

UFRRJ
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

TESE

**Desenvolvimento de Modelos Empíricos de Predição da
Seletividade e da Atividade de Inibidores da Fosfatase
Shp2 Utilizando o Método Semi-Empírico PM7**

SHEISI FONSECA LEITE DA SILVA ROCHA

2019



UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO

INSTITUTO DE QUÍMICA

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**Desenvolvimento de Modelos Empíricos de Predição da
Seletividade e da Atividade de Inibidores da Fosfatase
Shp2 Utilizando o Método Semi-Empírico PM7**

SHEISI FONSECA LEITE DA SILVA ROCHA

Sob a Orientação do Professor

Dr. Carlos Mauricio Rabello de Sant'Anna

e Coorientação da Professora

Dra. Cristiane Martins Cardoso de Salles

Tese submetida como requisito parcial
para obtenção do grau de Doutora em
Química, no Programa de Pós-
Graduação em Química, Área de
Concentração em Química.

Seropédica, RJ

Janeiro de 2019

R672d Rocha, Sheisi Fonseca Leite da Silva, 1987-
Desenvolvimento de Modelos Empíricos de Predição da Seletividade e da Atividade de Inibidores da Fosfatase Shp2 Utilizando o Método Semi-Empírico PM7 / Sheisi Fonseca Leite da Silva Rocha. - 2019. 108 f.

Orientador: Carlos Mauricio Rabello de Sant'Anna.
Coorientador: Cristiane Martins Cardoso de Salles.
Tese(Doutorado). -- Universidade Federal Rural do Rio de Janeiro, Programa de Pós Graduação em Química, 2019.

1. Shp2. 2. Seletividade. 3. Docagem. 4. PM7. 5. Efeito eletrolítico. I. Sant'Anna, Carlos Mauricio Rabello de , 1965-, orient. II. Salles, Cristiane Martins Cardoso de , -, coorient. III Universidade Federal Rural do Rio de Janeiro. Programa de Pós Graduação em Química. IV. Título.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

SHEISI FONSECA LEITE DA SILVA ROCHA

Tese submetida como requisito parcial para obtenção do grau de Doutora em Química, no Programa de Pós-Graduação em Química, Área de Concentração em Química.

Tese aprovada em 29 de janeiro de 2019.

Prof. Dr. Carlos Mauricio Rabello de Sant'Anna, UFRRJ
(Orientador)

Prof. Dr. Glauco Favilla Bauerfeldt, UFRRJ

Profa. Dra. Cristina Maria Barra, UFRRJ

Profa. Dra. Nelilma Correia Romeiro, UFRJ

Prof. Dr. Harold Hilarion Fokoue, UFRJ

“Seja bendito o nome de Deus para todo o sempre, porque Dele é a sabedoria e a força.”

(Bíblia Sagrada, Daniel 2:20)

AGRADECIMENTOS

Agradeço a Deus por ter me capacitado e por ter me fortalecido todos os dias. Por ter me dado sabedoria e paciência. Sem Seu fiel cuidado e amor incomparável eu não teria conseguido ir até o fim.

Ao meu esposo, José Geraldo (Kiko), pela amizade e pela empatia. Por todo apoio, carinho e compreensão. Por estar sempre ao meu lado, me incentivando e sempre me ensinando um pouco de química também. Pelas conversas, pelas ideias e por sempre me transmitir esperança. Eu também agradeço aos meus filhos Gabriel e João Pedro por terem suportado comigo uma das partes mais difíceis deste doutorado: as horas de ficarmos longe; mas também por terem tornado esses anos muito mais divertidos. Sem vocês esta conquista não seria tão grande para mim.

À minha mãe, que de forma doce e amorosa sempre se doou e sempre esteve pronta a me socorrer. Por sempre ter investido nos meus estudos e me incentivado a ir além. À minha irmã Shênia, meu cunhado Robson e sobrinho Lucas pela amizade, incentivo e torcida. Aos meus cunhados Danielle e Rodrigo e aos meus sobrinhos, Aleph, Rebecca e Kaleb pela amizade, pelos momentos de descontração e de incentivo.

À Ana Helena, pelo cuidado e carinho dedicado aos meus filhos durante este período, por não ser apenas uma babá, mas uma grande amiga de nossa família. Aos meus amigos da pós-graduação (Luana, Erika, Victor, Andreza, Carol, Gabriel e Renan) pela amizade, apoio e agradável convivência. Aos meus pastores e irmãos em Cristo, pelo cuidado, incentivo, conselhos e orações.

Ao meu professor e orientador Carlos Maurício, por todo conhecimento transmitido ao longo destes anos. Pela amizade, compreensão, paciência, gentileza e orientação segura. À professora Cristiane pela coorientação, pelo esforço dedicado e pelo conhecimento prático transmitido. A todos do LabMol/UFRRJ (Catarina, Larissa, Carolina, Nataly, Bruno, Daniel, Alessandra e Luan) pela amizade e por todo apoio prestado durante esta jornada.

Aos docentes do Instituto de Química, pelo incentivo e por todo conhecimento transmitido, dentro e fora da sala de aula. Agradeço também aos professores da banca por aceitarem o convite e pelas contribuições realizadas.

Ao CNPq pela concessão da bolsa de estudos e aos órgãos de fomento CAPES, INCT-INOVAR e FAPERJ.

Ao PPGQ/UFRRJ por proporcionar este acontecimento.

RESUMO

ROCHA, S. F. L. S. Desenvolvimento de um Modelo Empírico de Predição da Seletividade e da Atividade de Inibidores da Shp2 Utilizando o Método Semi-Empírico PM7. 2018. 94p. Tese (Doutorado em Química). Instituto de Química, Departamento de Química Orgânica, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2018.

A Shp2, juntamente com a Shp1, forma uma pequena família de proteínas tirosina fosfatases. Estudos sugerem que, embora a inibição da Shp2 seja vantajosa para o tratamento de alguns tipos de câncer, a inibição da Shp1 pode ter o efeito oposto, pois atua como supressora de tumores. Desta forma, buscou-se desenvolver uma metodologia *in silico* capaz de identificar inibidores da Shp2 mais seletivos. Neste trabalho, mostramos que apesar da complexidade termodinâmica envolvida na interação enzima/inibidor, foi possível correlacionar a seletividade de duas séries (76 compostos) com a diferença das entalpias de interação calculadas em ambas as enzimas. Os perfis de interação dos inibidores com a Shp2 e a Shp1 foram inicialmente obtidos por docagem molecular. Após o refinamento das geometrias dos complexos enzima/inibidor com o método do orbital molecular semi-empírico PM7, foram obtidos os valores de entalpia de interação. Para a série 1, composta por 52 inibidores seletivos da Shp2, demonstramos que a entalpia de interação pode ser usada como um critério confiável para a identificação de inibidores seletivos para a Shp2, pois foi significativamente mais favorável para Shp2 do que para a Shp1 com um nível de confiança de 99%. Para a série 2, composta por 24 compostos, uma correlação satisfatória ($R = 0,70$) pôde ser obtida entre a seletividade e a diferença percentual relativa das entalpias de interação calculadas em ambas as enzimas. Outro objetivo deste trabalho foi construir um modelo de predição da atividade de inibidores da Shp2 utilizando como base empírica a série 1 e posteriormente, validar com a série 2. Devido à presença de inibidores carregados negativamente dentro das séries estudadas, foi necessário considerar o efeito eletrolítico, corrigindo os valores experimentais de atividade inibitória (CI_{50}), uma vez que tais dados se referem a concentrações formais e a constante termodinâmica envolve concentrações efetivas. Para isso foi necessário calcular a força iônica do meio reacional e estimar os coeficientes de atividade das espécies envolvidas no equilíbrio de dissociação enzima/inibidor através da equação de Guntelberg. A construção do modelo se baseou em propostas da literatura sobre o uso de ciclos termodinâmicos para se calcular a energia livre de interação entre ligantes e enzimas. Neste sentido, foram obtidos termos referentes à entalpia de interação do complexo enzima/inibidor, a energia de solvatação do ligante e as perdas entrópicas devido a restrições rotacionais após a interação do mesmo com a enzima. Estes termos foram correlacionados através de regressão múltipla linear com dados experimentais de inibição. Desta forma foi possível desenvolver um modelo de predição da atividade de inibidores da Shp2 com boa correlação com dados experimentais ($R = 0,83$). Este modelo foi validado de forma satisfatória ($R = 0,73$) através da série 2 e utilizado na predição da atividade relativa de novos compostos.

Palavras-chave: Shp2, Seletividade, Docagem, PM7, Efeito eletrolítico.

ABSTRACT

Shp2, along with Shp1, forms a small family of protein tyrosine phosphatases. Studies suggest that although inhibition of Shp2 is advantageous for the treatment of some types of cancer, inhibition of Shp1 may have the opposite effect because it acts as a tumor suppressor. In this way, we sought to develop an *in silico* methodology capable of identifying more selective Shp2 inhibitors. In this work, we showed that in spite of the thermodynamic complexity involved in the enzyme/inhibitor interaction, it was possible to correlate the selectivity of two series (76 compounds) with the difference of the enthalpy of interaction calculated in both enzymes. The interaction profile of the inhibitors with Shp2 and Shp1 were initially obtained by molecular docking. After the refinement of the geometries of the enzyme / inhibitor complexes with the semi-empirical molecular orbital PM7 method, the enthalpy values of the interaction were obtained. For the series 1, composed of 52 selective inhibitors of Shp2, we demonstrated that the enthalpy of interaction can be used as a reliable criterion for the identification of selective inhibitors for Shp2, since it was significantly more favorable for Shp2 than for Shp1 with a confidence level of 99%. For series 2, composed of 24 compounds, a satisfactory correlation ($R = 0.70$) could be obtained between the selectivity and the relative percentage difference of the calculated enthalpies of interaction in both enzymes. Another objective of this work was to construct a model of prediction of the activity of inhibitors of Shp2 using as empirical basis the series 1 to validate it later with the series 2. Due to the presence of negatively charged inhibitors within the series, it was necessary to consider the electrolytic effect, correcting the experimental values of inhibitory activity, since such data refer to formal concentrations and the thermodynamic constant involves effective concentrations. For this it was necessary to calculate the ionic strength of the reaction medium and to estimate the activity coefficients of the species involved in the enzyme /inhibitor dissociation equilibrium through the Guntelberg equation. The construction of the model was based on literature proposals on the use of thermodynamic cycles to calculate the free energy of interaction between ligands and enzymes. In this sense, terms related to the enthalpy of interaction of the enzyme / inhibitor complex, the energy of solvation of the ligand and the entropic losses due to rotational restrictions were obtained after their interaction with the enzyme. These terms were correlated through linear multiple regression with experimental data of inhibition. In this way it was possible to develop a prediction model of the activity of inhibitors of Shp2 with good correlation with experimental data ($R = 0.83$). This model was validated satisfactorily ($R = 0.73$) with series 2 and used in the prediction of the relative activity of new compounds.

Keywords: Shp2, Selectivity, Docking, PM7, Electrolytic effect.

ÍNDICE DE TABELAS

Tabela 1. Grupamentos R ₁ e R ₂ dos derivados da indolina (série 1) e seus respectivos valores de <i>CI</i> ₅₀ experimentais para a Shp2.	36
Tabela 2. Grupamentos R ₁ e R ₂ para os derivados do oxindol (série 2) juntamente com os valores experimentais de <i>CI</i> ₅₀ para a Shp2 e para a Shp1, em µmol/L.	38
Tabela 3. Resíduos de aminoácidos usados para compor os sítios de interação da Shp2 e da Shp1	40
Tabela 4. Grupamento R e Ar dos compostos <i>N</i> -arilpiperazínicos (série 3).	50
Tabela 5. Pontuações obtidas através da docagem molecular com a função <i>Chemscore</i> para os compostos da série 1 na Shp1 e na Shp2.	55
Tabela 6. Entalpias de formação (H _f) em kcal mol ⁻¹ , para os ligantes da série 1 e para os complexos sítio-ligante obtidos com a Shp2 (4RDD) e a Shp1 (4GS0)	56
Tabela 7. Entalpias de interação (ΔH _{int}) calculadas pelo método PM7, em kcal mol ⁻¹ , para os complexos da série 1 obtidos com a Shp2 (4RDD) e com a Shp1 (4GS0).	57
Tabela 8. Entalpias de formação (H _f) em kcal mol ⁻¹ , para os ligantes da série 2 e os complexos sítio-ligante obtidos com a Shp2 (4RDD) e a Shp1 (4GS0)	60
Tabela 9. Entalpias de interação (ΔH _{int}) em kcal mol ⁻¹ , calculadas pelo método PM7 dos compostos da série 2 na Shp2 e na Shp1, diferença entre os valores de entalpias (ΔΔH _{int}), diferença percentual relativa (%ΔΔH _{int}) e seletividade relativa experimental (SR).....	61
Tabela 10. Contribuição de cada íon para a força iônica (μ) do meio reacional	64
Tabela 11. Coeficientes de atividades (γ) calculados através da equação de Guntelberg e a razão dos coeficientes de atividades (γ*) em relação à carga (z) dos inibidores.	65
Tabela 12. Valores experimentais de <i>CI</i> ₅₀ em µmol/L, razão dos coeficientes de atividade (γ*) e valores de p <i>CI</i> ₅₀ * para os compostos da série 1.....	65
Tabela 13. Energia de solvatação (E _{solv}) em kcal mol ⁻¹ , Entalpia de interação (ΔH _{int}) em kcal mol ⁻¹ e número de ligações rotacionáveis congeladas (N _{LRC}) para os compostos da série 1. ..	67
Tabela 14. Valores experimentais e preditos de p <i>CI</i> ₅₀ * e a diferença entre eles (Δp <i>CI</i> ₅₀ *), para a série 1.....	70
Tabela 15. Contribuição de cada termo da equação 23 para os compostos da série 1	73
Tabela 16. Valores experimentais de <i>CI</i> ₅₀ , razão dos coeficientes de atividade (γ*), valores de p <i>CI</i> ₅₀ * e dados teóricos necessários para a predição da atividade inibitória dos compostos da série 2.....	75
Tabela 17. Valores experimentais e preditos (equação 23) de p <i>CI</i> ₅₀ * e a diferença entre eles (Δp <i>CI</i> ₅₀ *), para a série 2.	76
Tabela 18. Contribuição de cada termo da equação 23 para os compostos da série 2	77
Tabela 19. Dados teóricos necessários para a predição da atividade inibitória, p <i>CI</i> ₅₀ * e <i>CI</i> ₅₀ (µmol L ⁻¹) preditos para a série 3.....	79
Tabela 20. Dados teóricos necessários para a predição da atividade inibitória, p <i>CI</i> ₅₀ * e <i>CI</i> ₅₀ preditos para os compostos propostos.	81

ÍNDICE DE FIGURAS

Figura 1. Fosforilação/desfosforilação da tirosina.	4
Figura 2. Estrutura química de alguns inibidores da PTK. Dasatinibe (A), imatinibe (B), regorafenibe (C), gefitinibe (D) e erlotinibe (E).....	5
Figura 3. Papel positivo e negativo da fosforilação da tirosina na sinalização celular.	5
Figura 4. Domínio PTP da Shp2 (código 3B7O do PDB) com as principais regiões destacadas e nomeadas. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)	8
Figura 5. Mecanismo geral para a desfosforilação de um resíduo de tirosina pelas PTP [SCOTT <i>et al.</i> , 2010].....	10
Figura 6. Auto inibição da PTP através do domínio SH2, ativação e interação com o substrato.....	11
Figura 7. Potenciais da superfície eletrostática dos domínios catalíticos da Shp2, código 3B7O (A) e da SHP-1, código 4GRY (B). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC).....	13
Figura 8. Sobreposição das estruturas cristalográficas dos domínios PTP da Shp2 na forma Apo (código do PDB 3ZM0, na cor azul) e SHP-1 na forma Apo (código do PDB 1GWZ, na cor verde). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)	14
Figura 9. Flexibilidade da alça WPD da Shp1 e rigidez da alça WPD da Shp2 (A) Shp1 (código do PDB 1GWZ em sua forma Apo) na cor verde; Shp1 (código 4GRY, com o ânion sulfato cocrystalizado) na cor rosa; Shp1 (código 4GRZ, com o ânion fosfato cocrystalizado) na cor azul; Shp1 (código 1FPR, com um peptídeo cocrystalizado) na cor laranja; Shp1 (código 4GS0, com a proteína JAK1 cocrystalizada) na cor amarela. (B) Shp2 (código 3ZM0 em sua forma Apo) na cor verde; Shp2 (código 3B7O com o ligante D-malato cocrystalizado) na cor azul; Shp2 (código 3MOW com um peptídeo cocrystalizado) na cor laranja; Shp2 (código 3O5X com derivado do ácido salicílico cocrystalizado) na cor rosa. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)	15
Figura 10. Proteína tirosina fosfatase Shp2 (código 2SHP do PDB). Domínio da PTP em azul com o P-loop em amarelo (região que contém resíduo de cisteína 459 responsável pela desfosforilação). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)....	16
Figura 11. Em resposta à fatores de crescimento, a Shp2 interage através de seus domínios SH2 com receptores autofosforilados ou com proteínas de ancoragem (como por exemplo a Gab-1, mostrada na figura) que são ativadas por receptores de tirosinas cinases ou por cinases da família Src. Tais interações resultam na ativação da Shp2 e consequentemente na ativação da Ras e de toda a via ERK1/2, levando ao crescimento celular.....	18
Figura 12. Número de estruturas cristalizadas da Shp2 depositadas nos últimos anos.....	21
Figura 13. Alguns inibidores da Shp2	23
Figura 14. (A) Estrutura geral dos derivados da indolina (série 1). (B) Estrutura geral dos derivados do oxindol (série 2)	36
Figura 15. Esquema da superfície eletrostática da Shp2 e representação da região próxima à entrada do sítio.	46

Figura 16. Ciclo termodinâmico com as principais etapas da formação do complexo proteína-ligante. O valor $\epsilon = 4,0$ refere-se à representação do restante da proteína como um meio contínuo de constante dielétrica uniforme.....	48
Figura 17. Estrutura geral dos compostos <i>N</i> -arilpiperazínicos (série 3)	50
Figura 18. Sobreposição da estrutura original da Shp2 (4RDD) com o derivado da cefsulodina cocristalizado (em tubos, átomos de C em verde) com as estruturas provenientes da redocagem com as seguintes funções de pontuação: <i>ChemPLP</i> (átomos de C em ciano), <i>Goldscore</i> (magenta), <i>Chemscore</i> (amarelo) e <i>ASP</i> (cinza).....	52
Figura 19. (A) Sobreposição dos modos de interação do composto 10 da série 1 (em tubos, carbonos em verde) com o sítio da Shp2 (4RDD) e do derivado da cefsulodina cocristalizado (em tubos, carbonos em cinza). (B) Sobreposição dos modos de interação do composto 10 da série 1 (em tubos, carbonos em verde) com o sítio da Shp1 (4GS0) e do ligante cocristalizado (em tubos, carbonos em cinza). As interações estão representadas por linhas pontilhadas pretas. Principais resíduos envolvidos nas interações em tubos, carbonos em cinza. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC).....	54
Figura 20. (A) Pose de melhor pontuação do complexo formado entre o composto 70 e a Shp2 (4RDD). (B) Pose escolhida por ser aquela que mais se assemelha ao perfil de interação do ligante cocristalizado. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)	59
Figura 21. Correlação entre a seletividade relativa (SR), baseada nos dados de CI_{50} experimentais e a diferença relativa percentual ($\% \Delta \Delta H_{int}$) das entalpias de interação dos compostos da série 2.....	62
Figura 22. (A) Cavidade entre a alça $\beta 5$ - $\beta 6$ e o resíduo de Tyr279 da alça pTyr da Shp2 (4RDD); (B) Composto 71 (carbonos em laranja) e composto 56 (carbonos em verde) interagindo com o sítio da Shp1 (4GS0); (C) Composto 56 (carbonos em verde) interagindo com o sítio da Shp2 (4RDD); (D) Composto 71 (carbonos em laranja) interagindo com o sítio da Shp2 (4RDD). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC) ..	63
Figura 23. Correlação entre ΔH_{int} e pCI_{50}^*	66
Figura 24. Composto 1 (em tubos, carbonos em verde) e composto 2 (em tubos, carbonos em laranja) interagindo com o sítio da Shp2 (4RDD). Observa-se que o grupamento R2 se localiza para fora do sítio ativo. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC).....	69
Figura 25. Correlação entre os valores experimentais de pCI_{50}^* e os valores preditos para os compostos da série 1 (equação 23)	72
Figura 26. Correlação entre os valores experimentais de pCI_{50}^* e os valores preditos (equação 23) para os compostos da série 2.	76
Figura 27. Perfil de interação do composto 77 da série 3 (carbonos em verde) com o sítio ativo da Shp2 (4RDD). Sobreposição com o composto 1 da série 1 (carbonos em laranja.	79
Figura 28. Estruturas propostas para inibidores da Shp2	80

ÍNDICE DE ABREVIACÕES E SÍMBOLOS

AG	Algoritmo Genético
AM1	Austin Model 1
AMBER	Assisted Model Building and Energy Refinement
COSMO	Conductor-like Screening Model
CI ₅₀	Concentração inibitória responsável por inibir 50% da atividade da enzima
E _{solv}	Energia de solvatação
FEP	Perturbação da energia livre
HF	Hartree-Fock
ΔH_{int}	Entalpia de interação
IT	Integração Termodinâmica
K _i	Constante de dissociação ou de inibição
LCAO	Linear Combination of Atomic Orbitals
LIE	Linear Interaction Energy
LMO	Orbitais Moleculares Localizados
MM2	Molecular Mechanics 2
MMFF	Merck Molecular Force Field
MNDO	Modified Neglect of Diatomic Overlap
MNDO/d	Modified Neglect of Diatomic Overlap, incluindo orbitais d
MOPAC	Molecular Orbital Package
NDDO	Neglect of Diatomic Differential Overlap
N _{LRC}	Número de Ligações Rotacionáveis
OM	Orbital Molecular
PDB	Protein Data Bank
PM3	Parametric Method 3
PM6	Parametric Method 6
PM7	Parametric Method 7
PMF	Potencial of Mean Force
PTK	Protein Tyrosine Kinase
PTP	Protein Tyrosine Phosphatase
RM1	Recife Model 1
RMN	Ressonância Magnética Nuclear
RMSD	Root-Mean-Squared Deviation
SEP	Superfície de Energia Potencial
SR	Seletividade Relativa
STO	Slater Type Orbitals
UHF	Unrestricted Hartree-Fock
ZDO	Zero Differential Overlap
γ	Coefficiente de atividade

SUMÁRIO

1	INTRODUÇÃO	1
2	REVISÃO DA LITERATURA.....	4
2.1	Fosforilação/desfosforilação de resíduos de tirosina	4
2.2	Classificação e estrutura das PTP	7
2.3	Mecanismo de desfosforilação das PTP e a movimentação da alça WPD	9
2.4	Proteínas Tirosina Fosfatases da família SHP	11
2.5	Principais diferenças entre a Shp2 e a Shp1	12
2.6	Estrutura e função da Shp2	15
2.7	Shp2 e o câncer	18
2.8	Alguns inibidores da Shp2	21
2.9	Modelagem molecular	23
2.10	Métodos clássicos.....	24
2.11	Métodos quânticos.....	25
2.12	Métodos semi-empíricos	28
2.13	Docagem.....	30
3	OBJETIVOS.....	34
4	METODOLOGIA	35
4.1	Redocagem.....	35
4.2	Estudo da seletividade dos inibidores da Shp2.....	35
4.3	Construção do modelo de predição da atividade de inibidores da Shp2.....	42
5	RESULTADOS E DISCUSSÃO	52
5.1	Redocagem.....	52
5.2	Estudo da seletividade dos inibidores da Shp2.....	53
5.3	Construção do modelo de predição da atividade de inibidores da Shp2.....	64
5.4	Validação do modelo de predição de atividade para inibidores da Shp2	74
5.5	Predição da atividade inibitória dos compostos <i>N</i> -arilpiperazínicos	78
6	CONCLUSÃO	83
7	REFERÊNCIAS BIBLIOGRÁFICAS	85
8	APÊNDICES	94

1 INTRODUÇÃO

As enzimas Shp2 e Shp1 formam uma pequena família de proteína tirosina fosfatases. A hiperativação da Shp2 é bastante frequente nos casos de leucemia juvenil mielomonocítica (34%) e nos casos de câncer de fígado (65,9%) e sua ativação anormal parece estar relacionada com o início do estado cancerígeno em pacientes infectados pela bactéria *Helicobacter pylori* CagA positiva [BENTIREN-ALJ *et al.*, 2004; DONG *et al.*, 2012; HAN *et al.*, 2015]. Além disso, estudos mais recentes demonstraram que a Shp2 é um dos poucos alvos já descobertos até o momento cuja inibição reativa o processo de senescência das células, desativando a proliferação celular e eliminando as células danificadas [SERRANO *et al.*, 2015; LAN *et al.*, 2015].

Muitos estudos sugerem que, embora a inibição da Shp2 possa ser benéfica no tratamento do câncer, a inibição da Shp1 adequadamente regulada pode prejudicar o tratamento, pois esta desempenha um papel de supressora de tumores, por regular negativamente a via de sinalização JAK/STAT [SHI *et al.*, 2000; QU *et al.*, 2000; ZHOU *et al.*, 2008]. Alguns inibidores da Shp2 já foram descobertos, porém, um dos principais desafios encontrados até o momento é o desenvolvimento de compostos que inibam a Shp2 sem inibir a Shp1 [LIU *et al.*, 2013; ZHANG *et al.*, 2010; WU *et al.*, 2012].

Atualmente, métodos computacionais de modelagem molecular têm sido usados com sucesso no auxílio do planejamento de candidatos a fármacos e, por isso, buscou-se explorar neste trabalho alguns deles como ferramenta para identificar inibidores mais seletivos e mais potentes para a Shp2.

O processo de reconhecimento molecular enzima/ligante é dirigido por fatores termodinâmicos que podem ser estimados através da energia livre de Gibbs da interação enzima/ligante. Esta energia de interação está diretamente relacionada à constante termodinâmica de dissociação enzima/inibidor, também chamada de constante de inibição, que representa uma razão entre as concentrações efetivas (atividades) das espécies envolvidas no equilíbrio. Porém, na determinação desta constante, os métodos analíticos comumente utilizados são espectrométricos e por isso determinam a constante de inibição em função das concentrações analíticas. Desta forma, é necessário estimar a

razão dos coeficientes de atividades das espécies envolvidas no equilíbrio a fim de se realizar a devida correção da constante de inibição.

Diferentes métodos computacionais já foram desenvolvidos a fim de se prever a constante termodinâmica de inibição através de cálculos da energia livre de interação enzima/inibidor, tais como os métodos de integração termodinâmica (IT) [KOLLMAN, 1993], da perturbação da energia livre (FEP, do inglês *free energy perturbation*) [BROOKS *et al.*, 1988] e o método da energia de interação linear (LIE do inglês *linear interaction energy*) [ASI *et al.*, 2004, HANSSON *et al.*, 1998], mas a obtenção de estimativas precisas de energia livre de Gibbs com estes métodos envolve um custo computacional bastante elevado para sistemas moleculares com grande número de átomos.

Uma maneira mais simples que pode ser explorada para se solucionar este problema é a construção de um ciclo termodinâmico representando o processo de interação enzima/ligante e o desenvolvimento de um modelo empírico baseado neste ciclo, ajustado através da regressão múltipla linear entre uma série de termos entálpicos e entrópicos calculados e valores experimentais da atividade inibitória [WANG *et al.*, 1994].

O encaixe do ligante no interior do sítio ativo da enzima é uma etapa importante para se estimar corretamente os termos entálpicos e entrópicos envolvidos no ciclo termodinâmico de interação enzima/ligante. A técnica de docagem molecular é bastante rápida e geralmente é capaz de prever com considerável precisão a geometria de um ligante ao interagir com uma determinada proteína. No entanto, as funções de pontuação usadas para quantificar a interação têm um desempenho ruim quando seus resultados são comparados com dados experimentais de bioatividade. Além da complexidade dos fenômenos associados à interação ligante-proteína, esse fraco desempenho das funções de pontuação pode estar associado, pelo menos em parte, à sua incapacidade de incorporar efeitos quânticos como a transferência de carga entre a proteína e o ligante e a ampla variação nas cargas atômicas entre diferentes estruturas do ligante e proteína [TACSAY *et al.*, 2015].

Desta forma, foi empregado neste trabalho a docagem molecular para a obtenção das estruturas iniciais dos complexos enzima/ligante de duas séries (76 compostos). Em razão das aproximações envolvidas nesta técnica, aplicou-se em seguida um método quântico aos melhores resultados obtidos por docagem, a fim de se obter resultados mais precisos tanto do ponto de vista estrutural quanto energético. Utilizou-se nesta

etapa o programa Mopac2016 (Stewart Computational Chemistry), que dispõe do método quântico semi-empírico PM7 e que, graças ao uso de orbitais moleculares localizados (LMO), possibilita a modelagem molecular com muito mais rapidez do que os métodos de orbitais moleculares *ab initio* [STEWART, 1996].

Os dados de entalpia de interação foram empregados neste trabalho de duas formas. Inicialmente, buscou-se correlacionar a seletividade relativa experimental com os valores obtidos de entalpia de interação provenientes dos complexos formados entre os inibidores e as proteínas tirosinas fosfatases Shp2 e Shp1. Esses dados também foram usados no desenvolvimento de um modelo matemático termodinamicamente consistente de predição da atividade de inibidores da Shp2, calibrado com valores experimentais de atividade inibitória. Para isto foi necessário, primeiramente, considerar o efeito eletrolítico sobre os dados experimentais de atividade inibitória que se encontravam em função de concentrações analíticas. No desenvolvimento do modelo, procurou-se combinar, além da entalpia de interação, parâmetros teóricos importantes no fenômeno de interação enzima-ligante, como a energia de solvatação e a perda entrópica ocorrida após o encaixe do ligante no sítio ativo. Finalmente, o modelo foi validado e utilizado na predição da atividade relativa de 20 novos compostos *N*-arilpiperazínicos sintetizados pelo Dr. Ciro Gonçalves e Sá no Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio®) [SÁ, 2017]. Dois novos compostos *N*-arilpiperazínicos foram então propostos.

2 REVISÃO DA LITERATURA

2.1 Fosforilação/desfosforilação de resíduos de tirosina

A fosforilação/desfosforilação de resíduos de tirosina em proteínas foi descoberta há cerca de quatro décadas e atualmente é reconhecida como o maior mecanismo regulatório de sinalização e ativação celular. É um processo dinâmico governado pela ação equilibrada das proteínas tirosinas quinases (PTK, do inglês *Protein Tyrosine Kinase*) e das proteínas tirosinas fosfatases (PTP, do inglês *Protein Tyrosine Phosphatase*) (Figura 1). Este evento representa um ponto crítico em numerosos processos fisiológicos, atuando também como um excelente interruptor no controle das principais vias de sinalização oncogênica envolvidas na apoptose, proliferação celular, migração e invasão. A fosforilação anormal dos resíduos de tirosina pode ser tanto a principal causa do câncer humano quanto pode ser necessária para a manutenção do estado cancerígeno [LABBÉ *et al.*, 2012; SCOTT *et al.*, 2010].

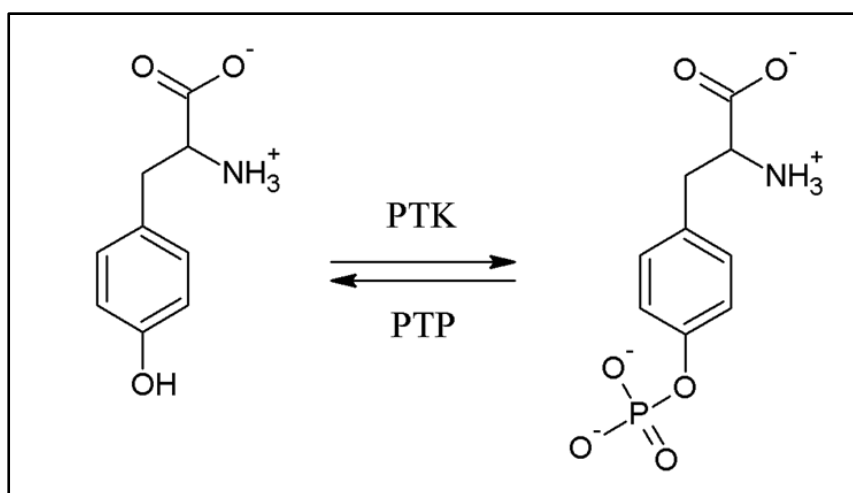


Figura 1. Fosforilação/desfosforilação da tirosina.

Nas últimas três décadas, muitos estudos demonstraram o importante papel das proteínas tirosinas quinases no câncer, justamente por ser essa uma classe de proteínas responsável por catalisar a fosforilação dos resíduos de tirosina nas proteínas. Alguns inibidores de PTK, como o dasatinibe, imatinibe, regorafenibe, gefitinibe e o erlotinibe, já são bastante conhecidos nos dias atuais e usados rotineiramente no tratamento do câncer (Figura 2) [KUREBAYASHI *et al.*, 2006, FAN *et al.*, 2014].

