. D.; Fox, P. C. Statistical variation in progressive scrambling. *J. Comput.-Aided Mol. Des.* **2004**, *17*, 1–14.
- (57) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Cross-Validation, Bootstrapping, and Partial Least-Squares Compared with Multiple-Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
- (58) Wehrens, R.; van der Linden, W. E. Bootstrapping Principal Component Regression Models. *J. Chemom.* **1997**, *11*, 157–171.

CI700453J

Virtual Screening and Its Integration with Modern Drug Design Technologies

Rafael V.C. Guido, Glaucius Oliva and Adriano D. Andricopulo*

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

Abstract: Drug discovery is a highly complex and costly process, which demands integrated efforts in several relevant aspects involving innovation, knowledge, information, technologies, expertise, R&D investments and management skills. The shift from traditional to genomics- and proteomics-based drug research has fundamentally transformed key R&D strategies in the pharmaceutical industry addressed to the design of new chemical entities as drug candidates against a variety of biological targets. Therefore, drug discovery has moved toward more rational strategies based on our increasing understanding of the fundamental principles of protein-ligand interactions. The combination of available knowledge of several 3D protein structures with hundreds of thousands of small-molecules have attracted the attention of scientists from all over the world for the application of structure- and ligand-based drug design approaches. In this context, virtual screening technologies have largely enhanced the impact of computational methods applied to chemistry and biology and the goal of applying such methods is to reduce large compound databases and to select a limited number of promising candidates for drug design. This review provides a perspective of the utility of virtual screening in drug design and its integration with other important drug discovery technologies such as high-throughput screening (HTS) and QSAR, highlighting the present challenges, limitations, and future perspectives in medicinal chemistry.

Keywords: Drug design, virtual screening, QSAR, HTS, binding affinity.

APÊNDICE C

INTRODUCTION

The identification of promising hits and the generation of high-quality leads are crucial steps in the early stages of any drug discovery project. Recent advances in medicinal chemistry at the interface of chemistry and biology have created an important foundation in the search for new drug candidates possessing a combination of optimized pharmacodynamic and pharmacokinetic properties. Despite the impact of the recent technological and scientific advances, drug discovery has become more expensive and time consuming over the same period of time [1,2].

The widespread use of combinatorial chemistry and high-throughput screening (HTS) for the discovery of lead compounds has created a large demand for small organic molecules that act on specific drug targets. These technologies focus on the generation of a huge number of molecules integrated with the biological screening of a very large number of samples. However, due to the ever increasing pressure to reduce drug development time and costs, there is a clear paradigm shift from the random screening of collections of compounds to a more rational process, which would directly affect the success rate of new chemical entity (NCE) generation and, therefore, improve pharmaceutical research and development (R&D) productivity. The definition and assessment of both chemical and biological space have revitalized the screening process model and emphasized the importance of exploring the intrinsic complementary nature of classical and modern methods in drug research. In this context, computational tools play an increasingly critical role in medicinal chemistry research programs [3,4].

Drug discovery is currently driven by innovation and knowledge employing a combination of experimental and computational methods. One of the most important challenges for the pharmaceutical industry is the identification of innovative NCEs from an incredibly large reservoir of real and virtual possible compounds. Several steps of the drug discovery process (e.g., hit identification, lead optimization, pharmacokinetic profile) can be improved in a rational way with the application of computational methods. Over the past decade, the high-performance computers, algorithms, methods and expertise have evolved and transformed structure-based drug design (SBDD) methods in tools of large impact in

modern drug discovery [5,6]. Although several efforts have been made to improve our understanding of the three-dimensional interactions involved in ligand-receptor binding and molecular recognition, it remains a major challenge for the computational technologies to accurately predict the binding affinity of new drug candidates [7].

In the process of ligand binding, the Gibbs free energy is governed by a combination of complex intermolecular interactions which determines drug-receptor affinity. Therefore, in order to be able to understand the several steps involved and to predict ligand-binding affinity, it is very useful a partition of the free energy of binding into individual, physically interpretable terms. This is not, of course, an easy issue to tackle, particularly concerning the relative calibration of the individual contributions against each other. On one hand, the enthalpic contributions to binding constant are reasonably well characterized in molecular descriptors that encode structural and shape information or empirical functions, such as molecular mechanics approximations and statistically derived atom-type and distance-dependent pair potentials that explore the capabilities of molecules to perform intermolecular interactions with a putative receptor [5,8,9]. On the other hand, entropic contributions are much more difficult to describe, particularly regarding solvent-to-protein transfer, where both the size of the hydrophobic surface area and the release of water molecules from the active site should be considered. Furthermore, ligand conformational flexibility has long been recognized as an important issue, since immobilization of rotational bonds at the binding site involves important entropy changes.

In order to estimate free energies associated with interactions between small-molecules and drug targets (e.g., enzymes, receptors), *ab initio* calculations and free energy perturbation are the preferred computational methods [10]. However, the complexity of the calculations associated with the time-consuming procedures make these methods rarely applicable in drug design [5,10]. In contrast, protein-ligand docking is an area of intense interest to both academia and industry [11]. Computer programs dedicated to docking small-molecules into protein binding sites have been receiving considerable attention in recent years because of their wide applications in medicinal chemistry. Docking methods such as DOCK [12], GOLD [13], FlexX [14], GLIDE [15,16], AUTODOCK [17], and Surflex-Dock [18] are widely used for the high-throughput sampling of ligands into protein binding pockets, with concomitant determination of the most likely binding modes and the estimation of the relative binding affinity [19].

*Address correspondence to this author at the Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil; Tel: + 55 16 3373-8095; Fax: +55 16 3373-9881; E-mail: andricopulo@ifsc.usp.br

Fig. (1). General procedure used in SBVS and LBVS.

VIRTUAL SCREENING

In general, the search for new biologically active molecules from large compound databases by means of computer-assisted techniques is a process known as virtual screening (VS). VS methods have rapidly become an essential component of the modern drug discovery process [20]. High-performance hardware and specialized software, combined with advanced knowledge of 3D pro-

tein structure and small-molecule binding modes, have made this technology a useful complement, and in some cases, a reasonable alternative to HTS. There are, fundamentally, two approaches to VS studies: i. structure-based virtual screening (SBVS), which requires knowledge of the 3D structures of target proteins to prioritize compounds by their complementarity to the binding site; and, ii. ligand-based virtual screening (LBVS), where no information on the protein is needed, instead, compounds known to bind to the protein are

used as queries to search databases for new molecules possessing biological activity [21-25]. Fig. (1) summarizes some important steps in both VS approaches.

STRUCTURE-BASED VIRTUAL SCREENING (SBVS)

In SBVS approaches, knowledge about the 3D structures of the target proteins is essential to perform *in silico* high-throughput receptor-ligand docking [26]. These macromolecular structures are usually determined by X-ray crystallography, NMR, and homology modeling [27-29]. Owing to its knowledge-based feature, VS strongly depends on the amount and quality of information available about the system under investigation. Regardless of what kind of protein will be employed as molecular target, important issues such as druggability of the target receptor, selection of the most relevant geometry, receptor flexibility, suitable assignment of protonation states and consideration of water molecules in binding site must be properly consider [24]. Another important step in SBVS is the appropriate design of databases of small-molecule candidates for the screening process. Usually, pharmaceutical companies possess their own private compound collections, which guarantee that the detected hits will be exclusive and will cover molecules for which the synthesis is well established. In addition, hundreds of thousands of commercially available compounds for VS can also be found in non-commercial databases or *in-house* collections of natural products and synthetic compounds, and so forth [24,30,31].

Screening libraries generally contain a large number of molecules with broad chemical diversity [32]. Members of such libraries are usually large in terms of size, with an average molecular weight falling close to typical drug-size molecules. However, several examples from drug discovery programs have demonstrated that small core fragments known to bind with substantial affinity are more suitable starting points for further lead optimization [33,34]. There-

fore, in order to generate a suitable screening subset, several molecular filters are available for selecting those molecules with required features. As a general criteria, compounds are primarily studied with respect to their drug-like (molecules which generally obey Lipinski's rule of five [35]) or lead-like (molecules which have lower molecular complexity when compared to drugs [36-38]) properties. Similarly, other important physicochemical features may be considered, as well as the pharmacokinetic profile (often referred to as ADME properties: Absorption, Distribution, Metabolism and Excretion).

Pharmacophores derived from receptor mapping play an important role in the design of focused collections of compounds. Due to an ever increasing number of structures solved and stored in the Protein Data Bank (PDB) [39], and also due to the development of methods that accurately probe and map ligand binding pockets in protein structures (GRID [40]; LUDI [41]; SuperStar [42]; Drug-Score [9,43]), the use of pharmacophores in SBDD studies has been remarkably increased [6,44,45]. For the generation of structure-based models, 3D pharmacophore hypotheses (Fig. 2) can be derived by considering the information gathered by the superposition of X-ray crystallographic structures (Fig. 3) and also from the analysis of the requirements imposed by the binding site through the identification of favorable regions for intermolecular interactions [46] (Fig. 4). The 3D pharmacophore queries are used to screen databases of compounds with the assistance of appropriate software, such as UNITY (Chemical Information Software, version 4.1, Tripos), Catalyst [47], or FlexX-Pharm [48], and only those molecules that carry the pharmacophoric features in the 3D space are retrieved and selected for the next VS steps.

One of the major challenges facing SBVS programs is the selection of an appropriate docking tool. The ligands (orientation and conformation) shall achieve the highest possible degree of

Fig. (2). Representation of a structure-based 3D pharmacophore model. Acceptor-site features of the protein and the complementary corresponding partner donor features are shown in blue; donor-site features of the protein and the complementary corresponding partner acceptor features are shown in red; and, the donor/acceptor features and the corresponding donor/acceptor site features of the threonine residues are shown in magenta.

Fig. (3). Structural superposition of three crystal structures of *T. cruzi* GAPDH in complex with an analogue of 1,3-bisphospho-d-glyceric acid (PDB code: 1QXS, in cyan), with an irreversible inhibitor 2-(2-phosphono-ethyl)-acrylic acid 4-nitro-phenyl ester (PDB code: 1ML3, in magenta), and with chalepin, a coumarin derivative inhibitor (PDB code: 1K3T, in green).

Fig. (4). Mapping of putative binding "hot spots" in the active site of *T. cruzi* GAPDH. "Hot spots" indicate areas within the protein binding site where certain functional groups should preferably bind. Panels A and D highlight properties of bulky and hydrophobic groups in a putative ligand: sp^3 carbon (orange) and DRY (yellow) as probes in DrugScore and GRID, respectively. Panels B and E highlight properties of a putative hydrogen-bond donor group in a ligand: amide nitrogen (blue) as probe in DrugScore and GRID. Panels C and F highlight properties of a putative hydrogen-bond acceptor group in a ligand: sp^2 oxygen and carbonyl oxygen (both in red) as probes in DrugScore and GRID, respectively.

molecular complementarity with respect to all binding sites of the receptor active site. Hence, the selection of the docking procedure is a crucial step for the success of the VS process. Currently available docking tools follow slightly different concepts, which make individual programs more suitable for a specific task. It is estimated that there are approximately 30 docking programs available [31]. The docking process involves the prediction of ligand orientation and conformation into the active site, followed by a measure of its fitness into the binding site (Fig. 5).

In general, there are two main complex issues that must be addressed during docking simulations: i. accurate structural modeling; and, ii. correct prediction of activity. Predictions of ligand binding

modes for small molecules in macromolecule binding sites are perhaps the most straightforward step, where, in fact, a considerable degree of success has been achieved. The docking programs employ different search methods (algorithms) to treat ligand flexibility. The search methods are divided into three basic categories: i. Systematic methods, such as incremental construction, conformational search, Hammerhead algorithm and databases, which explore all degrees of freedom in a molecule in order to place ligands (molecular fragments) into active sites of proteins; ii. Random or Stochastic methods, such as Monte Carlo and genetic algorithms (GA). The former generates several ligand configurations into the protein binding site and subsequently score the configurations in a multi-

Fig. (5). General docking procedure. Binding mode of the high-affinity selective inhibitor N⁶-(1-naphthalenemethyl)-2'-deoxy-2'-(3,5-dimethoxybenzamido) adenosine (NMD) to the GAPDH from *L. mexicana* (PDB code: 1I32) pocket. NMD minimum energy conformer (in white); NMD crystallographic conformer (in yellow); NMD docking solution (in green).

step procedure, while the latter deals with principles of biological competition and population dynamics to place the ligand in active site; iii. Simulation methods, such as molecular dynamics (MD) and energy minimization, which describe the molecular docking process in more detail. The MD method is based on Newton's equation of motion for an atomic system where atomic forces and masses are used to determine atomic positions over series of very small time steps. Energy minimization procedures are usually employed coupled with other search methods, since only a local energy minimum is reached. For example, minimization steps are carried out after each fragment addition followed by a final minimization before scoring [49].

Docking algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and protein active sites. Once the configurations of a system are sampled, the docking programs score them to identify the most likely candidates for the true structure [31,49]. Essentially, three types or classes of scoring functions are currently applied: i. Force field-based, which usually employs molecular mechanics force fields to quantify the sum of energies related to the receptor-ligand interaction energies and the internal ligand energies; ii. Empirical, that is based on the concept that binding energies can be approximated by a sum of individual uncorrelated terms. This type of scoring function is fit to reproduce experimental data, such as binding energies and/or conformations, as a sum of several parameterized functions; iii. Knowledge-based scoring functions, which uses atomic interaction-pair potentials to derive statistical potentials of mean force from large sets of protein-ligand complexes.

Scoring functions implemented in docking programs make different assumptions and simplifications in the evaluation of complexes and do not fully account for a number of physical phenomena that determine molecular recognition. For example, ligand-binding events are driven by a combination of enthalpic and entropic effects, where either entropy or enthalpy can dominate specific interactions. This often presents a conceptual problem for contemporary scoring functions, because most of them are much more focused on capturing energetic than entropic effects [50]. Several scoring functions have been designed to reproduce binding energies of protein-ligand systems. Usually, databases of 50-300 protein-ligand complexes are employed for validating the docking programs. Over the past years, a variety of scoring functions have been developed and assessed [11,51,52]. One of the most important parameters during the assessment procedures is the discriminatory power of the applied scoring function for ranking and enriching of

potentially active binders at the top list of the docking solutions. It is, however, a hard issue to address most likely because many physicochemical parameters and the general phenomena involved at the molecular level of the binding process are not yet fully understood. Therefore, these relevant features have not been incorporated into scoring functions in a proper way. It can be illustrated with the knowledge-based scoring function DrugScore [9], which was developed based on crystal structure information from the Cambridge Structural Database (CSD) [53] and protein-ligand complexes from the PDB [39]. This scoring function was employed to validate a test set of 56 crystallography complexes yielding an r^2 value of 0.56 for the correlation between experimental and predicted binding affinities [43]. This indicates that the generation of models possessing improved predictive ability (higher r^2 values) would require the incorporation of a better description of the complex molecular events related to binding affinity.

Scoring function programs have different parameters which rely on distinct atom type schemes and atomic partial charges calculation methods, and have been trained on diverse ligand-protein data sets. As a result, each program returns a particularly different estimate of relative binding affinity, and comparisons are nontrivial. To overcome this limitation, several approaches that assign low ranks to most of inactive compounds while assigning high ranks to most of active compounds have been proposed as alternative methods. One of the most common and employed strategy is to combine estimates from a variety of scoring functions into a single consensus score [20]. Several reports show successful applications of this approach to improve hit rates significantly [54,55]. The impact of consensus scoring strategies in the enrichment of true positives (leads) in SBVS can be explained by the fact that the mean of repeated samplings tends to be closer to the true value than any single sampling, thus, since useful scoring functions perform well, different methods will vote for some of the same actives [55,56]. This process contributes to a better understanding of the chemistry involved in ligand binding and also improves the enrichment of true positives. Consensus scoring is a relatively recent field of research in drug design, with a history of about 10 years. Rapidly, it has become an important tool in the field of *in silico* technologies for drug discovery. However, significant improvements are required in the efficiency of the consensus scoring methods, mainly regarding the full description of the molecular events involved in the prediction of binding affinity. Even though challenging issues such as the consideration of water molecules and protein flexibility are somewhat implemented by some docking tools, they still require substantial development to become more useful in drug design. Re-

ardless of what kind of docking tool is considered, docking and scoring are critical issues in virtual drug screening methods.

SUCCESSFUL APPLICATIONS OF SBVS

A number of recent successful applications of SBVS can be found in the literature for an impressive variety of drug targets from different therapeutic areas [8,20,24,30,57]. For example, VS studies were performed on a potential antimalarial drug target of the reductase family, an enoyl-acyl carrier protein reductase (ENR), which plays an important role in *Plasmodium falciparum* membrane construction and energy production and does not have a human homolog [58]. In summary, employing a Monte Carlo global energy optimization for flexible ligand docking, a database of 336,600 compounds from the ChemBridge Express Library (San Diego, CA) was fully screened. After applying some ADME filters, 169 compounds with suitable pharmacokinetic properties were retrieved and experimentally tested for their ability to inhibit the ENR activity. In this investigation, 16 inhibitors were identified with three of them having IC_{50} values in the micromolar range (5, 10 and 25 μM , Fig. 6A), which fall close to the inhibitory activity of triclosan, a potent *Pf*-ENR inhibitor ($IC_{50} = 4 \mu M$). In another example, a target protein of the STAT (Signal Transducer and Activator of Transcription) family was studied [59]. Irregular activity of one of the family members (Stat3) contributes to carcinogenesis and tumor progression by up-regulating gene expression and promoting dysregulated growth. Described as a critical step in STAT activation, the dimerization between two Stat3 monomers presents an attractive target to inhibit Stat3 DNA-binding and transcriptional activity. Investigations were conducted with the GLIDE (Grid-based Ligand Docking from Energetics) software [15], which employs an incremental construction algorithm to explore ligand flexibility, for the screening of 150,829 compounds from the National Cancer Institute (NCI) chemical libraries. These studies identified the high-scoring compound NSC 74859 (Fig. 6B), which selectively inhibits Stat3 DNA-binding activity *in vitro* with an IC_{50} value of 86 μM (IC_{50} values toward Stat1 and Stat5 were $>300 \mu M$ and 170 μM , respectively; no interaction with the Src protein family was observed, no significant effect on Erk1/2 or Shc, and low toxicity to cells with no aberrant Stat3). Furthermore, the compound induces growth inhibition and apoptosis of malignant cells as well as induces human breast tumor regression in xenograft models.

LIGAND-BASED VIRTUAL SCREENING (LBVS)

Ligand-based methods play an important role in medicinal chemistry and modern hit generation and lead identification. The discovery of a large number of novel biologically active compounds has been possible with the use of LBVS methods, for which no 3D information on the target protein is required [22,25]. In this approach, the screening process is based on the use of active compounds as templates. Ligand-based screening techniques essentially focus on comparative molecular similarity analysis of compounds with known and unknown activity, regardless of the methods or algorithms used. LBVS methods can be divided into three major

classes: i. Similarity searching, including 2D fingerprints (Tanimoto, Cosine, Hamming, Russel-Rao, and Forbes coefficients) [60], rely on the Similar Property Principle. This principle states that molecules that are structurally related are likely to have similar properties, thus, databases of structurally similar molecules which have not been tested for biological activity are likely to have compounds that exhibit the activity of interest; ii. Pharmacophore methods based on the pharmacophore concept, which is commonly defined as a three-dimensional arrangement of molecular features or fragments forming a necessary, but not necessarily sufficient, condition required for binding [61]. These methods involve the identification of the pharmacophoric pattern common to a set of known active compounds, and the use of this pattern in 3D substructure search; and, iii. Machine learning methods, in which a classification rule is developed from a training set containing known active and inactive molecules. For a more comprehensive review of similarity searching, pharmacophore generation and search, machine learning and artificial intelligence approaches, the readers are referred to recent excellent reviews [22,25,62-65].

