



RODRIGO MANCINI SANTOS

**METASTABLE STATES AND SPECTROSCOPIC
PARAMETERS OF PI3K ENZYME INHIBITORS BIASED
SIMULATIONS AND QUANTUM CALCULATIONS FOR EARLY
CANCER DIAGNOSIS**

LAVRAS – MG

2024

RODRIGO MANCINI SANTOS

**METASTABLE STATES AND SPECTROSCOPIC PARAMETERS OF PI3K ENZYME
INHIBITORS:
BIASED SIMULATIONS AND QUANTUM CALCULATIONS FOR EARLY CANCER
DIAGNOSIS**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração Química/Bioquímica, para obtenção do título de Mestre.

Prof. Dr. Teodorico de Castro Ramalho
Orientador

**LAVRAS – MG
2024**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Santos, Rodrigo Mancini

Metastable states and spectroscopic Parameters of PI3K enzyme inhibitors : Biased simulations and quantum calculations for early cancer diagnosis / Rodrigo Mancini Santos. 1^a ed. rev., atual. e ampl. – Lavras : UFLA, 2024.

85 p. : il.

Dissertação(mestrado acadêmico)–Universidade Federal de Lavras, 2024.

Orientador: Prof. Dr. Teodorico de Castro Ramalho.

Bibliografia.

1. Metastable states. 2. MD simulation. 3. ESIPT. I. Ramalho, Teodorico de Castro. II. Título.

RODRIGO MANCINI SANTOS

**METASTABLE STATES AND SPECTROSCOPIC PARAMETERS OF PI3K ENZYME
INHIBITORS: BIASED SIMULATIONS AND QUANTUM CALCULATIONS FOR
EARLY CANCER DIAGNOSIS**

**ESTADOS METAESTÁVEIS E PARÂMETROS ESPECTROSCÓPICOS DE
INIBIDORES DA ENZIMA PI3K: SIMULAÇÕES ENVIESADAS E CÁLCULOS
QUÂNTICOS PARA UM DIAGNÓSTICO PRECOCE DE CÂNCER**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração Química/Bioquímica, para obtenção do título de Mestre.

APROVADA em 02 de agosto de 2024.

Prof. Dr. Nelson Henrique Morgon UNICAMP
Prof. Dr. Helio Anderson Duarte UFMG

Prof. Dr. Teodorico de Castro Ramalho
Orientador

**LAVRAS – MG
2024**

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a minha família, por todo o carinho e apoio durante minha vida. Ao meu pai e minha mãe, agradecer pelo fato de sempre priorizarem a minha educação acima de qualquer outra coisa, por me incentivar a estudar o que gosto e a alcançar meus sonhos.

Gostaria de agradecer também os amigos que fiz durante a graduação e o mestrado. Agradecer o apoio mútuo, o carinho, a parceria, as risadas e os desabafos. Foi um privilégio estudar e conviver com pessoas engajadas, divertidas e companheiras.

Por fim, agradeço a todos aqueles que, de alguma forma, somaram à minha vida para que hoje eu pudesse estar aqui.

Agradeço também a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

Agradeço também a UFLA, por toda infraestrutura e formação.

Muito obrigado!

RESUMO

Em todo o mundo, o câncer afeta a vida de diversas pessoas. Não apenas com sintomas graves, podendo levar ao óbito do paciente, mas também por apresentar tratamentos de alto custo. Uma das grandes contribuições para a problemática apresentada reside no fato de que o diagnóstico precoce desta doença é de grande complexidade, diminuindo as chances de um tratamento ou intervenção bem sucedida. Assim sendo, o uso de sondas espectroscópicas, como os sensores fluorescentes, em exames clínicos de imagem, se mostram como uma boa estratégia a fim de aprimorar o diagnóstico em estágio inicial da enfermidade mencionada. Atualmente, existe uma gama de moléculas disponíveis para tal tarefa, sendo elas baseadas em fenômenos fotofísicos como a Transferência Intramolecular de Prótons no Estado Excitado (ESIPT). Entretanto, pouco se estuda sobre os efeitos de estados metaestáveis de sensores na fluorescência ESIPT. Dessa maneira, o estudo apresentado visou investigar como conformações em estado de não equilíbrio afetam a fluorescência ESIPT da molécula 2-(2'-hidroxi-4'-aminofenil)benzotiazol (HABT), um composto da classe dos fenilbenzotiazóis (PBTs). Para isso, três simulações de Dinâmica Molecular (DM) enviesadas foram realizadas, de forma a obter arquivos de trajetória para posterior análise utilizando a Análise de Componentes Principais (PCA). Os três sistemas estudados foram: HABT no vácuo (1), HABT em água (2) e HABT ancorado na proteína PI3K em água (3). Com isso, foi observado como o solvente e a ancoragem em proteína marcadora afetam o acesso de estados metaestáveis do HABT, em que estes foram caracterizados em termos conformacionais. Foi observado que, em um estado de não equilíbrio e considerando as interações entre a molécula HABT e a proteína PI3K, a maior parte das conformações acessadas mantiveram sua performance no ESIPT. Portanto, ao se trazer a complexidade de não equilíbrio para a modelagem de um sistema biológico, sendo esta de grande importância no desenvolvimento de novos fármacos, o estudo foi capaz de elucidar a performance do sensor proposto para um diagnóstico precoce de câncer. Em suma, o presente trabalho contribui de maneira significativa na busca por soluções para o câncer, em que mostrou ser possível a utilização da abordagem complexa de não equilíbrio na proposição de sensores fluorescentes para o diagnóstico de doenças.

Palavras-chave: Câncer; simulação de DM; ESIPT; estados metaestáveis; sondas espectroscópicas; sensores fluorescentes; fenilbenzotiazóis; HABT.

ABSTRACT

Cancer affects the life of many people worldwide. Not only by the severe symptoms, which can lead the patient to death, but also due to the high cost of the treatments. One of the major contributions for the presented problem lies in the fact that the early diagnosis of the mentioned disease is of great complexity, reducing the chances of a successful treatment or intervention. Hence, the use of spectroscopic probes, like fluorescent sensors, with imaging clinical examination, shows to be a good strategy for enhancing the early stage diagnosis of the mentioned illness. Nowadays, there is a range of molecules available for such a task, being based on photophysical phenomena like the Excited State Intramolecular Proton Transfer (ESIPT). Despite that, there are few works studying metastable states effects on the ESIPT fluorescence. Therefore, the present study main goal was to investigate how different conformations in a state of nonequilibrium affects the ESIPT fluorescence of the 2-(2'-hidroxi-4'-aminophenyl)benzothiazole (HABT), a molecule of the class of the phenylbenzothiazoles (PBTs). Thereunto, two biased MD simulations were carried out, obtaining the trajectory files for further analysis with the Principal Component Analysis (PCA). The three systems studied were: HABT only (1), HABT in water reservoir (2) and HABT anchored in protein PI3K in the presence of water (3). Therewith, it was observed how the solvent and protein anchoring affects the access of metastable states of the HABT molecule, in which they were characterized in conformational terms. It was observed that, by considering a nonequilibrium state and the interactions between the HABT molecule with PI3K protein, the most part of the accessed conformations were capable of performing the ESIPT process. Therefore, by bringing the complexity of a nonequilibrium approach for modelling a biological system, being this perspective of great importance in modern drug development, the present study sheds light on the performance of the proposed sensor in early cancer diagnosis. In that regard, the present work significantly contributes to the search for better solutions for cancer disease, in which it was shown to be possible to use the nonequilibrium approach in order to propose new fluorescent sensors for disease diagnosis.

Keywords: Cancer; MD simulation; ESIPT; metastable states; spectroscopic probes; fluorescent sensors; phenylbenzothiazoles; HABT.

INDICADORES DE IMPACTO

O câncer é uma doença extremamente complexa, e que acomete muitas pessoas em todo o mundo. É estimado que 1 em cada 5 pessoas irão desenvolver câncer ao longo de suas vidas, em que um montante considerável dessas pessoas virão a óbito pela enfermidade. Nesse sentido, o desenvolvimento de estratégias para o combate desta doença é de suma importância. Dessa maneira, o presente estudo se mostra como mais uma contribuição para o esforço global realizado neste combate. É notório como o diagnóstico precoce da enfermidade em questão aumenta consideravelmente a chance de sobrevivência do paciente, uma vez que, a partir de um diagnóstico em estágios iniciais da doença, as chances de uma intervenção e tratamento bem sucedido aumentam consideravelmente. Nesse sentido, a partir de uma abordagem ainda pouco explorada na literatura, o presente trabalho propõe a utilização de compostos químicos para fins diagnósticos. Com a abordagem utilizada, tal proposição seria capaz de considerar a complexidade enfrentada em sistemas biológicos, complexidade essa que se mostra um grande desafio quando se deseja desenvolver estratégias para o combate ao câncer e demais doenças. Assim sendo, o presente trabalho possui o potencial de impactar a vida de muitas pessoas em todo o planeta, sendo este capaz de propor estratégias para o diagnóstico de câncer, além de servir como um ponto de partida para o desenvolvimento de novas estratégias para esta doença.

IMPACT INDICATORS

Cancer is an extremely complex disease, which affects the life of many people worldwide. It is estimated that 1 in 5 people will develop cancer during its lifetime, where the outcome can be death. In this sense, developing new strategies against this disease is of great importance. Therefore, the present study is a contribution for this global effort against cancer. It is notorious that an early diagnosis of this illness substantially increases the chances of survival of the patient, once in early stages of the disease, the rate of a successful medical intervention and treatment is considerably higher. Hence, from a little explored approach, the present work proposes the utilization of chemical compounds for an early diagnosis of cancer. With the used approach, such proposal would be able to consider the complexity faced when working with biological applications, complexity which is a strong challenge when aiming to develop new strategies against cancer and other diseases. Hence, the present work has the potential of impacting the life of many people worldwide, being capable of proposing strategies for cancer diagnosis, and serve as a starting point for the development of new strategies against this disease.

LISTA DE FIGURAS

Figura 3.1 – A figura apresenta a estrutura do benzotiazol, em que esta é constituída da porção benzeno (em vermelho) e porção tiazol (em azul).	23
Figura 3.2 – Estrutura do HABT, um derivado de benzotiazol, constituído da porção benzeno (vermelho), tiazol (azul) e fenil (preto).	24
Figura 3.3 – Estrutura do ácido salicílico, em que em azul tem-se o grupo hidroxila e em vermelho, o grupo carboxílico.	25
Figura 3.4 – Esquema representativo do processo ESIPT utilizando a molécula de HABT para exemplificação.	26
Figura 3.5 – Ilustração do mecanismo geral proposto pelo modelo de Eigen-Weller para a transferência de prótons intra e intermolecular.	27
Figura 3.6 – Ilustração de uma curva de energia potencial. No ponto V_1 , em azul, tem-se um mínimo local (estado metaestável), separado do ponto V_0 , o mínimo global (equilíbrio), por uma barreira de energia ΔV	32
Figura 4.1 – HABT structure, a benzothiazole derivative, constituted from a benzene (red), thiazole (blue), and phenyl (black).	51
Figura 4.2 – RMSD (\AA) vs Time (ns). The RMSD profile of HABT is shown in red, and PI3K in blue. In the figure caption, it appears values of: average / standard deviation.	54
Figura 4.3 – Captured interactions between HABT and protein PI3K. The ASP-HABT interaction occurs between the carboxylic group of ASP and the -OH group of HABT. The LYS-HABT interaction occurs between the LYS carbonic chain and the benzene group of HABT.	55
Figura 4.4 – Chosen descriptors for further analysis, being (a) dihedral 1, and (b) dihedral 2.	57
Figura 4.5 – Biplot graph made from PCA analysis for system 1. Metastates were labeled according to $(N)\text{Meta } n$, where N denotes the system label, and n , the metastate label in that system.	57
Figura 4.6 – Biplot graph made from PCA analysis for system 2. Metastates were labeled according to $(N)\text{Meta } n$, where N denotes the system label, and n , the metastate label in that system.	59

Figura 4.7 – Biplot graph made from PCA analysis for system 3. Metastates were labeled according to (N)Meta <i>n</i> , where <i>N</i> denotes the system label, and <i>n</i> , the metastate label in that system.	60
Figura 4.8 – Shows the stabilizing interactions that a) stabilizes conformations of (3)Meta 1A and 3A, b) stabilizes conformations of (3)Meta 1B and 3B, and c) stabilizes conformations of (3)Meta 2A and 2B.	62
Figura 4.9 – The figure presents the six selected conformations, being them: 1A, 1B, 2A, 2B, 3A e 3B. Also, the equilibrium conformation is also shown.	65
Figura 4.10 – Shows the ESIPT scheme, where $\Delta E1$, $\Delta E2$ and $\Delta E3$ represents the energy variation between the states, values which are presented in Table 4.2.	66
Figura 4.11 – Shows a summary of the obtained spectroscopic results.	69
Figura 6.1 – RMSD analysis of HABT done for a) system 1, and b) system 2.	83
Figura 6.2 – Shows the -OH an -N= intramolecular h-bond interaction of HABT.	83
Figura 6.3 – Shows the competitive intermolecular h-bond interaction observed after solvation.	84
Figura 6.4 – Shows the labels assigned for each accessed metastate.	84

LISTA DE TABELAS

Tabela 4.1 – Table shows the values of dihedral 1, 2 and distance N···H of each representative conformation selected.	64
Tabela 4.2 – Shows the obtained energy differences between states, in eV, according to the shown in Figure 4.10.	67
Tabela 6.1 – Table shows the mean values of dihedral 1, 2 and distance N···H.	85

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS DA PESQUISA	17
2.1	Objetivo geral	17
2.2	Objetivos específicos	17
3	REFERENCIAL TEÓRICO	19
3.1	Câncer	19
3.2	Sondas espectroscópicas	21
3.3	Benzotiazóis e derivados	23
3.4	Transferência Intramolecular de Prótons no Estado Excitado (ESIPT)	24
3.5	Sistemas fora do equilíbrio	27
3.6	Estados metaestáveis	31
3.7	Métodos para acessar estados metaestáveis	33
3.7.1	Simulação de Dinâmica Molecular (<i>MD Simulation</i>)	33
3.7.2	Simulação de Dinâmica Molecular enviesada (<i>biased MD Simulation</i>)	35
3.7.3	PCA na identificação e interpretação de estados metaestáveis	42
3.8	Estudando o ESIPT	44
4	ARTICLE	48
4.1	Introduction	48
4.2	Methodology	50
4.2.1	HABT structure and parametrization	50
4.2.2	Unbiased Molecular Dynamics (MD) simulations	51
4.2.3	Biased Molecular Dynamics (MD) simulations	51
4.2.4	Principal Component Analysis (PCA) and conformational selection	52
4.2.5	QM calculations setup	52
4.3	Results and discussions	53
4.3.1	Unbiased MD simulation analysis	53
4.3.2	Biased MD simulations: assessing metastates distribution and evaluating the solvent effect	56
4.3.3	Biased MD simulations: assessing the distribution of metastable states and evaluating the effect of the protein environment	59
4.3.4	Conformational selection	63

4.3.5	Fluorescence properties in the protein environment	66
4.4	Conclusion	69
5	CONCLUSÃO	72
	REFERÊNCIAS	81
6	SUPPORTING INFORMATION	83
6.1	Unbiased MD simulation analysis	83
6.2	Labeling metastable states	84
6.3	Conformational selection	85

1 INTRODUÇÃO

O câncer é uma doença extremamente complexa, e acomete diversas pessoas em todo o mundo, levando a óbito um montante considerável de indivíduos (U. S. CENTERS FOR DISEASE CONTROL AND PREVENTION, 2023; INSTITUTO NACIONAL DE CÂNCER - INCA, 2023). De acordo com o Instituto Nacional de Câncer - INCA (2023), no Brasil, 231.694 pessoas vieram a óbito por câncer em 2021. Parte da contribuição para essa estatística vem do fato que o diagnóstico precoce desta enfermidade é complexo. Dessa maneira, o uso de sondas espectroscópicas se mostra bastante vantajoso, uma vez que possibilita um diagnóstico não invasivo e eficiente, utilizando-se de proteínas marcadoras (HAN et al., 2022; ZHOU et al., 2020).

Nesse sentido, a técnica de Ressonância Magnética de Imagem (RMI) se mostra a altura da complexidade da problemática apresentada, uma vez que permite um diagnóstico não invasivo de doenças como o câncer (GONÇALVES et al., 2015; HEQUET et al., 2020; KHALIGHINEJAD et al., 2020). Assim sendo, de forma a aumentar a eficiência da técnica de RMI, os chamados agentes de contraste são utilizados, uma vez que são capazes de aumentar a razão sinal/ruído, melhorando assim a eficiência e exatidão do diagnóstico (KHALIGHINEJAD et al., 2020; HEQUET et al., 2020; HOBSON et al., 2019). Dessa maneira, diversos agentes de contraste estão disponíveis comercialmente, sendo os mais utilizados aqueles a base de gadolínio (Gd^{3+}), uma vez que estes apresentam propriedades magnéticas únicas (WYSKOCCA-GAJDA et al., 2021; WANG et al., 2019; ROHRER et al., 2005).

Entretanto, devido à alta toxicidade do íon Gd^{3+} , e ao fato de que a longo prazo são liberados no organismo a partir da utilização sistemática de agentes de contraste, sua utilização é restringida, e novas estratégias devem ser pensadas no anseio de um diagnóstico rápido e eficiente (WANG et al., 2019; YI et al., 2019). A partir disso, pode-se pensar na utilização de sondas espectroscópicas ópticas, como os sensores fluorescentes (KIM et al., 2020; GEORGIEV et al., 2023). Quando comparado à técnica de RMI, sondas fluorescentes apresentam alta seletividade, baixo custo e maior biocompatibilidade, mostrando-se assim como uma boa alternativa (GEORGIEV et al., 2023; ZHOU et al., 2020).

Na temática de sondas fluorescentes, diversos fenômenos fotofísicos podem explicar seu mecanismo de ação, sendo um dos mais famosos a Transferência Intramolecular de Prótons no Estado Excitado (ESIPT) (GEORGIEV et al., 2023; SEDGWICK et al., 2018). Moléculas ca-

pazes de performar o ESIPT, dentre elas os derivados de benzotiazóis como os fenilbenzotiazóis (PBTs), se apresentam como excelentes sondas espectroscópicas, uma vez que devido ao fato de emitirem luz em dois comprimentos de onda diferentes, apresentam alta seletividade e eficiência (SEDGWICK et al., 2018). Neste contexto, a diferença entre os dois comprimentos de onda emitido, chamada de desvio de Stokes, é de suma importância quando se pensa na eficiência do sensor (PADALKAR et al., 2012; SEDGWICK et al., 2018).

Contudo, as diferenças conformacionais podem afetar a eficiência do processo ESIPT, aumentando ou diminuindo a fluorescência ESIPT de um dado sensor fluorescente (HU et al., 2017; QI et al., 2020). Portanto, estudos como o publicado por Hu et al. (2017) buscam entender como diferentes conformações de uma molécula afetam o processo ESIPT, de forma a ampliar a gama de moléculas disponíveis como sensores fluorescentes a partir da otimização do sensor. Não apenas isso, é importante ressaltar que sistemas em equilíbrio são raros na natureza, uma vez que constantemente estão trocando matéria e energia com o ambiente (DEMIREL, 2014; GHASSEMI; SHAHIDIAN, 2017; ZHANG; OUYANG, 2021). Um grande exemplo de sistema com diversos processos em um estado de não equilíbrio são os sistemas biológicos, estes em constante troca com o ambiente, absorvendo energia e nutrientes necessários para realizar reações bioquímicas, e expelindo resíduo e calor proveniente destas reações (SCHRÖDINGER, 1944; ZHANG; OUYANG, 2021). Portanto, o fato de sistemas biológicos apresentarem inúmeros processos em um estado de não equilíbrio termodinâmico faz com que estes assumam estados não estáveis ou metaestáveis, fazendo com que assim, conformações metaestáveis tenham grande relevância no estudo de sensores fluorescentes (DEMIREL, 2014; SRINIVASAN, 2023). Entretanto, devido a falta de literatura performando estudos desse tipo, trabalhos como este mostram um avanço na busca por sensores fluorescentes mais eficientes, baratos e de fácil utilização.

Assim sendo, este estudo visa investigar como os efeitos conformacionais de sistemas fora do equilíbrio afetam a fluorescência ESIPT do composto 2-(2'-hidroxi-4'-aminofenil) benzotiazol (HABT), molécula esta pertencente a classe dos PBTs e de grande interesse como sensor fluorescente (SEDGWICK et al., 2018; MANCINI et al., 2015). Com isso, o estudo visa também avaliar a eficiência deste sensor no diagnóstico de câncer, a partir da proteína marcador PI3K. A via PI3K está associada com a progressão da enfermidade mencionada (PAUL; MUKHOPADHYAY, 2004; JIANG; JI, 2019; LIU et al., 2009; JHA et al., 2015; YU; KOH, 2017; RAZANI et al., 2021), em que existem evidências de que o composto HABT interage

fortemente com essa enzima (ANDOLPHO et al., 2022; MANCINI et al., 2014; MAVROIDI et al., 2016; MACHURA et al., 2013). De forma a avaliar os efeitos conformacionais de sistemas fora do equilíbrio no processo ESIPT, uma metodologia robusta foi utilizada para acessar, identificar e caracterizar diferentes estados metaestáveis do HBT, em que diferentes conformações puderam ser investigadas. O intuito é que, a partir do estudo, a fluorescência ESIPT possa ser compreendida em uma perspectiva de um sistema fora do equilíbrio.

2 OBJETIVOS DA PESQUISA

2.1 Objetivo geral

Investigar os efeitos dos estados metaestáveis do sensor fluorescente 2-(2'-hidroxi-4'-aminofenil)benzotiazol (HABT) no processo ESIPT para fins de diagnóstico de câncer.

2.2 Objetivos específicos

Assim sendo, os objetivos específicos do presente trabalho são:

- a) Empregar uma metodologia para identificação e caracterização de estados metaestáveis;
- b) Selecionar conformações representativas dos estados metaestáveis para posteriores cálculos quânticos;
- c) Investigar como conformações metaestáveis afetam o processo ESIPT e, posteriormente, avaliar os parâmetros espectroscópicos do HABT dentro do sítio ativo da enzima PI3K.

3 REFERENCIAL TEÓRICO

3.1 Câncer

De acordo com o U. S. National Cancer Institute (2021), o câncer é uma doença em que algumas células do corpo crescem descontroladamente e espalham para outras partes do corpo. Dessa maneira, essa rápida e anormal reprodução de células pode formar o que são chamados tumores, que nada mais são que pedaços de tecidos. Entretanto, nem todos os tumores são cancerígenos. Para que um tumor seja considerado cancerígeno ele tem de ser capaz de se espalhar ou invadir tecidos próximos, viajando longas distâncias no corpo e formando novos tumores, processo esse chamado metástase.

No ano de 2020, 1.603.844 novos casos de câncer foram reportados nos Estados Unidos, em que 602.347 pessoas morreram de câncer (U. S. CENTERS FOR DISEASE CONTROL AND PREVENTION, 2023). Já no Brasil, de acordo com o Instituto Nacional de Câncer - INCA (2023), até 18 de julho de 2023, foram reportados 704.080 novos casos de câncer, e em 2021, 231.694 pessoas vieram a óbito pela doença. Portanto, faz-se extremamente necessário pesquisas capazes de aprimorar estratégias para o combate dessa doença, responsável por um número considerável de óbitos em todo o mundo.

