



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós Graduação em Química

Dissertação de Mestrado

ESTUDO TEÓRICO DE PROPRIEDADES DO ESTADO EXCITADO DE PROTEÍNAS FLUORESCENTES

Breno Xavier Porto Alves

Orientadora: Prof.^a Dr.^a Karen Cacilda Weber

Coorientador: Prof. Dr. Gerd Bruno Rocha

João Pessoa, Paraíba

Março de 2024



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós Graduação em Química

Dissertação de Mestrado

ESTUDO TEÓRICO DE PROPRIEDADES DO ESTADO EXCITADO DE PROTEÍNAS FLUORESCENTES

Breno Xavier Porto Alves

Dissertação submetida ao Programa de Pós Graduação em Química da Universidade Federal da Paraíba, como parte dos requisitos para obtenção do título de Mestre em Química.

Orientadora: Prof.^a Dr.^a Karen Cacilda Weber

Coorientador: Prof. Dr. Gerd Bruno Rocha

João Pessoa, Paraíba

Março de 2024

Catálogo na publicação
Seção de Catalogação e Classificação

A474e Alves, Breno Xavier Porto.

Estudo teórico de propriedades do estado excitado de proteínas fluorescentes / Breno Xavier Porto Alves. - João Pessoa, 2024.

53 f. : il.

Orientação: Karen Cacilda Weber.

Coorientação: Gerd Bruno Rocha.

Dissertação (Mestrado) - UFPB/CCEN.

1. Proteínas fluorescentes. 2. Estados excitados. 3. GFP. 4. TD-DFT. 5. ONIOM. I. Weber, Karen Cacilda. II. Rocha, Gerd Bruno. III. Título.

UFPB/BC

CDU 547(043)

Estudo teórico de propriedades do estado excitado de proteínas fluorescentes.

Dissertação de Mestrado apresentada pelo aluno **Breno Xavier Porto Alves** e aprovada pela banca examinadora em 09 de abril de 2024.

Documento assinado digitalmente
 **KAREN CACILDA WEBER**
Data: 21/04/2024 23:04:54-0300
Verifique em <https://validar.itl.gov.br>

Profa. Dra. Karen Cacilda Weber
DQ-UFPB
Orientadora/Presidente

Documento assinado digitalmente
 **GERD BRUNO DA ROCHA**
Data: 12/04/2024 15:33:15-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Gerd Bruno da Rocha
DQ-UFPB
Co-Orientador

Documento assinado digitalmente
 **PAULA HOMEM DE MELLO**
Data: 10/04/2024 10:00:02-0300
Verifique em <https://validar.itl.gov.br>

Profa. Dra. Paula Homem de Mello
UFdoABC/Santo André-SP
Examinadora externa

Documento assinado digitalmente
 **OTAVIO LUIS DE SANTANA**
Data: 09/04/2024 19:43:20-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Otávio Luís de Santana
DQ-UFPB
Examinador interno

À minha família.

AGRADECIMENTOS

Aos meus pais e irmãos, por me incentivarem a trilhar o caminho dos estudos, por acreditarem em meus sonhos, por nunca deixarem me faltar amor, acolhimento e colo, que tantas vezes foram necessários para que eu me mantivesse firme nessa jornada.

À minha orientadora, Prof^a Dr^a Karen Weber, por acreditar que eu daria conta dos desafios da área, pelos ensinamentos, pela paciência, pela parceria, pelos conselhos, pelos *happy hours*, pela preocupação, pela forma humana de lidar com as adversidades e por não ter desistido de mim quando eu já havia o feito. Sem você, eu jamais concluiria esse mestrado.

Ao meu coorientador, Prof. Dr. Gerd Rocha, nosso mestre Yoda, pelos vastos conhecimentos e ideias compartilhadas e por muitas vezes ter sido o oráculo a ser consultado em momentos de dificuldades, sempre trazendo clareza aos problemas que surgiam.

Ao Prof. Dr. Otávio e Prof. Dr. Wagner por terem aceitado participar da banca de pré-defesa e por suas valiosas contribuições ao trabalho.

Ao Programa de Pós-Graduação em Química da UFPB, que através de corpo docente e administrativo, me proporcionou uma memorável vivência científica e profissional rica de aprendizados.

Aos colegas do Laboratório de Química Quântica e Computacional (LQQC), por todos os momentos de aprendizado e descontração. Gratidão especial a Igor, pela força com o PRIMoRDIA.

Ao CNPq pela bolsa de mestrado concedida.

Aos meus amigos, de longa data e os que fiz nesse percurso, por ouvirem minhas lamentações frequentes, por terem trazido leveza a essa fase da minha vida e por constantemente me resgatarem de mim mesmo.

RESUMO

Neste trabalho a proteína fluorescente verde (GFP) foi escolhida como objeto de estudo devido às características fotofísicas de seu cromóforo a fim de investigar propriedades de estado excitado em sistemas biomoleculares. Sendo assim, buscou-se obter dados de energia de excitação para estruturas do cromóforo da proteína e da proteína otimizada em nível PM6/Amber com o método ONIOM, além de estruturas representativas de uma trajetória de dinâmica molecular de 200 ns. Nestas estruturas, foram realizados cálculos single-point com o método TD-DFT variando os funcionais, conjuntos de funções de base e tamanho das camadas tratadas com o método ONIOM. Com o funcional e conjunto de base escolhido, foram calculadas as energias de excitação para cromóforos de proteínas fluorescentes variantes da GFP. As análises estatísticas de RMSD e RMSF indicaram estruturas com baixas flutuações em torno do cromóforo. Os funcionais que desempenharam a melhor performance foram o CAM-B3LYP e ω B97XD com o conjunto de bases 6-311+G*. Os dados obtidos por cálculos com TD-DFT na abordagem ONIOM são bastante próximos dos valores reportados na literatura, apresentando variações significativas para algumas estruturas oriundas da dinâmica molecular. No entanto, a média dos valores de energia de excitação das estruturas significativas de cada cluster da trajetória de dinâmica molecular são mais próximos aos valores experimentais. Não se verificou influência significativa da proteína sobre os níveis de energia do cromóforo da GFP. Quanto aos cromóforos análogos, as energias de excitação seguem a tendência da literatura, embora os valores calculados tenham sido superestimados em relação aos dados experimentais.

Palavras-chave: proteínas fluorescentes; estados excitados; GFP; TD-DFT; ONIOM.

ABSTRACT

In this work, green fluorescent protein (GFP) was chosen as the object of study due to the photophysical characteristics of its chromophore in order to investigate excited state properties in biomolecular systems. Therefore, the goal was to obtain excitation energy data for the chromophore structures of the protein and the protein optimized at the PM6/Amber level using the ONIOM method, as well as representative structures from a 200 ns molecular dynamics trajectory. Single-point calculations were performed on these structures using the TD-DFT method, varying the functionals, basis sets, and the size of the layers treated with the ONIOM method. With the chosen functional and basis set, excitation energies were calculated for chromophores of fluorescent protein variants of GFP. Statistical analyses of RMSD and RMSF indicated structures with low fluctuations around the chromophore. The functionals that performed best were CAM-B3LYP and ω B97XD with the 6-311+G* basis set. The data obtained from TD-DFT calculations using the ONIOM approach are quite close to the values reported in the literature, showing significant variations for some structures arising from molecular dynamics. However, the average excitation energy values of the significant structures from each cluster of the molecular dynamics trajectory are closer to the experimental values. No significant influence of the protein on the energy levels of the GFP chromophore was observed. Regarding analogous chromophores, the excitation energies follow the trend of the literature, although the calculated values were overestimated compared to the experimental data.

Keywords: Fluorescent proteins; excited states; GFP; TD-DFT; ONIOM.

LISTA DE FIGURAS

- Figura 1.** Estrutura da GFP, com o cromóforo representado em licorice, gerada pelo VMD. 14
- Figura 2.** Representação dos resíduos que formam o cromóforo da GFP, antes e depois da ciclização. 14
- Figura 3.** Exemplos de cromóforos de proteínas fluorescentes, com respectivos comprimentos de onda de excitação (barra superior) e de emissão (barra inferior). Fonte: (ACHARYA, A. et al., 2017). 15
- Figura 4.** Representação do ciclo fotocatalítico de uma proteína fluorescente. A sigla GS se refere ao estado fundamental (do inglês, Ground State) e a sigla EES se refere ao estado eletronicamente excitado (do inglês, Excited Electronic State). Fonte: adaptado de Acharya et al (2017). 16
- Figura 5.** Caixa de simulação para a dinâmica molecular. Ao centro da caixa está representada a estrutura da GFP. Ao redor da proteína estão distribuídas as moléculas de água, em transparente, e os íons Cl⁻, em laranja, e K⁺, em verde. 28
- Figura 6.** Modelos de separação das regiões QM (em licorice) e MM da proteína para os cálculos com o método ONIOM. 30
- Figura 7.** Estrutura dos cromóforos de proteínas variantes da GFP. 32
- Figura 8.** (a) RMSD e (b) RMSF da etapa de produção da dinâmica molecular em triplicata. 34
- Figura 9.** Desvio relativo da energia de excitação do cromóforo da GFP no vácuo, em relação ao valor experimental de 3,14 eV, para diferentes funcionais. 36
- Figura 10.** Desvio relativo da energia de excitação do cromóforo da GFP com a estrutura cristalográfica, em relação ao valor experimental de 3,14 eV, para diferentes funcionais. 37
- Figura 11.** Desvio relativo da energia de excitação do cromóforo da GFP da estrutura oriunda da dinâmica molecular, em relação ao valor experimental de 3,14 eV, para diferentes funcionais. 37
- Figura 12.** Densidade de estados de energia para os orbitais moleculares da GFP, para: (a) o cromóforo isolado; (b) uma região de 5 Å da proteína centrada no cromóforo; (c) uma região de 10 Å da proteína centrada no cromóforo; (d) uma região de 15 Å da proteína centrada no cromóforo; e (e) a proteína inteira. 42

Figura 13. Energias de excitação vertical para cada proteína fluorescente.43

LISTA DE TABELAS

Tabela 1. Estruturas utilizadas para cálculos de potenciais híbridos QM/MM.	29
Tabela 2. Valores da energia total, temperatura média e pressão média após cada etapa de equilíbrio da dinâmica molecular.	33
Tabela 3. Clusters obtidos e a quantidade de frames agrupados em cada um.	35
Tabela 4. Valores obtidos para a energia de excitação, ΔE , em eV, e o desvio percentual, δ , para as estruturas A, B_1 e C0_1, calculados com diferentes conjuntos de funções de base, levando em consideração o acoplamento eletrônico.	39
Tabela 5. Valores obtidos para a energia de excitação, ΔE , em eV, e o desvio percentual, δ , para as estruturas B e C0 com diferentes modelos de particionamento QM/MM, com o método QM CAM-B3LYP/6-311+G* e acoplamento eletrônico.....	40
Tabela 6. Valores obtidos para a energia de excitação, ΔE , em eV, e o desvio percentual médio, $\bar{\delta}$, para as estruturas Cy com diferentes modelos de particionamento QM/MM, com o método QM CAM-B3LYP/6-311+G* e acoplamento eletrônico.....	41

Sumário

Capítulo 1: INTRODUÇÃO	10
Capítulo 2: OBJETIVOS	12
2.1 Objetivo Geral	12
2.2 Objetivos específicos	12
Capítulo 3: REFERENCIAL TEÓRICO	13
3.1 Proteína fluorescente verde	13
3.2 Cálculos de estados excitados e TD-DFT	17
3.3 O método ONIOM	19
3.3.1 Acoplamento mecânico e eletrônico	21
3.3.2 Fronteira entre as camadas QM e MM	23
3.4 Dinâmica Molecular.....	24
Capítulo 4: METODOLOGIA	27
Capítulo 5: RESULTADOS E DISCUSSÃO	33
Capítulo 6: CONCLUSÕES	46
REFERÊNCIAS	47

Capítulo 1: INTRODUÇÃO

Frente aos desafios apresentados pela pandemia da COVID-19, diversas áreas da ciência se empenharam na compreensão do vírus SARS-CoV-2, a fim de desenvolver tratamentos, fármacos, vacinas e métodos de detecção do material viral. A respeito da detecção de vírus, técnicas espectroscópicas como a FRET (do inglês, *Förster Resonance Energy Transfer*), quando aplicadas a sistemas biológicos, podem ser bastante úteis, além de permitir localizar células de interesse e monitorar eventos intracelulares.

No contexto de combate ao vírus, a química computacional contribui especialmente na triagem virtual de fármacos e no estudo detalhado de proteínas e ácidos nucleicos fundamentais para a replicação do vírus. Além disso, estudos teóricos podem permitir avanços na caracterização de sistemas biológicos em estado excitado, possibilitando, por exemplo, o planejamento de sondas espectroscópicas com propriedades adequadas para finalidades específicas.

A importância de estados eletronicamente excitados (EES, do inglês *Electronically Excited States*) em ciência e tecnologia vem aumentando consideravelmente nos últimos anos. Levar em consideração a influência desses estados sobre as propriedades da matéria é de suma importância para a elucidação de diversos fenômenos da natureza, tais como a fluorescência e a fosforescência, a conversão de energia em células solares, a terapia fotodinâmica e processos biológicos (JACQUEMIN; MENNUCI; ADAMO, 2011). Neste sentido, algumas sondas espectroscópicas adotadas em experimentos de FRET, como a proteína fluorescente verde (GFP, do inglês *Green Fluorescent Protein*), se apresentam como modelos de interesse para investigar propriedades do estado excitado em sistemas biológicos (ACHARYA *et al.*, 2017).

