



Universidade Estadual de Campinas
Instituto de Química

ALVARO JHOVALDO LOPEZ AYME

ESTUDO COMPUTACIONAL DA FLUORESCÊNCIA DO
TRIPTOFANO EM PROTEÍNAS POR MEIO DE SIMULAÇÕES
DE DINÂMICA MOLECULAR

Campinas

2019

ALVARO JHOVALDO LOPEZ AYME

**ESTUDO COMPUTACIONAL DA FLUORESCÊNCIA DO
TRIPTOFANO EM PROTEÍNAS POR MEIO DE SIMULAÇÕES
DE DINÂMICA MOLECULAR**

Tese de Doutorado apresentada ao Instituto de
Química da Universidade Estadual de Campinas
como parte dos requisitos exigidos para a obtenção
do título de Doutor em Ciências

Orientador: Prof. Dr. Leandro Martínez

O arquivo digital corresponde à versão final da Tese defendida pelo aluno Alvaro Jhovaldo Lopez Ayme e orientada pelo Prof. Dr. Leandro Martínez.

CAMPINAS
2019

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Química
Camila Barleta Fullin - CRB 8462

L881e Lopez Ayme, Alvaro Jhovaldo, 1988-
Estudo computacional da fluorescência do triptofano em proteínas por meio de simulações de dinâmica molecular / Alvaro Jhovaldo Lopez Ayme. – Campinas, SP : [s.n.], 2019.

Orientador: Leandro Martínez.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Química.

1. Triptofano. 2. Dinâmica molecular. 3. Fluorescência. I. Martínez, Leandro, 1979-. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Computational study of tryptophan fluorescence in proteins by molecular dynamics simulations

Palavras-chave em inglês:

Tryptophan

Molecular dynamics

Fluorescence

Área de concentração: Físico-Química

Titulação: Doutor em Ciências

Banca examinadora:

Leandro Martínez [Orientador]

Francisco Benedito Teixeira Pessine

Miguel Angel San Miguel Barrera

Kaline Rabelo Coutinho

Amando Siuiti Ito

Data de defesa: 31-05-2019

Programa de Pós-Graduação: Química

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: 000-0002-6073-6058

- Currículo Lattes do autor: <http://lattes.cnpq.br/9043941458377307>

BANCA EXAMINADORA

Prof. Dr. Leandro Martínez (Orientador)

Prof. Dr. Francisco Benedito Teixeira Pessine (IQ - UNICAMP)

Prof. Dr. Miguel Angel San Miguel Barrera (IQ - UNICAMP)

Profa. Dra. Kaline Rabelo Coutinho (IF - USP)

Prof. Dr. Amando Siuiti Ito (FFCLRP -USP)

A Ata da defesa assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Este exemplar corresponde à redação final da Tese de Doutorado defendida pelo aluno **Alvaro Jhovaldo Lopez Ayme**, aprovada pela Comissão Julgadora em 31 de maio de 2019.

Dedico esta tese a

Cassia e Rafaela

Agradecimentos

Quero usar este espaço para agradecer muitíssimo a meu orientador, o Prof. Dr. Leandro Martínez, pois foi graças a ele que eu tive a oportunidade de desenvolver este trabalho e conhecer mais de perto a pesquisa em Química Computacional. Sou muito grato a você professor e quero dizer que até o momento o senhor tem sido o melhor professor e orientador que eu conheci.

Também quero agradecer aos todos os colegas com os que pude compartilhar vários momentos agradáveis, em especial ao Ivan, Mariana, Luciano e Antonio.

Por último, agradeço a agencia de fomento CNPq pelo apoio financeiro (Processo CNPq 161789/2014-5).