SUCCESSFUL APPLICATIONS OF LBVS

There are numerous possible ways of applying LBVS strategies and several examples are available in the literature showing the usefulness of these methods to drug design [22,66]. For example, ligand-based methods were applied for the drug target enzyme 5-lipoxygenase (5-LO), that catalyzes the first steps in the conversion of arachidonic acid into leukotrienes, which has been associated with atherosclerosis, cancer, and osteoporosis [67]. In the search for new bioactive compounds, 43 known 5-LO inhibitors were assembled and used in similarity searching in the speedCATS software (Fig. 7A), which employs chemically advanced template search (CATS) descriptors [68]. Eighteen compounds were selected based on the smallest distance to their query molecules, and their 5-LO inhibitory activities were evaluated in a cell-based assay. Finally, two inhibitors from different structural classes exhibited submicromolar inhibitory activity in an intact PMNL cell assay (Fig. 7A). Another example involves the search for new ligands of the ATP-sensitive potassium channels (K_{ATP} channels). K_{ATP} channels couple changes in blood glucose concentrations to insulin secretion in pancreatic β -cells and are considered promising drug targets for the treatment of disorders resulting from excessive insulin release [69]. Accordingly, potassium channel openers (KCOs) comprise chemically heterogeneous classes of compounds and have demonstrated significant potential for the treatment of diabetes. In order to search for new KCOs chemotypes, LBVS studies were carried out employing a pharmacophore model comprising compounds from several structurally diverse classes [69], and a subset of the ZINC database with approximately 65,000 compounds. Pharmacokinetic properties were modeled by molecular interaction fields (MIF) based on Vol-Surf descriptors, which retained about 1,900 compounds with drug-like characteristics. Subsequently, pharmacodynamic properties were computed by GRIND (GRid-INdependent Descriptors) [70], FLAP (Fingerprints for Ligands and Proteins) [71] and TOPP (Triplets Of Pharmacophoric Points) [72] methods, which are based on

Fig. (6). Inhibitors discovered by SBVS.

MIF and molecular fingerprints. Six potent ligands of the K_{ATP} channels were used as a template for deriving the pharmacophore hypothesis (Fig. 7B). The top-ranked compounds according to the pharmacophore-based approaches were checked for chemical stability and toxicity by visual inspection, and a final set of 32 molecules were retrieved. The effects on membrane potentials and glucose-stimulated insulin release were assessed by HEK 293 and INS-1 cell assays, respectively, revealing three compounds that were able to inhibit insulin release with micromolar potency (Fig. 7B).

INTEGRATING VS AND HTS

Modern drug discovery involves the integration of a wide variety of technologies and expertise in multidisciplinary research teams. The influence of the synergy effects between HTS and VS on the selection of hits and lead compounds is a good example of the possible integration of advanced technologies in drug discovery programs. However, it is worth noting that the two approaches are quite distinct in nature and procedures. On one hand, HTS is largely phenomenological and technology-driven, being much more influenced by advances in automation and miniaturization [24,73]. This experimental procedure is costly, complex and labor intensive, with the estimated cost for a basic HTS process, without assay development, being approximately US\$75,000. Alone, biological screening and preclinical pharmacological testing account for about 15% of the total R&D expenditures of the pharmaceutical industry [73]. On the other hand, VS, a knowledge-driven approach, depends on the available information concerning target structure and small-molecule chemical space. As a starting point, either the 3D structure of the macromolecular target or a known active ligand is a prerequisite for VS studies. This computational method offers economy, speed and flexibility to drive drug discovery projects.

These technologies (VS and HTS) are complementary in the sense that they have mutual goals, that is, finding new hits and leads in drug research programs [24,74]. However, an important difference between them is related to the knowledge about the

mechanism of action of ligands upon biological receptors. In the HTS process, no or little information is provided about the possible target binding site or mechanism of action of the selected ligands. In this situation, further structural and kinetic studies would be required to elucidate them. Conversely, VS methods are capable of selecting molecules based on a specific 3D target binding site, therefore, useful information about the binding mode and mechanism of action could also be explored [75]. HTS commonly involves running a primary screen assay (single replicate, single compound concentration) on a large collection of compounds, followed by subsequent rounds of single-shot and dose-response screening. These rapid large-scale assays are monitored by spectroscopic methods such as fluorescence, absorbance and luminescence, being frequently performed in microtiter plates with 96–1536 wells. Since the biochemical assays are designed to screen hundreds of thousands of compounds per run, each plate usually contains internal control wells to ensure quality and comparability of results between plates [76]. Despite the fact that several strategies have been employed to monitor and control the quality and accuracy of the *in vitro* assays, key problems in HTS are the occurrence of false-positive hits (molecules that appear to inhibit the target but turn out to be uninteresting compounds), and the prevalence of nonspecific or promiscuous inhibitors [77]. VS methods are valuable tools designed for the search of large databases of compounds and selection of a reduced number of candidates for biological evaluation. The integration of VS techniques either previously or in connection with HTS methods has three fundamental objectives: i. to extend the scope of the screening to external databases; ii. to identify a larger number and structurally diverse hits; and, iii. to reduce the assay-to-lead attrition rate observed from HTS [78]. VS concepts complement HTS by including compound-filtering techniques based on functions ranging from simple rule-based to more complex neural network architectures. This approach aims to enrich libraries with molecules that have desirable or drug-like properties, eliminating those compounds that have unwanted characteristics for new leads. Alternatively, VS also provides tools for HTS database analysis, being capable of extracting useful information for database mining.

Fig. (7). Inhibitors discovered by LBVS strategies in medicinal chemistry research programs.

Even though structures of target proteins are becoming increasingly available, mainly owing to genomics, proteomics and bioinformatics advances, LBVS methods are still dominating the VS field, basically due to the fact that hit and lead information are the predominant source of knowledge in many cases [73]. The most frequently applied VS methods for HTS purposes rely on fast 2D descriptors employed by similarity-based methods. Thus, HTS hits are compared to each other to generate a hypothesis on the underlying lead structure, and then structure-activity relationships (SAR) can be generated to guide future medicinal chemistry efforts. Basically, two main computer-aided strategies have been described for the identification of active compounds: i. sequential screening; and, ii. one-shot screening. Sequential screening aims to reduce the

amount of compounds that need to be tested by selecting a subset from the original database for experimental evaluation and SAR studies. As for the one-shot screening, all of the database compounds are experimentally evaluated, and the iterative construction of a model for an active compound is then carried out based on the classical train-and-test paradigm [22].

The integration of HTS and VS shows that the combination of both experimental and computational efforts is feasible at many different levels of drug design, including library generation, compound prioritization, and data analysis [79]. VS techniques can be used to analyze the growing number of noisy data points from HTS experiments, whereas HTS techniques can be used to validate VS

MES_{EXP} = maximal electroshock experiment. Method for the identification of new drug candidates for partial and generalized seizures [85].

Fig. (8). Hits discovered using a combination VS and QSAR methods.

results [22]. It is important to emphasize that, in contrast to HTS, VS methods are more easily accessible to academic laboratories around the world, without the need for expensive materials, equipments, and complex infrastructure that an empirical screen demands.

INTEGRATING VS AND QSAR

Another important example of integration of modern technologies in drug discovery is the combination of VS and QSAR (*Quantitative Structure-Activity Relationships*). QSAR techniques generate descriptors based on molecular structures and uses computational algorithms to relate the key descriptors to the dependent property value of interest [80-85]. Medicinal chemistry studies aimed at elucidating fundamental aspects of the relationships between structural or property descriptors and biological activity are important in the understanding of the activity of interest and may enable the prediction of the biological property for new compounds [86-90]. QSAR has a long history in the drug discovery field, and reached a tremendous impact in the optimization of promising leads that act on specific targets. The availability of advanced molecular modeling techniques and several 2D and 3D QSAR methods has attracted the attention of many scientists around the world for the integration of computational drug design tools. Aiming at overcoming the inherent limitations, the integration of VS and QSAR strategies provides useful opportunities to partially fulfill each method limitation, as well as allows the capture and incorporation of valuable information for the design of new small-molecule drug candidates. A variety of studies describing the integration of these drug design techniques has been reported in the literature [91-96]. For example, investigations were conducted for the discovery of inhibitors of 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGCR) [91], through the integration of 3D QSAR CoMFA (Comparative Molecular Field Analysis) models [97] and FlexE [98] pre-filters based on energy score. As a result, eight novel non-statin-like scaffolds with promising biological activity (Fig. 8A) were identified. Another important example is the development of 3D QSAR CoMSIA (Comparative Molecular Similarity Indices Analysis) models [99] for a series of structurally diverse modulators of the nonsteroidal progesterone receptor (PR) [92]. The QSAR models generated were useful for VS of novel selective nonsteroidal modulators as an alternative receptor-specific scoring function for predicting binding affinities. Instead of searching for models capable of predicting the actual biological activity value, QSAR studies were performed to create models that enable assessment of hypothetical docking scores [93]. In this approach, QSAR models would be capable of identifying most probable nonbinders in docking databases, and also be used together with other pre-docking filters. Similarly, MLR (Multiple Linear Regression) models were developed for a series of chemokine receptor (CCR5) modulators, in order to create a useful alternative to filter out dissimilar compounds and to identify novel potent compounds [94].

Models designed for the prediction of new ligands even when no 3D receptor model is available have been described using a combination of QSAR and VS [95]. Owing to the wide application of QSAR methods, a huge number of molecular descriptors is available. Based on that, techniques that avoid the conformational and alignment ambiguities inherent to 3D QSAR methods have also been employed to produce predictive models useful for database mining and VS [94-96]. In this context, 3D and 2D QSAR models (CoMFA [97], VolSurf [100] and Hologram QSAR [101]) were assessed for the prediction of new ligands towards the nuclear receptor aryl hydrocarbon receptor (AhR) [95]. This study showed the robustness of "structural alignment free" methods as an interesting alternative for VS. HQSAR and VolSurf are fast and highly predictive techniques capable of rapidly generating suitable models for VS of large databases of compounds. In addition, QSAR models derived by *k* nearest neighbor (*k*NN) and molecular topological

descriptors were employed for the discovery of novel anticonvulsant agents [96]. In this study, ten QSAR models were used for the analysis of both Maybridge and National Cancer Institute (NCI) databases. Twenty-two compounds with high predicted activity values were selected and evaluated *in vivo*, leading to the discovery of seven new bioactive compounds (Fig. 8B).

CONCLUSION

Computational methods play a crucial role in modern medicinal chemistry, presenting a unique potential for transforming the early phases of drug research, particularly in terms of time and cost savings. In this scenario, SBVS and LBVS in combination with other techniques, such as HTS and QSAR, are useful synergistic approaches in drug discovery. VS techniques have been widely employed in the design of focused libraries, compound-filtering, database-mining, and for the rapid analysis of large databases subjected to HTS procedures. Similarly, QSAR methods have been employed as useful tools for guiding the selection of compounds for high-throughput analysis. VS strategies are influencing traditional strategies to analyze large databases of chemical compounds in order to identify possible lead candidates with good pharmacodynamic and pharmacokinetic profiles. In summary, it is always important to understand the limits and scope of any computational tool, and to distinguish between those opportunities which are appropriate to apply the VS technology to pursue a new drug candidate.

ACKNOWLEDGEMENTS

We gratefully acknowledge financial support from the State of São Paulo Research Foundation (FAPESP) and from the National Council for Scientific and Technological Development (CNPq), Brazil.

REFERENCES

- [1] Hajduk, P.J.; Greer, J. *Nat. Rev. Drug Discov.* **2007**, *6*, 211-219.
- [2] Hopkins, A.L.; Groom, C.R. *Nat. Rev. Drug Discov.* **2002**, *1*, 727-730.
- [3] Overington, J.P.; Al-Lazikani, B.; Hopkins, A.L. *Nat. Rev. Drug Discov.* **2006**, *5*, 993-996.
- [4] Andricopulo, A.D.; Montanari, C.A. *Mini-Rev. Med. Chem.* **2005**, *5*, 585-593.
- [5] Tame, J.R. *J. Comput. Aided. Mol. Des.* **1999**, *13*, 99-108.
- [6] Gruneberg, S.; Stubbs, M.T.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 3588-3602.
- [7] Sippl, W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 825-830.
- [8] Klebe, G. *Perspect Drug Discov. Des.* **1998**, *12-14*, 87-104.
- [9] Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337-356.
- [10] Reddy, M.R.; Singh, U.C.; Erion, M.D. *J. Comput. Chem.* **2007**, *28*, 491-494.
- [11] Hartshorn, M.J.; Verdonk, M.L.; Chessari, G.; Brewerton, S.C.; Mooij, W.T.; Mortenson, P.N.; Murray, C.W. *J. Med. Chem.* **2007**, *50*, 726-741.
- [12] Shoichet, B.K.; Kuntz, I.D. *J. Mol. Biol.* **1991**, *22*, 327-346.
- [13] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-748.
- [14] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470-489.
- [15] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. *J. Med. Chem.* **2004**, *47*, 1739-1749.
- [16] Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pol-lard, W.T.; Banks, J.L. *J. Med. Chem.* **2004**, *47*, 1750-1759.
- [17] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- [18] Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499-511.
- [19] Clark, R.D.; Strizhev, A.; Leonard, J.M.; Blake, J.F.; Matthew, J.B. *J. Mol. Graph. Model.* **2002**, *20*, 281-295.
- [20] Feher, M. *Drug Discov. Today* **2006**, *11*, 421-428.
- [21] Perola, E.; Walters, W.P.; Charifson, P.S. *Proteins* **2004**, *56*, 235-249.
- [22] Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M. *Drug Discov. Today* **2004**, *9*, 27-34.
- [23] Cummings, M.D.; DesJarlais, R.L.; Gibbs, A.C.; Mohan, V.; Jaeger, E.P. *J. Med. Chem.* **2005**, *48*, 962-976.
- [24] Klebe, G. *Drug Discov. Today* **2006**, *11*, 580-594.
- [25] Eckert, H.; Bajorath, J. *Drug Discov. Today* **2007**, *12*, 225-233.
- [26] Waszkowycz, B. *Curr. Opin. Drug Discov. Dev.* **2002**, *5*, 407-413.
- [27] Davis, A.M.; Teague, S.J.; Kleywegt, G.J. *Angew. Chem. Int. Ed.* **2003**, *42*, 2718-2736.

- [28] Kasimova, M.R.; Kristensen, S.M.; Howe, P.W.A.; Christensen, T.; Matthiesen, F.; Petersen, J.; Sorensen, H.H.; Led, J.J. *J. Mol. Biol.* **2002**, *318*, 679-695.
- [29] Oshiro, C.; Bradley, E.K.; Eksterowicz, J.; Evensen, E.; Lamb, M.L.; Lantot, J.K.; Putta, S.; Stanton, R.; Grootenhuys, P.D.J. *J. Med. Chem.* **2004**, *47*, 764-767.
- [30] Irwin, J.J.; Shoichet, B.K. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
- [31] Leach, A.R.; Shoichet, B.K.; Peishoff, C.E. *J. Med. Chem.* **2006**, *49*, 5851-5855.
- [32] Skarzynski, T.; Thorpe, J. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 102-107.
- [33] Erlanson, D.A.; McDowell, R.S.; O'Brien, T. *J. Med. Chem.* **2004**, *47*, 3463-3482.
- [34] Rees, D.C.; Congreve, M.; Murray, C.W.; Carr, R. *Nat. Rev. Drug Discov.* **2004**, *3*, 660-672.
- [35] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26.
- [36] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug Discov. Today* **2003**, *8*, 876-877.
- [37] Hann, M.M.; Oprea, T.I. *Curr. Opin. Chem. Biol.* **2004**, *8*, 255-263.
- [38] Oprea, T.I.; Allu, T.K.; Fara, D.C.; Rad, R.F.; Ostopovici, L.; Bologna, C.G. *J. Comput. Aided Mol. Des.* **2007**, *21*, 113-119.
- [39] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- [40] Goodford, P.J. *J. Med. Chem.* **1985**, *28*, 849-857.
- [41] Böhm, H.J. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593-606.
- [42] Verdonk, M.L.; Cole, J.C.; Taylor, R. *J. Mol. Biol.* **1999**, *289*, 1093-1108.
- [43] Gohlke, H.; Hendlich, M.; Klebe, G. *Perspect. Drug Discov. Des.* **2000**, *20*, 115-144.
- [44] Brenk, R.; Naerum, L.; Gradler, U.; Gerber, H.D.; Garcia, G.A.; Reuter, K.; Stubbs, M.T.; Klebe, G. *J. Med. Chem.* **2003**, *46*, 1133-1143.
- [45] Evers, A.; Klebe, G. *J. Med. Chem.* **2004**, *47*, 5381-5392.
- [46] Sottriffer, C.; Klebe, G. *Farmaco* **2002**, *57*, 243-251.
- [47] Kurogi, Y.; Güner, O.F. *Curr. Med. Chem.* **2001**, *8*, 1035-1055.
- [48] Hindle, S.A.; Rarey, M.; Buning, C.; Lengauer, T. *J. Comput. Aided Mol. Des.* **2002**, *16*, 129-149.
- [49] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
- [50] Salaniwal, S.; Manas, E.S.; Alvarez, J.C.; Unwalla, R.J. *Proteins* **2007**, *66*, 422-435.
- [51] Stahl, M.; Rarey, M. *J. Med. Chem.* **2001**, *44*, 1035-1042.
- [52] Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287-2303.
- [53] Allen, F.H. *Acta Crystallogr.* **2002**, *B58*, 380-388.
- [54] Jenkins, J.L.; Kao, R.Y.; Shapiro, R. *Proteins* **2003**, *50*, 81-93.
- [55] Baber, J.C.; Shirley, W.A.; Gao, Y.; Feher, M. *J. Chem. Inf. Model.* **2006**, *46*, 277-288.
- [56] Wang, R.; Wang, S. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422-1426.
- [57] Kubinyi, H. *Success Stories of Computer-Aided Design, Computer Applications in Pharmaceutical Research and Development*; John Wiley & Sons: New York, **2006**, pp. 377-424.
- [58] Nicola, G.; Smith, C.A.; Lucumi, E.; Kuo, M.R.; Karagyzov, L.; Fidock, D.A.; Sacchettini, J.C.; Abagyan, R. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 686-691.
- [59] Siddiquee, K.; Zhang, S.; Guida, W.C.; Blaskovich, M.A.; Greedy, B.; Lawrence, H.R.; Yip, M.L.; Jove, R.; McLaughlin, M.M.; Lawrence, N.J.; Sebt, S.M.; Turkson, J. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7391-7396.
- [60] Johnson, M.A.; Maggiora, G.M. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons: New York, **1990**, pp. 239-288.
- [61] Ghose, A.K.; Wendoloski, J.J. *Perspect. Drug Discov. Des.* **1998**, *9-11*, 253-271.
- [62] Patel, Y.; Gillet, V.J.; Bravi, G.; Leach, A.R. *J. Comput. Aided Mol. Des.* **2002**, *16*, 653-681.
- [63] Willett, P. *Drug Discov. Today* **2006**, *11*, 1046-1053.
- [64] Duch, W.; Swaminathan, K.; Meller, J. *Curr. Pharm. Des.* **2007**, *13*, 1497-1508.
- [65] Khedkar, S.A.; Malde, A.K.; Coutinho, E.C.; Srivastava, S. *Med. Chem.* **2007**, *3*, 187-197.
- [66] Hert, J.; Willett, P.; Wilton, D.J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. *J. Chem. Inf. Model.* **2006**, *46*, 462-470.
- [67] Franke, L.; Schwarz, O.; Müller-Kuhrt, L.; Hoernig, C.; Fischer, L.; George, S.; Tanrikulu, Y.; Schneider, P.; Werz, O.; Steinhilber, D.; Schneider, G. *J. Med. Chem.* **2007**, *50*, 2640-2646.
- [68] Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. *J. Comput. Aided Mol. Des.* **2003**, *17*, 687-698.
- [69] Carosati, E.; Mannhold, R.; Wahl, P.; Hansen, J.B.; Fremming, T.; Zamora, I.; Cianchetta, G.; Baroni, M. *J. Med. Chem.* **2007**, *50*, 2117-2126.
- [70] Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. *J. Med. Chem.* **2000**, *43*, 3233-3243.
- [71] Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J.S. *J. Chem. Inf. Model.* **2007**, *47*, 279-294.
- [72] Sciabola, S.; Morao, I.; de Groot, M.J. *J. Chem. Inf. Model.* **2007**, *47*, 76-84.
- [73] Bajorath, J. *Nat. Rev. Drug Discov.* **2002**, *11*, 882-894.
- [74] Mestres, J. *Biochem. Soc. Transact.* **2002**, *30*, 797-799.
- [75] Klebe, G.; Alvarez, J.; Shoichet, B. *Virtual Screening in Drug Discovery*; Taylor & Francis: New York, **2005**; pp. 3-4.
- [76] Wu, G.; Doberstein, S.K. *Drug Discov. Today* **2006**, *11*, 718-724.
- [77] Harper, G.; Pickett, S.D. *Drug Discov. Today* **2006**, *11*, 694-699.
- [78] Stahura, F.L.; Bajorath, J. *Comb. Chem. High Throughput Screen.* **2004**, *7*, 259-269.
- [79] Parker, C.N.; Shamu, C.E.; Kraybill, B.; Austin, C.P.; Bajorath, J. *Drug Discov. Today* **2006**, *11*, 863-865.
- [80] Honorio, K.M.; Garratt, R.C.; Andricopulo, A.D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119-3125.
- [81] Andricopulo, A.D.; Yunes, R.A. *Chem. Pharm. Bull.* **2001**, *49*, 10-17.
- [82] Castilho, M.S.; Postigo, M.P.; de Paula C.B.V.; Montanari, C.A.; Oliva, G.; Andricopulo, A.D. *Bioorg. Med. Chem.* **2006**, *14*, 516-527.
- [83] Leitão, A.; Andricopulo, A.D.; Oliva, G.; Pupo, M.T.; de Marchi, A.A.; Vieira, P.C.; da Silva, M.F.G.F.; Ferreira, V.F.; de Souza, M.C.B.V.; Sa, M.M.; Moraes, V.R.S.; Montanari, C.A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2199-2204.
- [83] Krogh, R.; Yunes, R.A.; Andricopulo, A.D. *Farmaco* **2000**, *55*, 730-735.
- [84] Andricopulo, A.D.; Yunes, R.A.; Cechinel, V.; Nunes, R.J.; Frazer, J.W.; Cordes, E.H. *Pharmazie* **1999**, *54*, 698-704.
- [85] Castilho, M.S.; Guido, R.V.C.; Andricopulo, A.D. *Lett. Drug Des. Discov.* **2007**, *4*, 106-113.
- [86] Andricopulo, A.D.; Akoachere, M.B.; Krogh, R.; Nickel, C.; McLeish, M.J.; Kenyon, G.L.; Arscott, L.D.; Williams, C.H.; Davioud-Charvet, E.; Becker, K. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2283-2292.
- [87] Andricopulo, A.D.; Willain, A.; Correa, R.; Santos, A.R.S.; Nunes, R.J.; Yunes, R.A.; Cechinel, V. *Pharmazie* **1998**, *53*, 493-494.
- [88] Cechinel, V.; Correa, R.; Vaz, Z.; Calixto, J.B.; Nunes, R.J.; Pinheiro, T.R.; Andricopulo, A.D.; Yunes, R.A. *Farmaco* **1998**, *53*, 55-57.
- [89] Andricopulo, A.D.; Yunes, R.A.; Nunes, R.J.; Savi, A.O.S.; Correa, R.; Cruz, A.B.; Cechinel, V. *Quím. Nova* **1998**, *21*, 573-577.
- [90] Lima, E.O.; Queiroz, E.F.; Andricopulo, A.D.; Nunes, R.J.; Yunes, R.A.; Correa, R.; Cechinel, V. *Bol. Soc. Chil. Quím.* **1999**, *44*, 185-189.
- [91] Zhang, Q.Y.; Wan, J.; Xu, X.; Yang, G.F.; Ren, Y.L.; Liu, J.J.; Wang, H.; Guo, Y. *J. Comb. Chem.* **2007**, *9*, 131-138.
- [92] Söderholm, A.A.; Lehtovuori, P.T.; Nyronen, T.H. *J. Med. Chem.* **2006**, *49*, 4261-4268.
- [93] Cherkasov, A.; Ban, F.; Li, Y.; Fallahi, M.; Hammond, G.L. *J. Med. Chem.* **2006**, *49*, 7466-7478.
- [94] Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P.A.; Markopoulos, J.; Igglei-Markopoulou, O. *J. Comput. Aided Mol. Des.* **2006**, *20*, 83-95.
- [95] Lo Piparo, E.; Koehler, K.; Chana, A.; Benfenati, E. *J. Med. Chem.* **2006**, *49*, 5702-5709.
- [96] Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J.P.; Kohn, H.; Tropsha, A. *J. Med. Chem.* **2004**, *47*, 2356-2364.
- [97] Cramer III, R.D.; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
- [98] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377-395.
- [99] Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130-4146.
- [100] Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000**, *11*, S29-S39.
- [101] Tong, W.; Lowis, D.R.; Perkins, R.; Chen, Y.; Welsh, W.J.; Goddette, D.W.; Heritage, T.W.; Sheehan, D.M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669-677.