Sendo assim, uma proposta de pesquisa relevante nessa temática consiste na elaboração de um modelo de detecção do RNA do SARS-CoV-2 através de sondas

fluorescentes, por meio da técnica FRET, a partir de um sistema formado por proteínas que interagem com o RNA viral. No entanto, um sistema desse tipo apresenta alta complexidade e, diante disso, a GFP se apresenta como um bom modelo de estudo para avaliar a viabilidade técnica do método proposto, devido às propriedades fotofísicas únicas de seu cromóforo (TSIEN, 1998).

Essa proteína desperta especial interesse nas ciências da vida, visto que há numerosos estudos experimentais e teóricos motivados pela fotofísica das proteínas dessa classe (WACHTER, 2006; MEECH, 2009; VAN THOR, 2009; SAMPLE; NEWMAN; ZHANG, 2009; NEMUKHIN; GRIGORENKO; SAVITSKY, 2009; SEWARD; BAGSHAW, 2009; LUKYANOV et al., 2010; BRAVAYA et al., 2012; SUBACH; VERKHUSHA, 2012). Devido à complexidade do sistema, muitos aspectos do fotociclo das proteínas fluorescentes e da formação dos respectivos cromóforos permanecem inexplorados. No entanto, a compreensão desses processos em nível molecular oferece uma vantagem crucial na modelagem de novas proteínas fluorescentes com propriedades para atender a diversas finalidades (ACHARYA et al., 2017).

Nesses estudos, os comprimentos de onda de absorção e fluorescência estão entre os principais parâmetros investigados. Proteínas fluorescentes vermelhas são de particular importância como marcadores adequados para imagens de tecidos profundos (SHCHERBAKOVA; VERKHUSHA, 2014). Além disso, proteínas fluorescentes de cores diferentes podem ser usadas como marcadores para diferentes proteínas (imagens multicoloridas) e para construir pares para espectroscopia FRET (GANESAN et al., 2006; PETTIKIRIARACHCHI et al., 2012).

O presente trabalho adota a GFP como modelo de estudo para avaliar a viabilidade de modelagens computacionais de sistemas biológicos mais complexos em estado excitado.

Capítulo 2: OBJETIVOS

2.1 Objetivo Geral

Estabelecer um protocolo computacional para determinar propriedades do EES para proteínas fluorescentes, por meio de dinâmica molecular e cálculos de potencial híbrido QM/MM usando a TD-DFT.

2.2 Objetivos específicos

- Verificar a estabilidade do modelo por meio de análises da trajetória da dinâmica molecular;
- Realizar uma amostragem conformacional da estrutura da proteína ao longo da dinâmica molecular;
- Escolher o melhor funcional da DFT para cálculos de potencial híbrido QM/MM;
- Obter valores da energia de excitação para modelos da proteína e cromóforo empregando o método ONIOM;
- Investigar a influência do ambiente da proteína sobre os valores de energia de excitação do cromóforo;
- Calcular a energia de excitação de cromóforos análogos ao da GFP.

Capítulo 3: REFERENCIAL TEÓRICO

3.1 Proteína fluorescente verde

As propriedades da proteína verde fluorescente (GFP) revolucionaram muitas áreas das ciências da vida, possibilitando observações *in vivo* da localização e de interações em proteínas, medições intracelulares de concentrações de íons fisiologicamente importantes, mapeamento de expressões genéticas, entre outros (TSIEN, 1998; DAY; DAVIDSON, 2009; ZIMMER, 2009; CHUDAKOV et al., 2010; MISHIN et al., 2015). A importância das proteínas fluorescentes e das tecnologias relacionadas foi reconhecida com o Prêmio Nobel de Química de 2008.

A GFP foi caracterizada pela primeira vez em nível proteico em extratos da água-viva *Aequorea victoria* em 1962 (JOHNSON et al., 1962). Foram necessários mais de 30 anos para que fosse possível clonar o gene da GFP e demonstrar que a GFP funcional pode ser expressa em vários organismos (PRASHER et al., 1992; CHALFIE et al., 1994). Essa descoberta inaugurou a era das aplicações da GFP como um marcador fluorescente totalmente codificado por um único gene. As proteínas naturais semelhantes à GFP demonstram uma ampla diversidade espectral, incluindo proteínas fluorescentes ciano, verde, amarela, laranja e vermelha. Análises filogenéticas e reconstrução de genes ancestrais mostraram que o fenótipo fluorescente verde provavelmente era característico de proteínas evolutivamente antigas, enquanto outras cores apareceram mais tarde na evolução, independentemente.

Sua estrutura conta com 238 aminoácidos distribuídos em 11 folhas- β que formam um cilindro atravessado longitudinalmente por uma hélice- α , onde se localiza o cromóforo da proteína, como pode ser observado na Figura 1. O β -barril tem diâmetro aproximado de 24 Å e altura de 42 Å, com peso molar entre 25 kDa e 30 kDa (TSIEN, 1998).

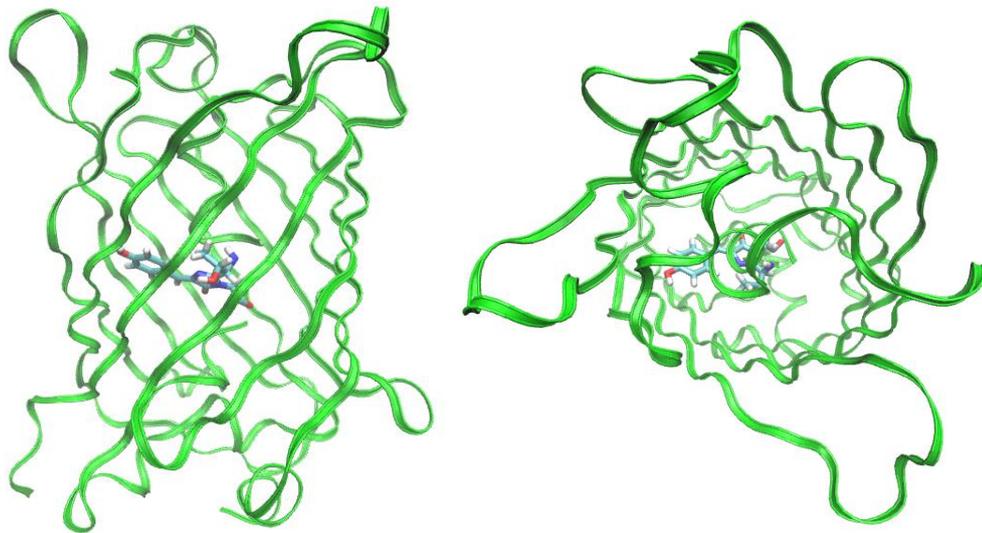


Figura 1. Estrutura da GFP, com o cromóforo representado em licorice, gerada pelo VMD.

É reportado que a GFP do tipo selvagem absorve luz majoritariamente em 395 nm, com um pequeno pico em 475 nm, devido aos estados de protonação do cromóforo (neutro e aniônico, respectivamente), e emite em torno de 508 nm (PATKNAIK; TROHALAKI; PACHTER, 2004; MA et al., 2012). O cromóforo, 4-(p-hidroxibenzilideno)-imidazolidin-5-ona (HBI), é formado a partir de uma reação de ciclização autocatalítica entre três resíduos do *backbone* da proteína (Ser65, Tyr66 e Gly67), seguida de uma reação de oxidação, sem necessidade de cofatores. Os resíduos pré-ciclização e pós formação do cromóforo podem ser observados no esquema apresentado na Figura 2.

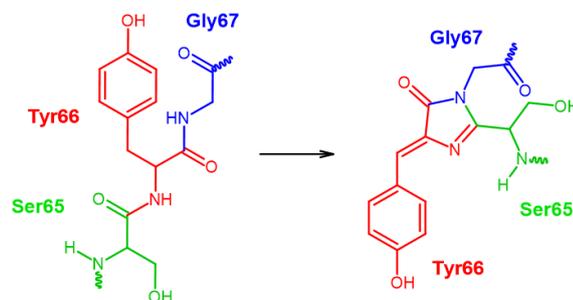


Figura 2. Representação dos resíduos que formam o cromóforo da GFP, antes e depois da ciclização.

Devido à complexidade do sistema, muitos aspectos do ciclo fotoquímico das proteínas fluorescentes e formação do cromóforo ainda estão em grande parte inexplorados. No entanto, o entendimento em nível molecular desses processos oferece uma vantagem crucial na modelagem de novas proteínas fluorescentes com propriedades adequadas para aplicações específicas, sendo os comprimentos de onda de absorção e emissão os parâmetros mais importantes.

Proteínas fluorescentes de diferentes cores podem ser usadas para marcar diferentes proteínas e para construir pares de FRET. Atualmente, as proteínas fluorescentes abrangem toda a faixa de luz visível, incluindo o extremo vermelho do espectro, conforme ilustrado na Figura 3.

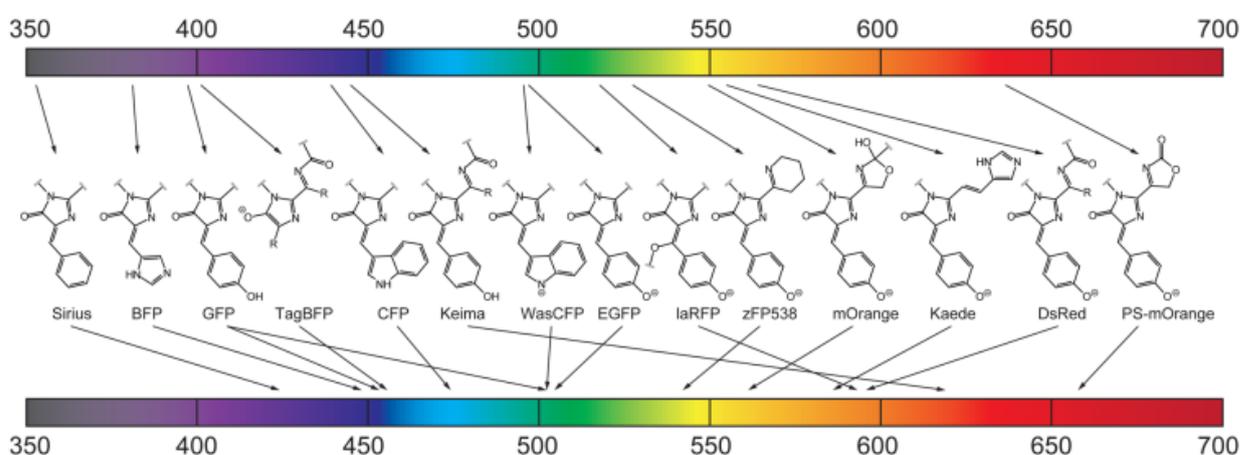


Figura 3. Exemplos de cromóforos de proteínas fluorescentes, com respectivos comprimentos de onda de excitação (barra superior) e de emissão (barra inferior). Fonte: (ACHARYA, A. et al., 2017).

O ajuste de cor em proteínas fluorescentes pode ser alcançado por vários mecanismos distintos, incluindo a variação da extensão do sistema π , mudança do estado de protonação do cromóforo, empilhamento π , interações eletrostáticas e outras interações específicas com resíduos próximos. O brilho é outro fator importante: proteínas fluorescentes mais brilhantes, ou seja, aquelas com maiores coeficientes de extinção e rendimentos quânticos de fluorescência, funcionam melhor como marcadores fluorescentes (ACHARYA, A. et al., 2017).

A fotofísica e fotoquímica das proteínas fluorescentes têm considerável semelhança com as dos corantes sintéticos. Do ponto de vista químico, os cromóforos típicos das proteínas fluorescentes (Figura 3) são semelhantes aos corantes cianina,

devido a uma característica estrutural comum: uma ponte metina conectando grupos aromáticos conjugados (BEREZIN; ACHILEFU, 2010; HA; TINNEFELD, 2012). No entanto, a presença do barril protéico leva a diferenças significativas. O ambiente protéico rígido restringe o alcance de movimento do cromóforo e limita sua acessibilidade ao solvente e a outras espécies presentes na solução (oxigênio ambiente, íons de sal, agentes oxidantes e redutores, etc). De fato, as propriedades fotofísicas dos cromóforos modelo em solução diferem drasticamente daquelas das respectivas proteínas fluorescentes: os cromóforos solvatados não fluorescem, sendo observada uma rápida recuperação do estado fundamental após a excitação, na escala de tempo de picossegundos (LIRVINENKO; WEBBER; MEECH, 2001; MARTIN; NEGRI; OLIVUCCI, 2004).

O ciclo fotocatalítico de uma proteína fluorescente é iniciado pela absorção de luz, produzindo um estado eletronicamente excitado inicial do cromóforo. A Figura 4 esboça vários processos no estado excitado em proteínas fluorescentes.