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

Resumo

A fluorescência do triptofano é altamente sensível ao ambiente, motivo pelo qual é até os dias de hoje amplamente usada para investigar a estrutura e dinâmica das proteínas. O comprimento de onda de emissão de fluorescência (λ_{em}) do triptofano em proteínas é muito variável. Existem proteínas com λ_{em} baixos (*e.g.* Azurina, 308 nm), e outras com λ_{em} altos (*e.g.* Glucagon, 355 nm). Essa diversidade de λ_{em} motivou estudos teóricos que pudessem explicar o fenômeno do desvio de Stokes do triptofano em proteínas. Um dos estudos mais relevantes foi o de Vivian e Callis (2004), que propuseram um modelo híbrido QM-MM com capacidade de prever os λ_{em} . O modelo de Vivian e Callis (2004) teve um desempenho muito bom, com desvio padrão das predições com respeito aos dados experimentais (SD) de 6,87 nm, e um coeficiente de correlação de $R^2 = 0,69$. O uso desses métodos requer métodos computacionais sofisticados restringindo assim seu uso na interpretação de dados experimentais. Nesta tese foram desenvolvidos modelos clássicos parametrizados baseados na área acessível ao solvente e baseados nas interações eletrostáticas para a predição dos λ_{em} do triptofano. Esses modelos clássicos, permitem melhores predições se comparadas ao modelo baseado nos primeiros princípios de Vivian e Callis (2004), a um custo e complexidade computacionais muito menores. Por exemplo, um modelo baseado nas interações eletrostáticas prediz os comprimentos de onda de emissão com um SD = 4,89 nm e $R^2 = 0,81$. Estes modelos parametrizados servem como uma metodologia simples e rápida para auxiliar na interpretação de dados experimentais. Nossos modelos foram utilizados para estudar a inconsistência encontrada entre a dinâmica reorientacional calculada por dinâmica molecular e deduzida de experimentos de anisotropia de fluorescência resolvida no tempo do Trp113 da enzima subtilisina Carlsberg. Essa inconsistência era devida a que as estruturas amostradas pela dinâmica molecular e pelos experimentos eram distintas. Assim a dinâmica molecular amostrava preferencialmente conformações com o Trp113 parcialmente protegido do solvente (correspondente à conformação da estrutura cristalográfica), enquanto o experimento aparenta ser consistente com conformações com o Trp113 exposto ao solvente. Essas diferenças foram quantificadas pela comparação dos λ_{em} experimentais e calculados, estes últimos com os modelos parametrizados. Simulações de dinâmica molecular mantendo o Trp113 exposto ao solvente durante todo o tempo foram realizadas mediante o uso de restrições

Resumo

estruturais. Os λ_{em} foram calculados para as estruturas amostradas nessas simulações, sendo similares aos observados nos experimentos. Como era previsto, a dinâmica reorientacional calculada para o Trp113 exposto ao solvente foi muito mais rápida quando comparada à dinâmica observada para a estrutura cristalográfica. Desta forma, foi possível interpretar estruturalmente o resultado experimental, e sugerir que a conformação cristalográfica não é representativa do conjunto de estruturas amostrado em solução.

Abstract

Tryptophan fluorescence is highly sensitive to the environment, reason why it is up to the present day widely used to investigate the structure and dynamics of proteins. The fluorescence emission wavelength (λ_{em}) of tryptophan in proteins is very variable. There are proteins with low λ_{em} (e.g. Azurine, 308 nm), and others with high λ_{em} (e.g. Glucagon, 355 nm). The diversity of λ_{em} motivated theoretical studies to explain the tryptophan Stokes shift phenomenon of in proteins. One of the most relevant studies was performed by Vivian and Callis (2004), which proposed a hybrid QM-MM model with the capacity to predict the λ_{em} . The performance of the model of Vivian and Callis (2004) was good, with a standard deviation of the predictions with respect to experimental data (SD) of 6,87 nm, and a correlation coefficient of $R^2 = 0,69$. The use of these methods requires sophisticated computational protocols restricting their use in the interpretation of experimental data. In this thesis were developed classical parametric models based on the solvent accessible surface area, and based on the electrostatic interactions for the prediction of tryptophan λ_{em} . These classical models allow better predictions compared to the model based on first principles of Vivian and Callis (2004), at a much lower computational cost and complexity. For example, a model based on electrostatic interactions predicts the emission wavelengths with a SD = 4,89 nm and $R^2 = 0,81$. These parameterized models represent a simple and fast methodology that can be used in the interpretation of experimental data. Our models were used to study the inconsistency found between the reorientational dynamics calculated by molecular dynamics and deduced from time-resolved fluorescence anisotropy experiments of the Trp113 of the enzyme subtilisin Carlsberg. This inconsistency was due to the fact that the structures sampled by molecular dynamics and experiments were different. The molecular dynamics sampled preferably conformations with the Trp113 partially protected from the solvent (corresponding to the crystallographic structure), while the experiment appears to be consistent with conformations with Trp113 exposed to the solvent. These differences were quantified by comparing the experimental and calculated λ_{em} , the later obtained with the parameterized models. Molecular dynamics simulations keeping the Trp113 exposed to the solvent all the time were carried out through the use of structural constraints. The λ_{em} were calculated for the structures sampled

Abstract

in these simulations, being similar to those observed in experiments. As expected, the calculated reorientational dynamics for the solvent-exposed Trp113 was much faster when compared to the observed for the crystallographic structure. In this way, it was possible to structurally interpret the experimental results, and suggest that the crystallographic conformation is not representative of the set of structures sampled in solution.