Classical and Hologram QSAR Studies on a Series of Tacrine Derivatives as Butyrylcholinesterase Inhibitors

APÊNDICE D

M.S. Castilho^{*,a}, R.V.C. Guido^b and A.D. Andricopulo^b

^aLaboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-290, Salvador-BA, Brazil

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

Received September 20, 2006; Revised October 30, 2006; Accepted November 02, 2006

Abstract: Alzheimer's disease (AD) is a major neurodegenerative disease that affects mainly people over 65 years of age. The main therapeutic approach to AD is based on the use of acetylcholinesterase (AChE) inhibitors. Recent studies have shown that butyrylcholinesterase (BuChE) can also be considered an attractive target for the development of novel drugs for AD therapy. The design of new potent and selective BuChE inhibitors is of great importance in drug discovery. 2D quantitative structure-activity relationship studies were conducted on a series of potent inhibitors of human BuChE using classical and hologram QSAR (HQSAR) approaches. A training set of 40 compounds was employed to derive the models. Classical QSAR models showed moderate correlation ($r^2 = 0.836$, $q^2 = 0.750$), with no substantial predictive power for untested compounds. On the other hand, the best HQSAR model ($r^2 = 0.928$, $q^2 = 0.723$) was used to predict the potency of 10 test set compounds, and the predicted values were in good agreement with the experimental results, showing the potential of this model for new untested compounds.

Keywords: 2D QSAR, HQSAR, Alzheimer, Tacrine derivatives, Butyrylcholinesterase, Inhibitors.

1. INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative disease that accounts for approximately 60% of all cases of dementia among people over 65 years of age [1]. Although the AD etiology remains elusive, its morphological and biochemical characteristics are well documented, that is,

deposits of β -amyloid plaques, intracellular formation of neurofibrillar tangles, loss of neuronal synapses and pyramidal neurons [1-3]. Overall, these alterations result in loss of memory and cognitive ability. These symptoms are thought to be related to cholinergic neurons in the cerebral cortex and subcortical structures [4]. For this reason, the main

Fig. (1). Chemical structure of drugs available for AD therapy.

*Address correspondence to this author at the Laboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-290, Salvador-BA, Brazil; Tel: +55 71 3332-1580; Fax: +55 71 3235-9350; E-mail: castilho@ufba.br

therapeutic approach to AD involves the improvement of cholinergic transmission. Some acetylcholinesterase (AChE) inhibitors (Fig. (1)), such as tacrine (A), donepezil (B), rivastigmine (C) and galanthamine (D), are commercially available drugs for AD therapy [1-3].

Table 1. Chemical Structures and IC₅₀ Values for a Series of Butyrylcholinesterase Inhibitors

Training Set Molecules														
Cpd	Structure	IC ₅₀ [nM]	Cpd	Structure	IC ₅₀ [nM]	Cpd	Structure	IC ₅₀ [nM]						
1		14	10		41	23		5,900						
2		2,300	11		88,000	24		2,580						
3		36,000	12		6,700	25		26,400						
Cpd	R ¹	IC ₅₀ [nM]	Cpd	R ¹	R ²	R ³	R ⁴	IC ₅₀ [nM]	Cpd	R ¹	R ²	R ³	R ⁴	IC ₅₀ [nM]
4	H	33,000	13	F	H	H	H	5,700	26	Me	H	H	H	126
5	Cl	10,300	14	H	H	H	OMe	3,000	27	Et	H	H	H	79
			15	H	Benz	H	H	500	28	Me	H	Me	H	449
			16	H	Me	H	H	1,600	29	Me	H	F	H	197
						30	Et	H	Me	H	208			
									31	Et	Me	H	H	512
									32	Me	F	F	H	963
									33	Me	H	Cl	H	236
									34	Et	F	H	H	109
Cpd	R ¹	IC ₅₀ [nM]	Cpd	n	IC ₅₀ [nM]	35	Et	H	F	H	57			
6	Me	12,200	17	2	48	36	Et	F	F	H	194			
7	Bu	81,000	18	4	5.04	37	Et	Cl	H	H	331			
			19	5	1.48	38	Et	H	Cl	H	16			
			20	6	3.24	39	Et	Cl	Cl	H	1,210			
			21	7	8.58									

(Table 1). contd.....

22				50	40				329
Cpd	R ¹	IC ₅₀ (nM)										
8	Me	148										
9	Et	63										

Test Set Molecules					
Cpd	Structure	IC ₅₀ [nM]	Cpd	Structure	IC ₅₀ [nM]
41		52	42		7,100
43		3,100	44		33
45		18,900	46		762
47		42,000	48		543
49		12	50		11

Butyrylcholinesterase (BuChE), a second cholinesterase, has been identified as a potential target for drug intervention. Recent studies have shown that AChE activity decreases to about 15% of normal levels in AD patients, while BuChE activity remains unchanged or even increases [5]. In addition, cognitive improvement correlates more strongly with BuChE inhibition than AChE inhibition [6]. This leads to new perspectives related to the therapeutic use of selective BuChE inhibitors or, at least, drugs that inhibit

both cholinesterases [4,7]. However, it must be pointed out that BuChE is not a validated nor a consensual target [8]. Therefore, it requires further investigation in order to assess its real value in AD treatment.

There is a large number of quantitative structure-activity relationship (QSAR) studies concerning the chemical structure and corresponding biological information for AChE inhibitors [9-13]. However, the lack of comparable studies

associated with BuChE prompted us to develop QSAR models for this class of attractive enzyme inhibitors. In the present study, a combination of classical QSAR and hologram QSAR (HQSAR) has been employed to generate 2D QSAR models for a series of tacrine derivatives as inhibitors of BuChE.

2. EXPERIMENTAL SECTION

Data Set

The data set was selected from the literature (Table 1) [14-20]. The QSAR modeling analysis, calculations and visualizations were performed with SYBYL 7.0 package (Tripos Inc., St. Louis, USA), running on Red Hat Enterprise Linux workstations. The data set of 50 inhibitors was divided into training (compounds 1-40, Table 1) and test (compounds 41-50, Table 1) sets in the ratio of 4:1 (20%). A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA).

Descriptors Calculation and Selection

2D molecular descriptors such as topological descriptors, connectivity indices, 2D autocorrelation descriptors, Burden eigenvalues indices, etc were computed using the software DRAGON 5.4 and used as independent variables in QSAR modeling. A total of 929 molecular descriptors was then subjected to the following selection criteria; descriptors with constant values or found to have poor correlation to biological property were discarded ($r^2 < 0.10$). This strategy afforded 536 descriptors.

Classical QSAR Studies

BUILDQSAR software was employed to systematically search for model of up to 4 variables that give rise to

multiple linear regression (MLR) models with $r^2 > 0.70$. All descriptors present in MLR models were pooled together and used for PLS analysis in PIROUETTE software.

HQSAR Studies

Statistical HQSAR modeling was carried out as previously described [21]. In this method, each molecule in the data set is broken down into several unique structural fragments, which are arranged to form a molecular hologram. HQSAR encodes all possible molecular fragments (i.e., linear, branched, and overlapping) [21, 22]. With the transformation of the chemical representation of a molecule into its corresponding molecular hologram, this method requires no explicit 3D information for the ligands (e.g., determination of 3D structure, putative binding conformations, and molecular alignment). HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction [21]. Several combinations of fragment distinction were considered during the QSAR modeling runs. Holograms were generated using 4 distinct fragment sizes. HQSAR analysis was performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. All models generated in our studies were investigated using full cross-validated r^2 (q^2) Partial Least Squares (PLS) Leave-One-Out (LOO) method. The predictive ability of the models was assessed by their q^2 values and further, in order to evaluate the predictive ability of the best model attained, an external validation method was employed.

3. RESULTS AND DISCUSSION

The data set of 50 BuChE inhibitors used in the 2D QSAR studies was selected from the literature [14-20]. The

Table 2. Descriptors Selected for Classical 2D QSAR Model Generation

Symbol	Definition and description
BEHm8	Highest eigenvalue n. 8 of Burden matrix weighted by atomic masses
BELm2	Lowest eigenvalue n. 2 of Burden matrix weighted by atomic masses
BELe2	Lowest eigenvalue n. 2 of Burden matrix weighted by Sanderson electronegativities
BELe5	Lowest eigenvalue n. 5 of Burden matrix weighted by Sanderson electronegativities
GGI2	Topological charge index of order 2
GGI7	Topological charge index of order 7
JGI2	Mean topological charge index of order 2
JGI7	Mean topological charge index of order 7
J	Balaban distance connectivity index
Rww	Reciprocal hyper-detour index
D/Dr07	Distance/detour ring index of order 7
MWC04	Molecular walk count of order 4
MWC06	Molecular walk count of order 6
MWC08	Molecular walk count of order 8
MWC09	Molecular walk count of order 9
piPC04	Molecular multiple path count of order 4
X4	Connectivity index chi-4
X4A	Average connectivity index chi-4

chemical structures and corresponding biological properties for the complete data set is shown in Table 1. The IC_{50} values vary from 1.48 to 88,000 nM (a factor of potency of about 60 000) and were determined under the same experimental conditions. The IC_{50} values were converted to the corresponding pIC_{50} ($-\log IC_{50}$) and used as dependent variables in the QSAR investigations. From the original data set, 40 compounds (compounds 1–40, Table 1) were selected as members of the training set for model generation, and the other 10 compounds (compounds 41–50, Table 1) were held as members of the test set for external model validation. A statistical cluster analysis carried out with Tsr 3D (Accelrys, San Diego, USA) confirmed that structurally diverse molecules possessing activities of wide range were included in both training and test sets. Thus, the data set is suitable for QSAR model development.

Classical 2D QSAR studies require the calculation of molecular descriptors, such as connectivity indices, 2D autocorrelation descriptors and Burden eigenvalues, which are used as independent variables in QSAR modeling. DRAGON 5.4 software (Taletto SRL, Milan, Italy) was used to generate the descriptors for QSAR model development. The following strategy for descriptor selection was applied in our studies. Descriptors possessing constant values as well as those with poor correlation to biological property ($r^2 < 0.10$) were discarded. To further reduce the number of descriptors, BUILDQSAR software was employed to systematically search for multiple linear regression (MLR) models of up to 4 variables with correlation coefficients $r^2 > 0.70$. At this point, none of the obtained MLR models succeed to correlate the biological property within the training set with standard deviations $SD < 1.0$.

These results lead us to consider exploring the statistical principal component analysis (PCA) and partial least squares

(PLS) methods for a set of 19 descriptors (previously selected from the MLR analysis), using the PIROUETTE software (Infometrix, Washington, USA) (Table 2).

According to the PCA results, 4 principal components accounted for 93% of total variance. The first PC explains for molecular shape features, whereas the second PC was able to discriminate the BuChE inhibitors into 3 separate groups according to the relative potency. (graphic not shown). These preliminary results encourage us to use the selected descriptors for QSAR modeling studies. PLS QSAR models were created using the leave-many-out (LMO) cross-validation procedure. As it can be seen in Fig. (2), the best statistical model shows moderate correlation within the training set ($r^2 = 0.836$, $q^2 = 0.750$, with 2 components), with no substantial predictive power for the 10 test set compounds.

As a strategy to create more robust 2D QSAR models for this data set, the HQSAR method was used as previously described [21–24]. HQSAR is a modern technique that has shown predictive ability comparable to those of widely used 3D QSAR techniques [22, 24–25].

In HQSAR, each molecule in the data set is broken down into several unique structural fragments, which are arranged to form a molecular hologram. HQSAR encodes all possible 2D molecular fragments (i.e., linear, branched, and overlapping) [21,25], and then requires no explicit 3D information for the ligands. HQSAR results are strongly linked to parameters concerning hologram generation, including hologram length, fragment size, and fragment distinction [21]. The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). Several

Fig. (2). Plot of predicted versus experimental values of pIC_{50} for the 50 BuChEI (training and test sets).

Table 3. Results of HQSAR Analysis for Various Fragment Distinctions on the Key Statistical Parameters Using Fragment Size Default (4-7)

Model	Fragment Distinction	q^2	r^2	SEE	HL	N
1	A/B	0.695	0.862	0.490	71	5
2	A/B/C	0.723	0.928	0.318	61	5
3	A/B/C/H	0.653	0.866	0.482	307	5
4	A/B/C/H/Ch	0.623	0.716	0.683	71	3
5	A/B/C/H/Ch/DA	0.714	0.927	0.361	61	6
6	A/B/H	0.581	0.888	0.447	257	6
7	A/B/C/Ch	0.663	0.885	0.453	53	6
8	A/B/DA	0.624	0.860	0.486	199	4
9	A/B/C/DA	0.577	0.889	0.424	83	6
10	A/B/H/DA	0.582	0.724	0.673	353	3
11	A/B/C/H/DA	0.649	0.893	0.437	61	6
12	A/B/H/Ch/DA	0.557	0.870	0.474	401	5

q^2 , cross-validated correlation coefficient; SEP, cross-validated standard error; r^2 , noncross-validated correlation coefficient; SEE, noncross-validated standard error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

combinations of these parameters were considered using the fragment size default (4–7), as follows: 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/C/DA, A/B/H/DA, A/B/C/H/DA, and A/B/H/Ch/DA. HQSAR analysis was performed over the 12 default series of hologram lengths values ranging from 53 to 401 bins. The patterns of fragment counts from the 40 training set inhibitors were then related to the experimental biological data using PLS analysis. The statistical results are summarized in Table 3.

As shown in Table 3, fragment distinction parameters have considerable effects on the quality of the models as expected. The best statistical result among all models was obtained for model 2 ($r^2 = 0.928$, $q^2 = 0.723$, with 5 components), derived using A/B/C as fragment distinction. The influence of different fragment sizes, which control the minimum and maximum length of fragments to be included in the hologram, was further investigated for the best HQSAR model as depicted in Table 4. As it can be seen, the variation of fragment size did not result in any improvements in the statistical parameters.

It is important to note that high q^2 values do not imply automatically that the models would possess high predictive

ability for external compounds [21,27], as noted in the classical QSAR studies. The most valuable test of a QSAR model is its ability to predict the property value for new structurally related compounds [21,27]. The predictive power of the HQSAR model 2 was assessed by predicting pIC_{50} values for 10 test set molecules (compounds 41–50, Table 1), which were completely excluded during the training of the model. The results of the external validation are listed in Table 5, and the graphic results for the experimental versus predicted activities of both training set and test set are displayed in Fig. (3). The good agreement between experimental and predicted values for the test set compounds indicates the reliability of the HQSAR model ($r^2 = 0.928$, and $\text{SEE} = 0.318$). The predicted values fall close to the experimental pIC_{50} values, deviating by less than 0.8 log units, without the presence of any outliers (Fig. (3)).