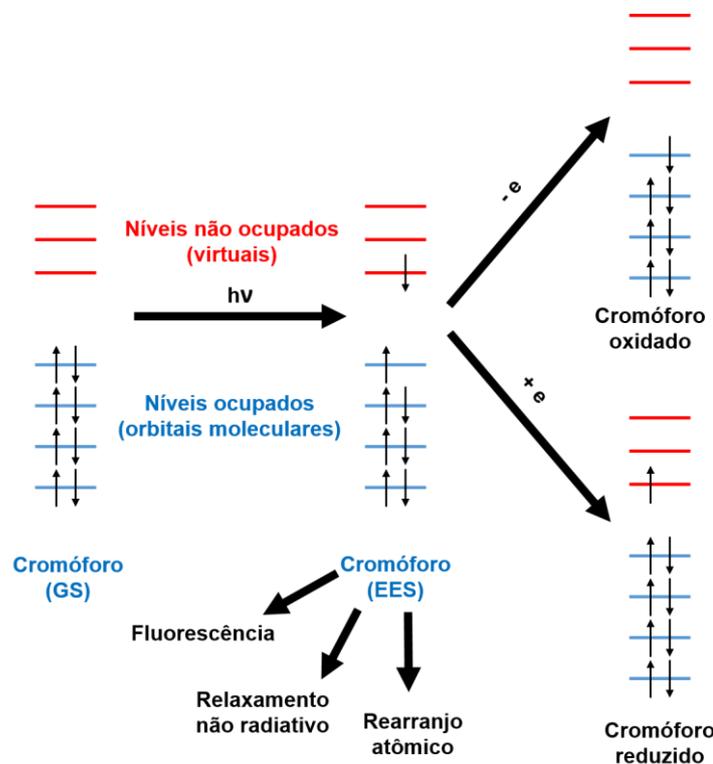


Figura 4. Representação do ciclo fotocatalítico de uma proteína fluorescente. A sigla GS se refere ao estado fundamental (do inglês, *Ground State*) e a sigla EES se refere ao estado eletronicamente excitado (do inglês, *Excited Electronic State*). Fonte: adaptado de Acharya et al (2017).

O principal canal de relaxamento que restaura o cromóforo no estado fundamental é a fluorescência. Alternativamente, o cromóforo pode retornar ao estado fundamental dissipando a energia eletrônica em movimentos nucleares via relaxamento não radiativo. Uma vez que o padrão de ligação nos estados excitados é diferente, a excitação eletrônica pode iniciar várias transformações químicas do cromóforo, como isomerização, formação ou quebra de ligações covalentes, foto-oxidação/foto-redução ou reações com resíduos próximos ou pequenas moléculas (por exemplo, oxigênio molecular). As escalas de tempo dos diferentes canais são crucialmente importantes para entender as proporções e rendimentos de ramificação. Um tempo de vida finito no estado excitado limita o escopo dos processos que podem ocorrer a partir deste estado. Típico de sistemas fluorescentes, os tempos de vida no estado excitado em proteínas fluorescentes são de 1 a 10 nanossegundos. Portanto, para ter um efeito perceptível no ciclo fotocatalítico, um processo no estado excitado deve ser iniciado em uma escala de tempo comparável ao tempo de vida do estado excitado (ACHARYA, A. et al., 2017).

O cromóforo da GFP tem sido estudado extensivamente, tanto por métodos experimentais como métodos teóricos (WEBBER; LITVINENKO; MEECH, 2001; HELMS, 2002; ZIMMER, 2002; PATNAIK; TROHALAKI; PACHTER, 2004; REMINGTON, 2006; MATSUURA et al., 2010; MA et al., 2012; THOMPSON, 2014; PETRONE et al., 2016; JIN; GLOVER; HE, 2020). Há estudos reportados na literatura de cálculos estáticos de energias de excitação vertical, interseções cônicas de energia mínima e trajetória de reação usando métodos *ab initio* (MARTIN; NEGRI; OLIVUCCI, 2004; SINICROPI et al., 2005), teoria do funcional da densidade (MARQUES et al., 2003; LAINO; NIFOSI; TOZZINI, 2004; PATNAIK et al., 2007) e métodos semiempíricos (VOITYUK; MICHEL-BEYERLE; ROSCH, 1998; WEBER et al., 1999; VOITYUK et al., 2001; TONIOLO; GRANUCCI; MARTINEZ, 2003).

3.2 Cálculos de estados excitados e TD-DFT

A Teoria do Funcional da Densidade Dependente do Tempo (TD-DFT, do inglês *Time Dependent Density Functional Theory*) é provavelmente a abordagem teórica

mais amplamente utilizada para calcular energias de transição e propriedades de estados excitados, como momentos de dipolo e energias de excitação vertical. A popularidade dessa abordagem se deve ao fato de suas concorrentes exigirem alta demanda computacional. Os métodos semiempíricos tendem a fornecer resultados extremamente dependentes do sistema, enquanto que as abordagens *ab initio* geralmente são muito mais exatas que a TD-DFT, mas sofrem de um custo computacional extremamente elevado (JACQUEMIN; MENNUCCI; ADAMO, 2011).

Em cálculos de TD-DFT é preciso levar em consideração os efeitos do ambiente, uma vez que as propriedades dos estados eletronicamente excitados são muito mais dependentes desses efeitos que suas contrapartes do estado fundamental. Salvo os casos de estruturas microsolvatadas, que podem ser calculadas por completo apenas com TD-DFT, um sistema pode ser dividido em duas partes, onde o grupo cromofórico é simulado com TD-DFT e o ambiente é tratado em um nível de teoria mais simplificado (ADAMO; JACQUEMIN, 2013). A vizinhança de um cromóforo pode ser apenas o solvente ou um ambiente químico complexo, como uma proteína.

Para o caso do solvente, há duas abordagens principais para estimar o efeito sobre o grupo cromofórico de interesse: modelos explícitos e modelos implícitos. Enquanto no primeiro, todas as moléculas do solvente são tratadas explicitamente, geralmente com abordagens de mecânica molecular, no segundo, todo o solvente é tratado como um contínuo sem estrutura visando a reprodução de propriedades macroscópicas (tal como a sua constante dielétrica). Por um lado, os modelos explícitos tendem a fornecer uma descrição completa dos efeitos do solvente a um custo computacional mais elevado, enquanto que os modelos implícitos requerem menos custo computacional, mas são capazes de fornecer apenas uma média dos efeitos das interações entre soluto e solvente (ADAMO; JACQUEMIN, 2013).

Para o caso de ambientes mais complexos, o desafio está em reproduzir corretamente as interações entre o cromóforo e a vizinhança, visto que essas interações podem ser consideradas específicas ou não específicas. O primeiro caso ocorre quando o ambiente participa diretamente da excitação eletrônica, como numa transferência de carga entre orbitais localizados no cromóforo e orbitais localizados na vizinhança, enquanto que no segundo caso o ambiente apenas exerce o papel de polarizar a densidade eletrônica das moléculas excitadas (ADAMO; JACQUEMIN, 2013).

Além disso, os efeitos da vizinhança podem ser considerados diretos, limitados à polarização da densidade eletrônica do cromóforo, ou indiretos, quando há modificações estruturais induzidas pelas interações entre o cromóforo e o ambiente. Desde que a excitação possa ser considerada localizada no cromóforo, uma abordagem eficaz é tratar as duas regiões, cromóforo e ambiente, com modelos diferentes, onde a região de interesse pode ser tratada com TD-DFT e o ambiente com métodos de mecânica molecular ou semiempíricos, por exemplo (ADAMO; JACQUEMIN, 2013).

Nesses modelos, os efeitos indiretos já são incluídos considerando um acoplamento mecânico, que implicam em restrições estruturais do ambiente na geometria do cromóforo. No entanto, os efeitos diretos são levados em consideração apenas se os dois níveis do modelo (QM e MM ou QM e QM) forem acoplados por meio de um termo eletrônico, adicionando a perturbação induzida pelo ambiente, com termos e aproximações adequadas, no Hamiltoniano do cromóforo. Dentre as diversas abordagens QM/QM ou QM/MM, o ONIOM permite levar em consideração o acoplamento eletrônico de forma eficaz (ADAMO; JACQUEMIN, 2013).

3.3 O método ONIOM

Calcular com exatidão a estrutura e as propriedades de sistemas moleculares grandes e complexos usando apenas métodos QM é difícil e extremamente demorado. Muitos esforços foram feitos para tornar tais cálculos viáveis, a exemplo dos métodos híbridos. Em um método híbrido, uma grande molécula é dividida em vários fragmentos. Diferentes métodos teóricos, com diferenças em custo computacional e poder de previsão, são então aplicados aos diferentes fragmentos (CHUNG et al., 2015).

O ONIOM pode ser considerado como um método híbrido baseado em um conceito um pouco diferente do método QM/MM de duas camadas. Embora o ONIOM possa ser usado como um método QM/MM de duas camadas, ele também pode, de forma única, combinar diferentes métodos QM e pode ser facilmente estendido para múltiplas camadas (SENN; THIEL, 2007).

Maseras e Morokuma (1995) propuseram e desenvolveram um esquema alternativo QM/MM, denominado Método Integrado de Orbital Molecular + Mecânica Molecular (IMOMM). O IMOMM é um método "subtrativo" ou "extrapolativo": a energia total de todo o sistema ("real") é avaliada como a energia QM da região de interesse ("modelo") ($E_{QM,modelo}$) mais a energia MM do sistema real ($E_{MM,real}$), descontando-se a energia MM do sistema modelo ($E_{MM,modelo}$):

$$E_{IMOMM} = E_{ONIOM2(QM:MM)} = E_{QM,modelo} + E_{MM,real} - E_{MM,modelo} \quad (1)$$

A operação subtrativa remove as contribuições MM duplamente contabilizadas para a região modelo (MASERAS; MOROKUMA, 1995; MATSUBARA; SIEBER; MOROKUMA, 1996). O método IMOMM pode ser considerado como uma extrapolação de tamanho a partir do cálculo QM preciso do pequeno sistema modelo ($E_{QM,modelo}$) para o grande sistema real ($E_{QM,real}$), adicionando o efeito aproximado do ambiente químico avaliado pelo método MM de menor custo. De um ponto de vista alternativo, o método IMOMM também pode ser considerado como a extrapolação do método (ou seja, $E_{QM,modelo} - E_{MM,modelo}$) do nível MM para o nível QM de maior exatidão a partir da energia MM do sistema real, $E_{MM,real}$ (HUMBEL; SIEBER; MOROKUMA, 1996).

Comparado ao esquema aditivo QM/MM a ausência do Hamiltoniano de acoplamento QM-MM adicional (E_{QM-MM}) no método extrapolativo IMOMM facilita sua implementação e evita a subcontagem ou supercontagem das interações QM-MM. Diferentemente do esquema aditivo, cálculos de MM do sistema modelo são necessários no método IMOMM (ou posteriormente denominado ONIOM2(QM:MM)). Portanto, é necessário um bom conjunto de parâmetros de campo de força de MM para o sistema modelo. No entanto, devido ao cancelamento de termos MM aditivos, apenas parâmetros não ligados para o sistema modelo, como parâmetros de van der Waals e cargas atômicas (no caso do acoplamento mecânico), bem como parâmetros ligados para a fronteira QM/MM, são importantes (VREVEN et al., 2006).

Em geral, no método ONIOM de duas camadas a energia de alto nível do sistema real, $E_{alto,real}$, que é a energia alvo, poderia ser aproximada por

$$E_{alto,real} \approx E_{ONIOM2(alto:baixo)} = E_{alto,modelo} + E_{baixo,real} - E_{baixo,modelo} \quad (2)$$

É importante destacar que a equação (2) é uma generalização da equação (1), no sentido que o método “baixo” pode ser um método QM de menor custo computacional.

O método ONIOM é aplicado para obter uma diferença de energia relativa confiável (entre estruturas ou estados), que são quimicamente e fisicamente mais importantes do que a energia absoluta do sistema (CHUNG et al., 2015).

3.3.1 Acoplamento mecânico e eletrônico

Um esquema de acoplamento mecânico (ME, do inglês *Mechanical Embedding*) foi implementado pela primeira vez no método ONIOM original (ONIOM-ME). Este esquema é o método mais simples e menos exigente computacionalmente para tratar as interações eletrostáticas QM/MM de forma clássica, utilizando cargas pontuais atômicas fixas no sistema modelo e no sistema real. No entanto, a polarização da função de onda QM pelo ambiente MM não é incluída. Portanto, a precisão das interações eletrostáticas QM/MM, que podem ser críticas para a energia total do sistema real, depende fortemente das cargas atômicas do sistema modelo, que devem ser definidas pelo usuário. A diferença de energia entre estados eletrônicos distintos, bem como diferentes geometrias (como reagente e estado de transição), é mais sensível à polarização da função de onda QM, e a realização de um cálculo de energia no esquema de acoplamento eletrônico (EE, do inglês *Electronic Embedding*) do ONIOM é recomendada (CHUNG et al., 2015).

Para melhorar as interações eletrostáticas QM/MM e incluir a polarização da função de onda QM pelo ambiente MM, um termo de um elétron descrevendo as cargas atômicas pode ser adicionado ao Hamiltoniano QM que é aplicado ao sistema modelo. Neste caso, a energia do sistema inteiro é escrita como mostrada na equação 3 (VREVEN et al., 2006):

$$E_{\text{ONIOM(QM:MM)-EE}}^{\text{V}} = E_{\text{QM,modelo}}^{\text{V}} + E_{\text{MM,real}} - E_{\text{MM,modelo}}^{\text{V}} \quad (3)$$

onde o hamiltoniano é expresso como:

$$H_{\text{QM,modelo}}^{\text{V}} = H_{\text{QM,modelo}} - \sum_I \sum_N \frac{s_N q_N}{R_{IN}} + \sum_J \sum_N \frac{z_J s_N q_N}{R_{JN}} \quad (4)$$

é usado para o cálculo QM do sistema modelo. N , J e I denotam o átomo MM, átomos QM e elétrons QM, respectivamente, e q_N e s_N são a carga embutida e um fator de escala para o átomo MM N ; Z_j é a carga nuclear do átomo QM J . Aqui, as interações eletrostáticas QM/MM (com base em cargas pontuais MM fixas) e a densidade QM são computadas autoconsistentemente para a região QM. Este é o chamado esquema EE. Deve-se notar que a polarização do ambiente MM causada pela região QM não é considerada no esquema EE.