Dessa maneira, uma das estratégias fundamentais para o combate ao câncer consiste no aprimoramento do rápido diagnóstico, aumentando assim a chance de sucesso do tratamento, em que para isso a utilização de sondas espectroscópicas para detecção de proteínas marcadoras de tumores apresenta grande relevância (HAN et al., 2022). Nesse sentido, tem-se as tirosina quinases (TKs), um grupo de proteínas com grande relevância na patofisiologia do câncer (PAUL; MUKHOPADHYAY, 2004). Além disso, a superexpressão de TKs é um indicativo do desenvolvimento e progressão dos mais diversos tipos de câncer (HICKS et al., 2010; PAUL; MUKHOPADHYAY, 2004). Portanto, a utilização desse marcador, aliado com compostos fluorescentes, um tipo específico de sondas espectroscópicas, chamado de sensores fluorescentes, como os derivados de benzotiazóis, se mostra uma boa estratégia no diagnóstico precoce e tratamento do câncer (HICKS et al., 2010; PAUL; MUKHOPADHYAY, 2004). Isso ocorre uma vez que os derivados de benzotiazóis apresentam atividade antitumoral (YADAV et al., 2023), em que a ação antitumoral desses compostos reside na sua interação com TKs (HICKS et al., 2010; PAUL; MUKHOPADHYAY, 2004).

Outro importante marcador relacionado ao câncer é a fosfatidilinositol-3-quinase (PI3K), uma enzima que pertence à família das lipídio quinases, estando ela relacionada ao funcionamento e crescimento de células (D'ANGELO et al., 2011; COLLIER et al., 2015). Nesse cenário, a via da PI3K é uma das vias mais conhecidas no desenvolvimento e progressão do câncer humano, uma vez que tem papel central na ploriferação, diferenciação e sobrevivência das células (PAUL; MUKHOPADHYAY, 2004; JIANG; JI, 2019; LIU et al., 2009). Além disso, a ativação dessa via geralmente está relacionada a receptores tirosina quinase (RTKs), cuja ativação constitutiva se dá por meio de mutações de ganho de função, amplificação gênica, entre outros (JIANG; JI, 2019).

Nesse sentido, como diversos cânceres em humanos estão associados a uma alteração genética somática, a qual resulta na ativação da sinalização da PI3K, é notório a importância da enzima mencionada como alvo no desenvolvimento de novas drogas no combate ao câncer (D'ANGELO et al., 2011; COLLIER et al., 2015). Dessa forma, estudos vêm direcionando sua atenção em utilizar a PI3K como alvo de pequenas moléculas inibidoras ATP-competidoras, uma vez que a inibição dessa via se mostra de suma importância no combate ao desenvolvimento da patologia em questão (D'ANGELO et al., 2011; COLLIER et al., 2015; KARAKAS et al., 2006; VOGT et al., 2007; YAP et al., 2008). Portanto, no anseio do desenvolvimento de novos sensores a fim de aprimorar o diagnóstico de câncer em humanos, é razoável inferir a importância do alvo PI3K no planejamento de novas estratégias.

Assim sendo, os derivados de fenilbenzotiazóis se apresentam como um possível sensor quando considerado o alvo PI3K (ANDOLPHO et al., 2022). Isso ocorre uma vez que, derivados de benzotiazol, em especial o composto HABT, interagem fortemente com a enzima PI3K, logo, sua alta seletividade e atividade antitumoral fazem deste composto uma interessante proposta como sensor fluorescente para o diagnóstico de câncer (ANDOLPHO et al., 2022; MANCINI et al., 2014; MAVROIDI et al., 2016; MACHURA et al., 2013). Este fato motivou trabalhos, como o publicado por Andolpho et al. (2022), o qual propôs um complexo de Rênio (Re), conjugado com o composto HABT, sendo o complexo resultante chamado de ReABT, como possível sonda fluorescente no diagnóstico de câncer. Portanto, o uso de derivados de benzotiazóis, como o HABT, com o marcador PI3K, se faz de grande interesse para a busca de um diagnóstico aprimorado para os mais diversos tipos de câncer.

Nesse sentido, é notório a relevância das sondas espectroscópicas, em especial os sensores fluorescentes, na busca por um diagnóstico precoce de câncer. Assim sendo, a fim de

elucidar seu papel no desenvolvimento de estratégias para o combate ao câncer, é necessário explorar o que seriam sondas espectroscópicas. Além disso, também faz-se necessário avaliar seu papel no desenvolvimento de um diagnóstico mais rápido e acurado.

3.2 Sondas espectroscópicas

As sondas espectroscópicas podem ser definidas como substâncias, geralmente exógenas, que ao interagir com analitos ou fatores ambientais têm suas propriedades espectroscópicas alteradas, sendo estas: absorção e emissão de luz, fluorescência e propriedades magnéticas (ZHOU; MA, 2016; BARTOSZ, 2006). Portanto, devido às alterações mencionadas, alvos de interesse podem ser determinados, permitindo assim diversas aplicações (ZHOU; MA, 2016). Entretanto, para que uma sonda espectroscópica seja considerada ideal, a mesma tem de interagir de forma eficiente com seu alvo, além de ser altamente específica (BARTOSZ, 2006). Essas características permitirão com que a sonda em utilização possa ser utilizada em baixas concentrações, impedindo a perturbação do sistema em estudo e evitando imprecisões (BARTOSZ, 2006).

Dessa maneira, sondas espectroscópicas podem ser utilizadas para as mais diversas aplicações, sendo uma destas as aplicações biomédicas para detecção de doenças (ZHOU; MA, 2016). No contexto de diagnóstico de patologias, a Ressonância Magnética de Imagem (RMI) se tornou uma das técnicas mais amplamente utilizadas, uma vez que é capaz, de forma não invasiva, diagnosticar com alta resolução estágios iniciais de doenças como câncer e Alzheimer (GONÇALVES et al., 2015; HEQUET et al., 2020; KHALIGHINEJAD et al., 2020). O RMI usual consiste na detecção de sinais dos prótons de hidrogênio da água, entretanto, este sinal pode por muita das vezes não ser o suficiente para uma detecção com boa resolução, sendo necessário a utilização de agentes de contraste (KHALIGHINEJAD et al., 2020; HEQUET et al., 2020; HOBSON et al., 2019).

Agentes de contraste em RMI são compostos capazes de alterar o tempo de relaxação dos hidrogênios da água, fazendo com que a resolução do sinal aumente, facilitando assim a distinção na imagem (SHARMA; PERVUSHIN, 2020; CHO et al., 2023). Dessa maneira, agentes de contraste são geralmente divididos em dois tipos, aqueles que alteram o tempo de relaxação longitudinal (T_1) e aqueles que alteram o tempo de relaxação transversal (T_2) quando aplicado um pulso de rádio apropriado, ambos sendo úteis para os mais diversos diagnósticos (YI et al., 2019; HEQUET et al., 2020). Atualmente, muitos agentes de contraste estão disponíveis comer-

cialmente, sendo a maioria composto por íons inorgânicos como Gd^{3+} ou Mg^{2+} , devido a sua alta capacidade de alterar o tempo de relaxação nas suas vizinhanças (WYSKOCCA-GAJDA et al., 2021; WANG et al., 2019; ROHRER et al., 2005).

Entretanto, agentes de contraste a base de gadolínio possuem uma considerável toxicidade a longo prazo, uma vez que os íons Gd^{3+} liberados no organismo apresentam alta toxicidade, limitando sua bioaplicabilidade (WANG et al., 2019; YI et al., 2019). Portanto, diversas estratégias vem sendo desenvolvidas de forma a contornar este problema, sendo uma delas, a utilização de sondas espectroscópicas ópticas (KIM et al., 2020; GEORGIEV et al., 2023).

Quando uma luz incide na superfície de um tecido biológico, uma combinação de processos ópticos ocorrem, como absorção, reflexão e fluorescência (KIM et al., 2020). Dessa forma, a análise dessas combinações fornece informações quanto ao estado metabólico e patológico desses tecidos, provendo uma possibilidade de diagnóstico para diversas doenças (WU et al., 2021; KIM et al., 2020). Com isso, o diagnóstico baseado em tecidos permite a avaliação do estado de uma doença a partir da avaliação dos tecidos do corpo (KIM et al., 2020). Neste contexto, a utilização de exames de imagem óptica permitem um diagnóstico não invasivo e em tempo real das mais diversas doenças, uma vez que a visualização dos tecidos é possível, em que sondas espectroscópicas podem ser utilizadas para aumentar a razão sinal-ruído, aumentando a eficiência do diagnóstico (WU et al., 2021).

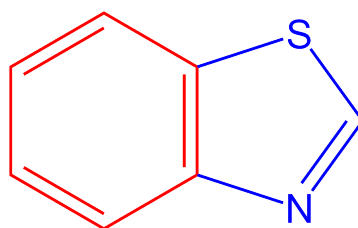
Nesse cenário, a utilização de sondas fluorescentes se encaixa perfeitamente bem, possuindo alta seletividade, eficiência e simplicidade, além de um baixo custo e toxicidade quando comparado a técnicas como a RMI (GEORGIEV et al., 2023; ZHOU et al., 2020). Assim sendo, o estudo feito por Zhou et al. (2020) mostra como sondas fluorescentes podem ser utilizadas para o diagnóstico de doenças extremamente complexas, como a doença de Alzheimer. Outro estudo, publicado por Han et al. (2022), mostra os recentes avanços na utilização de sondas espectroscópicas para o diagnóstico dos mais variados tipos de câncer a partir de enzimas marcadoras. Este importante avanço se faz estritamente necessário, uma vez que o diagnóstico precoce de doenças neurodegenerativas, como o Alzheimer e Parkinson, e também de diversos tipos de câncer, levam à um rápido início do tratamento, aumentando assim sua chance de sucesso (ZHOU et al., 2020; HAN et al., 2022). Dessa forma, o estudo de sondas fluorescentes é de suma importância no anseio de novas tecnologias que permitam o rápido diagnóstico de patologias extremamente complexas, o que consequentemente aumenta a chance de sucesso do tratamento, melhorando a qualidade de vida da população.

Contudo, diversos fenômenos fotofísicos servem de base para entender a detecção e exames de imagem utilizando fluorescência, em que dentre eles, temos a Transferência Intramolecular de Prótons no Estado Excitado (ESIPT) (GEORGIEV et al., 2023; SEDGWICK et al., 2018). Dado a importância deste fenômeno na performance e aplicação de sensores fluorescentes, uma seção será dedicada a este tema ao longo deste referencial teórico. No contexto de moléculas capazes de realizar o processo ESIPT, temos os derivados dos benzotiazóis, classe de compostos explorada na próxima seção.

3.3 Benzotiazóis e derivados

Um benzotiazol é um composto heterocíclico aromático e bicíclico, no qual um anel benzeno é ligado a um anel tiazol, como mostra a Figura 3.1 (SUMIT et al., 2021). Nesse cenário, os derivados de benzotiazóis se mostram como moléculas de grande interesse, uma vez que moléculas que exibem fluorescência ESIPT são aquelas que possuem uma interação por ligação de hidrogênio intramolecular entre um doador de ligação de hidrogênio (como -OH e NH_2) e um aceptor de ligação de hidrogênio (como $=\text{N}-$ e $\text{C}=\text{O}$) (SEDGWICK et al., 2018). Além disso, foi observado que o desvio de Stokes, sendo esta a diferença entre o comprimento de onda absorvido e emitido pela molécula, é mais intenso em moléculas contendo a unidade benzotiazol em sua estrutura (PADALKAR et al., 2012).

Figura 3.1 – A figura apresenta a estrutura do benzotiazol, em que esta é constituída da porção benzeno (em vermelho) e porção tiazol (em azul).

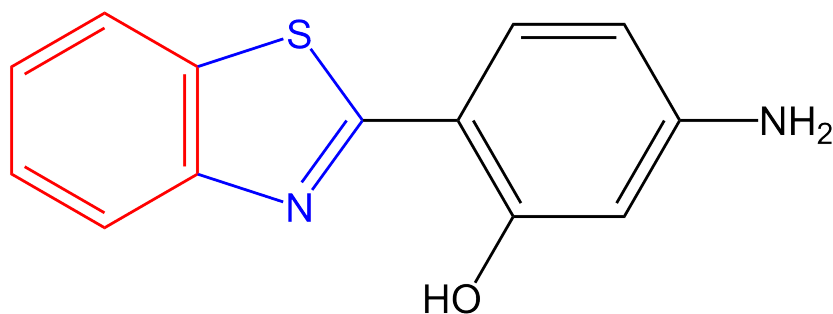


Fonte: Do autor (2023)

Dessa forma, no contexto de sondas espectroscópicas fluorescentes cujo mecanismo de ação consiste na capacidade de performar o ESIPT, temos diversos derivados de benzotiazóis (SEDGWICK et al., 2018; PADALKAR et al., 2012; LONG et al., 2020; SUMIT et al., 2021; MANCINI et al., 2015). Dentre eles, tem-se o composto 2-(2'-hidroxi-4'-aminofenil)benzotiazol (HABT), cuja estrutura é mostrada na Figura 3.2, em que este composto contém unidade benzotiazol e interação de ligação de hidrogênio intramolecular entre um grupo -OH e um grupo

=N-, pertencendo a classe dos fenilbenzotiazóis (PBTs) (SEDGWICK et al., 2018; MANCINI et al., 2015).

Figura 3.2 – Estrutura do HABT, um derivado de benzotiazol, constituído da porção benzeno (vermelho), tiazol (azul) e fenil (preto).



Fonte: Do autor (2023)

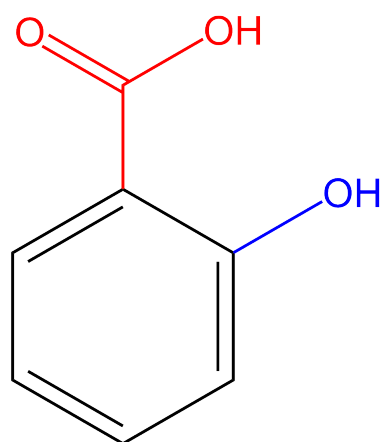
O esqueleto benzotiazol está presente em vários heterociclos bioativos e produtos naturais, sendo esta unidade considerada a principal porção em diversos compostos biologicamente ativos (SUMIT et al., 2021). Portanto, os derivados de benzotiazóis apresentam diversas atividades terapêuticas, como atividade antitumoral, antimicrobiana, analgésica, antifúngica, entre outras (SUMIT et al., 2021; MANCINI et al., 2015; LAW; K.Y., 2022; YADAV et al., 2023). Assim sendo, estudos como o publicado por Mancini et al. (2015) são de grande interesse, uma vez que ao utilizar a capacidade de performar o ESIPT por parte de derivados de benzotiazóis, o diagnóstico de diversas patologias pode ser aprimorado, e o monitoramento de tratamentos também é possibilitado.

3.4 Transferência Intramolecular de Prótons no Estado Excitado (ESIPT)

Sensores fluorescentes são utilizados para as mais diversas aplicações, sejam elas farmacológicas, biológicas, fisiológicas, entre outras (SEDGWICK et al., 2018). Nesse sentido, um dos principais mecanismos para se entender a fluorescência de sondas é a Transferência Intramolecular de Prótons no Estado Excitado (ESIPT) (SEDGWICK et al., 2018; GEORGIEV et al., 2023; LIU et al., 2021; JOSHI; ANTONOV, 2021). No geral, moléculas que exibem fluorescência ESIPT são aquelas que apresentam um doador de ligação de hidrogênio (-OH e NH₂) e um aceptor de ligação de hidrogênio (=N- e C=O), uma vez que no estado excitado existe o aumento da acidez e basicidade dos componentes da molécula capaz de performar o ESIPT (SEDGWICK et al., 2018; JOSHI; ANTONOV, 2021; GRANDJEAN et al., 2021).

O ESIPT foi primeiramente reportado por Weller (1955), sendo observado a partir do estudo com ácido salicílico (SA). Foi observado na estrutura do SA, apresentada na Figura 3.3, que no estado excitado, o hidrogênio do grupo hidroxila se torna mais eletropositivo, enquanto o oxigênio do grupo carboxílico se torna mais eletronegativo, facilitando a transferência de prótons intramolecular (JOSHI; ANTONOV, 2021; GRANDJEAN et al., 2021).

Figura 3.3 – Estrutura do ácido salicílico, em que em azul tem-se o grupo hidroxila e em vermelho, o grupo carboxílico.

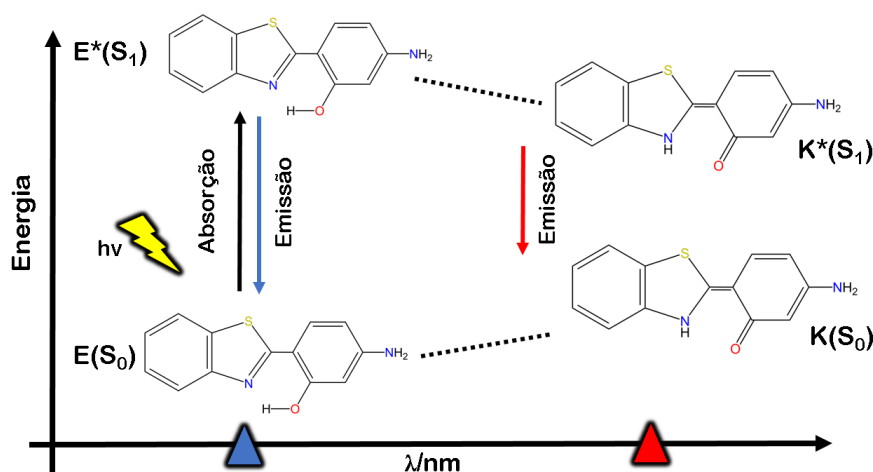


Fonte: Do autor (2023)

Dessa maneira, o processo ESIPT é classificado como um processo que ocorre em duas etapas (LIU et al., 2021). Partindo de uma molécula na sua forma enol, tem-se que após a excitação ocorre uma transferência de prótons em uma curta distância, entre o grupo doador de ligação de hidrogênio e o grupo aceptor de ligação de hidrogênio, formando um contato entre pares de íons (SEDGWICK et al., 2018; GRANDJEAN et al., 2021; LIU et al., 2021). Após isso, o que ocorre é a separação do contato entre os pares de íons, formando assim pares de íons livres, configurando este momento como o fim da formação da configuração ceto (SEDGWICK et al., 2018; LIU et al., 2021). Por fim, a emissão de luz é feita, entretanto, a emissão feita agora pelo composto na sua configuração ceto é menos energética, ou seja, a luz emitida possui um maior comprimento de onda λ do que o da luz absorvida (SEDGWICK et al., 2018). A Figura 3.4 mostra um esquema representativo do processo ESIPT com a molécula HABT, como descrito acima.

O Mecanismo do processo ESIPT é descrito pelo modelo de Eigen-Weller (LIU et al., 2021; WELLER, 1955; EIGEN, 1964). Dessa forma, conforme mostra a Figura 3.5, o mecanismo proposto segue da seguinte forma: excitação do fotoácido ROH por absorção de luz, reduzindo drasticamente seu valor de pK_a (1), transferência ultrarápida do hidrogênio do fotoá-

Figura 3.4 – Esquema representativo do processo ESIPT utilizando a molécula de HABT para exemplificação.



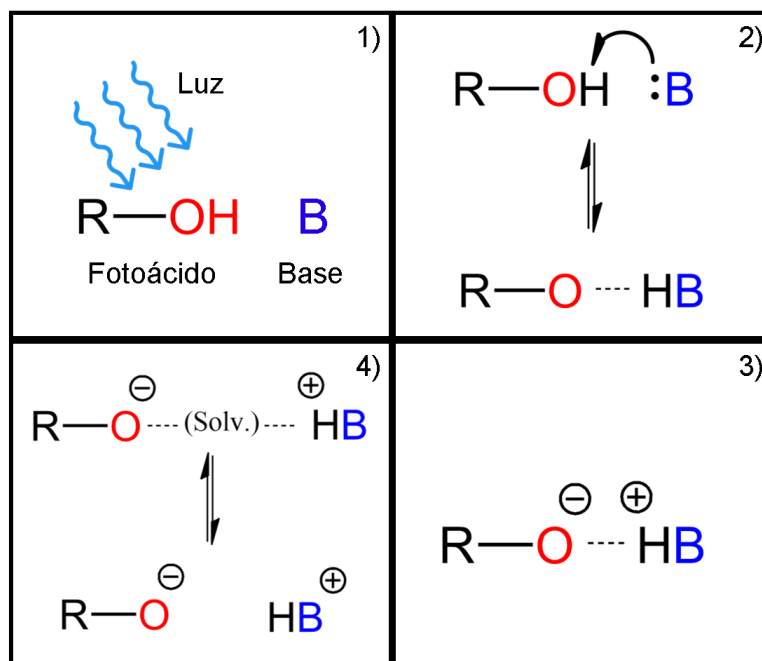
Fonte: Do autor (2023)

cido para moléculas de solvente ou uma base presente na vizinhança (2), formação do par de íons em contato (3), separação do par de íons por moléculas de solvente (4) (GRANDJEAN et al., 2021). Vale ressaltar que a transferência de prótons ultrarápida ocorre na ordem de femtossegundos (fs), ou seja, 10^{-15} segundos (GRANDJEAN et al., 2021; LIU et al., 2021).

A fotofarmacologia é uma disciplina inovadora no âmbito da química médica, uma vez que utiliza de técnicas para controlar a atividade biológica de moléculas através da luz (VELEMA et al., 2014; BROICHHAGEN et al., 2015). Nesse sentido, o estudo de revisão publicado por Kobauri et al. (2023) mostra os avanços nessa área, e como a luz pode ser utilizada para "ativar" conformações biologicamente ativas de moléculas.

Dessa maneira, o mecanismo ESIPT vem sendo estudado em aplicações biológicas para o diagnóstico de diversas doenças, como câncer e Alzheimer, além de outras aplicações de sinalização fluorescente (ZHOU et al., 2020; HAN et al., 2022; SEDGWICK et al., 2018; LONG et al., 2020). Entretanto, em alguns estudos, como os publicados por Hu et al. (2017) e Qi et al. (2020), vem sendo investigado como diferentes conformações de uma molécula capaz de performar o processo ESIPT afetam este processo. Neste escopo, é importante entender como diferentes conformações afetam o ESIPT, uma vez que, como exemplo, a salicilidenoanilina apresenta fraca fluorescência ESIPT, em que a principal motivação é devido a isomerização conformacional via rotação (HU et al., 2017). Assim sendo, o estudo feito por Hu et al. (2017) buscou, a partir da análise de derivados de salicilidenoanilina com conformações travadas, in-

Figura 3.5 – Ilustração do mecanismo geral proposto pelo modelo de Eigen-Weller para a transferência de prótons intra e intermolecular.



Fonte: Adaptado de Grandjean et al. (2021)

investigar quais conformações apresentam maior fluorescência ESIPT, fornecendo alternativas de moléculas com propriedades ópticas mais atraentes.

Portanto, entender como diferentes conformações apresentam diferentes fluorescências ESIPT é uma boa estratégia para ampliar a gama de moléculas disponíveis como sensores fluorescentes, uma vez que novas sínteses podem ser desenvolvidas a partir deste conhecimento. Além disso, diversos processos químicos em seres vivos ocorrem fora do estado de equilíbrio termodinâmico (ZHANG; OUYANG, 2021; SRINIVASAN, 2023). Nesse sentido, a maior parte das interações droga-alvo de pequenas moléculas clinicamente aprovadas consistem em um sistema em não equilíbrio (SRINIVASAN, 2023). Dessa maneira, estudar o ESIPT, utilizando conformações fora do equilíbrio, pode ser de grande relevância no desenvolvimento de novos sensores fluorescentes. Assim sendo, as próximas duas seções se comprometem em explorar a relevância de sistemas fora do equilíbrio e metaestabilidade nesta temática.