4. CONCLUSION

The creation of predictive QSAR models is a key point in modern drug design. The use of classical and HQSAR approaches for the same data set was a valuable tool to the better understanding of the limitations and possibilities for the creation of predictive models exploring different 2D

Table 4. Influence of Different Fragment Sizes on the Statistical Parameters of the HQSAR Models

Fragment Distinction	Fragment Size	q^2	r^2	SEE	HL	N
A/B/C	2-5	0.585	0.798	0.584	59	4
	3-6	0.628	0.822	0.548	61	4
	4-7	0.723	0.928	0.318	61	5
	5-8	0.641	0.879	0.466	53	6

Table 5. Experimental and Predicted Activities (pIC_{50}) with Residual Values for the Test Set Compounds

Model	Test Set	pIC_{50}		
	Compounds	Experimental	Predicted	Residual
A/B/C	41	7.28	6.50	0.78
	42	5.15	5.10	0.05
	43	5.51	5.30	0.21
	44	7.49	7.04	0.45
	45	4.72	5.36	-0.64
	46	6.12	6.63	-0.51
	47	4.38	5.06	-0.68
	48	6.27	6.66	-0.39
	49	7.92	7.97	-0.05
	50	7.97	7.45	0.52

Fig. (3). Plot of predicted versus experimental values of pIC_{50} for the 50 BuChEI (training and test sets).

QSAR strategies. The final HQSAR model should be useful for the design of novel BuChE inhibitors having improved potency and selectivity within this structural class.

5. ACKNOWLEDGEMENTS

We gratefully acknowledge financial support from FAPESP (The State of São Paulo Research Foundation) and FAPESB (The state of Bahia Research Foundation), Brazil.

REFERENCES

- [1] Dorronsoro, I.; Alonso, D.; Castro, A.; Del Monte, M.; Garica-Palomero, W.; Martínez, A. *Arch. Pharm.* **2005**, *338*, 18.
- [2] Rosini, M.; Andrisano, V.; Bartolini, M.; Bolognesi, M.L.; Hrelia, P.; Minarini, A.; Tarozzi, A.; Melchiorre, C. *J. Med. Chem.* **2005**, *48*, 360.
- [3] Orhan, G.; Orhan, I.; Sener, B. *Lett. Drug Des. Discov.* **2006**, *3*, 268.
- [4] Liston, D.R.; Nielsen, J.A.; Villalobos, A.; Chapin, D.; Jones, S.B.; Hubbard, S.T.; Shalaby, I. A.; Ramirez, A.; Nason, D.; White, W. F. *Eur. J. Pharmacol.* **2004**, *486*, 9.

- [5] Nicolet, Y.; Lockridge, O.; Masson, P.; Fontecilla-Camps, J.C.; Nachon, F. *J. Biol. Chem.* **2003**, 278, 41141.
- [6] Gauthier, S.; Emre, M.; Farlow, M.R.; Bullock, R.; Grossberg, G.T.; Potkin, S.G.; *Curr. Med. Res. Opin.* **2003**, 19, 707.
- [7] Giacobini, E. *Pharmacol. Res.* **2004**, 50, 433.
- [8] Kuhl, D.E.; Koeppe, R.A.; Snyder, S.E.; Minoshima, S.; Frey, K.A.; Killbourn, M.R.; *Ann. Neurol.* **2005**, 59, 13.
- [9] Recanatini, M.; Cavalli, A.; Belluti, F.; Piazzi, L.; Rampa, A.; Bisi, A.; Gobbi, S.; Valeti, P.; Andrisano V.; Bartolini, M.; Cavrini, V. *J. Med. Chem.* **2000**, 43, 2007.
- [10] Sippl, W.; Contreras, J.M.; Parrot, I.; Rival, Y.M.; Wermuth, C.G.; *J. Comput. Aided Mol. Des.* **2001**, 15, 395.
- [11] El-Yazal, J.; Rao, S.N.; Mehl, A.; Schlikker-Jr, W. *Toxicol. Sci.* **2001**, 63, 223.
- [12] Zaheer-ul-Haq, Wellenzohn, B.; Tonmunphean, S.; Khalid, A.; Choudhary, M.I.; Rode, B.M. *Bioorg. Med. Chem. Lett.* **2003**, 13, 4375.
- [13] Chandrasekaran, V.; McGaughey, G.B.; Cavallite, C.J.; Bowen, J.P. *J. Mol. Graph. Model.* **2004**, 23, 69.
- [14] McKenna, M.T.; Proctor, G.R.; Young, L.C.; Harvey, A.L. *J. Med. Chem.* **1997**, 40, 3516.
- [15] Camps, P.; El-Achab, R.; Görbig, D.M.; Morral, J.; Muñoz-Torrero, D.; Badia, ; Baños, J.E.; Vivas, N.M.; Barril, X.; Orozco, M.; Luque, F.J. *J. Med. Chem.* **1999**, 42, 3227.
- [16] Camps, P.; El-Achab, R.; Morral, J.; Muñoz-Torrero, D.; Badia, A.; Baños, J.E.; Vivas, N.M.; Barril, X.; Orozco, M.; Luque, F.J. *J. Med. Chem.* **2000**, 43, 4657.
- [17] Rampa, A.; Bisi, A.; Belluti, F.; Gobbi, S.; Valenti, P.; Andrisano, V.; Cavrini, V.; Cavalli, A.; Recantini, M. *Bioorg. Med. Chem.* **2000**, 8, 497.
- [18] Marco, J.L.; Los-Rios, C.; Carreiras, M.C.; Baños, J.E.; Badía, A.; Vivas, N.M. *Bioorg. Med. Chem.* **2001**, 9, 727.
- [19] Camps, P.; Gómez, E.; Muñoz-Torrero, D.; Badia, A.; Vivas, N.M.; Barril, X.; Orozco, M.; Luque, F.J. *J. Med. Chem.* **2001**, 44, 4733.
- [20] Camps, P.; Formosa, X.; Muñoz-Torrero, D.; Petrignet, J.; Badia, A.; Clos, M.V. *J. Med. Chem.* **2005**, 48, 1701.
- [21] Honorio, K.M.; Garratt, R.C.; Andricopulo, A.D. *Bioorg. Med. Chem. Lett.* **2005**, 15, 3119-3125.
- [22] Castilho, M.S.; Postigo, M.P.; de Paula, C.B.; Montanari, C.A.; Oliva, G.; Andricopulo, A.D. *Bioorg. Med. Chem.* **2006**, 14, 516.
- [23] Honorio, K.M.; Garratt, R.C.; Polikarpov, I.; Andricopulo, A.D. *Lett. Drug Des. Discov.* **2006**, 3, 261.
- [24] Honorio, K.M.; Garratt R.C.; Polikarpov, I.; Andricopulo A.D.; *J. Mol. Graph. Model.* **2006**, in press.
- [25] Seel, M.; Turner, D.B.; Willet, P. *Quant. Struct-Act. Relat.* **1999**, 18, 245.
- [26] Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, 20, 269-276.
- [27] Lowis, D.R.; *Tripes Technical Notes*, **1997**, 1, 1.

2D Quantitative structure–activity relationship studies on a series of cholesteryl ester transfer protein inhibitors

Marcelo S. Castilho,^{a,*} Rafael V. C. Guido^b and Adriano D. Andricopulo^b

^aLaboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-290 Salvador, BA, Brazil

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

Received 11 May 2007; revised 6 June 2007; accepted 8 June 2007
Available online 13 June 2007

APÊNDICE E

Abstract—Coronary heart disease (CHD) is one of the major causes of human death. The most successful therapeutic approach available is based on the reduction of low density-lipoprotein cholesterol (LDL-C). However, it is believed that the next paradigm in CHD treatment will rely on increased HDL-C levels. One of the most promising strategies for this goal is the inhibition of cholesteryl ester transfer protein (CETP). In the present work, robust classical 2D QSAR ($r^2 = 0.76$, $q^2 = 0.72$) and hologram QSAR ($r^2 = 0.88$, $q^2 = 0.70$) models were developed for a series of 85 CETP inhibitors (*N*-*N*-disubstituted trifluoro-3-amino-2-propanol derivatives). These models are complementary in nature and highlight important structural features for the design of novel CETP inhibitors with improved potency.

© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Although coronary heart disease (CHD) mortality has been diminishing in Western Europe and North America for the past decades, it remains one of the major causes of human death.^{1,2} In the past two decades, the most successful therapeutic approach for the treatment of this disease is based on the reduction of low density-lipoprotein cholesterol (LDL-C) levels.³ Accordingly, statins have become the gold standard treatment for patients with, or at risk for, CHD.⁴ However, despite the great benefits of statins, the outcome of many treated patients is still unsatisfactory, requiring immediate attention.⁵ Furthermore, epidemiologic studies have identified that low levels of high density-lipoprotein cholesterol (HDL-C) are a higher risk for CHD than LDL-C, total cholesterol or plasma triglycerides (TG).⁶ Therefore, it is believed that the next frontier in CHD treatment should rely on increased HDL-C levels.⁷

Nevertheless, the available drugs for CHD management have poor effects on HDL-C levels (Fig. 1): niacin has by far the best profile, raising the level of HDL-C by as much as 30%, however, its cutaneous flushing side-effect limits the patient compliance⁸; Fibrates (e.g., gemfibrozil, fenofibrate) increase HDL-C by 15–25% in patients with hypertriglyceridemia, but have little effect (<10%) on other patients; statins (e.g., atorvastatin, simvastatin) produce only modest effect on HDL-C levels (approximately 5%).^{9–11} This fact highlights the importance of probing more effective strategies to increase HDL-C levels.^{5,6,8} This goal can be achieved through the inhibition of cholesteryl ester transfer protein (CETP). However, recent clinical trials moderated the perspectives on CETP inhibition as a good target for CHD patients. Although torcetrapib, an irreversible inhibitor of CETP, in association with atorvastatin substantially increased HDL-C and decreased LDL-C when compared with atorvastatin monotherapy, no significant reduction in the progression of coronary atherosclerosis was observed.¹² Although the use of torcetrapib was associated with increased blood pressure, there are a number of CETP inhibitors that do not show similar effects. Therefore, there is an increasing interest in the study of new classes of CETP inhibitors aimed at developing therapies to raise HDL cholesterol levels.¹²

Keywords: QSAR; Coronary heart disease; Trifluoro-3-amino-2-propanol derivatives; CETP; Inhibitors.

* Corresponding author. Tel.: +55 71 3332 1580; fax: +55 71 3235 9350; e-mail: castilho@ufba.br

Figure 1. Drugs currently used for CHD management.

CETP is a glycoprotein that binds to HDL and is involved in the transfer of lipoprotein particles and neutral lipids, including cholesteryl ester, phospholipids, and triglyceride. Accordingly, CETP crystallographic structure revealed a noteworthy size and highly hydrophobic binding site capable of simultaneously binding four lipid molecules.¹³ Among the several inhibitors described in the literature,⁸ *N,N*-disubstituted trifluoro-3-amino-2-propanol derivatives pose themselves as interesting small-molecule targets for drug design due to their synthetic accessibility and high potency. The first crystallographic structure of CETP was made available very recently,¹³ thus allowing new possibilities for QSAR studies associated with structure-based drug design strategies. This was not possible previously, leading to the appearance of QSAR studies employing 3D alignments based on spatial conformations of minimum energy.¹⁴ In the present study, a data set of *N,N*-disubstituted trifluoro-3-amino-2-propanol derivatives, considered as CETP inhibitors, was used to develop 2D QSAR models employing easily derived topological descriptors and hologram QSAR (HQSAR), which afforded reliable and predictive QSAR models.

2. Results and discussion

The data set of 85 *N,N*-disubstituted trifluoro-3-amino-2-propanol derivatives used for the QSAR analyses was selected from the literature.^{15–18} Chemical structures and biological properties for the complete set of compounds (divided in the corresponding training and test sets) are listed in Table 1. The IC_{50} values employed in this work (varying from 20 to 65,000 nM), measured under the same experimental conditions, are acceptably distributed across the range of values (Fig. 2). Thus, the data set is appropriate for the purposes of QSAR model development. The IC_{50} values were converted to the corresponding pIC_{50} ($-\log IC_{50}$) and used as dependent variables in the QSAR investigations.

From the original data set of 85 CETP inhibitors, 68 compounds (compounds 1–68, Table 1) were selected as members of the training set for QSAR model development, and the remaining 17 compounds (compounds 69–85, Table 1) were considered as members of the test set for external validation. A statistical cluster analysis carried out with PIROUETTE confirmed that structurally diverse molecules possessing activities of wide range were included in both training and test sets (data not shown).

Classical 2D QSAR studies require the calculation of molecular descriptors, such as connectivity indices, 2D autocorrelation descriptors, and Burden eigenvalues, which are used as independent variables in QSAR modeling. The DRAGON 5.4 software was used to generate the descriptors for the QSAR studies. This procedure afforded 929 descriptors which were subjected to the following selection strategy. Descriptors possessing constant values as well as those with poor correlation to biological property ($r^2 < 0.10$) or that are more than 0.99 correlated were discarded. To further reduce the number of descriptors, BuildQSAR software was employed to systematically search for multiple linear regression (MLR) models of up to four variables with correlation coefficients $r^2 > 0.70$.

The best model shows moderate statistical values ($r^2 = 0.73$, $q^2 = 0.69$, $s = 0.50$, $F = 43.60$, SPRESS = 0.54) but lacked predictive ability when challenged against test set compounds (data not shown). This scenario could be the result of an incomplete description of the binding event by the selected descriptors or an optimistic evaluation of leave-one-out cross-validation procedure. At this point, none of the obtained MLR models succeed to correlate the biological property for the training set compounds within 1.0 log unit errors. These results lead us to consider exploring the statistical principal component analysis (PCA) and partial least squares (PLS) methods. For that purpose, all

Table 1. Chemical structures and corresponding IC₅₀ values for a series of CETP inhibitors

Training set

Compound	R ₁	IC ₅₀ (μM)	Compound	R ₁	IC ₅₀ (μM)	Compound	R ₁	IC ₅₀ (μM)	Compound	R ₁	IC ₅₀ (μM)
1		0.14	10		0.20	19		0.10	28		0.40
2		0.02	11		0.20	20		0.50	29		2.60
3		0.03	12		0.30	21		1.30	30		0.40
4		0.03	13		0.08	22		0.06	31		0.09
5		0.40	14		0.10	23		0.10	32		0.09
6		1.00	15		0.08	24		0.20	33		0.20
7		2.40	16		0.08	25		0.30	34		0.20
8		3.50	17		0.10	26		0.30	35		0.20
9		0.20	18		0.08	27		0.30			

Training set

Compound	R ₂	IC ₅₀ (μM)	Compound	R ₂	IC ₅₀ (μM)	Compound	R ₂	IC ₅₀ (μM)	Compound	R ₂	IC ₅₀ (μM)
36		0.48	41		0.30	46		1.00	50		5.20
37		0.66	42		22.90	47		15.00	51		0.42
38		4.08	43		6.72	48		6.79	52		16.00
39		0.91	44		1.12	49		4.42	53		14.30
40	OCH ₂ CH ₃	1.60	45	OCF ₃	1.00						

(continued on next page)

Table 1 (continued)

Training set			
Compound Structure	IC ₅₀ (μM)	Compound Structure	IC ₅₀ (μM)
54	40.0	59	12.00
55	15.0	60	10.00
56	25.0	61	10.00
57	15.0	62	65.00
58	50.0	63	0.72
		64	15.0
		65	1.50
		66	12.0
		67	20.0
		68	45.8

Test Set

Compound	R ₁	IC ₅₀ (μM)	Compound	R ₁	IC ₅₀ (μM)
69		0.70	75		0.30
70		2.50	76		0.10
71		0.40	77		0.08
72		0.09	78		7.80
73		0.20	79		6.24
74		0.06			

Compound	Structure	IC ₅₀ (μM)
80		1.30
81		1.00
82		0.07
83		12
84		3.60
85		3.96

descriptors from the MLR models with $r^2 > 0.70$ (Table 2) were pooled together, autoscaled, and employed for PCA and PLS analysis within the PIROUETTE software.

According to the PCA results, four principal components accounted for 85% of total variance, whereas further components contribute with less than 5%. The first PC explains molecular shape features, whereas the relative potency varies according to both first and second PC. Less bulk compounds have negative values in PC1, while bulkier compounds have positive values. The most potent compounds are grouped in the far right-hand side of PC1 and have negative values in PC2. Thus, training set compounds can be broadly separated into three groups. These preliminary results prove that selected descriptors have good discriminating power and can be used for QSAR modeling studies.

In addition, PLS QSAR models were created using the leave-many-out (LMO) cross-validation procedure. As seen in Figure 3, the best statistical model shows moderate correlation within the training set ($r^2 = 0.76$, $q^2 = 0.72$, with two components).

The predictive power of the best QSAR model derived using the 68 training set molecules was assessed by predicting pIC_{50} values for 17 test set compounds (69–85, Table 1), not used for QSAR model development. The external validation process can be considered the most reliable validation method, as cross-validation procedures may lead to very optimistic statistics.^{19–21} The results of the external validation are listed in Table 3, and the graphic results for the experimental versus predicted activities of both training set and test set are displayed in Figure 3.

Besides demonstrating statistical significance, QSAR models should also provide useful chemical insights for drug design. For this reason, an acceptable interpretation of the QSAR results is provided below. There is some criticism about using topological descriptors in QSAR modeling, as some of them lack adequate physi-

Figure 2. Distribution of the pIC_{50} values for the data set compounds.

Table 2. Descriptors selected for QSAR model development

Symbol	Definition and description
SPI	Superpendentic index
S1K	1-Path Kier alpha-modified shape index
D/Dr06	Distance/detour ring index of order 6
T(N...O)	Sum of topological distance between N...O
MATS3p	Moran autocorrelation—lag3/weighted by atomic polarizabilities
EEig04r	Eigenvalue 04 from edge adjacency matrix weighted by resonance integrals
MATS3v	Moran autocorrelation—lag3/weighted by atomic van der Waals volumes
BELe4	Lowest eigenvalue N.4 of Burden matrix/weighted by atomic Sanderson electronegativities
GGI10	Topological charge index of order 10
VEA1	Eigenvector coefficient sum from adjacency matrix
BEHm5	Highest eigenvalue N.5 of Burden matrix/weighted by atomic masses
QZZe	Qzz COMMA2 value/weighted by Sanderson electronegativities
RDF060v	Radial distribution function —6.0/weighted by atomic van der Waals volumes
RDF065p	Radial distribution function —6.5/weighted by atomic polarizabilities
SPAN	Span R
RDF065v	Radial distribution function —6.5/weighted by atomic van der Waals volumes
RDF060p	Radial distribution function —6.0/weighted by atomic polarizabilities
RDF030m	Radial distribution function —3.0/weighted by atomic masses
RDF060e	Radial distribution function —6.0/weighted by atomic Sanderson electronegativities

Figure 3. Predicted versus observed values of pIC_{50} values (filled diamonds, training set compounds; open squares, test set compounds).

cal interpretation. However, traditional descriptors (e.g., $\log P$ and pK_a) have also their own limitations, because some may be more correlated with pharmacokinetic properties, whereas others may be more related to affinity. Besides, traditional descriptors often are whole-molecule descriptors that convey little information on particular structural differences relevant for enzyme affinity and inhibitory potency. Nevertheless, it would be cumbersome to explain the influence of all topological descriptors over CETP inhibitors, potency. Instead,

Table 3. Classical 2D QSAR experimental and predicted activities (pIC_{50}) with residual values for the 17 test set compounds

Compound	pIC_{50}		
	Experimental	Predicted	Residual ^a
69	6.15	6.44	−0.28
70	5.60	6.54	−0.94
71	6.40	6.61	−0.21
72	7.05	6.86	0.18
73	6.70	6.83	−0.13
74	7.22	6.64	0.58
75	6.52	6.32	0.21
76	7.00	6.53	0.47
77	7.10	6.71	0.39
78	5.11	5.53	−0.42
79	5.20	5.16	0.05
80	5.89	6.45	−0.56
81	6.00	5.84	0.16
82	7.15	6.91	0.24
83	4.92	5.24	−0.32
84	5.44	5.02	0.43
85	5.40	5.75	−0.35

^a The difference between experimental and predicted values.