Existem dois problemas inevitáveis de fronteira QM-MM para o esquema EE: (1) superpolarização da função de onda QM por cargas MM próximas e (2) subcontagem ou supercontagem de interações eletrostáticas QM/MM.

Para evitar a superpolarização do sistema modelo QM, as cargas atômicas do ambiente MM dentro de três ligações da fronteira QM/MM são removidas nos dois cálculos modelo no ONIOM-EE padrão. Notadamente, mesmo que as cargas MM ao redor da fronteira sejam excluídas nos dois cálculos modelo ONIOM, as interações eletrostáticas ausentes com essas cargas MM ainda são consideradas por uma equação de Coulomb clássica no cálculo do sistema real ($E_{MM,real}$) no ONIOM-EE. Isso é único para o ONIOM-EE e pode remediar tanto a superpolarização quanto os problemas de subcontagem/supercontagem. Outra diferença importante entre os métodos QM/MM-EE e ONIOM-EE é que as cargas atômicas para o sistema modelo devem ser adicionadas usando o ONIOM (VREVEN; MOROKUMA, 2003; VREVEN et al., 2006).

O segundo problema se deve ao fato de que interações eletrostáticas entre átomos próximos são implicitamente incluídas nos termos parametrizados de estiramento, dobramento e diédrico de campos de força biomoleculares gerais. Vreven *et al.* (2003) desenvolveram e também implementaram o ONIOM-(QM:MM)-EE no programa Gaussian. O método ONIOM-(QM:MM)-EE, equação 3, é diferente dos esquemas QM/MM-EE aditivos. As cargas pontuais MM são incorporadas aos dois cálculos modelo ($E_{QM,modelo}^V$ e $E_{MM,modelo}^V$, equação 3), no esquema ONIOM-EE, enquanto as cargas pontuais MM são incorporadas apenas ao cálculo modelo QM e o termo eletrostático QM/MM clássico é removido no esquema QM/MM-EE genérico (EUCHINGER et al., 1999; DAS et al., 2002; KÖNIG et al., 2005).

3.3.2 Fronteira entre as camadas QM e MM

No ONIOM, todo o sistema é dividido em modelos ou regiões. Por exemplo, no ONIOM2, é trivial separar os sistemas "modelo" e "ambiente" se não houver ligação covalente entre eles. O ONIOM2-(QM:MM) deve fornecer resultados idênticos aos do esquema QM/MM aditivo, devido ao cancelamento dos termos MM (VREVEN et al., 2006). Usando a mesma abordagem para o caso de haver ligação covalente entre os sistemas, seria necessário cortar uma ligação; isso levará a ligações pendentes ou radicais e pode introduzir complicações devido à fronteira não física. Tal corte de fronteira QM/MM é inevitável em muitos casos. Muitas abordagens diferentes foram propostas para lidar com esse problema de "fronteira" em abordagens multicamadas. Em geral, existem três aproximações de fronteira: átomo de ligação (LA, do inglês *Link Atoms*), átomo de fronteira (ou pseudoligação) e orbital localizado congelado (SENN; THIEL, 2007; HU; YANG, 2008).

No ONIOM, a abordagem de átomo de ligação tem sido empregada no programa Gaussian. O ONIOM usa átomos de hidrogênio como átomos de ligação por padrão: todas as ligações σ pendentes no sistema modelo são cobertas com átomos de hidrogênio (FIELD; BASH; KARPLUS, 1990). As posições dos LAs podem ser independentes daquelas do átomo hospedeiro de ligação (LAH, o átomo substituído pelo átomo de ligação) e da conexão do átomo de ligação (LAC, o átomo ligado ao LAH no "sistema real" e ao LA no "sistema modelo"), como o tratamento mais simples de LA. No entanto, isso aumenta artificialmente os graus de liberdade no sistema e leva a complicações desnecessárias, como nas implementações anteriores de métodos QM/MM genéricos. Uma abordagem alternativa, usada nas primeiras implementações do ONIOM, é fixar as distâncias de ligação LAC-LAH e LAC-LA em valores razoáveis (os ângulos de ligação e diédricos envolvendo LA também são fixados), o que significa que a posição de LA é independente da posição de LAH. Esta implementação inicial resulta em menos graus de liberdade e, portanto, a análise vibracional não pode ser realizada (MASERAS; MOROKUMA, 1995; MATSUBARA; SIEBER; MOROKUMA, 1996; HUMBEL; SIEBER; MOROKUMA, 1996).

Para minimizar erros de átomos de ligação de hidrogênio em métodos QM/MM e ONIOM, o corte de ligação através da fronteira QM/MM deve ser examinado minuciosamente. Para ser consistente com a eletronegatividade do átomo de

hidrogênio LA, idealmente, uma ligação $C_{sp^3}-C_{sp^3}$ inerte deve ser cortada. Cortar ligações polares com fortes efeitos indutivos (por exemplo, C–O, Si–O) e ligações conjugadas com fortes efeitos de conjugação deve ser evitado em cálculos QM/MM e ONIOM (VREVEN et al., 2006; CLEMENTE; VREVEN; FRISCH, 2010).

3.4 Dinâmica Molecular

A Dinâmica Molecular (DM) é uma técnica computacional amplamente reconhecida por sua versatilidade no estudo de macromoléculas biológicas (HANSSON; OOSTENBRINK; VAN GUNSTEREN, 2002). As simulações de DM têm sido um método bastante popular para investigar a estrutura e função dessas biomoléculas.

A DM é empregada para analisar o movimento de partículas em diversos sistemas, levando em consideração os potenciais de interação entre elas e as equações que regem seus deslocamentos. Baseada nos princípios da Mecânica Clássica, essa abordagem permite examinar a evolução ao longo do tempo dos átomos individuais que constituem o sistema. Ao gerar sequências de posições, é factível obter as propriedades macroscópicas de interesse por meio da aplicação dos conceitos da mecânica estatística (MORGON; COUTINHO, 2007). As quantidades macroscópicas calculadas por meio das funções da mecânica estatística incluem a pressão, energia interna, volume, temperatura, entropia e energia livre (NAMBA; SILVA; da SILVA, 2008).

Nesse método, as partículas são descritas por um conjunto completo de potenciais de interação, denominado “campo de força” (ACKERMANN et al., 1988). Estes potenciais estabelecem as forças que atuam em cada partícula e, conseqüentemente, determinam a evolução do sistema ao longo do tempo, resultando nas trajetórias correspondentes. Com base nas condições iniciais t_0 , posição e velocidade, é possível calcular as forças líquidas que atuam sobre cada partícula (Equação 5), devido às interações com as outras partículas. Ao aplicar as equações newtonianas do movimento, são então determinadas as posições e velocidades em um momento subsequente. Essas novas posições são empregadas para calcular as novas forças, o que por sua vez atualiza as posições e velocidades. Esse

procedimento é repetido iterativamente, gerando trajetórias moleculares para todo o sistema (MORGON; COUTINHO, 2007).

$$F_i = -\nabla V_{total}(t) \quad (5)$$

sendo

$$V_{total} = V_{intra} + V_{inter} \quad (6)$$

onde V_{intra} corresponde aos potenciais de interações para os átomos ligados e V_{inter} para os átomos não ligados. Sendo:

$$V_{intra} = \sum_{lig} k_r (r - r_0)^2 + \sum_{ang} k_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{n,diedro} V_n [1 - (-1)^n \cos(\eta\varphi + \gamma_n)] \quad (7)$$

e

$$V_{inter} = \sum_{i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (8)$$

Os potenciais ligados, Equação 7, descrevem as deformações moleculares. O primeiro termo corresponde aos estiramentos de uma ligação química, o segundo e terceiro termos correspondem às deformações angulares e às deformações dos ângulos diedros, respectivamente. Os estiramentos e deformações angulares são descritos por potenciais harmônicos, onde os parâmetros r_0 e θ_0 correspondem ao comprimento e ângulo de equilíbrio da ligação, e k_r e k_θ às constantes de força. Essas constantes são obtidas de medidas espectroscópicas ou cálculos quânticos. Já as torções diedrais não são adequadamente modeladas segundo um potencial harmônico, uma vez que não são descritos por um potencial parabólico no intervalo $[0, 2\pi]$. Os parâmetros η , V e γ são, respectivamente, a periodicidade da função, a constante de força diedral e a fase angular.

Os potenciais não ligados, Equação 8, consistem na soma das interações de van der Waals e eletrostáticas entre pares de átomos de moléculas distintas, i e j . Os potenciais são representados pelos potenciais de Lennard-Jones e Coulomb, primeiro e segundo termo, respectivamente.

Além disso, existem vários campos de força desenvolvidos por vários grupos de pesquisa em todo o mundo. Entre os campos de força bem estabelecidos e amplamente utilizados para biomoléculas, destacam-se o AMBER, o GROMOS e o

CHARMM (MORGON; COUTINHO, 2007). Para conduzir a simulação de um sistema por meio da dinâmica molecular, além da utilização do campo de força, também são requeridas técnicas específicas que envolvem a implementação de condições periódicas de contorno, a definição de um raio de corte para interações de longo alcance, a seleção de um intervalo de integração para as equações de movimento, bem como o controle de temperatura e pressão por meio de termostatos e barostatos, respectivamente.

Capítulo 4: METODOLOGIA

As coordenadas atômicas da GFP foram obtidas de sua estrutura cristalográfica encontrada no banco de dados *Protein Data Bank* (PDB) sob o código de entrada 1W7S, uma estrutura determinada por raio-X com resolução de 1,8 Å, já estudada na literatura para cálculos com o método ONIOM (THOMPSON, 2014). Essa estrutura é composta por quatro cadeias protéicas, sendo escolhida a cadeia A para os estudos do presente trabalho. Embora a estrutura não possuísse o resíduo 1 e a sequência de resíduos do 232 ao 238, não foi relevante para o presente trabalho modelar os resíduos ausentes, uma vez que estes se localizam em regiões da proteína espacialmente distantes do cromóforo. Por fim, a adição de átomos de hidrogênio foi realizada usando o servidor CHARMM-GUI (JO et al., 2008; BROOKS et al., 2008).

As coordenadas dessa estrutura, após adição de hidrogênios, foram adotadas como ponto inicial para a simulação de dinâmica molecular e cálculos de energia do estado excitado.

A preparação da caixa de simulação e a parametrização do cromóforo foram realizadas utilizando o servidor CHARMM-GUI (JO et al., 2008; BROOKS et al., 2008). A proteína foi confinada em uma caixa de solvatação ortorrômbica, contendo uma camada de moléculas de água do tipo TIP3P, com distância de 10 Å da proteína até as bordas da caixa, Figura 5. Foram adicionados seis íons K^+ para neutralizar o sistema e uma concentração de $0,15 \text{ mol L}^{-1}$ de KCl, seguindo protocolo já adotado pelo grupo (PEREIRA, 2020).

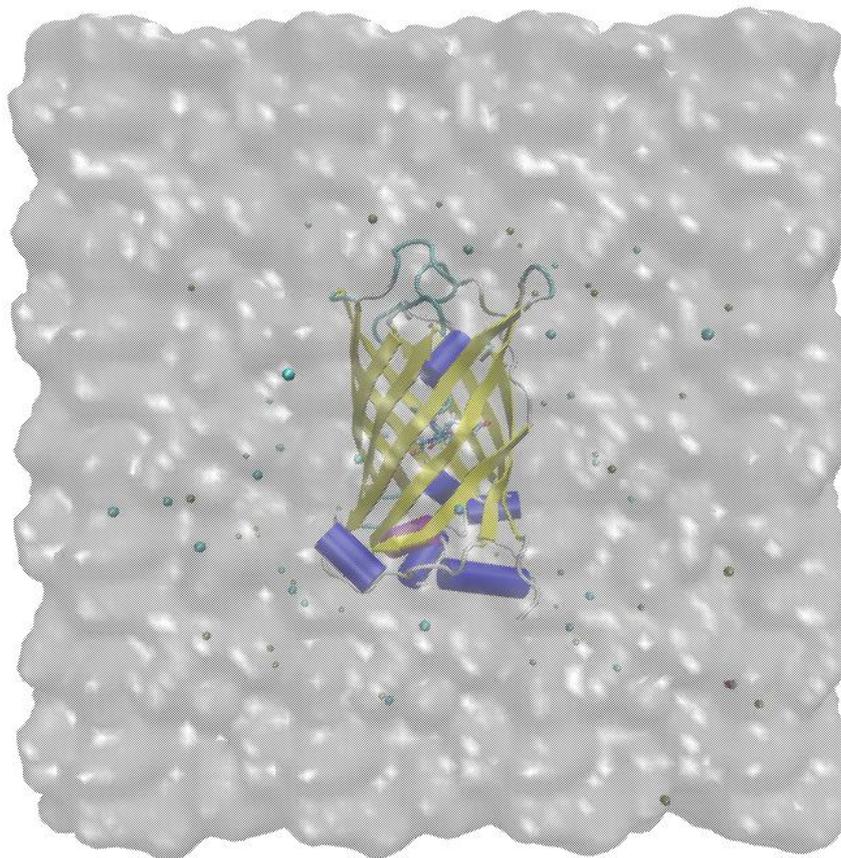


Figura 5. Caixa de simulação para a dinâmica molecular. Ao centro da caixa está representada a estrutura da GFP. Ao redor da proteína estão distribuídas as moléculas de água, em transparente, e os íons Cl^- , em laranja, e K^+ , em verde.