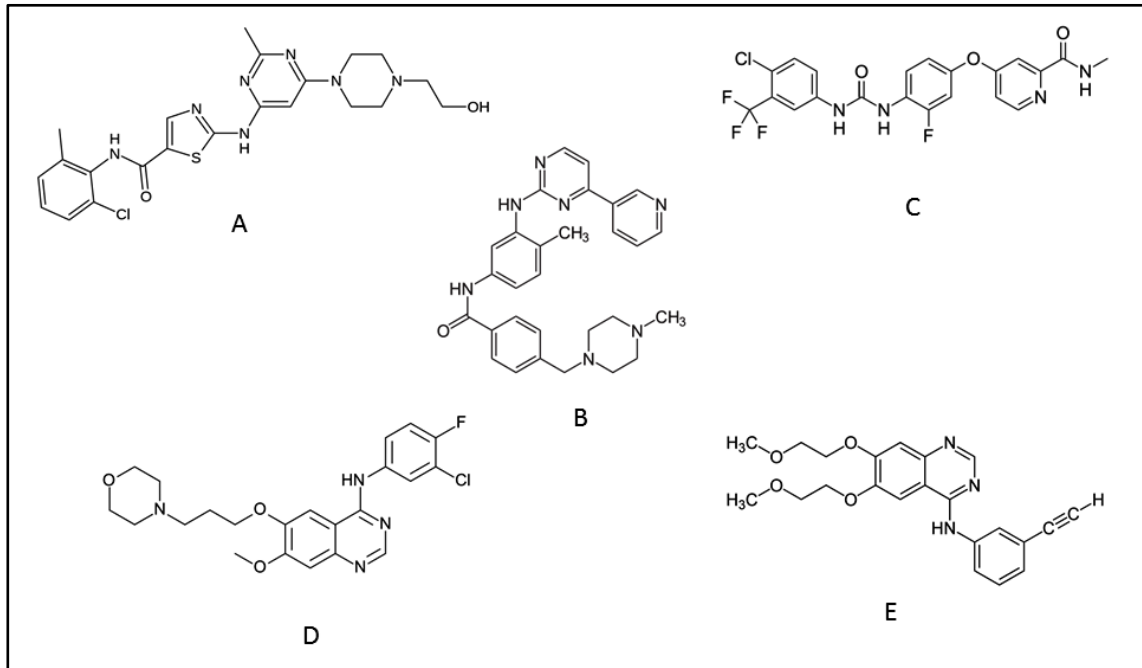


Figura 2. Estrutura química de alguns inibidores da PTK. Dasatinibe (A), imatinibe (B), regorafenibe (C), gefitinibe (D) e erlotinibe (E).

Uma vez que a ativação anormal de vários tipos de PTK está associada com diversos tipos de câncer humano, a simples observação da reação esquematizada na Figura 1 poderia fornecer um conceito errado acerca da inibição das PTP, uma vez que estas são responsáveis pela hidrólise do fosfato ligado aos resíduos de tirosina em proteínas, ou seja, por catalisar a reação inversa das PTK. O primeiro pensamento poderia ser de que as PTP colaboram para o não desenvolvimento do câncer e que sua inibição poderia resultar no surgimento ou no avanço da oncogênese. Entretanto, foi provado que este conceito está errado [TONKS, 2006]. Os sistemas biológicos são muito mais complexos do que uma simples reação química, como pode ser observado do esquema ilustrativo da Figura 3.

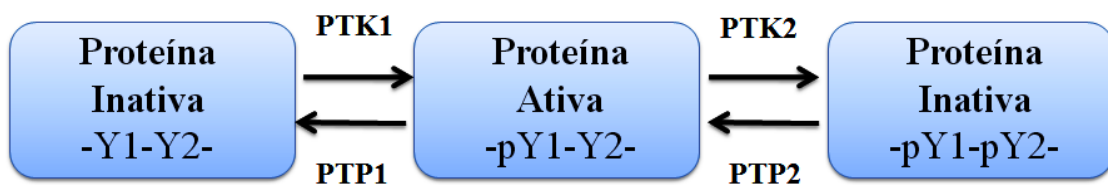


Figura 3. Papel positivo e negativo da fosforilação da tirosina na sinalização celular.

Nesta ilustração (Figura 3), Y1 e Y2 são resíduos de tirosina de uma única proteína. Enquanto que a fosforilação de Y1 aumenta a atividade da proteína, a fosforilação de Y2 juntamente com Y1 inativa a mesma. Dessa forma, a desfosforilação de Y1 pela PTP1 inativa a proteína e a desfosforilação de Y2 é necessária para sustentar a atividade da mesma. Logo, a PTP1 atua como regulador negativo e a PTP2 como regulador positivo [SCOTT *et al.*, 2010].

Sabe-se que algumas PTK fosforilam certos resíduos de tirosina de outras PTK ou de outras proteínas, desativando-as (o que chamamos de regulador negativo, ou papel negativo). Logo, a desfosforilação desses resíduos pelas PTP específicas conduz a ativação dessas PTK (o que chamamos de regulador positivo, ou papel positivo). Dessa forma, em alguns casos a inibição de uma PTP específica garante a inativação de uma PTK ou de uma proteína específica, o que poderia ser vantajoso no tratamento do câncer, caso esta esteja ativada de forma anormal (ou altamente expressa nas células cancerígenas), sendo responsável pela oncogênese ou pelo desenvolvimento do estado cancerígeno.

Crescentes evidências sugerem que, na verdade, a sinalização celular requer uma ação coordenada de ambas as enzimas (PTK e PTP). Assim, as PTP podem cooperar com as PTK como também podem antagonizá-las na promoção do crescimento do câncer.

Enquanto os inibidores das PTK já são atualmente bastante usados no tratamento do câncer, os inibidores das PTP ainda estão em fase de desenvolvimento, justamente por causa de conceitos errados acerca dessa família de enzimas. Segundo Scott e colaboradores (2010), outro dogma que contribuiu para o início lento das pesquisas em busca de inibidores das PTP foi o pensamento de que as PTK eram altamente específicas enquanto apenas algumas poucas PTP não específicas tinham a capacidade de antagonizar a função das PTK. Porém, sabe-se atualmente que muitas PTP são altamente reguladas e com ação específica em vários tipos de células humanas [SCOTT *et al.*, 2010]. Segundo revisão feita por Labbé e colaboradores (2012), a inibição de uma PTP específica pode ser benéfica ou não no tratamento de tumores e vai depender do tipo de tecido no qual se encontra o tumor [LABBÉ *et al.*, 2012]. A inibição seletiva da PTP Shp2, por exemplo, garantiria bons resultados no tratamento do câncer gástrico, da leucemia e de alguns tipos de câncer de mama [ZHANG *et al.*, 2010]. Segundo três critérios de avaliação estabelecidos por Scott e colaboradores (2010), existem atualmente pelo menos sete PTP que podem ser consideradas como alvos para o

desenvolvimento de novos fármacos anticâncer: Shp2, PTP1B, Cdc14, PRL-3, LMW-PTP, Cdc25 e Eya [SCOTT *et al.*, 2010].

No entanto, algumas PTP se comportam de fato como supressoras do tumor, justamente por catalisar a reação inversa de PTK cancerígenas. Estudos envolvendo mutações possibilitaram identificar algumas PTP com essas características: PTEN, PTPRF, PTPRG, PTPRO, PTPRT, PTPN3, PTPN6 (Shp1), PTPN13, PTPN14 e DUSP6 [ANDERSEN *et al.*, 2001]. Num estudo recente, envolvendo testes *in vivo* e mutações, foi constatado que a p-STAT3^{tyr705}, ou seja, a forma fosforilada da tirosina 705 da proteína STAT3 é altamente expressa em células cancerígenas e que a PTP Shp1 é capaz de desfosforilar a p-STAT3^{tyr705} desativando-a, ou seja, atuando como supressora do tumor. Neste mesmo estudo, os autores conseguiram comprovar que o regorafenibe (Figura 2) já usado regularmente no tratamento do câncer por inibir múltiplas PTK, também atua como ativador da PTP Shp1, o que o torna muito eficiente [FAN *et al.*, 2014].

2.2 Classificação e estrutura das PTP

O genoma humano contém 107 PTP, sendo que as PTP de classe I, baseadas em cisteínas, constituem o maior grupo. Este grupo pode ser subdividido em 61 fosfatases de dupla especificidade e 38 PTP específicas, conhecidas como clássicas. As PTP clássicas foram subdivididas em receptoras (R1-R8) e não receptoras (NT1-NT9). Estudos bioquímicos e estruturais têm contribuído para o entendimento detalhado do mecanismo catalítico das PTP. Aproximadamente 280 resíduos de aminoácidos nos domínios catalíticos das PTP consistem em estruturas em hélices α ou em folhas β . As características chaves do domínio catalítico das PTP da classe I podem ser observadas na Figura 4 e incluem a alça P, a alça móvel Trp-Pro-Asp (alça WPD, compreendida entre a folha β 11 e a hélice α 3), a porção responsável pelo reconhecimento da fosfotirosina (alça pTyr compreendida entre a hélice α 1 e a folha β 1) e a alça Q (região compreendida entre a hélice α 5 e a hélice α 6) [ALONSO *et al.*, 2004; ANDERSEN *et al.*, 2001].

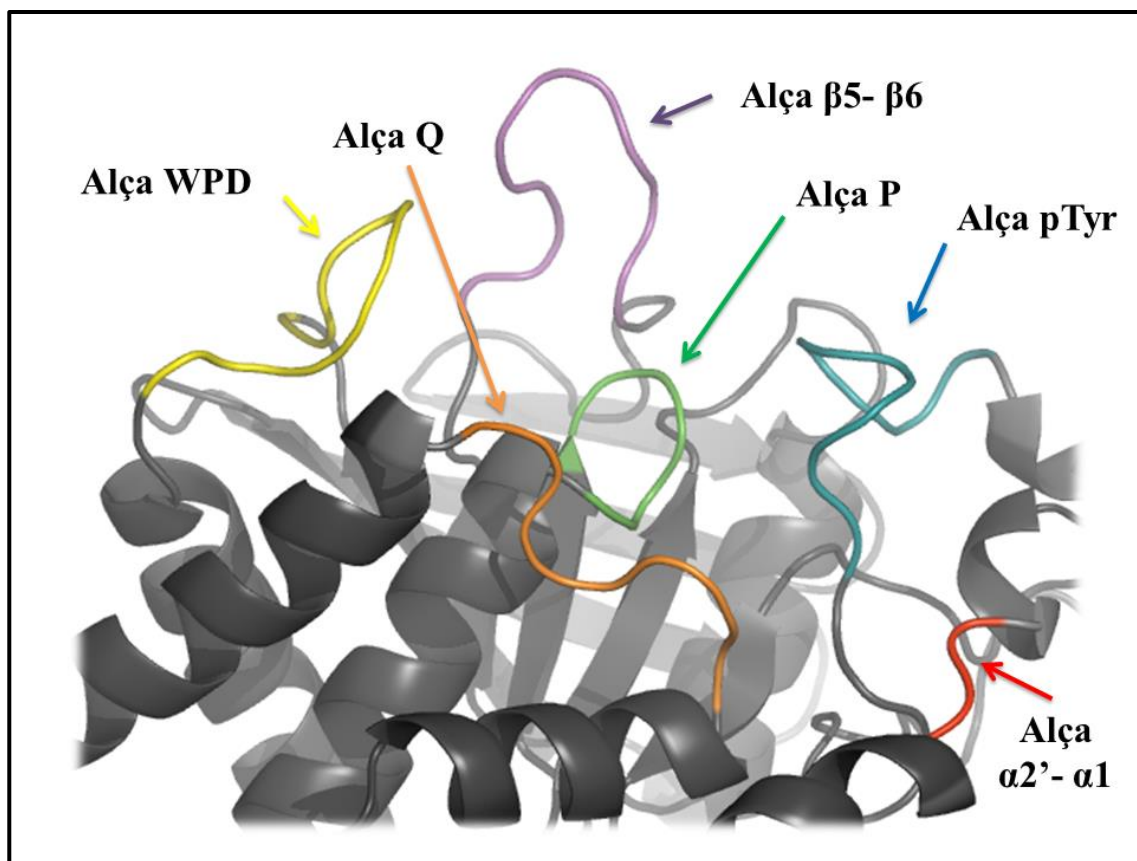


Figura 4. Domínio PTP da Shp2 (código 3B7O do PDB) com as principais regiões destacadas e nomeadas. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)

A alça P (em verde) é a porção que contém o resíduo de cisteína nucleofílico e o resíduo de arginina, responsáveis pelo reconhecimento da porção fosfatada nos substratos. A alça WPD (em amarelo) contém o resíduo de aspartato conservado, responsável pela catálise ácido/base. A alça pTyr (em azul) contribui para o reconhecimento seletivo da fosfotirosina frente a fosfoserina e a fosfotreonina. Por fim, a alça Q (em laranja) contém um resíduo conservado de glutamina responsável por posicionar e ativar uma molécula de água para a hidrólise do intermediário da reação [ALONSO *et al.*, 2004; ANDERSEN *et al.*, 2001].

A alça que liga a folha $\beta 5$ e a folha $\beta 6$ (alça $\beta 5$ - $\beta 6$, em lilás) é a região da superfície da fenda catalítica com o maior número de resíduos de aminoácidos não conservados entre as PTP e por isso possui grande influência na especificidade do substrato [YANG *et al.*, 2000].

Outra característica estrutural importante observada em algumas PTP é a existência de uma alça denominada como “segundo sítio de interação” próximo ao sítio

ativo (Figura 4). Esta alça, que conecta a hélice $\alpha 2'$ com a hélice $\alpha 1$ (em vermelho), possui a capacidade de interagir com a fosfotirosina, porém não de forma reativa. A presença deste segundo sítio de interação, identificado primeiramente na PTP1B, não depende apenas dos resíduos presentes na região da alça, mas da sua conformação. A importância deste segundo sítio de interação reside na possibilidade do desenvolvimento de inibidores bidentados altamente seletivos que se liguem simultaneamente aos dois sítios [BARR *et al.*, 2009]. Já foram identificados alguns inibidores da PTP1B com essa característica [SHEN *et al.*, 2001].

Apesar das PTP apresentarem estruturas secundárias e terciárias bastante conservadas, elas exibem uma forte especificidade por seus substratos *in vivo*. Esta especificidade está atribuída a alguns fatores, como [YANG *et al.*, 2000; SCOTT *et al.*, 2010; BARR *et al.*, 2009]:

- Superfície eletrostática ao redor do domínio catalítico;
- Conformação, flexibilidade e estabilidade da alça WPD;
- Alguns poucos resíduos de aminoácidos não conservados na superfície do domínio catalítico (principalmente na alça $\beta 5$ - $\beta 6$);
- Outros fatores regulatórios, como por exemplo, o domínio regulatório SH2 da família SHP.

2.3 Mecanismo de desfosforilação das PTP e a movimentação da alça WPD

A fosfotirosina entra na fenda catalítica através do reconhecimento feito pela alça pTyr. A alça P possui um resíduo de cisteína nucleofílico que ataca o átomo de fósforo da fosfotirosina para quebrar a ligação P-O através de uma reação S_N2 . Numa segunda etapa, o resíduo ácido de aspartato localizado na alça WPD atua como doador de próton para protonar a tirosina fenólica e liberá-la na forma de tirosina desfosforilada. Isto gera o intermediário fosforil-cisteína que é hidrolisado com auxílio do resíduo ácido de aspartato para reestabelecer o resíduo de cisteína da PTP e liberar o íon fosfato (Figura 5) [SCOTT *et al.*, 2010].

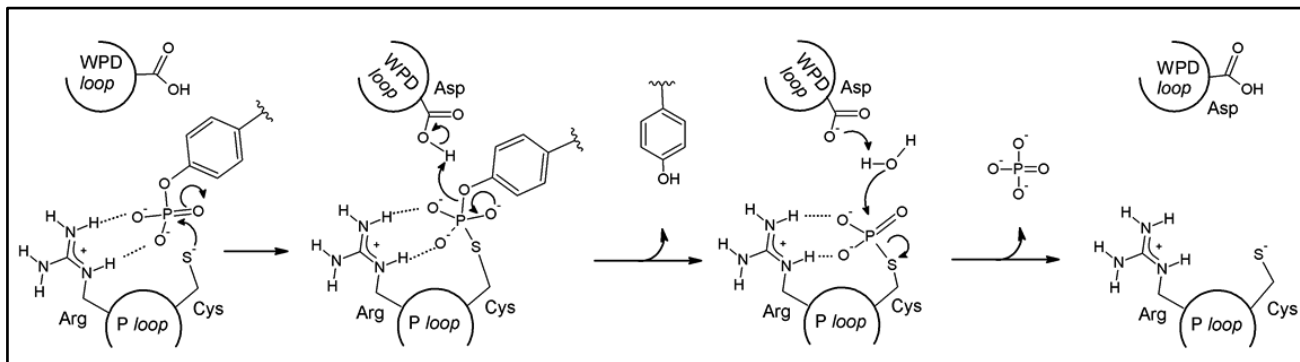


Figura 5. Mecanismo geral para a desfosforilação de um resíduo de tirosina pelas PTP [SCOTT *et al.*, 2010].

Conforme foi mencionado, a alça WPD possui um resíduo conservado de aspartato que atua como um catalisador ácido/básico na reação de desfosforilação. Esse resíduo de aspartato só fica posicionado de forma conveniente para a catálise quando a alça WPD se encontra fechada e nesta posição, diz-se que a PTP se encontra na conformação ativa.

É importante esclarecer a diferença entre “conformação ativa” e a forma “ativada”. As PTP se encontram “ativadas” quando não estão inibidas por uma molécula ou, no caso das PTP que possuem um domínio regulatório, quando este não está auto inibindo a PTP, ou seja, bloqueando a entrada do sítio ativo. Já a definição “conformação ativa” diz respeito à conformação fechada da alça WPD e a “conformação inativa”, conseqüentemente, à conformação aberta da alça WPD.

Barr e colaboradores (2009), através de uma análise estrutural envolvendo a estrutura cristalográfica de 22 PTP, identificaram quatro tipos de conformações para a alça WPD (fechado, aberto, intermediário e atípico). Os autores também observaram que uma molécula de água, próxima à alça Q, interage fortemente com dois resíduos de glutamina nas PTP cuja alça WPD se encontra fechada e que nas PTP cuja alça WPD se encontra na conformação aberta ou atípica esta molécula de água não foi observada ou se encontrava deslocada, sugerindo que ela é parte fundamental do mecanismo de fechamento da alça WPD [BARR *et al.*, 2009]. Yang e colaboradores demonstraram que a atividade catalítica das PTP é influenciada pela flexibilidade e estabilidade da alça WPD [YANG *et al.*, 1998]. Seus estudos também sugerem que a formação do intermediário fosforil-cisteína (mediante o ataque nucleofílico do resíduo de cisteína na fosfotirosina) não é a etapa lenta da reação de desfosforilação. A etapa lenta da reação

seria a hidrólise desse intermediário, por requerer o fechamento da alça WPD para o devido posicionamento do resíduo de aspartato [YANG *et al.*, 2002].

2.4 Proteínas Tirosina Fosfatases da família SHP

Fosfatases que contêm o domínio regulatório SH2 compõem uma pequena subfamília de proteína-tirosina fosfatases não receptoras, com dois membros na espécie humana, Shp1 e Shp2. Estas enzimas estão localizadas no citoplasma e possuem papel regulatório em citosinas e na sinalização de fatores de crescimento. A Shp1 é predominantemente expressa em células hematopoiéticas e epiteliais atuando basicamente como um regulador que diminui os sinais de transdução. A Shp2 é expressa ubiquamente e tem sido associada principalmente com a regulação do aumento dos eventos de sinalização celular. Estas duas enzimas contêm dois domínios SH2 acoplados ao domínio da PTP, o SH2 C-terminal (C-SH2) e o SH2 N-terminal (N-SH2). Estes domínios exercem tanto uma função auto-inibitória sobre a enzima como também são sítios de ligação para a ativação da mesma. A arquitetura geral da família SHP, além dos dois domínios SH2, também possui o domínio catalítico da PTP seguido de uma cauda C-terminal. Depois de rompida a auto-inibição através da interação de uma molécula ativadora em algum dos dois sítios do domínio SH2, há uma alteração na conformação da enzima que possibilita a entrada do substrato no domínio catalítico da PTP, onde ocorre a reação de desfosforilação (Figura 6) [ALCIEA-VELAZQUEZ & BOGON, 2013].

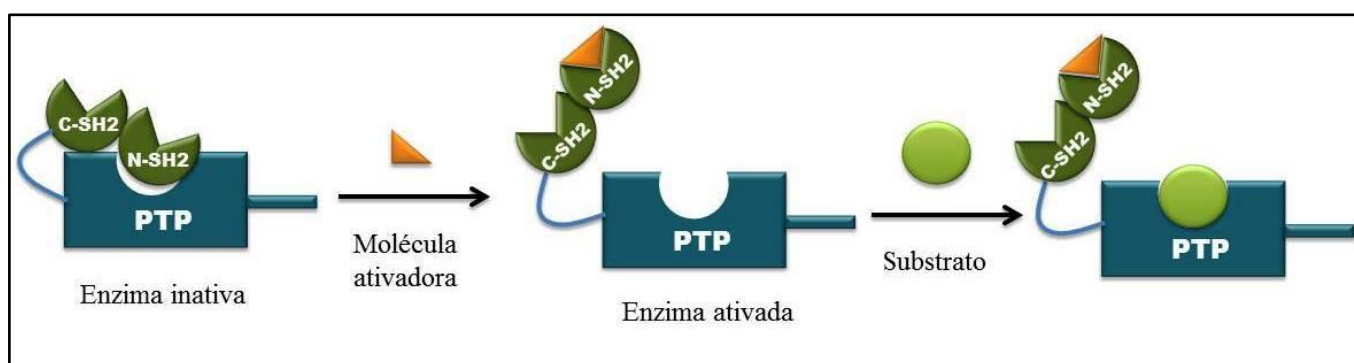


Figura 6. Auto inibição da PTP através do domínio SH2, ativação e interação com o substrato.

2.5 Principais diferenças entre a Shp2 e a Shp1

Por serem da mesma família de PTP, as enzimas Shp1 e Shp2 possuem alta similaridade estrutural (60% de identidade e 75% de similaridade), e a sobreposição dos Ca de seus domínios catalíticos mostra uma similaridade extremamente alta em relação à estrutura secundária e terciária (RMSD de 0,6 Å) [YANG *et al.*, 1998].

Apesar dessa alta similaridade, a Shp1 e a Shp2 possuem papéis bem distintos e específicos na sinalização celular. Enquanto a inibição da Shp2 poderia ser, em muitos casos, benéfica no tratamento do câncer, a inibição da Shp1 poderia prejudicar o tratamento, pois, como já foi relatado anteriormente, ela possui papel de supressora de tumor, por regular negativamente a p-STAT3^{tyr705} [FAN *et al.*, 2014]. Assim, uma estratégia correta seria o desenvolvimento de ativadores para a enzima Shp1 e não de inibidores. Apesar de não ser impossível, o desenvolvimento de pequenas moléculas ativadoras de enzimas é bem mais difícil de ser conseguido do que o desenvolvimento de inibidores. Dessa forma, atualmente, muitas pesquisas têm investido esforços no desenvolvimento de inibidores seletivos para a Shp2, que não inibam a Shp1 [YU *et al.*, 2011; LIU *et al.*, 2013].

O desenvolvimento de inibidores específicos para a Shp2 é complicado por causa da sua similaridade com a Shp1. Porém, os domínios catalíticos dessas duas enzimas possuem substratos específicos diferentes, sugerindo que a fenda catalítica não é idêntica. Além disso, a superfície eletrostática potencial da fenda catalítica é mais positiva na Shp2 do que na Shp1 humana, como pode ser visto na Figura 7, onde os potenciais eletrostáticos negativos e positivos estão representados pelas cores vermelha e azul, respectivamente. A localização do sítio catalítico está indicada pela seta amarela. [Wu *et al.*, 2011; YANG *et al.*, 1998].

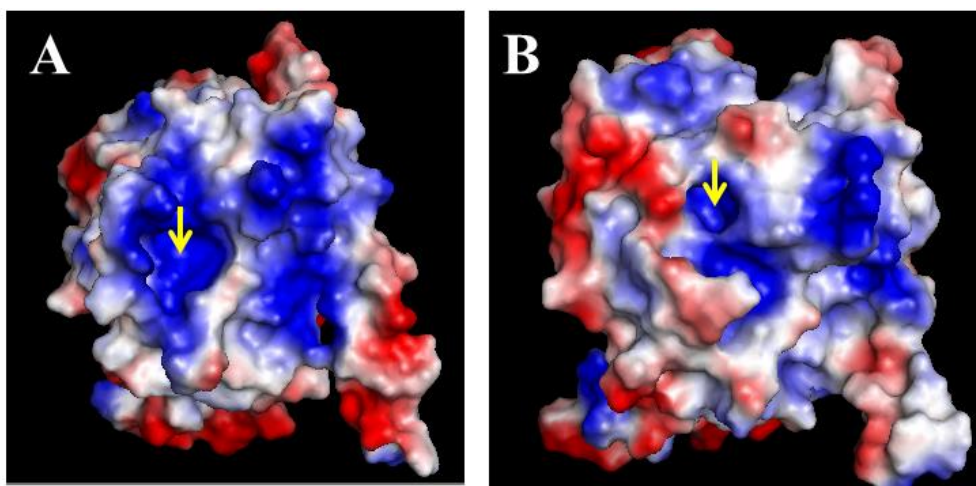


Figura 7. Potenciais da superfície eletrostática dos domínios catalíticos da Shp2, código 3B7O (A) e da SHP-1, código 4GRY (B). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)

Estudos estruturais evidenciaram que a conformação mais estável para a alça WPD da Shp2 é a conformação aberta e inativa [SCOTT *et al.*, 2010]. As evidências sugerem que, somente após a entrada do substrato no domínio catalítico da Shp2, a alça WPD se movimentaria para a posição fechada e ativa. De forma contrária, a conformação mais estável para a alça WPD da Shp1 seria a conformação fechada e ativa [YANG *et al.*, 2002]. Estudos de cinética enzimática demonstraram que o domínio catalítico da Shp1 é muito mais ativo do que o domínio catalítico da Shp2 *in vitro*, reafirmando a hipótese de que em solução aquosa o domínio catalítico da Shp2 esteja na conformação aberta (inativa) e o domínio catalítico da Shp1 esteja na conformação fechada (ativa), favorecendo a reação [NIU *et al.*, 1999]. As conformações distintas da alça WPD dessas duas PTP contribuem para as diferentes distribuições de carga superficial ao redor do sítio catalítico, como visto na Figura 7, o que provavelmente influencia na seletividade do inibidor. A análise da estrutura cristalográfica do domínio PTP da Shp1 em sua forma apo revela claramente que a alça WPD da Shp1 se encontra na conformação fechada (ativa), posicionando o resíduo de aspartato para perto da cisteína catalítica. Por outro lado, conforme pode ser observado na Figura 8, a análise da estrutura cristalográfica do domínio PTP da Shp2 em sua forma apo revela que a alça WPD se encontra aberta (inativa) e que o resíduo de aspartato se posiciona de forma desfavorável para a catálise [YANG *et al.*, 2002].

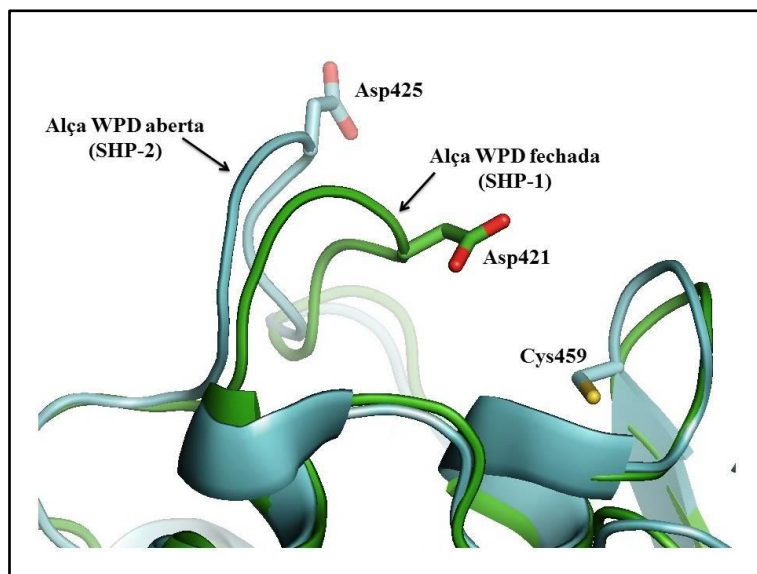


Figura 8. Sobreposição das estruturas cristalográficas dos domínios PTP da Shp2 na forma Apo (código do PDB 3ZM0, na cor azul) e SHP-1 na forma Apo (código do PDB 1GWZ, na cor verde). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)

A alça WPD da Shp1 parece ser mais flexível e ter sua conformação mais facilmente influenciada por ligantes. Como pode ser observado na Figura 9, a Shp1 cocrystalizada com uma molécula de fosfato apresenta sua alça WPD numa conformação ainda mais fechada em relação à Shp1 em sua forma apo. No entanto, outros ligantes como o sulfato e alguns peptídeos se encaixam no sítio resultando na abertura da alça WPD (Figura 9A). Por outro lado, a alça WPD da Shp2 parece ser mais rígida e sua conformação mais estável. Conforme pode ser visto na Figura 9B, a análise da conformação da alça WPD de algumas estruturas cristalográficas da Shp2 cocrystalizada com diferentes ligantes revela que não há mudança conformacional desta alça em relação à estrutura cristalográfica da Shp2 em sua forma apo [YANG *et al.*, 2002].

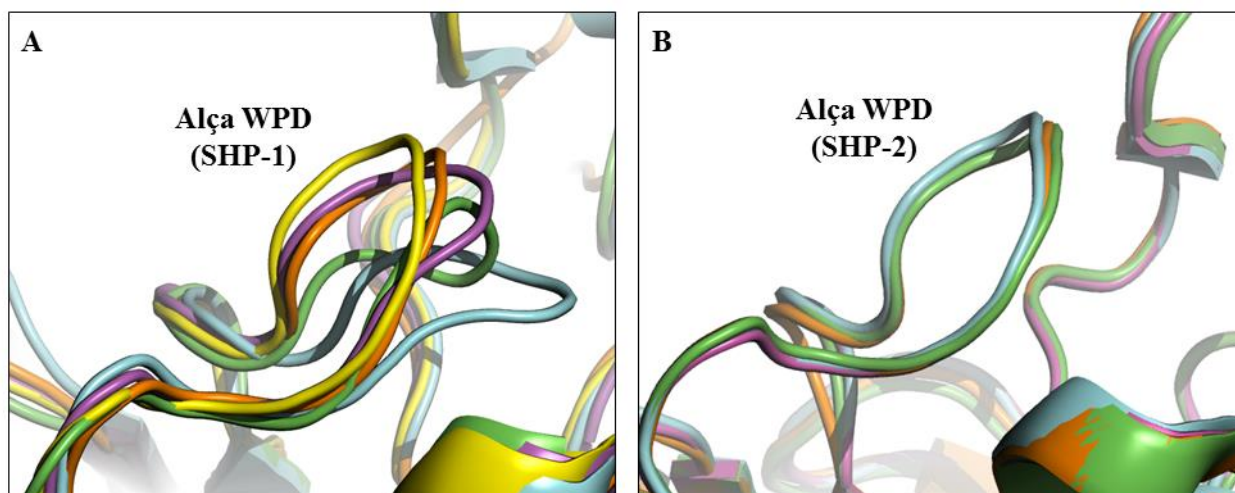


Figura 9. Flexibilidade da alça WPD da Shp1 e rigidez da alça WPD da Shp2 (A) Shp1 (código do PDB 1GWZ em sua forma Apo) na cor verde; Shp1 (código 4GRY, com o ânion sulfato cocrystalizado) na cor rosa; Shp1 (código 4GRZ, com o ânion fosfato cocrystalizado) na cor azul; Shp1 (código 1FPR, com um peptídeo cocrystalizado) na cor laranja; Shp1 (código 4GS0, com a proteína JAK1 cocrystalizada) na cor amarela. (B) Shp2 (código 3ZM0 em sua forma Apo) na cor verde; Shp2 (código 3B7O com o ligante D-malato cocrystalizado) na cor azul; Shp2 (código 3MOW com um peptídeo cocrystalizado) na cor laranja; Shp2 (código 3O5X com derivado do ácido salicílico cocrystalizado) na cor rosa. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)

Yang e colaboradores (2000) cristalizaram duas estruturas do domínio catalítico da Shp1 complexada com derivados de substratos peptídicos a fim de analisar a questão da especificidade. Os autores concluíram que a alça $\beta 5\text{-}\beta 6$ e a alça pTyr determinam a especificidade da Shp1, porém apenas a alça $\beta 5\text{-}\beta 6$ difere entre a Shp1 e a Shp2. O resíduo 360 da alça $\beta 5\text{-}\beta 6$ é uma arginina na Shp1 e uma lisina na Shp2. Segundo os autores, essa diferença é uma das principais razões para a especificidade de substratos entre estas duas enzimas. Além do resíduo 360, outros resíduos (Arg358 e Ser361) na alça $\beta 5\text{-}\beta 6$ também diferem entre estas duas enzimas e são importantes na questão da especificidade [YANG *et al.*, 2000].

2.6 Estrutura e função da Shp2

Duas estruturas diferentes de Shp2 são encontradas em humanos e roedores: uma tem 593 resíduos de aminoácidos e a outra tem 597 resíduos de aminoácidos, diferindo em 4 resíduos de aminoácidos no domínio catalítico. Estas duas Shp2 são derivadas de um único gene e acredita-se que sejam idênticas funcionalmente [SCOTT *et al.*, 2010].

A estrutura cristalográfica da Shp2 humana na forma apo (aa 1-525, sem a região C-terminal, código 2SHP do PDB) mostra que os sítios de ligação com a

fosfotirosina dos domínios SH2 se posicionam virados para fora (para o solvente), enquanto a parte posterior do domínio N-SH2 se posiciona dentro do domínio catalítico da PTP (através da alça D'E). Esta interação intra-enzimática, além de imitar algumas características da interação com o substrato, bloqueia o acesso deste ao sítio ativo (Figura 10) [HOF *et al.*, 1998].