3.5 Sistemas fora do equilíbrio

De acordo com Ghassemi e Shahidian (2017), a termodinâmica de não equilíbrio é um ramo da termodinâmica que estuda sistemas físicos que não estão em equilíbrio termodinâmico, entretanto podem ser descritos por variáveis de estado que representam uma extrapolação da-

que as variáveis que são utilizadas para descrever um sistema em equilíbrio termodinâmico. Assim sendo, inúmeros sistemas encontrados na natureza não estão em equilíbrio termodinâmico, indicando que sistemas que permanentemente apresentam comportamento estável e reversível são raros (DEMIREL, 2014; GHASSEMI; SHAHIDIAN, 2017). Isso ocorre uma vez que, sistemas reais estão sujeitos a mudarem ao longo do tempo, uma vez que ocorre constantemente um fluxo de matéria e energia para dentro e fora do sistema, além de estarem sujeitos a ocorrência de reações químicas (GHASSEMI; SHAHIDIAN, 2017). Nesse sentido, enquanto sistemas próximos do equilíbrio termodinâmico apresentam pequenas flutuações, aqueles que estão fora do equilíbrio termodinâmico têm suas flutuações amplificadas, o que pode levar o sistema a um estado não estável ou metaestável (DEMIREL, 2014).

Afinal, o que seriam sistemas fora do equilíbrio termodinâmico? Para responder essa pergunta, pode-se recorrer à uma explicação entrópica. A fim de exemplificar, considere a mistura de dois gases ideais inertes (A e B), estando eles a temperatura e pressão constantes, onde assume-se este sendo um sistema termodinâmico isolado, ou seja, não há fluxo de matéria nem troca de energia entre o sistema e o meio. O movimento das moléculas dos gases é completamente aleatório, e suas moléculas não interagem uma com as outras, uma vez que foram considerados dois gases ideais (LEVINE, 2008).

Assumindo o estado inicial como um recipiente, em que à esquerda tem-se o gás A, e à direita o gás B, separados por uma parede rígida, impermeável e adiabática, o que ocorre quando os gases A e B são colocados em contato pela remoção da parede que os separa? O estado final deste sistema, sendo esta a posição de equilíbrio termodinâmico, é caracterizado pela completa mistura desses gases (LEVINE, 2008). Ou seja, aproximadamente o mesmo número de moléculas de A que estarão na porção esquerda do recipiente também estarão na porção direita, e o mesmo ocorre com o gás B (LEVINE, 2008). Entretanto, apesar de esta ser a posição de equilíbrio termodinâmico, é curioso o fato de esta ser uma posição tão definida. Isso pois, de certa maneira, é contraintuitivo pensar que o movimento aleatório de moléculas levariam à uma posição definida e não aleatória ao longo do tempo. Dito isso, a fim de explicar o porquê isso ocorre, faz-se necessário recorrer à segunda lei da termodinâmica.

Dessa forma, a segunda lei da termodinâmica diz que, para um sistema isolado, à medida que o mesmo avança em direção ao equilíbrio termodinâmico, a entropia S do sistema aumenta (LEVINE, 2008). Portanto, retomando o exemplo criado, na posição de equilíbrio termodinâmico, assumindo 1 mol (6×10^{23} moléculas) do gás A e do gás B, tem-se $\approx 3 \times 10^{23}$ moléculas

do gás A na porção esquerda do recipiente, e a mesma quantia na porção direita do recipiente. O mesmo pode ser dito para as moléculas do gás B. Logo, uma vez que as moléculas não apresentam restrição de movimento, já que movimentam-se randomicamente, a resposta para o porquê o descrito seria a posição do equilíbrio termodinâmico do sistema reside na probabilidade de ocorrência desse estado (LEVINE, 2008). Assim sendo, no equilíbrio termodinâmico de um sistema isolado, o qual é caracterizado pela maximização da entropia, a entropia seria interpretada como uma medida da probabilidade (\mathbf{p}) de um estado do sistema, ou seja, uma função da probabilidade $S = f(p)$ (LEVINE, 2008). Logo, o equilíbrio termodinâmico pode ser compreendido pela entropia do sistema, em que este tende a assumir o estado de maior probabilidade.

Dessa maneira, definido o equilíbrio termodinâmico como o estado de máxima entropia, ou seja, estado de maior probabilidade de ocorrência, faz-se necessário entender agora o que seria o não equilíbrio termodinâmico. Para isso, é necessário desviar da premissa central, a qual foi utilizada a fim de definir o equilíbrio termodinâmico, sendo esta a premissa de um sistema estar isolado do ambiente externo. Isso pois, quando considerado sistemas "abertos", como sistemas biológicos, as implicações são diferentes, uma vez que as flutuações são de grande relevância, devido a troca de matéria e energia que ocorre com o ambiente (LEVINE, 2008; DEMIREL, 2014; GHASSEMI; SHAHIDIAN, 2017; ZHANG; OUYANG, 2021). Em seu livro *What is Life? The Physical Aspect of the Living Cell* (SCHRÖDINGER, 1944), Erwin Schrödinger proporciona uma contribuição dos físicos para compreensão da vida como conhecida, uma vez que o mesmo relata como químicos e biólogos fizeram contribuições essenciais para esta temática. O autor propõe que a diferença crucial entre um sistema "vivo" e um sistema "morto" consiste no seu ordenamento, ou melhor dizendo, no comportamento da sua entropia (SCHRÖDINGER, 1944; ZHANG; OUYANG, 2021).

Nesse sentido, em um sistema sem atividade biológica, como a mistura de dois gases ideais em um recipiente isolado, exemplo discutido em parágrafos anteriores, é uma questão de tempo até que este atinja uma posição estável, desordenada, em que todas as suas atividades intrínsecas irão cessar uma vez atingido o equilíbrio termodinâmico, estado este em que a entropia é maximizada (LEVINE, 2008; ZHANG; OUYANG, 2021). Enquanto isso, a maioria dos sistemas biológicos são mantidos ordenados, em um curso estável, e dificilmente atingem o equilíbrio termodinâmico (SCHRÖDINGER, 1944; ZHANG; OUYANG, 2021). Como discutido, para um sistema isolado, o estado de equilíbrio é aquele cuja entropia é maximizada (LEVINE, 2008), logo, o que confere a sistemas biológicos a constante posição de não equilí-

brio é o fato de que estes trocam matéria e energia com o ambiente externo, absorvendo energia e nutrientes necessários para realizar reações bioquímicas, e expelindo resíduo e calor proveniente destas reações, retardando a maximização da entropia (SCHRÖDINGER, 1944; ZHANG; OUYANG, 2021). Portanto, a posição de não equilíbrio termodinâmico seria definida por um ordenamento, este sendo possível, com mais facilidade, em sistemas onde há troca com o ambiente, seja de energia ou matéria. Definindo agora em termos entrópicos, esta posição seria aquela em que a entropia não foi maximizada, em que a troca de energia e matéria com o ambiente é crucial para retardar o aumento de sua entropia e evitar o estado de maior probabilidade de ocorrência.

Dessa maneira, o ponto chave para que um sistema biológico se mantenha vivo reside na sua capacidade de se manter com "baixa entropia", em que para isso é necessário absorver "entropia negativa" do ambiente externo para anular o aumento de sua entropia (SCHRÖDINGER, 1944; ZHANG; OUYANG, 2021). Portanto, na tendência de atingir o estado de equilíbrio termodinâmico, sendo este o estado mais provável (LEVINE, 2008), sistemas biológicos buscam atrasar este inevitável curso, em que este, segundo Schrödinger (1944), seria a morte do sistema vivo. A contribuição do cientista mencionado teve grande impacto nos campos da biologia e física, apresentando grande aceitação na comunidade científica, uma vez que o metabolismo, processo esse descrito como a troca de matéria e energia em sistemas biológicos, é essencial para a vida (ZHANG; OUYANG, 2021). Por fim, o que torna um sistema biológico vivo é o fato de tentarem se manter em um estado constante de não equilíbrio termodinâmico, retardando a maximização de sua entropia.

Assim sendo, a termodinâmica de não equilíbrio foi aplicada, de maneira bem sucedida, para explicar processos biológicos, como a dobra e desdobra de proteínas e transporte através de membranas (EBELING; SOKOLOV, 2005; GHASSEMI; SHAHIDIAN, 2017). Além disso, conceitos de não equilíbrio podem ser utilizados para desvendar as condições genéticas e epigenéticas para o desenvolvimento de tumores, e assim, explorar novas estratégias para o combate da doença (FANG et al., 2019; WELTER; RIEGER, 2013). Nesse sentido, fatores ambientais podem levar a mudanças ou desequilíbrios na regulação de genes, em que algumas dessas condições podem favorecer o desenvolvimento do câncer (FANG et al., 2019).

Na temática do desenvolvimento de novos fármacos para o combate das mais diversas doenças, é necessário compreender que a maior parte das interações droga-alvo de pequenas moléculas clinicamente aprovadas são de natureza não equilibrada (SRINIVASAN, 2023). Isso

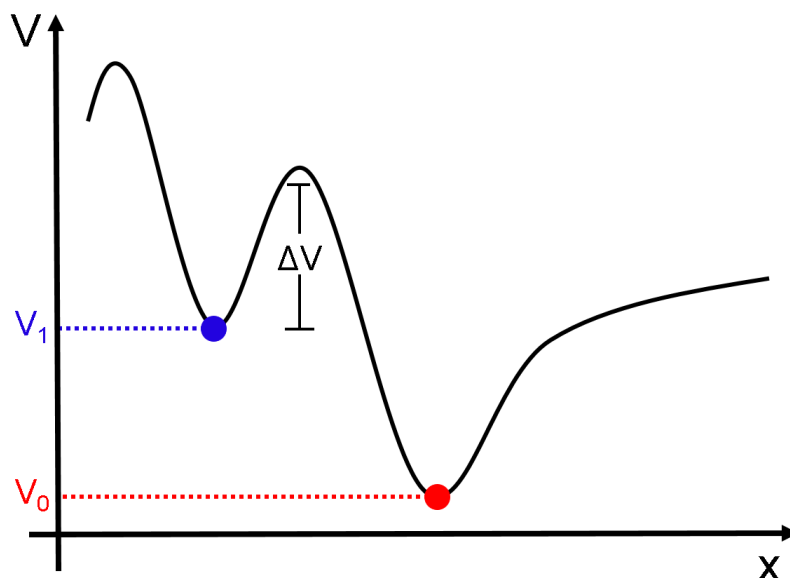
pois, a matéria viva consiste de sistemas longe do equilíbrio, em que o uso de energia externa é de suma importância para manter o sistema fora do equilíbrio, utilizando-se de barreiras cinéticas adequadas (SRINIVASAN, 2023). Portanto, para o planejamento de um novo fármaco, pode ser relevante considerar abordagens fora do equilíbrio (SRINIVASAN, 2023; COPELAND, 2013; COPELAND, 2016; COPELAND, 2021). Assim sendo, para um desenvolvimento moderno de sensores fluorescentes, levar em consideração o não equilíbrio termodinâmico, faz-se necessário, englobando assim a compreensão de estados metaestáveis de um sistema, estes caracterizados por estarem fora do equilíbrio e apresentarem barreiras cinéticas que os separam dos demais estados de um sistema.

3.6 Estados metaestáveis

Por definição, um estado metaestável é aquele que se encontra de forma estável por período de tempo relativamente longo, entretanto, este não se encontra no estado de equilíbrio, estado esse verdadeiramente estável (CHENG, 2008). A Figura 3.6 mostra uma curva de energia potencial, onde é possível reconhecer o estado de metaestabilidade e o estado de equilíbrio. É possível notar que, como o estado metaestável representa um mínimo local na curva de energia potencial, e que sistemas tendem ao equilíbrio (mínimo global), a tendência é que um sistema que se encontre em um estado metaestável não ficará no mesmo para sempre (CHENG, 2008). Entretanto, para que um sistema passe de um estado metaestável para o estado de equilíbrio, uma barreira energética ΔV deve ser vencida, assim sendo, a depender da energia necessária para superar essa barreira, um metaestado pode apresentar um tempo de vida mais longo ou mais curto (CHENG, 2008).

Dessa forma, a identificação de estados metaestáveis pode ser de grande interesse para compreender as diferenças conformacionais de compostos em sistemas complexos, como por exemplo, de proteínas (GHOSH; RANJAN, 2020). Assim sendo, um estado metaestável de uma proteína é uma estrutura cineticamente presa a um mínimo local de energia livre, em que este se encontra separado do mínimo global por uma barreira de energia (GHOSH; RANJAN, 2020). Logo, nem toda conformação pertence a um metaestado, uma vez que para que isso ocorra duas condições devem ser satisfeitas: ser um estado com estabilidade diferente dos demais estados (1) e ter uma barreira energética a ser vencida para que possa assumir outros metaestados ou o estado de equilíbrio (2) (GHOSH; RANJAN, 2020). Contudo, os diferentes metaestados de uma molécula, como uma proteína, são caracterizados por fenômenos que estabilizam uma

Figura 3.6 – Ilustração de uma curva de energia potencial. No ponto V_1 , em azul, tem-se um mínimo local (estado metaestável), separado do ponto V_0 , o mínimo global (equilíbrio), por uma barreira de energia ΔV .



Fonte: Do autor (2023)

nova estrutura, como criação e quebra de interações intramoleculares, portanto, a identificação dos estados metaestáveis se mostra uma ferramenta útil para entender e investigar diferenças conformacionais (GHOSH; RANJAN, 2020).

Apesar de estados metaestáveis serem de grande interesse na compreensão das diferenças conformacionais da estrutura de proteínas, moléculas menores também podem ser investigadas dessa maneira, como mostra o estudo publicado por Marinova et al. (2021). Nesse sentido, vale ressaltar que no mundo moderno, a compreensão de conformações e suas características é de suma importância no desenvolvimento de novos fármacos, uma vez que a interação de um potencial fármaco e seu alvo é controlada pelo seu arranjo conformacional (MARINOVA et al., 2021; PISANI et al., 2014).

Um estudo publicado por Mandal e Panda (2023) mostrou ao estudar alguns oligômeros de 2-fenilpiridina que, a diferença conformacional desse oligômero tem um efeito significativo nas propriedades do estado excitado do composto em estudo. Em outros dois estudos, sendo eles feitos por Hu et al. (2017) e Qi et al. (2020), foram observados como diferentes arranjos conformacionais de moléculas orgânicas pequenas apresentam diferentes resultados para a fluorescência ESIPT. Apesar disso, existe uma carência na literatura investigando como o ESIPT é afetado por diferentes metaestados, caracterizados por diferentes arranjos conformacionais.

Assim sendo, o estudo, através da identificação e caracterização de metaestados, de como o ESIPT é afetado por diferentes arranjos conformacionais, e conseqüentemente qual o efeito dessas conformações na sua qualidade enquanto sensor fluorescente, se faz necessário. Isso pois, no desenvolvimento moderno de novos fármacos, o entendimento do efeito conformacional, considerando o não equilíbrio termodinâmico, apresenta grande relevância como discutido na seção 3.5, em que a utilização da estratégia descrita se faz uma novidade na comunidade científica.

3.7 Métodos para acessar estados metaestáveis

Abaixo serão descritos os métodos que podem ser utilizados para identificação e caracterização de estados metaestáveis, assim como posteriores análises.

3.7.1 Simulação de Dinâmica Molecular (*MD Simulation*)

O estudo de estruturas macromoleculares a nível atômico, como por exemplo o estudo de proteínas, se faz de grande interesse no descobrimento e estudo de novas drogas (HOSPITAL et al., 2015; HOLLINGSWORTH; DROR, 2018). Dessa maneira, a simulação de dinâmica molecular (DM) é uma importante ferramenta para contornar este problema, devido à sua alta precisão e bom custo benefício ao analisar fenômenos moleculares (SRIVASTAVA et al., 2021). Através da simulação de dinâmica molecular é possível investigar, em nível atômico, como um sistema molecular complexo se comporta, contornando assim as limitações tecnológicas vigentes para este tipo de estudo (HOLLINGSWORTH; DROR, 2018; DURRANT; MCCAMMON, 2011).

As simulações de DM tem a capacidade de prever como cada átomo de um sistema molecular irá se mover ao longo do tempo, em que para isso utiliza-se um modelo físico clássico para delinear os potenciais interatômicos (HOLLINGSWORTH; DROR, 2018; KARPLUS; MCCAMMON, 2002). Nesse cenário, como as funções moleculares e as interações intermoleculares dependem da dinâmica do sistema, as simulações de DM são uma interessante ferramenta para compreender fenômenos a nível atômico (HOLLINGSWORTH; DROR, 2018).

Para tal predição, são utilizados os chamados campos de força, que nada mais são que a descrição de uma série de potenciais clássicos do sistema, cujos parâmetros são provenientes de cálculos quânticos e/ou dados experimentais (HOLLINGSWORTH; DROR, 2018; SRIVASTAVA et al., 2021; ZHANG et al., 2022). O campo de força é uma série de parâmetros

necessários para descrever as interações em um sistema molecular, os quais são descritos por funções de energia potencial das contribuições estruturais, sendo elas: comprimento de ligação (V_{bonds}), ângulos de ligação (V_{angles}), ângulos diedros ($V_{\text{dihedrals}}$) e interações entre átomos não ligados (V_{Coulomb} e $V_{\text{Lennard-Jones}}$) (MARTÍNEZ et al., 2007). Assim sendo, a Equação 3.1 mostra a expressão implícita de um campo de força genérico, levando em consideração as contribuições mencionadas acima

$$V_{\text{total}} = V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{Coulomb}} + V_{\text{Lennard-Jones}} \quad (3.1)$$

em que os termos ligados, sendo eles V_{bonds} , V_{angles} e $V_{\text{dihedrals}}$, descrevem a contribuição do estiramento de ligação, estiramento de ângulo de ligação, e potencial do diedro, respectivamente. Já os termos não ligados são divididos em potencial eletrostático intra e intermolecular e interações de Van der Waals, V_{Coulomb} e $V_{\text{Lennard-Jones}}$ respectivamente (KOSTAL, 2016). Dessa maneira, a Equação 3.1 pode ser reescrita explicitamente como a Equação 3.2

$$V_{\text{total}} = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi[\cos(n\phi - \sigma) + 1] + \sum_{\text{Coulomb}} \left[\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] + \sum_{\text{Lennard-Jones}} 4\epsilon_{ij} \left[\frac{\sigma^{12}}{r_{ij}^{12}} - \frac{\sigma^6}{r_{ij}^6} \right] \quad (3.2)$$

sendo k_b , k_θ e k_ϕ as constantes de força, em que b e θ correspondem ao comprimento e ângulo de ligação, respectivamente, e b_0 e θ_0 os valores de equilíbrio. Já n é a periodicidade, ϕ o ângulo diedro, σ o ângulo de fase, r_{ij} a distância entre os átomos i e j , q_i e q_j as cargas pontuais de cada átomo, ϵ_0 a permissividade elétrica do vácuo, ϵ a profundidade do poço potencial e σ a distância a qual o potencial de Lennard-Jones é 0. Dessa maneira, dado a quantidade de parâmetros a serem descritos, existe uma dificuldade inerente em modelar um sistema molecular a partir de um campo de força (MARTÍNEZ et al., 2007).

Entretanto, através dos potenciais descritos, dado as coordenadas iniciais do sistema, é possível calcular a força exercida em cada átomo devido a presença dos outros átomos (HOLLINGSWORTH; DROR, 2018). Com isso, as equações do movimento de Newton podem ser utilizadas para prever a posição espacial de cada átomo ao longo do tempo (HOLLINGSWORTH;

DROR, 2018; DURRANT; MCCAMMON, 2011). Dessa forma, as Equações 3.3 e 3.4 são resolvidas simultaneamente, fornecendo um conjunto de equações para cada partícula do sistema, o que implica a necessidade de resolução por meio métodos numéricos de integração (ALAVI, 2020; MARTÍNEZ et al., 2007; PEREIRA, 2020).

$$\vec{F} = m_i \vec{a}_i(t) \quad (3.3)$$

$$\vec{F} = - \frac{\partial V_{total}(r_1, r_2, \dots, r_n)}{\partial r_1} \quad (3.4)$$

Nas equações apresentadas logo acima, \vec{F} é a força agindo na partícula i , m_i a massa dessa partícula, \vec{a}_i a sua aceleração e V_{total} o potencial total dependente a posição da partícula, o qual é representado por r_i . Dessa forma, ao calcular as forças e velocidade dos átomos, repetidamente ao longo do tempo, as simulações de DM são capazes de fornecer uma trajetória tridimensional do sistema molecular, em que a mesma funciona como um filme, descrevendo a configuração atômica do sistema em cada ponto ao longo do tempo de simulação (HOLLINGSWORTH; DROR, 2018; DURRANT; MCCAMMON, 2011; HOSPITAL et al., 2015).

Portanto, simulações de DM vem sendo utilizadas, por exemplo, para o desenvolvimento de potenciais sondas espectroscópicas para o diagnóstico de doenças, como mostrado nos trabalhos publicados por Santos et al. (2023a) e Gonçalves et al. (2019). Além disso, as simulações de DM também podem ser utilizadas no estudo de sondas fluorescentes, também para o diagnóstico de doenças a partir da interação de pequenas moléculas orgânicas com alvos biológicos, motivação essa que levou ao desenvolvimento do estudo publicado por Mancini et al. (2015). Dessa forma, o uso de simulações de DM se mostra como uma excelente estratégia no desenvolvimento de sondas espectroscópicas, uma vez que, ao se comparar simulações feitas em diferentes condições, é possível avaliar os efeitos, a nível atômico, das mais diversas perturbações (HOLLINGSWORTH; DROR, 2018).

3.7.2 Simulação de Dinâmica Molecular enviesada (*biased MD Simulation*)

Como descrito em seções anteriores, identificar e caracterizar estados metaestáveis de moléculas orgânicas pequenas se mostra uma boa estratégia para compreender efeitos conformacionais de não equilíbrio em mecanismos como o ESIPT, fornecendo assim informações de como a qualidade de um sensor fluorescente é afetada. Além disso, foi também discutido como

as simulações de DM podem ser úteis no estudo e desenvolvimento de novas sondas espectroscópicas. Entretanto, não foi discutido como, a partir desta técnica, é possível identificar estados metaestáveis para que posteriormente os mesmos possam ser caracterizados e interpretados. Dessa maneira, esta seção se propõe a isso.

Enquanto o espaço configuracional de um sistema físico é extremamente grande, o que ocorre, por muita das vezes, é que apenas uma fração desse espaço é visitada com uma probabilidade razoável (NOVELLI et al., 2022). Isso muitas vezes leva a uma distribuição multimodal de Boltzmann, o que acaba por refletir a presença de estados metaestáveis com longa duração (NOVELLI et al., 2022). Exemplos disso são os estados iniciais e finais de um processo químico, dobra e desdobra de proteínas ou até mesmo estados físicos de um material (NOVELLI et al., 2022; GHOSH; RANJAN, 2020). Dito isso, é importante ressaltar que a transição entre esses estados é rara, e ocorrem em uma escala de tempo muito maior do que outros processos característicos de um sistema molecular (NOVELLI et al., 2022).

Atualmente, a possibilidade de se fazer longas e acuradas simulações se deve muito ao fato de que houve um avanço significativo em algoritmos e *hardwares* (NOVELLI et al., 2022). Com isso, simulações mais longas puderam ser realizadas, abrangendo inclusive escalas de tempo onde transições entre metaestados podem ser acessadas (NOVELLI et al., 2022; JAIN; STOCK, 2012). Entretanto, performar longas simulações de sistemas complexos pode ser computacionalmente caro, além de que uma grande quantidade de dados é gerado, em que analisar esse grande volume de dados não é nada trivial, sendo interessante a busca por novas alternativas (NOVELLI et al., 2022).

Nesse cenário, o estudo feito por Ivernizzi e Parrinello (2020) desenvolveu uma técnica de amostragem, esta chamada *on-the-fly probability enhanced sampling* (OPES) expandido, a fim de observar esses metaestados em simulações de MD, e por fim, acessar a barreira de energia livre que os separam. Em resumo, a técnica criada pelos autores segue da seguinte forma: um conjunto expandido é obtido somando um conjunto de conjuntos em condições termodinâmicas ligeiramente diferentes, ou com hamiltonianos ligeiramente diferentes. Dessa maneira, tais conjuntos podem ser amostrados, permitindo então que estados metaestáveis, os quais provavelmente não seriam acessados em simulações de MD curtas, possam ser observados. Entretanto, a técnica não é nada trivial, fazendo com que tal sofisticação necessite ser melhor explorada.