Os cálculos das simulações de dinâmica molecular de equilíbrio e produção foram realizados no ambiente computacional do CENAPAD-SP. As simulações de dinâmica molecular clássica foram realizadas com o software NAMD v2.14 (PHILLIPS et al., 2020) com o campo de força CHARMM36 (BEST et al., 2012). Foram aplicadas condições periódicas de contorno e utilizado o algoritmo SHAKE para restringir as ligações dos átomos de hidrogênio. As interações eletrostáticas, calculadas com o método Particle-Mesh Ewald (PME) e as interações do tipo van der Waals foram calculadas num raio de corte de 10 Å. A simulação foi realizada para a pressão de 1 atm e temperatura de 300 K utilizando o barostato e o termostato de Langevin (FELLER et al., 1995; DAVIDCHAK; HANDEL; TRETAKOV, 2009), respectivamente. O cálculo foi realizado com passos de integração de 2 fs e foi dividido em três etapas:

- 1. Minimização:** esta etapa foi realizada utilizando o algoritmo do gradiente conjugado por 80 ps, com restrição harmônica com força constante de 5 kcal mol⁻¹ Å⁻² para os átomos do *backbone* da proteína.

2. Equilíbrio: após a minimização, o sistema foi submetido a uma etapa de equilíbrio, no *ensemble* NPT, que foi realizada em três estágios:

- a) Estágio 1: 200 ps de simulação com restrições harmônicas com força constante de $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ para os átomos do *backbone* da proteína.
- b) Estágio 2: 200 ps de simulação com restrições harmônicas com força constante de $3 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ para os átomos do *backbone* da proteína.
- c) Estágio 3: 200 ps de simulação com restrições harmônicas com força constante de $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ para os átomos do *backbone* da proteína.

3. Produção: por fim, o sistema foi conduzido à etapa de produção por 200 ns, sem restrições, e com coordenadas coletadas e armazenadas a cada 20 ps.

As informações da trajetória foram obtidas utilizando o programa CPPTRAJ (ROE, CHEATHAM III, 2013), incluso no pacote do AmberTools22 (CASE et al., 2023). Para este trabalho, foram analisados RMSD e RMSF e foi realizada uma clusterização das conformações da proteína ao longo da trajetória.

Para os cálculos de energia com potenciais híbridos QM/MM, efetuados no cluster LMMRQ da UFPB e realizados com o software Gaussian 16 (FRISCH et al., 2016), foram utilizados três modelos de estrutura, apresentados na Tabela 1.

Tabela 1. Estruturas utilizadas para cálculos de potenciais híbridos QM/MM.

Código	Descrição
A	Estrutura do cromóforo otimizada com PM6.
B_x	Estrutura da GFP cristalográfica otimizada com PM6/Amber. O índice x indica o modelo de particionamento QM/MM.
Cy_x	Estrutura representativa da GFP oriunda da dinâmica molecular. O índice x indica o modelo de particionamento QM/MM e o índice y indica o cluster de referência.

Para as estruturas B e C, foram realizados cálculos *single point* adotando o método ONIOM para separar as regiões de interesse. Foram preparados três modelos para a região QM seguindo as orientações da literatura (THOMSON et al., 2014): modelo 1, contendo apenas o cromóforo; modelo 2, contendo o cromóforo mais os

resíduos Arg96 e His148; e modelo 3, contendo o cromóforo mais os resíduos Arg96, His148, Ser205 e Glu222, conforme apresentado na Figura 6.

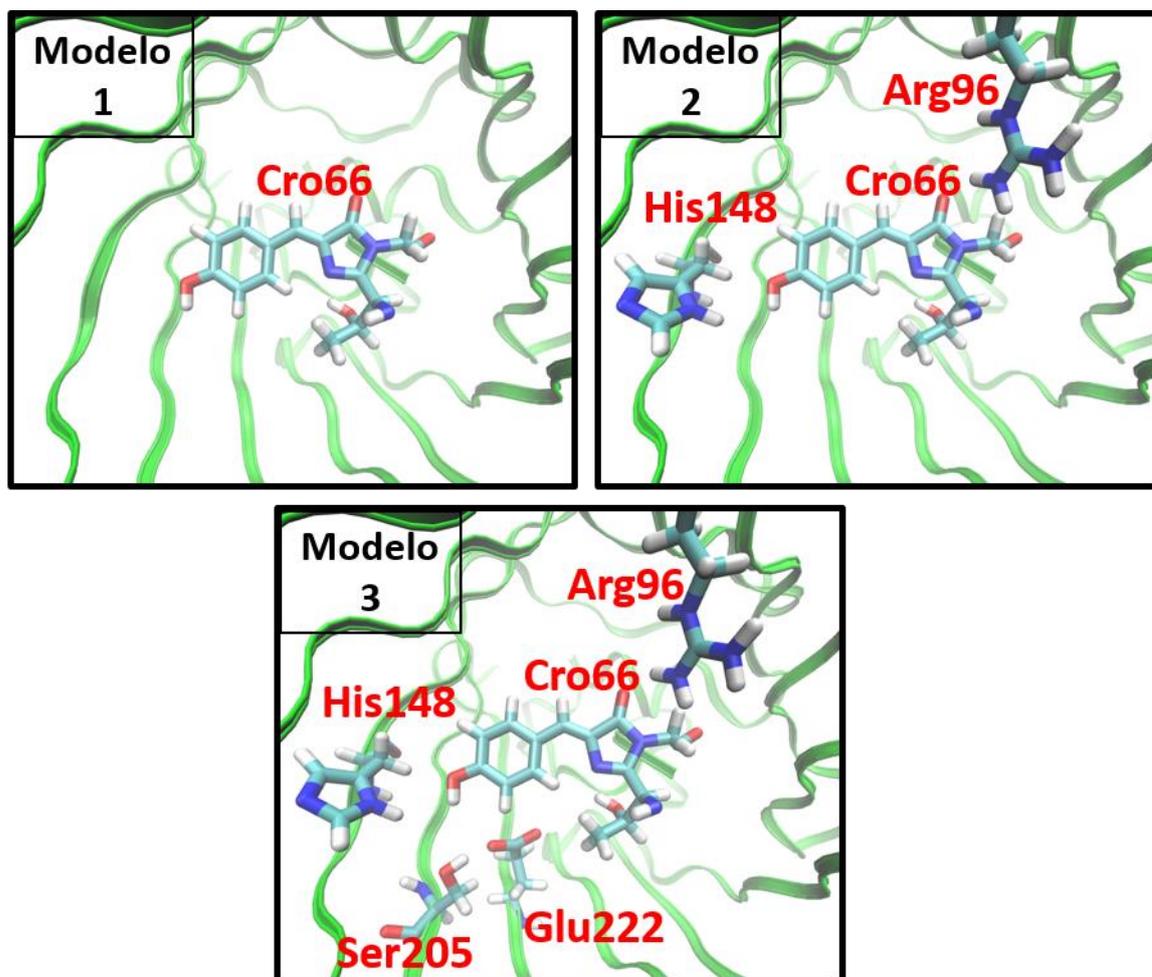


Figura 6. Modelos de separação das regiões QM (em *licorice*) e MM da proteína para os cálculos com o método ONIOM.

A região MM foi tratada com o campo de força Amber ff19SB (CORNELL et al., 1995) e foi feito um estudo comparativo para selecionar o funcional e o conjunto de funções de base mais apropriados para a região QM do sistema de interesse.

Foram realizados testes com os funcionais B3LYP, CAM-B3LYP, PBE0 e ω B97XD. Com o funcional com o melhor desempenho em relação à reprodução do espectro de excitação, foram testados os seguintes conjuntos de funções de base: 3-21G, 6-31G, 6-311G, 6-311+G* e 6-311++G**, em busca de maior acurácia em relação ao espectro e menor custo computacional. Esses testes foram realizados com as estruturas A, B_1 e C0_1. Em seguida, foram replicados cálculos de energia para

os demais modelos de particionamento e estruturas representativas de cada cluster, com o funcional e o conjunto de base escolhidos.

Para avaliar o efeito da proteína sobre as densidades de energia da proteína com o cromóforo, foram utilizados diferentes recortes da estrutura cristalográfica da proteína, sendo eles: o cromóforo isolado; uma região de raio de 5 Å com centro no cromóforo; uma região de raio de 10 Å com centro no cromóforo; uma região de raio de 15 Å com centro no cromóforo; e toda a proteína. A estrutura eletrônica foi calculada com o hamiltoniano semi-empírico PM7 (STEWART, 2013) com a técnica de escalonamento linear de orbitais moleculares localizados MOZYME (STEWART, 1996), implementados no programa MOPAC2016 (STEWART 2007). Utilizando o PRIMoRDIA (GRILLO; URQUIZA-CARVALHO; ROCHA, 2020), foram gerados gráficos de densidade de estados de energia para os orbitais da proteína em cada um dos recortes.

Por fim, os cálculos de energia de excitação foram realizados para os cromóforos de proteínas semelhantes à GFP, adotando o funcional e o conjunto de base mais apropriados, selecionados com base nas etapas anteriores. Mutações nos resíduos Ser65, Tyr66 e Gly67 da proteína fluorescente verde selvagem, wtGFP (com energia de excitação de 3,14 eV para estrutura neutra e 2,61 eV para o ânion), pré-ciclicização, dão origem a diversos resíduos cromofóricos. Assim, foram modelados os cromóforos da proteína fluorescente verde melhorada, eGFP (Thr65-Tyr66-Gly67), com energia de excitação de 2,54 eV; da proteína fluorescente amarela, YFP (Gly65-Tyr66-Gly67), com energia de excitação de 3,16 eV para a estrutura neutra e 2,41 eV para o ânion; da proteína fluorescente ciano melhorada, eCFP (Thr65-Trp66-Gly67), com energia de excitação de 2,85 eV; e da proteína fluorescente azul, BFP (Ser65-His66-Gly67), conforme apresentado na Figura 7 (BEEREPOOT et al., 2013).

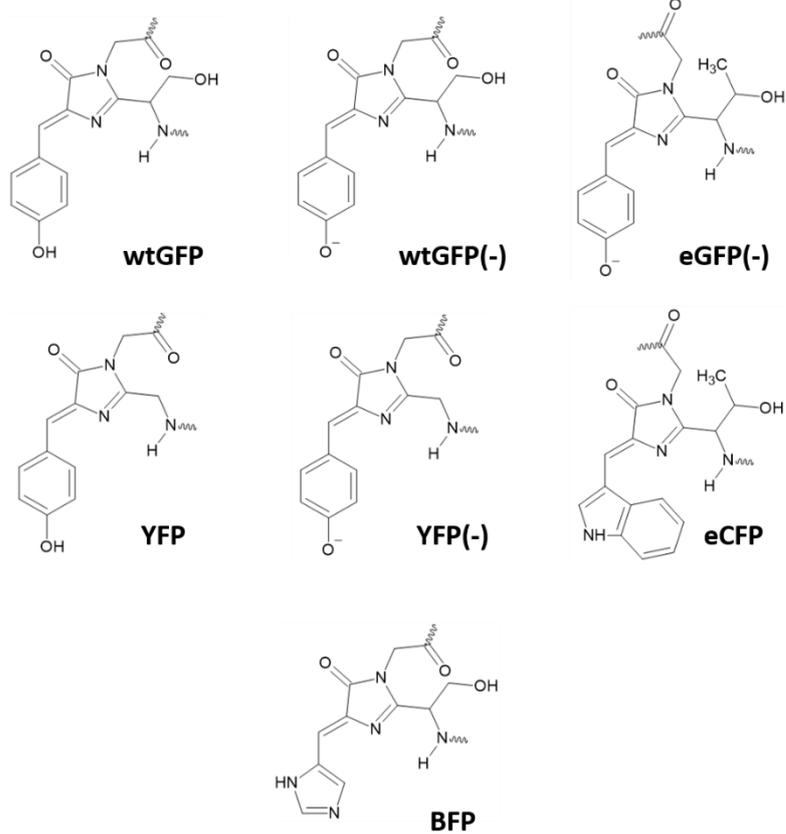


Figura 7. Estrutura dos cromóforos de proteínas variantes da GFP.

A preparação de *inputs* e as análises das trajetórias e de demais resultados foram realizados nos computadores do LQQC/UFPB.

Capítulo 5: RESULTADOS E DISCUSSÃO

Ao final de cada etapa da equilibração da dinâmica molecular, foram coletados os valores de energia total, temperatura média e pressão média, apresentados na Tabela 2.

Tabela 2. Valores da energia total, temperatura média e pressão média após cada etapa de equilibração da dinâmica molecular.

Etapa da equilibração	Propriedades		
	Energia total (kcal mol ⁻¹)	Temperatura média (K)	Pressão média (Bar)
1	-93927,71	299,08	-38,32
2	-93772,66	299,52	-10,25
3	-93918,01	298,06	10,17

A partir desses dados, é possível observar baixa variação nos valores de energia e temperatura média ao longo das três etapas da equilibração. No entanto, os valores de pressão média oscilam consideravelmente, além de não convergir para o valor de 1 atm, estipulado para a dinâmica. Tal comportamento pode ser atribuído à ausência de uma etapa de pressurização do sistema, visto que toda a simulação foi realizada no ensemble NPT.