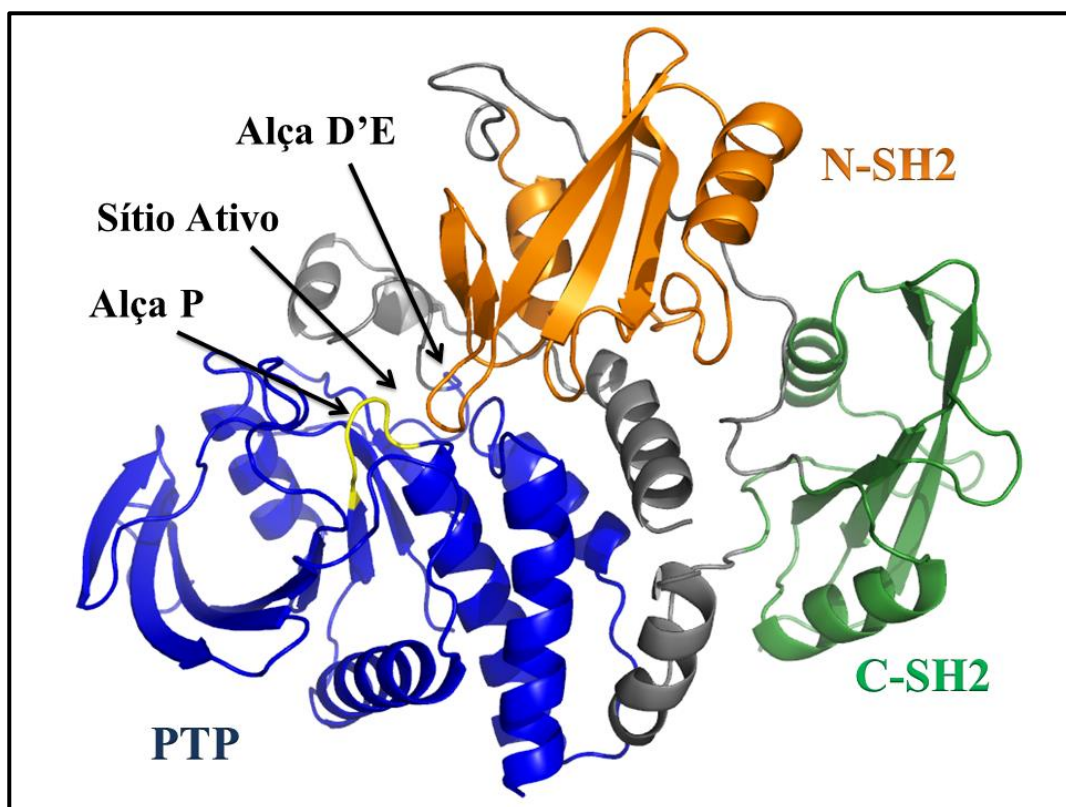


Figura 10. Proteína tirosina fosfatase Shp2 (código 2SHP do PDB). Domínio da PTP em azul com o P-loop em amarelo (região que contém resíduo de cisteína 459 responsável pela desfosforilação). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC).

Alguns ligantes polipeptídicos contendo o resíduo de fosfotirosina (pY) possuem a capacidade de ativar a Shp2 através da ligação com os domínios SH2 (principalmente com o domínio N-SH2), o que faz com que a auto inibição entre os domínios SH2 e o domínio da PTP seja rompida, havendo dessa forma a exposição do sítio ativo [TONKS *et al.*, 2006; HOF *et al.*, 1998].

Recentemente, um novo mecanismo baseado na dimerização da Shp2 foi descrito. De fato, 15% da Shp2 nas células em repouso foram encontradas na forma de dímero resultando no decaimento da atividade da fosfatase. É importante destacar que os domínios SH2 não estão envolvidos no processo de dimerização da Shp2 e que a

relação entre dímero/monômero não é estática e é regulada por fatores de crescimento e pelo estado redox da célula. A Shp2 possui um papel regulatório significativo na maioria das vias de sinalização celular e, por esse motivo, manter sua atividade sob controle é crucial para a homeostase celular [NARDOZZA *et al.*, 2012].

Em alguns estudos envolvendo a Shp2, foi realizada a troca da cisteína catalítica 459 por uma serina, a fim de se criar uma Shp2 mutante cataliticamente inerte, porém com a capacidade de interagir com seus intermediários de sinalização via domínio SH2. Nestes estudos, buscou-se compreender a função da Shp2 nas vias de sinalização e foi descoberto que a Shp2 regula positivamente a via de sinalização da insulina, fatores de crescimento da epiderme e fatores de crescimento do fibroblasto. De forma mais específica, foi constatado que a Shp2 está relacionada com a ativação das proteínas cinases ativadas por mitógenos (MAPK, do inglês *Mitogen Activated Protein kinase*) e das cinases reguladas por sinais extracelulares 1 e 2 (ERK 1/2) em resposta aos fatores de crescimento da epiderme, insulina, trombina e fibronectina [DANCE *et al.*, 2008; SHI *et al.*, 2000; QU *et al.*, 2000].

Existem fortes evidências de que a Shp2 é requerida para a ativação da via ERK/MAPK na maioria, senão em todos, os receptores de PTK, bem como pelos receptores de proteínas G e receptores de citocinas e integrinas (Figura 11). Os estudos também apontam que dentro da via que ativa a ERK1/2, a Shp2 deve atuar numa etapa anterior à proteína RAS (*rat sarcoma*), evidenciando assim a importância desta enzima para a ativação da RAS e de toda a via. A via da cinase Ras-ERK1/2 regulada pela Shp2 está envolvida na sobrevivência, proliferação, diferenciação, adesão e migração celular. A maneira como a Shp2 ativa essa via não é muito bem compreendida, porém algumas PTK receptoras envolvidas e alguns mecanismos já foram descobertos [TONKS *et al.*, 2006; SCOTT *et al.*, 2010; DANCE *et al.*, 2008; SHI *et al.*, 2000; QU *et al.*, 2000].

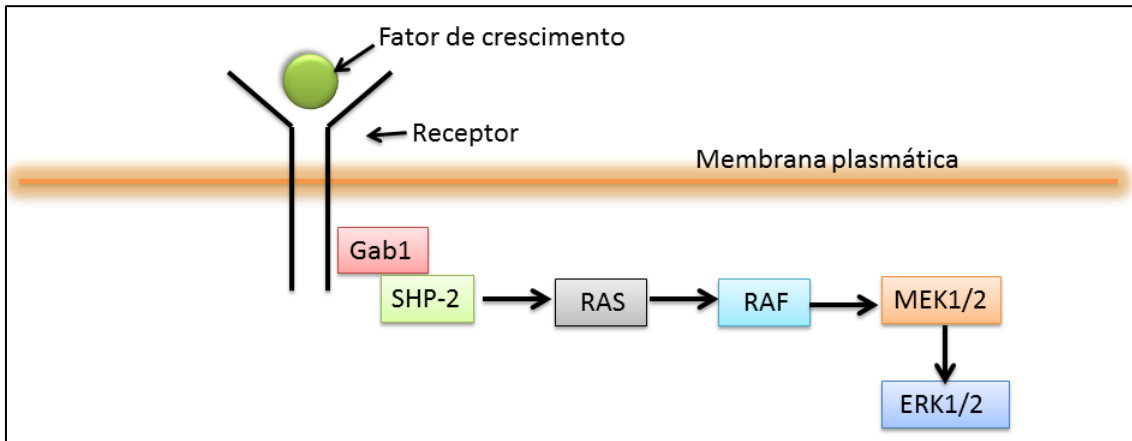


Figura 11. Em resposta à fatores de crescimento, a Shp2 interage através de seus domínios SH2 com receptores autofosforilados ou com proteínas de ancoragem (como por exemplo a Gab-1, mostrada na figura) que são ativadas por receptores de tirosinas cinases ou por cinases da família Src. Tais interações resultam na ativação da Shp2 e consequentemente na ativação da Ras e de toda a via ERK1/2, levando ao crescimento celular.

De forma resumida, pode-se dizer que a Shp2 é uma reguladora da adesão e da migração celular. Estas funções são mediadas pela desfosforilação da cinase de adesão focal (FAK), ativação da cinase Src e ativação da via RAS-ERK1/2. [TONKS *et al.*, 2006; QU *et al.*, 2000].

2.7 Shp2 e o câncer

Em seres humanos, mutações germinais no gene PTPN11 (que codifica a Shp2) foram identificadas em mais de 50% das crianças com a síndrome de Noonan, uma síndrome que está frequentemente associada com o desenvolvimento da leucemia juvenil mielomonocítica (JMML). Reciprocamente, mutações somáticas no gene PTPN11 também foram identificadas em 34% dos pacientes com JMML. A maioria das mutações somáticas da Shp2 localizam-se nos resíduos do domínio N-SH2 e afetam a interação com o domínio PTP resultando no aumento da atividade da enzima [BENTIREN-ALJ *et al.*, 2004].

Xu e colaboradores (2011) constataram que camundongos com o gene PTPN11 contendo a mutação mais comum E76K desenvolveram leucemia aguda [XU *et al.*, 2005]. Por estes motivos, a Shp2 foi a primeira tirosina fosfatase a receber a classificação de oncoproteína [CHAN *et al.*, 2007].

Enquanto a forma nativa da Shp2 atua como um gene não cancerígeno, um ganho de função em Shp2 mutantes tem sido considerado como cancerígeno. Porém, nenhum caso de perda de função em Shp2 mutante em humanos foi relacionada com o câncer [ROCCOGRANDI *et al.*, 2017]. As mutações que causam perda de função na Shp2, entretanto, estão relacionadas com a síndrome de Leopard [NODA *et al.*, 2016].

Uma expressão elevada de Shp2 tem sido observada em câncer de mama e em células leucêmicas primárias [BENTIREES-ALJ *et al.*, 2004; ZHOU *et al.*, 2008; XU *et al.*, 2005; LOH *et al.*, 2004]. Bentires-Alj e colaboradores (2004) investigaram as mutações do gene PTPN11 de 13 diferentes tipos de neoplasmas e descobriram 11 tipos de mutações responsáveis por aumentar a atividade da Shp2. Estas mutações, no entanto, ocorreram em baixa frequência nos mais variados tipos de tumores [BENTIREES-ALJ *et al.*, 2004].

Porém, a mutação não é o único meio pelo qual a hiperativação da Shp2 ocorre. Em 2002, através dos estudos de Higashi e colaboradores, foi relatado pela primeira vez a ativação da Shp2 pelo antígeno A associado à citotoxina (CagA) liberada pela bactéria cancerígena *Helicobacter pylori* (*H. pylori*). As cepas de *H. pylori* podem ser divididas em dois tipos: as que contêm o antígeno A associado à citotoxina (chamadas CagA-positivas) e as que não contêm (chamadas CagA-negativas). As cepas CagA-positivas são mais virulentas e causam inflamações na mucosa do estômago e úlceras mais graves, aumentando o risco de câncer gástrico. As proteínas CagA, após serem injetadas nas células epiteliais gástricas através do sistema de excreção tipo IV, ficam retidas na membrana plasmática e recrutam a Shp2, para induzir transformações nas células, desregulando a multiplicação celular e aumentando a mobilidade das mesmas [HIGASHI *et al.*, 2002].

A proteína CagA possui um resíduo de tirosina fosforilado que se liga ao domínio SH2 da Shp2 revertendo o mecanismo de auto inibição e aumentando a atividade da fosfatase o que resulta na ativação da via ERK/MAPK. Além disso, a infecção das células epiteliais gástricas através das cepas de *H. Pylori* CagA positiva foram relacionadas ao surgimento de um tipo alongado de célula dependente da interação entre a CagA e a Shp2 [HIGASHI *et al.*, 2004].

Num experimento conduzido por Ohnishi e colaboradores (2008), ratos transgênicos expressando um tipo selvagem de CagA apresentaram hiperplasia epitelial gástrica e alguns desenvolveram adenocarcinoma. Em contraste, essas patologias não foram observadas nos ratos transgênicos expressando o tipo CagA resistente à

fosforilação, ou seja, que não interage com a Shp2. Estes estudos apontam a importância da fosforilação da tirosina da CagA para desregular a atividade da Shp2 [OHNISHI *et al.*, 2008]. É importante destacar que a interação entre a CagA e a Shp2 foi observada em pacientes com gastrite atrófica, mas não em pacientes com metaplasia intestinal ou com câncer, sugerindo que este mecanismo é crítico apenas nas fases iniciais da oncogênese. Interessantemente, existe um aumento da expressão da Shp2 em cânceres gástricos independentemente da infecção por *H. Pylori*, sugerindo que a Shp2 pode sofrer a interferência de diferentes fatores e enfatizando que esta fosfatase é de importância crucial no câncer gástrico [DONG *et al.*, 2012].

Por fim, a regulação da atividade fosfatase da Shp2 ainda permanece desconhecida em doenças inflamatórias do intestino bem como em outras doenças gastrointestinais. A Shp2 é inativada durante o estresse oxidativo e na presença de óxido nítrico por produtos. Isto sugere que a perda de atividade da Shp2 e das vias celulares reguladas pela Shp2 pode coincidir com a lesão celular e o estresse oxidativo que ocorre durante a inflamação [FICHTNER-FEIGL *et al.*, 2015; CHEN *et al.*, 2009].

Alguns estudos demonstraram que a Shp2 nas células do fígado possui duas faces: apesar de ser uma oncoproteína, também atua como supressora de tumor. De fato, a deleção do gene PTPN11 nas células do fígado de ratos causou inflamação, necrose e inibição da apoptose, o que promoveu o desenvolvimento de tumores. Porém, as evidências apontam que a Shp2 só atua como supressora de tumores nas células do fígado sob um contexto específico (como a inflamação). Mais especificamente, a Shp2 atua como supressora da inflamação, por regular negativamente a STAT3 e a deleção total do gene PTPN11 nas células do fígado de ratos induziu um estado de inflamação crônica resultando na formação de tumor [BARD-CHAPEAU *et al.*, 2011]. Por outro lado, Han e colaboradores (2015) constataram uma expressão elevada da Shp2 em 65,9% (394/598) dos pacientes com câncer hepatocelular. A regulação negativa da Shp2 atenuou a adesão e a migração das células do fígado diminuindo a formação de metástase em ratos. Os estudos também apontaram que a Shp2 promoveu o crescimento do tumor e a metástase por coordenar a ativação da via RAS/ERK e a cascata PI3-K/Akt [HAN *et al.*, 2015].

Recentemente descobriu-se que a Shp2 pode ser um alvo no tratamento do câncer ainda que não se encontre hiperativada, pois sua inibição resulta na reativação da senescência celular [SERRANO *et al.*, 2015; LAN *et al.*, 2015]. A reativação do processo de senescência das células emergiu recentemente como uma estratégia

promissora na terapia do câncer. A senescência celular é uma resposta ao estresse que desativa a proliferação celular e induz um processo inflamatório que elimina as células danificadas. Em 2015, o FDA aprovou o primeiro medicamento para tratamento do câncer pró-senescência, o Palbociclib, um inibidor da CDK4 (quinase dependente da ciclina 4) e da CDK6 (quinase dependente da ciclina 6), indicado para o tratamento do câncer de mama avançado. Além da CDK4 e da CDK6, a Shp2 é um dos poucos alvos já descobertos cuja inibição resulta na senescência celular [SERRANO *et al.*, 2015; LAN *et al.*, 2015].

Nos últimos anos, houve um significativo aumento de estruturas cristalizadas da Shp2 depositadas no banco de dados de proteínas *Protein Data Bank*, contendo mutações de ganho de função relacionadas ao câncer ou contendo inibidores cocrystalizados (Figura 12), o que indica que houve um aumento de interesse dos pesquisadores nesta enzima para o tratamento de alguns tipos específicos de tumores.

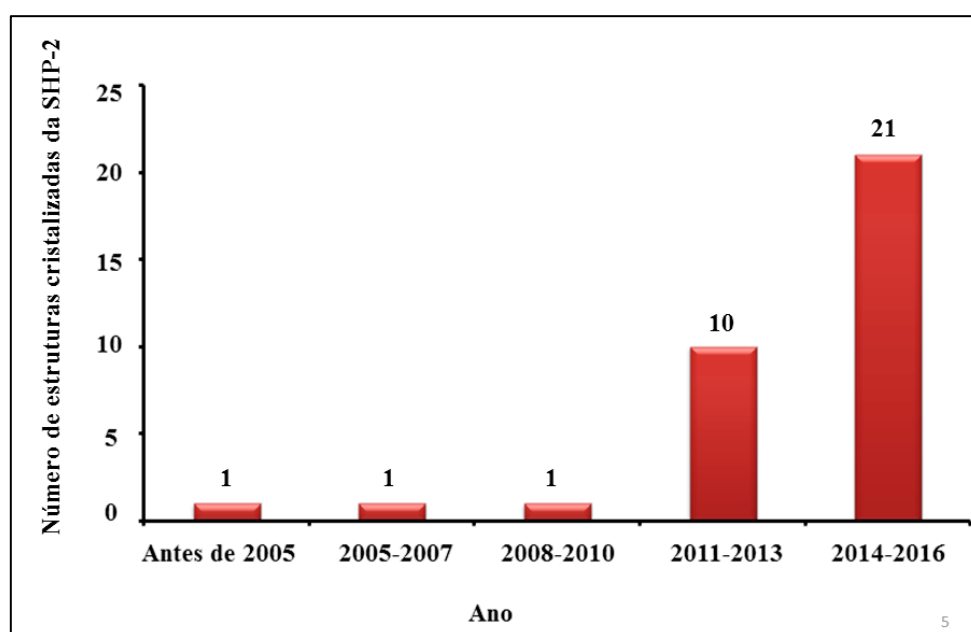


Figura 12. Número de estruturas cristalizadas da Shp2 depositadas nos últimos anos.

2.8 Alguns inibidores da Shp2

Através de uma triagem realizada na biblioteca de compostos do Instituto Nacional do Câncer, foi identificado o composto NCS-87877 como um potente inibidor da Shp2 (composto **F**, Figura 13). Este composto inibe a Shp2 com um CI_{50} de 0,318

μM , porém também inibe a Shp1 com um CI_{50} de $0,355 \mu\text{M}$. Ensaios envolvendo células mostraram que este inibidor é ativo e não foi detectado nenhum efeito “*off-target*” na via de ativação da ERK1/2 estimulada pela EGF [WU *et al.*, 2015].

Noren-Muller e colaboradores (2006) identificaram dois compostos derivados da furanodictina com atividade *in vitro* contra a Shp2 (compostos **G** e **H**, Figura 13). Porém, não são seletivos, pois interagem também com a PTP1B e com a Shp1 [NOREN-MULLER *et al.*, 2006]

Scott e colaboradores (2011) identificaram o fosfato de estramustina (composto **I**, Figura 13) como um inibidor potente da Shp2, porém pouco seletivo, já o fosfato de estradiol (composto **J**) demonstrou ter quase a mesma atividade contra a Shp2, porém apresentou-se bastante seletivo. A cinética de reação foi determinada como mista, ou seja, este inibidor pode estar interagindo tanto com os aminoácidos do sítio ativo quanto com os aminoácidos de outra região da enzima [SCOTT *et al.*, 2011].

Yu e colaboradores (2011) obtiveram bons resultados na inibição da Shp2 através de uma pequena molécula de esqueleto ainda não explorado na inibição desta enzima (composto **K**, Figura 13). Esta molécula não é seletiva, apresentando um CI_{50} de $8,3 \mu\text{M}$ contra a Shp2, e um CI_{50} de $14 \mu\text{M}$ contra a Shp1 [YU *et al.*, 2011].

He e colaboradores (2015), através de uma busca realizada numa biblioteca de fármacos aprovados pelo FDA, descobriram que a cefsulodina (composto **L**, Figura 13) inibe a enzima SHP-2. A fim de se estudar as principais interações responsáveis pela inibição da SHP-2 pela cefsulodina, os autores determinaram a estrutura cristalográfica do complexo com $1,6 \text{ \AA}$ de resolução (código 4RDD do PDB). De forma inesperada, um derivado da cefsulodina (estrutura **M**, Figura 13) e não a cefsulodina estava cocrystalizado no sítio ativo da SHP-2 devido à hidrólise do anel β -lactâmico. A análise da estrutura cristalográfica 4RDD também permitiu aos autores identificar o grupamento sulfonil acético amida como um novo composto mimético da fosfotirosina [HE *et al.*, 2015].

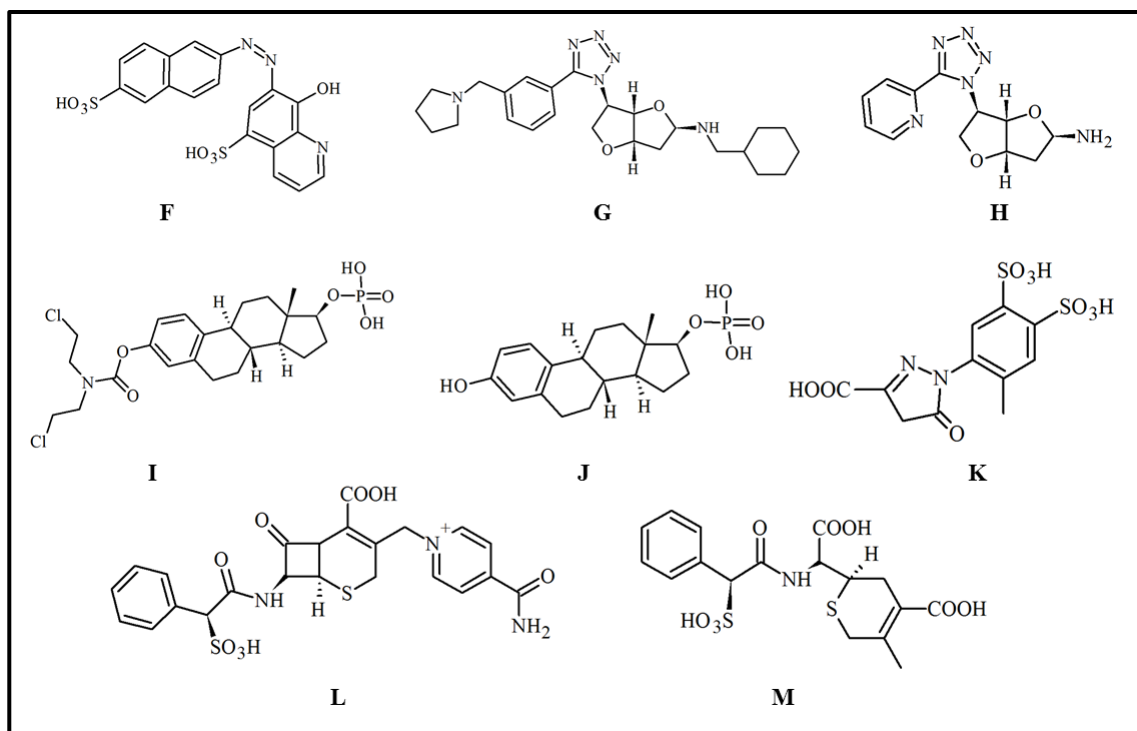


Figura 13. Alguns inibidores da Shp2

2.9 Modelagem molecular

Com o surgimento da Mecânica Quântica no início do século XX, as aplicações da Matemática aos problemas da Química entraram em um novo patamar, o que tornou possível prever algumas propriedades de moléculas antes mesmo de sua síntese. Além disso, após o surgimento e o desenvolvimento da computação, grandes progressos começaram a ocorrer na área dos estudos teóricos em Química, permitindo que sistemas moleculares cada vez maiores fossem estudados [CLARK,1985]. A partir do desenvolvimento de programas capazes de determinar a estrutura molecular por meio de cálculos teóricos, surgiu a modelagem molecular, que, segundo a IUPAC, trata da investigação das estruturas e das propriedades moleculares pelo uso da Química Computacional e de técnicas de visualização gráfica, visando fornecer uma representação tridimensional, sob um dado conjunto de circunstâncias [SANT'ANNA, 2002].

Quando se trata de proteínas ou de seus complexos com moléculas pequenas (ligantes), as técnicas experimentais como a cristalografia de raios-X, a ressonância magnética nuclear (RMN) e, mais recentemente, a microscopia eletrônica, têm

fornecido dados estruturais valiosos, tanto do alvo biológico quanto do ligante cocrystalizado. De posse de tais informações, a modelagem molecular pode ser usada como ferramenta no planejamento de novos ligantes, com propriedades estruturais otimizadas para um melhor reconhecimento molecular pelo alvo biológico, no chamado planejamento de fármacos baseado na estrutura (SBDD, do inglês *Structure Based Drug Design*). As técnicas de modelagem molecular podem contribuir em todos os estágios do planejamento de compostos candidatos a ensaios biológicos e testes clínicos, desde a descoberta de um protótipo até sua otimização, em termos de afinidade, seletividade, toxicidade e biodisponibilidade [OPREA, 2005].

Os dois métodos que têm sido usados predominantemente em estudos de modelagem molecular são o método clássico, que inclui os métodos da mecânica molecular e da dinâmica molecular, e o método quântico, que inclui métodos como os *ab initio* e os semi-empíricos. De modo geral, a escolha entre estes métodos depende das propriedades que se deseja avaliar, da precisão desejada e da capacidade computacional disponível para a realização dos cálculos [SANT'ANNA, 2009].

2.10 Métodos clássicos

Os métodos de mecânica molecular baseiam-se na proposta de que as moléculas são formadas por fragmentos estruturalmente similares e, por esta razão, algumas propriedades podem ser transferidas de uma molécula para a outra, tais como comprimentos de ligação, frequências vibracionais, entalpias de formação, entre outras [JENSEN, 1999].

O modelo é justificável porque os parâmetros associados a conjuntos de átomos permanecem razoavelmente constantes entre estruturas diferentes, desde que o tipo e a hibridação dos átomos envolvidos sejam os mesmos [BURKET, 1982]. De fato, tabelas de distâncias e ângulos padrão de ligação, determinados experimentalmente, são facilmente encontradas na literatura. Assim, no modelo da mecânica molecular, é desenvolvido o chamado campo de força, um conjunto de funções de energia que determinam penalidades energéticas para o afastamento desses valores padrão da estrutura [BURKET, 1982]. A minimização de energia é um processo no qual, através de um algoritmo matemático adequado, busca-se reduzir essas penalidades em conjunto, ou seja, reduzir a energia a um mínimo [SANT'ANNA, 2009].

Muitos campos de força estão disponíveis hoje em dia, a maioria deles parametrizados para sistemas orgânicos, fornecendo ótimos resultados qualitativos [ABRAHAM *et al.*, 2015; HUANG *et al.*, 2016; CHMIELA *et al.*, 2017]. Os resultados serão tanto melhores quanto mais próxima for a estrutura da molécula que está sendo estudada daquelas usadas como padrão. Um exemplo de campo de força muito difundido é o MM2 [ALLINGER, 1977], desenvolvido com o intuito de descrever pequenos sistemas orgânicos. Outro exemplo, aplicável principalmente a macromoléculas de interesse biológico, como proteínas e ácidos nucleicos, é o AMBER [WEINER *et al.*, 1984].

Uma vez que bons parâmetros estejam disponíveis, será possível alcançar boa predição das geometrias, das energias relativas e barreiras conformacionais de uma grande quantidade de moléculas em um tempo computacional pequeno. Porém, como os programas que usam métodos de mecânica molecular buscam uma conformação de energia mínima, isso pode trazer limitações ao seu uso em estudos de Química Medicinal, uma vez que na temperatura da biofase as moléculas possuem energia cinética suficiente para assumir muitas outras conformações.

O método de dinâmica molecular pode dar uma descrição do espaço conformacional muito mais ampla, pois utiliza os fundamentos da mecânica molecular com um algoritmo de busca determinístico para descrever o comportamento dinâmico do sistema ao longo do tempo [LEACH, 2001]. Neste método, a resolução das equações de movimento representa a evolução no tempo dos movimentos moleculares, chamada trajetória, que pode ser usada para o estudo de propriedades que dependem do tempo, tais como a difusão, o dobramento de cadeias moleculares (como acontece no enovelamento de proteínas) e a distribuição de moléculas de solvente ao redor de um soluto, entre outros. Em um cálculo de dinâmica molecular, as moléculas que dispõem de energia cinética, podem superar barreiras de energia potencial, o que permite a exploração da superfície de energia potencial (SEP) na busca de outras conformações estáveis [SANT'ANNA, 2009].

2.11 Métodos quânticos

Para a determinação das propriedades do sistema, os métodos quânticos usados em estudos de modelagem molecular obtêm soluções para a equação de Schrödinger independente do tempo. Para chegar a essa equação, Schrödinger concluiu que só seria

possível determinar corretamente a energia em um sistema com a inclusão do comportamento ondulatório proposto por de Broglie para todas as partículas (núcleos e elétrons) que constituem esse sistema. Inicialmente, Schrödinger definiu uma função de onda, Ψ , que depende das coordenadas de todos os núcleos e das coordenadas de todos os elétrons do sistema [SANT'ANNA, 2009].

As soluções da equação de Schrödinger estão sujeitas a restrições importantes. Em primeiro lugar, a função de onda Ψ deve ser normalizável e, em segundo lugar, tem que ser antissimétrica com respeito à permuta de dois elétrons quaisquer [ALCÁCER, 2007]. A equação é de solução complexa e uma forma de simplificar o problema é assumir a imobilidade dos núcleos atômicos. A massa nuclear é várias vezes maior que a massa dos elétrons, fazendo com que os elétrons se ajustem quase que instantaneamente a qualquer mudança de posição nuclear. Esta aproximação é conhecida como aproximação de Born-Oppenheimer, que permite separar a equação de Schrödinger em duas: uma para a parte eletrônica e outra para a nuclear [BORN & OPPENHEIMER, 1927].

Ainda assim, a equação de Schrödinger não pode ser resolvida exatamente para sistemas com mais do que um elétron [ALCÁCER, 2007]. Diversos métodos têm sido desenvolvidos para se obter a solução aproximada da equação de Schrödinger para sistemas de muitos elétrons. Dentre estes, o mais popular é o método de Hartree-Fock (HF) [HARTREE, 1923]. Além de ser capaz de fornecer uma solução para sistemas de muitos elétrons, o método HF tem a vantagem de servir como base para outros métodos: por um lado, os métodos semi-empíricos, onde aproximações adicionais são incluídas com o propósito de reduzir o custo computacional; por outro, métodos onde à solução HF são acrescentadas correções no sentido de aproximar a solução exata.

Uma segunda complicação diz respeito ao spin, uma forma intrínseca de momento angular do elétron. O spin é caracterizado pelo número quântico s , sendo um efeito relativístico. Como a equação de Schrödinger não é relativística, inclusões adicionais que computem tal efeito na forma da função de onda tiveram que ser feitas. A combinação da função de onda espacial do elétron com uma função de spin adequada produz a chamada função spin-orbital [CLARK, 1985].

No método HF, um sistema de N elétrons é descrito como uma soma de N sistemas de um elétron que se move num campo gerado pelos núcleos estacionários e num campo médio resultante da distribuição espacial de todos os outros elétrons; assim, o problema se reduz ao problema de N elétrons independentes. A teoria de HF, nesta

aproximação, obriga a construção de funções spin-orbitais χ para cada elétron, que minimizem a energia do estado fundamental do sistema (método variacional) [MORGON & COUTINHO, 2007].

Estas funções spin-orbitais devem fazer parte de uma função determinantal escolhida de forma ótima, uma vez que uma função de onda que descreve um sistema de muitos elétrons deve ser antissimétrica perante a troca de coordenadas de dois destes elétrons e a antissimetria é uma característica dos determinantes. Este determinante é conhecido como determinante de Slater [SLATER, 1929], onde as funções spin-orbitais moleculares são funções das coordenadas espaciais e de spin de cada elétron [MORGON & COUTINHO, 2007].

O que se deseja é obter uma função de onda que produza a menor energia para um sistema multieletrônico mantendo a formulação spin-orbital individual, mas em um sistema multieletrônico temos a presença das interações entre elétrons. O problema é encontrar uma solução que simultaneamente englobe todos os movimentos eletrônicos, bem como o efeito que um elétron em um orbital exerce no comportamento de um outro elétron em um outro orbital devido a estes movimentos eletrônicos.

A solução para este problema pode ser encontrada na equação de Hartree-Fock, onde se assume que cada elétron se move no campo médio produzido pelos demais elétrons e pelos núcleos [HOUSE, 2004]. Esta equação possui um operador (operador de Fock) que estabelece uma dependência entre a equação de cada spin-orbital com os outros e os spin-orbitais moleculares também dependem das soluções da própria equação.

Sendo assim, as equações devem ser resolvidas de forma acoplada, através de aproximações sucessivas (solução iterativa). As energias totais correspondentes a cada conjunto são comparadas até que a energia não varie entre duas estruturas consecutivas, dentro de um limite pré-determinado. A melhor função de onda é obtida de tal forma que a energia seja um mínimo [HØYVIK & JORGENSEN, 2016].

Uma aproximação para representar a função de onda eletrônica em sistemas moleculares é usar uma combinação linear de funções de onda de um elétron estendidas por toda a molécula. Estas funções são chamadas de orbitais moleculares (MO), e esta aproximação é chamada de LCAO (combinação linear dos orbitais atômicos). A distribuição eletrônica em uma molécula segue os mesmos princípios da distribuição em átomos, ou seja, cada MO pode conter no máximo 2 elétrons, com spins opostos (princípio de Pauli) [MORGON & COUTINHO, 2007].

Pelo princípio da exclusão de Pauli, nos sistemas de camada fechada, os spins dos elétrons em um mesmo orbital molecular devem ser opostos e, assim, se anulam mutuamente. Logo, não é necessário considerar o efeito do spin eletrônico na resolução da equação de Schrödinger nesses sistemas. Para sistemas com elétrons desemparelhados, como é o caso dos radicais livres, é necessário tratar de alguma forma a questão do spin. A forma mais comum é dividir os OM em 2 grupos, chamados alfa e beta, que são ocupados separadamente por elétrons com apenas um determinado spin. Esse método é chamado de UHF (do inglês *Unrestricted Hartree-Fock*) [SANT'ANNA, 2009].

2.12 Métodos semi-empíricos

O método HF é considerado um método *ab initio* (do princípio), no sentido em que na solução da equação de Schrödinger, os únicos dados fornecidos na entrada são constantes físicas fundamentais. Já nos métodos semi-empíricos, uma grande parte das integrais referentes à resolução da equação de Schrödinger são negligenciadas devido à introdução de parâmetros empíricos ou previamente calculados que permitem a eliminação do cálculo de algumas dessas integrais [SANT'ANNA, 2009]. Isso reduz consideravelmente o tempo de cálculo e, desta forma, métodos semi-empíricos são os mais indicados no estudo de sistemas grandes, como as proteínas e seus complexos com outras moléculas.

Em geral, os métodos semi-empíricos limitam-se ao tratamento dos elétrons de valência, não considerando explicitamente os elétrons do cerne, que contribuem muito menos, em geral, para o comportamento das moléculas [ALCÁCER, 2007]. Estes métodos procuram resolver, de forma autoconsistente, as equações de Hartree-Fock, utilizando parâmetros, obtidos por ajuste numérico ou derivados de resultados experimentais, para tornar mais rápida a resolução das equações. Na maioria das vezes, utilizam um conjunto de base mínimo de valência, formado por funções do tipo Slater (STO – *Slater type orbitals*), e fazem uso da teoria de orbitais moleculares para a construção da função de onda molecular, construindo estes orbitais moleculares a partir da aproximação LCAO.