Listas de Figuras

1.1	O diagrama de Jablonski. S_0 é o estado fundamental, S_1 e S_2 são os estados excitados singletos, T_1 e T_2 são os estados tripletos, IC é a conversão interna, e ISC é a conversão intersistema. As linhas verticais contínuas e as linhas onduladas indicam processos radiativos e não-radiativos, respectivamente.	31
1.2	Diagrama esquemático do fluorímetro, onde são mostrados os componentes básicos do aparelho. 1) Fonte de luz, 2) monocromador de excitação, 3) amostra, 4) monocromador de emissão, 5) sensor de detecção de luz e 6) sistema de aquisição dos dados.	36
1.3	Espectros de absorção (vermelho) e emissão (azul) dos aminoácidos A) fenilalanina, B) tirosina e C) triptofano em solução aquosa. Os dados experimentais dos espectros de absorção e emissão correspondem a Fasman et al. (1976) [19] e Chen et al. (1972) [18], respectivamente. Os espectros estão disponíveis no site da PhotochemCAD [20].	38
1.4	Diagrama esquemático do fluorímetro usado nas medidas de anisotropia de fluorescência resolvida no tempo. 1) Fonte de luz laser, 2) discriminador, 3) polarizador, 4) amostra, 5) divisor de feixe, 6) monocromador, 7) tubo fotomultiplicador, 8) instrumento de detecção integrada, 9) sinal de fluorescência, e 10) processamento do sinal de fluorescência	41
1.5	Probabilidade de excitação do fluoróforo em função do ângulo entre o vetor de campo elétrico da luz polarizada incidente e o dipolo de absorção do fluoróforo.	42
1.6	Diagramas das orientações dos momentos de dipolo de absorção (μ_a) e emissão (μ_e).	46
2.1	Parâmetros usados na construção dos potenciais ligantes. A) Estiramento das ligações, B) deformação angular, C) deformação diedral, D) potencial de Urey-Bradley e E) deformação de diedros impróprios	53
2.2	Representação esquemática do funcionamento do algoritmo Verlet-velocidade.	59

Lista de Figuras

2.3	Representação esquemática da superfície de energia potencial de uma proteína. Dois mínimos locais são mostrados. Uma ampliação do mínimo I revela alguns detalhes da superfície de energia potencial.	60
4.1	Estrutura da proteína apo-repressora do <i>trp</i> de <i>E.coli</i> . Os triptofanos Trp19 (em azul) e Trp99 (em vermelho) são destacados.	64
4.2	Dependência dos espectros de fluorescência dos triptofanos da proteína apo-repressora do <i>trp</i> de <i>E.coli</i> com respeito às condições desnaturantes do solvente. A) Apo-repressor W99F, e B) Apo-repressor W19F . O desnaturante usado é ureia e as concentrações são as seguintes: 0 M (—), 3 M (---), 5 M (.....), e 9 M (---). Adaptado de Royer et al. (1993) [7].	65
4.3	Evolução temporal da SASA do indol na Parvalbumina (azul), Nuclease de estafilococo (verde) e Glucagon (vermelho). O aumento da área acessível ao solvente está associado ao incremento no comprimento de onda de emissão de fluorescência.	69
4.4	Correlação entre o comprimento de onda de emissão de fluorescência do triptofano experimental (λ_{em}) e o valor médio da SASA do indol em 19 proteínas. As proteínas correspondem aos da Tabela 4.1. A linha cinza mostra a tendência dos dados.	70
4.5	Comprimento de onda de emissão de fluorescência experimental (λ_{em}^{exp}) versus o calculado (λ_{em}^{calc}) usando os modelos baseados na área acessível ao solvente. A) Modelo 1: Baseado na $\langle SASA \rangle$ do grupo indol do triptofano. B) Modelo 2: Baseado na decomposição da $\langle SASA \rangle$ dos grupos benzeno incompleto e pirrol completo do indol. C) Modelo 3: Baseado na decomposição da $\langle SASA \rangle$ dos grupos benzeno completo e pirrol incompleto do indol. D) Modelo 4: Baseado na decomposição da $\langle SASA \rangle$ dos grupos benzeno completo e pirrol completo do indol.	72
4.6	Representação esquemática do indol. Os átomos de carbono e nitrôgenio estão numerados para identificar os átomos que farão parte do anel benzênico e pirrólico durante os cálculos da SASA ou interações eletrostáticas.	72