A etapa da simulação de produção foi realizada em triplicata e, para cada uma das trajetórias obtidas, foram realizadas análises de RMSD (Figura 8) dos átomos do *backbone* da proteína e RMSF por resíduo da proteína, em relação a estrutura inicial da simulação, a fim de avaliar a sua estabilidade ao longo da dinâmica.

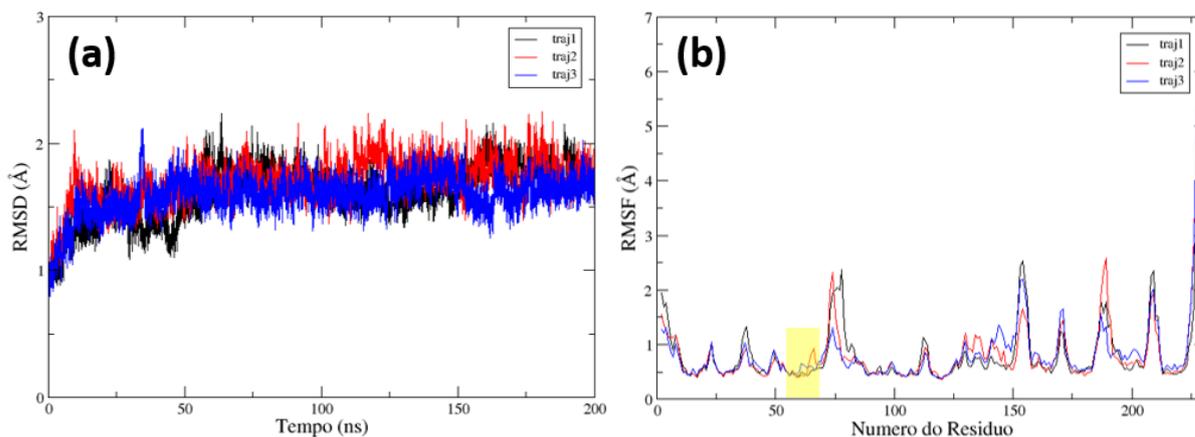


Figura 8. (a) RMSD e (b) RMSF da etapa de produção da dinâmica molecular em triplicata.

Na Figura 8 (a) é possível observar, para as três trajetórias, um aumento progressivo do RMSD no início da etapa, atingindo um limite de aproximadamente 2 Å, indicando a estabilização do sistema e a conservação da estrutura secundária e terciária da proteína.

Na Figura 8 (b), nota-se que, nas três trajetórias, os picos de flutuações coincidem com os resíduos que compõem os loops, sendo estas as regiões de maior flexibilidade da estrutura secundária da proteína durante a dinâmica. A região do cromóforo (Cro66), destacada na Figura 8 (b), apresenta um RMSF médio em torno de 0,5 Å, visto que se encontra na região interna do β -barril entre hélices- α , o que justifica a baixa mobilidade dos resíduos desta região.

Com base nessas trajetórias, foi realizada uma clusterização das conformações da GFP, obtendo-se um total de 10 clusters, apresentados na Tabela 3.

Tabela 3. Clusters obtidos e a quantidade de frames agrupados em cada um.

Cluster	Quantidade de frames (percentual da trajetória)
0	6051 (20,2)
1	4172 (13,9)
2	3878 (12,9)
3	2792 (9,3)
4	2765 (9,2)
5	2449 (8,2)
6	2393 (8,0)
7	2135 (7,1)
8	1744 (5,8)
9	1621 (5,4)

Para avaliar o desempenho dos funcionais de correlação e troca selecionados, foi escolhida a estrutura representativa do cluster mais populoso, C0, a fim de representar as estruturas oriundas da trajetória da dinâmica molecular. Além desta também foram adotadas estruturas cristalográficas otimizadas da proteína e do cromóforo, A e B, respectivamente.

Na Figura 9 são apresentados os valores obtidos para a energia de excitação apenas para a estrutura do cromóforo isolado, que pôde ser tratado com TD-DFT. O desvio percentual, δ , foi calculado em relação ao o valor de referência de 3,14 eV (BEEREPOOT et al., 2013).

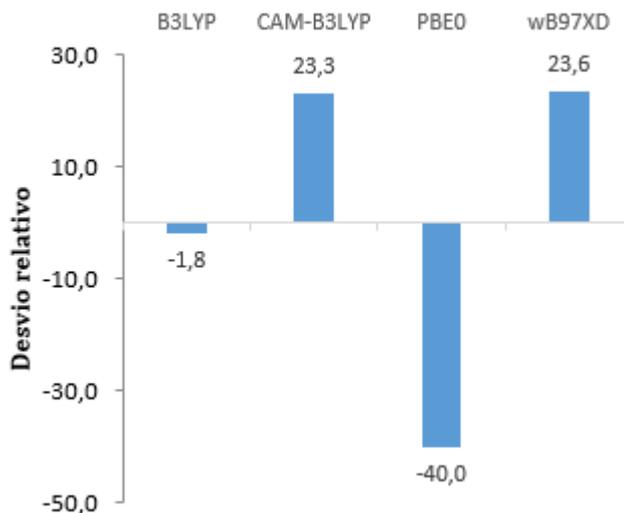


Figura 9. Desvio relativo da energia de excitação do cromóforo da GFP no vácuo, em relação ao valor experimental de 3,14 eV, para diferentes funcionais.

Com base nesses dados, pode-se perceber que o funcional B3LYP apresentou o menor desvio percentual em relação ao valor experimental para a energia de excitação do cromóforo da GFP. Os funcionais CAM-B3LYP e ω B97XD apresentaram desvios percentuais bastante próximos. Em contrapartida, o funcional PBE0 subestimou significativamente o valor da energia de excitação.

Como um dos objetivos do trabalho é realizar esses cálculos para o cromóforo no ambiente da proteína, não é possível concluir que o funcional B3LYP é o mais apropriado, com base apenas nesse resultado.

Assim, o estudo foi realizado, de modo semelhante, para a estrutura B, a fim de investigar o efeito do ambiente da proteína sobre o comportamento do cromóforo. Para tanto, o modelo de partição QM/MM considerado foi o mais simples, onde há apenas o cromóforo na região QM. Além disso, foram levados em consideração o acoplamento mecânico e o acoplamento eletrônico. A Figura 10 apresenta os valores relativos obtidos para a energia de excitação para a estrutura B.

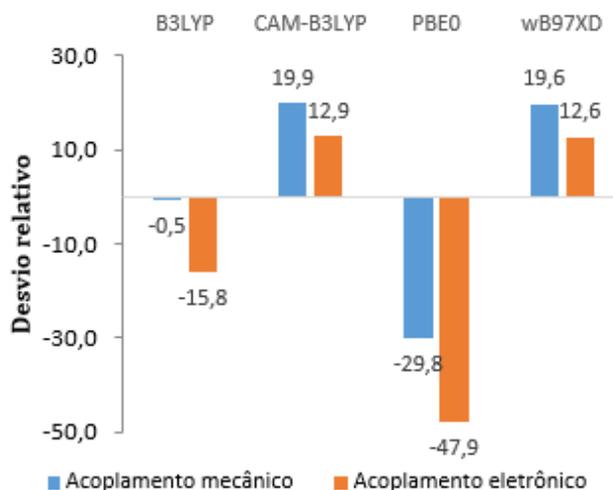


Figura 10. Desvio relativo da energia de excitação do cromóforo da GFP com a estrutura cristalográfica, em relação ao valor experimental de 3,14 eV, para diferentes funcionais.

Assim como para a estrutura B, o estudo com os funcionais também foi realizado para a estrutura C. A Figura 11 apresenta os valores relativos obtidos para a energia de excitação para a estrutura C0.

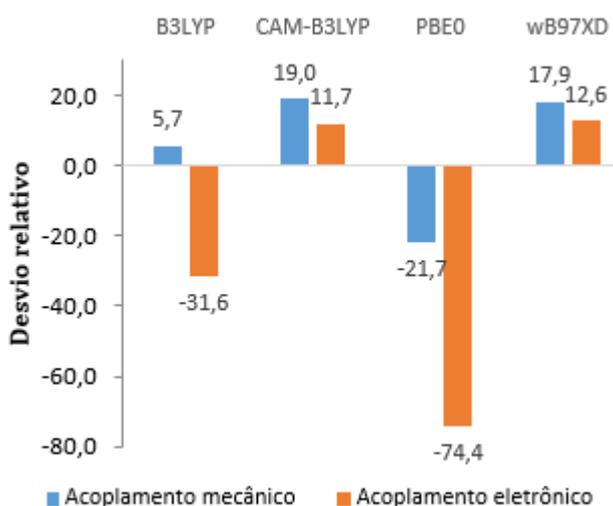


Figura 11. Desvio relativo da energia de excitação do cromóforo da GFP da estrutura oriunda da dinâmica molecular, em relação ao valor experimental de 3,14 eV, para diferentes funcionais.

Pode-se observar que, tanto para a estrutura B quanto para a estrutura C0, o desempenho dos funcionais têm comportamento semelhante ao observado para a estrutura A. Os funcionais que apresentam melhores resultados são o CAM-B3LYP e

o ω B97XD, enquanto que o PBE0 desempenhou a pior performance para esse tipo de cálculo. Considerando o acoplamento eletrônico o valor obtido para a energia de excitação é ainda mais próximo ao valor experimental, que os obtidos com o acoplamento mecânico. Sendo assim, nos cálculos de potencial híbrido subsequentes foram todos realizados adotando o acoplamento eletrônico.

Após a avaliação dos funcionais para cálculos com TD-DFT foi realizado um estudo a fim de selecionar o conjunto de funções de base mais completo, que fornecesse o menor desvio do valor de energia de excitação com o menor custo computacional. Os valores obtidos para cada conjunto de base estão apresentados na Tabela 4.

Tabela 4. Valores obtidos para a energia de excitação, ΔE , em eV, e o desvio percentual, δ , para as estruturas A, B_1 e C0_1, calculados com diferentes conjuntos de funções de base, levando em consideração o acoplamento eletrônico.

Estrutura	Base	CAM-B3LYP		ω B97XD	
		$\Delta E/\text{eV}$	$\delta/\%$	$\Delta E/\text{eV}$	$\delta/\%$
A	3-21G	3,87	23,3	3,88	23,6
	6-31G	3,83	21,9	3,84	22,4
	6-311G	3,80	21,0	3,82	21,6
	6-311+G*	3,00	4,5	3,72	18,4
	6-311++G**	3,69	17,6	3,71	18,2
B_1	3-21G	3,55	12,9	3,54	12,6
	6-31G	3,63	15,6	3,63	15,5
	6-311G	3,62	15,5	3,63	15,5
	6-311+G*	3,60	14,6	3,61	14,9
	6-311++G**	3,59	14,4	3,60	14,7
C0_1	3-21G	3,51	11,7	3,54	12,6
	6-31G	3,70	17,9	3,71	18,3
	6-311G	3,72	18,4	3,73	18,9
	6-311+G*	3,73	18,9	3,75	19,3
	6-311++G**	3,73	18,8	3,74	19,2

Para as três estruturas estudadas e para os dois funcionais adotados, os conjuntos de funções 6-311+G* e 6-311++G** apresentaram resultados bastante próximos, sendo que os resultados com o funcional CAM-B3LYP foram ligeiramente melhores. Levando em consideração o menor custo computacional associado, selecionou-se o conjunto de funções de base 6-311+G* juntamente com o funcional CAM-B3LYP.

Uma vez realizada a escolha do funcional e do conjunto de base, foi possível explorar o efeito da inclusão de resíduos próximos ao cromóforo na região QM. Sendo assim, foram realizados cálculos com as estruturas da proteína, B e C0, adotando os três diferentes modelos de divisão entre as regiões QM e MM. A Tabela 5 apresenta os valores obtidos para a energia de excitação.

Tabela 5. Valores obtidos para a energia de excitação, ΔE , em eV, e o desvio percentual, δ , para as estruturas B e C0 com diferentes modelos de particionamento QM/MM, com o método QM CAM-B3LYP/6-311+G* e acoplamento eletrônico.

Estrutura	$\Delta E/eV$	$\delta/\%$
B_1	3,60	14,6
B_2	3,60	14,8
B_3	3,60	14,6
C0_1	3,73	18,9
C0_2	3,17	0,9
C0_3	2,41	23,1

É possível notar que, para a estrutura B, otimizada, os valores da energia de excitação foram bastante próximos para os três modelos de particionamento. Em contraponto, a estrutura C0 apresenta variações significativas à medida que são adicionados resíduos da proteína à região QM. Dada essa discrepância, pode-se supor que o problema seja a estrutura oriunda da dinâmica molecular.

Estes cálculos foram replicados para as demais estruturas representativas dos clusters obtidos da trajetória da dinâmica molecular. A Tabela 6 apresenta os valores obtidos para a energia de excitação de cada estrutura para cada modelo de particionamento QM/MM.

Tabela 6. Valores obtidos para a energia de excitação, ΔE , em eV, e o desvio percentual médio, δ , para as estruturas Cy com diferentes modelos de particionamento QM/MM, com o método QM CAM-B3LYP/6-311+G* e acoplamento eletrônico.