Existe uma grande variedade de métodos semi-empíricos, nos quais se fazem várias aproximações adicionais de modo a reduzir as dificuldades de cálculo. Dentre estas aproximações, uma das que têm impacto no custo computacional é conhecida

como aproximação ZDO (*Zero Differential Overlap*). Esta aproximação ignora as integrais de sobreposição de dois elétrons que se encontram em orbitais atômicos diferentes, simplificando bastante as equações de Hartree-Fock, mas que pode levar a grandes erros nas várias propriedades moleculares calculadas, ficando então seu uso restrito ao tratamento de sistemas π de moléculas planares [MORGON & COUTINHO, 2007]. A fim de que os métodos semi-empíricos pudessem tratar outros tipos de sistema, Pople e colaboradores (1967) propuseram novos métodos: CNDO (*Complete Neglect of Differential Overlap*), INDO (*Intermediate Neglect of Differential Overlap*) e NDDO (*Neglect of Diatomic Differential Overlap*). Estes métodos possuem níveis de aproximações diferentes. A aproximação NDDO é a que considera um maior número de integrais para serem calculadas e, desta forma é a melhor das três aproximações integrais [POPLE *et al.*, 1967].

Os métodos semi-empíricos paramétricos apoiam-se na proposta de Dewar e colaboradores (1985) na qual a parametrização é feita de modo a reproduzir da melhor forma possível medidas experimentais [DEWAR *et al.*, 1985]. São selecionados conjuntos de átomos que pertencem a conjuntos de moléculas cujas entalpias de formação padrão, geometria molecular e momento dipolar sejam conhecidos experimentalmente. A determinação do melhor conjunto de parâmetros é feita por minimização dos erros nos cálculos de entalpia de formação, geometria e momento dipolar [ALCÁCER, 2007].

Como exemplos de métodos semi-empíricos seguindo essa proposta, têm-se o MNDO (Modified Neglect of Diatomic Overlap) [DEWART & THIEL, 1977], AM1 (Austin Model 1) [DEWAR *et al.*, 1985], PM3 (Parametric Method 3) [STEWART, 1989], o MNDO/d (Modified Neglect of Diatomic Overlap, incluindo orbitais d) [THIEL & VOITYUK, 1992] e, mais recentemente, os métodos RM1 [ROCHA *et al.*, 2006], PM6 [STEWART, 2007] e PM7 [STEWART, 2013]. O PM7 é um dos mais recentes métodos orbitais semi-empíricos e apresenta algumas melhorias em relação aos seus antecessores: uma modificação foi feita para melhorar a descrição de interações não covalentes e dois pequenos erros no formalismo do NDDO presentes em métodos semi-empíricos anteriores foram corrigidos. Devido a mudanças feitas nas aproximações, o PM7 apresenta um aumento significativo de precisão, levando a uma redução de erros nas geometrias numa ordem de 30% quando comparado ao seu antecessor PM6. O método PM7 é rápido o suficiente para ser aplicado rotineiramente em sistemas contendo centenas de átomos [STEWART, 2013].

A extensão dos métodos semi-empíricos para metais de transição enfrentou algumas dificuldades [STEWART, 2007]: a escassez de dados de referência e o fato de que é comum que compostos de metais de transição apresentem camada aberta; é mais difícil resolver as equações do método auto consistente para sistemas de camada aberta porque o *spin* total é diferente de zero e precisa ser considerado de algum modo na resolução (como através do método UHF, descrito anteriormente). Finalmente, estudos preliminares indicam que cálculos *ab initio* de alto nível para metais de transição não são úteis para gerar dados de referência termoquímica, embora sejam úteis para gerar dados de geometrias [MORGON & COUTINHO, 2007].

A maioria dos grandes avanços obtidos na pesquisa de métodos semi-empíricos foi essencialmente incorporada em um único programa: o MOPAC (Molecular Orbital PACKage). Iniciado em 1983, no grupo do Prof. M. Dewar, com a participação do Prof. J. Stewart, que também é o responsável pelo desenvolvimento de muitos métodos paramétricos bem sucedidos [STEWART, 1996].

2.13 Docagem

Os sítios ativos das enzimas e os sítios de interação presentes em receptores exibem alta seletividade no reconhecimento de seus bioligantes. Se, de alguma forma, estas interações pudessem ser explicadas e preditas por algum método, a modelagem de moléculas inéditas seria realizada de forma muito mais eficiente. Com isso surgiu a técnica de atracamento molecular, também conhecida como docagem (ou ancoramento) molecular, na década de 1980 [KUNTZ *et al.*, 1982], atualmente bastante utilizada quando as estruturas dos sítios ativos e das moléculas estão disponíveis, ou ainda para o atracamento entre duas proteínas.

Muitos programas de docagem empregam métodos de inteligência artificial, como algoritmos genéticos (AG), que são particularmente úteis na resolução de problemas com um vasto número de variáveis, por sua rapidez e eficiência no encontro de soluções [LINDEN, 2006]. Este algoritmo foi proposto nos anos 60 por John H. Holland, pesquisador da Universidade de Michigan (EUA), e foi primeiramente utilizada na Química por Lucasius e Kateman (1991), em um trabalho para seleção de comprimentos de onda em espectros de ultravioleta. O AG otimiza as soluções encontradas em várias situações, sendo considerado um dos melhores algoritmos de otimização estocástica, encontrando soluções muito próximas às verdadeiras (as

encontradas em métodos experimentais, como a cristalografia de raios-X) [COSTA FILHO & POPPI, 1999; MORRIS *et al.*, 1998].

Os AG simulam matematicamente idéias genéticas, sendo necessárias 5 etapas básicas: a codificação das variáveis, a criação da população inicial, a avaliação da resposta (aptidão), o cruzamento (“*crossover*”) e a mutação [PAGADALA *et al.*, 2017]. A codificação das variáveis consiste na criação dos “cromossomos”, formados por possíveis soluções (ou seja, pelas conformações e orientações de um ligante frente a uma proteína/sítio ativo), que podem ser definidas por um grupo de valores (geralmente por codificação binária) descrevendo a translação, orientação e conformação do ligante atracado, chamado genótipo; as características atômicas de um ligante correspondem ao seu fenótipo. Com isto, temos a criação da população inicial, constituída por vários cromossomos. Quanto maior é a diversidade da população, maior o número de soluções diferentes. Se a população inicial foi gerada usando alguma técnica heurística para obtenção de indivíduos de boa qualidade, o AG pode chegar a soluções mais rapidamente quando comparado ao processo de geração aleatória de indivíduos [REEVES, 1995].

A aptidão de um indivíduo da população é então calculada a partir dos valores descritos acima, baseados no mérito relativo das soluções encontradas. Pares randômicos de cromossomos são então cruzados, onde novas gerações de cromossomos herdarão características de seus “pais”. Alguns passarão pelo processo de mutação, onde ocorrem alterações nas informações contidas nos genes. A seleção da descendência de tais gerações é baseada na aptidão dos novos cromossomos: aqueles que melhor interagem com o seu ambiente (proteína) são selecionados, e aqueles que não se adaptam bem são abandonados [COSTA FILHO & POPPI, 1999; MORRIS *et al.*, 1998].

As funções de escore (“*scoring functions*”) ou funções de pontuação mensuram o grau de aptidão dos indivíduos, permitindo a classificação dos chamados posicionamentos ou “poses”, que são as combinações de orientações e conformações do ligante no sítio receptor [MORGON & COUTINHO, 2007]. Existem três tipos de funções de pontuação: as baseadas em campos de força, as empíricas e as baseadas em conhecimento [KITCHEN *et al.*, 2004].

As funções de pontuação baseadas em campos de força quantificam a soma de duas energias: a de interação entre o receptor e o ligante (energia “externa”) e a energia

“interna” do ligante. As interações entre o ligante e a proteína são quase sempre descritas utilizando termos de energia eletrostática e de van der Waals.

O termo eletrostático é uma formulação de Coulomb, com uma função dielétrica dependente de distância, para o tratamento de interações carga-carga [KITCHEN *et al.*, 2004]. Várias funções são baseadas em diferentes parâmetros de campos de força; a do programa *AutoDock* [MORRIS *et al.*, 1998], por exemplo, é baseada no campo de força AMBER [WEINER *et al.*, 1984]. Algumas funções possuem extensões dos termos de energia para englobar termos entrópicos. Como exemplo, temos a inclusão de termos de entropia torcional ao “*GScore*” e “*Goldscore*” [VERDONK *et al.*, 2003], e a inclusão de termos para ligação hidrogênio entre ligante e proteína no “*Goldscore*” [VERDONK *et al.*, 2003] e “*AutoDock*” [MORRIS *et al.*, 1998].

As funções de pontuação empíricas reproduzem dados experimentais, como energias de interação e conformações, como uma soma de várias funções parametrizadas, baseado na aproximação das energias de interação por uma soma de termos individuais não-correlacionados. Tais termos são obtidos de uma análise por regressão, usando energias de interação determinadas experimentalmente, com informações estruturais (obtidas por cristalografia de raios-X). Suas formas funcionais são frequentemente mais simples que as funções de score obtidas por campos de força, com termos de avaliação mais simples. Sua desvantagem reside na dependência dos conjuntos de dados moleculares, o que leva a diferentes fatores (quando comparados a funções de campos de força), que não podem ser facilmente recombinaados em uma nova função de aptidão [KITCHEN *et al.*, 2004].

E por último, também existem as funções de pontuação baseadas em conhecimento, que foram feitas para reproduzir estruturas experimentais ao invés de energias de ligação. Os complexos proteína-ligante são modelados usando potenciais de interação atômica relativamente simples. Implementações destas funções incluem o “*PMF*” (“*Potencial of Mean Force*”) [MUEGGE & MARTIN, 1999] e o “*DrugScore*” [GOHLKE *et al.*, 2000], que também incluem correções como a acessibilidade ao solvente, dentre outras. Tais funções primam pela simplicidade, fazendo com que grandes conjuntos de moléculas possam ser avaliadas com rapidez. A desvantagem deste método reside no fato de que sua derivação é baseada na informação codificada implicitamente em conjuntos limitados de complexos proteína-ligante.

A etapa de reconhecimento molecular entre o ligante e a macromolécula é um processo complexo, pois envolve fatores entálpicos e entrópicos e que incluem a

flexibilidade do ligante e da proteína, o efeito do ambiente protéico na distribuição de cargas do ligante e possíveis interações com moléculas de água presentes no meio [KITCHEN *et al.*, 2004]. Assim, em razão desses fatores, é necessário avaliar qual função de pontuação se adequa melhor para cada sistema proteína-ligante a ser estudado.

Os métodos de docagem molecular são geralmente conhecidos por serem eficazes na reprodução das geometrias de interação, mas simplificações adotadas em suas funções de pontuação podem reduzir o nível de correlação entre as pontuações e as informações experimentais relacionadas à interação ligante-enzima, como K_i ou CI_{50} [TACSAY & KESERU, 2015; KAWASAKI & FREIRE, 2011]. Além da complexidade dos fenômenos associados à ligação ligante-proteína, esse baixo desempenho das funções de pontuação pode estar associado, pelo menos em parte, à sua incapacidade de incorporar efeitos quânticos como a transferência de carga entre proteína e ligante e a grande variação nas cargas atômicas entre diferentes estruturas de ligantes e proteínas. [ZHOU & HUANG, 2008]. Claramente, o cálculo da energia associada à ligação ligante-proteína se beneficiaria do uso de abordagens teóricas mais sofisticadas, incluindo métodos baseados em mecânica quântica. De fato, há exemplos na literatura que utilizam cálculos de mecânica quântica para produzir valores de entalpia que, juntamente com outros termos termodinâmicos calculados, reproduzem dados de energia livre que se correlacionam bem com dados experimentais [WANG *et al.*, 1994; OLIVEIRA *et al.*, 2006; FANFRLIK *et al.*, 2010; DOBES *et al.*, 2011].

3 OBJETIVOS

O objetivo deste trabalho é contribuir para o planejamento de inibidores da Shp2 mais seletivos e potentes, fazendo uso da modelagem molecular. Dentro deste objetivo geral, este trabalho buscou alcançar os seguintes objetivos específicos:

- Desenvolver um procedimento para identificação de inibidores seletivos da Shp2, combinando a docagem molecular de duas séries de compostos (totalizando 76 estruturas) na Shp2 e na Shp1, com o refinamento das geometrias resultantes e cálculo das entalpias de interação através do método do orbital molecular semi-empírico PM7 em ambas as enzimas.
- Desenvolver uma metodologia que permita considerar a influência do efeito eletrolítico na correção dos valores experimentais de CI_{50} através da determinação da força iônica do meio e da estimativa da razão dos coeficientes de atividade das espécies envolvidas no equilíbrio de dissociação enzima/inibidor.
- Construir um modelo de previsão da atividade de inibidores da Shp2 a partir dos principais parâmetros termodinâmicos associados à inibição, conforme proposta de Wang e colaboradores (1994), calibrado com dados obtidos experimentalmente para uma série de inibidores, por meio de um modelo matemático de correlação múltipla.
- Validar o modelo construído com uma segunda série de inibidores da Shp2.
- Utilizar o modelo construído na predição da atividade relativa dos novos compostos *N*-arilpiperazínicos sintetizadas pelo Dr. Ciro Gonçalves e Sá no Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio®) da UFRJ.

4 METODOLOGIA

4.1 Redocagem

Com o objetivo de avaliar a eficiência de cada função de pontuação disponível no programa GOLD 5.5 (CCDC) em prever a pose de um determinado ligante no interior do sítio ativo da Shp2, foi realizado um estudo de redocagem. Primeiramente, foi feita uma busca no *Protein Data Bank* a fim de selecionar uma estrutura cristalográfica da Shp2 apropriada para este procedimento, ou seja, de boa resolução e que contivesse um ligante cocrystalizado estruturalmente mais parecido com as séries de compostos que seriam utilizados como base empírica no trabalho. A estrutura escolhida foi a Shp2 de código 4RDD (resolução de 1,6 Å), que possui o ligante cocrystalizado 1 - ((2*R*) -4-carboxi-2 - [(*R*) -carboxi {[(2*R*) -2- fenil-2-sulfoacetil] amino} metil] -3,6-di-hidro-2*H*-1,3- tiazin-5-il} metil) piridínio (Composto **M**, Figura 13) [HE *et al.*; 2015].

Átomos de hidrogênio foram adicionados na estrutura cristalográfica escolhida (4RDD) por meio do programa GOLD 5.5 (CCDC). O ligante co-cristalizado foi removido do sítio e então foi realizada a redocagem, ou seja, a docagem do ligante co-cristalizado novamente no sítio. Este procedimento foi feito testando-se cada uma das funções de pontuação disponíveis no programa: ChemPLP [KORB *et al.*, 2007, 2009], Chemscore [ELDRIDGE *et al.*, 1997], Goldscore [VERDONK *et al.*, 2003], e ASP [BONET *et al.*, 1997]. Foi escolhido como centro do sítio de interação o átomo de enxofre do resíduo catalítico Cys459 localizado na alça P. O sítio foi padronizado com um raio de 15 Å ao redor deste átomo. O valor de 10 Å, que é o raio padrão do programa GOLD 5.5 (CCDC) não foi utilizado por não abranger a alça β5-β6. As funções de pontuação foram então avaliadas de acordo com o valor do desvio médio quadrático (RMSD) entre as estruturas docadas de maior pontuação e a estrutura experimental.

4.2 Estudo da seletividade dos inibidores da Shp2

Duas séries de compostos foram selecionadas da literatura para serem avaliadas nesta etapa. A série 1 é composta por 52 derivados da indolina identificados por inibirem seletivamente a Shp2 [WU *et al.*, 2012]. A estrutura geral destes compostos está mostrada na Figura 14-A e os grupamentos R₁ e R₂ estão descritos na Tabela 1

juntamente com os dados experimentais de CI_{50} . Para esta série, apenas foi reportado que os derivados eram seletivos para a Shp2 em relação à Shp1, porém somente os valores de CI_{50} para a Shp2 foram disponibilizados [WU *et al.*, 2012]. Os compostos desta série que apresentavam carbonos quirais não foram incluídos neste estudo pois não foi identificado qual enantiômero era o eutômero, ou seja, o responsável pela atividade inibitória apresentada. A série 2 foi composta por 24 derivados do oxindol, e sua estrutura geral está mostrada na Figura 14-B [WU *et al.*, 2008, LAWRENCE *et al.*, 2008]. Os grupamentos R_1 e R_2 bem como os valores de CI_{50} desta série, disponíveis tanto para a Shp2 quanto para a Shp1, estão apresentados na Tabela 2.

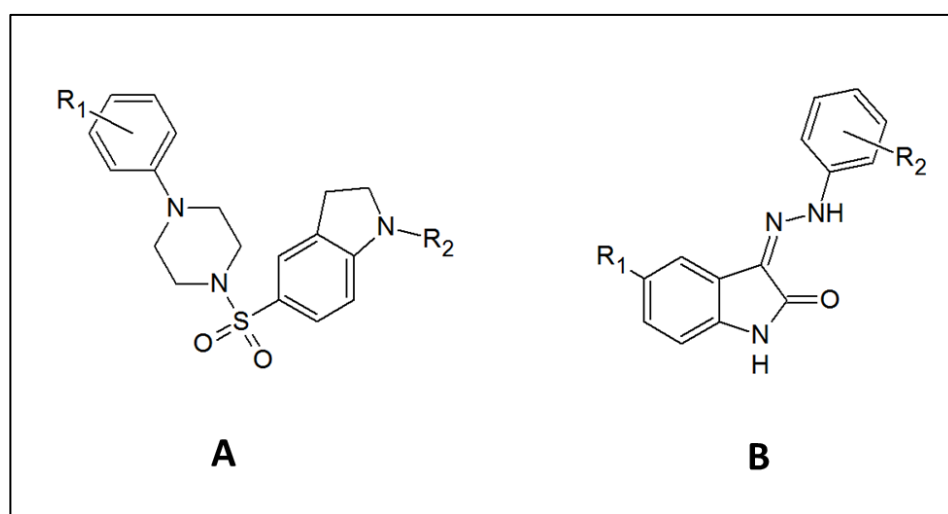


Figura 14. (A) Estrutura geral dos derivados da indolina (série 1). (B) Estrutura geral dos derivados do oxindol (série 2)

Tabela 1. Grupamentos R_1 e R_2 dos derivados da indolina (série 1) e seus respectivos valores de CI_{50} experimentais para a Shp2.

Composto	R_1	R_2	CI_{50} ($\mu\text{mol/L}$)
1	3-Cl	$\text{COC}_2\text{H}_4\text{CO}_2\text{H}$	$0,7 \pm 0,3$
2	3-Cl, 6- CH_3	$\text{COC}_2\text{H}_4\text{CO}_2\text{H}$	$1,5 \pm 0,0$
3	3-Cl	$\text{COC}_2\text{H}_2\text{CO}_2\text{H}$	$2,8 \pm 0,7$
4	3-Cl	$\text{COC}_3\text{H}_6\text{CO}_2\text{H}$	$2,8 \pm 0,7$
5	3-Cl	$\text{CONH}(3\text{-CO}_2\text{H-C}_6\text{H}_4)$	$3,0 \pm 0,8$
6	3- CH_3	$\text{COC}_2\text{H}_4\text{CO}_2\text{H}$	$3,4 \pm 2,3$

Tabela 1 (continuação). Grupamentos R₁ e R₂ dos derivados da indolina (série 1) e seus respectivos valores de *CI*₅₀ experimentais para a Shp2.

Composto	R ₁	R ₂	<i>CI</i> ₅₀ (µmol/L)
7	3,5-Cl	COC ₂ H ₄ CO ₂ H	4,5 ± 1,2
8	3,4-Cl	COC ₂ H ₄ CO ₂ H	4,6 ± 2,4
9	3-Cl	CONHCH ₂ CO ₂ H	4,6 ± 0,3
10	3-OCH ₃	COC ₂ H ₄ CO ₂ H	4,7 ± 4,4
11	3-Cl	COCO ₂ H	4,8 ± 2,8
12	3-Cl	CONHCH ₂ CH ₂ CO ₂ H	4,8 ± 3,3
13	4-COOH	COC ₂ H ₄ CO ₂ H	6,1 ± 1,8
14	3-Cl	CONH(2-CO ₂ H-C ₆ H ₄)	8,3 ± 5,6
15	4-OCH ₃	COC ₂ H ₄ CO ₂ H	9,0 ± 5,9
16	3,5-CF ₃	COC ₂ H ₄ CO ₂ H	9,1 ± 1,6
17	3-CF ₃	COC ₂ H ₄ CO ₂ H	11,3 ± 3,1
18	3-F	COC ₂ H ₄ CO ₂ H	13,2 ± 3,6
19	H	COC ₂ H ₄ CO ₂ H	14,8 ± 7,3
20	3-Cl, 4-CH ₃	COC ₂ H ₄ CO ₂ H	16,6 ± 3,6
21	4-F	COC ₂ H ₄ CO ₂ H	31,9 ± 10
22	2-F	COC ₂ H ₄ CO ₂ H	34,6 ± 12,6
23	2,3-Cl	COC ₂ H ₄ CO ₂ H	143,6 ± 89,5
24	3-Cl	CONH(2-Cl-C ₆ H ₄)	3,2 ± 0,5
25	3-Cl	CONH(4- ⁱ Pr-C ₆ H ₄)	4,1 ± 1,0
26	3-Cl	CONH(3,4-Cl-C ₆ H ₃)	5,7 ± 2,5
27	3-Cl	CO(3-Cl-C ₆ H ₄)	5,7 ± 3,9
28	3-Cl	CO(4-OCH ₃ -C ₆ H ₄)	6,3 ± 2,9
29	3-Cl	CO(C ₅ H ₉)	6,9 ± 2,2
30	3-Cl	CO(3,4-Cl-C ₆ H ₃)	7,1 ± 1,8
31	3-Cl	CO(2,4-Cl-C ₆ H ₃)	7,6 ± 1,3
32	3-Cl	COC ₃ H ₇	7,8 ± 2,1
33	3-Cl	COCH ₂ (4-Cl-C ₆ H ₄)	8,0 ± 1,9
34	3-Cl	CO(3-OCH ₃ -C ₆ H ₄)	8,1 ± 0,8
35	3-Cl	CONH(3-F-C ₆ H ₄)	8,1 ± 1,7
36	3-Cl	CO(3,5-Cl-C ₆ H ₃)	8,2 ± 2,6
37	3-Cl	CONH(3-Cl-C ₆ H ₄)	8,3 ± 3,4
38	3-Cl	CO(C ₆ H ₅)	8,4 ± 2,3

Tabela 1 (continuação). Grupamentos R₁ e R₂ dos derivados da indolina (série 1) e seus respectivos valores de *CI*₅₀ experimentais para a Shp2.

Composto	R ₁	R ₂	<i>CI</i> ₅₀ (μmol/L)
39	3-Cl	CO(C ₃ H ₅)	8,5 ± 2,0
40	3-Cl	CONH(2-CF ₃ -C ₆ H ₄)	9,0 ± 1,7
41	3-Cl	CONHC ₂ H ₄ CO ₂ C ₂ H ₅	9,3 ± 2,2
42	3-Cl	COCH ₂ OCH ₃	9,6 ± 2,4
43	3-Cl	CONH(3-OCH ₃ -C ₆ H ₄)	9,9 ± 3,5
44	3-Cl	CO(4-Cl-C ₆ H ₄)	10,4 ± 2,0
45	3-Cl	CONH(C ₆ H ₅)	11,7 ± 1,4
46	3-Cl	CONH(2-F-C ₆ H ₄)	12,2 ± 2,0
47	3-Cl	CONH(4-F- C ₆ H ₄)	12,6 ± 5,4
48	3-Cl	CONH(2-CO ₂ CH ₃ - C ₆ H ₄)	14,3 ± 4,1
49	3-Cl	CONH(4-Cl-C ₆ H ₄)	19,1 ± 5,1
50	3-Cl	COCH ₂ ⁱ Pr	20,7 ± 10,8
51	3-Cl	COCH ₂ (C ₆ H ₅)	22,8 ± 4,5
52	3-Cl	CONHCH ₂ CO ₂ C ₂ H ₅	60,3

Tabela 2. Grupamentos R₁ e R₂ para os derivados do oxindol (série 2) juntamente com os valores experimentais de *CI*₅₀ para a Shp2 e para a Shp1, em μmol/L.

Estrutura	R ₁	R ₂	<i>CI</i> ₅₀ ^{Shp2}	<i>CI</i> ₅₀ ^{Shp1}
53	SO ₂ NH ⁱ Pr	2-NO ₂	49,6 ± 25,5	110,0 ± 40,8
54	SO ₂ NH ⁱ Pr	2-CO ₂ H	7,94 ± 4,56	10,75 ± 7,81
55	SO ₃ H	2-NO ₂	46,8 ± 10,2	68 ± 17,1
56	CO ₂ H	2-CO ₂ H	0,8 ± 0,2	15,4 ± 2,1
57	CO ₂ H	3-CO ₂ H	15,8 ± 6,6	72,5 ± 19,3
58	CONHCH ₂ (4-ClC ₆ H ₄)	3-CO ₂ H	22,3 ± 5,2	19,4 ± 1,8
59	SO ₂ NH ⁱ Pr	3-CO ₂ H	4,5 ± 1,7	15,7 ± 1,7
60	SO ₂ NH ⁱ Pr	4-CO ₂ H	4,5 ± 1,0	27,7 ± 9,0
61	SO ₂ NH ₂	2-NO ₂	11,9 ± 1,5	103,7 ± 17,4
62	SO ₂ NHCH ₂ (4-ClC ₆ H ₄)	2-NO ₂	4,4 ± 2,4	40,9 ± 15,3
63	SO ₂ NHCH ₂ (4-MeC ₆ H ₄)	3-CO ₂ H	5 ± 1,9	32,6 ± 8,5
64	SO ₂ NHCH ₂ C ₆ H ₄	4-CO ₂ H	1,3 ± 0,9	7,1 ± 1,1
65	SO ₂ NHCH ₂ (3-ClC ₆ H ₄)	3-CO ₂ H	3,8 ± 0,2	42,5 ± 7,2

Tabela 2 (continuação). Grupamentos R₁ e R₂ para os derivados do oxindol (série 2) juntamente com os valores experimentais de *CI*₅₀ para a Shp2 e para a Shp1, em μmol/L.

Estrutura	R ₁	R ₂	<i>CI</i> ₅₀ ^{Shp2}	<i>CI</i> ₅₀ ^{Shp1}
66	SO ₂ NHCH ₂ (4-ClC ₆ H ₄)	3-CO ₂ H	1,4 ± 0,6	18 ± 7,2
67	SO ₂ NHCH ₂ (4-ClC ₆ H ₄)	4-CO ₂ H	5,9 ± 1,4	18,1 ± 3,7
68	SO ₂ NHCH ₂ (2-ClC ₆ H ₄)	4-CO ₂ H	7,4 ± 2,4	60,8 ± 26,1
69	SO ₂ NHCH ₂ (2-ClC ₆ H ₄)	3-CO ₂ H	5,5 ± 0,4	15,9 ± 1,7
70	SO ₂ NHCH ₂ (3-CF ₃ -4-Cl-C ₆ H ₃)	2-CO ₂ H	22,9 ± 7	32,4 ± 17,9
71	SO ₂ NHCH ₂ (4-FC ₆ H ₄)	3-CO ₂ H	1 ± 0,2	18,3 ± 5,2
72	SO ₂ NHCH ₂ (4-FC ₆ H ₄)	4-CO ₂ H	6,3 ± 2,3	10,8 ± 4,1
73	SO ₂ NHCH ₂ (3-Cl-4-F-C ₆ H ₃)	3-CO ₂ H	4,8 ± 0,6	122,8 ± 74,5
74	SO ₂ NH(CH ₂) ₂ (2-Cl-4-ClC ₆ H ₃)	3-CO ₂ H	5,2 ± 1,7	73,4 ± 36,9
75	SO ₂ NHCH ₂ (3-CF ₃ -4-Cl-C ₆ H ₃)	4-CO ₂ H	8,3 ± 1,7	43,1 ± 7,2
76	SO ₂ NHCH ₂ (3-CF ₃ -4-Cl-C ₆ H ₃)	3-CO ₂ H	10,6 ± 1,5	74,2 ± 8,0

As estruturas dos 76 ligantes usados neste trabalho (série 1 e 2) foram preparadas com o programa Spartan'14 (Wavefunction, Inc.) e submetidas a uma análise conformacional usando o método de Monte Carlo com o campo de força MMFF (*Merck Molecular Force Field*), resultando em uma distribuição de conformeros. A geometria do conformero mais estável para cada composto foi então reotimizada com o método semi-empírico PM3 [STEWART, 1989].

Os ligantes reotimizados foram então docados no sítio ativo da Shp2 (código 4RDD) e da Shp1 (código 4GS0) usando a função de pontuação selecionada através do estudo de redocagem.

Os complexos proteína-ligante contendo as “poses” (orientação e conformação) de maior pontuação para cada composto foram exportados e usados para preparar os arquivos de entrada dos cálculos semi-empíricos. Para reduzir o tempo de cálculo, os resíduos de aminoácidos contendo pelo menos um átomo dentro de um raio de 6,0 Å ao redor do composto mais volumoso da série 1 (composto **25**) foram selecionados para compor o sítio de interação da Shp2; os resíduos de aminoácidos correspondentes foram selecionados para compor o sítio de interação da Shp1 (Tabela 3). Para manter o sistema com camada fechada, átomos de hidrogênio foram adicionados nas ligações truncadas, através do programa Babel [O'BOYLE *et al.*, 2011].

Tabela 3. Resíduos de aminoácidos usados para compor os sítios de interação da Shp2 e da Shp1

Enzima	Resíduos de aminoácidos
Shp2 (4RDD)	278, 279, 280, 281, 282, 361, 362, 363, 364, 365, 366, 423, 424, 425, 426, 427, 428, 459, 460, 461, 462, 463, 464, 465, 466, 505, 506, 507, 508, 509, 510
Shp1 (4GS0)	275, 276, 277, 278, 279, 355, 356, 357, 358, 359, 360, 417, 418, 419, 420, 421, 422, 453, 454, 455, 456, 457, 458, 459, 460, 499, 500, 501, 502, 503, 504

As estruturas foram otimizadas com o método do orbital molecular PM7 [STEWART, 2013], disponível no programa MOPAC 2016 (Stewart Computational Chemistry), usando orbitais moleculares localizados para as estruturas contendo um grande número de átomos [STEWART, 1996]. O método PM7 é bastante rápido e pode ser aplicado rotineiramente em sistemas contendo centenas de átomos, e possui erro médio atribuído ao calor de formação de compostos orgânicos de apenas 4,0 kcal.mol⁻¹ [STEWART, 2013]. Para evitar uma possível desestruturação do sítio de interação, os átomos das ligações peptídicas (N-C α -C=O) tiveram suas coordenadas fixadas durante a otimização da estrutura, que foi efetuada seguindo os critérios padrão do programa, mas com o uso de orbitais moleculares localizados, que não alteram na prática os resultados em comparação com o método SCF, mas aceleraram significativamente os cálculos [STEWART, 1996]. A inclusão de efeitos de campo na resolução da equação de Schrödinger foi baseada no modelo COSMO, desenvolvido por Klamt e Schüümann (1993), no qual o soluto é imerso em um contínuo de constante dielétrica (ϵ) adequada [KLAMT & SCHUUMAN, 1993]. Para o interior das proteínas, existem muitos valores sugeridos na literatura, sendo um dos mais aceitos $\epsilon = 4,0$ [GILSON & HONIG, 1986], adotado neste trabalho.

A otimização das geometrias moleculares foi realizada em duas etapas: na primeira, apenas as coordenadas dos átomos de hidrogênio foram otimizadas; esta etapa foi seguida pela otimização das coordenadas de todas as cadeias laterais dos aminoácidos e de todos os átomos do ligante. A entalpia de interação (ΔH_{int}) foi então calculada para cada composto de acordo com a equação 1:

$$\Delta H_{int} = \Delta H_f^{complexo} - (\Delta H_f^{sítio} + \Delta H_f^{ligante}) \quad (1)$$

Na equação 1, $\Delta H_f^{complexo}$ é a entalpia de formação do complexo formado pelo sítio e o ligante usando $\epsilon=4,0$, $\Delta H_f^{sítio}$ é a entalpia de formação do sítio na conformação apo (código 4HJP para a Shp1 e código 3ZM1 para a Shp2) usando $\epsilon=4,0$ e $\Delta H_f^{ligante}$ é a entalpia de formação do ligante em água ($\epsilon=78,4$), pois o ligante fica solubilizado no meio aquoso antes de sua entrada no sítio ativo da enzima.

Para a série 1, a eficiência de predição da seletividade dos derivados de indolina com base nos valores de ΔH_{int} obtidos foi verificada por um teste estatístico pareado com nível de confiança de 99%. O valor de t da distribuição de *Student* (t_{calc}) foi calculado da seguinte forma [SKOOG, 2013]:

$$t_{calc} = \frac{\overline{\Delta\Delta H_{int}}}{s_d/\sqrt{N}} \quad (2)$$

O termo $\overline{\Delta\Delta H_{int}}$ representa o valor médio das diferenças entre os valores de ΔH_{int} obtidos para Shp2 e Shp1 e deve ter uma distribuição normal. O termo s_d representa o desvio padrão estimado de $\Delta\Delta H_{int}$ obtido pela equação 3 e N é o grau de liberdade.

$$\Delta\Delta H_{int} = \Delta H_{int}^{Shp2} - \Delta H_{int}^{Shp1} \quad (3)$$

A hipótese nula (H_0) para comparação estabelece que os valores de ΔH_{int}^{Shp2} são estatisticamente iguais aos valores de ΔH_{int}^{Shp1} e será aceita se t_{calc} estiver entre os valores tabelados de $t_{crítico}$ ($t_{crítico} = \pm 2,676$, teste de duas caudas, N= 51 graus de liberdade, nível de confiança de 99%). Desta forma, duas hipóteses alternativas podem ser formuladas (H_1 e H_2). A hipótese alternativa H_1 estabelece que os valores de ΔH_{int}^{Shp2} são estatisticamente menores (mais favoráveis) aos valores de ΔH_{int}^{Shp1} e será aceita se t_{calc} for menor que $t_{crítico}$. A hipótese alternativa H_2 estabelece que os valores de ΔH_{int}^{Shp2} são estatisticamente maiores (menos favoráveis) aos valores de ΔH_{int}^{Shp1} e será aceita se t_{calc} for maior que $t_{crítico}$.