Figure 4. Regression vectors to PC1.

we decided to focus on the three most significant ones to PC1 (50.38% of variance), which are SPI, BELe4 and RDF030m (Fig. 4).

- BELe4^{22,23} is a low-dimensional metric, derived from the Burden matrix, which accounts for electrostatic interactions. Analysis of BELe4 values reveals a bell-shaped correlation to potency, weak inhibitors lie in the left (BELe4 value < 1.20) and right (BELe4 value > 1.50) tails and strong potent inhibitors are placed in the center (BELe4 value = 1.46). This result suggests that CETP inhibition depends on a delicate charge balance for interaction with key residues in the CETP active site.
- SPI²⁴ is a topological descriptor derived from a submatrix of distance matrix that accounts for shape features in CETP inhibitors. This descriptor is linearly correlated to potency and indicates that inhibitors that fit in the active site have a better inhibitory profile. However, some compounds have already exceeded the optimal volume (compounds **64**, **67**, **104** (SPI value > 30) vs compound **2** (SPI value = 27.10)).

- RDF030m²⁵ is a 2D descriptor calculated from radial distribution function code and represents the probability distribution to find an atom in a spherical volume of radius r . Thus, this descriptor also accounts for steric and shape aspects of CETP inhibitors. As a consequence, it is correlated to SPI ($r^2 = 0.59$) and behaves accordingly.

Although the classical 2D QSAR model provided some useful information and showed a good predictive ability, topological descriptors used for model development convey little information on which moieties are particularly important to CETP inhibition. In order to shed some light on this subject, we resorted to a modern approach named hologram QSAR (HQSAR), a 2D QSAR method that has shown predictive ability comparable to those of more sophisticated 3D QSAR techniques.^{26–29} HQSAR generates specialized molecular holograms that incorporate information about each 2D fragment (i.e., linear, branched, and overlapping), and each of its constituent subfragments, implicitly encoding 3D structural information that is important for binding affinity.^{26–29}

HQSAR results are strongly linked to parameters concerning hologram generation, including hologram length, fragment size, and fragment distinction.²⁶ The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). Several combinations of these parameters were considered using the fragment size default (4–7), as follows: 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/C/DA, A/B/H/DA, A/B/C/H/DA, and A/B/H/Ch/DA. HQSAR analysis was performed over the 12 default series of hologram length values ranging from 53 to 401 bins. The patterns of fragment counts from the 68 training set inhibitors were then related to the experimental biological data using PLS analysis. The statistical results are summarized in Table 4.

As shown in Table 4, fragment distinction parameters have considerable effects on the quality of the models as expected. For instance, models 8 and 10 were derived using A/B/DA and A/B/H/DA, respectively. Accordingly, the inclusion of hydrogen as distinction parameters was detrimental for biological activity prediction of excluded molecules during the internal validation procedure (LOO) as indicated by the decrease of ~5% in q^2 value. This result is in good agreement with previous results^{26,28,30} which suggest that both options should not be used together due to the substantial increase in the number of fragments generated when both fragment distinctions are considered in the model construction.

The best statistical result among all models (Table 4) was obtained for model 8 ($r^2 = 0.87$, $q^2 = 0.70$, with five components), derived using A/B/DA as fragment distinction. The influence of different fragment sizes, which control the minimum and maximum length of fragments to be included in the hologram, was further investigated for the best HQSAR model as depicted in Table 5.

Table 4. Results of HQSAR analyses for various fragment distinctions on the statistical parameters using fragment size default (4–7)

Model	Fragment distinction	q^2	r^2	SEE	HL	N
1	A/B	0.64	0.87	0.35	257	5
2	A/B/C	0.57	0.87	0.33	401	5
3	A/B/C/H	0.59	0.83	0.38	401	5
4	A/B/C/H/Ch	0.62	0.83	0.39	151	5
5	A/B/C/H/Ch/DA	0.67	0.88	0.33	307	5
6	A/B/H/	0.59	0.85	0.37	353	5
7	A/B/C/Ch	0.58	0.86	0.35	151	5
8	A/B/DA	0.70	0.87	0.33	257	5
9	A/B/C/DA	0.67	0.90	0.30	401	5
10	A/B/H/DA	0.65	0.89	0.32	257	5
11	A/B/C/H/DA	0.63	0.81	0.41	61	5
12	A/B/H/Ch/DA	0.67	0.89	0.32	307	5

Table 5. HQSAR analyses for the influence of various fragment sizes on the key statistical parameters

Model	Fragment size	q^2	r^2	SEE	HL	N
13	2–5	0.62	0.83	0.39	199	5
14	3–6	0.66	0.86	0.35	257	5
15	4–7	0.70	0.87	0.33	257	5
16	5–8	0.69	0.80	0.42	59	4
17	6–9	0.70	0.88	0.32	353	4
18	7–10	0.70	0.87	0.34	353	4

This analysis revealed that, fragment size variation did not improve the statistical parameters. However, slightly simpler models were generated when fragment sizes of 6–9 ($r^2 = 0.88$, $q^2 = 0.70$ and $SEE = 0.32$) and 7–10 ($r^2 = 0.87$, $q^2 = 0.70$, and $SEE = 0.34$) were investigated. In order to select the best and more robust model, an external validation procedure was carried out. Accordingly, the predictive ability of models 17 and 18 was assessed by predicting pIC_{50} values for the same test set molecules previously used in classical 2D QSAR studies (compounds **69–85**, Table 1). Model 17 presented predictive- r^2 value of 0.71, whilst model 18 presented predictive- r^2 value of 0.61 (data not shown). The results of both training and test set for the best generated model (model 17) are displayed in Figure 5 and Table 6. The good agreement between experimental and predicted values for the test set compounds indicates the reliability of the HQSAR model ($r^2 = 0.88$

Figure 5. Predicted versus experimental values of pIC_{50} for the training and test of CETP inhibitors obtained by the best HQSAR model. Filled diamonds, training set compounds; open squares, test set compounds.**Table 6.** HQSAR experimental and predicted activities (pIC_{50}) with residual values for the 17 test set compounds

Compound	pIC_{50}		
	Experimental	Predicted	Residual ^a
69	6.15	6.11	0.04
70	5.60	6.40	−0.80
71	6.40	6.87	−0.47
72	7.05	6.93	0.12
73	6.70	7.00	−0.30
74	7.22	6.99	0.23
75	6.52	6.28	0.24
76	7.00	6.59	0.41
77	7.10	6.76	0.34
78	5.11	5.88	−0.77
79	5.20	5.50	−0.30
80	5.89	5.56	0.33
81	6.00	5.73	0.27
82	7.15	7.21	−0.06
83	4.92	4.94	−0.02
84	5.44	4.64	0.80
85	5.40	5.82	−0.42

^a The difference between experimental and predicted values.

and $SEE = 0.32$). The predicted values fall close to the experimental pIC_{50} values, within 0.80 log units. This result is slightly better than those obtained for the classical 2D QSAR model, which fails to predict compound **70** by almost a log unit.

Besides improved statistical robustness, the HQSAR model is able to provide useful insights into the relationship between structural fragments and biological activity, which can be visualized through contribution maps. For example, analysis of the HQSAR contribution maps (Fig. 6) for compounds **56** ($IC_{50} = 25.0 \mu M$) and **45** ($IC_{50} = 1.0 \mu M$) clearly revealed different fragment contributions to biological activity.

Figure 6. Contribution maps of compounds **56** (left panel) and **45** (right panel). The color scheme represents fragment contribution to CETP inhibitors, potency.

This result shows structural features that are related to activity, but does not provide a straightforward physico-chemical explanation for the differences in potency of these compounds, for instance, the different color scheme might be a consequence of steric or electrostatic contribution or even both. However, when classical QSAR descriptors are considered as well, it can be seen that both molecules have optimal electrostatic properties (BELe4 value = 1.46), but different steric properties, as observed for their RDF30m values (compound **45** = 8.23 and compound **56** = 12.1). Accordingly, the data suggest that the lower inhibitory activity of compound **56** is possibly associated to steric restraints within the active site. This confirms previous results from Durlay and colleagues^{15,17} indicating that the *para* substitution pattern of the benzamine moiety is detrimental to activity. In addition, our QSAR models afforded a quantitative stand point of view on this subject and underscored the facts that due to sterical restraints compound **56** is prevented from binding properly in CETP active site.

3. Conclusion

The use of both molecular holograms and topological descriptors, along with the standard and reproducible high quality biological data, allowed the generation of robust QSAR models for this series of CETP inhibitors. A synergic approach, using information from the HQSAR contribution maps and also from a careful interpretation of the topological descriptors from classical 2D QSAR, shed some light on the effect of the substitution pattern over fragments that should be sterically constrained. Therefore, the QSAR models described herein are complementary in nature and shall be useful for the design of structurally related CETP inhibitors with improved potency.

4. Materials and methods

4.1. Data set

The data set of 85 CETP inhibitors used in the QSAR analyses was collected from the literature.^{15–18} A statis-

tical cluster analysis was carried out with PIROUETTE software (Infometrix, Washington, USA). Accordingly, the data set was split into training (compounds **1–68**, Table 1) and test (compounds **69–85**, Table 1) sets in the ratio of 5:1 (20%). The test set of 17 compounds was selected in such a way to provide an appropriate representation of the training set in terms of structural composition and activity range. The pIC₅₀ values were treated as dependent variables during the Partial Least Squares (PLS) analyses.

4.2. Classical QSAR studies

The 2D molecular descriptors, such as topological descriptors, connectivity indices, 2D autocorrelation descriptors, Burden eigenvalue indices, and so on, were computed using the software DRAGON 5.4 (Taletto SRL, Milan, Italy) and used as independent variables in the classical QSAR studies. A total of 929 molecular descriptors was then subjected to the following selection criteria; descriptors with constant values or found to have poor correlation to biological property were discarded ($r^2 < 0.10$). This strategy afforded 536 descriptors. BuildQSAR¹⁹ software was employed to systematically search for models of up to four variables that give rise to multiple linear regression (MLR) models with $r^2 > 0.70$. All descriptors present in MLR models were pooled together, autoscaled, and used for PLS analysis in PIROUETTE software.

4.3. HQSAR modeling

The HQSAR modeling analyses, calculations, and visualizations were performed with SYBYL 7.2 package (Tripos Inc., St. Louis, USA), running on Red Hat Enterprise Linux workstations. Several combinations of fragment distinction were considered during the QSAR modeling runs. Holograms were generated using six distinct fragment sizes over the 12 default series of hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins). The molecular holograms generated were used as independent variable during the PLS regression analyses to derive the HQSAR models.

4.4. Statistical analyses

The several statistical models generated in our studies were investigated using the PLS leave-many-out (LMO) method. The predictive ability of the models was assessed by their full cross-validated r^2 (q^2) values. Internal validation methods leave-one-out and LMO (with 10 groups of compounds) were used to determine the optimum number of PLS components and the stability of the models. The number of components used in the final nonvalidated model was optimized to give the highest q^2 value and the lowest standard error of prediction. The noncross-validated models were assessed by the conventional correlation coefficient r^2 and F -values. External validation was performed with a test set of 17 compounds, which were not included in the training set during the process of QSAR model generation.

References and notes

- Abildstrom, S. Z.; Rasmussen, S.; Rosén, M.; Madsen, M. *Heart* **2003**, 89, 507.
- Jemal, A.; Ward, E.; Hao, Y.; Thun, M. *JAMA* **2005**, 294, 1255–1259.
- Ong, H. T. *Q. J. Med.* **2005**, 98, 599.
- Shah, P. K. *Eur. Heart J.* **2007**, 28, 5.
- Shinkai, H. *Mini Rev. Med. Chem.* **2002**, 2, 271.
- Barter, P. J.; Kastelein, J. J. P. *J. Am. Coll. Cardiol.* **2006**, 47, 492.
- Milani, R. V.; Lavie, C. J. O. *J. Am. Coll. Cardiol.* **2006**, 48, 1791.
- Sikorski, J. A. *J. Med. Chem.* **2006**, 49, 2.
- Milani, R. V.; Lavie, C. J. In *Cardiovascular Drug Therapy*; Messerli, F. H., Ed.; Saunders: Philadelphia, PA, 1996; pp 1098–1110.
- Jones, P. H.; Davidson, M. H.; Stein, E. A.; Bays, H. E.; McKenney, J. M.; Miller, E.; Cain, V. A.; Blasetto, J. W. *Am. J. Cardiol.* **2003**, 92, 152.
- Deedwania, P. C.; Hunnigdale, D. B.; Bays, H. E.; Jones, P. H.; Cain, V. A.; Blasetto, J. W. *Am. J. Cardiol.* **2005**, 95, 360.
- Nissen, S. E.; Tardif, J. C.; Nicholls, S. J.; Revkin, J. H.; Shear, C. L.; Duggan, W. T.; Ruzyllo, W.; Bachinsky, W. B.; Lasala, G. P.; Tuzcu, E. M. *ILLUSTRATE Investigators. N. Engl. J. Med.* **2007**, 356, 1304.
- Qiu, X.; Mistry, A.; Ammirati, M. J.; Chrnyk, B. A.; Clark, R. W.; Cong, Y.; Culp, J. S.; Danley, D. E.; Freeman, T. B.; Geoghegan, K. F.; Griffor, M. C.; Hawrylik, S. J.; Hayward, C. M.; Hensley, P.; Hoth, L. R.; Karam, G. A.; Lira, M. E.; Lloyd, D. B.; McGrath, K. M.; Stutzman-Engwall, K. J.; Subashi, A. K.; Subashi, T. A.; Thompson, J. F.; Wang, I. K.; Zhao, H.; Seddon, A. P. *Nat. Struct. Mol. Biol.* **2007**, 14, 106.
- Kelkar, M. A.; Pednekar, D. V.; Pimple, S. R.; Akamanchi, K. G. *Med. Chem. Res.* **2004**, 13, 590.
- Durley, R. C.; Grapperhaus, M. L.; Massa, M. A.; Mischke, D. A.; Parnas, B. L.; Fobian, Y. M.; Rath, N. P.; Honda, D. D.; Zeng, M.; Connolly, D. T.; Heuvelman, D. M.; Witherbee, B. J.; Glenn, K. C.; Krul, E. S.; Smith, M. E.; Sikorski, J. A. *J. Med. Chem.* **2000**, 43, 4576.
- Massa, M. A.; Spangler, D. P.; Durley, R. C.; Hickory, B. S.; Connolly, D. T.; Witherbee, B. J.; Smith, M. E.; Sikorski, J. A. *Bioorganic. Med. Chem. Lett.* **2001**, 11, 1625.
- Durley, R. C.; Grapperhaus, M. L.; Hickory, B. S.; Massa, M. A.; Wang, J. L.; Spangler, D. P.; Mischke, D. A.; Parnas, B. L.; Fobian, Y. M.; Rath, N. P.; Honda, D. D.; Zeng, M.; Connolly, D. T.; Heuvelman, D. M.; Witherbee, B. J.; Melton, M. A.; Glenn, K. C.; Krul, E. S.; Smith, M. E.; Sikorski, J. A. *J. Med. Chem.* **2002**, 45, 3891.
- Reinhard, E. J.; Wang, J. L.; Durley, R. C.; Fobian, Y. M.; Grapperhaus, M. L.; Hickory, B. S.; Massa, M. A.; Norton, M. B.; Promo, M. A.; Tollefson, M. B.; Vernier, W. F.; Connolly, D. T.; Witherbee, B. J.; Melton, M. A.; Regina, K. J.; Smith, M. E.; Sikorski, J. A. *J. Med. Chem.* **2003**, 46, 2152.
- Oliveira, D. B.; Gaudio, A. C. *Quant. Struct.–Activ. Relat.* **2003**, 19, 599.
- Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, 20, 269.
- Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O. *J. Comput. Aided Mol. Des.* **2006**, 20, 83.
- Pearlman, R. S.; Smith, K. M. In *3D-QSAR in drug design*, Kubinyi, H., Ed.; Kluwer/Escom, 1998; Vol. 2, pp 339–353.
- Pearlman, R. S.; Smith, K. M. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 28.
- Gupta, S.; Singh, M.; Madan, A. K. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 272.
- Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vibrat. Spectrosc.* **1999**, 19, 151.
- HQSAR™ Manual, SYBYL 6.9.2, Tripos Inc., St. Louis, MO, 2003.
- Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 669.
- Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. *Bioorg. Med. Chem. Lett.* **2005**, 15, 3119.
- Castilho, M. S.; Guido, R. V. C.; Andricopulo, A. D. *Lett. Drug. Des. Discov.* **2007**, 4, 106.
- Valadares, N. F.; Castilho, M. S.; Polikarpov, I.; Garratt, R. C. *Bioorg. Med. Chem.* **2007**, 15, 4609.

Volume 29, Number 2, April 2005

**Journal of
ENZYME INHIBITION
and
MEDICINAL CHEMISTRY**

Editor in Chief: H. John Smith
Regional Editors: J. Garcia, J. Birkbeck, B. Ueda, G. Li, R. Bae

ISSN: 1473-0194
Taylor & Francis
Taylor & Francis Group

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*

RAFAEL V. C. GUIDO¹, GUSTAVO H. G. TROSSINI², MARCELO S. CASTILHO³, GLAUCIUS OLIVA¹, ELIZABETH I. FERREIRA², & ADRIANO D. ANDRICOPULO¹

¹Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, São Carlos-SP 13560-970, Brazil, ²Laboratório de Planejamento e Síntese de Quimioterápicos Potenciais Contra Endemias Tropicais, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Av. Professor Lineu Prestes 580, São Paulo-SP 05508-900, Brazil, and ³Laboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, Salvador-BA 40170-290, Brazil

(Received 28 August 2007; accepted 25 October 2007)

APÊNDICE F

Abstract

Chagas' disease is a parasitic infection widely distributed throughout Latin America, with devastating consequences in terms of human morbidity and mortality. Cruzain, the major cysteine protease from *Trypanosoma cruzi*, is an attractive target for antitrypanosomal chemotherapy. In the present work, classical two-dimensional quantitative structure-activity relationships (2D QSAR) and hologram QSAR (HQSAR) studies were performed on a training set of 45 thiosemicarbazone and semicarbazone derivatives as inhibitors of *T. cruzi* cruzain. Significant statistical models (HQSAR, $q^2 = 0.75$ and $r^2 = 0.96$; classical QSAR, $q^2 = 0.72$ and $r^2 = 0.83$) were obtained, indicating their consistency for untested compounds. The models were then used to evaluate an external test set containing 10 compounds which were not included in the training set, and the predicted values were in good agreement with the experimental results (HQSAR, $r_{\text{pred}}^2 = 0.95$; classical QSAR, $r_{\text{pred}}^2 = 0.91$), indicating the existence of complementary between the two ligand-based drug design techniques.

Keywords: Chagas' disease, enzyme inhibitors, drug design, chemotherapy, QSAR

Introduction

Enzymes are extremely attractive targets for small-molecule drug intervention in human diseases [1–4]. The ability to efficiently design or discover novel, patentable new chemical entities which are potent and specific inhibitors of enzymes is, therefore, of great importance [5,6]. Chagas' disease or American trypanosomiasis is a major cause of morbidity and mortality in Latin America, where about 16–18 million people are infected with *Trypanosoma cruzi* [7]. The two nitroheterocyclic drugs (nifurtimox and benznidazole) available for the treatment of this parasitic disease have severe limitations, including poor

efficacy, high toxicity, and long courses of parenteral administration. Therefore, there is an urgent need for new drugs for chemotherapy of the disease [8–10]. This has led to a new paradigm for the discovery of drug candidates that act on specific enzymes or metabolic pathways. Cysteine proteases are essential components for the survival of several parasitic protozoa. Cruzain, the major cysteine protease of the protozoan parasite *T. cruzi*, has been identified as an attractive target for the development of a new generation of antitrypanosomal agents [11,12].