Cluster	Modelo de particionamento		
	1	2	3
C0	3,73	3,17	2,42
C1	3,40	3,43	3,38
C2	3,33	3,37	3,36
C3	3,41	3,31	2,30
C4	3,37	3,40	3,26
C5	3,39	3,50	3,12
C6	3,45	3,48	3,34
C7	3,44	3,62	2,65
C8	3,39	3,46	3,43
C9	3,37	3,50	1,98
Média	3,39	3,44	3,19
δ/%	8,10	9,64	1,70

Analisando os dados apresentados para as estruturas oriundas da dinâmica molecular, embora seja possível notar variações entre as energias de excitação de cada modelo, o valor médio é significativamente mais próximo do valor experimental do que os valores apresentados pela estrutura representativa do cluster mais populoso.

Ainda sobre a influência da proteína sobre o cromóforo, a Figura 12 apresenta os gráficos de densidade de estados de energias para os orbitais moleculares da GFP calculados com o método PM7 junto com algoritmo MOZYME.

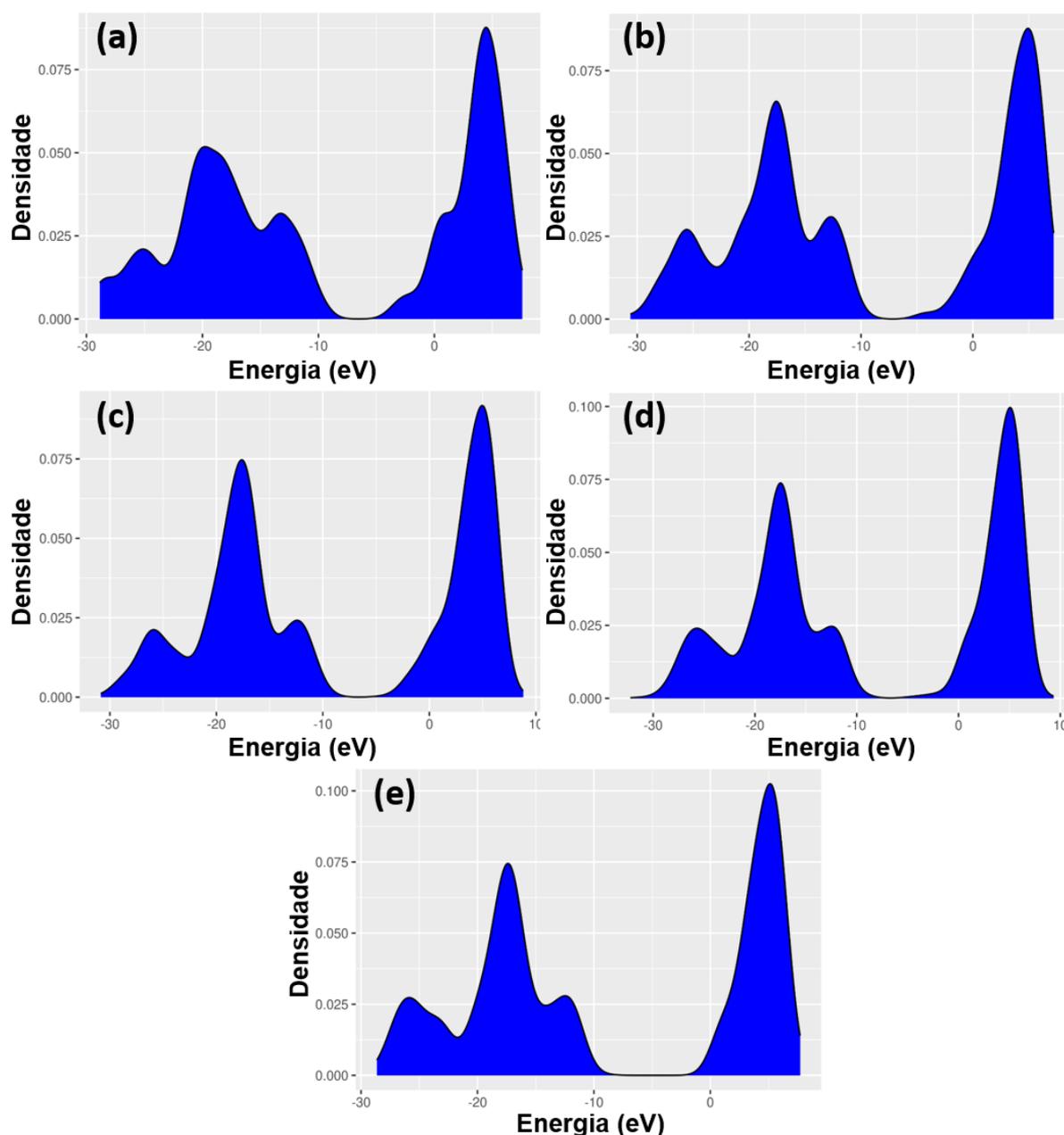


Figura 12. Densidade de estados de energia para os orbitais moleculares da GFP, para: (a) o cromóforo isolado; (b) uma região de 5 Å da proteína centrada no

cromóforo; (c) uma região de 10 Å da proteína centrada no cromóforo; (d) uma região de 15 Å da proteína centrada no cromóforo; e (e) a proteína inteira.

Observando o perfil da distribuição da densidade de energia de todas as estruturas calculadas, não é possível perceber influência significativa do ambiente da proteína sobre o cromóforo. No entanto, pode-se perceber que, quando tratado isoladamente, o cromóforo apresenta menor concentração de energia, Figura 12 (a), quando comparado com estruturas que incluem porções maiores da proteína, Figura 12 (d) e (e). Como consequência disso, o gap HOMO-LUMO é ligeiramente maior neste último caso.

Em relação aos cromóforos de proteínas semelhantes à GFP, os valores obtidos para as energias de excitação, Figura 13, foram calculados com ONIOM(CAM-B3LYP:AMBER) para a região das proteínas contendo apenas os respectivos cromóforos, com a região QM apenas para o cromóforo.

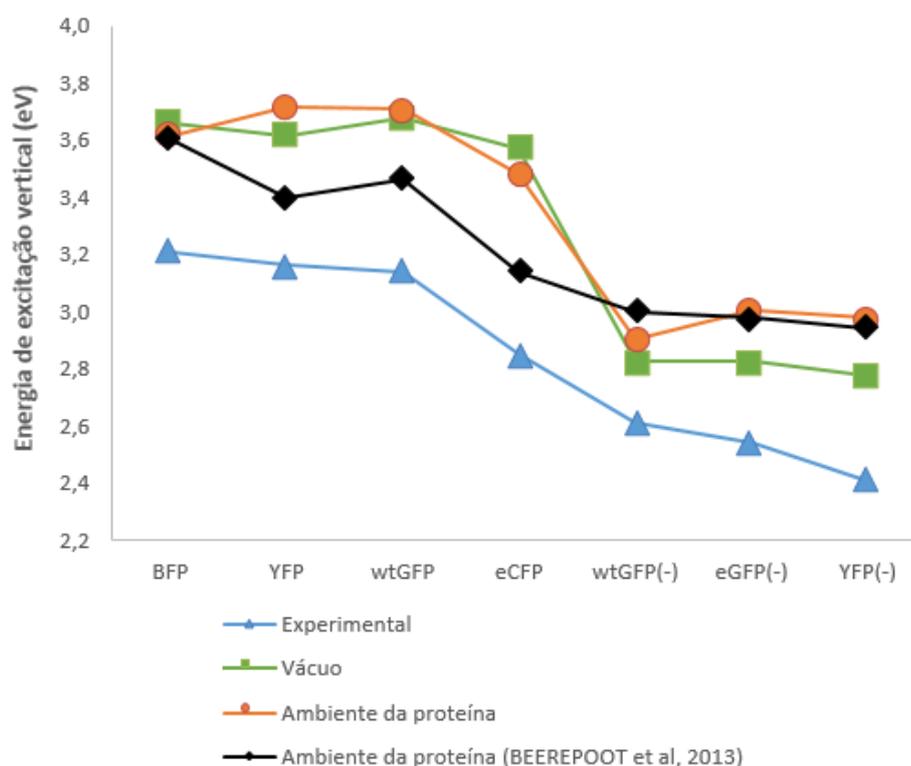


Figura 13. Energias de excitação vertical para cada proteína fluorescente.

Na Figura 13 é possível perceber que todas as energias de excitação foram superestimadas quando comparadas com os valores experimentais, reproduzindo um comportamento já reportado na literatura para cálculos com CAM-B3LYP (LIST et al 2012; BEEREPOOT et al., 2013). Apesar disso, os valores calculados reproduzem a

tendência experimental para as diferentes proteínas. Neste sentido, é necessário pontuar que os valores obtidos para os modelos de cromóforos no vácuo, no geral, renderam resultados mais próximos aos valores experimentais que os calculados com o ONIOM.

Para avaliação do efeito da alteração dos resíduos que dão origem ao cromóforo das proteínas fluorescentes, o cromóforo da wtGFP é adotado como estrutura de referência. Para esta proteína, a energia de excitação calculada com o método ONIOM foi equivalente a 0,57 eV acima do valor experimental, de 3,14 eV.

O estado de ionização para o cromóforo da wtGFP prevê uma diminuição de 0,53 eV na energia de excitação do cromóforo neutro para o cromóforo aniônico. No entanto, os valores calculados para essas estruturas levaram a uma diferença de 0,80 eV entre os dois estados. A estrutura aniônica da wtGFP rendeu a estimativa calculada com o ONIOM mais próxima do valor experimental, excedendo 0,30 eV.

Quando a mudança na estrutura do cromóforo consiste em substituir o resíduo Tyr66, da wtGFP, por His65, na BFP, o valor calculado prevê uma diminuição na energia de excitação em torno de 0,10 eV, enquanto que na observação dos dados experimentais é esperado que essa mudança provoque uma elevação na energia correspondente a 0,07 eV. Em face ao valor experimental da energia de excitação para a BFP, o valor calculado excede em 0,40 eV.

No caso do cromóforo da eGFP, além da substituição do resíduo Ser65 por Thr65, o cromóforo está no estado aniônico. Assim, para efeito de verificar a influência dessas substituições sobre o valor da energia de excitação, a melhor comparação a ser feita é com o cromóforo aniônico da wtGFP. Experimentalmente, o cromóforo da eGFP tem energia de excitação menor que a do cromóforo aniônico da wtGFP em torno de 0,07 eV. A diferença entre os valores calculados, no entanto, prevê que a energia de excitação para a eGFP exceda o valor da wtGFP aniônica em 0,10 eV. Em relação ao valor experimental da energia de excitação, o valor calculado excedeu 0,46 eV.

A alteração do resíduo Ser65, da wtGFP, por Gly65, na YFP, não resultou em diferença significativa na energia de excitação, assim como observado nos valores experimentais da energia de excitação para essas proteínas. Para essa proteína, o valor calculado excede em 0,56 eV do valor experimental.

Embora o valor calculado para a estrutura aniônica da YFP exceda em 0,57 eV o valor experimental, a diferença entre os valores de energia de excitação calculados

para estrutura neutra e a aniônica é equivalente à diferença observada para o valor experimental dessas estruturas, equivalente a 0,75 eV.

Na eCFP, há dois resíduos diferentes da wtGFP: o cromóforo apresenta a sequência Thr65-Trp66 ao invés de Ser65-Tyr66. Como consequência, experimentalmente é observado que o cromóforo da eCFP apresenta energia de excitação diminuída em 0,29 eV com relação ao cromóforo da wtGFP. Os valores calculados se aproximam desse resultado, alcançando uma diferença de 0,23 eV entre as energias de excitação para os dois cromóforos. Em relação ao valor experimental, a estimativa calculada pelo método utilizado foi 0,63 eV superior.

Capítulo 6: CONCLUSÕES

A simulação de dinâmica molecular apresentou estruturas com baixas flutuações em torno do cromóforo. Os funcionais de correlação e troca que apresentaram melhores desempenhos para os cálculos da energia de excitação eletrônica foram o CAM-B3LYP e o ω B97XD, levando em consideração o acoplamento eletrônico, com pequenas variações entre si. Para este estudo foi selecionado o funcional CAM-B3LYP com o conjunto de base 6-311+G*.

Ao considerar os modelos de particionamento, foi observado uma variação significativa do valor da energia de excitação entre os três modelos da estrutura C0, em relação à estrutura B. No entanto, a média dos valores da energia de excitação das estruturas representativas da dinâmica apresentam maior compatibilidade com dados experimentais.

Acerca da influência da proteína sobre os níveis de energia do cromóforo, não foi observado efeito significativo. Desta forma, espera-se que realizar os cálculos de energia em nível computacional mais rigoroso apenas para o cromóforo não compromete a confiabilidade do resultado.

Os valores calculados para a energia de excitação dos cromóforos análogos ao cromóforo da GFP superestimaram o valor experimental, previsto com o funcional CAM-B3LYP. Entretanto, as tendências observadas estão em consonância com as já reportadas na literatura.