Para os inibidores da série 2 estão disponíveis na literatura os valores de CI_{50} tanto para a Shp2 quanto para a Shp1 [WU *et al.*, 2008, LAWRENCE *et al.*, 2008]. A disponibilidade de valores de CI_{50} para as duas enzimas permite que se determine a seletividade de forma quantitativa e, por este motivo, foi adotado um procedimento diferente para verificar a eficiência da previsão da seletividade destes compostos. A seletividade relativa (SR) para a série 2 foi calculada da seguinte maneira:

$$SR = \frac{CI_{50}^{Shp1}}{CI_{50}^{Shp2}} \pm s_{SR} \quad (4)$$

com o desvio padrão (s_{SR}) determinado pela equação 5:

$$s_{SR} = \left(\sqrt{\left(\frac{s_{CI_{50}^{Shp1}}}{CI_{50}^{Shp1}} \right)^2 + \left(\frac{s_{CI_{50}^{Shp2}}}{CI_{50}^{Shp2}} \right)^2} \right) SR \quad (5)$$

Depois a obtenção dos valores de $\Delta\Delta H_{int}$ através da equação 3, a diferença percentual relativa das entalpias de interação ($\% \Delta\Delta H_{int}$) foi calculada através da equação 6:

$$\% \Delta\Delta H_{int} = \left(\frac{\Delta\Delta H_{int}}{\Delta H_{int}^*} \right) \times 100 \quad (6)$$

Onde ΔH_{int}^* representa o valor de ΔH_{int} mais favorável (mais negativo) obtido para cada composto. Os valores de $\Delta\Delta H_{int}$ foram comparados com os dados de SR.

4.3 Construção do modelo de previsão da atividade de inibidores da Shp2

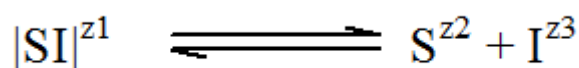
A ligação seletiva de moléculas para formar complexos é de importância central para a farmacologia e a química medicinal. Apesar de fatores cinéticos poderem influenciar os rendimentos de diferentes complexos moleculares em ambientes celulares, os fatores primordiais que devem ser considerados na análise do reconhecimento molecular são termodinâmicos.

Uma reação química avança espontaneamente no sentido da diminuição da energia livre e, portanto, o equilíbrio termodinâmico é alcançado quando a variação da energia livre (ΔG) da reação é igual à zero.

$$\Delta G = \Delta G^\theta + RT \ln \frac{\{Produtos\}}{\{Reagentes\}} = 0 \quad (7)$$

Onde ΔG^θ é a variação de energia livre padrão (temperatura (T) de 25°C, concentração de 1 mol L⁻¹ de todos os reagentes e pressão de 1 atm) e R é a constante fundamental dos gases (R= 8,314 J mol⁻¹ K⁻¹).

Neste caso, quando o equilíbrio termodinâmico é alcançado ($\Delta G=0$), a relação entre os produtos e reagentes torna-se uma constante, chamada constante de equilíbrio ou constante termodinâmica (K). Tradicionalmente, nas reações de inibição enzimática a constante refere-se ao equilíbrio de dissociação do complexo enzima-inibidor, sendo chamada de constante de inibição. Neste trabalho, o termo K_i^* será utilizado para representar a constante de inibição termodinâmica, que é uma função das concentrações efetivas (atividades) das espécies envolvidas no equilíbrio, e assim diferenciá-la da constante de inibição obtida através de métodos espectroscópicos (K_i) que é uma função das concentrações formais dos reagentes e produtos. Existem diferentes tipos de inibição enzimática, mas assumiremos, por simplicidade, que a inibição é do tipo competitiva, em que a interação entre a enzima e o inibidor ocorre no sítio ativo enzimático. Assim, podemos representar o equilíbrio da seguinte forma:



Onde |SI| representa o complexo sítio-inibidor com carga total z1, S representa o sítio ativo vazio com carga z2 e I representa o inibidor com carga z3, em solução.

Com base neste equilíbrio, a constante de equilíbrio de inibição (K_i^*) pode ser escrita da seguinte forma:

$$K_i^* = \frac{\{S\}\{I\}}{\{SI\}} \quad (8)$$

Onde {X} representa a chamada atividade das espécies envolvidas no equilíbrio. Podemos definir atividade como a “concentração ativa” ou “efetiva” das espécies

envolvidas no equilíbrio. Neste trabalho, adotar-se-á a expressão “concentração efetiva” para $\{X\}$, pois em assuntos envolvendo inibição enzimática, o termo “atividade” já é empregado quando se refere à potência de um determinado inibidor.

Nos métodos de análise eletroanalíticos é possível determinar as concentrações efetivas $\{X\}$ das espécies envolvidas num determinado equilíbrio. Porém, na maioria dos ensaios enzimáticos são utilizados métodos espectroscópicos de análise e com estes não é possível determinar as concentrações efetivas, mas apenas as concentrações formais $[X]$. Desta forma, a constante de inibição obtida é uma constante de equilíbrio em função de concentrações formais:

$$K_i = \frac{[S][I]}{[SI]} \quad (9)$$

A concentração efetiva $\{X\}$ e a concentração formal $[X]$ estão relacionadas da seguinte maneira:

$$\{X\} = \gamma [X] \quad (10)$$

onde γ é definido como o coeficiente de atividade da espécie em questão. Os coeficientes de atividade são utilizados para descrever quantitativamente o efeito dos eletrólitos sobre as espécies carregadas presentes no equilíbrio. Quando as soluções são suficientemente diluídas (diluição infinita), tem-se uma solução idealmente diluída. Neste caso $\gamma = 1$ e podemos dizer que $\{X\} = [X]$. De fato, a equação 9 pode ser usada para representar K_i em forças iônicas baixas, próximas de zero, pois neste caso o K_i^* se iguala ao K_i . Com isso, diz-se que as interações iônicas e as interações íon-solvente não são levadas em conta e as espécies iônicas coexistem como cargas pontuais muito afastadas. Em forças iônicas mais altas, no entanto, é necessário considerar o efeito eletrolítico na concentração efetiva das espécies envolvidas no equilíbrio. Assim, em soluções iônicas reais, mais concentradas, deve-se calcular os coeficientes de atividade das espécies envolvidas no equilíbrio.

Substituindo a equação 10 e 9 na equação 8, temos:

$$K_i^* = \frac{\gamma_S \gamma_I}{\gamma_{SI}} K_i \quad (11)$$

Os coeficientes de atividade podem ser calculados conhecendo-se a força iônica (μ) do meio, uma medida de sua concentração de íons. Desta forma, a força iônica (μ) foi calculada através da equação **12**:

$$\mu = \frac{1}{2} \sum C_i Z_i^2 \quad (12)$$

onde C_i corresponde à concentração de cada íon em solução e Z_i corresponde à carga pontual de cada íon.

Para o cálculo dos coeficientes de atividade usamos a equação de Guntelberg (equação **13**), que pode ser aplicada em forças iônicas de até $0,1 \text{ molL}^{-1}$ [GUNTELBERG, 1926].

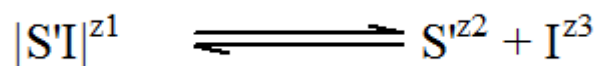
$$-\log \gamma_i = \frac{A |z_i|^2 \sqrt{\mu}}{1 + \sqrt{\mu}} \quad (13)$$

onde A é uma constante que depende da constante dielétrica ε da solução e da temperatura T do experimento:

$$A = 1,825 \times 10^6 (\varepsilon T)^{-3/2} \quad (14)$$

Na superfície do sítio ativo da Shp2 existem dois resíduos de aminoácidos básicos o suficiente para estarem protonados no pH fisiológico e, portanto, carregados positivamente, Arg361 e Lys363. Suas cadeias laterais se posicionam para fora do sítio, de forma a interagir com o meio aquoso. Devido à presença dessas cargas, é possível que ocorra um gradiente na concentração iônica e, portanto, na força iônica do meio aquoso, sendo a força iônica próxima à entrada do sítio maior do que a do interior do solvente. Porém, estimar o aumento da força iônica nesta região é bastante complicado. Desta forma, decidiu-se simplificar este problema assumindo que a “atmosfera” iônica é homogênea em todo o meio reacional, gerando uma força iônica igual a $0,075 \text{ mol L}^{-1}$ (cálculo no item 5.3).

Assumindo que os íons em solução não entram no interior do sítio ativo, iremos tratar o efeito eletrolítico apenas no meio aquoso, através do seguinte equilíbrio:



onde S' corresponde à região aquosa imediatamente adjacente à superfície do sítio. Esta região possui carga $z_2 = +2$ devido à presença das cadeias laterais dos resíduos Arg361 e Lys363, que interagem com a solução. |S'I| corresponde à entrada do inibidor I na região S'. Desta forma, a entrada de um inibidor de carga igual a -1 tornaria $z_1 = +1$. As cargas presentes na região S' (z_1 e z_2) serão consideradas como pontuais conferindo, assim, uma penalidade pelo aumento desconhecido da força iônica nesta região (Figura 15).

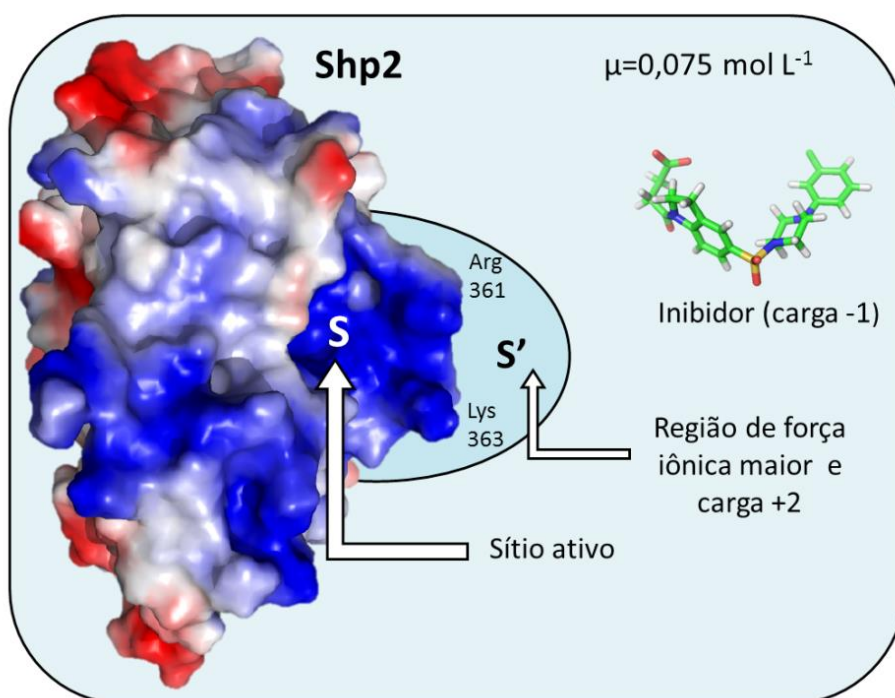


Figura 15. Esquema da superfície eletrostática da Shp2 e representação da região próxima à entrada do sítio.

Apesar de não estarem disponíveis na literatura os valores de K_i dos derivados da indolina (série 1), sabe-se que os valores CI_{50} se correlacionam aos valores de K_i através da equação de Cheng-Prusoff [CHENG & PRUSOFF, 1973]:

$$CI_{50} = K_i \left(1 + \frac{S}{K_m} \right) \quad (15)$$

Onde S corresponde à concentração de substrato e K_m corresponde à constante de Michaelis-Menten.

Substituindo a equação 11 na equação 15, tem-se:

$$CI_{50} = \frac{K_i^*}{\gamma^*} \left(1 + \frac{S}{K_m} \right) \quad (16)$$

onde o termo γ^* representa a razão dos coeficientes de atividade das espécies envolvidas no equilíbrio de dissociação sítio-inibidor:

$$\gamma^* = \frac{\gamma_{SI} \gamma_I}{\gamma_S \gamma_I} \quad (17)$$

Desta forma, é possível calcular a concentração efetiva de inibidor responsável por inibir 50% da atividade enzimática (CI_{50}^*):

$$\gamma^* CI_{50} = K_i^* \left(1 + \frac{S}{K_m} \right) = CI_{50}^* \quad (18)$$

Depois de obtidos os valores de pCI_{50}^* a partir dos valores experimentais de CI_{50} (em mol L⁻¹) dos derivados da indolina para a Shp2, buscou-se construir um modelo termodinamicamente consistente para a predição da atividade de novos inibidores da Shp2.

De forma geral, a formação do complexo sítio-inibidor pode ser conceitualmente dividida em duas etapas: Na primeira ocorre o transporte do inibidor da fase aquosa até o interior da enzima e na segunda etapa ocorre a interação entre o inibidor e o sítio ativo, conforme mostrado no esquema da Figura 16.

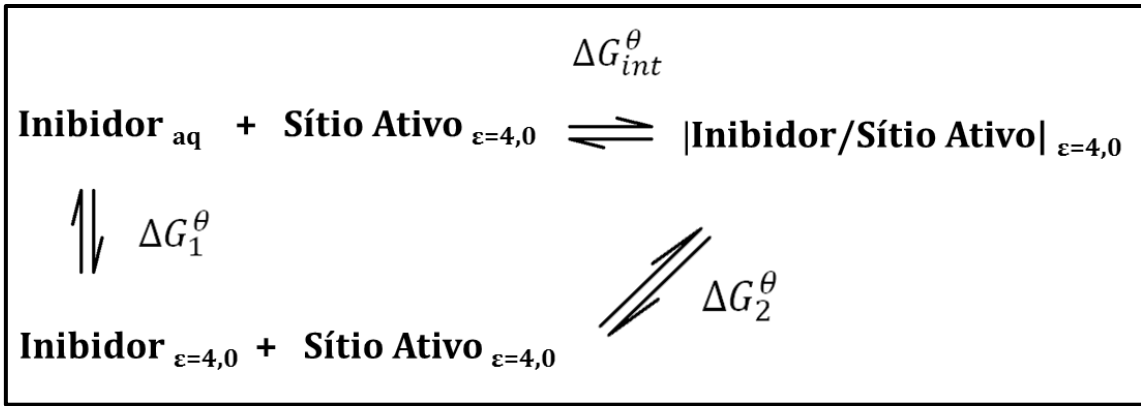


Figura 16. Ciclo termodinâmico com as principais etapas da formação do complexo proteína-ligante. O valor $\varepsilon = 4,0$ refere-se à representação do restante da proteína como um meio contínuo de constante dielétrica uniforme.

Dessa forma, a variação de energia livre padrão associada à interação entre a enzima e o inibidor que leva à formação do complexo enzima-inibidor, ΔG_{int}^{θ} , pode ser escrita como o somatório da energia livre de Gibbs padrão da partição do ligante entre a fase aquosa e a proteína (ΔG_1^{θ}) e da energia livre de Gibbs padrão da interação entre o ligante e o sítio ativo no interior da proteína (ΔG_2^{θ}):

$$\Delta G_{int}^{\theta} = \Delta G_1^{\theta} + \Delta G_2^{\theta} \quad (19)$$

No caso de complexos enzima/inibidor, ΔG_{int}^{θ} corresponde ao oposto da variação da energia livre de Gibbs padrão de dissociação (ΔG_{diss}^{θ}), a qual pode ser relacionada com a constante de inibição termodinâmica (K_i^*) da seguinte forma:

$$\Delta G_{int}^{\theta} = -\Delta G_{diss}^{\theta} = RT \ln K_i^* = 2,303 RT \log K_i^* \quad (20)$$

Apesar dos valores de K_i^* para os derivados da indolina não estarem disponíveis na literatura, eles podem ser considerados proporcionais aos valores de CI_{50}^* (equação 18) e, portanto, a equação acima deve se manter válida para os dados de CI_{50}^* a menos de um fator multiplicativo.

Substituindo ΔG_{int}^{θ} na equação 20 e substituindo K_i^* por CI_{50}^* , temos:

$$2,303 RT \log K_i^* = 2,303 RT \log CI_{50}^* + B = \Delta G_1^{\theta} + \Delta G_2^{\theta} \quad (21)$$

onde B é uma constante, igual a $2,303RT\log C$, onde C é a constante de proporcionalidade entre K_i^* e CI_{50}^* .

O termo ΔG_1^θ corresponde à energia livre associada com o processo de transferência do inibidor da fase aquosa para o interior do sítio ativo. Assumindo que esta transferência seja um processo passivo, a eficiência com a qual o inibidor é transferido para o interior do sítio pode ser relacionada com a energia livre de solvatação (E_{solv}) do inibidor. Quanto mais favorável (negativo) for o valor de E_{solv} mais afinidade o composto terá pelo meio aquoso e menos afinidade terá pelo interior do sítio ativo da enzima, que é um meio com um caráter mais lipofílico. Os valores de E_{solv} foram calculados através do método SM5.4 [CRAMER *et al.*, 1992; CHAMBERS *et al.*, 1996; GIESEN *et al.*, 1997] do programa Spartan'14 (Wavefunction Inc.). O método SM5.4 baseia-se na estimativa de efeitos eletrostáticos pelo modelo de Born generalizado com efeitos de blindagem e de desblindagem dielétrica dependente da forma molecular e nas estimativas do efeito da primeira camada de solvatação a partir de tensões superficiais atômicas dependentes das vizinhanças, combinadas com áreas de superfície acessíveis ao solvente para cada átomo no soluto. Parâmetros são obtidos por ajuste empírico a dados experimentais. O modelo é aplicável nas otimizações de geometrias realizadas com o método PM3. Além de moléculas contendo os átomos de hidrogênio e carbono, também estão incluídos na parametrização do método SM5.4 moléculas contendo heteroátomos (N, O, S, F, Cl, Br, I).

O termo ΔG_2^θ é composto de um termo entálpico, ΔH_{int} , e um termo entrópico, $-T\Delta S_{\text{int}}$. A variação da entalpia de interação (ΔH_{int}) foi calculada conforme descrito no item 4.2. A entropia de um ligante está associada às entropias translacional, vibracional, rotacional e conformacional. Assumindo-se as perdas de entropias translacional, vibracional e rotacional no processo de interação com a enzima como similares para todos os compostos avaliados, pois estes têm estruturas bastante semelhantes, e como a função de predição será posteriormente calibrada com dados experimentais de atividade, o termo entrópico $-T\Delta S_{\text{int}}$ referente à ΔG_2^θ será considerado como sendo essencialmente conformacional. Sendo assim, este termo será considerado como apenas proporcional ao número de ligações rotacionáveis que se tornaram “congeladas” (N_{LRC}) como resultado da interação do inibidor com o sítio ativo, conforme proposto por Wang e colaboradores (1994). Os demais termos entrópicos não são nulos, mas por serem aproximadamente

constantes na série avaliada, serão incluídos como uma constante na equação **21**, gerada pela calibração. Logo, pode-se reescrever a equação **21** da seguinte forma:

$$RT\log CI_{50}^* = c_1 E_{solv} + c_2 \Delta H_{int} + c_3 N_{LRC} + c_4 \quad (22)$$

na qual os coeficientes c_1 - c_4 podem ser determinados ajustando-se a equação aos valores experimentais de $\log CI_{50}$ para a série de compostos da literatura, através de uma análise por regressão linear múltipla. O coeficiente c_4 representa o valor onde a reta intercepta o eixo da variável dependente e inclui a constante B e os termos entrópicos translacionais e rotacionais, aproximadamente constantes.

Os inibidores da série 2 foram utilizados para validar o modelo construído e, por fim, o modelo foi utilizado na predição da atividade relativa dos novos compostos *N*-arilpiperazínicos (série 3) sintetizados pelo Dr. Ciro Gonçalvez e Sá no Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio®) [SÁ, 2017]. A estrutura geral dos compostos da série 3 está mostrada na Figura 17 e os grupamentos R e Ar para cada derivado estão listados na Tabela 4.

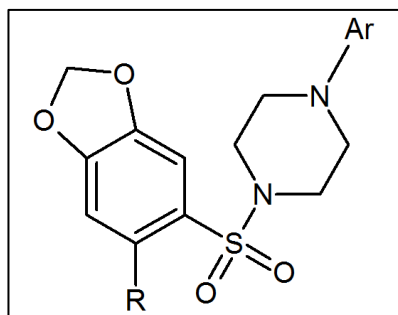


Figura 17. Estrutura geral dos compostos *N*-arilpiperazínicos (série 3)

Tabela 4. Grupamento R e Ar dos compostos *N*-arilpiperazínicos (série 3).

Composto	R	Ar
77	-CH ₃	Pirimidin-2-il
78	-CH ₂ CH ₃	Pirimidin-2-il
79	-CH ₂ CH ₂ COOH	Pirimidin-2-il
80	-CH ₂ CH ₂ COOCH ₃	Pirimidin-2-il
81	-CH ₂ CH ₂ COOCH ₂ CH ₃	Pirimidin-2-il

Tabela 4 (continuação). Grupamento R e Ar dos compostos *N*-arilpiperazínicos (série 3).

Composto	R	Ar
82	-CH ₃	2,3 diclorofenila
83	-CH ₂ CH ₃	2,3 diclorofenila
84	-CH ₂ CH ₂ COOH	2,3 diclorofenila
85	-CH ₂ CH ₂ COOCH ₃	2,3 diclorofenila
86	-CH ₂ CH ₂ COO CH ₂ CH ₃	2,3 diclorofenila
87	-CH ₃	Fenila
88	-CH ₂ CH ₃	Fenila
89	-CH ₂ CH ₂ COOH	Fenila
90	-CH ₂ CH ₂ COOCH ₃	Fenila
91	-CH ₂ CH ₂ COOCH ₂ CH ₃	Fenila
92	-CH ₃	2-metoxofenila
93	-CH ₂ CH ₃	2-metoxofenila
94	-CH ₂ CH ₂ COOH	2-metoxofenila
95	-CH ₂ CH ₂ COOCH ₃	2-metoxofenila
96	-CH ₂ CH ₂ COOCH ₂ CH ₃	2-metoxofenila

5 RESULTADOS E DISCUSSÃO

5.1 Redocagem

O estudo de redocagem mostrou que todas as funções de pontuação disponíveis no programa GOLD 5.5 (CCDC) foram eficientes em prever a pose (orientação e conformação) do ligante cocristalizado (Figura 18), mas a função de pontuação *Chemscore* apresentou um resultado ligeiramente melhor que os outros. Os valores do desvio médio quadrático (RMSD) para cada uma das funções testadas foram de 0,45 Å; 0,70 Å; 0,42 Å e 0,45 Å para as funções de pontuação *ChemPLP*, *Goldscore*, *Chemscore* e *ASP*, respectivamente. Com base nesse resultado, a função *Chemscore* foi escolhida para as próximas etapas do trabalho.

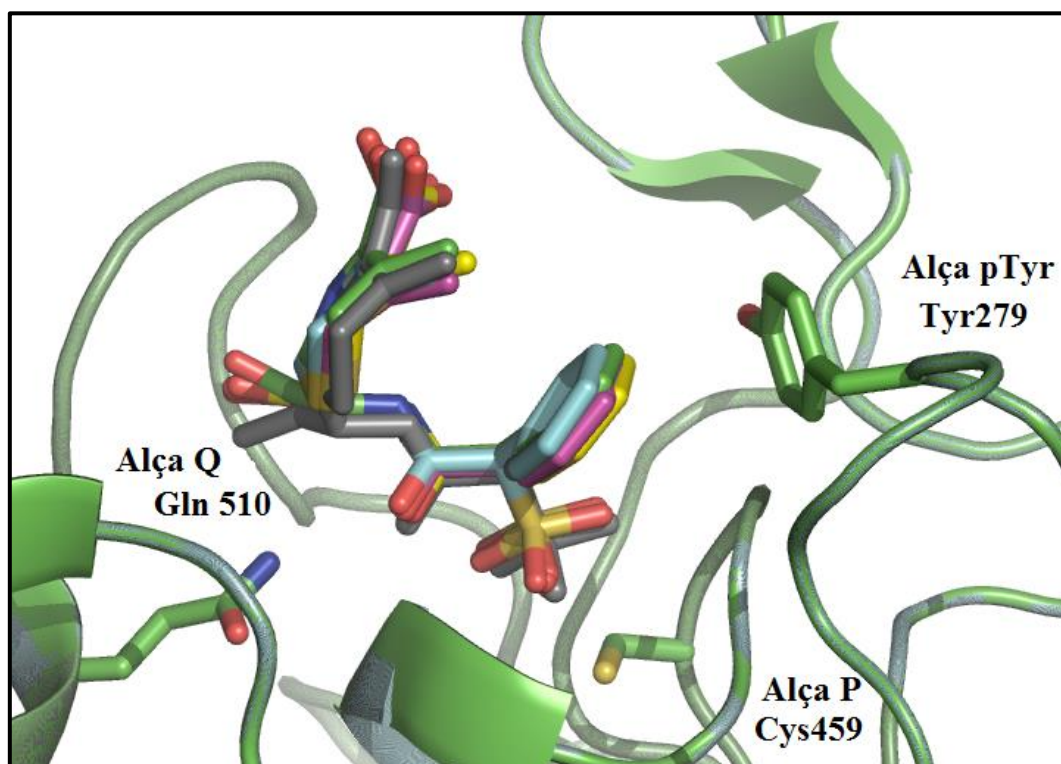


Figura 18. Sobreposição da estrutura original da Shp2 (4RDD) com o derivado da cefsulodina cocristalizado (em tubos, átomos de C em verde) com as estruturas provenientes da redocagem com as seguintes funções de pontuação: *ChemPLP* (átomos de C em ciano), *Goldscore* (magenta), *Chemscore* (amarelo) e *ASP* (cinza).

5.2 Estudo da seletividade dos inibidores da Shp2

Após a docagem dos derivados da indolina na Shp2 (4RDD) e na Shp1 (4GS0), as poses dos compostos docados foram consideradas satisfatórias em ambas as enzimas. O perfil de interação foi considerado satisfatório por se assemelhar ao perfil de interação observado entre o derivado da cefsulodina cocristalizado na Shp2 (4RDD). Como pode ser observado na Figura 19A, o derivado da cefsulodina cocristalizado na Shp2 (4RDD) possui um anel aromático interagindo via empilhamento π com o resíduo de aminoácido Tyr279 da alça pTyr e os compostos da série 1 também possuem esta interação. Além disso, o derivado da cefsulodina apresenta o grupamento sulfona interagindo via ligação de hidrogênio com a cadeia principal de resíduos de aminoácidos localizados na alça P (porção mais interna do sítio) e os compostos da série 1 também apresentaram este perfil de interação. Alguns compostos da série 1 também interagiram via ligação de hidrogênio com o resíduo de Gln510 da alça P e tal interação também foi observada no cocristal (Figura 19A).

Devido à semelhança do sítio ativo da Shp2 com o sítio ativo da Shp1, o perfil de interação entre os compostos da série 1 com o sítio da Shp1 (4GS0) também apresentou algumas características semelhantes ao perfil de interação destes compostos com o sítio da Shp2 (4RDD). Também se observou alguns aspectos semelhantes entre os modos de interação dos compostos da série 1 com os do ligante cocristalizado na Shp1 (4GS0) como, por exemplo, a interação via empilhamento π do anel aromático do ligante com o resíduo de tirosina 276 da alça pTYr; além disso, o grupamento sulfona dos compostos da série 1 ficou localizado próximo do grupamento fosfonato do ligante cocristalizado na Shp1 (4GS0), interagindo via ligação de hidrogênio com a cadeia principal de resíduos de aminoácidos localizados na alça P (Figura 19B).

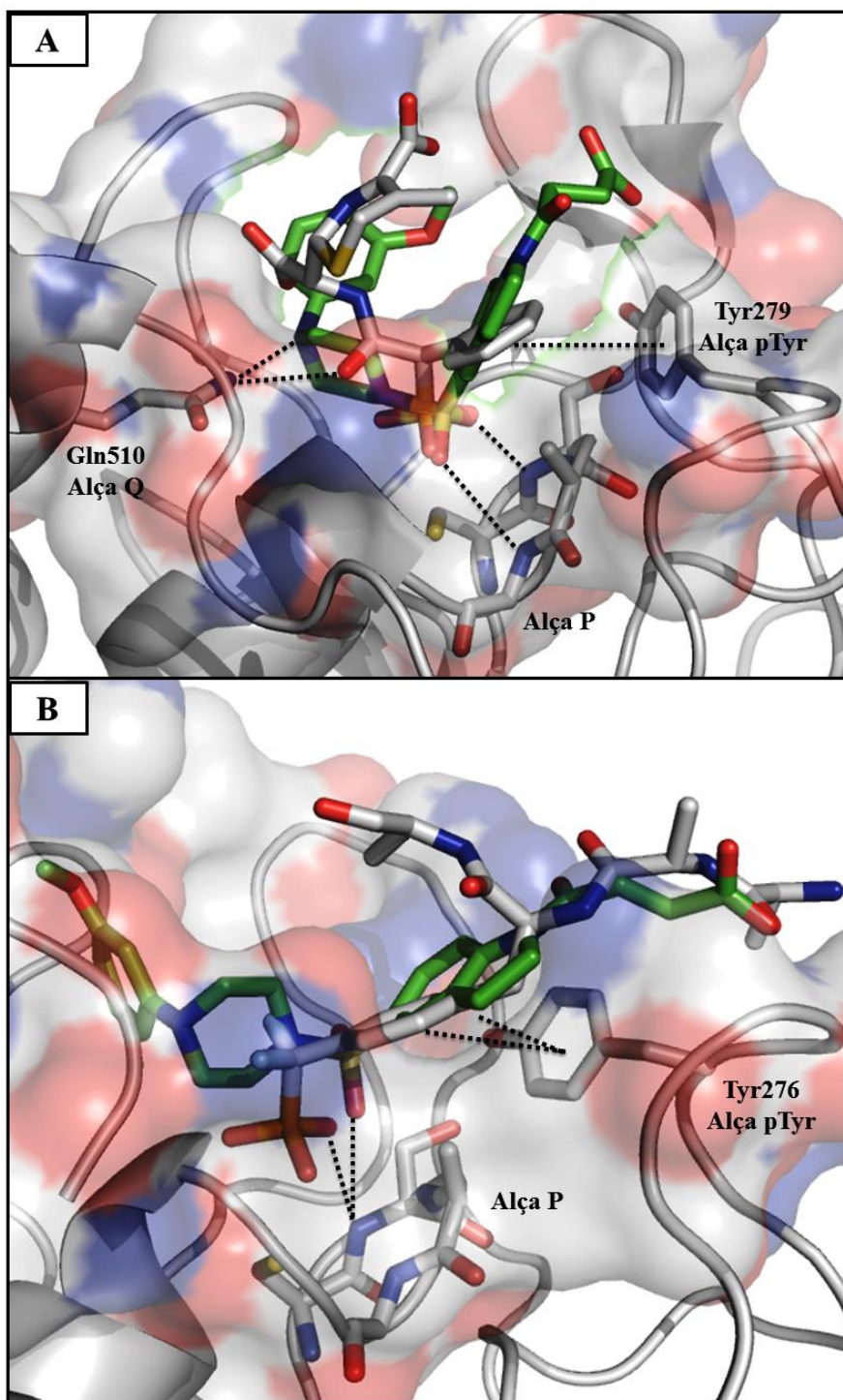


Figura 19. (A) Sobreposição dos modos de interação do composto **10** da série 1 (em tubos, carbonos em verde) com o sítio da Shp2 (4RDD) e do derivado da cefsulodina cocristalizado (em tubos, carbonos em cinza). (B) Sobreposição dos modos de interação do composto **10** da série 1 (em tubos, carbonos em verde) com o sítio da Shp1 (4GS0) e do ligante cocristalizado (em tubos, carbonos em cinza). As interações estão representadas por linhas pontilhadas pretas. Principais resíduos envolvidos nas interações em tubos, carbonos em cinza. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC).

No entanto, não houve diferença significativa entre os valores de pontuação entre a Shp2 e a Shp1, com um nível de confiança de 99% (Tabela 5). Os métodos de docagem molecular são geralmente conhecidos por serem eficazes na reprodução das geometrias de interação, mas simplificações adotadas em suas funções de pontuação podem reduzir o nível de correlação entre as pontuações obtidas e os dados experimentais de atividade, como K_i ou CI_{50} [TACSAY *et al.*, 2015].

Tabela 5. Pontuações obtidas através da docagem molecular com a função *Chemscore* para os compostos da série 1 na Shp1 e na Shp2.

Composto	Pontuação (Shp1)	Pontuação (Shp2)	Composto	Pontuação (Shp1)	Pontuação (Shp2)
1	29,24	30,53	27	32,00	30,17
2	29,77	29,14	28	29,54	30,47
3	31,33	28,20	29	30,80	29,13
4	28,71	28,92	30	31,77	30,88
5	35,72	29,91	31	31,11	31,47
6	28,52	28,89	32	30,36	30,24
7	30,00	30,02	33	31,73	32,73
8	29,01	28,04	34	28,59	28,31
9	28,80	29,80	35	32,31	29,37
10	29,72	28,71	36	32,84	30,55
11	28,34	30,58	37	30,03	28,61
12	28,05	27,67	38	29,07	30,04
13	34,11	27,85	39	31,73	30,03
14	29,56	27,80	40	29,01	28,62
15	25,58	26,55	41	29,87	28,64
16	23,95	27,58	42	31,47	28,73
17	29,28	29,53	43	31,37	31,04
18	27,87	28,96	44	32,72	29,39
19	27,65	28,78	45	26,99	27,95
20	28,68	28,67	46	27,28	29,64
21	25,85	27,36	47	32,92	31,11
22	28,51	28,39	48	29,67	28,68
23	29,12	30,61	49	33,11	29,41

Tabela 5 (continuação). Pontuações obtidas através da docagem molecular com a função *Chemscore* para os compostos da série 1 na Shp1 e na Shp2.