Structure- and ligand-based drug design techniques are vital components of modern drug discovery. Success in this endeavor has been driven by scientific

Correspondence: A. D. Andricopulo, Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, São Carlos-SP 13560-970, Brazil. Tel: 55 16 3373 8095. Fax: 55 16 3373 9881. E-mail: aandrico@ifsc.usp.br

and technological innovation that enabled the discovery of new therapeutic agents [13,14]. Computational techniques have provided valuable information for the design of enzyme inhibitors with a wide variety of pharmacological properties, however, designing compounds that are effective against a single target is a difficult task [15]. The understanding of the chemical basis involved in the complex process of molecular recognition and biological activity, remains a major challenge. One of the key strategies in inhibitor drug design is the development of quantitative structure-activity relationships (QSAR) from a correct interpretation of the mutual binding-mode and mechanism of action within series of structurally related inhibitors. In the present study, hologram QSAR (HQSAR) and classical QSAR methods have been employed to explore both the molecular basis underlying the selective modulation of *T. cruzi* cruzain, and to investigate the QSAR of a series of chemically versatile thiosemicarbazone and semicarbazone derivatives as selective inhibitors of this trypanosomal cysteine protease.

Materials and methods

Data sets

The data set used for the QSAR studies contains 55 thiosemicarbazone and semicarbazone derivatives as inhibitors of the trypanosomal cysteine protease cruzain, which have been selected from the literature [16]. The structures and corresponding IC_{50} values for the whole set of inhibitors are included in Table I. The IC_{50} (concentration required for 50% inhibition of cruzain) values were converted to pIC_{50} ($-\log IC_{50}$) values and used as dependent variables in the QSAR analyses. The complete data set was divided into training (compounds 1–45, Table I) and test (compounds 46–55, Table I) sets. A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) using the complete linkage clustering method (Euclidean distances) with no data standardization.

HQSAR analysis

The HQSAR modelling analyses, calculations and visualisations were performed using the SYBYL 7.3 package (Tripos Inc., St. Louis, USA) running on Red Hat Linux 7.3 workstations, as previously described [17–21]. Basically, HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction. The generation of the molecular holograms was carried out using several combinations of the following fragment distinction: Atoms (A), Bonds (B), Connections (C), Hydrogen atoms (H), Chirality (Ch), and Donor & Acceptor (DA).

The influence of fragment size, which controls the minimum and maximum length of fragments to be included in the hologram, was further investigated by using six distinct fragment sizes over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using the partial least squares (PLS) analysis.

Classical QSAR analysis

Classical QSAR studies were carried out using the DRAGON 5.4 (Taletto SRL, Milan, Italy), BuildQSAR [22] and PIROUETTE 3.11 software (Infomatrix, Washington, USA), as previously described [19,20]. Briefly, 2D molecular descriptors, including topological descriptors, connectivity indices, 2D autocorrelation descriptors, and Burden eigenvalues indices, among others, were computed using the software DRAGON 5.4 and used as independent variables in the QSAR analyses. Next, about 950 molecular descriptors were subjected to the following selection criteria. Descriptors with zero variance as well as those with poor correlation to biological property ($r^2 < 0.10$) or those more than 0.97 pairwise correlated were discarded. This strategy afforded a total of 247 descriptors. Subsequently, the BuildQSAR software was employed to search for multiple linear regressions (MLR) models of up to 4 variables using genetic algorithm (GA). The selection protocol was applied to 50 individuals (models with up to 4 variables) that were evaluated through 200,000 generations. During this processes mutation (10%) and cross-over were allowed in order to increase model fitness, which was evaluated according to the correlation coefficient criteria. All descriptors present in the 10 best MLR models were pooled together, autoscaled and used for the PLS analyses performed with the PIROUETTE software.

QSAR models validation

The QSAR models were investigated using full cross-validated r^2 (q^2) PLS. Leave-one-out (LOO) cross-validation has been applied to determine the number of components that yield optimally predictive models. Leave-many-out (LMO) validation procedures with either 10 (LMO_{10}) or 5 (LMO_5) randomly selected groups were used as more rigorous tests to assess model stability and statistical significance. Each random cross-validation run was repeated 25 times to obtain mean values for q^2 and the corresponding standard error of prediction (SDEP). External validation was conducted employing a test set of 10 compounds, which were not considered for QSAR model generation.

Table I. Chemical structures and corresponding pIC₅₀ values for a series of inhibitors of cruzain from *Trypanosoma cruzi*.

Training Set Inhibitors								
Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀
1		6.55	10		7.30	19		6.70
2		5.15	11		7.70	20		7.22
3		6.85	12		5.00	21		6.34
4		6.55	13		5.41	22		7.30
5		5.55	14		6.25	23		5.72
6		5.00	15		5.00	24		6.70
7		5.00	16		5.89	25		7.10
8		5.55	17		6.49	26		6.57
9		6.66	18		6.32	27		6.64
28		7.22	34		5.00	40		5.00
29		7.15	35		5.00	41		5.00
30		7.10	36		5.80	42		5.00
31		5.00	37		6.00	43		5.00
32		5.00	38		5.85	44		5.00

Table I – continued

Training Set Inhibitors								
Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀
33		5.00	39		5.00	45		5.00
Test Set Inhibitors								
46		7.00	51		5.42			
47		6.52	52		7.40			
48		6.25	53		5.00			
49		7.70	54		5.77			
50		6.77	55		5.00			

Results and discussion

Chemical and biological data

HQSAR and classical QSAR models were derived for a data set of 55 inhibitors of *T. cruzi* cruzain (Table I). The series of chemically attractive thiosemicarbazone and semicarbazone derivatives examined has substantial structural diversity. The bulk of the structural diversity lies in the nature of the substituent linked to the N1, N4 and C5 of the thiosemicarbazone and semicarbazone moieties (Figure 1).

The *in vitro* IC₅₀ values employed in this work were measured under the same experimental conditions [16], a fundamental requirement for QSAR studies [5,17,20]. The IC₅₀ values of the data set inhibitors vary from 20 to 10,000 nM, and are acceptably distributed across the range of property values (Figure 2).

The generation of consistent statistical models is dependent on the adequacy of the training and test sets used. Therefore, from the original data set, 45 inhibitors (compounds 1–45, Table I) were selected as members of the training set for model generation, whereas the other 10 inhibitors (compounds 46–55, Table I) were held as part of the test set for external validation.

A statistical cluster analysis revealed an appropriate composition of both training and test sets in terms of chemical diversity and distribution of the biological property across the range of IC₅₀ values. Thus, the data set is suitable for QSAR model development.

HQSAR analysis

HQSAR analyses require as input only 2D chemical structures and the corresponding biological activity data. Basically, the 2D QSAR analyses carried out

Figure 1. General scaffold of thiosemicarbazone and semicarbazone derivatives.

Figure 2. Distribution of pIC₅₀ values for the data set compounds.

in this work involved three main steps: (i) the generation of substructural fragments for each *T. cruzi* cruzain inhibitor member of the data set; (ii) the encoding of these fragments into a molecular hologram; (iii) the statistical generation of PLS QSAR models [20,23]. The molecular hologram (a specialised fragment fingerprint) generated consists of several substructural fragments of the thiosemicarbazone and semicarbazone derivatives that were counted into bins of a fixed length array. The bins represent all of the unique fragments included within a particular molecule and were assigned by a cyclic redundancy check (CRC) algorithm. The bin occupancies of the molecular hologram are structural descriptors (independent variables) encoding compositional and topological molecular information, that is, a fundamental component of molecular basis for the process of target recognition and biological activity. Thus, the molecular hologram includes information on the quantity and type of each fragment in each molecule of the data set, encoding all linear, branched, and overlapping molecular fragments without requiring the 3D alignment of the inhibitors.

HQSAR uses the molecular holograms and PLS to generate fragment-based structure-activity relationships. This investigation requires selecting values for parameters that specify the length of the hologram, as well as the size and type of the fragments that are to be encoded [24–26]. Therefore, several combinations of fragment distinction were considered during the QSAR modelling runs using the fragment size default

(4–7), as follows: A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/C, A/C/Ch, A/C/DA, A/C/H/Ch, A/C/Ch/DA, A/B/H, A/B/C/Ch, A/B/DA, A/B/C/DA, A/B/H/DA, A/B/C/H/DA, and A/B/H/Ch/DA. The HQSAR analyses were performed over the 12 default series of hologram lengths ranging from 53 to 401 bins. The statistical results from the PLS analyses for the 45 training set inhibitors of *T. cruzi* cruzain using several fragment distinction combinations are summarized in Table II.

As can be seen, the best statistical results were obtained using A/C/Ch/DA as distinction information (model 9, $q^2 = 0.72$ and $r^2 = 0.94$, with $SEE = 0.23$). During the development of the several models, a considerable influence of the fragment distinction parameter on the predictive ability of the models was observed, as indicated by the values of q^2 in Table II (models 1–8, 10–15). This is probably due to an increase in poor information generated by the descriptors used for model development, which adds substantial noise to the cross-validation process [27]. Conversely, the inclusion of the fragment distinction Ch and DA has dramatically improved the predictive ability of the model 9 when compared to models 6 and 7, suggesting that steric and hydrogen-bonding properties play a major role towards the inhibitory potency in this series of cruzain inhibitors. These results highlight the importance of selecting the most appropriate combination of molecular descriptors for the generation of predictive models, which reflects the quantitative description of the structure-activity relationships within

Table II. HQSAR analysis for various fragment distinctions on the key statistical parameters using fragment size default (4-7).

Model	Fragment Distinction	Statistical Parameters				
		q^2	r^2	SEE	HL	PC
1	A/B	0.45	0.83	0.39	307	6
2	A/B/C	0.54	0.87	0.35	401	6
3	A/B/C/H	0.49	0.88	0.33	353	6
4	A/B/C/H/Ch	0.57	0.88	0.33	353	6
5	A/C	0.41	0.79	0.44	199	6
6	A/C/Ch	0.41	0.83	0.38	353	6
7	A/C/DA	0.24	0.49	0.65	97	3
8	A/C/H/Ch	0.38	0.73	0.49	151	4
9	A/C/Ch/DA	0.72	0.94	0.23	353	6
10	A/B/H	0.41	0.86	0.36	401	6
11	A/B/C/Ch	0.56	0.84	0.38	53	6
12	A/B/DA	0.27	0.52	0.64	83	4
13	A/B/C/DA	0.26	0.50	0.65	307	3
14	A/B/H/DA	0.52	0.80	0.42	353	5
15	A/B/C/H/DA	0.53	0.82	0.40	353	6
16	A/B/H/Ch/DA	0.67	0.92	0.27	151	6

q^2 , cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; SEE, non-cross-validated standard error; HL, hologram length; PC, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

the data set. Subsequently, the influence of different fragment sizes on the key statistical parameters was further investigated for the best HQSAR model generated (Table III). Fragment size parameters control the minimum and maximum lengths of fragments to be included in the hologram fingerprint. The results for the different fragment sizes evaluated (2-5, 3-6, 4-7, 5-8, 6-9 and 7-10) show that the statistical quality of models improved with the increasing of the fragment size up to 6-9 (model 21, $q^2 = 0.75$ and $r^2 = 0.96$, with $SEE = 0.20$).

The internal statistical consistency of the model 21 was evaluated through the analysis of values of q^2 and SDEP obtained by more consistent cross-validation procedures using about 10% and 20% of training set compounds (LMO₁₀ and LMO₅, respectively). The cross-validation using 10 random groups yielded an average q^2_{10} of 0.72 with SDEP₁₀ of 0.50, while an average q^2_5 of 0.70 with SDEP₅ of 0.52 was obtained for the cross-validation using 5 random groups. These results confirm the good stability and robustness of the model.

Although q^2 values may give a suitable representation of the model internal consistency and predictive power, the most valuable test of a QSAR model is its ability to predict the property value for new compounds, not included in the training set for model generation [24,28-30]. As the structure encoded in the 2D specialized fingerprints is directly related to the biological activity, HQSAR models should be useful to predict the activity of new structurally related molecules from its fingerprint [25,26]. Thus, the predictive power of the best QSAR model derived using the 45 training set inhibitors was assessed by predicting the biological property for 10 test set compounds (inhibitors 46-55, Table I). The results of the external evaluation are shown in Table IV, and the graphic results for the predicted versus experimental activities of both training set and test set are displayed in Figure 3.

The good agreement between experimental and predicted pIC₅₀ values for the test set cruzain inhibitors indicates the high quality of the HQSAR model ($r^2_{pred} = 0.95$). The predicted values fall close to the experimental pIC₅₀ values, deviating by no more

Table III. HQSAR analysis for the influence of different fragment sizes on the statistical parameters.

Fragment Distinction	Model	Fragment Size	Statistical Parameters				
			q^2	r^2	SEE	HL	PC
A/C/Ch/DA	17	2-5	0.67	0.86	0.36	59	6
	18	3-6	0.70	0.92	0.26	307	6
	19	4-7	0.71	0.94	0.23	353	6
	20	5-8	0.73	0.95	0.21	353	6
	21	6-9	0.75	0.96	0.20	401	6
	22	7-10	0.73	0.95	0.22	401	6

Table IV. Experimental and predicted activities (pIC_{50}) with residual values for the test set inhibitors.

Cpd	Experimental	HQSAR		Classical QSAR	
		Predicted	Residual ^a	Predicted	Residual ^a
46	7.00	6.75	0.25	6.59	0.41
47	6.52	6.51	0.01	6.60	-0.08
48	6.25	5.81	0.44	6.52	-0.27
49	7.70	7.97	-0.27	7.53	0.17
50	6.77	6.91	-0.14	6.75	0.02
51	5.42	5.62	-0.20	5.62	-0.20
52	7.40	7.25	0.15	6.98	0.42
53	5.00	4.85	0.15	4.93	0.07
54	5.77	5.78	-0.01	6.35	-0.58
55	5.00	5.15	-0.15	4.97	0.03

^a The difference between experimental and predicted values.

than 0.44 log units. From the low residual values, it can be seen that the HQSAR model obtained is substantially reliable and can be employed to predict the biological activity of novel cruzain inhibitors within this structural diversity.

Besides predicting the activities of untested molecules, the HQSAR model plays another important role in providing hints about the influence of different molecular subfragments to biological activity. It includes the identification of substructural patterns related to potency (IC_{50} values) in the data set of cruzain inhibitors. Therefore, the HQSAR model is graphically represented in the form of very intuitive contribution maps where the colours of the atoms reflect their relative contribution to the overall activity of the molecule. In this context, an accurate interpretation of the contribution maps revealed important molecular fragments related to biological

potency, as well as providing further insights into the underlying mechanism of action of this series of cruzain inhibitors. For the analysis of the contribution maps, colours at the red end of the spectrum (e.g., red, red-orange, and orange) reflect poor (or negative) contributions, while colours at the green end of the spectrum (e.g., yellow, green-blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are coloured white.

For example, the contribution of different subfragments for the most potent inhibitor of the series (compound **11**, Table I) is shown in Figure 4. As can be seen, the contribution map indicates that presence of the thiosemicarbazone moiety is favorable to the inhibitory activity. This is supported by the fact that the replacement of the thiosemicarbazone with a semicarbazone moiety (compounds **39-44**) did considerably affect the inhibitory potency, showing that the thiosemicarbazone moiety is essential for this series of enzyme inhibitors. In addition, the contribution map also highlights the importance of the sulfur atom (thiosemicarbazone) for intermolecular interactions with the catalytic cysteine of the *T. cruzi* cruzain binding pocket. Concomitantly, the model also shows the importance of the N4 and C5 alkyl substituents for activity. This is in agreement with the fact that compounds **9** and **10**, with ethyl substituent at C5, have enhanced inhibitory activity when compared to molecules without substituents on C5 (inhibitors **14-18**) [16].

Classical QSAR analysis

In order to obtain further insight into the structure-activity relationships for this series of cruzain inhibitors, we resorted to a classical QSAR approach. Recently, the integration between HQSAR and classical QSAR methods has been successfully applied [19,20], illustrating the complementary nature of these approaches as a powerful tool for medicinal chemistry studies and drug design.

Figure 3. Plot of predicted values of pIC_{50} versus the corresponding experimental values for the training (open squares) and test (solid circles) set inhibitors for the final HQSAR model.

Figure 4. HQSAR contribution map for the most potent cruzain inhibitor 11.

Classical 2D QSAR studies require the calculation of molecular descriptors and the construction of mathematical models. In this work, 2D-topological, 2D-BCUTs, 2D-walk and path counts, 2D-autocorrelations, 2D-connectivity indices, 2D-information indices, 2D topological charge indices, and 2D-eigenvalue-based indices, were computed in the DRAGON 5.4 software and then used as independent variables in the QSAR modelling investigations. In order to reduce redundant and useless information, descriptors possessing zero variance, poor correlation to biological activity ($r^2 < 0.10$) as well as those with pairwise correlation superior to 0.97 were discarded. The descriptor space was further reduced through the selection of descriptors closely related to the biological activity. In order to achieve this task, BuildQSAR software was employed to search for the 10 best multiple linear regression (MLR) models of up to 4 variables using GA [31]. This strategy led to the selection of 12 descriptors (gathered from the MLR models) that were pooled together, autoscaled and used in the principal components analysis (PCA) and PLS analysis with the PIROUETTE software.

According to the PCA results, 4 principal components accounted for 91% of total variance, while additional components were considered irrelevant. A graphical analysis of the PCA results indicates that some molecular features of the data set inhibitors could be satisfactorily described by PC2. Less bulky compounds have negative values in PC2, while bulkier compounds have positive values. Weak inhibitors ($\text{pIC}_{50} = 5.0$) are broadly distributed along PC1 and PC2; however, none of them is located in the lower-right quadrant. In contrast, 75% of the most potent inhibitors ($\text{pIC}_{50} > 7.0$) are located in this region. These results suggest that the selected descriptors can be used for deriving QSAR models. Thus, PLS QSAR models were created using the LOO cross-validation procedure. As can be seen in Figure 5, the best statistical model shows good correlation within the training set ($q^2 = 0.72$ and $r^2 = 0.83$). The stability

and robustness of the model was further evaluated by the LMO cross-validation procedures. In both cases, no significant differences were obtained ($q_{10}^2 = 0.71$, with $\text{SDEP}_{10} = 0.52$; and $q_5^2 = 0.69$, with $\text{SDEP}_5 = 0.54$).

The predictive ability of the model was assessed through the prediction of the biological property for 10 test set inhibitors (compounds 46-55, Table I). The results, of similar quality compared to the HQSAR investigation, are listed in Table IV ($r_{\text{pred}}^2 = 0.91$). Besides demonstrating statistical significance, QSAR models should also provide useful chemical insights for drug design. This goal can be accomplished through a careful interpretation of the importance of the descriptors for the QSAR model. Figure 6 illustrates the influence of the descriptors over the regression vector, which can be regarded as a weighted

Figure 5. Plot of predicted values of pIC_{50} versus the corresponding experimental values for the training (open squares) and test (solid circles) set inhibitors for the final classical QSAR model.

Figure 6. Descriptors score to regression vector.

sum of the loadings included in the model. Accordingly, descriptors with small coefficients (PW3, PiPC09, PiPC10, X5A, EEig03d, EEig02r, BELe4, BELe3, SEIgv, C-039) did not significantly affect the quality of the QSAR model. Conversely, BELm3 and BEHp3 are very important topological descriptors that can be derived from the connectivity matrix, the matrix of eigenvectors, and a diagonal matrix of eigenvalues weighted by an atomic property (atomic masses and atomic polarizabilities, respectively). These descriptors are collectively known as BCUT metric and contain contributions from all atoms, thus reflecting the topology of the whole molecule. The opposite contribution of BELm3 and BEHp3 to the final model indicates that highly polarized under-sized molecules have a better interaction profile to the cruzain binding site. For instance, the integrated analyses of classical QSAR and HQSAR models can be regarded as complementary in nature by suggesting that the inhibitory potency of compounds **9** and **10** might be improved in the presence of more polarizable groups of similar molecular mass in positions N4 and C5 adjacent to thiosemicarbazone moiety.