Lista de Figuras

4.7	Representação esquemática do método de validação da robustez dos modelos. O número total de dados é $N = 4$. Para este caso em particular, o processo de validação considera deixar um dado fora ($k = 1$) para construir os modelos, assim o conjunto de dados de treinamento serão de $N - k = 3$ dados. O dado deixado fora será usado para avaliar a robustez do modelo em relação aos dados de treinamento usados para a parametrização.	80
5.1	A) Orientação dos dipolos do estado fundamental (GS) e L_a do indol. B) Correlação entre os λ_{em} experimentais e calculados, segundo o modelo híbrido QM-MM desenvolvido por Vivian e Callis (2004) [74].	86
5.2	Evolução temporal das interações eletrostáticas do indol com o ambiente (água + proteína) nas proteínas: Parvalbumina (azul), Nuclease de Estafilococo (verde) e Glucagon (vermelho).	88
5.3	Correlação entre o comprimento de onda de emissão de fluorescência experimental e o valor médio das interações eletrostáticas do indol com o ambiente (água + proteína) para 19 proteínas que contém um único resíduo de triptofano.	88
5.4	Comprimento de onda de emissão de fluorescência experimental (λ_{em}^{exp}) versus o calculado (λ_{em}^{calc}) usando os modelos baseados nas interações eletrostáticas. A) Modelo 1: Baseado na interação eletrostática do grupo indol com o ambiente (água + proteína). B) Modelo 2: Baseado nas interações eletrostáticas do grupo indol com a água e a proteína. C) Modelo 3: Baseado nas interações eletrostáticas do benzeno incompleto-ambiente (água + proteína) e pirrol completo-ambiente (água + proteína). D) Modelo 4: Baseado na decomposição das interações eletrostáticas do benzeno incompleto e pirrol completo com a água e a proteína.	96

Lista de Figuras

5.5	Comprimento de onda de emissão de fluorescência experimental (λ_{em}^{exp}) versus o calculado (λ_{em}^{calc}) usando os modelos baseados nas interações eletrostáticas. A) Modelo 5: Baseado nas interações eletrostáticas do benzeno completo - ambiente (água + proteína) e pirrol incompleto - ambiente (água + proteína). B) Modelo 6: Baseado na decomposição das interações eletrostáticas do benzeno completo e pirrol incompleto com a água e a proteína. C) Modelo 7: Baseado nas interações eletrostáticas do benzeno completo - ambiente (água + proteína) e pirrol completo - ambiente (água + proteína). D) Modelo 8: Baseado na decomposição das interações eletrostáticas do benzeno completo e pirrol completo com a água e a proteína.	101
6.1	A) Representação esquemática da estrutura secundária da enzima subtilisina Carlsberg (sC). As α -hélices estão representadas por cilindros azuis, as folhas- β como flechas vermelhas. As regiões conservadas (linhas contínuas em cinza) e variáveis (linhas descontínuas em cinza) estão mostradas de acordo ao reportado por Siezen, <i>et al.</i> para a família das subtilases [86]. B) Estrutura cristalográfica da enzima subtilisina Carlsberg (PDB id. 1SBC) mostrando as α -hélices (cilindros azuis) e folhas- β (flechas vermelhas), assim como a tríade catalítica, Asp32, His64 e Ser221 (verde).	109
6.2	Fluorescência do Trp livre em água. As distribuições dos λ_{em} do Trp em água foram calculados segundo os modelos parametrizados baseados na A) SASA do benzeno/pirrol e B) interação eletrostática do indol com o sistema (água + proteína). C) Anisotropia de fluorescência resolvida no tempo do Trp livre em água Os dados experimentais e calculados estão desenhados em azul e vermelho, respectivamente.	118
6.3	A) Conformação da enzima subtilisina Carlsberg com o Trp113 parcialmente protegido do solvente. B) Representação da acessibilidade ao solvente do resíduo Trp113. C) Representação da distância característica entre os átomos Ala29-C β e Trp113-CZ3 na conformação da sC com o Trp113 parcialmente protegido do solvente ($\sim 5 \text{ \AA}$).	119