Quanto aos princípios fundamentais da amostragem aprimorada, estes seguem basicamente duas estratégias (IVERNIZZI et al., 2020). A primeira seria a chamada *umbrella sam-*

pling, as quais foram primeiramente apresentadas por Torrie e Valleau (TORRIE; VALLEAU, 1974; TORRIE; VALLEAU, 1977; IVERNIZZI et al., 2020). Esta estratégia se inicia ao se identificar parâmetros de ordem, ou variáveis coletivas (CVs), $\mathbf{s} = \mathbf{s}(\mathbf{x})$, as quais são função da configuração microscópica do sistema físico, e são responsáveis por codificar os modos lentos do sistema, ou seja, descrever estados raros do sistema. Assim sendo, uma vez obtido os descritores capazes de descrever os modos mais lentos do sistema, um potencial enviesado, esta função das CVs, é aplicado ao potencial total do sistema, sendo assim capaz de acelerar a amostragem desses eventos raros.

A respeito da segunda estratégia, esta podendo ser utilizada conjuntamente com a primeira, consiste no chamado *parallel tempering* (SWENDSEN; WANG, 1986; MARINARI; PARRISI, 1992; IVERNIZZI et al., 2020). A estratégia consiste em combinar, no mesmo conjunto generalizado, configurações do sistema as quais foram acessadas em diferentes temperaturas (IVERNIZZI et al., 2020). Sua utilização pode aprimorar a amostragem pois, em temperaturas mais altas é mais provável que os modos lentos do sistema sejam acessados, uma vez que transições raras são aceleradas e é menos provável que o sistema se encontre preso em um ou outro estado metaestável (IVERNIZZI et al., 2020).

Explicitado as duas principais estratégias para aprimorar a amostragem de modos lentos de um sistema, faz-se necessário ressaltar que estas técnicas foram vistas na literatura como distintas e complementares (IVERNIZZI et al., 2020). Entretanto, apesar de em alguns estudos estas técnicas serem utilizadas de forma combinada, elas tipicamente foram vistas como abordagens híbridas (IVERNIZZI et al., 2020). Dessa maneira, o trabalho desenvolvido por Ivernizzi e Parrinello (2020), em que a técnica OPES expandido foi apresentada, visou unificar as duas abordagens e evidenciar como o seu uso conjunto amplia a amostragem de eventos raros. Não apenas isso, a técnica OPES é tida como uma evolução da *metadynamics*, método desenvolvido por Laio e Parrinello (2002) baseado na estratégia de *parallel tempering*.

O método *on-the-fly probability enhanced sampling* (OPES) expandido é um método baseado em variáveis coletivas (CVs), em que estas são funções da configuração microscópica do sistema, $\mathbf{s} = \mathbf{s}(\mathbf{x})$, a qual é capaz de descrever, em menores dimensões, um sistema físico (IVERNIZZI et al., 2020). Assim sendo, o OPES expandido foca na reconstrução da distribuição de probabilidade física de \mathbf{s} , $P(\mathbf{s})$ (IVERNIZZI et al., 2020; IVERNIZZI; PARRINELLO, 2020). A partir desta reconstrução é possível então atingir uma distribuição de probabilidade alvo, $p^{\text{tg}}(\mathbf{s})$ (IVERNIZZI et al., 2020; IVERNIZZI; PARRINELLO, 2020; RIZZI et al., 2023).

De forma a atingir este objetivo, um potencial enviesado deve ser adicionado ao potencial do sistema, em que este é definido abaixo pela Equação 3.5 como

$$V(\mathbf{s}) = -\frac{1}{\beta} \log \frac{P^{tg}(\mathbf{s})}{P(\mathbf{s})} \quad (3.5)$$

em que β é a temperatura inversa, esta advinda da estatística termodinâmica. Como o OPES expandido é tido como uma evolução do *metadynamics*, o método a princípio utilizou-se da distribuição alvo *well-tempered*, definida como $p^{WT}(\mathbf{s}) \propto [P(\mathbf{s})]^{1/\gamma}$, em que $\gamma > 1$ é conhecido como o *bias factor*, uma espécie de fator de enviesamento (IVERNIZZI et al., 2020). Dessa maneira, a distribuição alvo descrita tem como finalidade o aumento da taxa de transições entre metaestados do sistema, em que para isso, diminui em um fator γ a barreira de energia livre que separam as CVs em metaestados (IVERNIZZI et al., 2020). No limite $\lim_{\gamma \rightarrow \infty}$ tem-se o equivalente a escolher uma distribuição alvo uniforme (IVERNIZZI et al., 2020).

Entretanto, a princípio $P(\mathbf{s})$ não é conhecido, assim sendo, o método OPES expandido consiste de um esquema iterativo (IVERNIZZI et al., 2020; IVERNIZZI; PARRINELLO, 2020). A ideia por trás do método é então estimar a distribuição de probabilidade em cada etapa n , $P_n(\mathbf{s})$, reponderando de forma iterativa, conhecido também como um procedimento *on the fly*, uma simulação enviesada com $V_n(\mathbf{s})$, a qual é construída a partir da distribuição estimada $P_n(\mathbf{s})$ como mostra a Equação 3.5 (IVERNIZZI et al., 2020). A $P_n(\mathbf{s})$ é obtida através de um *kernel* estimador de densidade ponderado, adicionando um novo *kernel* em um pequeno intervalo fixado (IVERNIZZI et al., 2020).

Em termos mais simples, o método expressa a distribuição alvo como uma função da distribuição de probabilidade não enviesada $p^{WT}(\mathbf{s}) \propto [P(\mathbf{s})]^{1/\gamma}$, estima $P_n(\mathbf{s})$ via reponderação e, por fim, calcula o fator de enviesamento (IVERNIZZI; PARRINELLO, 2020). Para isso, a distribuição de probabilidade é estimada iterativamente, ou *on the fly*, em que periodicamente são adicionadas gaussianas (IVERNIZZI; PARRINELLO, 2020). Nesse sentido, cada nova gaussiana adicionada é ponderada de acordo com o potencial enviesado previamente utilizado, como mostra a Equação 3.6

$$\tilde{P}_n(\mathbf{s}) = \frac{\sum_k^n w_k G(\mathbf{s}, \mathbf{s}_k)}{\sum_k^n w_k} \quad (3.6)$$

em que os pesos w_k são dados por $w_k = e^{\beta V_{k-1}(\mathbf{s}_k)}$ e $G(\mathbf{s}, \mathbf{s}_k)$ são as gaussianas adicionadas (IVERNIZZI; PARRINELLO, 2020). Ao final, a expectativa é que o fator de enviesamento atinja a convergência (IVERNIZZI; PARRINELLO, 2020).

Acima foi descrito como opera o método OPES, entretanto, no trabalho publicado por Ivernizzi et al. (2020) os autores propõem outro avanço com relação a métodos de aprimoração de amostragem, como o *metadynamics*. No trabalho mencionado, diferentemente do feito para o *metadynamics*, que consiste em combinar, no mesmo conjunto generalizado, configurações do sistema as quais foram acessadas em diferentes temperaturas (IVERNIZZI et al., 2020), os autores propõem expandir os conjuntos, ou seja, amostrar múltiplos conjuntos termodinâmicos em uma mesma simulação.

Dessa forma, primeiramente assume-se $u(\mathbf{x})$ como sendo o potencial reduzido adimensional que contém todos os termos dependentes das restrições termodinâmicas, tais como temperatura, pressão, ou outros (IVERNIZZI et al., 2020). Vale ressaltar que \mathbf{x} indica as coordenadas atômicas e qualquer outra variável configuracional a qual o potencial reduzido pode vir a depender (IVERNIZZI et al., 2020). Como exemplo, no caso do conjunto canônico, tem-se $u(\mathbf{x}) = \beta U(\mathbf{x})$, em que β é a temperatura termodinâmica inversa e $U(\mathbf{x})$ é a energia potencial do sistema (IVERNIZZI et al., 2020). Assim sendo, considerando um sistema com potencial reduzido $u_\lambda(\mathbf{x})$, a qual é função de λ , em que λ pode ser um único parâmetro ou um conjunto de parâmetros, e pode indicar um parâmetro termodinâmico, como temperatura, no equilíbrio sua distribuição de probabilidade segue a distribuição de Boltzmann, como mostra a Equação 3.7 (IVERNIZZI et al., 2020)

$$P_\lambda(\mathbf{x}) = \frac{e^{-u_\lambda(\mathbf{x})}}{Z_\lambda} \quad (3.7)$$

em que Z_λ é a função de partição, $Z_\lambda = \int e^{-u_\lambda(\mathbf{x})} d\mathbf{x}$.

Dessa maneira, o interesse do método OPES expandido consiste em amostrar configurações em um intervalo $\Delta\lambda$ de valores de λ . Portanto, ao invés de performar múltiplas simulações independentes em diferentes valores de λ , o método é capaz de amostrar um conjunto generalizado o qual contém todas as configurações microscópicas relevantes, e então reponderá-los de formar a retornar a estatística correta para qualquer $\lambda \in \Delta\lambda$ (IVERNIZZI et al., 2020).

Nesse cenário, uma distribuição alvo, a qual contém todas as configurações microscópicas relevantes para o intervalo $\Delta\lambda$ escolhido, deve ser escolhida (IVERNIZZI et al., 2020). Portanto, um conjunto λ de $N_{\{\lambda\}}$ valores $\lambda \in \Delta\lambda$, tal que haja uma boa sobreposição entre $P_\lambda(\mathbf{x})$

e distribuições adjacentes, é escolhido (IVERNIZZI et al., 2020). Portanto, a distribuição alvo é definida como segue na Equação 3.8

$$p_{\lambda}(\mathbf{x}) = \frac{1}{N_{\{\lambda\}}} \sum_{\lambda} P_{\lambda}(\mathbf{x}) \quad (3.8)$$

em que essa classe de distribuições de probabilidade alvo são usualmente chamadas de distribuições alvo de conjunto expandido (IVERNIZZI et al., 2020). É importante mencionar que para o método OPES expandido o mesmo peso $1/N_{\{\lambda\}}$ é dado para todos os subconjuntos. Seguindo adiante, sem perder a generalização do método, é possível assumir $\lambda = 0$ como pertencente ao intervalo $\Delta\lambda$, portanto, é possível performar uma simulação em $\lambda = 0$ e usar o esquema OPES para iterativamente otimizar um fator de enviesamento capaz de amostrar $p_{\{\lambda\}}(\mathbf{x})$ (IVERNIZZI et al., 2020). Entretanto, antes definir explicitamente a distribuição alvo e o potencial enviesado, define-se $P_{\lambda}(\mathbf{x})$ como expresso pela Equação 3.9

$$P_{\lambda}(\mathbf{x}) = P_0(\mathbf{x}) e^{-u_{\lambda}(\mathbf{x})+u_0(\mathbf{x})} \frac{Z_0}{Z_{\lambda}} = P_0(\mathbf{x}) e^{-u_{\lambda}(\mathbf{x})+\Delta F(\lambda)} \quad (3.9)$$

em que $\Delta u_{\lambda}(\mathbf{x}) = u_{\lambda}(\mathbf{x}) - u_0(\mathbf{x})$ é a diferença de energia potencial e $\Delta F(\lambda)$ é como expresso na Equação 3.10

$$\Delta F(\lambda) = -\log \frac{Z_{\lambda}}{Z_0} = -\log \langle e^{-\Delta u_{\lambda}} \rangle_{u_0} \quad (3.10)$$

sendo $\Delta F(\lambda)$ a diferença adimensional de energia livre com relação ao sistema de referência u_0 , a qual também pode ser expressa como a média de um conjunto (IVERNIZZI et al., 2020). Com isso, a Equação 3.11 mostra como fica, explicitamente, o alvo expandido

$$P_{\lambda}(\mathbf{x}) = P_0(\mathbf{x}) \frac{1}{N_{\{\lambda\}}} \sum_{\lambda} e^{-\Delta u_{\lambda}(\mathbf{x})+\Delta F_n(\lambda)} \quad (3.11)$$

Dessa maneira, de forma a obter o fator de enviesamento alvo, primeiro reescreve-se a Equação 3.5 como mostrado na Equação 3.12.

$$v(\mathbf{s}) = -\log \frac{p^{tg}(\mathbf{s})}{P(\mathbf{s})} \quad (3.12)$$

Por fim, o *bias* adimensional necessário para amostrar o alvo expandido $p_{\{\lambda\}}(\mathbf{x})$ é expresso na Equação 3.13

$$v(\mathbf{x}) = -\log \left(\frac{1}{N_{\{\lambda\}}} \sum_{\lambda} e^{-\Delta u_{\lambda}(\mathbf{x}) + \Delta F(\lambda)} \right) \quad (3.13)$$

em que o valor do potencial enviesado $v(\mathbf{x})$ depende das coordenadas \mathbf{x} , apenas através das $N_{\{\lambda\}}$ quantidades de Δu_{λ} , sendo essas as variáveis coletivas de expansão, em que $\Delta u_{\lambda}(\mathbf{x}) = u_{\lambda}(\mathbf{x}) - u_0(\mathbf{x})$ representa a diferença de energia potencial, e $\Delta F_n(\lambda)$ a diferença adimensional de energia livre em relação ao sistema de referência u_0 (IVERNIZZI et al., 2020). Nesse sentido, as ECVs caracterizam completamente a expansão, uma vez que não apenas o *bias* ($v(\mathbf{x})$), mas também $\Delta F(\lambda)$ são definidos a partir dessas variáveis, em que $\Delta F(\lambda) = -\log \langle e^{-\Delta u_{\lambda}} \rangle_{u_0}$ (IVERNIZZI et al., 2020).

É possível observar na Equação 3.13 a dependência de $v(\mathbf{x})$ da $\Delta F(\lambda)$, este definido como $\Delta F(\lambda) = -\log \langle e^{-\Delta u_{\lambda}} \rangle_{u_0}$. Assim sendo, é praticamente impossível estimar $\langle e^{-\Delta u_{\lambda}} \rangle_{u_0}$ a partir de uma trajetória não enviesada, isso pois $e^{-\Delta u_{\lambda}}$ pode ser significativamente grande em regiões onde P_0 é extremamente pequeno, estando assim inapropriadamente amostradas (IVERNIZZI et al., 2020). Dessa maneira, é utilizado assim um procedimento de reponderação para escrever este valor como uma média sobre o conjunto enviesado, como mostra a Equação 3.14

$$e^{-\Delta F(\lambda)} = \langle e^{-\Delta u_{\lambda}} \rangle_{u_0} = \frac{\langle e^{-\Delta u_{\lambda} + v} \rangle_{u_0 + v}}{\langle e^v \rangle_{u_0 + v}} \quad (3.14)$$

em que a média do conjunto $\langle \cdot \rangle_{u_0 + v}$ é calculada como um tempo médio ao longo de uma trajetória enviesada (IVERNIZZI et al., 2020). Entretanto, como visto na equação 3.13, o *bias* alvo é ele próprio uma função de $\Delta F(\lambda)$, assim sendo um conjunto iterativo de equações é montado para resolver tal problemática. O esquema iterativo consiste em realizar uma simulação enviesada cujo *bias* é baseado na estimativa de $\Delta F(\lambda)$ obtido pela Equação 3.14 (IVERNIZZI et al., 2020). Na etapa n , a simulação segue com potencial $u_0(\mathbf{x}) + v_n(\mathbf{x})$, em que

$$v_n(\mathbf{x}) = -\log \left(\frac{1}{N_{\{\lambda\}}} \sum_{\lambda} e^{-\Delta u_{\lambda}(\mathbf{x}) + \Delta F_n(\lambda)} \right) \quad (3.15)$$

A estimativa reponderada de $\Delta F_n(\lambda)$ é atualizada em cada etapa n , em que entre essas etapas existe um passo curto fixo em que a simulação segue com $\Delta F_n(\lambda)$ e $v_n(\mathbf{x})$ constantes (IVERNIZZI et al., 2020). A estimativa da diferença de energia livre na n -ésima etapa pode ser escrita, explicitamente, como sendo

$$\Delta F_n(\lambda) = -\log \left(\frac{\sum_{k=1}^n e^{-\Delta u_\lambda^{(k)} + v_{k-1}^{(k)}}}{\sum_{k=1}^n e^{v_{k-1}^{(k)}}} \right) \quad (3.16)$$

em que é utilizado as notações $\Delta u_\lambda^{(k)} = \Delta u_\lambda(\mathbf{x}_k)$ e $v_n^{(k)} = v_n(\mathbf{x}_k)$, com \mathbf{x}_k sendo a configuração da k -ésima etapa iterativa. A medida que o *bias* se aproxima da convergência, o conjunto amostrado se aproxima do conjunto alvo, e os valores de $\Delta F_n(\lambda)$ se tornam cada vez mais acurados (IVERNIZZI et al., 2020). Dessa maneira, as Equações 3.15 e 3.16 são as equações iterativas explícitas utilizadas pelo método OPES para amostragem de conjuntos expandidos definidos pela distribuição alvo $p_{\{\lambda\}}(\mathbf{x})$ (IVERNIZZI et al., 2020).

Portanto, a partir desta técnica é possível não somente observar estados metaestáveis em simulações de menor duração, mas também quantificar a barreira energética que separa esses estados. Dessa forma, a partir da técnica de amostragem, obtendo-se assim simulações enviesadas, é possível aplicar métodos estatísticos para identificar, caracterizar e interpretar esses estados metaestáveis (NOVELLI et al., 2022).

3.7.3 PCA na identificação e interpretação de estados metaestáveis

Após obtidos os dados provenientes de uma simulação de DM enviesada, utilizando técnicas de amostragem, o principal objetivo é identificar e interpretar os estados metaestáveis. Para isso, técnicas de aprendizado de máquina são de suma importância, permitindo maior rapidez e facilidade na análise (NOVELLI et al., 2022). Alguns trabalhos vem concentrando esforços em propor métodos, utilizando aprendizado de máquina, para atingir este objetivo, cobijado por diversos pesquisadores (NOVELLI et al., 2022; JAIN; STOCK, 2012; SITTEL; STOCK, 2018).

Em outros estudos, é mostrado também como a Análise de Componentes Principais (PCA) pode ser utilizada para contribuir nesta análise (JAIN; STOCK, 2012; SITTEL; STOCK, 2018). A PCA é uma popular ferramenta exploratória, ou seja, um método estatístico não supervisionado, o qual pode ser utilizado para reduzir a dimensionalidade dos dados, aumentando a interpretabilidade e ao mesmo tempo minimizando a perda de informação (JOLLIFFE; CADIMA, 2016; JAIN; STOCK, 2012; SITTEL; STOCK, 2018). Dessa maneira, a PCA consiste em uma transformação linear a qual é capaz de extrair os elementos mais importantes do conjunto de dados, em que para isso utiliza-se a matriz de covariância ou a matriz de correlação (BALSERA et al., 1996; DAVID; JACOBS, 2014).

A PCA faz isso a partir da criação de novas variáveis não correlacionadas, as quais maximizam a variância contida na matriz de dados originais (JOLLIFFE; CADIMA, 2016). Ao encontrar essas variáveis, as componentes principais, a solução consiste na resolução de um problema de autovalor/autovetor, e as novas variáveis são definidas pelo conjunto de dados (JOLLIFFE; CADIMA, 2016). Dessa maneira, a PCA funciona de forma que, a primeira componente principal (PC1) representa a direção com maior variância na matriz de dados (LI et al., 2011). Já a segunda componente principal (PC2), representa a direção que contém a segunda maior variação entre os dados, sendo ortogonal a anterior, em que as próximas PCs seguem a mesma lógica (LI et al., 2011).

Nesse cenário, a PCA começa com o conjunto de dados de geometria, ou matriz de geometria $\mathbf{G}_{n \times m}$, em que n representa o número de amostras, e m a quantidade de variáveis descritoras (DU; ZHU, 2019; LI et al., 2011). Assim sendo, a matriz de geometria nada mais seria do que a matriz original de dados. Com isso, a PCA pode ser computada ao calcular os autovalores e autovetores da matriz de covariância da geometria centrada \mathbf{G} (LI et al., 2011). Dessa maneira, como mostrado por Li et al. (2011), \mathbf{G} pode ser decomposta de acordo com as Equações 3.17 e 3.18

$$\mathbf{G} = \mathbf{S}\mathbf{P} \quad (3.17)$$

$$\mathbf{S} = \mathbf{G}\mathbf{P}^T \quad (3.18)$$

em que \mathbf{S} é uma matriz $n \times l$ chamada de *scores* da PCA e \mathbf{P} são os autovetores de \mathbf{G} , que é uma matriz normalizada $l \times l$. É notório que, como a matriz de *scores* \mathbf{S} tem dimensões $n \times l$, isso implica que para cada amostra n haverá um valor de *scores* associado. Como mencionado anteriormente a respeito da distribuição da variância explicada pelas PCs calculadas, a matriz \mathbf{G} pode ser aproximada, para contendo mais de 90% da variância cumulativa, ao se pegar as primeiras k PCs calculadas, como mostra a Equação 3.19 (LI et al., 2011)

$$\mathbf{G}^* = \mathbf{S}_k\mathbf{P}_k \quad (3.19)$$

em que \mathbf{G}^* é a aproximação de \mathbf{G} , \mathbf{S}_k é a primeira coluna k de \mathbf{S} , e \mathbf{P}_k são as primeiras linhas de \mathbf{P} .

Dessa forma, a utilização da PCA também pode ser de grande interesse na busca por observar e interpretar estados metaestáveis, sendo estes descritos por variações estruturais. Assim sendo, o estudo publicado por Buslaev et al. (2016) mostrou, a partir da investigação das diferenças estruturais em uma molécula de lipídio, e utilizando-se da PCA e de dados provenientes de simulações de DM, como esta ferramenta se mostra uma boa estratégia para investigar diferenças conformacionais de moléculas. Além disso, o artigo publicado por Palma e Pierdominici-Sottile (2023) faz uma revisão extensa acerca da utilização da PCA, em que a finalidade da utilização do método seria a de caracterizar macromoléculas biológicas a partir de dados provenientes de simulações de DM. Portanto, a utilização da PCA para caracterizar sistemas moleculares a partir de dados de simulações de DM se faz recorrente na literatura, uma vez que de forma prática e rápida é capaz de fornecer interpretabilidade para sistemas moleculares.

3.8 Estudando o ESIPT

O uso de simulações de DM e métodos estatísticos são de suma importância para identificação e interpretação de estados metaestáveis. Entretanto, dado a maior acurácia dos cálculos quânticos, sua utilização é interessante em um estudo apropriado de reações químicas. Apesar disso, o custo computacional de cálculos quânticos é relativamente mais alto, fazendo necessário a seleção de algumas poucas geometrias para estudo. Dessa maneira, como visto na seção anterior, a PCA pode ser uma boa ferramenta para esta tarefa, permitindo a exploração e a seleção de geometrias a partir da distribuição gerada no método. Dito isso, esta seção se dedica a descrever o próximo passo, o qual consiste na utilização de cálculos quânticos para o estudo de reações químicas como o ESIPT.

Nesse sentido, a química computacional é uma importante ferramenta para compreender fenômenos que, por muita das vezes, são extremamente difíceis de serem investigados experimentalmente (MA, 2020). O desenvolvimento da teoria e de métodos computacionais fazem com que, atualmente, a química deixe de ser apenas uma ciência empírica, se tornando também uma ciência teórica rigorosa (MCARDLE et al., 2020). A princípio, qualquer sistema químico pode ser estudado teoricamente através de métodos teóricos computacionais, entretanto, devido ao fato de que a complexidade desses métodos aumenta rapidamente com o aumento do número de elétrons do sistema, cálculos acurados estão usualmente restritos ao estudo de pequenas moléculas (MCARDLE et al., 2020).