Conclusion

The use of fragment-based molecular holograms and topological descriptors, along with a standard and reproducible high quality biological data, allowed the generation of QSAR models exhibiting both good internal and external consistency, with substantial predictive power for this series of *T. cruzi* cruzain inhibitors. The information gathered from the HQSAR contribution maps, integrated with a careful analysis of the topological descriptors from the classical QSAR, shed some light on the effects of the substitution pattern related to biological activity within this series of thiosemicarbazone and semicarbazone derivatives. The consistent results confirmed that investigations can be carried out concomitantly to search for synergies between 2D QSAR technologies and drug design. In sum, as the models reported herein have been internally and

externally validated employing different 2D QSAR approaches, they appear to be valuable tools for ligand-based enzyme inhibitor design.

Acknowledgements

We gratefully acknowledge financial support from FAPESP (The State of São Paulo Research Foundation), FAPESB (The State of Bahia Research Foundation) and CNPq (The National Council for Scientific and Technological Development), Brazil.

References

- [1] Mestres J. Representativity of target families in the Protein Data Bank: Impact for family-directed structure-based drug discovery. *Drug Discov Today* 2005;10:1629–1637.
- [2] Robertson JG. Mechanistic basis of enzyme-targeted drugs. *Biochemistry* 2005;44:5561–5571.
- [3] Lespinet O, Labedan B. Orphan enzymes could be an unexplored reservoir of new drug targets. *Drug Discov Today* 2006;11:300–305.
- [4] Supuran CT, Scozzafava A. Carbonic anhydrases as targets for medicinal chemistry. *Bioorg Med Chem* 2007;15:4336–4350.
- [5] Farutin V, Masterson L, Andricopulo AD, Cheng J, Riley R, Hakimi R, Frazer J, Cordes EH. Structure-activity relationships for a class of inhibitors of purine nucleoside phosphorylase. *J Med Chem* 1999;42:2422–2431.
- [6] Andricopulo AD, Montanari CA. Structure-activity relationships for the design of small-molecule inhibitors. *Mini Rev Med Chem* 2005;5:585–593.
- [7] Beyrer C, Villar JC, Suwanvanichkij V, Singh S, Baral SD, Mills EJ. Neglected diseases, civil conflicts, and the right to health. *Lancet* 2007;370:619–627.
- [8] Barrett MP, Burchmore RJ, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, Krishna S. The trypanosomiasis. *Lancet* 2003;362:1469–1480.
- [9] Nwaka S, Hudson A. Innovative lead discovery strategies for tropical diseases. *Nat Rev Drug Discov* 2006;5:941–955.
- [10] Nwaka S, Ridley RG. Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nat Rev Drug Discov* 2003;2:919–928.
- [11] McKerrow JH, Engel JC, Caffrey CR, Related Articles, Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg Med Chem* 1999;7:639–644.
- [12] Doyle PS, Zhou YM, Engel JC, McKerrow JH. A Cysteine Protease Inhibitor Cures Chagas' Disease in an Immuno-deficient Murine Model of Infection. *Antimicrob Agents Chemother* 2007, doi: 10.1128/AAC.00436-07.
- [13] Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: Foundations, limitations and novel approaches. *Drug Discov Today* 2007;12:225–233.
- [14] Klebe G. Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discov Today* 2006;11:580–594.
- [15] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat Rev Drug Discov* 2004;3:935–949.
- [16] Du X, Guo C, Hansell E, Doyle PS, Caffrey CR, Holler TP, McKerrow JH, Cohen FE. Synthesis and structure-activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain. *J Med Chem* 2002;45:2695–2707.
- [17] Honorio KM, Garratt RC, Andricopulo AD. Hologram quantitative structure-activity relationships for a series of farnesoid X receptor activators. *Bioorg Med Chem Lett* 2005; 15:3119–3125.

- [18] Honorio KM, Garratt RC, Polikarpov I, Andricopulo AD. Hologram QSAR studies on farnesoid X receptor activators. *Lett Drug Des Discov* 2006;3:261–267.
- [19] Castilho MS, Guido RVC, Andricopulo AD. 2D Quantitative structure activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. *Bioorg Med Chem* 2007;15: 6242–6252.
- [20] Castilho MS, Guido RVC, Andricopulo AD. Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors. *Lett Drug Des Discov* 2007; 4:106–113.
- [21] Ferreira LG, Leitao A, Montanari CA, Andricopulo AD. Hologram quantitative structure-activity relationships for a class of inhibitors of HIV-1 protease. *Lett Drug Des Discov* 2007;4:356–364.
- [22] Oliveira DB, Gaudio AC. BuildQSAR: A new computer program for QSAR analysis. *Quant Struct-Activ Relat* 2003; 19:599–601.
- [23] HQSARTM Manual. SYBYL 6.9.2, Tripos Inc., St. Louis, MO, 2003.
- [24] Honorio KM, Garratt RC, Polikarpov I, Andricopulo AD. 3D QSAR comparative molecular field analysis on Nonsteroidal farnesoid X receptor activators. *J Mol Graph Modell* 2007;25: 921–927.
- [25] Castilho MS, Postigo MP, de Paula CB, Montanari CA, Oliva G, Andricopulo AD. Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg Med Chem* 2006; 14:516–527.
- [26] Salum LD, Polikarpov I, Andricopulo AD. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. *J Mol Graph Modell* 2007;25:434–442.
- [27] Wold S, Johansson E, Cocchi M. PLS — Partial Least Square projections to latent structures. In: Kubinyi H, editor. *3D QSAR in Drug Design*. Leiden: ESCOM; 1993. p 523–550.
- [28] Doweiko AM. 3D-QSAR illusions. *J Comput Aided Mol Des* 2004;18:587–596.
- [29] Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* 2003; 17:241–253.
- [30] Golbraikh A, Tropsha A. Beware of q^2 ! *J Mol Graph Model* 2002;20:269–276.
- [31] Liu H, Gramatica P. QSAR study of selective ligands for the thyroid hormone receptor beta. *Bioorg Med Chem* 2007; 15:5251–5261.

The development of an immobilized enzyme reactor containing glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma cruzi*: the effect of species' specific differences on the immobilization

Carmen Lúcia Cardoso,^a Marcela Cristina de Moraes,^b Rafael Victorio Carvalho Guido,^c Glaucius Oliva,^c Adriano Defini Andricopulo,^c Irving William Wainer^d and Quezia Bezerra Cass^{*b}

Received 23rd July 2007, Accepted 17th September 2007

First published as an Advance Article on the web 8th October 2007

DOI: 10.1039/b711145b

APÊNDICE G

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) plays an important role in the life cycle of the *Trypanosoma cruzi*, and an immobilized enzyme reactor (IMER) has been developed for use in the on-line screening for GAPDH inhibitors. An IMER containing human GAPDH has been previously reported; however, these conditions produced a *T. cruzi* GAPDH-IMER with poor activity and stability. The factors affecting the stability of the human and *T. cruzi* GAPDHs in the immobilization process and the influence of pH and buffer type on the stability and activity of the IMERs have been investigated. The resulting *T. cruzi* GAPDH-IMER was coupled to an analytical octyl column, which was used to achieve chromatographic separation of NAD⁺ from NADH. The production of NADH stimulated by D-glyceraldehyde-3-phosphate was used to investigate the activity and kinetic parameters of the immobilized *T. cruzi* GAPDH. The Michaelis–Menten constant (K_m) values determined for D-glyceraldehyde-3-phosphate and NAD⁺ were $K_m = 0.5 \pm 0.05$ mM and 0.648 ± 0.08 mM, respectively, which were consistent with the values obtained using the non-immobilized enzyme.

Introduction

Chagas' disease, also known as American trypanosomiasis, is an infection caused by the protozoan parasite *Trypanosoma cruzi*. Chagas' disease is endemic in 15 countries of Latin America, and, according to the World Health Organization the disease affects 16–18 million people, about 40 million people are at risk and 200 000 new cases are registered each year.^{1,2} The drugs currently available for the treatment of this disease (e.g. benznidazole and nifurtimox) are inadequate and their use is limited by serious side effects, toxicities, and ineffectiveness. Therefore, the discovery of new drugs with different mechanisms of action for the treatment of Chagas' disease is a critical need and a challenging task.³

One promising approach to accomplish this task is the selective inhibition of enzymes that participate in the glycolytic pathway of the parasite. The trypanosomatids are highly dependent on glycolysis for ATP production,^{4,5} and the reaction catalyzed by glycosomal glyceraldehyde-3-phosphate dehydrogenase (GAPDH) plays a central role in

controlling ATP production in pathogenic parasites.^{6,7} GAPDH (EC 1.2.1.12) catalyzes the oxidative phosphorylation of D-glyceraldehyde-3-phosphate into 1,3-diphosphoglycerate in the presence of NAD⁺ and inorganic phosphate.

Crystallographic studies showed that GAPDH from *T. cruzi* and human GAPDH differ by a substitution of Asp²¹⁰ (*T. cruzi*) by Leu¹⁹⁴ (human).⁸ Based on this difference, it is possible that a selective inhibitor of *T. cruzi* GAPDH could be developed to treat Chagas' disease.⁹ The discovery and development of a selective *T. cruzi* GAPDH inhibitor is a challenging task, which requires the development of methods to rapidly identify lead compounds in complex chemical and biological mixtures, and to assess the specificity for GAPDH of the target (*T. cruzi*) relative to the host (human). One such approach is on-line screening using an immobilized enzyme reactor (IMER).

IMERs have been prepared from a wide variety of enzymes and have been used in high performance liquid chromatographic systems for carrying out on-line synthesis, in the study of enzyme kinetics for the determination Michaelis–Menten constant and in the identification of enzyme inhibitors. The development and use of IMERs have recently been reviewed.¹⁰ As part of our program to develop new treatments for Chagas' disease, this laboratory initially developed an IMER containing human GAPDH immobilized within a fused silica capillary.¹¹ However, when the same immobilization procedure was followed using *T. cruzi* GAPDH in place of human GAPDH, the resulting IMER had poor enzymatic activity and stability. This work reports the results from a systematic study of factors, such as pH and buffer type, that affect *T. cruzi* GAPDH stability and activity. The optimized conditions were

^aDepartamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, São Paulo, Brazil

^bDepartamento de Química, Universidade Federal de São Carlos, Cx. Postal 676, São Carlos, 13565-905, São Paulo, Brazil.
E-mail: quezia@dq.ufscar.br; Fax: +55-16-3351-8350;
Tel: +55-16-3351-8087

^cCentro de Biotecnologia Molecular Estrutural – CBME, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil

^dNational Institute of Aging, National Institutes of Health, Baltimore, MD, USA

used to prepare a *T. cruzi* GAPDH-IMER, which was placed in a multidimensional high performance chromatographic system, and the resulting system was used to characterize the activity of the immobilized enzyme. The data from this study demonstrate that subtle changes in protein structure can require significant alterations in the procedures required to immobilize the protein in a chromatographic environment.

Experimental

Reagents and materials

D,L-Glyceraldehyde-3-phosphate free acid (GA3P), β -nicotinamide adenine dinucleotide reduced form (NADH), and β -nicotinamide adenine dinucleotide (NAD^+) were purchased from Sigma Aldrich (St. Louis, MO, USA). Buffer components and all chemical materials used during the immobilization procedure were of analytical grade and were supplied by Sigma or by Merck (Darmstadt, Germany). All solvents were HPLC grade and were purchased from J.T. Baker (Phillipsburg, USA). Water was purified with a Millipore Milli-Q system (Millipore, São Paulo, Brazil) and was used for all experiments. The mobile phases were prepared daily. The fused silica capillary for electrophoresis (0.375 mm $\times$ 0.10 mm) used to immobilize the enzyme and to prepare the IMERs was purchased from Polymicro Technologies (Phoenix, AZ, USA). The Luna[®] octyl silica (10.0 μm , 100 Å) was supplied by Phenomenex (Torrance, CA, USA). Before their use for HPLC analysis, the buffer solutions were filtered through cellulose nitrate membrane filters (0.45 μm) purchased from Phenomenex. Dialysis and concentration of enzymatic solution was carried out using an Amicon concentrator (30 mL, Millipore, Billerica, MA) and a centrifuge (Eppendorf Instruments, Enfield, USA).

Buffers

The purification and storage buffer was: triethanolamine (100.0 mM, pH 7.5), containing 1.0 mM β -mercaptoethanol, 1.0 mM ethylenediaminetetraacetic acid (EDTA), 30.0 mM sodium arsenate heptahydrate ($\text{NaHAsO}_4 \cdot 7\text{H}_2\text{O}$), 1.0 mM phenylmethylsulfonyl fluoride (PMSF), 50.0 mM NAD^+ , 1.0 mM pepstatin, 1.0 mM leupeptin, 0.5 M KCl, and glycerol 10%.

Buffer 1: triethanolamine (100.0 mM, pH 7.5), containing 1.0 mM β -mercaptoethanol, 1.0 mM EDTA, 30.0 mM $\text{NaHAsO}_4 \cdot 7\text{H}_2\text{O}$, 1.0 mM PMSF, and 0.5 M KCl.

Buffer 2: phosphate buffer (50.0 mM, pH 7.0).

Buffer 3: 20.0 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8.2.

Buffer 4: 20.0 mM ammonium acetate, pH 8.0.

Buffer 5: 20.0 mM sodium borate, pH 8.6.

Buffer 6: triethylamine (TEA) (1% in water v/v acidified with AcOH, pH = 6.0).

Buffer 7: triethanolamine (100.0 mM, pH 7.5), containing 1.0 mM EDTA, 1.0 mM PMSF, 1.0 mM β -mercaptoethanol.

Buffer 8: Tris-HCl (50.0 mM, pH 8.6), containing 1.0 mM β -mercaptoethanol, 30.0 mM $\text{NaHAsO}_4 \cdot 7\text{H}_2\text{O}$, and 1.0 mM EDTA.

T. cruzi GAPDH

T. cruzi GAPDH was over-expressed and purified as previously reported.⁹

Chromatographic systems

The immobilization of the enzyme was carried out using a syringe-pump 341B (Sage Instruments, Boston, USA).

Two modular HPLC systems were setup in order to carry out the on-line studies and the systems were connected as depicted in Fig. 1. The chromatographic experiments were carried out using a Shimadzu HPLC system (Shimadzu, Kyoto, Japan), which consisted of the two LC 10 AD VP pumps with one of the pumps having a valve FCV-10AL for selecting solvent, a UV-Vis detector (SPD-M10AV VP), an autosampler equipment with a 500 μL loop (SIL 10 AD VP). The column containing the immobilized GAPDH enzyme (GAPDH-IMER) was coupled on-line to an octyl column (Luna-Phenomenex[®], 100 Å, 10 μm , 10 cm $\times$ 0.46 mm I.D.). A six-way switching sample-valve, Valco Nitronic 7000 EA (Supelco, St. Louis, MO, USA), was used to connect the two columns. Data acquisition was done on a Shimadzu SCL 10 AVP system interface with a computer equipped with Shimadzu-LCsolution (LCsolution 2.1) software (Shimadzu, Kyoto, Japan).

Free GAPDH storage conditions

Conditions of storage were evaluated for the free soluble GAPDH. After purification, the enzyme was stored at low temperature (-80°C) in the purification buffer. The enzyme activity and stability were evaluated by weekly measurements of the activity of the stored enzyme. The procedure for measuring the activity is described in the section 'Free GAPDH activity assays'.

Optimization of the immobilization conditions for GAPDH on fused silica capillary

Previously reported immobilization procedure was evaluated and then modified¹¹ as follows: three different immobilization buffers were tested: buffer 3, buffer 4, and buffer 5.

The fused capillary cleaning procedure was changed from NaOH 2.0 mol L^{-1} to HCl 2.0 mol L^{-1} and the effect on the silanization procedure was investigated by passing

Fig. 1 Schematic diagram of multidimensional chromatographic system.

3-aminopropyltriethoxysilane (APTS) solution, once or twice, thorough the capillary.

The modified immobilization procedures for GAPDH

Using a syringe-pump with a flow rate of $130 \mu\text{L min}^{-1}$, the fused silica capillary tube ($100 \mu\text{m I.D.} \times 0.375 \text{ mm} \times 50 \text{ cm}$) was cleaned by washing with 2.0 mL of a 2.0 mol L^{-1} HCl solution, followed by 1.0 mL of distilled water. After rinsing, the capillary was dried in an oven at 95°C for at least 1 h, and then 1.0 mL of a solution of 3-aminopropyltriethoxysilane [10% (v/v)] in water was passed through the capillary and subsequently placed in an oven at 95°C for 30 min. The capillary was stored overnight at room temperature.

The enzymatic solution was exhaustively dialyzed from the storage buffer to buffer 3 and concentrated, using an Amicon concentrator, to a final concentration of the 1.0 mg mL^{-1} used in the immobilization step.

A glutaraldehyde solution 1% (v/v), in buffer 2 (2.0 mL), was passed through the aminopropylsilica (APS) capillary by syringe-pump at $130 \mu\text{L min}^{-1}$ flow rate. In order to remove free glutaraldehyde and thus avoid polymerization, the capillary tubing was rinsed with buffer 2 (0.5 mL at $130 \mu\text{L min}^{-1}$). After this, the capillary tube was rinsed with buffer 3 (0.5 mL), immediately followed by 0.5 mL of GAPDH enzyme solution (1.0 mg mL^{-1}) in buffer 3. The enzyme solution was passed through the capillary a second time, and then the capillary tube was rinsed with 1.0 mL of buffer 1. When not in use the GAPDH-IMER was kept at 4°C with the two ends of the capillary tubing immersed in buffer 1.

GAPDH-IMER storage conditions and mobile phase

Two different buffers – buffer 1 and buffer 8 – were evaluated in order to select the best working buffer and optimal conditions of storage for the GAPDH-IMERS. The pH and temperature effects on the enzyme stability were evaluated with the free enzyme using both buffers. This was done daily in order to estimate the decrease in activity at each day. The experimental details are specified in the section ‘Free GAPDH activity assays’.

Chromatographic conditions

The analytical columns were packed by the ascending slurry method, using methanol for the preparation of the slurry (50.0 mL) and also for the packing. The packing was carried out at a pressure of 7500 p.s.i. (1 p.s.i. = 6894.76 Pa).

The analytical columns listed below were evaluated in different chromatographic conditions and temperatures: Column A: diol-silica Spherex[®]-OH ($100 \text{ \AA}$, $10.0 \text{ cm} \times 0.46 \text{ mm I.D.}$, $10 \mu\text{m}$); Column B: octyl silica Luna[®] ($100 \text{ \AA}$, $10 \mu\text{m}$, $10.0 \text{ cm} \times 0.46 \text{ mm I.D.}$). Mobile phases evaluated: (a) KH_2PO_4 10 mM, pH 6.0, flow rate 0.8 mL min^{-1} ; (b) TEA (1% in water v/v, pH 6.0)–MeOH (98 : 2; 91 : 9; 90 : 10, 97 : 3 v/v) flow rate 0.8, 0.6 mL min^{-1} ; (c) ammonium acetate 10.0 mM, pH 6.0, flow rate 0.8 mL min^{-1} ; (d) HEPES 10.0 mM, pH 6.0, flow rate 0.8 mL min^{-1} ; (e) TEA 10.0 mM, pH 6.0, flow rate 0.8 mL min^{-1} ; (f) TEA (1% in water v/v, pH 6.0)–MeCN (98 : 2; 97 : 3; 96 : 4, 96.5 : 3.5, 91 : 9, 90 : 10, v/v) flow rate 0.8, 0.6 mL min^{-1} . Temperatures evaluated: (a) 22°C , (b) 25°C , (c) 28°C , and (d) 35°C .