Lista de Figuras

6.4	Fluorescência da conformação da subtilisina Carlsberg com o resíduo Trp113 parcialmente protegido do solvente. A) e B) Espectros de fluorescência experimentais (azul) e distribuições dos λ_{em} calculados (vermelho) do Trp113 da sC em água. Os cálculos dos λ_{em} foram feitos usando o modelo parametrizado baseado na SASA do benzeno/pirrol (em A) ou o modelo baseado nas interações eletrostáticas do indol com a água/proteína (em B). C) Anisotropia de fluorescência resolvida no tempo do Trp113 da sC em água. As curvas azuis contínuas e tracejadas correspondem aos dados experimentais reportados por Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93], respectivamente.	120
6.5	A) Conformação da enzima subtilisina Carlsberg com o Trp113 completamente exposto ao solvente. B) Representação da acessibilidade ao solvente do resíduo Trp113. C) Representação da distância característica entre os átomos Ala29-C β e Trp113-CZ3 na conformação da sC com o Trp113 completamente exposto ao solvente ($\sim 12 \text{ \AA}$).	122
6.6	Fluorescência da conformação da subtilisina Carlsberg com o resíduo Trp113 exposto ao solvente. A) e B) Espectro de fluorescência experimental (azul) e distribuição dos λ_{em} calculados (vermelho) do Trp113 da sC em água. Os cálculos dos λ_{em} foram feitos usando o modelo parametrizado baseado na SASA do benzeno/pirrol (em A) ou o modelo baseado nas interações eletrostáticas do indol com a água/proteína (em B). C) Anisotropia de fluorescência resolvida no tempo do Trp113 da sC em água. As curvas azuis contínuas e tracejadas correspondem aos dados experimentais reportados por Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93], respectivamente.	124
A.1	Gráficas do RMSD do backbone da proteína com respeito as estructuras iniciais usadas na produção das simulações. A) Azurina, B) Parvalbumina, C) Mioglobina, D) Ribonuclease T1, E) FKBP12, F) T4-lisozima W126Y W138Y, G) T4-lisozima W126Y W158Y, H) T4-lisozima W138Y W158Y, I) subtilisina Carlsberg, J) Nucleasa de estafilococo.	142

Lista de Figuras

Listas de Tabelas

1.1	Rendimento quântico (Φ_F), e comprimentos de onda máximo de absorção (λ_{abs}) e emissão (λ_{em}) dos aminoácidos fenilalanina (Phe), tirosina (Tyr) e triptofano (Trp) em solução aquosa.	38
1.2	Diversidade dos comprimentos de onda de emissão de fluorescência (λ_{em}) do triptofano em proteínas. Todas as proteínas possuem um único resíduo de triptofano. Os números de acesso das estruturas no Protein Data Bank (PDB id.) são mostrados.	39
4.1	Proteínas usadas, números de acesso das estruturas (PDB id.), comprimentos de ondas de emissão experimentais (λ_{em}^{exp}), e áreas acessíveis ao solvente calculadas a partir das simulações de dinâmica molecular. $SASA_{ind}$, $SASA_{ben,i}$, $SASA_{pyr,i}$, $SASA_{ben,c}$, $SASA_{pyr,c}$ são as áreas acessíveis ao solvente do indol, benzeno incompleto, pirrol incompleto, benzeno completo e pirrol completo, respectivamente. $\langle \rangle$ representa a média sobre a simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as SASAs em \AA^2	76
4.2	Comprimentos de onda de emissão de fluorescência experimental e calculado do triptofano: λ_{em}^{exp} é o comprimento de onda de emissão de fluorescência experimental. $\lambda_{em}^{(1-4)}$ e $\Delta\lambda_{em}^{(1-4)}$ são o comprimento de onda de emissão de fluorescência calculado e o desvio relativo ao valor experimental, respectivamente, usando os modelos baseados na área acessível ao solvente 1 ao 4. As unidades dos comprimentos de onda e os desvios são os nm.	77
4.3	Modelos baseados na área acessível para a predição de comprimentos de onda de emissão de fluorescência por simulações de dinâmica molecular. R^2 é o coeficiente de correlação, SD é o desvio padrão das predições com respeito aos dados experimentais. $SASA_{ind}$, $SASA_{ben,c}$, $SASA_{ben,i}$, $SASA_{pyr,c}$, $SASA_{pyr,i}$ são as áreas acessíveis ao solvente do indol, benzeno completo, benzeno incompleto, pirrol completo e pirrol incompleto, respectivamente. As unidades das SASAs são \AA^2	78