Nesse sentido, a química computacional é construída baseada nas teorias e descobertas da mecânica quântica (MCARDLE et al., 2020). Dessa forma, o método de Hartree-Fock (HF) é o método mais básico de cálculo *ab initio* baseado em funções de onda, sendo o precursor de métodos mais modernos e acurados, chamados *post* Hartree-Fock (MCARDLE et al., 2020; MA, 2020). No método HF, uma aproximação de primeira ordem foi proposta assumindo que elétrons não interagem uns com os outros, resolvendo assim o problema apresentado pela equação de Schrödinger para sistemas multieletrônicos, separando o hamiltoniano em operadores de um elétron, como mostra a Equação 3.20 (MCARDLE et al., 2020; MA, 2020).

$$\hat{H} = \sum_{i=1}^N h_i \quad (3.20)$$

Entretanto, em um sistema real a correlação eletrônica é significativa, assim sendo, como a aproximação de HF não leva isso em conta, métodos posteriores, os *post* Hartree-Fock, foram criados para resolver essa problemática (MCARDLE et al., 2020; MA, 2020). Estes métodos começam com cálculos HF, aplicando posteriormente uma correção para contabilizar a correlação eletrônica (MCARDLE et al., 2020). Apesar disso, métodos baseados na função de onda apresentam um problema inerente: o custo computacional à medida que o sistema cresce (aumenta o número de elétrons) aumenta exponencialmente (MCARDLE et al., 2020).

Dessa forma, o desenvolvimento de cálculos baseado na função de densidade eletrônica foi um grande salto no aprimoramento de cálculos *ab initio*, em que esses métodos estão atrelados a Teoria do Funcional de Densidade (DFT) (MCARDLE et al., 2020). A DFT é baseada nos teoremas de Hohenberg-Kohn, os quais descrevem que: a densidade eletrônica do estado fundamental descreve todas as propriedades do estado fundamental, incluindo energia e função de onda, e a função de densidade eletrônica que minimiza a energia do sistema é a função de densidade eletrônica real do sistema (MCARDLE et al., 2020). Portanto, o funcional descrito por Hohenberg-Kohn pode ser expresso na forma da função de onda de um único elétron, como mostra a Equação 3.21 (MCARDLE et al., 2020).

$$E[\{\Psi_i\}] = E_{\text{known}}[\{\Psi_i\}] + E_{xc}[\{\Psi_i\}] \quad (3.21)$$

Em que E_{xc} é o funcional de correlação e troca, o qual deve ser aproximado, e a parte conhecida da expressão (E_{known}), é descrita por potenciais que podem ser calculados a partir da função de densidade eletrônica (MCARDLE et al., 2020). Dessa forma, o uso de cálculos base-

ado na densidade eletrônica do sistema pode ser bastante útil para compreender os mais diversos sistemas químicos, com um custo computacional apreciável. Nesse cenário, sua utilização para o estudo dos mais diversos sistemas se mostra bastante atrativo.

Um exemplo consiste no estudo de propriedades ópticas de sistemas químicos, este crucial em diversos campos, como por exemplo, para o estudo de sensores fluorescentes através do mecanismo ESIPT (MOURIK et al., 2011; ROOHI; POURYAHYA, 2023). Extensões da DFT convencional foram desenvolvidas de forma a acessar as propriedades do estado excitado, como por exemplo, a TDDFT (MOURIK et al., 2011). Assim sendo, o estudo desenvolvido por Roohi e Pouryahya (2023) investigou o mecanismo ESIPT performado por corantes cumarina-benzotiazol para potencial aplicação em sistemas de entrega de drogas. Portanto, com o uso da DFT, o estudo do mecanismo ESIPT pode ser realizado de forma acurada e com custo computacional razoável. Não apenas isso, a utilização deste método na investigação dos efeitos conformacionais da molécula HABT na sua capacidade de performar o processo ESIPT se mostra uma interessante estratégia.

4 ARTICLE

Molecular Dynamics-assisted interaction between HABT and PI3K enzyme: Metastable States to biological application for Cancer diagnosis

4.1 Introduction

In 1944, Schrödinger added a new perspective about how living systems work with his book *What is life? The Physical Aspect of the Living Cell* (SCHRÖDINGER, 1944). From its perspective, the living system stays alive because it is constantly struggling against its entropy increase (DEMIREL, 2014; GHASSEMI; SHAHIDIAN, 2017; ZHANG; OUYANG, 2021). However, once the inevitable entropy increase through the equilibrium is reached (ZHANG; OUYANG, 2021), it causes the death of the living system.

Schrödinger's approach was quite innovative in that moment, and the main reason is that because he considered living systems as open systems, which keeps many processes in a local nonequilibrium regime (ZHANG; OUYANG, 2021; ARISTOV, 2019). Therefore, from that scientific foundation, this perspective gained much attention and had a great impact on natural sciences (ZHANG; OUYANG, 2021).

Local nonequilibrium states have recently become a valuable tool for modeling biological systems (ARISTOV, 2019). From this scenario, it was possible to investigate the fold and unfold events of proteins, through membrane transport, cell signaling, and many other fundamental processes that occur in living systems (EBELING; SOKOLOV, 2005; GHASSEMI; SHAHIDIAN, 2017; HERNÁNDEZ-LEMUS, 2012). In addition to that, the local nonequilibrium approach was used in many works to investigate ligands interaction with target enzymes, a crucial step in drug design and optimization (CHAN et al., 2023; WOLF et al., 2019; MACCHIAGODENA et al., 2021).

Furthermore, one of the primary objectives of using a nonequilibrium perspective was to unravel the genetic and epigenetic causes of tumor development, which can lead to cancer (FANG et al., 2019; WELTER; RIEGER, 2013). It is well known that cancer is one of the leading causes of death worldwide, being estimated that 1 in 5 people will develop cancer during its lifetime (WORLD HEALTH ORGANIZATION (WHO), 2024; U. S. NATIONAL CANCER INSTITUTE, 2021). Therefore, efforts to understand and combat this illness is of great value.

In this scenario, it is known that cancer treatments are not efficient enough for giving full protection against this disease, being also very expensive, limiting its access (SERRANO et al., 2019; CHAKRABORTY; RAHMAN, 2012). In addition to that, developing new treatments is of great complexity, facing strong knowledge barriers (KAUR et al., 2023; STROBL et al., 2023). An alternative strategy is to develop an early and effective diagnosis. In fact, an early diagnosis considerably improves the survival rate of the patients affected by this disease, allowing intervention in an early stage, preventing development and lethality of cancer (CROSBY et al., 2022).

That said, in the attempt of developing new compounds for the cancer diagnosis, previous works shed light on a very important subject for drug development: the drug-target interaction. It was verified that a great amount of drug-target interactions of small molecules clinically approved are of a nonequilibrium nature (SRINIVASAN, 2023; COPELAND, 2013; COPELAND, 2016; COPELAND, 2021). Thus, bringing this knowledge to an early diagnosis improvement may be of great value, adding the needed complexity when facing this challenge.

Therefore, aiming for a better, biocompatible, and early diagnosis of cancer, fluorescent spectroscopic probes may be a great choice for this job (HAN et al., 2022; KIM et al., 2020; GEORGIEV et al., 2023; ZHOU et al., 2020). When compared to other spectroscopic probes, the so called fluorescent sensors present high selectivity, low cost, and better biocompatibility (GEORGIEV et al., 2023; ZHOU et al., 2020). For this, many photophysical phenomena can explain its action, like the Excited State Intramolecular Proton Transfer (ESIPT) (GEORGIEV et al., 2023; SEDGWICK et al., 2018).

Molecules capable of performing ESIPT, such as phenylbenzothiazoles (PBTs), can emit light at two different wavelengths (SEDGWICK et al., 2018). In some cases, dual emission fluorescence can occur when the ESIPT process takes place (GRANDJEAN et al., 2021). It is important to note that this proton transfer is often referred to as ultrafast, as it occurs on a femtosecond (10^{15} s) scale (GRANDJEAN et al., 2021; LIU et al., 2021). Therefore, considering the nonequilibrium nature of biological systems and the far-from-equilibrium time scale of the ESIPT process, it is crucial to evaluate the effects of metastable states of the sensor in the ESIPT process when aiming to develop a better fluorescent sensor for cancer diagnosis.

In that regard, fluorescent sensors capable of performing ESIPT could be used for signaling enzyme targets related to cancer development. In this sense, the Phosphoinositide 3-kinase (PI3K) protein may be a great target, once it is widely associated with the progression of the

mentioned illness (PAUL; MUKHOPADHYAY, 2004; JIANG; JI, 2019; LIU et al., 2009; JHA et al., 2015; YU; KOH, 2017; RAZANI et al., 2021). Now, considering a sensor for this target, the 2-(2'-hydroxy-4'-aminophenyl)benzothiazole (HABT) can be an appropriate probe, once it is capable of performing ESIPT, and there are evidences of strong interaction between the HABT molecule and the PI3K protein (ANDOLPHO et al., 2022; MANCINI et al., 2014; MAVROIDI et al., 2016; MACHURA et al., 2013). Therefore, evaluating HABT in signaling the PI3K target, adding the complexity of a local nonequilibrium regime, can be a powerful tool when aiming to improve early diagnosis of cancer.

However, from a methodological point of view, the non-triviality of this task considerably hampers such a goal. Despite that, the work published by Ivernizzi e Parrinello (2020) provided a fresh new methodology for accessing metastable states of molecular systems, where this could be done by enhancing the sampling of rare events in MD simulations. The proposed technique, which is a collective variable based method (IVERNIZZI et al., 2020), was called on-the-fly probability enhanced sampling (OPES). As shown by (ANSARI et al., 2022), the OPES method was successfully used in a complex biological system, being able to enable the evaluation of benzamidine residence time in trypsin enzyme.

Therefore, from a fresh new perspective and methodology, the present work's main goal is to evaluate how the metastable states of HABT affects the signaling of the PI3K protein target.

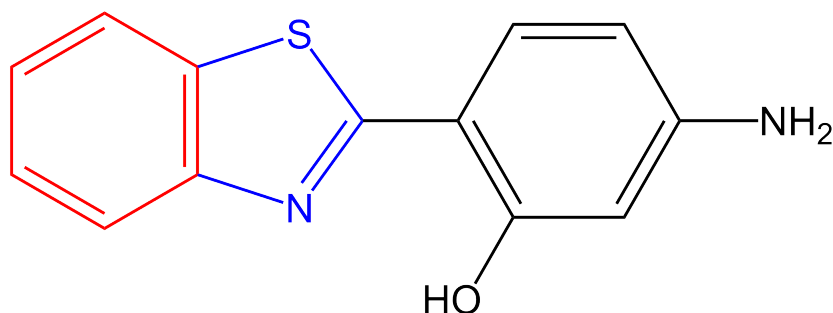
4.2 Methodology

4.2.1 HABT structure and parametrization

The studied molecule was the 2-(2'-hydroxy-4'-aminophenyl)benzothiazole (HABT), whose structure is shown above in Figure 4.1. In order to perform MD simulations, first was needed to construct the HABT molecule. For that, the software GaussView 6 (DENNINGTON et al., 2019) was used.

After that, it was necessary to optimize the structure and generate parameters for the force field. Thus, the Automated Force Field Topology Builder (ATB) 3.0, available at <<https://atb.uq.edu.au/>> (MALDE et al., 2011; STROET et al., 2018), was used for both tasks, where ATB used B3LYP functional and 6-311g basis set for such a calculation.

Figura 4.1 – HABT structure, a benzothiazole derivative, constituted from a benzene (red), thiazole (blue), and phenyl (black).



4.2.2 Unbiased Molecular Dynamics (MD) simulations

With the HABT structure and force field parameters in hand, unbiased MD simulations were carried out in order to obtain the needed data from trajectory files to perform further analysis. This occurs in a scenario where the main goal is to achieve physical interpretation, thus the unbiased simulation is of great value.

In this scenario, three MD simulations of 100 ns were carried out. Therefore, three different systems were studied, being: HABT in vacuum (1), HABT in water (2), and protein-docked HABT in water (3). The protein studied was the PI3K (PDB ID: 3QJZ) (D'ANGELO et al., 2011), where the docking position was provided by the work published by Santos et al. (2023b). In order to simplify the text, the systems were named as: HABT+VAC (1), HABT+WAT (2), and HABT+WAT+PROT (3).

The simulations were carried out using GROMACS 2021.4 (LINDAHL et al., 2021), where GROMOS 54a7 (SCHMID et al., 2011) force field was used to describe the system's potential. For water, the simple point-charge (SPC) model was used. The performed simulations consisted of the steps of minimization, equilibration and production.

4.2.3 Biased Molecular Dynamics (MD) simulations

From unbiased MD simulations, it is possible to obtain physical information about a certain studied molecular system. Now, from biased MD simulations, it is possible to enhance sampling of metastable states and characterize them. Therefore, three biased MD simulations of 100 ns were carried out, each one for each of the three studied systems. From the biased trajectory, it was possible not only to access different metastates and characterize them, but also to understand the effects of solvent and protein docking in their distribution and population. The

technique used to bias the simulation was the on-the-fly probability enhanced sampling (OPES) expanded, developed by (IVERNIZZI et al., 2020).

In order to perform the biased MD simulations, PLUMED 2 v2.8.0 software (TRIBELLO et al., 2014), patched with GROMACS 2021.4, was used. In order to expand the simulation to sample multiple temperatures simultaneously, a multi thermal expansion collective variable (ECV_MULTITHERMAL) was used, where a maximum temperature of 600K was set, with a pace of 500. With the obtained trajectories for all three systems, it was possible to identify and characterize the accessed metastates with Principal Component Analysis (PCA).

4.2.4 Principal Component Analysis (PCA) and conformational selection

With the calculated descriptors from biased MD trajectories, PCA analysis was carried out for all the three studied systems. From PCA analysis, it was possible to analyze how the accessed metastable states were distributed, and then characterize them in terms of the chosen descriptors. In order to observe their distribution, their PCA scores were observed, and to characterize them, it was observed the variable loadings. Thus, with a biplot graph from PCA analysis it was possible to see how conformations grouped in metastates and how the descriptors behaved in each metastable state.

Therefore, after identification and characterization of the accessed metastable states, representative conformations were selected from each metastate of system 3. For that, the mean value of the descriptors was calculated in each identified metastate, accessed from biased MD trajectories. After that, in the unbiased MD trajectories, the conformations whose descriptor values were closest to the calculated averages of its respective metastate, were selected. In this scenario, with the selected conformations, it was possible to move forward for QM calculations.

4.2.5 QM calculations setup

In order to study how metastable states affect the ESIPT process, QM calculations were carried out with the selected conformations. With the selected HABT structure from the unbiased trajectory, it was also collected residues to simulate the protein environment in QM calculations. The residues selection was based on an h-bond and RMSD analysis, conducted for system 3. The two residues which most interacted with HABT were collected. Thus, in the selected frame of the unbiased trajectory, it was collected the coordinates of HABT and of the two major interacting residues verified by h-bond and RMSD analysis.

In this scenario, DFT and TDDFT calculations were carried out using Gaussian 09 software (FRISCH et al., 2009), in order to obtain conformation energies. The calculations were performed in order to construct the following path: enol configuration at ground state (1), enol configuration at the first excited state (2), keto configuration at the first excited state (3), and keto configuration at ground state (4). From this path it was possible to analyze how the accessed metastates affected the ESIPT process.

For the calculations in the ground state, the B3LYP functional (LEE et al., 1988) was used with the tzvp basis set (SCHAFER et al., 1992). Now, for the excited states calculation, CAM-B3LYP functional (YANAI et al., 2004) was used with the tzvp basis set. All conformation calculations were carried out as a single point charge calculation, in order to evaluate the metastable conformational effect. Both B3LYP and CAM-B3LYP functionals are widely used to study benzothiazole derivatives, where the tzvp basis set follows the same (MANCINI et al., 2015; MANOJAI et al., 2017; WANG et al., 2024).

4.3 Results and discussions

4.3.1 Unbiased MD simulation analysis

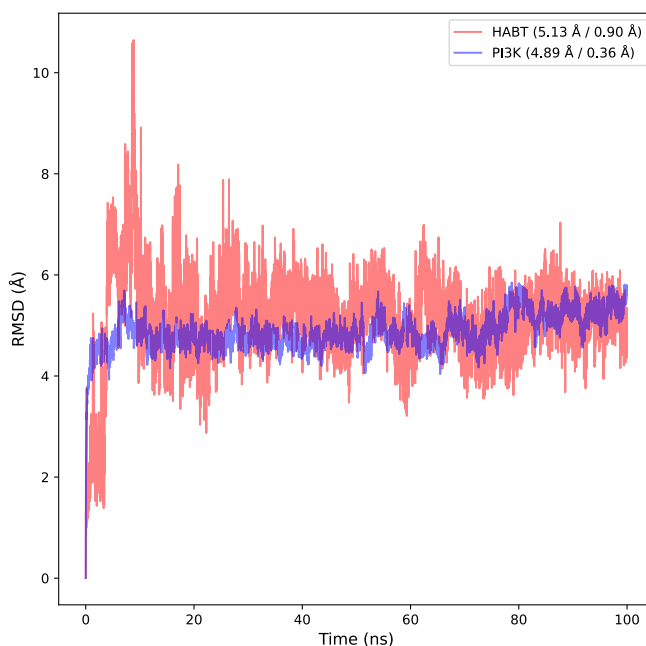
As a starting point, unbiased MD simulations were carried out for all studied three systems, being them: HABT in vacuum (1), HABT in water (2), and protein-docked HABT in water (3). From systems 1 and 2, it was possible to evaluate the solvent effect on HABT metastates distribution. In the other hand, from system 3, it would be possible to evaluate the proteic environment, to select conformations for the QM calculations, and then, obtain physical interpretation about how metastates of HABT affects its performance in signaling PI3K protein. In this scenario, a Root Mean Square Deviation (RMSD) analysis was performed in order to evaluate the simulation stability for all studied systems.

Therefore, Figure SI.6.1 shows the obtained RMSD values for HABT in system 1 and 2. For system 1, it was found a mean RMSD value of 0,13 Å, and a standard deviation of 0,05 Å. Now, for system 2, a mean RMSD value of 0,15 Å was obtained, and a standard deviation of 0,05 Å. The values found for both systems implies the simulation stability.

Now for system 3, Figure 4.2 shows the obtained RMSD values of HABT and PI3K for each simulation step. In that regard, it was obtained a RMSD average of 5.13 Å for the HABT molecule, with a standard deviation of 0.90 Å. For PI3K, a RMSD average of 4.89 Å was

obtained, with a standard deviation of 0.36 Å. Therefore, the obtained values are considered low for large systems, implying reasonable stability for the performed simulation.

Figura 4.2 – RMSD (Å) vs Time (ns). The RMSD profile of HABT is shown in red, and PI3K in blue. In the figure caption, it appears values of: average / standard deviation.



In addition to that, the obtained value for PI3K was close for the one obtained by Santos et al. (2023b), which was a RMSD mean of 3.68 Å, and a standard deviation of 0.65 Å. Despite the fact that the mentioned published work used a different force field for PI3K, the closer values reinforce the simulation stability. In that direction, from the unbiased trajectory it would be possible to collect reliable conformations for further QM calculations.

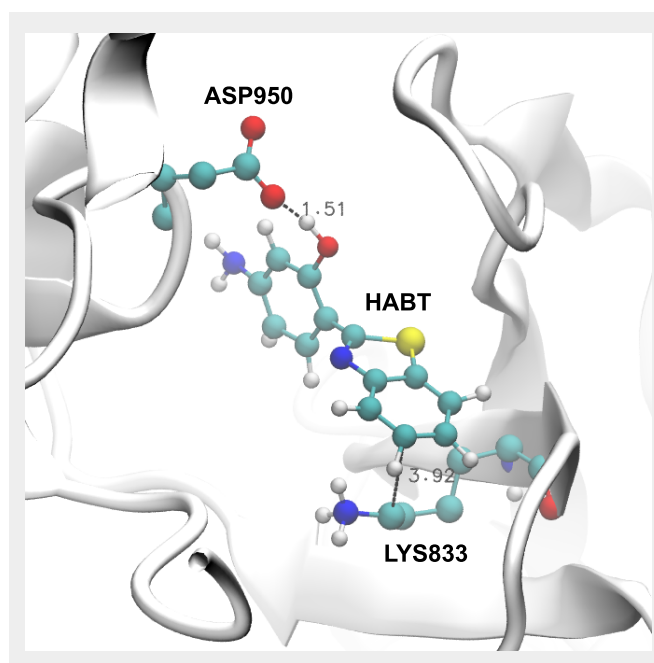
For a better understanding about the nature of the systems, it is important to evaluate interactions that may help stabilize the observed evolution of the systems. Since system 1 was simulated in vacuum, only intramolecular interactions can occur. In that scenario, by observing the HABT structure (Figure 4.1), the only h-bond interaction that can occur is the one between the -OH group of its phenyl portion and the -N= group of its benzothiazole portion, shown in Figure SI.6.2. Thus, this interaction was held for almost all simulated frames of system 1.

However, after the addition of water, intermolecular h-bond must be taken into account. In this scenario, despite for system 2 the -OH and -N= intramolecular h-bond interaction still occur with great relevance, competitive intermolecular h-bond can occur with water molecules.

Therefore, it was verified that for system 2 an intermolecular h-bond between the hydrogen of -OH group of HABT and oxygen of water molecules can take place of the intramolecular h-bond interaction (see Figure SI.6.3). Nevertheless, only a slight increase in its RMSD average was observed for system 2 when compared to system 1. Thus, it was possible to imply that the -OH and -N= intramolecular h-bond interaction still kept its great relevance.

On the other hand, for system 3, much more relevant competitive intermolecular h-bonds can occur, since in this system the HABT molecule was docked in the PI3K protein. However, the evaluation of h-bonds in this scenario is much more complex, once many amino acid residues can form interactions of this nature. Therefore, a h-bond analysis between the HABT and PI3K was performed using the VMD (Visual Molecular Dynamics) (HUMPHREY et al., 1996). A cutoff radius of 3.0 Å and cutoff angle of 30° were considered. It was verified one very important h-bond, formed between the hydrogen of the -OH group of HABT and the carboxylic group of the aspartic acid residue (ASP950), shown in Figure 4.3.

Figure 4.3 – Captured interactions between HABT and protein PI3K. The ASP-HABT interaction occurs between the carboxylic group of ASP and the -OH group of HABT. The LYS-HABT interaction occurs between the LYS carbonic chain and the benzene group of HABT.



The HABT:OH...O:ASP intermolecular h-bond interaction represented 69.75% of these interactions. This interaction starts occurring after 20 ns of simulation production, shown by the RMSD stabilization in Figure 4.2. Meanwhile, at the beginning of the simulation, it was verified h-bonds with the valine residue (VAL882), as pointed out by Santos et al. (2023b), re-

presenting 11.46% of the h-bond interactions. However, this interaction did not hold with the simulation evolution. Its occurrence was only observed before 20 ns of simulation, indicated by an increasing of RMSD until 20 ns, where the observed stabilization after this point is due to HABT:OH \cdots O:ASP interaction formation, shown in Figure 4.2.

Another important interaction, that is not captured in the h-bond analysis, is the interaction between the lysine residue (LYS833) and the benzene portion of HABT, shown in Figure 4.3. It is important to mention that this hydrophobic interaction is not as stabilizing as the mentioned h-bonds interactions. However, it follows from the simulation point that the HABT:OH \cdots O:ASP h-bond interaction starts occurring.

In this regard, exploring the interactions between the HABT molecule and its environment could provide valuable insights into the stabilization of the accessed metastable states.