The flow rate used in the GAPDH-IMER and time-width, to transfer the enzyme reaction products from the GAPDH-IMER to the analytical columns, was evaluated by injecting duplicate 15 μL aliquots of a solution containing NAD^+ (20 mM) and NADH (2.0 mM).

The chromatographic separations between NAD^+ and NADH were achieved by a multidimensional chromatography system, in which the GAPDH-IMER was used in the first dimension; at room temperature, coupled to the analytical octyl silica column ($100 \text{ \AA}$, $10 \mu\text{m}$, $10 \text{ cm} \times 4.6 \text{ cm I.D.}$) using a switching six way valve (Fig. 1). The chromatographic conditions are specified on Table 1.

Method validation

The NADH calibration curve was obtained using the appropriate standard solutions of NADH. Sample solutions were prepared in triplicate at the following concentrations: 5, 10, 20, 40, 80, 160, 280, and $320 \mu\text{mol L}^{-1}$. To prepare these solutions, aliquots (60 μL) of the appropriate standard solution of NADH were added to 40 μL of buffer 1. The solutions were vortex-mixed for 10 s and aliquots of 90 μL were transferred to an autoinjector vial. Samples of 15 μL were injected to the GAPDH-IMER at the HPLC system. NADH calibration curves were constructed by plotting the peak area against the concentration of NADH.

The intra- and inter-day precision and accuracy of the method were evaluated by analyzing quality control samples at three different concentrations: 12.0, 240.0 and $300.0 \mu\text{mol L}^{-1}$. Five samples of each concentration were prepared and analyzed on three non-consecutive days. The acceptance criteria for the limit of quantification were that the precision of three samples was under 20% of variability, while the limit of detection was calculated taking a signal-to-noise ratio of 3.

Table 1 Multidimensional chromatographic conditions to separate NAD^+ from NADH

Pump (eluent) ^a	Time/min	Event	Valve position
1 (A)	0.00–2.00	Elution of reagents through the GAPDH-IMER	1
2 (B)	0.00–2.00	Conditioning of the analytical column	1
1 (A)	2.01–8.50	Transfer of the analytes from the IMER to analytical column	2
2 (B)	2.01–8.50	Conditioning analytical column	2
1 (A)	8.51–20.00	GAPDH-IMER conditioning	1
2 (B)	8.51–20.00	Separation of the analytes at the analytical column	1

^a Pump 1: flow rate: 0.05 mL min^{-1} , eluent A, buffer 1. Pump 2: flow rate: 0.8 mL min^{-1} , eluent B, buffer 6: MeCN (96.5 : 3.5, v/v). $\lambda = 340 \text{ nm}$.

Chromatograms of blank buffer were analyzed to evaluate the selectivity of the method.

Free GAPDH activity assays

The enzymatic activity of soluble GAPDH was evaluated by measuring the formation of NADH by the following modified spectrophotometric method.⁹ In a quartz cuvette (500 μL) the reaction mixture was composed as follows (final concentration): to 5 μL of enzyme solution (20.0 nmol L^{-1}) was added 385 μL of buffer 7, 15 μL of $\text{NaHAsO}_4 \cdot 7\text{H}_2\text{O}$ (30.0 mM), 30 μL of D,L-GA3P (800 μM , final concentration of D-GA3P), and 5 μL NAD^+ (800 μM). The reaction was initiated with the addition of D,L-GA3P solution. The extent of the enzymatic conversion was monitored by following the increase in NADH at $\lambda = 340 \text{ nm}$. Enzymatic activity was calculated from the initial slope of the curve obtained during 30 s of reaction. A sample prepared with buffer 7, $\text{NaHAsO}_4 \cdot 7\text{H}_2\text{O}$, and NAD^+ was used as the reference sample.

Spectrometric determinations were performed using a Shimadzu (Shimadzu, Kyoto, Japan) UV-1650 PC spectrophotometer, with a computer equipped with a UV Probe (Kinetics) software version 1.10 for data collection.

Kinetic parameters were determined using the Sigma Plot software version 7.0.

Kinetics studies of the immobilized enzyme

The enzymatic activity of GAPDH in the IMER format was evaluated by using the multidimensional chromatographic system. Kinetic studies were performed in order to determine saturating conditions for NAD^+ and D-GA3P. Solutions with NAD^+ concentrations ranging from 0.10 to 12.5 mM and D-GA3P concentrations between 0.10 and 12.5 mM were injected in duplicate. The parameters for NAD^+ were determined under saturating conditions of D-GA3P (7.5 mM) while the parameters for D-GA3P were determined at saturating concentrations of NAD^+ (10.0 mM). Samples were injected in duplicate (15 μL) and the chromatographic conditions are the ones specified in Table 1.

Non-linear regression analysis using the Sigma Plot software version 7.0 was used to determine the Michaelis–Menten values (K_m) for the studied systems. GAPDH-IMER stability was determined every day by injecting 15 μL of saturating concentrations of NAD^+ (10.0 mM) and D-GA3P (7.5 mM).

Results and discussion

Free *T. cruzi* GAPDH storage conditions

Once the *T. cruzi* GAPDH was expressed and purified a crucial point was the long-term storage of this soluble enzyme with preservation of activity. The stability of GAPDH in the storage buffer at -80°C was evaluated weekly for an eight month period. The results demonstrated that 99% of the enzymatic activity was retained under the storage conditions, which made it possible to use the same batch of purified enzymes in the preparation of the *T. cruzi* GAPDH-IMERS used in this study.

Mobile phase and conditions of storage for GAPDH-IMERS

The stability and enzymatic activity are influenced by a number of factors including temperature, pH, and buffer composition. These variables were investigated using free *T. cruzi* GAPDH in order to determine the optimum immobilization and working buffers, temperature and storage conditions for the *T. cruzi* GAPDH-IMER. Two buffers were evaluated as working buffers: buffer 1 at pH 7.5 and buffer 8 at pH 8.6, at two different temperatures.

Aliquots of free enzyme, kept under the storage conditions, were exhaustively dialyzed against buffer 1 and 8 respectively, and concentrated to 1.0 mg mL^{-1} before use. Following this preparation, one set of enzyme solutions was maintained at room temperature and a second at 4°C . The enzymatic activity was evaluated daily at saturating concentrations of the substrate and cofactor.

With buffer 1, 99% of the enzymatic activity was retained after 24 h and 57% after 120 h, Fig. 2. Temperature had no significant effect. When buffer 8 was used, 55% of the enzymatic activity was retained after dialysis, and after 48 h, the calculated activity had fallen to 25% (after storage at 4°C) and 7% (after storage at room temperature), Fig. 2. Thus, both buffer composition and temperature had a significant effect on the enzymatic activity, and buffer 1 was utilized for the studies of the activity of free and immobilized *T. cruzi* GAPDH.

Multidimensional on-line GAPDH-IMER chromatographic system

In standard GAPDH assays, enzymatic activity is monitored by measuring the formation of NADH using UV-Vis detection. However, in the IMER format, a chromatographic separation of NAD^+ from NADH is necessary. As previously observed,¹¹ the GAPDH-IMER does not have sufficient chromatographic selectivity to achieve this separation and multidimensional chromatography is required. The column-switching system used in this study is illustrated in Fig. 1.

In our previous work, the best resolution of NAD^+ and NADH was achieved using a diol column.¹¹ However, the use of buffer 1 with the *T. cruzi* GAPDH-IMER altered the chromatographic selectivity of the diol column and efficient resolution of NAD^+ and NADH was not achieved. Therefore, a variety of analytical columns and chromatographic conditions was evaluated. The best chromatographic selectivity for

Fig. 2 Effect of buffers 1 and 8 on the free *T. cruzi* GAPDH enzymatic activity kept at two different temperatures and measured as the absorbance of NADH produced.

Fig. 3 (---) Representative chromatogram of the separation of NAD^+ and NADH by an octyl analytical column coupled to the GAPDH-IMER. (—) Representative chromatogram of the on-line reduction of NAD^+ . D-GA3P (7.5 mM), NAD^+ (10 mM). Separation obtained by an octyl analytical column coupled to the GAPDH-IMER, $\lambda_{\text{max}} = 340$ nm. Experimental conditions as described in the text.

NAD^+ and NADH was obtained using an octyl column using buffer 6 as the mobile phase and a flow rate of 0.8 mL min^{-1} .

In the optimized system, the retention factors (k) of NAD^+ and NADH were 0.81 and 1.55 respectively, with a separation factor (α) of 5.2 and a resolution (R_s) of 12.1 (Fig. 3). The identity of the NADH peak was confirmed by injecting separated NADH and NAD^+ standard solutions, at the same chromatographic conditions, and comparing retention times.

Method validation

NADH production by the *T. cruzi* GAPDH-IMER was evaluated using peak areas. A standard curve was constructed using NADH solutions ranging from 5 to $320 \mu\text{M}$ and a linear correlation was observed between injected concentration and peak area ($y = 0.000287122x + 2.54127$; $r = 0.99987$). Good precision and accuracy were obtained for triplicate analyses as the coefficient of variation (CV) ranged from 0.600 to 2.07% and the accuracy from 97 to 115%. The intra- and inter-day precision and accuracy of the method were determined by analyzing five replicates of three quality controls on three non-consecutive days. Precision was expressed as the CV and the accuracy was evaluated by back-calculation and expressed as the percentage deviation between the amount found and the amount prepared in the three concentrations examined. These results are shown in Table 2. The limit of quantification was of

Fig. 4 Stability of the free enzyme *T. cruzi* GAPDH in three different buffers.

$5.0 \mu\text{mol L}^{-1}$ (CV = 2.07%, and accuracy 115%, $n = 3$) while the limit of detection was $2.0 \mu\text{mol L}^{-1}$.

Immobilization conditions for *T. cruzi* GAPDH

Based on previous evaluation for the immobilization of human GAPDH enzyme,¹¹ potassium phosphate buffer (50 mM, pH 8.6) was used as the initial immobilization buffer, but with poor results. These results reflected the stability of *T. cruzi* GAPDH in this buffer as only 50% of the initial enzymatic activity was retained after 24 h.

Enzymatic stability is a key issue since, unlike human GAPDH enzyme, which is used as a lyophilized powder,¹¹ *T. cruzi* GAPDH is stored in triethanolamine. Since glutaraldehyde is used in the immobilization step, the enzyme must be dialyzed into a buffer that is free of reactive amino moieties. Therefore, the enzyme needs to be stable in the buffer during the dialysis and immobilization procedures.

In the selection of the immobilization buffer, two important characteristics were considered: (1) the buffer should not affect enzymatic activity during dialysis and; (2) the buffer should not react with glutaraldehyde during the immobilization process. Using these criteria, buffers 3, 4 and 5 were selected for study.

T. cruzi GAPDH in the storage buffer was dialyzed against each of the buffers followed by concentration to 1.0 mg mL^{-1} . The resulting solutions were evaluated for residual enzymatic activity immediately after dialysis and after storage for 4 and 7 h at room temperature and 4°C , Fig. 4.

The results demonstrate that buffer 3 produced the best results and it was used in the immobilization studies. The stability of *T. cruzi* in buffer 3 was consistent with the use of this buffer in previous studies of this enzyme.¹²

Previous studies on the development of an IMER containing human GAPDH¹¹ demonstrated that optimum activity and stability was obtained using a fused silica capillary. This approach was also used in this study. The initial step in the immobilization involves the cleaning and activation of the

Table 2 Accuracy and intra- ($n = 5$) and inter-day ($n = 3$) precision for the assay of quantification of the NADH . [GAPDH-IMER ($50.0 \text{ cm} \times 0.10 \text{ mM I.D.}$)]

Concentration/ $\mu\text{mol L}^{-1}$	First day		Second day		Third day		Pooled ($n = 15$) RSD (%)
	RSD (%)	Accuracy (%)	RSD (%)	Accuracy (%)	RSD (%)	Accuracy (%)	
12	0.600	108	2.25	96.5	1.03	98.2	1.29
240	1.13	108	0.370	97.3	1.55	102	1.01
300	1.35	102	2.73	94.1	6.57	103	3.55

capillary's surface, which can be accomplished using NaOH or HCl.^{13,14}

For the immobilization of GAPDH from *T. cruzi*, both conditions were investigated with the best results obtained when 2.0 mol L⁻¹ HCl was used which was then selected to activate the capillaries.

The amount of *T. cruzi* GAPDH immobilized on the surface of the capillary was estimated by the difference in the absorbance at 210 and 280 nm of the solution of *T. cruzi* GAPDH passed through the activated capillary. The results indicate that between 160 and 180 µg (16–18% of the original protein content) was immobilized on a 50 cm capillary. The activity of the immobilized *T. cruzi* GAPDH was established by injecting the substrate, D-GA3P, and cofactor, NAD⁺, onto the *T. cruzi* GAPDH-IMER and measuring the resulting NADH, cf. Fig. 3. These results indicate that the immobilized *T. cruzi* GAPDH retained its enzymatic activity.

Stability and storage of GAPDH-IMERS

The activity of a *T. cruzi* GAPDH-IMER was assessed daily over a 35 day period. During the first 10 days, the IMER activity dropped to 10% of the original level and then remained stable for the remainder of the test period. The results indicated that in the IMER format, the stability of *T. cruzi* GAPDH was increased from hours to at least one month, when compared to the free enzyme. The high enzymatic activity and stability permitted the performance of several on-line kinetic studies.

When the *T. cruzi* GAPDH-IMERS were not in use they were stored at 4 °C. The effect of storage on the enzymatic activity was examined using a second IMER that was used, then washed and stored for five days. This procedure was repeated during the same period of time that the first was in daily use. There were no significant differences in the time *versus* activity profiles of the two IMERS, indicating that the stability of the immobilized enzyme was independent of both use and storage. One possible explanation of these results is that only 10% of the estimated enzyme was actually covalently immobilized on the surface of the capillary and that the loss of activity reflects bleeding of the *T. cruzi* from the IMER.

Immobilization reproducibility

The reproducibility of the enzymatic activity in the *T. cruzi* GAPDH-IMER was investigated by simultaneously preparing six IMERS. The resulting IMERS exhibited nearly identical activities with production of NADH ranging from 360 to 270 µM.

Using this new protocol a human GAPDH-IMER was also prepared, but no increase in activity was observed when compared to the previously reported method. Thus, for the immobilization of human GAPDH the previously reported¹¹ conditions were maintained since the stability of the IMER was huge under the reported conditions.

Kinetics studies of the free and immobilized *T. cruzi* GAPDH

The Michaelis–Menten constants K_m of the IMER format (immobilized) were determined for both substrate and cofactor

Table 3 Determination of K_m values for immobilized *T. cruzi* GAPDH by non-linear curve-fitting regression analysis^a

System	$K_m/\mu\text{M}$ (D-GA3P)	$K_m/\mu\text{M}$ (NAD ⁺)
GAPDH-IMER (<i>T. cruzi</i>)	500 ± 50	674 ± 80
Free GAPDH (<i>T. cruzi</i>)	425 ± 17	258 ± 150
Ratio GAPDH-IMER/free GAPDH	1.18	2.61

^a The data shown are representative of three independent experiments (mean ± standard error of the mean).

following the experimental conditions and described in Table 1, respectively. The enzymatic activity of soluble GAPDH was evaluated by measuring the formation of NADH using the spectrophotometric method, as described in the Experimental section.

Non-linear curve-fitting regression analysis was applied in order to determine the K_m values from the collected experimental data. The results are presented in Table 3.

The data demonstrate that the immobilization did not affect the affinity of the immobilized GAPDH relative to the free enzyme in solution. In the case of NAD⁺, the K_m value for the immobilized enzyme was about two-fold higher than that measured for the enzyme in solution. However, when GA3P was used as the substrate, the K_m values were almost the same for the immobilized and the free enzyme. This indicates that the immobilization process had a slightly more pronounced effect on the binding of the cofactor.

These results differ from the previous ones employing the human GAPDH-IMER system.¹¹ Probably, because of the restriction in flexibility of structural components involved in the catalytic mechanism, the immobilization of human GAPDH reduced the binding-affinity for the substrate and the cofactor.

It is important to emphasize that the structural requirements for the binding of small molecules (inhibitor candidates of small molecular mass) to the target protein in the GAPDH-IMER are conserved as proved by the experiments carried out in this work. The covalent immobilization of the GAPDH enzyme not only retained the enzymatic activity, but also increased the enzyme stability. These are important achievements, allowing the biological screening of inhibitor candidates with improved accuracy and reproducibility. The differences in the K_m values for the free GAPDH and GAPDH-IMER are a consequence of the conformational changes caused by the immobilization of the receptor target. However, the GAPDH-IMER system has retained the structural requirements for the search of competitive inhibitors at both ligand sites (NAD⁺ and GA3P), as indicated by the kinetic studies on the free GAPDH and immobilized enzyme.

The comparison of the immobilization effect, for human and *T. Cruzii* GAPDH-IMERS will be discussed elsewhere.

Conclusions

In the present work, the process of enzyme immobilization, which resulted in the GAPDH-IMER, is of substantial interest in drug discovery. Considering the high costs and difficulties regarding the purification of enzymes (GAPDH and others, from human or other organisms), this technique represents a

useful way to preserve enzyme activity for a high number of enzyme assays. We consider this an important advance for the screening of synthetic and natural products in the search for new bioactive substances. This approach can also be useful for other target enzymes, considering that the structural requirements for the binding of substrates and small-molecule modulators could be preserved at both molecular and protein levels.

Acknowledgements

This work was funded by grants of the São Paulo State Research Foundation (FAPESP). Q. B. C and C. L. C acknowledge CNPq and FAPESP for research and post-doctoral fellowship.

References

- 1 World Health Organization Statistical Information System Web-site, at <http://www.who.ch/whosis/whosis.htm> (accessed 25 September 2007).
- 2 J. A. Urbina and R. Docampo, *Trends Parasitol.*, 2003, **19**, 495–501.
- 3 J. R. Coura and S. L. Castro, *Mem. Inst. Oswaldo Cruz*, 2002, **97**, 3–24.
- 4 F. R. Opperdoes, *Annu. Rev. Microbiol.*, 1987, **41**, 127–151.
- 5 F. R. Opperdoes and P. A. M. Michels, *Int. J. Parasitol.*, 2001, **31**, 482–490.
- 6 B. M. Bakker, P. A. M. Michels, F. R. Opperdoes and H. V. Westerhoff, *J. Biol. Chem.*, 1999, **274**, 14551–14559.
- 7 J. J. Harris and M. Walters, in *Enzymes*, ed. P. D. Boyer, Academic Press, New York, 1976, vol. 13, p.1.
- 8 B. M. Bakker, H. V. Westerhoff, F. R. Opperdoes and P. A. M. Michels, *Mol. Biochem. Parasitol.*, 2000, **106**, 1–10.
- 9 D. H. F. Souza, R. C. Garratt, A. P. U. Araújo, B. G. Guimarães, W. D. P. Jesus, P. A. M. Michelis, V. Hannaert and G. Oliva, *FEBS Lett.*, 1998, **424**, 131–135.
- 10 A. M. Girelli and E. Mattei, *J. Chromatogr., B*, 2005, **819**, 3–16.
- 11 C. L. Cardoso, V. V. Lima, A. Zottis, G. Oliva, A. D. Andricopulo, I. W. Wainer, R. Moaddel and Q. B. Cass, *J. Chromatogr., A*, 2006, **1120**, 151–157.
- 12 E. C. Schirmer, *J. Biol. Chem.*, 1998, **273**, 15546–15552.
- 13 Y. Shi and S. R. Crouch, *Anal. Chim. Acta*, 1999, **381**, 165–171.
- 14 Q. Yang, X.-Y. Liu, J. Miyake and H. Toyotama, *Supramol. Sci.*, 1998, **5**, 769–772.