Lista de Tabelas

4.4	Robustez do modelo baseado na SASA do indol (Modelo 1) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 1 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.	82
5.1	Proteínas usadas, códigos de acesso das estruturas (PDB id.), comprimentos de onda de emissão de fluorescência experimentais (λ_{em}^{exp}), e interações eletrostáticas calculadas a partir das simulações de dinâmica molecular. Q_T , Q_w e Q_p são as interações eletrostáticas entre o indol e o ambiente (água + proteína), o indol e a água, e o indol e a proteína, respectivamente. O símbolo $\langle \rangle$ representa a média sobre as simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as interações eletrostáticas em $kJ \times mol^{-1}$	90
5.2	Proteínas usadas, códigos de acesso das estruturas (PDB id.), comprimentos de onda de emissão de fluorescência experimentais (λ_{em}^{exp}), e interações eletrostáticas calculadas a partir das simulações de dinâmica molecular. $Q_{ben-i,w}$, $Q_{ben-i,p}$, $Q_{pyr-c,w}$ e $Q_{pyr-c,p}$ são as interações eletrostáticas entre o benzeno incompleto do indol e a água, o benzeno incompleto do indol e a proteína, o pirrol completo do indol e a água, o pirrol completo do indol e a proteína e respectivamente. As interações eletrostáticas entre o benzeno incompleto e o ambiente (água + proteína) são computadas como $Q_{ben-i} = Q_{ben-i,w} + Q_{ben-i,p}$. Da mesma maneira, as interações eletrostáticas entre o pirrol completo e o ambiente (água + proteína) são computadas como $Q_{pyr-c} = Q_{pyr-c,w} + Q_{pyr-c,p}$. O símbolo $\langle \rangle$ representa a média sobre as simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as interações eletrostáticas em $kJ \times mol^{-1}$. . .	91

Lista de Tabelas

5.3	Proteínas usadas, códigos de acesso das estruturas (PDB id.), comprimentos de onda de emissão de fluorescência experimentais (λ_{em}^{exp}), e interações eletrostáticas calculadas a partir das simulações de dinâmica molecular. $Q_{ben-c,w}$, $Q_{ben-c,p}$, $Q_{pyr-i,w}$ e $Q_{pyr-i,p}$ são as interações eletrostáticas entre o benzeno completo do indol e a água, benzeno completo do indol e a proteína, pirrol incompleto do indol e a água e pirrol incompleto do indol e a proteína. respectivamente. As interações eletrostáticas entre o benzeno completo e o ambiente (água + proteína) são computados como $Q_{ben-c} = Q_{ben-c,w} + Q_{ben-c,p}$. Da mesma maneira, as interações eletrostáticas entre o pirrol incompleto e o ambiente (água + proteína) são computados como $Q_{pyr-i} = Q_{pyr-i,w} + Q_{pyr-i,p}$. O símbolo $\langle \rangle$ representa a média sobre as simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as interações eletrostáticas em $kJ \times mol^{-1}$	92
5.4	Comprimentos de onda de emissão de fluorescência experimentais e calculados do triptofano: λ_{em}^{exp} é o comprimento de onda de emissão de fluorescência experimental. $\lambda_{em}^{(1-4)}$ e $\Delta\lambda_{em}^{(1-4)}$ são os comprimentos de onda de emissão de fluorescência calculados e os desvios relativos aos valores experimentais, respectivamente, usando os modelos baseados nas interações eletrostáticas 1 ao 4. Os comprimentos de onda e os desvios estão em nm.	97
5.5	Comprimentos de onda de emissão de fluorescência experimentais e calculados do triptofano: λ_{em}^{exp} é o comprimento de onda de emissão de fluorescência experimental. $\lambda_{em}^{(5-8)}$ e $\Delta\lambda_{em}^{(5-8)}$ são o comprimento de onda de emissão de fluorescência calculado e o desvio relativo ao valor experimental, respectivamente, usando os modelos baseados nas interações eletrostáticas 5 ao 8. Os comprimentos de onda e os desvios estão em nm.	102