4.3.2 Biased MD simulations: assessing metastates distribution and evaluating the solvent effect

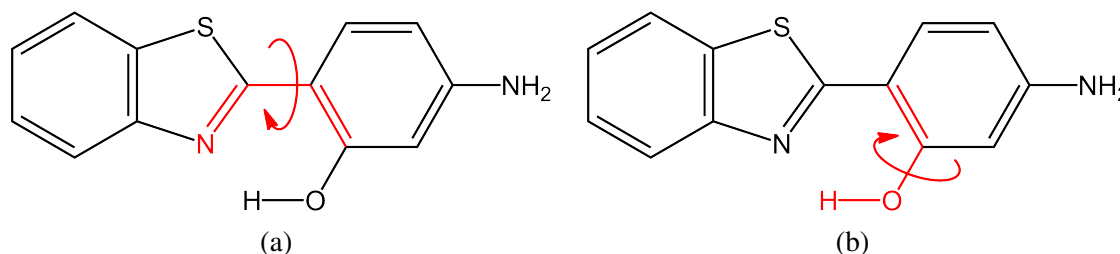
From biased trajectories it was possible to evaluate the accessed metastable states for all three systems. For that, PCA analysis was conducted in order to identify and characterize the accessed metastable states. The characterization was made in terms of descriptors, specifically chosen considering the nature of the HABT molecule.

The descriptor's source is the obtained trajectory files $(\mathbf{R})_{i=1}^N$ of the biased MD simulations, where i goes up to the N steps of the performed simulation. However, \mathbf{R} represents cartesian coordinates, which vary considerably from step to step, making it prudent to calculate more appropriate descriptors to encode the slow modes of the system. In this scenario, dihedral angles can be a good choice of descriptor, once internal coordinates are capable of providing a correct separation between the internal and overall motion (ALTIS et al., 2007).

Therefore, since the main goal by performing biased MD simulations is to identify and characterize the metastable states of the HABT molecule, and not the metastates of the entire system, only HABT descriptors were calculated. The HABT molecule consists of a conjugated heterocyclic structure, therefore mostly rigid. Thus, the dihedrals chosen as descriptors, shown in Figure 4.4, were chosen based on non-rigid portions of the molecule. Hence, dihedral 1 was chosen since the C-C bond that connects the benzothiazole and phenyl parts can rotate with a certain degree of freedom. Now, dihedral 2 was chosen based on the fact that the C-OH bond of the phenyl portion can also rotate with some degree of freedom. In addition to that, dihedral 2

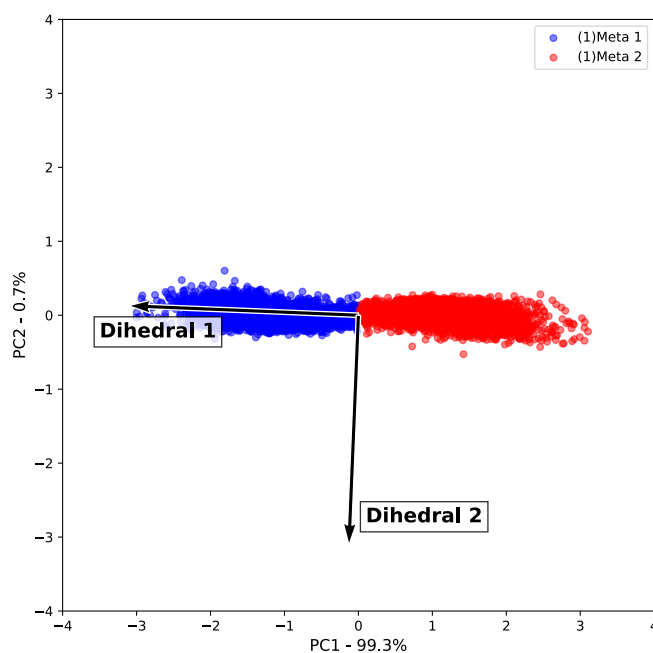
is of particular interest when studying the ESIPT process, once the hydrogen of the -OH group is transferred to the -N= group of the benzothiazole portion when ESIPT occurs. Therefore, it is expected that the distortion of the dihedral 2 may affect the ESIPT fluorescence.

Figura 4.4 – Chosen descriptors for further analysis, being (a) dihedral 1, and (b) dihedral 2.



With the calculated descriptors, it was possible then to perform PCA for all three systems. Hence, comparing the obtained metastates distribution of system 1 and 2, it was possible to evaluate the solvent effect. Starting from system 1, Figure 4.5 shows the biplot graph made for this system. It was possible to observe that two long lived metastable states were accessed ((1)Meta 1 and (1)Meta 2), being differentiated by PC1 (99.3% of explained variance).

Figura 4.5 – Biplot graph made from PCA analysis for system 1. Metastates were labeled according to (*N*)Meta *n*, where *N* denotes the system label, and *n*, the metastate label in that system.



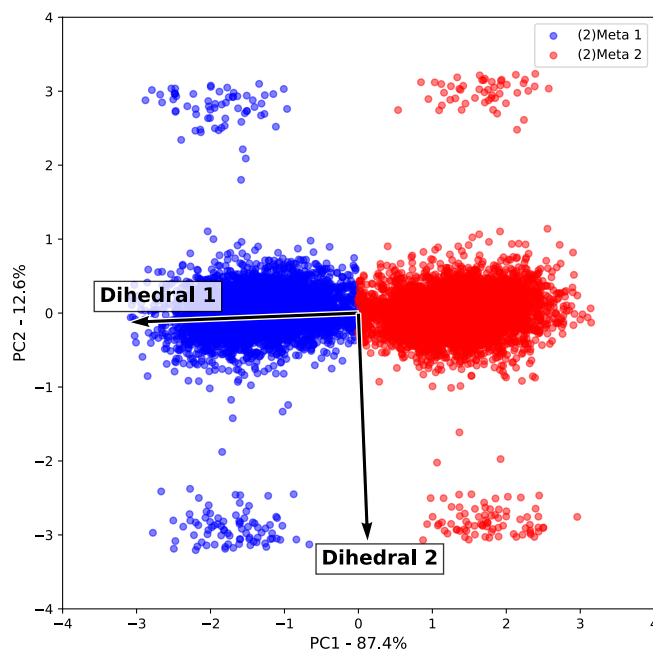
After performing a population analysis, it was observed that (1)Meta 1 and (1)Meta 2 contain almost the same population, being 5,012 conformations for (1)Meta 1, and 4,989 conformations for (1)Meta 2. In this scenario, these conformations were differentiated only by a dihedral 1 (Figure 4.4.a) fluctuation, presenting no considerable fluctuation of dihedral 2 (Figure 4.4.b). Therefore, for (1)Meta 1 it was observed more positive values of dihedral 1, and for (1)Meta 2, more negative ones.

Exploring the interactions that could explain the observed results, it is clear that system 1 does not present any intermolecular interactions, since this system was simulated in vacuum. Thus, only intramolecular interactions can stabilize the accessed metastable conformations. Therefore, one important interaction that prevents dihedral 2 fluctuation is the -OH and -N= intramolecular h-bond interaction, shown in Figure SI.6.2. Thus, once there is no other interaction capable of stabilizing the dihedral 2 fluctuation, the -OH and -N= groups are kept close enough to maintain this interaction.

Now, for system 2, Figure 4.6 shows the biplot graph made for this system. Differently from what was observed for system 1, after the solvation, it was possible to notice the appearance of new metastable states with a smaller population, being differentiated by PC2 (12.6% of explained variance). After population analysis, it was shown that the new appeared metastates presented a population that varies between 50 to 83 conformations. This new accessed metastates are characterized by its dihedral 2 fluctuation, which has no considerable fluctuation for the main group (located around PC2=0). Therefore, the main group maintained almost the same population as observed for system 1, consisting of 4,849 conformations for (2)Meta 1, and 4,876 conformations for (2)Meta 2. Hence, after the addition of water, a relevant dihedral 2 fluctuation could be stabilized.

In this scenario, intermolecular interactions should be taken into account to explain the stabilization of the accessed metastable conformations. It was verified that, for system 2, relevant dihedral 2 fluctuations are stabilized by intermolecular h-bond between the hydrogen of -OH group of HABT and oxygen of water molecules. This interaction is shown in Figure SI.6.3. However, the population analysis implies that the occurrence of this interaction is much lower than the -OH and -N= intramolecular h-bond interaction. Hence, it was verified that only water is not capable of effectively stabilizing a relevant dihedral 2 fluctuation in long-lived metastable state.

Figura 4.6 – Biplot graph made from PCA analysis for system 2. Metastates were labeled according to (N) Meta n , where N denotes the system label, and n , the metastate label in that system.



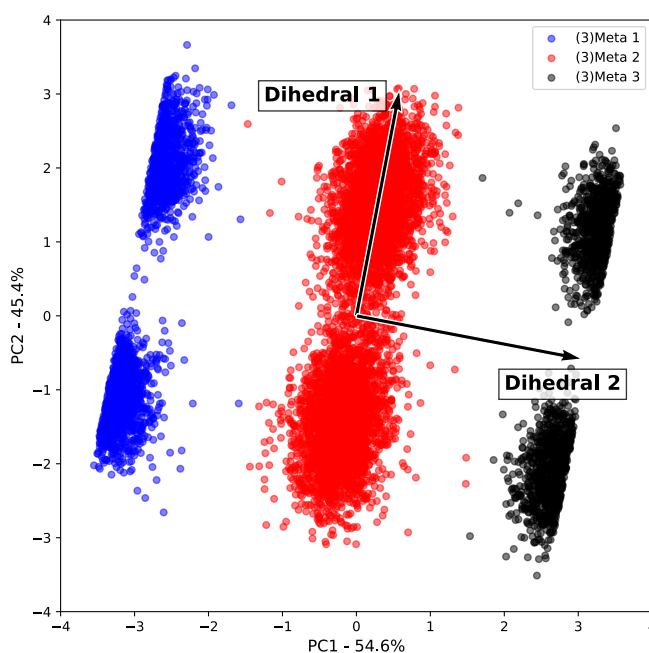
Therefore, by evaluating the solvent effect, valuable insights could be obtained about the descriptor's behavior. The discussed analysis showed that by adding possible intermolecular interactions to the system, dihedral 2 starts to gain relevance in explaining the metastates distribution of the HABT molecule. This is of great importance, once this knowledge led to a better comprehension of the obtained distribution for system 3. In this scenario, it was possible to move further and evaluate the HABT metastates in a much more complex scenario.

4.3.3 Biased MD simulations: assessing the distribution of metastable states and evaluating the effect of the protein environment

For system 3, a separate discussion was held, since this is the system of interest in this work, which will provide conformations for further ESIPT study with QM calculations. Hence, Figure 4.7 shows the biplot graph obtained from the PCA analysis done for system 3. At a first glance, it was noticed that differently from what was observed for system 1 and 2, for system 3, three long-lived metastates were accessed, being them separated in PC1 axis, which contains 54.6% of the explained variance. Also, there was a substantial increase in the variance

explained by PC2, which for system 3 is 45.4%, almost four times higher than the one observed for system 2.

Figura 4.7 – Biplot graph made from PCA analysis for system 3. Metastates were labeled according to (N) Meta n , where N denotes the system label, and n , the metastate label in that system.



In this scenario, by performing a population analysis, it was observed that (3)Meta 2 contains almost 63% of the explored conformations, where 6294 conformations belong to (3)Meta 2. For those metastates at the ends of PC1 axis ((3)Meta 1 and 3), approximately 37% belongs to them, where they are almost split in half, where (3)Meta 1 contains 1916 conformations, and (3)Meta 3, 1791 conformations. Furthermore, all three long-lived metastates are decomposed in two shorter-lived metastates, which is observed in the separation performed by the PC2 axis. This decomposition occurs in a way that the long-lived metastate were split in half, where both clusters separated by the PC2 axis represent approximately 50% of the conformations of its respective long-lived metastate.

Concerning the descriptor's behavior in the differentiation of the long and shorter-lived metastates, it was observed an inversion of relevance when compared to systems 1 and 2. For system 3, dihedral 2 was relevant to differentiate long-lived metastates, being more positive in (3)Meta 3, more negative in (3)Meta 1, and presenting irrelevant fluctuation in (3)Meta 2. Now, dihedral 1 was relevant to differentiate the decomposition of the three long-lived metastates,

in other words, to differentiate the shorter-lived metastates. For those metastates with shorter lifespan, the ones located more positively in the PC2 axis, the dihedral 1 was verified to be more positive, and for those more negative in PC2 axis, more negative.

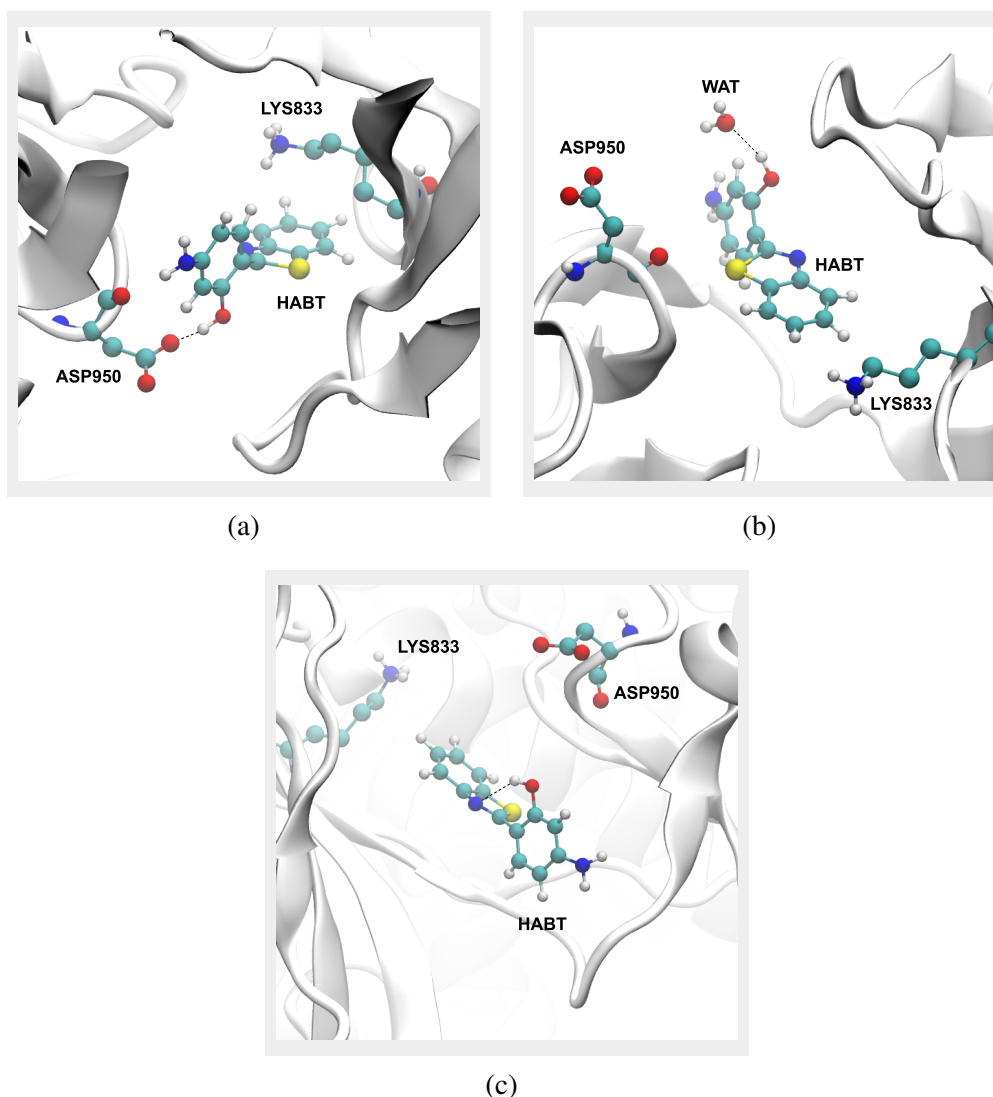
Therefore, after docking the HABT molecule in the protein PI3K, it was observed a considerable increase in dihedral 2 fluctuation. As discussed for system 2, the water molecules affect the fluctuation of dihedral 2, in which this occurrence is due to the presence of intermolecular interactions between the HABT and water molecules. These intermolecular interactions are also present in system 3, being part of the explanation for the observed, since HABT is docked in a water exposed cavity of PI3K. However, the main reason for an increase in the dihedral 2 fluctuation remains on the new intermolecular interactions of HABT and amino acid residues of the protein PI3K. Thus, exploring the interactions that occur between the HABT molecule and the PI3K protein is important to understand how such metastates could be accessed.

That said, the h-bond analysis conducted for system 3 showed the h-bond interactions between the HABT molecule and PI3K. It was verified that 69.75% of the h-bond interactions between the HABT molecule and PI3K occurs with the aspartic acid residue (ASP950). Despite the lower percentage of lysine residue (LYS833) interacting by h-bond with the HABT molecule, it was verified that lysine residues remains very close to the benzene part of the benzothiazole portion of HABT, implying hydrophobic interactions between carbon chains. Figure 4.3 shows both ASP-HABT and LYS-HABT interactions. The aspartic acid residue mainly interacts by h-bond with the -OH of the phenyl portion of HABT, and lysine residue mainly interacts by hydrophobic interactions with the benzene part of the benzothiazole portion of HABT.

In this scenario, both interactions may be relevant to explain the stabilization of the accessed metastable states. The aspartic acid residue interactions can stabilize the fluctuations of dihedral 2. Now, the lysine residue can help stabilize the fluctuations of dihedral 1, where this can occur by preventing the C-C bond that connects the benzothiazole to the phenyl part of HABT to rotate.

Therefore, an investigation was performed in order to verify specific interactions that stabilized each metastable conformations. In this sense, Figure 4.8 shows the observed interactions that stabilize the conformations of its respective metastates. In order to clearly explain the observed interactions, the accessed shorter-lived metastates were divided into (3)Meta nA and nB, where n is the number of the respective long-lived metastate, and the label A and B is associated with its location in PC2. Figure SI.6.4 shows how the labeling was defined.

Figura 4.8 – Shows the stabilizing interactions that a) stabilizes conformations of (3)Meta 1A and 3A, b) stabilizes conformations of (3)Meta 1B and 3B, and c) stabilizes conformations of (3)Meta 2A and 2B.



From Figure 4.8.a), it was possible to observe that the conformations of (3)Meta 1A and 3A stabilizes its dihedral 2 fluctuation by making an intermolecular h-bond interaction between the -OH group of HABT and the carboxylic group of ASP residue. Now, from Figure 4.8.b), it was possible to observe that the conformations of (3)Meta 1B and 3B stabilizes its dihedral 2 fluctuation by making an intermolecular h-bond interaction between the -OH group of HABT and water molecules. This difference occurs once in the metastates 1B and 3B, the dihedral 1 fluctuation prevents the h-bond interaction with the ASP residue. Finally, Figure 4.8.c) shows that the conformations of (3)Meta 2A and 2B are stabilized by the -OH an -N= intramolecular h-bond interaction. It is important to mention that, in all accessed metastable states, the LYS-HABT hydrophobic interaction is present, helping the stabilization of the dihedral 1 fluctuation.

Therefore, from population analysis done in this subsection, it was possible to observe that the -OH an -N= intramolecular h-bond interaction is the most stabilizing. This is possible to conclude in a scenario where almost 63% of the accessed conformations belongs to (3)Meta 2, metastable state whose conformations are stabilized based on the mentioned interaction. On the other hand, it is possible to observe that the protein environment was capable of stabilizing much more conformations than only the water environment. From system 2 to system 3, the conformations stabilized by h-bond intermolecular interactions increased substantially. In that way, the results presented so far show the challenge when dealing with biological environments, and the need for new approaches, like considering a nonequilibrium environment.

Despite MD simulations providing valuable information about metastates distribution and stabilization, it does not provide information about ESIPT spectroscopy. For that, more accurate calculations, like QM calculations, should be performed. However, due to the fact that QM calculations are much more computationally expensive, for a feasible analysis it is necessary to reduce the number of conformations that will go through expensive calculations. Therefore, selecting representative conformations of each accessed metastate may be a good strategy for ESIPT study.

4.3.4 Conformational selection

Once identified and characterized the accessed metastable states of system 3, it was possible to move forward and select representative conformations of these metastates. This selection is of great importance in order to study the effect of metastable states on the ESIPT reaction for PI3K signaling. In this scenario, all accessed metastates, being them long or shorter-live metastable states, were treated in accordance with the previously mentioned labeling, presented in Figure SI.6.4. Now, about the conformational selection, the criteria was based on the average values of dihedral 1 and 2.

In each metastate, the average of dihedral 1 and 2 were calculated, as shown by Table SI.6.1, and the conformation whose dihedral 1 and 2 values were the closest of the averages, were selected. In addition to that, the equilibrium conformation was also selected using the same criteria. However, the dihedral 1 and 2 were calculated from the unbiased trajectory. In this scenario, the average value was computed from all the 10000 frames of the unbiased trajectory. Then, the conformation with the closest dihedral 1 and 2 values from its average was selected. These values are also shown in Table SI.6.1.

In this scenario, Table 4.1 was constructed in order to show the values of dihedral 1 and 2 of all selected conformations, including a third variable, being it the distance between the group -N= and the hydrogen of the group -OH of HABT molecule. The mentioned distance is of great importance. Since both groups are involved in the ESIPT reaction, it is expected a hampered ESIPT for those conformations whose distance N...H are larger. The selected conformations for each metastate can be observed in Figure 4.9.

Tabela 4.1 – Table shows the values of dihedral 1, 2 and distance N...H of each representative conformation selected.

Conformation	Dihedral 1 (rad)	Dihedral 2 (rad)	Distance N...H (Å)
1A	1.52	-2.90	4.40
1B	-1.74	-2.91	4.49
2A	1.49	-0.04	2.94
2B	-1.46	0.02	2.61
3A	1.72	2.91	4.53
3B	-1.58	2.92	4.48
<i>EQ</i>	<i>0.03</i>	<i>-0.16</i>	<i>2.30</i>

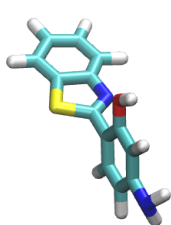
Note: EQ stands for the selected equilibrium conformation used for comparison with the selected metastable conformations.

Therefore, as a general description, it was possible to see that conformations of (3)Meta 2 are the most similar to the equilibrium one, shown in Figure 4.9.g). This occurs once in conformations of (3)Meta 2 the fluctuation of dihedral 2 is irrelevant, presenting only fluctuation for the dihedral 1.

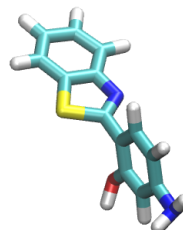
On the other hand, conformations of (3)Meta 1 and 3 are the most different of the equilibrium one, presenting fluctuations on both dihedrals, and a higher N...H distance. For the equilibrium conformation, as expected, no relevant fluctuation was verified for dihedral 1 and dihedral 2, being the closest of the optimized structure.

Thereby, from the selected conformations, it was possible to move further for QM calculations. With that, it was possible to evaluate the effect of the metastates of HABT in the ESIPT process. Not only that, this effect could be verified in spectroscopic terms of the HABT molecule.

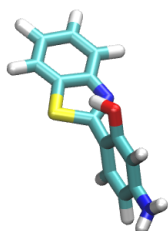
Figura 4.9 – The figure presents the six selected conformations, being them: 1A, 1B, 2A, 2B, 3A e 3B. Also, the equilibrium conformation is also shown.