REFERÊNCIAS

- ACHARYA, A. et al. Photoinduced chemistry in fluorescent proteins: curse or blessing?. **Chemical Reviews**, v. 117, n. 2, p. 758-795, 2017.
- ACKERMANN, T. et al. Proteins: a theoretical perspective of dynamics, structure and thermodynamics, volume LXXI. **advances in chemical physics, John Wiley & Sons**, New York, p. 96-96, 1988.
- ADAMO, C.; JACQUEMIN, D. The calculations of excited-state properties with Time-Dependent Density Functional Theory. **Chemical Society Reviews**, v. 42, n. 3, p. 845-856, 2013.
- BEEREPOOT, M. T. P et al. A polarizable embedding DFT study of one-photon absorption in fluorescent proteins. **Physical Chemistry Chemical Physics**, v. 15, n. 13, p. 4735-4743, 2013.
- BEREZIN, M. Y.; ACHILEFU, S. Fluorescence lifetime measurements and biological imaging. **Chemical reviews**, v. 110, n. 5, p. 2641-2684, 2010.
- BEST, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. **Journal of chemical theory and computation**, v. 8, n. 9, p. 3257-3273, 2012.
- BRAVAYA, K. B. et al. Quantum chemistry behind bioimaging: insights from ab initio studies of fluorescent proteins and their chromophores. **Accounts of chemical research**, v. 45, n. 2, p. 265-275, 2012.
- BROOKS, B. R. et al. CHARMM: the biomolecular simulation program. **Journal of computational chemistry**, v. 30, n. 10, p. 1545-1614, 2008.
- CHUDAKOV, D. M. et al. Fluorescent proteins and their applications in imaging living cells and tissues. **Physiological reviews**, v. 90, n. 3, p. 1103-1163, 2010.
- CHUNG, L. W. et al. The ONIOM method and its applications. **Chemical reviews**, v. 115, n. 12, p. 5678-5796, 2015.
- CLEMENTE, F. R.; VREVEN, T.; FRISCH, M. J.; MATTA, C. F. (Ed.). Quantum biochemistry. Wiley-VCH Verlag GmbH & Co, 2010.
- CORNELL, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. **Journal of the American Chemical Society**, v. 117, n. 19, p. 5179-5197, 1995.

D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, G. Giambaşu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, Q. Zhu, and P.A. Kollman. [Programa de Computador] Amber 2023, University of California, San Francisco, 2023.

DAS, D. et al. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. **The Journal of chemical physics**, v. 117, n. 23, p. 10534-10547, 2002.

DAVIDCHACK, R. L.; HANDEL, R.; TRETYAKOV, M. V. Langevin thermostat for rigid body dynamics. **The Journal of chemical physics**, v. 130, n.23, 2009.

DAY, R. N.; DAVIDSON, M. W. The fluorescent protein palette: tools for cellular imaging. **Chemical Society Reviews**, v. 38, n. 10, p. 2887-2921, 2009.

FELLER, S. E. et al. Constant pressure molecular dynamics simulation: The Langevin piston method. **The Journal of chemical physics**, v. 103, n. 11, p. 4613-4621, 1995.

FIELD, M. J.; BASH, P. A.; KARPLUS, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. **Journal of computational chemistry**, v. 11, n. 6, p. 700-733, 1990.

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. [Programa de Computador] Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford CT, 2016.

GANESAN, S. et al. A dark yellow fluorescent protein (YFP)-based Resonance Energy-Accepting Chromoprotein (REACH) for Förster resonance energy transfer with GFP. **Proceedings of the National Academy of Sciences**, v. 103, n. 11, p. 4089-4094, 2006.

GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; ROCHA, G. B. PRIMoRDIA: A software to explore reactivity and electronic structure in large biomolecules. **Journal of Chemical Information and Modeling**, v. 60, n. 12, p. 5885-5890, 2020.

HANSSON, T.; OOSTENBRINK, C.; VAN GUNSTEREN, W. Molecular dynamics simulations. **Current opinion in structural biology**, v. 12, n.2, p.190-196, 2002.

HELMS, V. Electronic excitations of biomolecules studied by quantum chemistry. **Current opinion in structural biology**, v. 12, n. 2, p. 169-175, 2002.

HU, H.; YANG, W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. **Annu. Rev. Phys. Chem.**, v. 59, p. 573-601, 2008.

HUMBEL, S.; SIEBER, S.; MOROKUMA, K. The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and SN 2 reaction: RCl+ Cl⁻. **The Journal of chemical physics**, v. 105, n. 5, p. 1959-1967, 1996.

JACQUEMIN, D.; MENNUCCI, B.; ADAMO, C. Excited-state calculations with TD-DFT: from benchmarks to simulations in complex environments. **Physical chemistry chemical physics**, v. 13, n. 38, p. 16987-16998, 2011.

JIN, X.; GLOVER, W. J.; HE, X. Fragment quantum mechanical method for excited states of proteins: Development and application to the green fluorescent protein. **Journal of Chemical Theory and Computation**, v. 16, n. 8, p. 5174-5188, 2020.

JO, S. et al. CHARMM-GUI: a web-based graphical user interface for CHARMM. **Journal of computational chemistry**, v. 29, n. 11, p. 1859-1865, 2008.

KHANDOGIN, J. YORK, D. M. Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. **Proteins: Structure, Function, and Bioinformatics**, v. 56, n. 4, p. 724-737, 2004.

KÖNIG, P. H. et al. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. **The Journal of Physical Chemistry B**, v. 109, n. 18, p. 9082-9095, 2005.

LAINO, T.; NIFOSÌ, R.; TOZZINI, V. Relationship between structure and optical properties in green fluorescent proteins: a quantum mechanical study of the chromophore environment. **Chemical physics**, v. 298, n. 1-3, p. 17-28, 2004.

LIST, N. H. et al. Performance of popular XC-functionals for the description of excitation energies in GFP-like chromophore models. **International Journal of Quantum Chemistry**, v. 112, n. 3, p. 789-800, 2012.

LITVINENKO, K. L.; WEBBER, N. M.; MEECH, S. R. An ultrafast polarisation spectroscopy study of internal conversion and orientational relaxation of the chromophore of the green fluorescent protein. **Chemical Physics Letters**, v. 346, n. 1-2, p. 47-53, 2001.

LUKYANOV, K. A. et al. Fluorescent proteins as light-inducible photochemical partners. **Photochemical & Photobiological Sciences**, v. 9, n. 10, p. 1301-1306, 2010.

MA, Y. et al. Theoretical studies of chromophore maturation in the wild-type green fluorescent protein: ONIOM (DFT: MM) investigation of the mechanism of cyclization. **The Journal of Physical Chemistry B**, v. 116, n. 4, p. 1426-1436, 2012.

MARQUES, M. A. et al. Time-dependent density-functional approach for biological chromophores: the case of the green fluorescent protein. **Physical review letters**, v. 90, n. 25, p. 258101, 2003.

MARTIN, M. E.; NEGRI, F.; OLIVUCCI, M. Origin, nature, and fate of the fluorescent state of the green fluorescent protein chromophore at the CASPT2//CASSCF resolution. **Journal of the American Chemical Society**, v. 126, n. 17, p. 5452-5464, 2004.

MASERAS, F.; MOROKUMA, K. IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. **Journal of Computational Chemistry**, v. 16, n. 9, p. 1170-1179, 1995.

MATSUBARA, T.; SIEBER, S.; MOROKUMA, K. A test of the new "integrated MO+ MM"(IMOMM) method for the conformational energy of ethane and n-butane. **International journal of quantum chemistry**, v. 60, n. 6, p. 1101-1109, 1996.

MATSUURA, A. et al. Theoretical study on the absorption maxima of real GFPs. **Chemical Physics Letters**, v. 484, n. 4-6, p. 324-329, 2010.

MEECH, S. R. Excited state reactions in fluorescent proteins. **Chemical Society Reviews**, v. 38, n. 10, p. 2922-2934, 2009.

MISHIN, A. S. et al. Novel uses of fluorescent proteins. **Current opinion in chemical biology**, v. 27, p. 1-9, 2015.

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

NAMBA, A. M. SILVA, V. B.; da SILVA, C. H. T. P. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, v. 33, p p. 13–24, 2008.

NEMUKHIN, A. V.; GRIGORENKO, B. L.; SAVITSKY, A. P. Computer modeling of the structure and spectra of fluorescent proteins. **Acta Naturae**, v. 1, n. 2 (2), p. 33-43, 2009.

PATNAIK, S. S. et al. Computational study of the absorption spectra of green fluorescent protein mutants. **Biopolymers: Original Research on Biomolecules**, v. 85, n. 3, p. 253-263, 2007.

PATNAIK, S. S.; TROHALAKI, S.; PACHTER, R. Molecular modeling of green fluorescent protein: structural effects of chromophore deprotonation. **Biopolymers: Original Research on Biomolecules**, v. 75, n. 6, p. 441-452, 2004.

PEREIRA, L. S. **Estudo teórico do mecanismo de ação da enzima RTA da Ricina com um ligante análogo ao substrato natural rRNA 28S**. Dissertação (Mestrado em Química) – Programa de Pós-graduação em Química, Universidade Federal da Paraíba. João Pessoa, p.65. 2020.

PETRONE, A. et al. On the driving force of the excited-state proton shuttle in the green fluorescent protein: a time-dependent density functional theory (TD-DFT) study of the intrinsic reaction path. **Journal of chemical theory and computation**, v. 12, n. 10, p. 4925-4933, 2016.

PETTIKIRIARACHCHI, A. et al. Ultramarine, a chromoprotein acceptor for Förster resonance energy transfer. **PloS one**, v. 7, n. 7, p. e41028, 2012.

PHILLIPS, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. **The Journal of chemical physics**, v. 153, n. 4, p. 044130, 2020.

REMYNGTON, S. James. Fluorescent proteins: maturation, photochemistry and photophysics. **Current opinion in structural biology**, v. 16, n. 6, p. 714-721, 2006.

ROE, D. R.; CHEATHAM III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. **Journal of chemical theory and computation**, v. 9, n. 7, p. 3084-3095, 2013.

SAMPLE, V.; NEWMAN, R. H.; ZHANG, J. The structure and function of fluorescent proteins. **Chemical Society Reviews**, v. 38, n. 10, p. 2852-2864, 2009.

SENN, H. M.; THIEL, W.; REIHER, M. (Ed.). Atomistic approaches in modern biology: from quantum chemistry to molecular simulations. Springer, 2007.

SEWARD, H. E.; BAGSHAW, C. R. The photochemistry of fluorescent proteins: implications for their biological applications. **Chemical Society Reviews**, v. 38, n. 10, p. 2842-2851, 2009.

SHCHERBAKOVA, D. M.; VERKHUSHA, V. V. Chromophore chemistry of fluorescent proteins controlled by light. **Current opinion in chemical biology**, v. 20, p. 60-68, 2014.

SINICROPI, A. et al. Properties of the emitting state of the green fluorescent protein resolved at the CASPT2//CASSCF/CHARMM level. **Journal of the American Chemical Society**, v. 127, n. 33, p. 11534-11535, 2005.

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

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

STEWART, J. J. P. Stewart computational chemistry. <http://openmopac.net/>, 2007.

SUBACH, F. V.; VERKHUSHA, V. V. Chromophore transformations in red fluorescent proteins. **Chemical reviews**, v. 112, n. 7, p. 4308-4327, 2012.

THOMPSON, L. M. et al. Analytical harmonic vibrational frequencies for the green fluorescent protein computed with ONIOM: chromophore mode character and its response to environment. **Journal of chemical theory and computation**, v. 10, n. 2, p. 751-766, 2014.

TONIOLO, A.; GRANUCCI, G.; MARTÍNEZ, T. J. Conical intersections in solution: A QM/MM study using floating occupation semiempirical configuration interaction wave functions. **The Journal of Physical Chemistry A**, v. 107, n. 19, p. 3822-3830, 2003.

TSIEN, R. Y. The green fluorescent protein. **Annual review of biochemistry**, v. 67, n. 1, p. 509-544, 1998.

VAN THOR, J. J. Photoreactions and dynamics of the green fluorescent protein. **Chemical Society Reviews**, v. 38, n. 10, p. 2935-2950, 2009.

VOITYUK, A. A. et al. Absorption spectra of the GFP chromophore in solution: comparison of theoretical and experimental results. **Chemical Physics**, v. 269, n. 1-3, p. 83-91, 2001.

VOITYUK, A. A.; MICHEL-BEYERLE, M.E.; RÖSCH, N. Structure and rotation barriers for ground and excited states of the isolated chromophore of the green fluorescent protein. **Chemical physics letters**, v. 296, n. 3-4, p. 269-276, 1998.

VREVEN, T. et al. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. **Journal of chemical theory and computation**, v. 2, n. 3, p. 815-826, 2006.

VREVEN, T.; MOROKUMA, K. Investigation of the S₀-S₁ excitation in bacteriorhodopsin with the ONIOM (MO: MM) hybrid method. **Theoretical Chemistry Accounts**, v. 109, p. 125-132, 2003.

WACHTER, R. M. The family of GFP-like proteins: structure, function, photophysics and biosensor applications. Introduction and perspective. **Photochemistry and Photobiology**, v. 82, n. 2, p. 339-344, 2006.

WEBBER, N. M.; LITVINENKO, K. L.; MEECH, S. R. Radiationless relaxation in a synthetic analogue of the green fluorescent protein chromophore. **The Journal of Physical Chemistry B**, v. 105, n. 33, p. 8036-8039, 2001.

ZIMMER, M. GFP: from jellyfish to the Nobel prize and beyond. **Chem. Soc. Rev.**, v. 38, p. 2823-2832, 2009.

ZIMMER, M. Green fluorescent protein (GFP): applications, structure and related photophysical behavior. **Chem. Rev.**, v. 102, p. 759–81, 2002.