Lista de Tabelas

5.6	Modelos baseados nas interações eletrostáticas para a predição dos comprimentos de onda de emissão de fluorescência (λ_{em}) do triptofano em proteínas por meio de simulações de dinâmica molecular. R^2 é o coeficiente de correlação quadrático de Pearson e SD é o desvio padrão das previsões com respeito aos dados experimentais. Q_T , Q_w , Q_p , Q_{ben-i} , Q_{ben-c} , Q_{pyr-i} , Q_{pyr-c} , $Q_{ben-i,w}$, $Q_{ben-i,p}$, $Q_{pyr-i,w}$, $Q_{pyr-i,p}$, $Q_{ben-c,w}$, $Q_{ben-c,p}$, $Q_{pyr-c,w}$ e $Q_{pyr-c,p}$ são as interações eletrostáticas entre o indol - ambiente (água + proteína), indol - água, indol - proteína, benzeno incompleto - ambiente (água + proteína), benzeno completo - ambiente (água + proteína), pirrol incompleto - ambiente (água + proteína), pirrol completo - ambiente (água + proteína), benzeno incompleto - água, benzeno incompleto - proteína, pirrol incompleto - água, pirrol incompleto - proteína, benzeno completo - água, benzeno completo - proteína, pirrol completo - água e pirrol completo - proteína, respectivamente. As unidades das interações eletrostáticas são $\text{kJ} \times \text{mol}^{-1}$	103
5.7	Robustez do Modelo 1 baseado nas interações eletrostáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 1 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.	106
6.1	Condições experimentais das medidas de fluorescência do Trp113 da subtilisina Carlsberg. A proteína sC foi adquirida comercialmente e passos de purificação foram realizados antes de realizar as medidas de fluorescência. HPLC (Cromatografia líquida de alta eficiência). FPLC (Cromatografia líquida de proteína rápida). λ_{ex} e λ_{em} são os comprimentos de onda de excitação e emissão, respectivamente.	110

Lista de Tabelas

6.2	Parâmetros experimentais e calculados da anisotropia de fluorescência resolvida no tempo do triptofano livre em água. λ_{em} é o comprimento de onda usado para registrar os decaimentos de anisotropia. O λ_{em} foi calculado usando ^(a) o modelo baseado na SASA do benzeno/pirrol e ^(b) nas interações eletrostáticas do indol com a água/proteína. O modelo mono-exponencial usado para o ajuste do decaimento da anisotropia de fluorescência é: $r(t) = r_0 \times e^{(-t/\tau_1)}$, sendo τ_1 o tempo de reorientação do fluoróforo. Os λ_{em} calculados correspondem a média dos comprimentos de onda calculados a partir de 30 simulações de dinâmica molecular. O ajuste exponencial da anisotropia foi aplicado à média das 30 curvas de decaimento de anisotropia de fluorescência calculadas a partir das simulações de dinâmica molecular.	119
6.3	Parâmetros experimentais e calculados da anisotropia de fluorescência resolvida no tempo do Trp113 da subtilisina Carlsberg em água. λ_{em} é o comprimento de onda usado para registrar os decaimentos de anisotropia. O λ_{em} foi calculado usando ^(a) o modelo baseado na SASA do benzeno/pirrol e ^(b) nas interações eletrostáticas do indol com a água/proteína. O modelo bi-exponencial usado para o ajuste do decaimento da anisotropia de fluorescência foi: $r(t) = r_0 \times (\beta_1 e^{-t/\tau_1} + \beta_2 e^{-t/\tau_2})$, sendo β_1 e β_2 os termos pré-exponenciais, e τ_1 e τ_2 os tempos de reorientação das componentes rápidas e lentas, respectivamente. Os λ_{em} calculados correspondem a média dos comprimentos de onda calculados a partir de 30 simulações de dinâmica molecular. O ajuste exponencial da anisotropia foi aplicado à média das 30 curvas de decaimento de anisotropia de fluorescência calculadas a partir das simulações de dinâmica molecular. ^(†) Conformação da sC com o Trp113 parcialmente protegido do solvente. ^(‡) Conformação da sC com o Trp113 completamente exposto ao solvente.	126

Lista de Tabelas

- B.1 Robustez do modelo baseado na SASA do benzeno incompleto e pirrol completo (Modelo 2) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 2 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm. 145
- B.2 Robustez do modelo baseado na SASA do benzeno completo e pirrol incompleto (Modelo 3) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 3 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm. 146
- B.3 Robustez do modelo baseado na SASA do benzeno completo e pirrol completo (Modelo 4) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 4 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm. 147

Lista de Tabelas

Lista de Tabelas

Lista de Tabelas

List of Abbreviations

λ_{em}	Comprimento de onda emissão de fluorescência
λ_{em}^{exp}	Comprimento de onda emissão de fluorescência experimental
λ_{em}^{calc}	Comprimento de onda emissão de fluorescência calculado
ABF	<i>Adaptive Biasing Force</i>
DNA	<i>Deoxyribonucleic acid</i>
MD	<i>Molecular dynamics</i>
NAMD	<i>Not (just) Another Molecular Dynamics program</i>
PDB	<i>Protein Data Bank</i>
Q_T	Interação eletrostática entre o indol e o ambiente (água + proteína)
Q_p	Interação eletrostática entre o indol e a proteína
Q_w	Interação eletrostática entre o indol e a água
QM-MM	Simulações híbridas mecânica quântica-mecânica molecular
RMSD	<i>Root-mean-square deviation</i>
SASA	<i>Solvent accessible surface area</i>
sC	subtilisina Carlsberg
Trp	Triptofano
Trp113	Triptofano da posição 113
TCSPC	<i>Time-correlated single-photon counting</i>