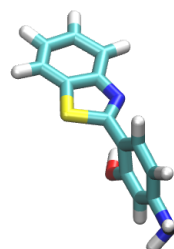
(a) Conformation 1A



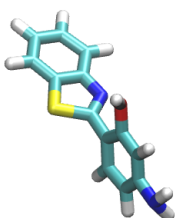
(b) Conformation 1B



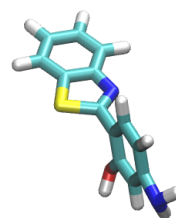
(c) Conformation 2A



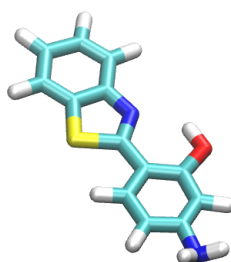
(d) Conformation 2B



(e) Conformation 3A



(f) Conformation 3B

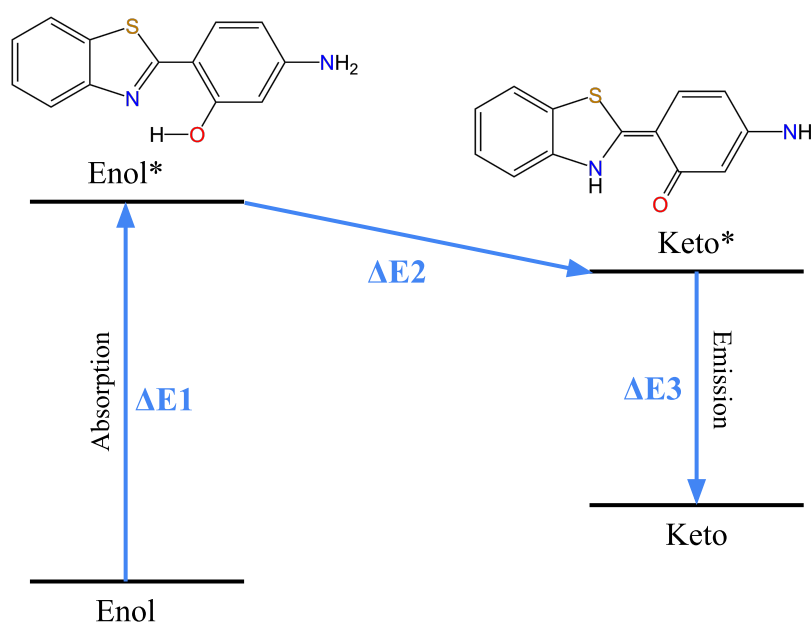


(g) Equilibrium conformation

4.3.5 Fluorescence properties in the protein environment

After QM calculations, it was possible to evaluate the ESIPT process for all selected metastable conformations. Therefore, in order to understand the obtained results, the collected data was summarized in Table 4.2, where the values contained in table follow the values shown in Figure 4.10, being them the energy difference between the states. In this scenario, it was verified that the equilibrium conformation is capable of performing ESIPT, thus following the expected, as shown by the work published by Mancini et al. (2015). However, when observing the obtained results for the meta conformations, it was observed that only the ones belonging to (3)Meta 2 (2A and 2B) were capable of performing ESIPT, indicating only enol emission for those belonging to (3)Meta 1 and 3.

Figura 4.10 – Shows the ESIPT scheme, where $\Delta E1$, $\Delta E2$ and $\Delta E3$ represents the energy variation between the states, values which are presented in Table 4.2.



Therefore, what was possible to conclude by observing the values in Table 4.2 and the conformations in Figure 4.9 is that, the ESIPT only occurs when there exists an intramolecular interaction by h-bond between OH \cdots N (Figure SI.2). This conclusion is in accordance with the proposed mechanism for this reaction by Eigen and Weller (EIGEN, 1964; WELLER, 1955), once it reinforces the need of the h-bond contact between the photoacid and the near basis. In that way, it was only verified the occurrence of ESIPT for the ones where the distance N \cdots HO is below 3.00 Å.

Tabela 4.2 – Shows the obtained energy differences between states, in eV, according to the shown in Figure 4.10.

Conformation	$\Delta E1$	$\Delta E2$	$\Delta E3$
1A	3.98	-	-
2A	3.17	-0.03	2.28
3A	3.56	-	-
1B	3.84	-	-
2B	3.81	-0.50	2.68
3B	3.70	-	-
<i>EQ</i>	<i>3.37</i>	<i>-0.43</i>	<i>2.43</i>

Note: EQ stands for the selected equilibrium conformation used for comparison with the selected metastable conformations.

Thus, the appropriate geometry for ESIPT is only achieved when a negligible fluctuation of dihedral 2 is observed. In other words, only conformations where there exists an intramolecular interaction by h-bond between OH \cdots N are capable of performing ESIPT. Hence, conformations 1A, 1B, 3A and 3B did not present ESIPT occurrence. In the other hand, despite the fact that dihedral 1 fluctuation had a minor impact on favoring or disfavoring the ESIPT, its fluctuation did not prevented the ESIPT occurrence, once it not inhibited the mentioned intramolecular interaction

Therefore, at closer distances of intramolecular interaction by h-bond between OH \cdots N, the ESIPT process is favored. Between the three conformations that verified the ESIPT occurrence, being them the conformations EQ, 2A, and 2B, the worst favored was the conformation 2A, with an $\Delta E2$ of -0.03 eV. For this conformation, the distance N \cdots HO was the highest, being 2.94 Å. Now for conformations 2B and EQ, it was verified to have a close value of $\Delta E2$, being -0.50 eV for 2B, and -0.43 eV for EQ. The distance N \cdots HO values were lower than the one for 2A, being 2.61 Å for 2B, and 2.30 Å for EQ.

Observing the spectroscopy of the conformations in which there is the occurrence of ESIPT ((3)Meta 2 and EQ), it was observed a near violet enol emission for all of those conformations ($\Delta E1$ of 3.17-3.81 eV). For conformations 2B and EQ, the enol emissions occurs in the UVA region, being 325.42 and 367.91 nm, respectively. However, for conformation 2A it was observed a higher wavelength of 391.12 nm, located in the visible violet.

For the keto emissions, it was observed for conformations 2A and EQ a near green emission ($\Delta E3$ of 2.28-2.43 eV), with 543.79 and 510.22 nm, respectively, while for the 2B conformation, a blue emission ($\Delta E3$ of 2.68 eV), with 462.63 nm. Therefore, conformations

2A and EQ presented the highest stokes shift, implying a higher redshift, with 152.67 and 142.31 nm respectively, while conformation 2B presented a lower stokes shift, implying a lower redshift, with 137.21 nm. Thus, all conformations where it was verified the occurrence of ESIPT presented a considerable stokes shift, appropriate for dual emission fluorescent sensors.

Now, for conformations of (3)Meta 1 and 3, competitive intermolecular h-bond interactions caused an inappropriate geometry for ESIPT occurrence, verified from the fact that no ΔE_3 emission was observed for those conformations. The inappropriate geometry could also be verified from the higher values of N...HO distance, where these values are kept around 4,50 Å for those conformations. This increase is caused by dihedral 2 fluctuations, which are stabilized by h-bonds with the surroundings.

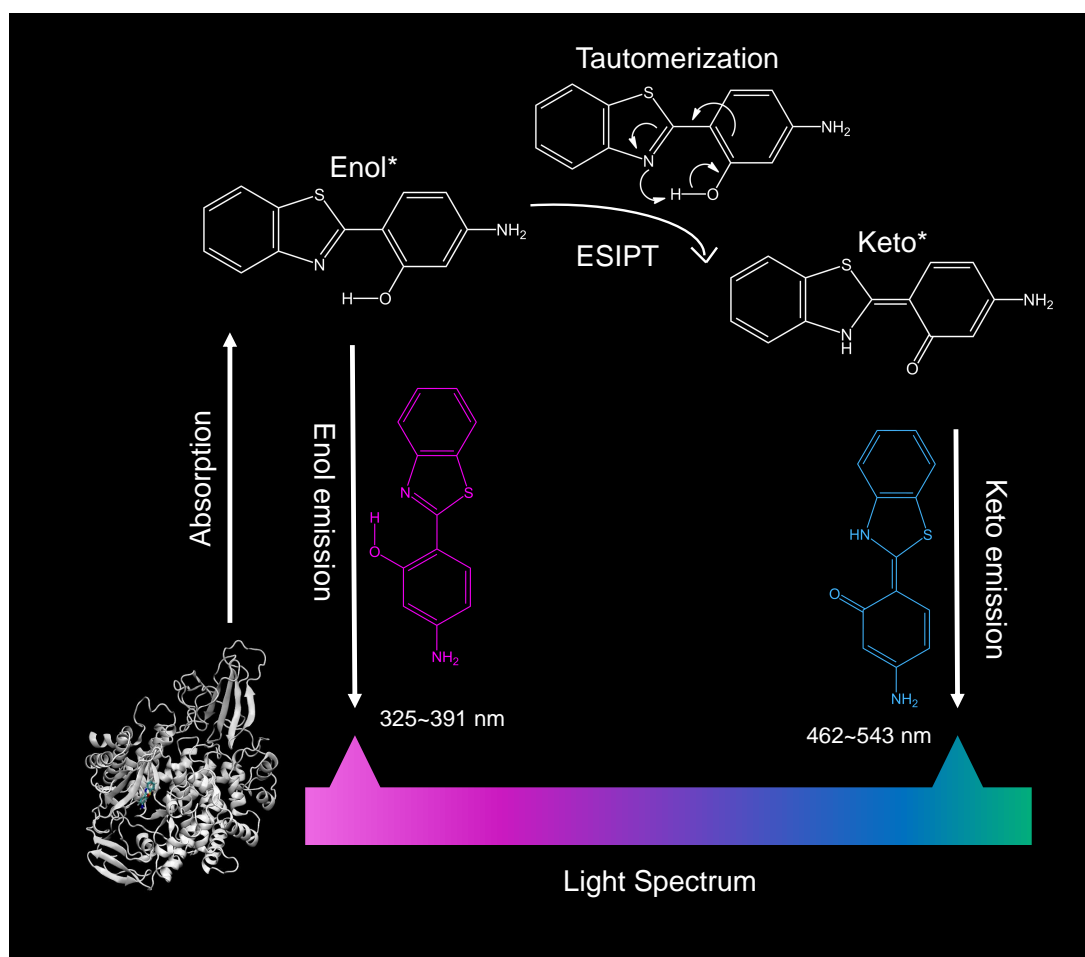
Therefore, it was verified that for the conformations of (3)Meta 1A and 3A, the dihedral 2 fluctuation is stabilized by an intermolecular h-bond interaction between the hydrogen of -OH group of HABT and the carboxylic group of ASP residue, interaction shown in Figure 4.8.a). On the other hand, dihedral 1 fluctuation to more negative values inhibits the OH...ASP intermolecular h-bond interaction. Therefore, the stabilization of dihedral 2 fluctuation in conformations 1B and 3B occurs due to intermolecular h-bond interaction between the hydrogen of -OH group of HABT and oxygen of water molecules. In Figure 4.8.b) it is shown the mentioned interaction. Therefore, conformations of (3)Meta 1 and 3 are not dual emission sensors, implying a much poorer performance as a fluorescent sensor.

Summarizing the obtained results, it was observed that the ESIPT only occurs when an intramolecular h-bond interaction between OH...N exists, since it keeps HABT in an appropriate geometry. In addition to that, closer interactions favored the ESIPT occurrence. Now, once this interaction is inhibited, there is no occurrence of ESIPT. Hence, this interaction ends once a dihedral 2 fluctuation is observed, where this fluctuation is stabilized by intermolecular h-bond interactions with the surroundings. Figure 4.11 summarizes the presented results in this subsection.

In this scenario, remembering the population analysis performed when observing metastates distribution, it was possible to evaluate the HABT performance as a dual emission fluorescent sensor. As discussed in this subsection, only conformations 2A and 2B presented an ESIPT occurrence. Those conformations represent almost 63% of all the accessed conformations in a nonequilibrium scenario. Therefore, since most conformations are capable of performing ESIPT, it is possible to imply that, despite a decrease of performance is observed,

the HABT molecule is still efficient as a dual emission fluorescent sensor in a nonequilibrium scenario. Thus, the HABT molecule would be capable of signaling the PI3K protein target for cancer diagnosis with reasonable efficiency.

Figura 4.11 – Shows a summary of the obtained spectroscopic results.



4.4 Conclusion

In the present work, it was possible to observe that the HABT remains as an efficient dual emission fluorescent sensor for PI3K signaling. However, when evaluated in a nonequilibrium scenario, a decrease in its efficiency was observed. This occurs once the ES IPT was observed in approximately 63% of the accessed conformations, and not observable in 37% of the accessed conformations.

In this scenario, it was verified that dihedral fluctuations had a considerable impact on the ES IPT occurrence. Both dihedral 1 and 2 fluctuations had an impact on the HABT performance in ES IPT, where the last was the most important. It was observed that fluctuations

of dihedral 2 were crucial in preventing the ESIPT occurrence. That occurred in a scenario where these fluctuations were the cause for an inappropriate geometry for ESIPT occurrence. From this, the intramolecular OH...N h-bond interaction could be inhibited, interaction that is of great relevance for the tautomerization process.

In this sense, the present work is a step forward through development of new fluorescent sensors for cancer diagnosis. From Schrödinger's point of view, explored in several works, it was possible to add a more realistic comprehension of biological systems, necessary to develop new probes for disease diagnosis. Therefore, when dealing with theoretical methodologies, adding the necessary complexity to models is the main goal. Thus, the present work shows that from an entirely theoretical methodology, it was possible to achieve fresh conclusions about the performance of a dual emission fluorescent sensor for cancer diagnosis.

In this sense, these results reveal very important data for future tests, considering that the cancer remains without cure, and the early diagnosis of this disease can save many lives. Besides that, this study can also serve as a basis for future experiments and other theoretical studies, as a starting point for the development of compounds derived from ciprofloxacin antibiotic, which can act as even more efficient fluorescent probes, and with lower levels of toxicity.

5 CONCLUSÃO

A partir da metodologia executada, foi possível concluir que a PCA, aliada a simulações de dinâmica molecular enviesadas, se mostrou uma boa ferramenta para explorar os estados metaestáveis acessados. Esta conclusão se fez possível principalmente quanto a sua capacidade de diferenciar agrupamentos com base nos descritores escolhidos, além de permitir uma análise populacional destes agrupamentos. Nas três PCAs feitas, uma para cada um dos três sistemas, foi evidenciado a importância dos diedros calculados na caracterização dos estados metaestáveis acessados.

Além disso, foi possível concluir que a adição de perturbações ao sistema, sendo elas a adição do solvente água e posteriormente da ancoragem da molécula de HBT na proteína PI3K, apresentaram um efeito significativo na distribuição das conformações em estados metaestáveis. Não apenas isso, foi observado que estas perturbações aumentam a relevância do descritor diedro 2 na diferenciação dos metaestados acessados, acompanhada de uma crescente queda na relevância do descritor diedro 1 para tal tarefa. Tal conclusão pôde ser feita a partir da observação de um decréscimo na variância explicada pela PC1 e aumento da variância explicada pela PC2 do sistema 1 para o 2, além de inversão de relevância dos diedros, o qual foi observado no sistema 3.

Para mais, foi possível concluir que a interação intramolecular de ligação de hidrogênio $N \cdots H$ foi fundamental para conter a flutuação do diedro 2, mantendo a performance do ESIPT. Isso ocorre pois, para as conformações cuja flutuação do diedro 2 se mostrou relevante, não foi observado a performance do ESIPT. Nesse sentido, a partir da avaliação espectroscópica com métodos quânticos, observou-se que 63% das conformações performam o ESIPT, em que 37% não performam devido à uma flutuação relevante do diedro 2, esta que proporciona uma geometria desfavorável para o processo.

Por fim, o presente estudo revela informações valiosas para futuros testes, considerando que o câncer se mantém um problema atual, e o diagnóstico precoce desta enfermidade é capaz de salvar muitas vidas. Além disso, o presente estudo pode servir como base para futuros experimentos e outros estudos teóricos, sendo um ponto de partida para o desenvolvimento de compostos derivados do antibiótico cicloproxacino, os quais podem agir como sensores fluorescentes ainda mais eficientes, e com níveis mais baixos de toxicidade.

REFERÊNCIAS

- ALAVI, S. Molecular simulations: Fundamentals and practice. In: _____. 1. ed. [S.l.]: Wiley-VCH, 2020. p. 326. ISBN 978-3-527-34105-4.
- ALTIS, A.; P.H., N.; STOCK, G. Dihedral angle principal component analysis of molecular dynamics simulations. **J Chem Phys**, v. 126, p. 244111, 2007.
- ANDOLPHO, G.; CUNHA, E. da; RAMALHO, T. Insights into the value of statistical models, solvent, and relativistic effects for investigating re complexes of 2-(4-aminophenyl)benzothiazole: a potential spectroscopic probe. **J Mol Model**, v. 28, 2022.
- ANSARI, N.; RIZZI, V.; PARRINELLO, M. Water regulates the residence time of benzamidine in trypsin. **Nat Commun**, v. 13, p. 5438, 2022.
- ARISTOV, V. Biological systems as nonequilibrium structures described by kinetic methods. **Results in Physics**, v. 13, p. 102232, 2019.
- BALSERA, M. A.; WRIGGERS, W.; OONO, Y.; SCHULTEN, K. Principal component analysis and long time protein dynamics. **J. Phys. Chem.**, v. 100, p. 2567–2572, 1996.
- BARTOSZ, G. Use of spectroscopic probes for detection of reactive oxygen species. **Clinica Chimica Acta**, v. 368, p. 53–76, 2006.
- BROICHHAGEN, J.; FRANK, J.; TRAUNER, D. A roadmap to success in photopharmacology. **Acc Chem Res**, v. 48, p. 1947–1960, 2015.
- BUSLAEV, P.; GORDELIY, V.; GRUDININ, S.; GUSHCHIN, I. Principal component analysis of lipid molecule conformational changes in molecular dynamics simulations. **J Chem Theory Comput**, v. 12, p. 1019–1028, 2016.
- CHAKRABORTY, S.; RAHMAN, T. The difficulties in cancer treatment. **Ecancermedicalscience**, v. 6, 2012.
- CHAN, H.; OLIVEIRA, A.; SCHOFIELD, C.; MULHOLLAND, A.; DUARTE, F. Dynamical nonequilibrium molecular dynamics simulations identify allosteric sites and positions associated with drug resistance in the sars-cov-2 main protease. **JACS**, v. 3, p. 1767–1774, 2023.
- CHENG, S. Z. Chapter 3 - concepts of metastable states. In: CHENG, S. Z. (Ed.). **Phase Transitions in Polymers**. Amsterdam: Elsevier, 2008. p. 61–76. ISBN 978-0-444-51911-5.
- CHO, M. et al. Sensitive t2 mri contrast agents from the rational design of iron oxide nanoparticle surface coatings. **Journal of Physical Chemistry C**, v. 127, n. 2, p. 1057–1070, 2023.
- COLLIER, P. et al. Structural basis for isoform selectivity in a class of benzothiazole inhibitors of phosphoinositide 3-kinase . **J Med Chem**, v. 58, p. 517–521, 2015.
- COPELAND, R. The drug–target residence time model: a 10-year retrospective. **Nat Rev Drug Discov**, v. 15, p. 87–95, 2016.
- COPELAND, R. Evolution of the drug-target residence time model. **Expert Opin Drug Discov**, v. 16, p. 1441–1451, 2021.

COPELAND, R. A. Evaluation of enzyme inhibitors in drug discovery: A guide for medicinal chemists and pharmacologists. In: _____. 2. ed. [S.l.]: John Wiley Sons, Ltd, 2013. ISBN 978-1-118-48813-3.

CROSBY, D. et al. Early detection of cancer. **Science**, v. 375, 2022.

DAVID, C.; JACOBS, D. Principal component analysis: a method for determining the essential dynamics of proteins. **Methods Mol Biol**, v. 1084, p. 193–226, 2014.

DEMIREL, Y. Chapter 3 - fundamentals of nonequilibrium thermodynamics. In: DEMIREL, Y. (Ed.). **Nonequilibrium Thermodynamics (Third Edition)**. Third edition. Amsterdam: Elsevier, 2014. p. 119–176. ISBN 978-0-444-59557-7.

DENNINGTON, R.; KEITH, T. A.; MILLAM, J. M. **GaussView Version 6**. 2019. Semichem Inc. Shawnee Mission KS.

DU, X.; ZHU, F. A novel principal components analysis (pca) method for energy absorbing structural design enhanced by data mining. **Adv Eng Softw**, v. 127, p. 17–27, 2019.

DURRANT, J.; MCCAMMON, J. Molecular dynamics simulations and drug discovery. **BMC Biol**, v. 9, 2011.

D'ANGELO, N. et al. Discovery and optimization of a series of benzothiazole phosphoinositide 3-kinase (pi3k)/mammalian target of rapamycin (mtor) dual inhibitors. **J Med Chem**, v. 54, p. 1789–1811, 2011.

EBELING, W.; SOKOLOV, I. M. **Statistical Thermodynamics and Stochastic Theory of Nonequilibrium Systems**. [S.l.]: WORLD SCIENTIFIC, 2005.

EIGEN, M. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. part i: Elementary processes. **Angewandte Chemie**, v. 3, p. 1–72, 1964.

FANG, X.; KRUSE, K.; LU, T.; WANG, J. Nonequilibrium physics in biology. **Rev Mod Phys**, v. 91, p. 045004, 2019.

FRISCH, M. J. et al. **Gaussian 09 Revision A.2**. 2009.

GEORGIEV, N.; BAKOV, V.; ANICHINA, K.; BOJINOV, V. Fluorescent probes as a tool in diagnostic and drug delivery systems. **Pharmaceuticals**, v. 16, p. 381, 2023.

GHASSEMI, M.; SHAHIDIAN, A. Chapter 2 - thermodynamics. In: GHASSEMI, M.; SHAHIDIAN, A. (Ed.). **Nano and Bio Heat Transfer and Fluid Flow**. Oxford: Academic Press, 2017. p. 9–30. ISBN 978-0-12-803779-9.

GHOSH, D.; RANJAN, A. The metastable states of proteins. **Protein Sci**, v. 29, p. 1559–1568, 2020.

GONÇALVES, M.; SANTOS, L.; PRATA, D.; PEIXOTO, F.; RAMALHO, T. Nmr relaxation and relaxivity parameters of mri probes revealed by optimal wavelet signal compression of molecular dynamics simulations. **Int J Quantum Chem**, 2019.

GONÇALVES, M. A.; CUNHA, E. F. F. da; PEIXOTO, F. C.; RAMALHO, T. C. Probing thermal and solvent effects on hyperfine interactions and spin relaxation rate of -feoh(100) and [mnh3buea(oh)]²⁻: Toward new mri probes. **Comput. Theor. Chem.**, v. 1069, p. 96–104, 2015.

GRANDJEAN, A.; LUSTRES, L.; JUNG, G. Solvent-controlled intermolecular proton-transfer follows an irreversible eigen-weller model from fs to ns. **Molecules**, v. 5, p. 1094–1105, 2021.

HAN, H. et al. Recent advances in organic fluorescent probes for tumor related enzyme detection. **Dyes and Pigments**, v. 204, p. 110386, 2022.

HEQUET, E. et al. Design, characterization and molecular modeling of new fluorinated paramagnetic contrast agents for dual h-1/f-19 mri. **Magnetochemistry**, v. 6, n. 1, 2020.

HERNÁNDEZ-LEMUS, E. Nonequilibrium thermodynamics of cell signaling. **J. Thermodyn.**, 2012.

HICKS, J.; VANBROCKLIN, H.; WILSON, A.; HOULE, S.; VASDEV, N. Radiolabeled small molecule protein kinase inhibitors for imaging with pet or spect. **Molecules**, v. 15, p. 8260–8278, 2010.

HOBSON, N. J. et al. Facile aqueous, room temperature preparation of high transverse relaxivity clustered iron oxide nanoparticles. **Colloids and Surfaces a-Physicochemical and Engineering Aspects**, v. 570, p. 165–171, 2019.