Composto	Pontuação (Shp1)	Pontuação (Shp2)	Composto	Pontuação (Shp1)	Pontuação (Shp2)
24	29,79	29,16	50	31,07	29,81
25	31,72	30,54	51	29,21	33,33
26	33,86	30,90	52	30,88	30,31

Um método melhor para quantificar a energia de interação enzima/ligante foi então explorado. Os métodos mecânicos quânticos semi-empíricos são rápidos o suficiente para serem aplicados a sistemas contendo centenas de átomos de forma rotineira. O método do orbital molecular PM7 é um dos mais recentes métodos semi-empíricos e possui erro médio para o calor de formação de compostos orgânicos de apenas 4,0 kcal.mol⁻¹ [STEWART, 2013]. O método PM7 possui algumas melhorias em relação aos seus antecessores: nele foi feita uma modificação para melhorar a descrição das interações não covalentes, uma característica essencial nesse estudo, e dois pequenos erros no formalismo do NDDO presentes nos métodos semi-empíricos anteriores foram corrigidos. [STEWART, 2013].

Depois de realizada a otimização da geometria dos sítios vazios, ligantes e dos complexos sítio/ligante, as entalpias de interação dos compostos da série 1 foram calculadas em ambas as enzimas (Shp2 e Shp1). Os valores de entalpias de formação estão listados na Tabela 6 e os valores de entalpias de interação (ΔH_{int}) correspondentes estão listados na Tabela 7.

Tabela 6. Entalpias de formação (H_f) em kcal mol⁻¹, para os ligantes da série 1 e para os complexos sítio-ligante obtidos com a Shp2 (4RDD) e a Shp1 (4GS0)

Ligante	$H_f^{complexo}$ (Shp2)	$H_f^{complexo}$ (Shp1)	$H_f^{ligante}$	Ligante	$H_f^{complexo}$ (Shp2)	$H_f^{complexo}$ (Shp1)	$H_f^{ligante}$
1	-2287,00	-2261,17	-296,56	27	-2035,53	-2031,46	-75,46
2	-2262,83	-2281,39	-301,76	28	-2077,08	-2073,23	-109,67
3	-2245,73	-2276,93	-263,84	29	-2068,45	-2064,68	-107,89
4	-2289,77	-2241,65	-298,79	30	-2059,18	-2035,96	-82,03
5	-2294,45	-2184,51	-260,04	31	-2042,29	-1984,51	-85,27
6	-2269,5	-2229,38	-296,5	32	-2077,53	-2048,43	-112,21

Tabela 6 (continuação). Entalpias de formação (H_f) em kcal mol⁻¹, para os ligantes da série 1 e para os complexos sítio-ligante obtidos com a Shp2 (4RDD) e a Shp1 (4GS0)

Ligante	$H_f^{complexo}$ (Shp2)	$H_f^{complexo}$ (Shp1)	$H_f^{ligante}$	Ligante	$H_f^{complexo}$ (Shp2)	$H_f^{complexo}$ (Shp1)	$H_f^{ligante}$
7	-2286,78	-2245,24	-298,4	33	-2059,03	-2038,49	-86,62
8	-2291,01	-2287,45	-300,08	34	-2056,84	-2013,14	-108,35
9	-2288,85	-2227,15	-287,49	35	-2076,26	-2083,74	-117,02
10	-2322,72	-2319,97	-326,43	36	-2055,75	-2013,21	-83,16
11	-2246,89	-2218,03	-281,64	37	-2034,24	-2029,16	-79,1
12	-2256,56	-2256,52	-292,80	38	-2009,19	-2019,25	-66,74
13	-2472,94	-2478,14	-472,78	39	-2044,48	-2023,53	-76,23
14	-2236,38	-2175,81	-258,07	40	-2194,64	-2162,23	-225,3
15	-2300,36	-2254,74	-316,16	41	-2165,91	-2148,46	-195,65
16	-2590,14	-2571,2	-601,27	42	-2083,2	-2069,26	-132,58
17	-2414,4	-2392,88	-448,16	43	-2074,05	-2049,38	-109,28
18	-2305,6	-2293,14	-332,24	44	-2045,95	-2027,71	-76,37
19	-2273,81	-2269,01	-285,82	45	-2026,87	-2001,81	-66,64
20	-2275,53	-2273,16	-304,21	46	-2069,21	-2069,31	-111,67
21	-2288,19	-2323,91	-332,52	47	-2077,14	-2080,1	-114,54
22	-2292,84	-2316,04	-328,69	48	-2110,69	-2104,78	-157,48
23	-2269,2	-2263,55	-302,69	49	-2048,59	-2017,98	-76,62
24	-2051,73	-2036,28	-75,69	50	-2082,15	-2082,56	-117,25
25	-2045,62	-2011,44	-86,8	51	-2037,01	-2005,25	-76,54
26	-2057,51	-1999,97	-84,03	52	-2161,07	-2139,41	-189,87

$H_f^{sítio (Shp2)} = -1892,21 \text{ kcal mol}^{-1}$ $H_f^{sítio (Shp1)} = -1890,38 \text{ kcal mol}^{-1}$

Tabela 7. Entalpias de interação (ΔH_{int}) calculadas pelo método PM7, em kcal mol⁻¹, para os complexos da série 1 obtidos com a Shp2 (4RDD) e com a Shp1 (4GS0).

Composto	ΔH_{int}^{Shp2}	ΔH_{int}^{Shp1}	Composto	ΔH_{int}^{Shp2}	ΔH_{int}^{Shp1}
1	-98,23	-74,23	27	-67,86	-65,62
2	-68,86	-89,25	28	-75,20	-73,18
3	-89,68	-122,71	29	-68,35	-66,41
4	-98,77	-52,48	30	-84,94	-63,55
5	-142,2	-34,09	31	-64,81	-8,86

Tabela 7 (continuação). Entalpias de interação (ΔH_{int}) calculadas pelo método PM7, em kcal mol⁻¹, para os complexos da série 1 obtidos com a Shp2 (4RDD) e com a Shp1 (4GS0).

Composto	ΔH_{int}^{Shp2}	ΔH_{int}^{Shp1}	Composto	ΔH_{int}^{Shp2}	ΔH_{int}^{Shp1}
6	-80,79	-42,5	32	-73,11	-45,84
7	-96,17	-56,46	33	-80,2	-61,49
8	-98,72	-96,99	34	-56,28	-14,41
9	-109,15	-49,28	35	-67,03	-76,34
10	-104,08	-103,16	36	-80,38	-39,67
11	-73,04	-46,01	37	-62,93	-59,68
12	-71,55	-73,34	38	-50,24	-62,13
13	-107,95	-114,98	39	-76,04	-56,92
14	-86,10	-27,36	40	-77,13	-46,55
15	-91,99	-48,20	41	-78,05	-62,43
16	-96,66	-79,55	42	-58,41	-46,3
17	-74,03	-54,34	43	-72,56	-49,72
18	-81,15	-70,52	44	-77,37	-60,96
19	-95,78	-92,81	45	-68,02	-44,79
20	-79,11	-78,57	46	-65,33	-67,26
21	-63,46	-101,01	47	-70,39	-75,18
22	-71,94	-96,97	48	-61,00	-56,92
23	-74,3	-70,48	49	-79,76	-50,98
24	-83,83	-70,21	50	-72,69	-74,93
25	-66,61	-34,26	51	-68,26	-38,33
26	-81,27	-25,56	52	-78,99	-59,16

Observou-se que 80% dos derivados da série 1 apresentaram valores de ΔH_{int}^{Shp2} mais favoráveis em relação aos valores de ΔH_{int}^{Shp1} calculados (Tabela 7). Após a verificação da existência de distribuição normal (teste do histograma) e aplicação do teste t bilateral pareado, verificou-se que há diferença estatística entre os valores de entalpia ($t_{calc} = -4,70 < t_{critico} = -2,40$). Portanto, é correto afirmar que os valores de ΔH_{int}^{Shp2} para os compostos da série 1 são significativamente menores (mais favoráveis) em relação aos valores de ΔH_{int}^{Shp1} em um nível de confiança de 99%. Desta forma, pode-se concluir que o cálculo da entalpia de interação pelo método PM7

mostrou-se eficiente em prever a seletividade desta série de compostos. Para esta série, ΔH_{int}^{Shp2} foi, em média, 17 kcal mol^{-1} mais favorável em relação a ΔH_{int}^{Shp1} .

No caso dos compostos da série 2, observou-se que na maioria dos complexos sítio/ligante de maior pontuação, os ligantes não estavam posicionados no sítio conforme esperado, de acordo como o modo de interação experimental observado na estrutura cristalográfica da Shp2 de código 4RDD. Essa estrutura possui um derivado da cefsulodina co-cristalizado, onde o grupamento sulfona interage com a alça P e um anel aromático interage com a alça Tyr (Figura 18). Por isso, para os arquivos de entrada dos cálculos semiempíricos para esta série, optou-se por usar um critério estrutural para a seleção das poses, usando-se aquelas que melhor se assemelhassem ao modo de interação experimental. Um exemplo é apresentado na Figura 20.

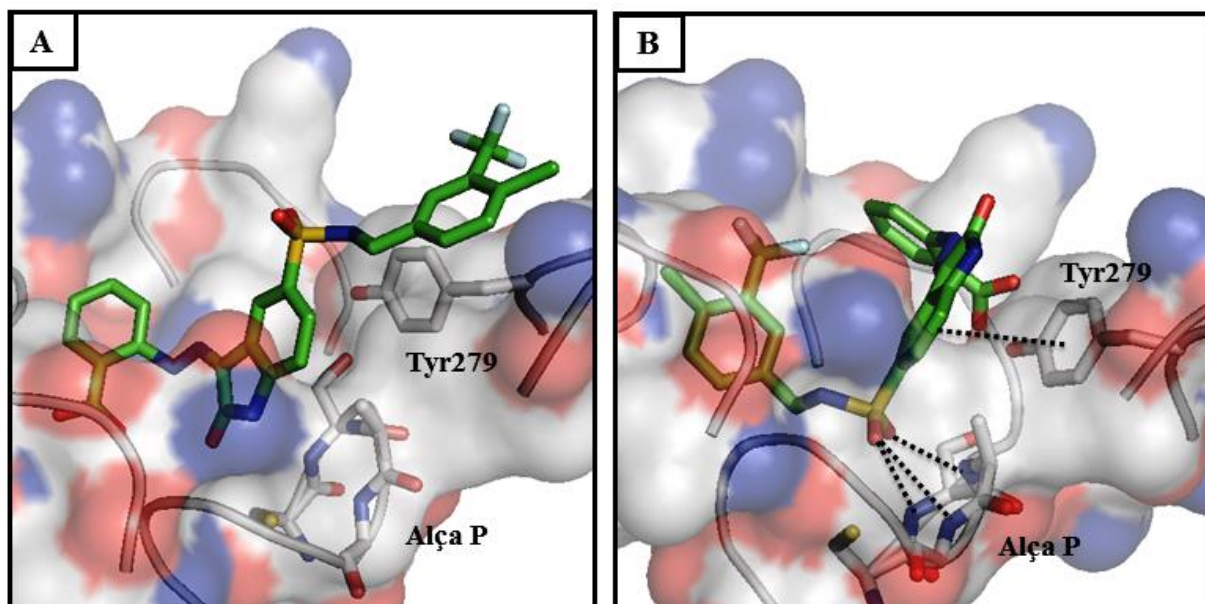


Figura 20. (A) Pose de melhor pontuação do complexo formado entre o composto **70** e a Shp2 (4RDD). (B) Pose escolhida por ser aquela que mais se assemelha ao perfil de interação do ligante cocristalizado. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC).

Os valores de ΔH_f para os ligantes da série 2 e para os complexos sítio-ligante estão listados na Tabela 8. Os valores de ΔH_{int}^{Shp2} e ΔH_{int}^{Shp1} para esta série estão listados na Tabela 9.

Tabela 8. Entalpias de formação (H_f) em kcal mol⁻¹, para os ligantes da série 2 e os complexos sítio-ligante obtidos com a Shp2 (4RDD) e a Shp1 (4GS0)

Composto	$H_f^{complexo}$ (Shp2)	$H_f^{complexo}$ (Shp1)	$H_f^{ligante}$
53	-2010,2	-1997,1	-70,6
54	-2227,0	-2224,8	-252,7
55	-2186,4	-2199,9	-202,1
56	-2328,2	-2255,2	-339,0
57	-2322,9	-2296,3	-341,2
58	-2151,4	-2162,7	-186,0
59	-2223,5	-2217,7	-255,2
60	-2216,6	-2218,0	-257,0
61	-2007,9	-1981,6	-59,5
62	-1991,2	-1985,6	-42,2
63	-2195,5	-2172,4	-226,3
64	-2183,1	-2166,2	-218,5
65	-2204,0	-2193,5	-227,3
66	-2204,0	-2162,1	-226,3
67	-2194,6	-2172,1	-228,3
68	-2203,1	-2185,2	-229,7
69	-2178,8	-2164,5	-226,3
70	-2366,7	-2336,1	-382,6
71	-2246,2	-2198,4	-262,8
72	-2240,0	-2228,8	-264,7
73	-2238,9	-2201,9	-270,8
74	-2246,8	-2192,2	-243,0
75	-2339,9	-2306,4	-391,7
76	-2381,7	-2369,0	-388,8

$$H_f^{sítio (Shp2)} = -1892,21 \text{ kcal mol}^{-1} \quad H_f^{sítio (Shp1)} = -1890,38 \text{ kcal mol}^{-1}$$

Tabela 9. Entalpias de interação (ΔH_{int}) em kcal mol⁻¹, calculadas pelo método PM7 dos compostos da série 2 na Shp2 e na Shp1, diferença entre os valores de entalpias ($\Delta\Delta H_{int}$), diferença percentual relativa ($\% \Delta\Delta H_{int}$) e seletividade relativa experimental (SR)

Composto	ΔH_{int}^{Shp1}	ΔH_{int}^{Shp2}	$\Delta\Delta H_{int}$	$\% \Delta\Delta H_{int}$	SR
53	-36,1	-47,4	-11,3	23,9	2,2 ± 1,4
54	-81,7	-82,1	-0,4	0,5	1,4 ± 1,3
55	-107,4	-92,1	15,3	-14,2	1,5 ± 0,5
56	-25,8	-97,0	-71,1	73,4	19,3 ± 5,5
57	-64,7	-89,5	-24,8	27,7	4,6 ± 2,3
58	-86,3	-73,2	13,1	-15,2	0,9 ± 0,2
59	-72,1	-76,1	-4,0	5,2	3,5 ± 1,4
60	-70,6	-67,4	3,2	-4,6	6,2 ± 2,4
61	-31,7	-56,2	-24,5	43,5	8,7 ± 1,8
62	-53,0	-56,8	-3,8	6,7	9,3 ± 6,1
63	-55,7	-77,0	-21,3	27,7	6,5 ± 3,0
64	-57,3	-72,4	-15,2	21,0	5,5 ± 3,9
65	-75,8	-84,5	-8,7	10,3	11,2 ± 2,0
66	-45,4	-85,5	-40,1	46,9	12,9 ± 7,5
67	-53,4	-74,1	-20,7	27,9	3,1 ± 1,0
68	-65,1	-81,2	-16,2	19,9	8,2 ± 4,4
69	-47,8	-60,3	-12,4	20,6	2,9 ± 0,4
70	-63,1	-91,9	-28,9	31,4	1,4 ± 0,9
71	-45,2	-91,2	-46,0	50,4	18,3 ± 6,4
72	-73,7	-83,1	-9,5	11,4	1,7 ± 0,9
73	-40,7	-75,9	-35,2	46,4	25,6 ± 15,8
74	-58,8	-111,6	-52,8	47,3	14,1 ± 8,5
75	-24,3	-56,0	-31,7	56,6	5,2 ± 1,4
76	-89,8	-100,7	-10,9	10,8	7,0 ± 1,2

Em média, o valor de $\% \Delta\Delta H_{int}$ para os derivados não seletivos, ou seja, aqueles com seletividade relativa experimental (SR) $\leq 1,5$ (**54**, **55**, **58** e **70**), foi de 15%. Por outro lado, para compostos altamente seletivos, isto é, aqueles com SR > 10 (**56**, **65**, **66**, **71**, **73** e **74**), $\% \Delta\Delta H_{int}$ foi em média 50%. O valor médio de $\% \Delta\Delta H_{int}$ entre esses dois extremos é de 33% e pode ser usado como valor de corte em ensaios de triagem virtual

para obter inibidores mais seletivos para Shp2. Considerando este valor de corte para a série 2, foi possível detectar 83% dos compostos mais seletivos e excluir 100% dos compostos não seletivos

Observou-se que a diferença percentual relativa entre ΔH_{int} ($\% \Delta \Delta H_{int}$) foi moderadamente proporcional ao grau de seletividade (SR) de cada composto ($R = 0,67$). Descartando-se a contribuição do composto **73**, que tem um desvio padrão de SR experimental muito alto, obtém-se uma melhor correlação ($R = 0,70$), como pode ser visto na Figura 15.

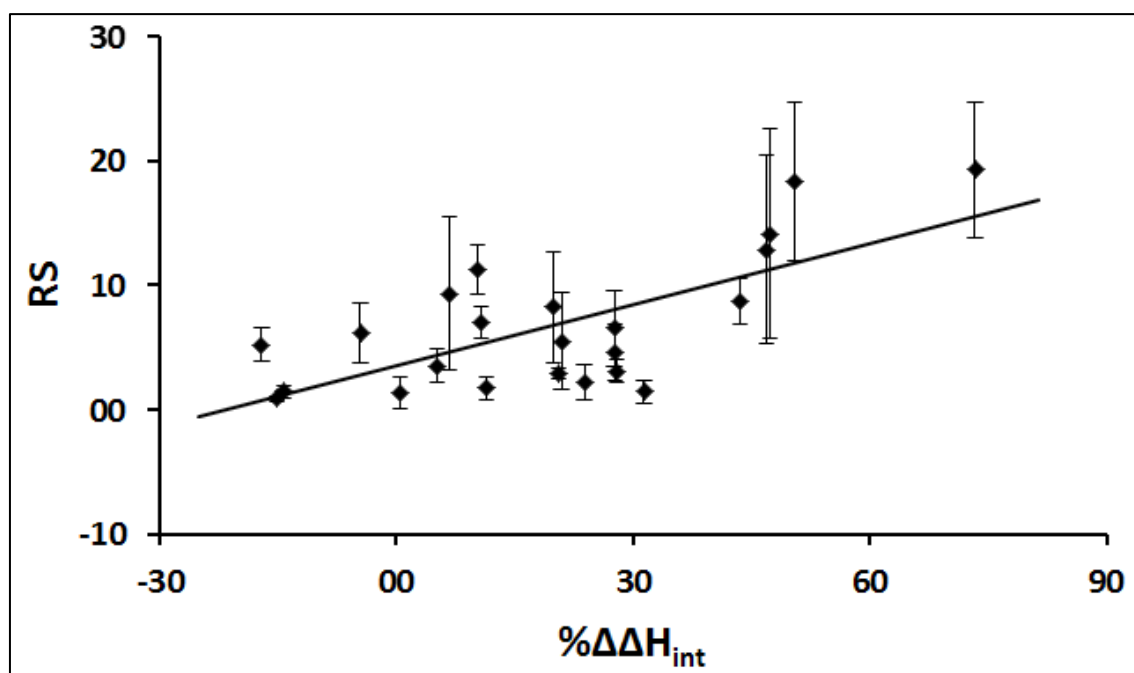


Figura 21. Correlação entre a seletividade relativa (SR), baseada nos dados de CI_{50} experimentais e a diferença percentual ($\% \Delta \Delta H_{int}$) das entalpias de interação dos compostos da série 2.

Em geral, observou-se que os diferentes grupamentos R_2 presentes nos compostos da série 2 se posicionam próximo aos resíduos de aminoácidos da alça $\beta 5$ - $\beta 6$ tanto na Shp2 quanto na Shp1. De fato, as principais diferenças entre a Shp2 e a Shp1 se localizam nessa região. A alça $\beta 5$ - $\beta 6$ é conhecida como a região da superfície do sítio ativo com o maior número de resíduos de aminoácidos não conservados entre as PTP e, portanto, tem uma grande influência na especificidade dos substratos [SCOTT *et al.*, 2010]. Como pode ser observado nas Figuras 22A e 22B, existe um resíduo de Lys364 na alça $\beta 5$ - $\beta 6$ da Shp2 e, ocupando o mesmo lugar, um resíduo de Arg358 na Shp1. Observa-se que, na Shp1, o resíduo de Arg358, por ter uma cadeia lateral mais

volumosa e mais posicionada para o interior do sítio, parece prejudicar a entrada dos grupamentos aromáticos dos compostos **56** e **71** na região entre a alça $\beta 5$ - $\beta 6$ e o resíduo de Tyr279 da alça pTyr (figura 22B). Estes compostos (**56** e **71**) são bastante seletivos, e conforme era esperado, apresentaram entalpia de interação bem mais favorável para a Shp2 em relação à Shp1. A melhor interação destes compostos com a Shp2 parece estar relacionada com a presença de uma cavidade entre a alça $\beta 5$ - $\beta 6$ e o resíduo de Tyr279 da alça pTyr na Shp2 (Figura 22A). A presença desta cavidade parece tornar possível para o anel aromático do composto **56** interagir via empilhamento π com o resíduo Tyr279 e para o grupamento R2 do composto **71** (-COOH) interagir via ligação de hidrogênio com a hidroxila do resíduo Tyr279 (Figura 22C e 22D).

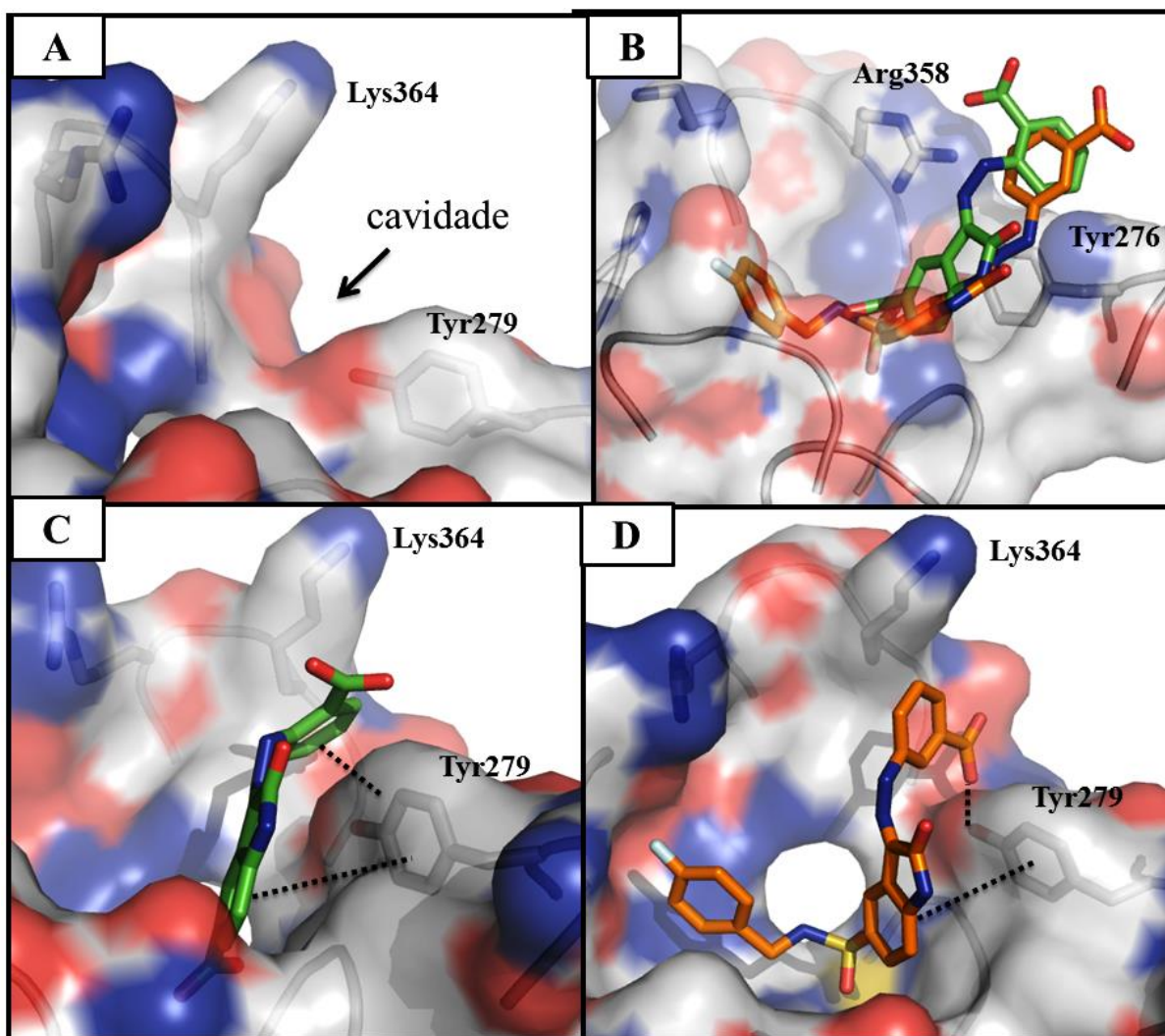


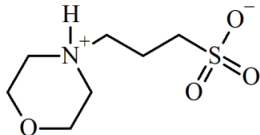
Figura 22. (A) Cavidade entre a alça $\beta 5$ - $\beta 6$ e o resíduo de Tyr279 da alça pTyr da Shp2 (4RDD); (B) Composto **71** (carbonos em laranja) e composto **56** (carbonos em verde) interagindo com o sítio da Shp1 (4GS0); (C) Composto **56** (carbonos em verde) interagindo com o sítio da Shp2 (4RDD); (D) Composto **71** (carbonos em laranja) interagindo com o sítio da Shp2 (4RDD). Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)

5.3 Construção do modelo de predição da atividade de inibidores da Shp2

Os resultados anteriores indicam o cálculo da entalpia de interação pelo método PM7 a partir de geometrias de interação definidas por docagem molecular com a função de pontuação *Chemscore* como um método efetivo para se definir a seletividade relativa de inibidores sobre as fosfatases Shp2 e Shp1. Entretanto, seria mais informativo dispor de um método para o cálculo da atividade dos inibidores, que serão obtidas neste trabalho através da equação **21**, após a devida calibração com dados experimentais. Para tanto, pelos argumentos apresentados neste trabalho, inicialmente é necessário se converter os dados experimentais de CI_{50} em dados de concentração *efetiva* de um inibidor capaz de inibir 50% da atividade enzimática (CI_{50}^*), a qual depende da força iônica do meio reacional utilizado durante a execução dos experimentos.

Segundo Wu e colaboradores (2012), o ensaio de inibição realizado com os derivados da indolina (série 1) foi feito de forma semelhante ao descrito por Chen e colaboradores (2006). Esta metodologia emprega basicamente dois reagentes que, em solução aquosa, produzem íons de forma considerável: o tampão MOPS (ácido 3-morfolinopropano-1-sulfônico) em concentração 25 mmol L⁻¹ (pH 7,0) e NaCl em concentração 50 mmol L⁻¹. Este tipo de tampão empregado é preparado através da mistura equimolar da espécie ácida (zwitterion) com sua base conjugada (dissociada em solução juntamente com um contra íon). Desta forma, a força iônica (μ) do meio reacional foi calculada através da equação **12**. Na Tabela 10 está listado a contribuição de cada íon para a força iônica total (0,075 mol L⁻¹).

Tabela 10. Contribuição de cada íon para a força iônica (μ) do meio reacional

Reagente	Íon	$C_i Z_i^2$	μ (mol L ⁻¹)
NaCl	Na ⁺	0,05	0,05
	Cl ⁻	0,05	
Tampão MOPS 	Espécie ácida	0,0125 + 0,0125	0,025
	Base conjugada	0,0125	
	Contra íon	0,0125	
			μ total = 0,075

Como o experimento de inibição enzimática foi realizado à temperatura ambiente (298 K), considerando-se a constante dielétrica igual a 78,4 para todos os casos (pois assumimos que os íons não entram no sítio ativo da enzima e por isso o

efeito eletrolítico será considerado apenas na solução), tem-se que valor da constante A (equação 14) que será usado na equação de Guntelberg (equação 13) para o cálculo dos coeficientes de atividades é igual a 0,51.

Desta forma, como a carga da superfície do sítio ativo foi assumida como igual à +2, através da equação 13, tem-se que o valor do coeficiente de atividade da superfície do sítio é de 0,36. Os valores de γ^* calculados para os derivados da série 1, de acordo com a equação 17, estão listados na Tabela 11 e variam em função da carga do ligante.

Tabela 11. Coeficientes de atividades (γ) calculados através da equação de Guntelberg e a razão dos coeficientes de atividades (γ^*) em relação à carga (z) dos inibidores.

Γ	$z= 0$	$z= -1$	$z= -2$
γ_I	1,0	0,78	0,36
$\gamma_{ S^+I }$	0,36	0,78	1,0
γ^*	1,0	0,36	0,13

Os valores de CI_{50}^* foram obtidos multiplicando-se os valores experimentais de CI_{50} (em mol L⁻¹) aos valores da razão dos coeficientes de atividade (γ^*) de acordo com a carga do composto (equação 16). Aplicando-se o negativo do logaritmo em base 10 a estes valores, foram obtidos os valores de pCI_{50}^* (Tabela 12).

Tabela 12. Valores experimentais de CI_{50} em $\mu\text{mol L}^{-1}$, razão dos coeficientes de atividade (γ^*) e valores de pCI_{50}^* para os compostos da série 1

Composto	CI_{50}	pCI_{50}	γ^*	pCI_{50}^*	Composto	CI_{50}	pCI_{50}	γ^*	pCI_{50}^*
1	0,70	6,15	0,36	6,60	27	5,70	5,24	1	5,24
2	1,50	5,82	0,36	6,27	28	6,30	5,20	1	5,20
3	2,80	5,55	0,36	6,00	29	6,90	5,16	1	5,16
4	2,80	5,55	0,36	6,00	30	7,10	5,15	1	5,15
5	3,00	5,52	0,36	5,97	31	7,60	5,12	1	5,12
6	3,40	5,47	0,36	5,91	32	7,80	5,11	1	5,11
7	4,50	5,35	0,36	5,79	33	8,00	5,10	1	5,10
8	4,60	5,34	0,36	5,78	34	8,10	5,09	1	5,09
9	4,60	5,34	0,36	5,78	35	8,10	5,09	1	5,09
10	4,70	5,33	0,36	5,77	36	8,20	5,09	1	5,09
11	4,80	5,32	0,36	5,76	37	8,30	5,08	1	5,08

Tabela 12 (continuação). Valores experimentais de CI_{50} em $\mu\text{mol/L}$, razão dos coeficientes de atividade (γ^*) e valores de pCI_{50}^* para os compostos da série 1

Composto	CI_{50}	pCI_{50}	γ^*	pCI_{50}^*	Composto	CI_{50}	pCI_{50}	γ^*	pCI_{50}^*
12	4,80	5,32	0,36	5,76	38	8,40	5,08	1	5,08
13	6,10	5,21	0,36	6,10	39	8,50	5,07	1	5,07
14	8,30	5,08	0,36	5,52	40	9,00	5,05	1	5,05
15	9,00	5,05	0,36	5,49	41	9,30	5,03	1	5,03
16	9,10	5,04	0,36	5,48	42	9,60	5,02	1	5,02
17	11,30	4,95	0,36	5,39	43	9,90	5,00	1	5,00
18	13,20	4,88	0,36	5,32	44	10,40	4,98	1	4,98
19	14,80	4,83	0,36	5,27	45	11,70	4,93	1	4,93
20	16,60	4,78	0,36	5,22	46	12,20	4,91	1	4,91
21	31,90	4,50	0,36	4,94	47	12,60	4,90	1	4,90
22	34,60	4,46	0,36	4,90	48	14,30	4,84	1	4,84
23	143,60	3,84	0,36	4,29	49	19,10	4,72	1	4,72
24	3,20	5,49	1	5,49	50	20,70	4,68	1	4,68
25	4,10	5,39	1	5,39	51	22,80	4,64	1	4,64
26	5,70	5,24	1	5,24	52	60,30	4,22	1	4,22

É interessante destacar que o termo ΔH_{int} já apresenta uma correlação moderada ($R=0,60$) com os dados experimentais de atividade inibitória efetiva (pCI_{50}^*), como pode ser visto no gráfico da Figura 23. Por outro lado, os valores de pCI_{50} apresentam uma correlação mais fraca ($R=0,40$) com os dados de ΔH_{int} , o que reforça a importância em se considerar o efeito eletrolítico.

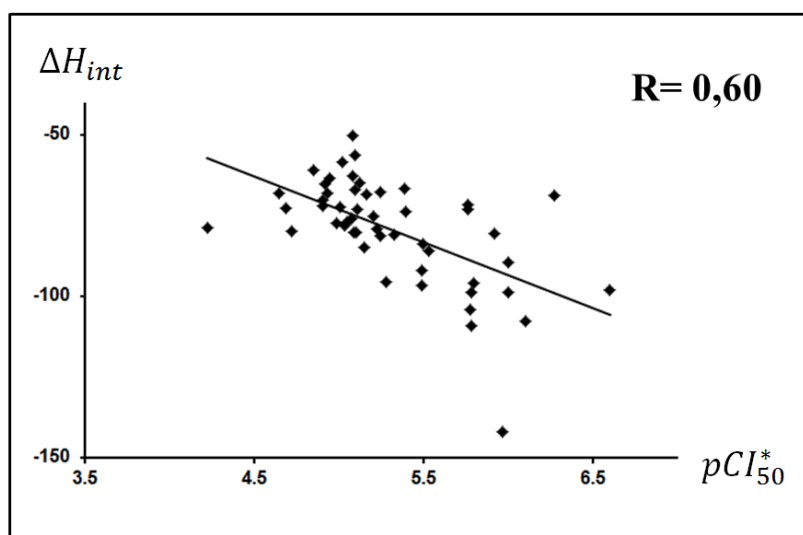


Figura 23. Correlação entre ΔH_{int} e pCI_{50}^*

Os dados teóricos (E_{solv} , ΔH_{int} e N_{LRC}) necessários para a análise de regressão múltipla linear (equação **22**), são apresentados na Tabela 13:

Tabela 13. Energia de solvatação (E_{solv}) em kcal mol⁻¹, Entalpia de interação (ΔH_{int}) em kcal mol⁻¹ e número de ligações rotacionáveis congeladas (N_{LRC}) para os compostos da série 1.