HOLLINGSWORTH, S.; DROR, R. Molecular dynamics simulation for all. **Neuron**, v. 99, p. 1129–1143, 2018.

HOSPITAL, A.; GOÑI, J.; OROZCO, M.; GELPI, J. Molecular dynamics simulations: advances and applications. **Adv Appl Bioinform Chem**, v. 8, p. 37–47, 2015.

HU, J. et al. Conformationally locked salicylideneaniline derivatives with strong esipt fluorescence. **Dyes and Pigments**, v. 145, p. 493–504, 2017.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Vmd: visual molecular dynamics. **J Mol Graph**, v. 14, p. 33–38, 1996.

INSTITUTO NACIONAL DE CÂNCER - INCA. **Estatísticas de câncer**. [S.l.], 2023. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/numeros>>.

IVERNIZZI, M.; PARRINELLO, M. Rethinking metadynamics: From bias potentials to probability distributions. **J Phys Chem Lett**, v. 11, p. 27312736, 2020.

IVERNIZZI, M.; PIAGGI, P.; PARRINELLO, M. Unified approach to enhanced sampling. **Phys Rev X**, v. 10, 2020.

JAIN, A.; STOCK, G. Identifying metastable states of folding proteins. **J Chem Theory Comput**, v. 8, p. 3810–3819, 2012.

JHA, S.; JHA, N.; KAR, R.; AMBASTA, R.; KUMAR, P. p38 mapk and pi3k/akt signalling cascades in parkinson's disease. **nt J Mol Cell Med**, v. 4, p. 67–86, 2015.

JIANG, W.; JI, M. Receptor tyrosine kinases in pi3k signaling: The therapeutic targets in cancer. **Semin Cancer Biol**, v. 59, p. 3–22, 2019.

JOLLIFFE, I.; CADIMA, J. Principal component analysis: a review and recent developments. **Phil Trans R Soc A**, v. 374, 2016.

- JOSHI, H.; ANTONOV, L. Excited-state intramolecular proton transfer: A short introductory review. **Molecules**, v. 26, p. 1475, 2021.
- KARAKAS, B.; BACHMAN, K.; PARK, B. Mutation of the pik3ca oncogene in human cancers. **Br J Cancer**, v. 94, p. 455–9, 2006.
- KARPLUS, M.; MCCAMMON, J. Molecular dynamics simulations of biomolecules. **Nat Struct Biol**, v. 9, p. 646–52, 2002.
- KAUR, R.; BHARDWAJ, A.; GUPTA, S. Cancer treatment therapies: traditional to modern approaches to combat cancers. **Mol Biol Rep**, v. 50, p. 9663–9676, 2023.
- KHALIGHINEJAD, P.; PARROTT, D.; SHERRY, A. D. Imaging tissue physiology in vivo by use of metal ion-responsive mri contrast agents. **Pharmaceuticals**, v. 13, n. 10, 2020.
- KIM, J.; WALES, D.; YANG, G. Optical spectroscopy for in vivo medical diagnosis—a review of the state of the art and future perspectives. **Prog. Biomed. Eng.**, v. 2, p. 042001, 2020.
- KOBAURI, P.; DEKKER, J.; SZYMANSKI, W.; FERINGA, B. Rational design in photopharmacology with molecular photoswitches. **Angewandte Chemie**, v. 62, p. e202300681, 2023.
- KOSTAL, J. Computational chemistry in predictive toxicology. In: _____. [S.l.: s.n.], 2016. v. 10, p. 139–186. ISBN 9780128047002.
- LAIO, A.; PARRINELLO, M. Escaping free-energy minima. **PNAS**, v. 99, p. 12562–12566, 2002.
- LAW, C.; K.Y., Y. Current trends of benzothiazoles in drug discovery: a patent review (2015-2020). **Expert Opin Ther Pat**, v. 32, p. 299–315, 2022.
- LEE, C.; YANG, W.; PARR, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. **Phys. Rev. B**, American Physical Society, v. 37, p. 785–789, Jan 1988. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevB.37.785>>.
- LEVINE, I. **Physical Chemistry, 6th edition**. 6. ed. [S.l.]: McGraw-Hill Science/Engineering/Math, 2008. ISBN 0072538627; 9780072538625.
- LI, Z. et al. Development, validation, and application of a parametric pediatric head finite element model for impact simulations. **Ann Biomed Eng**, v. 39, p. 2984–2997, 2011.
- LINDAHL; ABRAHAM; HESS; SPOEL van der. **GROMACS 2021.4 Source code**. Zenodo, 2021. Disponível em: <<https://doi.org/10.5281/zenodo.5636567>>.
- LIU, P.; CHENG, H.; ROBERTS, T.; ZHAO, J. Targeting the phosphoinositide 3-kinase pathway in cancer. **Nat Rev Drug Discov**, v. 8, p. 627–44, 2009.
- LIU, Y. et al. Excited-state intramolecular proton transfer with and without the assistance of vibronic-transition-induced skeletal deformation in phenol–quinoline. **RSC Adv.**, v. 11, p. 37299–37306, 2021.
- LONG, Y. et al. Cooperation of esipt and ict processes in the designed 2-(2-hydroxyphenyl)benzothiazole derivative: A near-infrared two-photon fluorescent probe with a large stokes shift for the detection of cysteine and its application in biological environments. **Anal. Chem.**, v. 92, p. 14236–14243, 2020.

MA, X. Development of computational chemistry and application of computational methods. **J Phys: Conf Ser**, v. 2386, p. 012005, 2020.

MACCHIAGODENA, M. et al. Nonequilibrium alchemical simulations for the development of drugs against covid-19. In: _____. **In Silico Modeling of Drugs Against Coronaviruses: Computational Tools and Protocols**. New York, NY: Springer US, 2021. p. 231–271. ISBN 978-1-0716-1366-5. Disponível em: <https://doi.org/10.1007/7653_2020_50>.

MACHURA, B.; WOLFF, M.; BENOIST, E.; COULAIS, Y. Tricarbonyl rhenium(i) complex of benzothiazole – synthesis, spectroscopic characterization, x-ray crystal structure and dft calculations. **J Organomet Chem**, v. 724, p. 82–87, 2013.

MALDE, A. et al. An automated force field topology builder (atb) and repository: Version 1.0. **J Chem Theory Comput**, v. 7, p. 4026–4037, 2011.

MANCINI, D.; SEN, K.; BARBATTI, M.; THIEL, W.; RAMALHO, T. Excited-state proton transfer can tune the color of protein fluorescent markers. **Chemphyschem**, v. 16, p. 3444–9, 2015.

MANCINI, D.; SOUZA, E.; CAETANO, M.; RAMALHO, T. ⁹⁹tc nmr as a promising technique for structural investigation of biomolecules: theoretical studies on the solvent and thermal effects of phenylbenzothiazole complex. **MRC**, v. 52, p. 129–137, 2014.

MANDAL, P.; PANDA, A. Conformational effect on the excitonic states of 2-phenylpyridine oligomers: Ab initio studies and analysis. **J Phys Chem A**, v. 127, p. 7898–7907, 2023.

MANOJAI, N.; DAENGGERN, R.; KERDPOL, K.; KUNGWAN, N.; NGAOJAMPA, C. Td-dft study of absorption and emission spectra of 2-(2'-aminophenyl)benzothiazole derivatives in water. **J Fluoresc**, v. 27, p. 745–754, 2017.

MARINARI, E.; PARISI, G. Simulated tempering: A new monte carlo scheme. **Europhysics Letters**, v. 19, p. 451, 1992.

MARINOVA, V. et al. Identifying conformational isomers of organic molecules in solution via unsupervised clustering. **J Chem Inf Model**, v. 61, p. 2263–2273, 2021.

MARTÍNEZ, L.; BORIN, I.; SKAF, M. Fundamentos de simulação por dinâmica molecular. In: _____. [S.l.]: Livraria da Física, 2007. p. 413–452.

MAVROIDI, B. et al. Palladium(ii) and platinum(ii) complexes of derivatives of 2-(4-aminophenyl)benzothiazole as potential anticancer agents. **Inorganica Chim Acta**, v. 444, p. 63–75, 2016.

MCARDLE, S.; ENDO, S.; ASPURU-GUZI, A.; BENJAMIN, S.; YUAN, X. Quantum computational chemistry. **Rev. Mod. Phys.**, v. 92, p. 015003, 2020.

MOURIK, T. van; BÜHL, M.; GAIGEOT, M. Density functional theory across chemistry, physics and biology. **Philos Trans A Math Phys Eng Sci**, v. 10, p. 20120488, 2011.

NOVELLI, P.; BONATI, L.; PONTIL, M.; PARRINELLO, M. Characterizing metastable states with the help of machine learning. **J Chem Theory Comput**, v. 18, p. 5195–5202, 2022.

- PADALKAR, V. et al. Synthesis and photo-physical characteristics of esipt inspired 2-substituted benzimidazole, benzoxazole and benzothiazole fluorescent derivatives. **J Fluoresc**, v. 22, p. 311–322, 2012.
- PALMA, J.; PIERDOMINICI-SOTTILE, G. On the uses of pca to characterise molecular dynamics simulations of biological macromolecules: Basics and tips for an effective use. **ChemPhysChem**, v. 24, p. e202200491, 2023.
- PAUL, M.; MUKHOPADHYAY, A. Tyrosine kinase - role and significance in cancer. **Int J Med Sci**, v. 1, p. 101–115, 2004.
- PEREIRA, A. F. Dissertação de Mestrado, **Development and biological application of a quantum mechanically derived force field: the case of a platinum (II) complex**. Lavras-MG: [s.n.], 2020.
- PISANI, P. et al. Describing the conformational landscape of small organic molecules through gaussian mixtures in dihedral space. **J Chem Theory Comput**, v. 10, p. 2557–2568, 2014.
- QI, Y. et al. A new interpretation of the esipt mechanism of 2-(benzimidazol-2-yl)-3-hydroxychromone derivatives. **SAA**, v. 224, p. 117359, 2020.
- RAZANI, E. et al. The pi3k/akt signaling axis in alzheimer's disease: a valuable target to stimulate or suppress? **Cell Stress Chaperones**, v. 26, p. 871–887, 2021.
- RIZZI, V.; AURELI, S.; ANSARI, N.; GERVASIO, F. Oneopes, a combined enhanced sampling method to rule them all. **J Chem Theory Comput**, v. 19, p. 57315742, 2023.
- ROHRER, M.; BAUER, H.; MINTOROVITCH, J.; REQUARDT, M.; WEINMANN, H. Comparison of magnetic properties of mri contrast media solutions at different magnetic field strengths. **Investigative Radiology**, v. 40, 2005.
- ROOHI, H.; POURYAHYA, T. Td-dft study of the excited state intramolecular proton transfer (esipt) mechanism and photophysical properties in coumarin–benzothiazole derivatives: substitution and solvent effects. **Mol Syst Des Eng**, v. 8, p. 647–665, 2023.
- SANTOS, R.; TAVARES, C.; SANTOS, T.; RASOULI, H.; RAMALHO, T. Md simulations to calculate nmr relaxation parameters of vanadium(iv) complexes: A promising diagnostic tool for cancer and alzheimer's disease. **Pharmaceuticals**, v. 16, p. 1653, 2023.
- SANTOS, T.; ANDOLPHO, G.; TAVARES, C.; GONÇALVES, M.; RAMALHO, T. Improving the path to obtain spectroscopic parameters for the pi3k—(platinum complex) system: Theoretical evidences for using 195pt nmr as a probe. **Magnetochemistry**, v. 89, 2023.
- SCHAFFER, A.; HORN, H.; AHLRICHS, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. **The Journal of Chemical Physics**, v. 97, n. 4, p. 2571–2577, 08 1992. ISSN 0021-9606. Disponível em: <<https://doi.org/10.1063/1.463096>>.
- SCHMID, N. et al. Definition and testing of the gromos force-field versions 54a7 and 54b7. **Eur Biophys J**, v. 40, p. 843–56, 2011.
- SCHRÖDINGER, E. **What is Life? The Physical Aspect of the Living Cell**. [S.l.]: Cambridge University Press, 1944.

- SEDGWICK, A. et al. Excited-state intramolecular proton-transfer (esipt) based fluorescence sensors and imaging agents. **Chem. Soc. Rev.**, v. 47, p. 8842–8880, 2018.
- SERRANO, D.; BONANNI, B.; BROWN, K. Therapeutic cancer prevention: achievements and ongoing challenges - a focus on breast and colorectal cancer. **Mol Oncol**, v. 13, p. 579–590, 2019.
- SHARMA, B.; PERVUSHIN, K. Magnetic nanoparticles as in vivo tracers for alzheimer's disease. **Magnetochemistry**, v. 6, n. 1, 2020.
- SITTEL, F.; STOCK, G. Perspective: Identification of collective variables and metastable states of protein dynamics. **J Chem Phys**, v. 149, p. 150901, 2018.
- SRINIVASAN, B. Chapter three - non-equilibrium modalities of inhibition: Characterizing irreversible inhibition for the erbb receptor family members. In: LLOYD, M. (Ed.). **Modern Methods of Drug Design and Development**. [S.l.]: Academic Press, 2023, (Methods in Enzymology, v. 690). p. 85–108.
- SRIVASTAVA, I.; KOTIA, A.; GHOSH, S. K.; ALI, M. K. A. Recent advances of molecular dynamics simulations in nanotribology. **Journal of Molecular Liquids**, v. 335, p. 116154, 2021.
- STROBL, M.; GALLAHER, J.; ROBERTSON-TESSI, M.; WEST, J.; ANDERSON, A. Treatment of evolving cancers will require dynamic decision support. **Annals of Oncology**, v. 34, p. 867–884, 2023.
- STROET, M. et al. Automated topology builder version 3.0: Prediction of solvation free enthalpies in water and hexane. **J Chem Theory Comput**, v. 14, p. 5834–5845, 2018.
- SUMIT; KUMAR, A.; MISHRA, A. Advancement in pharmacological activities of benzothiazole and its derivatives: An up to date review. **Mini Rev Med Chem**, v. 21, p. 314–335, 2021.
- SWENDSEN, R.; WANG, J. Replica monte carlo simulation of spin-glasses. **Phys Rev Lett**, v. 57, p. 2607–2609, 1986.
- TORRIE, G.; VALLEAU, J. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. **Chem Phys Lett**, v. 28, p. 578–581, 1974.
- TORRIE, G.; VALLEAU, J. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. **J Comput Phys**, v. 23, p. 187–199, 1977.
- TRIBELLO, G.; BONOMI, M.; BRANDUARDI, D.; CAMILLONI, C.; BUSSI, G. Plumed 2: New feathers for an old bird. **Comput Phys Commun**, v. 185, p. 604–613, 2014.
- U. S. CENTERS FOR DISEASE CONTROL AND PREVENTION. **Cancer Data and Statistics**. [S.l.], 2023. Disponível em: <<https://www.cdc.gov/cancer/dcpc/data/index.htm>>.
- U. S. NATIONAL CANCER INSTITUTE. **What Is Cancer?** [S.l.], 2021. Disponível em: <<https://www.cancer.gov/about-cancer/understanding/what-is-cancer#definition>>.
- VELEMA, W.; SZYMANSKI, W.; FERINGA, B. Photopharmacology: Beyond proof of principle. **J Am Chem Soc**, v. 136, p. 2178–2191, 2014.

- VOGT, P.; KANG, S.; ELSLIGER, M.; GYMNOPOULOS, M. Cancer-specific mutations in phosphatidylinositol 3-kinase. **Trends Biochem Sci**, v. 32, p. 342–9, 2007.
- WANG, S. et al. Gadolinium metallofullerene-based activatable contrast agent for tumor signal amplification and monitoring of drug release. **Small**, v. 15, n. 16, 2019.
- WANG, T. et al. Tddft study on the esipt properties of 2-(2-hydroxyphenyl)-benzothiazole and sensing mechanism of a derived fluorescent probe for fluoride ion. **Molecules**, v. 29, 2024.
- WELLER, A. Über die fluoreszenz der salizylsäure und verwandter verbindungen. **Naturwissenschaften**, v. 42, p. 175–176, 1955.
- WELTER, M.; RIEGER, H. Interstitial fluid flow and drug delivery in vascularized tumors: A computational model. **PLOS ONE**, v. 8, p. 1–23, 2013.
- WOLF, S. et al. Estimation of protein–ligand unbinding kinetics using non-equilibrium targeted molecular dynamics simulations. **J Chem Inf Model**, v. 59, p. 5135–5147, 2019.
- WORLD HEALTH ORGANIZATION (WHO). **Global cancer burden growing, amidst mounting need for services**. [S.l.], 2024. Disponível em: <<https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services>>.
- WU, L.; HUANG, J.; PU, K.; JAMES, T. Dual-locked spectroscopic probes for sensing and therapy. **Nat Rev Chem**, v. 5, p. 406–421, 2021.
- WYSKOCKA-GAJDA, M. et al. A step towards gadolinium-free bioresponsive mri contrast agent. **European Journal of Medicinal Chemistry**, v. 211, p. 113086, 2021.
- YADAV, K. P. et al. Synthesis and biological activities of benzothiazole derivatives: A review. **Intelligent Pharmacy**, v. 1, p. 122–132, 2023.
- YANAI, T.; TEW, D. P.; HANDY, N. C. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). **Chemical Physics Letters**, v. 393, n. 1, p. 51–57, 2004. ISSN 0009-2614. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0009261404008620>>.
- YAP, T. et al. Targeting the pi3k-akt-mtor pathway: progress, pitfalls, and promises. **Curr Opin Pharmacol**, v. 8, p. 393–412, 2008.
- YI, Z. G. et al. In vivo tumor visualization through mri off-on switching of nagdf4-caco3 nanoconjugates. **Advanced Materials**, v. 31, n. 37, 2019.
- YU, H.; KOH, S. The role of pi3k/akt pathway and its therapeutic possibility in alzheimer’s disease. **hmr**, v. 37, p. 18–24, 2017.
- ZHANG, D.; OUYANG, Q. Nonequilibrium thermodynamics in biochemical systems and its application. **Entropy (Basel)**, v. 23, p. 271, 2021.
- ZHANG, Z. H. et al. A force field for molecular dynamics simulations of iron oxide system. **Materials Science and Engineering B-Advanced Functional Solid-State Materials**, v. 283, 2022.
- ZHOU, J. et al. Fluorescent diagnostic probes in neurodegenerative disease. **Advanced Materials**, v. 32, p. 2001945, 2020.

ZHOU, J.; MA, H. Design principles of spectroscopic probes for biological applications. **Chem Sci**, v. 10, p. 6309–6315, 2016.

6 SUPPORTING INFORMATION

6.1 Unbiased MD simulation analysis

A RMSD analysis was conducted for system 1, HABT in vacuum, and 2, HABT in water. Figure 6.1 shows the analysis performed for both systems. For system 1, it was found a mean RMSD value of 0,13 Å, and a standard deviation of 0,05 Å. Now, for system 2, it was found a mean RMSD value of 0,15 Å, and a standard deviation of 0,05 Å. The values found for both systems implies the simulation stability.

Figura 6.1 – RMSD analysis of HABT done for a) system 1, and b) system 2.

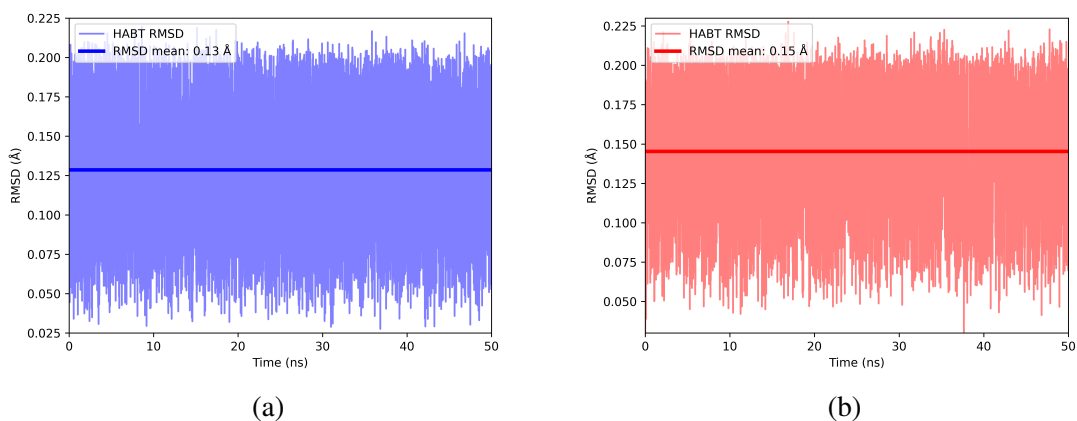


Figura 6.2 – Shows the -OH an -N= intramolecular h-bond interaction of HABT.

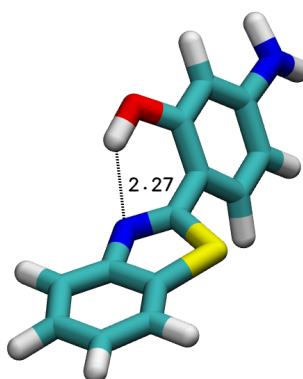
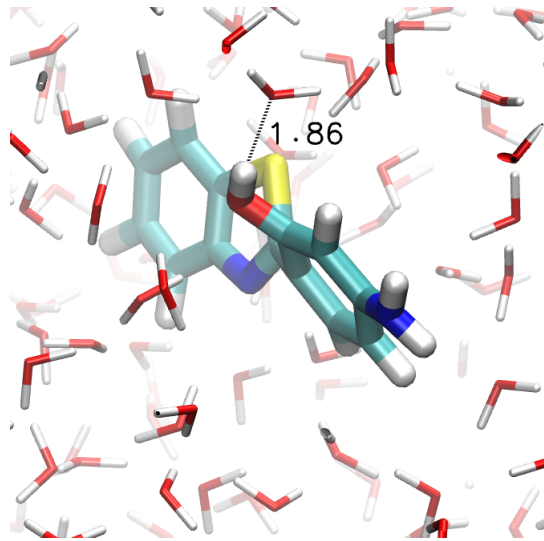
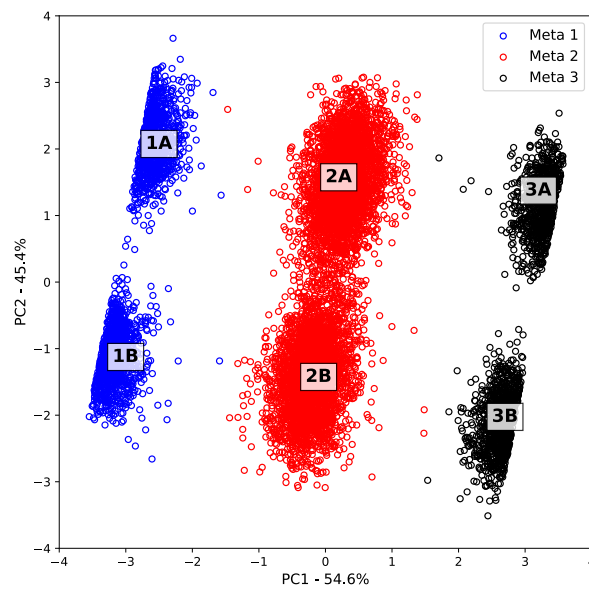


Figura 6.3 – Shows the competitive intermolecular h-bond interaction observed after solvation.



6.2 Labeling metastable states

Figura 6.4 – Shows the labels assigned for each accessed metastate.



6.3 Conformational selection

Tabela 6.1 – Table shows the mean values of dihedral 1, 2 and distance N···H.

Metastate	Dihedral 1 (rad)	Dihedral 2 (rad)	Distance N···H (Å)
1A	1,53	-2,92	4,46
1B	-1,75	-2,91	4,58
2A	1,50	-0,03	2,96
2B	-1,47	0,04	2,93
3A	1,75	2,91	4,59
3B	-1,61	2,93	4,49
EQ	0.07	-0.16	3.75