Composto	E_{solv}	ΔH_{int}	N_{LRC}	Composto	E_{solv}	ΔH_{int}	N_{LRC}
1	-75,39	-98,23	3	27	-18,95	-67,86	5
2	-73,91	-68,86	3	28	-19,46	-75,20	6
3	-63,95	-89,68	6	29	-18,18	-68,35	5
4	-62,47	-98,77	8	30	-18,89	-84,94	5
5	-68,69	-142,20	6	31	-18,70	-64,81	5
6	-65,80	-80,79	7	32	-18,24	-73,11	6
7	-62,93	-96,17	7	33	-19,60	-80,20	6
8	-63,27	-98,72	7	34	-20,12	-56,28	6
9	-66,27	-109,15	7	35	-21,14	-67,03	6
10	-64,50	-104,08	8	36	-18,59	-80,38	5
11	-68,70	-73,04	4	37	-21,91	-62,93	6
12	-62,77	-71,55	8	38	-18,84	-50,24	5
13	-159,59	-107,95	7	39	-18,26	-76,04	4
14	-59,20	-86,10	6	40	-17,59	-77,13	6
15	-61,41	-91,99	8	41	-18,07	-78,05	10
16	-62,15	-96,66	7	42	-19,55	-58,41	6
17	-58,15	-74,03	7	43	-22,44	-72,56	7
18	-61,04	-81,15	7	44	-18,87	-77,37	5
19	-64,11	-95,78	7	45	-21,34	-68,02	6
20	-60,59	-79,11	7	46	-19,57	-65,33	6
21	-63,32	-63,46	7	47	-20,71	-70,39	6
22	-74,68	-71,94	7	48	-19,27	-61,00	7
23	-59,92	-74,30	7	49	-21,41	-79,76	6
24	-19,78	-83,83	6	50	-17,52	-72,69	6
25	-20,22	-66,61	7	51	-19,08	-68,26	6
26	-21,13	-81,27	6	52	-21,96	-78,99	9

A análise por regressão múltipla linear demonstrou que existe uma baixa correlação ($R=0,41$ e erro padrão= $0,35$) entre os valores experimentais de pCI_{50} (Tabela 12) e os dados teóricos (Tabela 13) para os compostos da série 1. Esta baixa correlação novamente evidencia a necessidade da inclusão do efeito eletrolítico, uma vez que a série 1 engloba compostos neutros e carregados. Utilizando-se os valores de pCI_{50}^* ao invés dos valores de pCI_{50} na regressão múltipla linear com os mesmos dados teóricos, houve uma considerável melhora na correlação ($R=0,70$ e erro padrão= $0,35$). O valor F de significação foi de $2,91 \times 10^{-7}$, o que demonstra que o modelo tem capacidade preditiva significativa, pois está abaixo de $0,05$. O valor-p para cada variável do modelo foi: $0,019$ para a ΔH_{int} , $0,0002$ para a E_{solv} e $0,20$ para o N_{LRC} . Como o valor-p para a variável N_{LRC} se encontra acima de $0,05$ pode-se dizer que essa variável não possui capacidade preditiva no modelo. De fato, a exclusão da contribuição desta variável na regressão múltipla linear não afetou o coeficiente de correlação (R) que permaneceu $0,70$.

Conforme anteriormente mencionado, a variável N_{LRC} representa o número de ligações rotacionáveis que ficam “congeladas” após o encaixe do ligante no sítio da Shp2. Na contagem destas ligações assumiu-se que todas ficam congeladas após o encaixe no interior do sítio ativo, porém é possível que algumas se posicionem para fora do sítio, mantendo rotação parcial ou total. Desta forma, foi realizada uma análise cuidadosa de cada complexo sítio-inibidor para verificar o número de ligações rotacionáveis que de fato estivessem congeladas. Nesta análise observou-se que dentre os ligantes da série 1 apenas os compostos 1 e 2 possuíam o grupamento R_2 direcionados para fora do sítio (Figura 24). Desta forma, as ligações rotacionáveis presentes em R_2 foram desconsideradas na obtenção do valor de N_{LRC} dos compostos 1 e 2, resultando em uma perda entrópica menor do que a anteriormente considerada. Com isso, na análise de regressão múltipla linear houve uma melhora na correlação ($R=0,76$ e erro padrão= $0,32$) entre os valores de pCI_{50}^* e os dados teóricos (ΔH_{int} , N_{LRC} , E_{solv}). O valor F de significação do modelo foi de $3,24 \times 10^{-9}$. O valor-p para cada variável do modelo foi: $0,004$ para a ΔH_{int} , $6,43 \times 10^{-5}$ para a E_{solv} e $0,001$ para o N_{LRC} . A melhora na capacidade preditiva do modelo demonstrou que a variável N_{LRC} também é importante, apesar de requerer uma análise visual mais cuidadosa do complexo sítio/ligante.

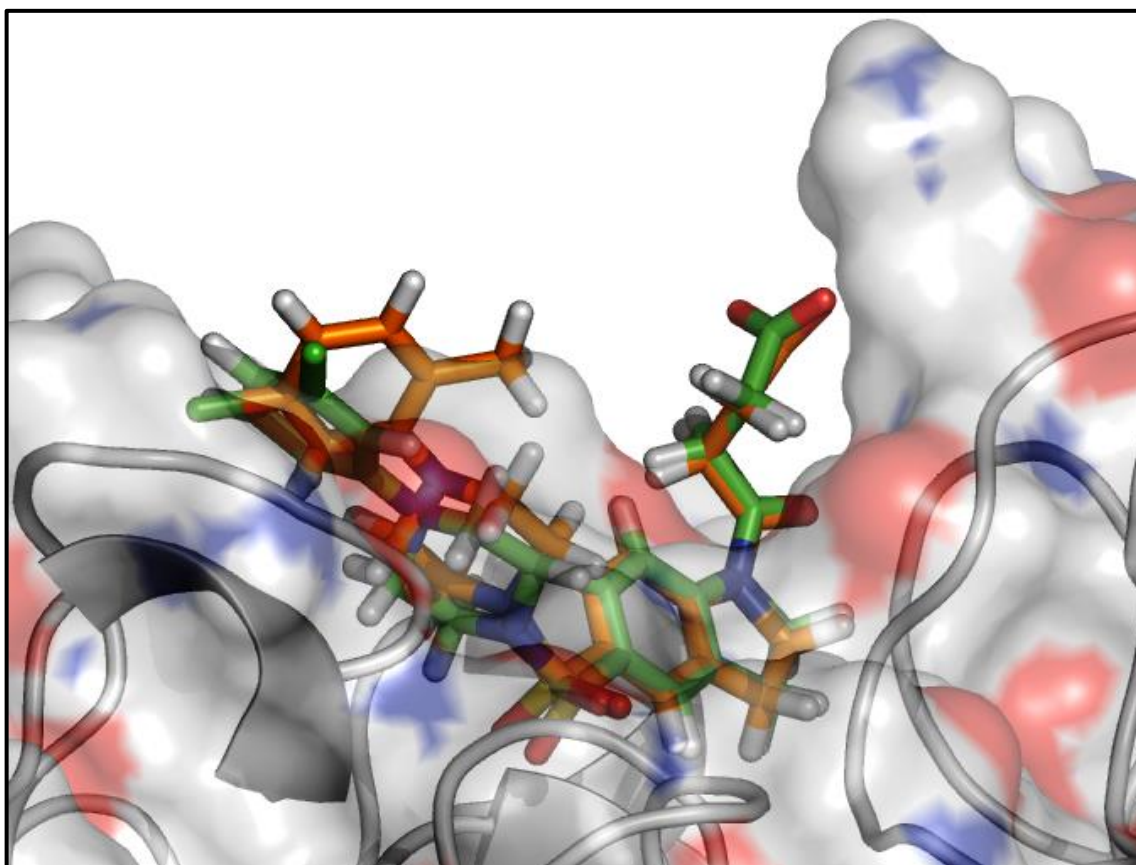


Figura 24. Composto **1** (em tubos, carbonos em verde) e composto **2** (em tubos, carbonos em laranja) interagindo com o sítio da Shp2 (4RDD). Observa-se que o grupamento R2 se localiza para fora do sítio ativo. Figura gerada com o programa PyMOL v0.99 (DeLano Scientific LLC)

O composto **23** foi identificado como um “*outlier*” e, de fato, dentre os compostos da série 1, é o composto com maior desvio médio experimental de CI_{50} ($143,6 \pm 89,5 \mu\text{mol/L}$). Após descartar a contribuição do composto **23**, uma melhor correlação ($R=0,80$ e erro padrão= $0,28$) foi obtida.

Uma dependência quadrática entre ΔG_I (energia livre associada com o processo de transferência do inibidor da fase aquosa para uma fase mais lipofílica) e a atividade inibitória foi proposta por Wang e colaboradores (1994) sob a justificativa de que, como a enzima analisada por eles era uma enzima de membrana, valores extremos de ΔG_I prejudicariam a atividade. Valores muito negativos para esse termo poderiam fazer com que o inibidor interagisse muito bem com a membrana ao invés de interagir com a enzima e valores muito positivos para este termo indicariam que a passagem do inibidor do meio aquoso para o interior da enzima não seria muito favorecida.

Apesar da Shp2 não ser uma enzima de membrana, é esperado uma melhor correlação ao se incorporar o termo quadrático do ΔG_I à equação, pois valores muito negativos para esse termo poderiam fazer com que o inibidor interagisse muito bem com regiões mais apolares da enzima, impedindo-o de alcançar a superfície do sítio que é bastante carregada, enquanto valores muito positivos para este termo indicariam que a passagem do inibidor do meio aquoso para o interior do sítio não seria muito favorecida. Dessa forma, a proposta feita por Wang e colaboradores (1994) também poderia ser aplicada à Shp2. O termo ΔG_I é representado, no modelo, pela energia de solvatação (E_{solv}). A inclusão do termo quadrático E_{solv}^2 na equação **22** resultou na nova equação **23**, que oferece melhor correlação ($R=0,83$ e erro padrão=0,27). Os coeficientes calibrados com os dados experimentais disponíveis são:

$$pCI_{50}^* = -6,2 \times 10^{-5} (E_{solv} + 141,45)^2 - 0,0078 \Delta H_{int} - 0,11649 N_{LRC} + 6,0895 \quad (23)$$

Este modelo possui boa capacidade preditiva, justificada através do F de significação ($5,45 \times 10^{-11}$) e o valor-p para cada variável (0,01 para a ΔH_{int} ; $3,67 \times 10^{-5}$ para a E_{solv} ; 0,0003 para o N_{LRC} e 0,02 para E_{solv}^2), todos abaixo de 0,05.

A aplicação da equação **23** na predição da atividade dos compostos da série 1 reproduziu bem os valores de pCI_{50}^* ($R=0,83$), com exceção do composto **23**, que foi eliminado (Tabela 14 e Figura 25)

Tabela 14. Valores experimentais e preditos de pCI_{50}^* e a diferença entre eles (ΔpCI_{50}^*), para a série 1.

Composto	pCI_{50}^* Exp.	pCI_{50}^* Predito	ΔpCI_{50}^*	Composto	pCI_{50}^* Exp.	pCI_{50}^* Predito	ΔpCI_{50}^*
1	6,60	6,24	0,36	27	5,24	5,11	0,14
2	6,27	5,99	0,27	28	5,20	5,05	0,15
3	6,00	5,72	0,28	29	5,16	5,10	0,06
4	6,00	5,54	0,46	30	5,15	5,24	-0,09
5	5,97	6,17	-0,20	31	5,12	5,08	0,04
6	5,91	5,55	0,36	32	5,11	5,02	0,09
7	5,79	5,64	0,15	33	5,10	5,10	0,00
8	5,78	5,67	0,12	34	5,09	4,92	0,17
9	5,78	5,77	0,01	35	5,09	5,02	0,08

Tabela 14 (continuação). Valores experimentais e preditos de pCI_{50}^* e a diferença entre eles (ΔpCI_{50}^*), para a série 1.

Composto	pCI_{50}^* Exp.	pCI_{50}^* Predito	ΔpCI_{50}^*	Composto	pCI_{50}^* Exp.	pCI_{50}^* Predito	ΔpCI_{50}^*
10	5,77	5,60	0,17	36	5,09	5,20	-0,11
11	5,76	5,87	-0,10	37	5,08	5,00	0,09
12	5,76	5,33	0,43	38	5,08	4,97	0,11
13	6,10	6,10	0,01	39	5,07	5,28	-0,21
14	5,52	5,64	-0,12	40	5,05	5,04	0,00
15	5,49	5,48	0,01	41	5,03	4,59	0,44
16	5,48	5,64	-0,15	42	5,02	4,92	0,09
17	5,39	5,42	-0,03	43	5,00	4,96	0,04
18	5,32	5,51	-0,18	44	4,98	5,18	-0,20
19	5,27	5,65	-0,38	45	4,93	5,03	-0,09
20	5,22	5,49	-0,26	46	4,91	4,98	-0,07
21	4,94	5,39	-0,45	47	4,90	5,04	-0,14
22	4,90	5,56	-0,65	48	4,84	4,82	0,02
23*	4,29	5,44	-1,15	49	4,72	5,12	-0,40
24	5,49	5,13	0,37	50	4,68	5,01	-0,32
25	5,39	4,88	0,50	51	4,64	4,99	-0,35
26	5,24	5,13	0,12	52	4,22	4,77	-0,55

*Composto “outlier”

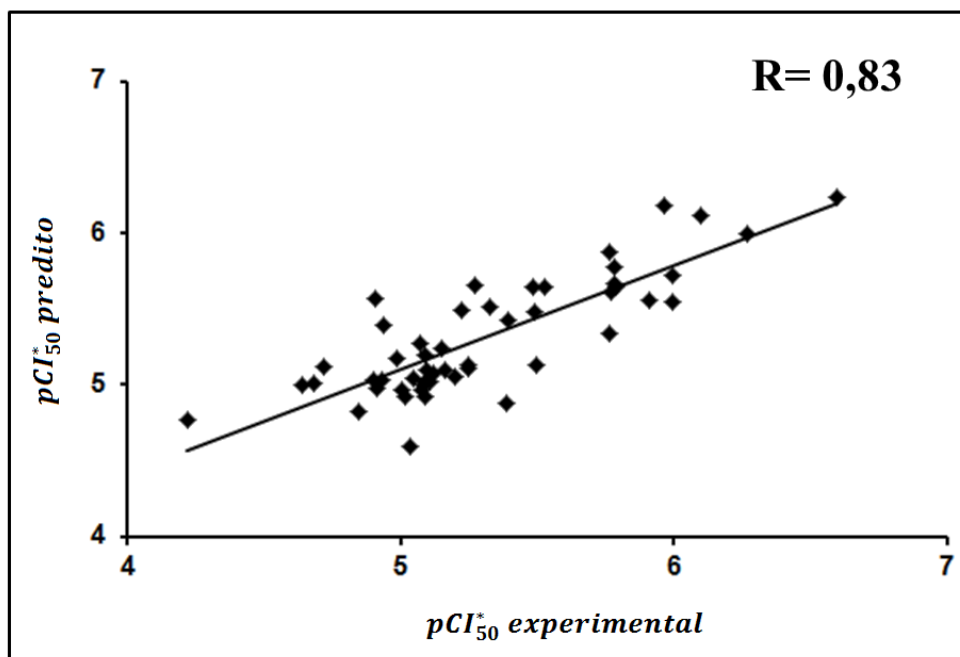


Figura 25. Correlação entre os valores experimentais de pCI_{50}^* e os valores preditos para os compostos da série 1 (equação 23)

A equação 23 demonstra que o valor de E_{solv} ideal para a obtenção de compostos mais ativos na inibição da enzima isolada é $-141,45 \text{ kcal mol}^{-1}$. Porém, pensando nos critérios mais amplos da Química Medicinal, é provável que compostos tão polares assim não apresentem boa atividade *in vivo* devido à dificuldade que enfrentarão na passagem em membranas celulares.

Embora todos os termos da equação sejam significativos para a determinação da atividade inibitória da série 1, a análise da contribuição de cada termo para a atividade demonstrou que o termo E_{solv} teve uma maior contribuição em 54% dos compostos, o termo N_{LRC} teve uma maior contribuição em 33% dos compostos e o termo ΔH_{int} teve uma maior contribuição 13% dos compostos (Tabela 15).

Tabela 15. Contribuição de cada termo da equação **23** para os compostos da série 1

Comp.	$-6,2 \times 10^{-5}(E_{solv} + 141,45)^2$	$-0,0078\Delta H_{int}$	$-0,11649N_{LRC}$
1	-0,27	0,77	-0,35
2	-0,28	0,54	-0,35
3	-0,37	0,70	-0,70
4	-0,39	0,77	-0,93
5	-0,33	1,11	-0,70
6	-0,35	0,63	-0,82
7	-0,38	0,75	-0,82
8	-0,38	0,77	-0,82
9	-0,35	0,85	-0,82
10	-0,37	0,81	-0,93
11	-0,33	0,57	-0,47
12	-0,38	0,56	-0,93
13	-0,02	0,84	-0,82
14	-0,42	0,67	-0,70
15	-0,40	0,72	-0,93
16	-0,39	0,75	-0,82
17	-0,43	0,58	-0,82
18	-0,40	0,63	-0,82
19	-0,37	0,75	-0,82
20	-0,41	0,62	-0,82
21	-0,38	0,49	-0,82
22	-0,28	0,56	-0,82
23	-0,41	0,58	-0,82
24	-0,92	0,65	-0,70
25	-0,91	0,52	-0,82
26	-0,90	0,63	-0,70
27	-0,93	0,53	-0,58
28	-0,92	0,59	-0,70
29	-0,94	0,53	-0,58
30	-0,93	0,66	-0,58
31	-0,93	0,51	-0,58
32	-0,94	0,57	-0,70
33	-0,92	0,63	-0,70

Tabela 15 (continuação). Contribuição de cada termo da equação **23** para os compostos da série 1

Comp.	$-6,2 \times 10^{-5}(E_{solv} + 141,45)^2$	$-0,0078\Delta H_{int}$	$-0,11649N_{LRC}$
34	-0,91	0,44	-0,70
35	-0,90	0,52	-0,70
36	-0,94	0,63	-0,58
37	-0,89	0,49	-0,70
38	-0,93	0,39	-0,58
39	-0,94	0,59	-0,47
40	-0,95	0,60	-0,70
41	-0,94	0,61	-1,16
42	-0,92	0,46	-0,70
43	-0,88	0,57	-0,82
44	-0,93	0,60	-0,58
45	-0,89	0,53	-0,70
46	-0,92	0,51	-0,70
47	-0,90	0,55	-0,70
48	-0,93	0,48	-0,82
49	-0,89	0,62	-0,70
50	-0,95	0,57	-0,70
51	-0,93	0,53	-0,70
52	-0,89	0,62	-1,05

5.4 Validação do modelo de predição de atividade para inibidores da Shp2

Obtida a equação **23** para predição das atividades de inibidores da Shp2, calibrada com os dados dos compostos da série 1, a etapa seguinte foi sua validação externa por meio da predição da atividade inibitória de compostos cujos dados não foram usados na sua construção, os compostos da série 2. Esta série contém 24 derivados do oxindol, dentre eles, dois compostos possuem carga -2, dezoito possuem carga -1 e três compostos não possuem carga. Na Tabela 16 estão listados os dados utilizados na predição.

Tabela 16. Valores experimentais de CI_{50} , razão dos coeficientes de atividade (γ^*), valores de pCI_{50}^* e dados teóricos necessários para a predição da atividade inibitória dos compostos da série 2.

Composto	CI_{50} ($\mu\text{mol L}^{-1}$)	γ^*	pCI_{50}^* Exp.	ΔH_{int} (kcal mol^{-1})	N_{LRC}	E_{solv} (kcal mol^{-1})
53	49,6	1	4,30	-47,4	6	-23,36
54	7,94	0,36	5,54	-82,1	6	-89,34
55*	46,8	0,36	4,77	-92,1	3	-77,36
56	0,8	0,13	6,98	-97,0	4	-172,47
57	15,8	0,13	5,69	-89,5	5	-171,79
58	22,3	0,36	5,10	-73,2	7	-77,22
59	4,5	0,36	5,79	-76,1	7	-84,74
60	4,5	0,36	5,79	-67,4	5	-85,52
61	11,9	1	4,92	-56,2	5	-28,07
62	4,4	1	5,36	-56,8	6	-25,32
63	5,0	0,36	5,74	-77,0	7	-86,90
64	1,3	0,36	6,33	-72,4	7	-84,62
65	3,8	0,36	5,86	-84,5	8	-86,00
66	1,4	0,36	6,30	-85,5	7	-87,33
67	5,9	0,36	5,67	-74,1	7	-88,65
68	7,4	0,36	5,57	-81,2	8	-84,47
69	5,5	0,36	5,70	-60,3	8	-85,82
70	22,9	0,36	5,08	-91,9	8	-69,68
71	1,0	0,36	6,44	-91,2	7	-85,07
72	6,3	0,36	5,64	-83,1	7	-86,05
73	4,8	0,36	5,76	-75,9	7	-85,42
74	5,2	0,36	5,73	-111,6	7	-86,62
75	8,3	0,36	5,52	-56,0	8	-84,24
76	10,6	0,36	5,42	-100,7	8	-82,93

Houve boa correlação ($R=0,73$) entre os valores de pCI_{50}^* experimentais e os valores de pCI_{50}^* preditos pela equação **23**, após o descarte da contribuição do composto **55** (Tabela 17) e Figura 26).

Tabela 17. Valores experimentais e preditos (equação 23) de pCI_{50}^* e a diferença entre eles (ΔpCI_{50}^*), para a série 2.

Composto	pCI_{50}^* Exp.	pCI_{50}^* Predito	ΔpCI_{50}^*	Composto	pCI_{50}^* Exp.	pCI_{50}^* Predito	ΔpCI_{50}^*
53	4,30	4,90	-0,59	65	5,86	5,63	0,24
54	5,54	5,86	-0,32	66	6,30	5,76	0,54
55*	4,77	6,20	-1,43	67	5,67	5,68	-0,01
56	6,98	6,32	0,66	68	5,57	5,59	-0,02
57	5,69	6,15	-0,46	69	5,70	5,44	0,27
58	5,10	5,59	-0,49	70	5,08	5,56	-0,47
59	5,79	5,67	0,12	71	6,44	5,79	0,66
60	5,79	5,84	-0,05	72	5,64	5,73	-0,09
61	4,92	5,15	-0,22	73	5,76	5,67	0,09
62	5,36	5,00	0,36	74	5,52	5,73	0,00
63	5,74	5,69	0,05	75	5,42	5,39	0,13
64	6,33	5,64	0,69	76	4,30	5,73	-0,31

*Composto “outlier”

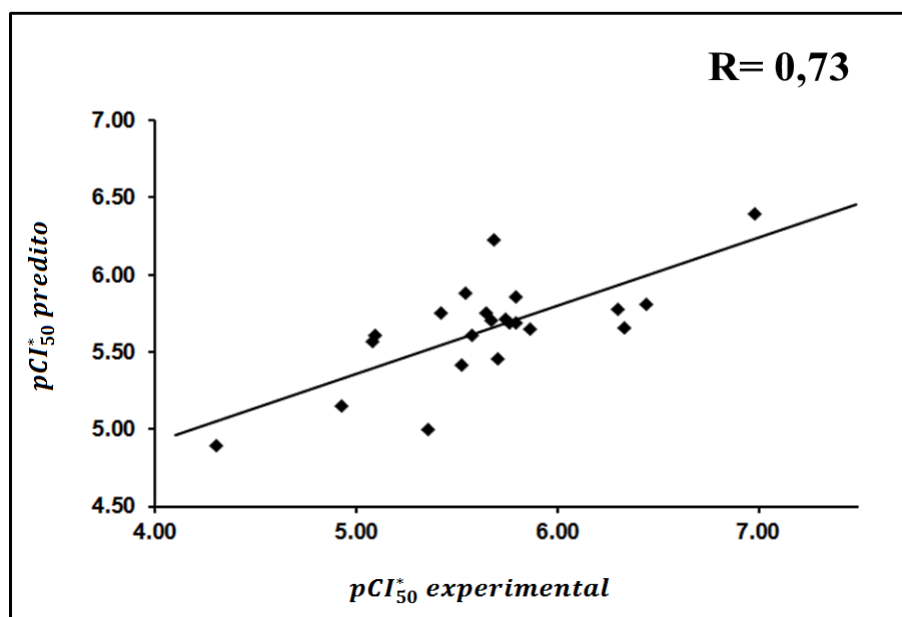


Figura 26. Correlação entre os valores experimentais de pCI_{50}^* e os valores preditos (equação 23) para os compostos da série 2.

A análise da contribuição de cada termo para a atividade inibitória da série 2 demonstrou que o termo de E_{solv} teve uma maior contribuição em 12,5% dos compostos, o termo N_{LRC} teve uma maior contribuição em 75% dos compostos e o termo ΔH_{int} teve uma maior contribuição 12,5% dos compostos (Tabela 18).

Tabela 18. Contribuição de cada termo da equação 23 para os compostos da série 2

Composto	$-6,2 \times 10^{-5}(E_{solv} + 141,45)^2$	$-0,0078\Delta H_{int}$	$-0,11649N_{LRC}$
53	-0,88	0,37	-0,69
54	-0,18	0,64	-0,69
55*	-0,27	0,72	-0,35
56	-0,05	0,76	-0,46
57	-0,05	0,70	-0,58
58	-0,27	0,57	-0,81
59	-0,21	0,59	-0,81
60	-0,21	0,53	-0,58
61	-0,82	0,44	-0,58
62	-0,85	0,44	-0,69
63	-0,20	0,60	-0,81
64	-0,21	0,56	-0,81
65	-0,20	0,66	-0,93
66	-0,19	0,67	-0,81
67	-0,18	0,58	-0,81
68	-0,21	0,63	-0,93
69	-0,20	0,47	-0,93
70	-0,33	0,72	-0,93
71	-0,21	0,71	-0,81
72	-0,20	0,65	-0,81
73	-0,21	0,59	-0,81
74	-0,21	0,44	-0,93
75	-0,22	0,79	-0,93
76	-0,88	0,37	-0,69

5.5 Predição da atividade inibitória dos compostos *N*-arilpiperazínicos

O modelo desenvolvido (equação **23**) foi utilizado para prever a atividade inibitória (CI_{50}) dos novos compostos *N*-arilpiperazínicos (série 3, Figura 17 e Tabela 2) sintetizadas pelo Dr. Ciro Gonçalves e Sá, durante seu doutorado no Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio®) da Universidade Federal do Rio de Janeiro (UFRJ) [SÁ, 2017]. A série 3 foi escolhida devido à sua semelhança estrutural com os compostos da série 1. O perfil de interação destes compostos (Figura 27) também se assemelhou com o perfil de interação do ligante cocristalizado na Shp2 (4RDD). Conforme mostrado anteriormente na Figura 18, os compostos da série 1 apresentam um perfil de interação semelhante ao derivado da cefsulodina cocristalizado na Shp2 (4RDD). Da mesma forma, nos compostos da série 3, por serem estruturalmente parecidos com os compostos da série 1, também se observou um perfil de interação semelhante. Como pode ser observado na Figura 27, tanto os compostos da série 1 quanto os compostos da série 3 possuem um anel aromático interagindo via empilhamento π com o resíduo de aminoácido Tyr279 da alça pTyr. Além disso, ambos também apresentam um grupamento sulfona interagindo via ligação de hidrogênio com a cadeia principal de resíduos de aminoácidos localizados na alça P.

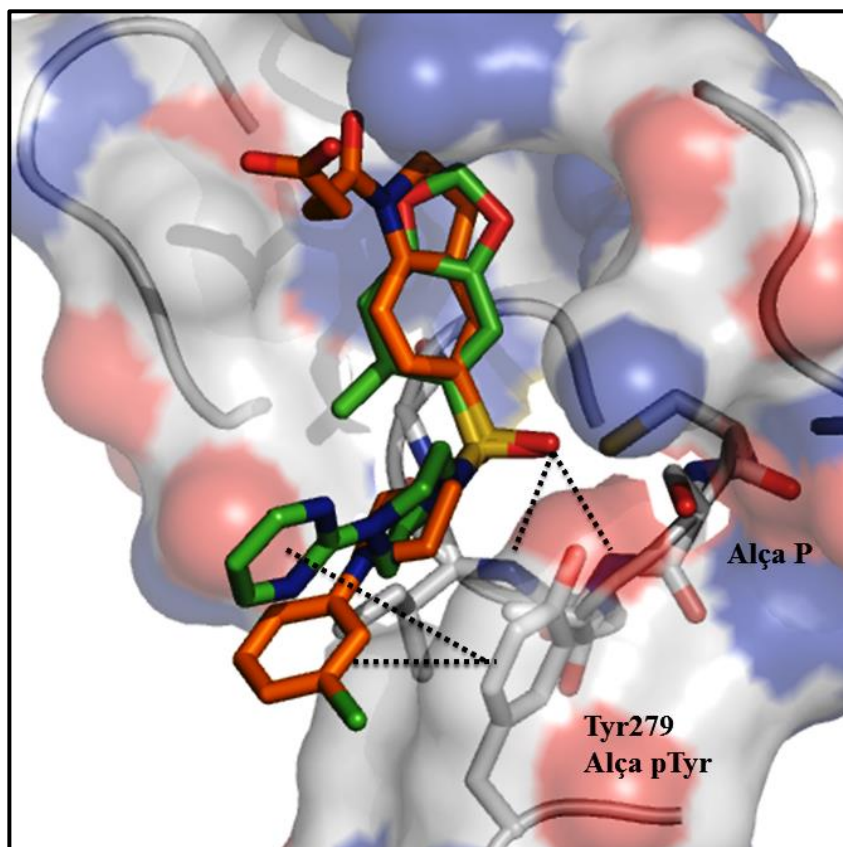


Figura 27. Perfil de interação do composto **77** da série 3 (carbonos em verde) com o sítio ativo da Shp2 (4RDD). Sobreposição com o composto **1** da série 1 (carbonos em laranja).

Na Tabela 19 estão listados os dados teóricos que foram calculados para esta série, os valores de pCI_{50}^* preditos através da equação **23** e os valores de CI_{50} calculados.

Tabela 19. Dados teóricos necessários para a predição da atividade inibitória, pCI_{50}^* e CI_{50} ($\mu\text{mol L}^{-1}$) preditos para a série 3.

Composto	E_{solv}	ΔH_{int}	N_{LRC}	pCI_{50}^* predito	CI_{50} Predito
77	-19,88	-72,00	3	5,39	4,12
78	-18,91	-84,38	4	5,35	4,46
79	-78,02	-107,09	6	5,98	2,93
80	-21,30	-57,31	7	4,83	14,93
81	-21,11	-71,71	8	4,82	15,17
82	-17,07	-75,45	3	5,37	4,27
83	-16,96	-89,43	5	5,24	5,71

Tabela 19 (continuação). Dados teóricos necessários para a predição da atividade inibitória, pCI_{50}^* e CI_{50} ($\mu\text{mol L}^{-1}$) preditos para a série 3.

Composto	E_{solv}	ΔH_{int}	N_{LRC}	pCI_{50}^* predito	CI_{50} Predito
84	-62,21	-108,84	6	5,85	3,92
85	-19,66	-72,19	7	4,92	12,09
86	-19,56	-80,85	8	4,87	13,58
87	-16,89	-85,90	3	5,45	3,56
88	-16,88	-81,51	5	5,18	6,60
89	-77,11	-63,75	6	5,63	6,49
90	-19,94	-71,92	7	4,92	12,03
91	-19,62	-72,88	8	4,81	15,64
92	-19,51	-79,30	4	5,32	4,78
93	-19,37	-64,59	6	4,97	10,71
94	-76,43	-105,12	7	5,83	4,09
95	-21,29	-84,83	8	4,92	11,91
96	-20,46	-57,90	9	4,59	25,99

Conforme pode ser observado na Tabela 19, os derivados da série 3 que possuem o grupamento $\text{CH}_2\text{CH}_2\text{COOH}$ (**79**, **84** e **94**) são os candidatos a inibidores da Shp2 mais promissores, pois possuem CI_{50} preditos abaixo de $5 \mu\text{mol L}^{-1}$. Os derivados contendo o grupamento CH_3 (**77**, **82**, **87** e **92**) também se mostraram promissores. Por outro lado, os derivados contendo o grupamento $\text{CH}_2\text{CH}_2\text{COOCH}_2\text{CH}_3$ e $\text{CH}_2\text{CH}_2\text{COOCH}_2\text{CH}_3$ (**81**, **91** e **96**) não foram tão promissores, pois os valores de CI_{50} preditos ficaram acima de $10 \mu\text{mol L}^{-1}$. Em relação aos grupamentos aromáticos, observou-se que a presença da 2-metoxifenila em alguns derivados (**93** e **96**) resultou em valores CI_{50} preditos mais desfavoráveis quando comparados aos outros grupamentos aromáticos.

Com o objetivo de contribuir com a síntese de *N*-arilpiperazínicos mais ativos na inibição da Shp2, foram propostas duas novas estruturas. O grupamento aromático 2,3-diclorofenila da estrutura **84** foi substituído pelo grupamento 3-clorofenila, pois este grupamento está presente no composto **1**, que é o mais ativo da série 1, resultando assim na estrutura proposta **97**. Além disso, o grupamento pirimidin-2-il da estrutura **77** foi

substituído pelo grupamento 4-carboxifenila para aumentar as interações com a alça Tyr, resultando na estrutura proposta **98** (Figura 28).

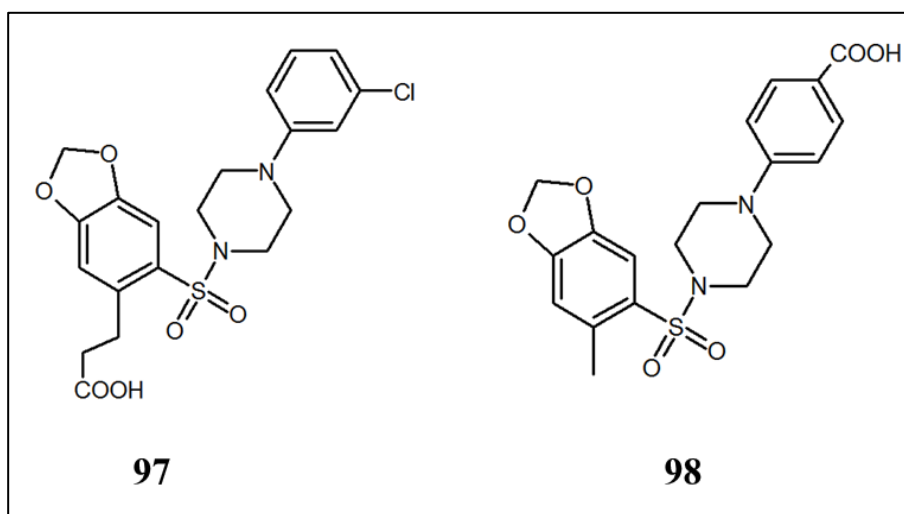


Figura 28. Estruturas propostas para inibidores da Shp2

Na Tabela 20 estão listados os dados teóricos e os valores de CI_{50} preditos através da equação **23** para as estruturas propostas.

Tabela 20. Dados teóricos necessários para a predição da atividade inibitória, pCI_{50}^* e CI_{50} preditos para os compostos propostos.

Composto	E_{solv}	ΔH_{int}	N_{LRC}	pCI_{50}^*	CI_{50} ($\mu\text{mol L}^{-1}$)
97	-75,55	-110,92	6	5,99	2,87
98	-83,84	-82,51	3	6,18	1,84

Conforme pode ser observado na Tabela 20 as estruturas propostas são promissoras na inibição da Shp2.

Entretanto, na pesquisa em Química Medicinal outros aspectos precisam ser considerados quando se trata do planejamento de compostos, já que um bom inibidor enzimático precisa atender outros critérios para ser um bom fármaco. Segundo a literatura, o composto **71** (série 2), mesmo apresentando uma potente atividade inibitória *in vitro* ($CI_{50} = 1,0 \mu\text{M}$), possui pouca atividade *in vivo* [CHEN *et al.*, 2010]. Chen e colaboradores substituíram o grupamento ácido carboxílico do composto **71** por um éster metílico e este novo composto não apresentou atividade *in vitro*, porém,

apresentou atividade *in vivo*, indicando que o novo composto consegue permear a membrana celular e dentro da célula é metabolizado e se transforma no composto **71** [CHEN *et al.*, 2010].

A energia de solvatação calculada para o composto **71** foi de $-85,07 \text{ kcal mol}^{-1}$, valor bem próximo dos valores de energia de solvatação dos compostos propostos (estruturas **97** e **98**) e dos compostos da série 3 que possuem o grupamento ácido carboxílico. Portanto, espera-se que estes também não apresentem boa atividade *in vivo*, por não conseguirem permear a membrana. Desta forma, a substituição do grupamento ácido carboxílico por um éster metílico é uma estratégia que também pode ser experimentada nos novos compostos *N*-arilpiperazínicos (série 3), a fim de se melhorar sua atividade *in vivo*.

6 CONCLUSÃO

O procedimento adotado neste trabalho, que combinou a docagem de 52 compostos (série 1), conhecidos por inibir seletivamente Shp2, com o refinamento das geometrias pelo método do orbital molecular semi-empírico PM7, revelou entalpias de interação significativamente mais favoráveis para Shp2 do que Shp1 ($\alpha = 0,01$), indicando que este fator desempenha um papel fundamental para a seletividade, ao menos para esta série. Para a série 2, composta por 24 derivados do oxindol com informações sobre a atividade sobre ambas as enzimas, houve uma correlação satisfatória ($R = 0,70$) entre a diferença percentual relativa da entalpia de interação ($\% \Delta \Delta H_{\text{int}}$) e a seletividade relativa (SR). Desta forma, este procedimento se mostrou confiável para ser aplicado no planejamento de inibidores mais seletivos para a Shp2.

Muitos trabalhos envolvendo o planejamento e a síntese de novos inibidores da Shp2 já utilizaram a técnica de triagem virtual para selecionar os compostos mais promissores e a técnica de docagem molecular para investigar as possíveis interações entre o inibidor e o sítio ativo da Shp2 [LIU *et al.*, 2013; ZHANG *et al.*, 2010]. Porém, antes deste, nenhum trabalho empregou modelagem molecular como uma ferramenta para a previsão da seletividade dos inibidores de Shp2.

A segunda parte deste trabalho se baseou no trabalho pioneiro de Wang e colaboradores (1994) e consistiu no desenvolvimento de um modelo empírico para a determinação da atividade inibitória baseado em um ciclo termodinâmico do processo de interação enzima/ligante, ajustado através da regressão múltipla linear entre os termos entálpicos e entrópicos calculados e os valores experimentais atividade inibitória.

Os valores de atividade inibitória preditos pelo modelo desenvolvido neste trabalho (equação **23**) apresentou uma boa correlação com os dados experimentais ($R=0,83$). O modelo foi construído com uma base empírica de 51 compostos (22 compostos com carga e 29 compostos sem carga) e apesar de não ter sido construído utilizando os dados dos termos calculados de todos os inibidores da série 1, devido a exclusão de um composto que possuía um alto desvio médio experimental de CI_{50} , ele ainda envolveu a participação de compostos com uma faixa razoável de valores experimentais de CI_{50} (0,7 a 60,0 μM). Além disso, o modelo se mostrou robusto, pois pôde ser validado de forma satisfatória ($R=0,73$) com uma série diferente de compostos (série 2), não empregados na construção do modelo.

Por se tratar de um modelo termodinamicamente consistente, demonstrou-se que a análise de regressão múltipla linear entre as variáveis independentes (E_{solv} , ΔH_{int} , N_{LRC}) e a variável dependente (pCI_{50}) apenas foi capaz de gerar boa correlação quando se considerou o efeito eletrolítico sobre os valores de pCI_{50} . Isto porque a “atmosfera” iônica da solução provoca um desvio da idealidade principalmente nos compostos contendo carga, ou seja, suas concentrações efetivas se distanciam de suas concentrações formais. A superfície eletrostática do sítio ativo da Shp2 é bastante positiva, o que torna complexo o cálculo exato do desvio da idealidade ocorrido sobre estes inibidores (principalmente os que contêm cargas). Porém, a estratégia abordada neste trabalho para o cálculo da razão dos coeficientes de atividades (γ^*) se mostrou satisfatória na obtenção de um modelo (equação **23**) de boa capacidade preditiva. Wang e colaboradores (1994), em seu trabalho, não realizaram nenhuma correção sobre os valores experimentais de atividade (K_i) que utilizaram, porém o modelo preditivo construído por estes autores baseou-se numa série de 18 compostos sem carga e por isso, o efeito eletrolítico pôde ser desprezado [WANG *et al.*, 1994].

Por fim, os valores de atividade inibitória (CI_{50}) predita através da equação **23** para os novos compostos *N*-arilpiperazínicos (série 3) e para os compostos propostos (**97** e **98**) neste trabalho, indicam que esta série é promissora na inibição da Shp2.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHAM, M. J., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J. C., HESS, B., LINDAHL, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1, p. 19-25, 2015.

ALCÁCER, Luís. Introdução à química quântica computacional. *Energia*, v. 265, n. 268, p. 273, 2007.

ALICEA-VELAZQUEZ, L. N., BOGGON, J. T.. SHP family protein tyrosine phosphatases adopt canonical active-site conformations in the apo and phosphate-bound states. **Protein and Peptide Letters**, v. 20, n. 9, p. 1039-1048, 2013.

ALONSO, A., SASIN, J., BOTTINI, N., FRIEDBERG, I., FRIEDBERG, I., OSTERMAN, A., MUSTELIN, T. Protein tyrosine phosphatases in the human genome. **Cell**, v. 117, n. 6, p. 699-711, 2004.

ALLINGER, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. **Journal of the American Chemical Society**, v. 99, n. 25, p. 8127-8134, 1977.

ANDERSEN, J. N., MORTENSEN, O. H., PETERS, G. H., DRAKE, P. G., IVERSEN, L. F., OLSEN, O. H., MØLLER, N. P. H. Structural and evolutionary relationships among protein tyrosine phosphatase domains. **Molecular and Cellular Biology**, v. 21, n. 21, p. 7117-7136, 2001.

ASI, A. M., RAHMAN, N. A., MERICAN, A. F. Application of the linear interaction energy method (LIE) to estimate the binding free energy values of Escherichia coli wild-type and mutant arginine repressor C-terminal domain (ArgRc)-l-arginine and ArgRc-l-citrulline protein-ligand complexes. **Journal of Molecular Graphics and Modelling**, v. 22, n. 4, p. 249-262, 2004.

BARD-CHAPEAU, E. A., LI, S., DING, J., ZHANG, S. S., ZHU, H. H., PRINCEN, F., VARKI, N. M. Ptpn11/Shp2 acts as a tumor suppressor in hepatocellular carcinogenesis. **Cancer cell**, v. 19, n. 5, p. 629-639, 2011.

BARR, A. J., UGOCHUKWU, E., LEE, W. H., KING, O. N., FILIPPAKOPOULOS, P., ALFANO, I., KNAPP, S. Large-scale structural analysis of the classical human protein tyrosine phosphatome. **Cell**, v. 136, n. 2, p. 352-363, 2009.

BENTIRES-ALJ, M., PAEZ, J. G., DAVID, F. S., KEILHACK, H., HALMOS, B., NAOKI, K., RICHARDS, W. G. Activating mutations of the noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. **Cancer Research**, v. 64, n. 24, p. 8816-8820, 2004.

BONET, B., LOERINCS, G., GEFFNER, H. A robust and fast action selection mechanism for planning. In: **AAAI/IAAI**. 1997. p. 714-719.

BORN, M., OPPENHEIMER, R. ROBERT. Zur quantentheorie der molekeln. **Annalen der Physik**, v. 389, n. 20, p. 457-484, 1927.

BROOKS, C. L. A theoretical perspective of dynamics, structure, and thermodynamics. *Advances in chemical physics*; John Wiley: New York, v.LXXI. 1988.

CHAMBERS, C. C., HAWKINS, G. D., CRAMER, C. J., TRUHLAR, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First Solvation Shell Effects. **Journal of Physical Chemistry**. v. 100, p. 16385-16298, 1996.

CHAN, R. J., FENG, G. S. PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. **Blood**, v. 109, n. 3, p. 862-867, 2007.

CHEN, C. Y., WILLARD, D., RUDOLPH, J. Redox regulation of SH2-domain-containing protein tyrosine phosphatases by two backdoor cysteines. **Biochemistry**, v. 48, n. 6, p. 1399-1409, 2009.

CHEN, L., PERNAZZA, D., SCOTT, L. M., LAWRENCE, H. R., REN, Y., LUO, Y., LAWRENCE, N. J. Inhibition of cellular Shp2 activity by a methyl ester analog of SPI-112. **Biochemical pharmacology**, v. 80, n. 6, p. 801-810, 2010.

CHENG, Y.; PRUSOFF, W. H. Relationship between the inhibition constant (KI) and the concentration which causes 50% inhibition of an enzymatic reaction. **Biochemical Pharmacology**, v. 22, p. 403-411, 1973.

CHMIELA, S., TKATCHENKO, A., SAUCEDA, H. E., POLTAVSKY, I., SCHÜTT, K. T., MÜLLER, K. R. Machine learning of accurate energy-conserving molecular force fields. **Science Advances**, v. 3, n. 5, p. e1603015, 2017.

CLARK, Tim. A handbook of computational chemistry: A practical guide to chemical structure and energy calculations. Wiley-Interscience, 1985.

COSTA FILHO, Paulo Augusto da; POPPI, Ronei Jesus. Genetic algorithm in chemistry. **Química Nova**, v. 22, n. 3, p. 405-411, 1999.

CRAMER, C. J.; TRUHLAR, D. G. PM3-SM3: A general parameterization for including aqueous solvation effects in the PM3 molecular orbital model. **Journal of Computational Chemistry**, v. 13, n. 9, p. 1089-1097, 1992.

DANCE, M., MONTAGNER, A., SALLES, J. P., YART, A., RAYNAL, P. The molecular functions of Shp2 in the Ras/Mitogen-activated protein kinase (ERK1/2) pathway. **Cellular signalling**, v. 20, n. 3, p. 453-459, 2008.

DEWAR, M. J., ZOEBISCH, E. G., HEALY, E. F., STEWART, J. J. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. **Journal of the American Chemical Society**, v. 107, n. 13, p. 3902-3909, 1985.

- DEWAR, M. J. S; THIEL, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. **Journal of the American Chemical Society**, v. 99, n. 15, p. 4899-4907, 1977.
- DOBEŠ, P., FANFRLÍK, J., ŘEZÁČ, J., OTYEPKA, M., HOBZA, P. Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. **Journal of Computer-Aided Molecular Design**, v. 25, n. 3, p. 223-235, 2011.
- DONG, S., LI, F. Q., ZHANG, Q., LV, K. Z., YANG, H. L., GAO, Y., YU, J. R. Expression and clinical significance of SHP2 in gastric cancer. **Journal of International Medical Research**, v. 40, n. 6, p. 2083-2089, 2012.
- ELDRIDGE, M. D., MURRAY, C. W., AUTON, T. R., PAOLINI, G. V., MEE, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. **Journal of Computer-Aided Molecular Design**, v. 11, n. 5, p. 425-445, 1997.
- FAN, L. C., TENG, H. W., SHIAU, C. W., LIN, H., HUNG, M. H., CHEN, Y. L., CHEN, K. F. SHP-1 is a target of regorafenib in colorectal cancer. **Oncotarget**, v. 5, n. 15, p. 6243, 2014.
- FANFRLÍK, J., BRONOWSKA, A. K., ŘEZÁČ, J., PŘENOSIL, O., KONVALINKA, J., HOBZA, P. A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands. **The Journal of Physical Chemistry B**, v. 114, n. 39, p. 12666-12678, 2010.
- FICHTNER-FEIGL, S., KESSELRING, R., STROBER, W. Chronic inflammation and the development of malignancy in the GI tract. **Trends in Immunology**, v. 36, n. 8, p. 451-459, 2015.
- GILSON, Michael K.; HONIG, Barry H. The dielectric constant of a folded protein. **Biopolymers: Original Research on Biomolecules**, v. 25, n. 11, p. 2097-2119, 1986.
- GOHLKE, H., HENDLICH, M., KLEBE, G. Knowledge-based scoring function to predict protein-ligand interactions. **Journal of Molecular Biology**, v. 295, n. 2, p. 337-356, 2000.
- GÜNTEMBERG, E. Untersuchungen über Ioneninteraktion. **Zeitschrift für Physikalische Chemie**, v. 123, n. 1, p. 199-247, 1926.
- HAN, T., X., D. M., SUN, W., LIU, N., SUN, H. L., WEN, W., CHENG, Z. PTPN11/Shp2 overexpression enhances liver cancer progression and predicts poor prognosis of patients. **Journal of Hepatology**, v. 63, n. 3, p. 651-660, 2015.

- HANSSON, T., MARELIUS, J., ÅQVIST, J. Ligand binding affinity prediction by linear interaction energy methods. **Journal of Computer-Aided Molecular Design**, v. 12, n. 1, p. 27-35, 1998.
- HARTREE, D. R. On some approximate numerical applications of Bohr's theory of spectra. In: **Proceedings of the Cambridge Philosophical Society**. 1923. p. 625-641.
- HE, R., YU, Z. H., ZHANG, R. Y., WU, L., GUNAWAN, A. M., LANE, B. S., WELLS, C. D. Exploring the existing drug space for novel pTyr mimetic and SHP2 inhibitors. **ACS Medicinal Chemistry Letters**, v. 6, n. 7, p. 782-786, 2015.
- HIGASHI, H., NAKAYA, A., TSUTSUMI, R., YOKOYAMA, K., FUJII, Y., ISHIKAWA, S., TANAKA, S. Helicobacter pylori CagA induces Ras-independent morphogenetic response through SHP-2 recruitment and activation. **Journal of Biological Chemistry**, v. 279, n. 17, p. 17205-17216, 2004.
- HØYVIK, I.M., JØRGENSEN, P. Characterization and generation of local occupied and virtual Hartree–Fock orbitals. **Chemical Reviews**, v. 116, n. 5, p. 3306-3327, 2016.
- HOF, P., PLUSKEY, S., DHE-PAGANON, S., ECK, M. J., SHOELSON, S. E. Crystal structure of the tyrosine phosphatase SHP-2. **Cell**, v. 92, n. 4, p. 441-450, 1998.
- HOUSE, J. E. **Fundamentals of Quantum Chemistry. Compllementary Science Series**. Elsevier Science & Technology, 2004.
- HUANG, J., RAUSCHER, S., NAWROCKI, G., RAN, T., FEIG, M., DE GROOT, B. L., MACKERELL JR, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. **Nature Methods**, v. 14, n. 1, p. 71, 2016.
- JENSEN, F. Introduction to computational chemistry. Chichester: John Wiley & Sons, p.429. 1999.
- KAWASAKI, Y., FREIRE, E. Finding a better path to drug selectivity. **Drug Discovery Today**, v. 16, n. 21-22, p. 985-990, 2011.
- KLAMT, A.; SCHUURMANN, G. J. Chem Soc Perkin Trans 2 1993, 5, 799;(b) Baldrige, K.; Klamt. **A. J Chem Phys**, v. 106, p. 6622, 1997.
- KITCHEN, D. B., DECORNEZ, H., FURR, J. R., BAJORATH, J. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature Reviews Drug Discovery**, v. 3, n. 11, p. 935, 2004.
- KUREBAYASHI, J., OKUBO, S., YAMAMOTO, Y., IKEDA, M., TANAKA, K., OTSUKI, T., SONOO, H. Additive antitumor effects of gefitinib and imatinib on anaplastic thyroid cancer cells. **Cancer Chemotherapy and Pharmacology**, v. 58, n. 4, p. 460-470, 2006.
- KOLLMAN, Peter. Free energy calculations: applications to chemical and biochemical phenomena. **Chemical Reviews**, v. 93, n. 7, p. 2395-2417, 1993.

KORB, O., STÜTZLE, T., EXNER, T. E. An ant colony optimization approach to flexible protein–ligand docking. **Swarm Intelligence**, v. 1, n. 2, p. 115-134, 2007.

KORB, O., STUTZLE, T., EXNER, T. E. Empirical scoring functions for advanced protein– ligand docking with PLANTS. **Journal of Chemical Information and Modeling**, v. 49, n. 1, p. 84-96, 2009.

KUNTZ, I. D., BLANEY, J. M., OATLEY, S. J., LANGRIDGE, R., FERRIN, T. E. A geometric approach to macromolecule-ligand interactions. **Journal of Molecular Biology**, v. 161, n. 2, p. 269-288, 1982.

LABBÉ, D. P., HARDY, S., TREMBLAY, M. L. Protein tyrosine phosphatases in cancer: friends and foes!. In: **Progress in Molecular Biology and Translational Science**. Academic Press, 2012. p. 253-306.

LAN, L., HOLLAND, J. D., QI, J., GROSSKOPF, S., VOGEL, R., GYÖRFFY, B., BIRCHMEIER, W. Shp2 signaling suppresses senescence in PyMT-induced mammary gland cancer in mice. **The EMBO journal**, p. e201489004, 2015.

LAWRENCE, H. R., PIREDDU, R., CHEN, L., LUO, Y., SUNG, S. S., SZYMANSKI, A. M., LAWRENCE, N. J. Inhibitors of Src homology-2 domain containing protein tyrosine phosphatase-2 (Shp2) based on oxindole scaffolds. **Journal of Medicinal Chemistry**, v. 51, n. 16, p. 4948-4956, 2008.

LEACH, A. R. *Molecular Modelling - Principles and Applications*. England: Person Prentice Hall, p.744. 2001.

LINDEN, R. *Algoritmos genéticos: uma importante ferramenta da inteligência computacional*. Brasport, 2006.

LIU, W., YU, B., XU, G., XU, W. R., LOH, M. L., TANG, L. D., QU, C. K. Identification of cryptotanshinone as an inhibitor of oncogenic protein tyrosine phosphatase SHP2 (PTPN11). **Journal of Medicinal Chemistry**, v. 56, n. 18, p. 7212-7221, 2013.

LOH, M. L., Reynolds, M. G., Vattikuti, S., Gerbing, R. B., Alonzo, T. A., Carlson, E., Meshinchi, S. PTPN11 mutations in pediatric patients with acute myeloid leukemia: results from the Children's Cancer Group. **Leukemia**, v. 18, n. 11, p. 1831, 2004.

MORGON, N. H., COUTINHO, K. *Métodos de química teórica e modelagem molecular*. Editora Livraria da Física, 2007.

MORRIS, G. M., GOODSSELL, D. S., HALLIDAY, R. S., HUEY, R., HART, W. E., BELEW, R. K., & OLSON, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1639-1662, 1998.

MUEGGE, I., MARTIN, Y. C. A general and fast scoring function for protein– ligand interactions: a simplified potential approach. **Journal of Medicinal Chemistry**, v. 42, n. 5, p. 791-804, 1999.

NARDOZZA, A. P., D'ORAZIO, M., TRAPANNONE, R., CORALLINO, S., FILOMENI, G., TARTAGLIA, M., CASTAGNOLI, L. ROS and EGF are antagonistic cues controlling SHP-2 dimerization. **Molecular and Cellular Biology**, p. MCB. 06674-11, 2012.

NIU, T., LIANG, X., YANG, J., ZHAO, Z., ZHOU, G. W. Kinetic comparison of the catalytic domains of SHP-1 and SHP-2. **Journal of Cellular Biochemistry**, v. 72, n. 1, p. 145-150, 1999.

NODA, S., TAKAHASHI, A., HAYASHI, T., TANUMA, S. I., HATAKEYAMA, M. Determination of the catalytic activity of LEOPARD syndrome-associated SHP2 mutants toward parafibromin, a bona fide SHP2 substrate involved in Wnt signaling. **Biochemical and Biophysical Research Communications**, v. 469, n. 4, p. 1133-1139, 2016.

O'BOYLE, N. M., BANCK, M., JAMES, C. A., MORLEY, C., VANDERMEERSCH, T., & HUTCHISON, G. R. Open Babel: An open chemical toolbox. **Journal of Cheminformatics**, v. 3, n. 1, p. 33, 2011.

OLIVEIRA, Fernanda G. *et al.* Molecular docking study and development of an empirical binding free energy model for phosphodiesterase 4 inhibitors. **Bioorganic & medicinal chemistry**, v. 14, n. 17, p. 6001-6011, 2006.

OHNISHI, Naomi *et al.* Transgenic expression of Helicobacter pylori CagA induces gastrointestinal and hematopoietic neoplasms in mouse. **Proceedings of the National Academy of Sciences**, v. 105, n. 3, p. 1003-1008, 2008.

OPREA, T. I. Chemoinformatics in drug discovery. Weinheim: Wiley-VCH, p.493. 2005.

PAGADALA, Nataraj S.; SYED, Khajamohiddin; TUSZYNSKI, Jack. Software for molecular docking: a review. **Biophysical Reviews**, v. 9, n. 2, p. 91-102, 2017.

POPLE, J. A.; BEVERIDGE, D. L.; DOBOSH, P. A. Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap. **The Journal of Chemical Physics**, v. 47, n. 6, p. 2026-2033, 1967.

QU, C. K. The SHP-2 tyrosine phosphatase: signaling mechanisms and biological functions. **Cell Research**, v. 10, n. 4, p. 279, 2000.

REEVES, C. R. Modern heuristic techniques for combinatorial problems. Advanced topics in computer science. Mc Graw-Hill, 1995.

- ROCCOGRANDI, L., BINDER, Z. A., ZHANG, L., ACETO, N., ZHANG, Z., BENTIREN-ALJ, M., O'ROURKE, D. M. SHP2 regulates proliferation and tumorigenicity of glioma stem cells. **Journal of Neuro-Oncology**, v. 135, n. 3, p. 487-496, 2017.
- ROCHA, G. B., FREIRE, R. O., SIMAS, A. M., STEWART, J. J. Rm1: A reparameterization of am1 for h, c, n, o, p, s, f, cl, br, and i. **Journal of Computational Chemistry**, v. 27, n. 10, p. 1101-1111, 2006.
- SÁ, C. G. Planejamento, síntese e avaliação farmacológica in vitro de novos compostos N-arilpiperazínicos candidatos a protótipos antipsicóticos. Tese de doutorado. UFRJ, 2017
- SANT'ANNA, C. M. R. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). **Química Nova**, v. 25, n. 3, p. 505-512, 2002.
- SANT'ANNA, C. M. R. Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução. **Revista Virtual de Química**, v. 1, n. 1, p. 49-57, 2009.
- SKOOG, D. A., WEST, D. M., HOLLER, F. J., CROUCH, S. Fundamentals of analytical chemistry. Nelson Education, 2013.
- SLATER, J. C. The theory of complex spectra. **Physical Review.**, v.34, p.1293-1322, 1929.
- SERRANO, M. SHP2: a new target for pro-senescence cancer therapies. **The EMBO journal**, v. 34, n. 11, p. 1439-1441, 2015.
- M SCOTT, L., R LAWRENCE, H., M SEBTI, S., J LAWRENCE, N., WU, J. Targeting protein tyrosine phosphatases for anticancer drug discovery. **Current Pharmaceutical Design**, v. 16, n. 16, p. 1843-1862, 2010.
- SCOTT, L. M., CHEN, L., DANIEL, K. G., BROOKS, W. H., GUIDA, W. C., LAWRENCE, H. R., WU, J. Shp2 protein tyrosine phosphatase inhibitor activity of estramustine phosphate and its triterpenoid analogs. **Bioorganic & Medicinal Chemistry Letters**, v. 21, n. 2, p. 730-733, 2011.
- SHI, Z. Q., YU, D. H., PARK, M., MARSHALL, M., FENG, G. S. Molecular mechanism for the Shp-2 tyrosine phosphatase function in promoting growth factor stimulation of Erk activity. **Molecular and Cellular Biology**, v. 20, n. 5, p. 1526-1536, 2000.
- STEWART, J. J. P. Optimization of parameters for semiempirical methods II. Applications. **Journal of Computational Chemistry**, v. 10, n. 2, p. 221-264, 1989.

STEWART, J. J. P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. **International Journal of Quantum Chemistry**, v. 58, n. 2, p. 133-146, 1996.

STEWART, J. J. P. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. **Journal of Molecular Modeling**, v. 13, n. 12, p. 1173-1213, 2007.

STEWART, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. **Journal of Molecular Modeling**, v. 19, n. 1, p. 1-32, 2013.

TARCSAY, Á., KESERŰ, G. M. Is there a link between selectivity and binding thermodynamics profiles?. **Drug Discovery Today**, v. 20, n. 1, p. 86-94, 2015.

THIEL, W., VOITYUK, A. A. Extension of the MNDO formalism to d orbitals: Integral approximations and preliminary numerical results. **Theoretica Chimica Acta**, v. 81, n. 6, p. 391-404, 1992.

TONKS, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. **Nature Reviews Molecular Cell Biology**, v. 7, n. 11, p. 833, 2006.

VERDONK, M. L., COLE, J. C., HARTSHORN, M. J., MURRAY, C. W., TAYLOR, R. D. Improved protein–ligand docking using GOLD. **Proteins: Structure, Function, and Bioinformatics**, v. 52, n. 4, p. 609-623, 2003.

WANG, S., MILNE, G. W., NICKLAUS, M. C., MARQUEZ, V. E., LEE, J., BLUMBERG, P. M. Protein kinase C. Modeling of the binding site and prediction of binding constants. **Journal of Medicinal Chemistry**, v. 37, n. 9, p. 1326-1338, 1994.

WEINER, S. J., KOLLMAN, P. A., CASE, D. A., SINGH, U. C., GHIO, C., ALAGONA, G., WEINER, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. **Journal of the American Chemical Society**, v. 106, n. 3, p. 765-784, 1984.

WU, J.; LAWRENCE, N. J.; SEBTI, S. M.; LAWRENCE, H. R.; University Of South Florida, H. Lee Moffitt Cancer Center and Research Institute, Inc. 2012. (US20120034186 A1)

WU, N. J. LAWRENCE, S. M. SEBTI, H. R. LAWRENCE. University Of South Florida, H. Lee Moffitt Cancer Center and Research Institute, Inc. 2008. (US2008/0176309 A1)

XU, R., YU, Y., ZHENG, S., ZHAO, X., DONG, Q., HE, Z., XU, X. Overexpression of Shp2 tyrosine phosphatase is implicated in leukemogenesis in adult human leukemia. **Blood**, v. 106, n. 9, p. 3142-3149, 2005.

YANG, J., LIANG, X., NIU, T., MENG, W., ZHAO, Z., ZHOU, G. W. Crystal structure of the catalytic domain of protein-tyrosine phosphatase SHP-1. **Journal of Biological Chemistry**, v. 273, n. 43, p. 28199-28207, 1998.

YANG, J., CHENG, Z., NIU, T., LIANG, X., ZHAO, Z. J., ZHOU, G. W. Structural basis for substrate specificity of protein-tyrosine phosphatase SHP-1. **Journal of Biological Chemistry**, v. 275, n. 6, p. 4066-4071, 2000.

YANG, J., NIU, T., ZHANG, A., MISHRA, A. K., ZHAO, Z. J., ZHOU, G. W. Relation between the flexibility of the WPD loop and the activity of the catalytic domain of protein tyrosine phosphatase SHP-1. **Journal of Cellular Biochemistry**, v. 84, n. 1, p. 47-55, 2002.

YU, Z. H., CHEN, L., WU, L., LIU, S., WANG, L., ZHANG, Z. Y. Small molecule inhibitors of SHP2 tyrosine phosphatase discovered by virtual screening. **Bioorganic & Medicinal Chemistry Letters**, v. 21, n. 14, p. 4238-4242, 2011.


ZHANG, X., HE, Y., LIU, S., YU, Z., JIANG, Z. X., YANG, Z., WANG, L. Salicylic acid based small molecule inhibitor for the oncogenic Src homology-2 domain containing protein tyrosine phosphatase-2 (SHP2). **Journal of Medicinal Chemistry**, v. 53, n. 6, p. 2482-2493, 2010.

ZHOU, X., COAD, J., DUCATMAN, B., AGAZIE, Y. M. SHP2 is up-regulated in breast cancer cells and in infiltrating ductal carcinoma of the breast, implying its involvement in breast oncogenesis. **Histopathology**, v. 53, n. 4, p. 389-402, 2008.

ZHOU, T., HUANG, D., CAFLISCH, A. Is quantum mechanics necessary for predicting binding free energy?. **Journal of Medicinal Chemistry**, v. 51, n. 14, p. 4280-4288, 2008.

8 APÊNDICES


- Artigo publicado:



RVQ

Revista Virtual de Química

ISSN 1984-6835



ANOS

Artigo

Proteína Tirosina Fosfatase SHP-2 e sua Relação com o Câncer

Rocha, S. F. L. S.; Salles, C. M. C.; Sant'Anna, C. M. R.*

Rev. Virtual Quim., 2017, 9 (3), 1342-1363. Data de publicação na Web: 6 de junho de 2017

<http://rvq.sbq.org.br>

Protein Tyrosine Phosphatase SHP-2 and its Relation with Cancer

Abstract: This paper presents an introduction to the cellular signaling mechanism governed by the balanced action of protein tyrosine kinases (PTK) and protein tyrosine phosphatases (PTP) by phosphorylation / dephosphorylation of tyrosine residues in proteins. For many years PTK were the main targets for the development of anticancer drugs and today many PTK inhibitors are already widely used in the cancer treatment. Inhibitors of PTP are still in the development stage because of misconceptions about this family of enzymes. Thus, this review seeks to clarify some important points that justify the current interest in PTP as a target in cancer treatment. An introductory description of the classification is made, together with a discussion on structure and PTP mechanism of action with particular focus on SHP family composed of two PTP: SHP-1 and SHP-2. This paper discusses the effects of over expression or gain of function mutations occurred in SHP-2 in the emergence and progression of cancerous state by activation of the RAS / ERK. Finally, some SHP-2 inhibitors that have been discovered to date are presented.

Keywords: Protein tyrosine phosphatase; SHP-2; cancer.

- Artigo enviado para publicação

A Reliable Procedure Combining Molecular Docking and Semiempirical Method PM7 for Identification of Selective Shp2 Inhibitors

Sheisi Fonseca Leite da Silva Rocha¹ Carlos Mauricio R. Sant'Anna^{1,2}

Correspondence to: Carlos Mauricio R. Sant'Anna (E-mail: santanna@ufrrj.br)

¹ Programa de Pós-Graduação em Química, Instituto de Química, Universidade Federal Rural do Rio de Janeiro, Rodovia BR 465, Km 07, Seropédica, RJ, 23890-000

² Departamento de Química Fundamental, Instituto de Química, Universidade Federal Rural do Rio de Janeiro, Rodovia BR 465, Km 07, Seropédica, RJ, 23890-000

ABSTRACT

Shp2 together with Shp1 makes up a small family of protein tyrosine phosphatases. Studies suggest that while the inhibition of Shp2 is advantageous for the treatment of some types of cancer, the inhibition of Shp1 may have the opposite effect because it acts as a suppressor of tumors. In this work, we demonstrated that, despite the complexity of the thermodynamics involved in the enzyme/inhibitor interaction, the selectivity for Shp2 of a series of 52 inhibitors could be effectively correlated with the difference in their interaction enthalpy values with both enzymes. To obtain these values, the interaction modes of the inhibitors with Shp2 and Shp1 were initially found by molecular docking. After refinement of the corresponding geometries of the enzymes/inhibitors complexes with the PM7 semi-empirical molecular orbital method, we demonstrated that the resulting interaction enthalpy values can be used as a reliable criterion for identification of selective Shp2 inhibitors, as they are significantly more favorable for Shp2 than Shp1 (significance level $\alpha = 0.01$). This method was successfully validated using another series of 24 compounds, showing moderate correlation ($R = 0.70$) between relative selectivity, based on IC_{50} data, and the relative percentage difference of the calculated interaction enthalpies with the two enzymes.

KEYWORDS: Shp2 inhibitors, PM7 method, molecular docking