Sumário

1	Fluorescência	30
1.1	Uma breve introdução à fluorescência	30
1.2	Conceitos básicos de fluorescência	31
1.2.1	Diagrama de Jablonski	31
1.2.2	Propriedades da fluorescência	32
1.3	Fluorescência em proteínas	35
1.3.1	Fluorescência estacionária	36
1.3.2	Anisotropia de fluorescência resolvida no tempo	39
2	Simulações de dinâmica molecular	47
2.1	Simulações de dinâmica molecular	47
2.1.1	Condições iniciais	48
2.1.2	Potenciais de interação	49
2.1.3	Integrando as equações de movimento	54
2.2	Simulações ABF	59
3	Objetivos	62
4	Modelos paramétricos baseados na área acessível ao solvente para computar os comprimentos de onda de emissão do triptofano	63
4.1	Área acessível ao solvente	63
4.2	Metodologia	66
4.2.1	Simulações de dinâmica molecular	66
4.2.2	Cálculo da área acessível ao solvente	67
4.3	Resultados e discussões	68
4.3.1	Relação entre o comprimento de onda de emissão e a área acessível ao solvente do triptofano	68
4.3.2	Construção dos modelos	70
4.3.3	Robustez dos modelos	79

Sumário

4.4	Conclusões	81
5	Modelos paramétricos baseados na interação eletrostática para computar os comprimentos de onda de emissão do triptofano	84
5.1	Introdução	84
5.1.1	O papel das interações eletrostáticas nos λ_{em} do Trp	84
5.2	Metodologia	86
5.2.1	Simulações de dinâmica molecular	86
5.2.2	Cálculo das interações eletrostáticas	86
5.3	Resultados e discussões	87
5.3.1	Interações eletrostáticas do indol	87
5.3.2	Construção dos modelos	93
5.3.3	Robustez dos modelos	104
5.4	Conclusões	106
6	Simulações de dinâmica molecular da anisotropia de fluorescência resolvida no tempo da subtilisina Carlsberg	108
6.1	Introdução	108
6.1.1	Estrutura e função da enzima subtilisina Carlsberg	108
6.1.2	Fluorescência da subtilisina Carlsberg	109
6.1.3	Anisotropia de fluorescência resolvida no tempo da subtilisina Carlsberg	112
6.2	Metodologia	113
6.2.1	Simulações de dinâmica molecular	113
6.2.2	Simulações ABF	114
6.2.3	Cálculos do λ_{em}	115
6.2.4	Cálculos dos decaimentos da anisotropia de fluorescência	115
6.3	Resultados e discussões	116
6.3.1	Decaimento da anisotropia do triptofano livre em água	116
6.3.2	Decaimento da anisotropia de fluorescência do Trp113 da subtilisina Carlsberg parcialmente protegido	117

Sumário

6.3.3	Decaimento da anisotropia de fluorescência do Trp113 da subtilisina Carlsberg completamente exposto ao solvente	121
6.4	Conclusões	125
7	Conclusões finais	128
8	Perspectivas	130
	Bibliografia	131
A	RMSD	141
B	Tabelas do Capítulo 4	144
C	Tabelas do Capítulo 5	148

Capítulo 1

Fluorescência

1.1 Uma breve introdução à fluorescência

O primeiro registro conhecido da fluorescência foi reportado pelo médico espanhol Nicolas Monardes. Em 1565, ele observou a emissão de luz da cor azul de uma infusão de uma planta chamada *Lignum Nephriticum*. Nos anos seguintes, muitas outras substâncias com capacidade de emitir luz foram descobertas, incluindo as soluções de quinina [1].

Em 1852, Sir George Stokes desenvolveu um experimento simples mas de grande importância, que permitiria explicar o processo de emissão de luz. No seu experimento a luz natural era dividida com a ajuda de um prisma, logo a solução de quinina, a qual estava contida num tubo, era colocada em diversas regiões do espectro de luz. Quando a solução de quinina era colocada na região da luz visível nenhuma luz era emitida pela substância. No entanto, quando a solução era colocado na região não-visível da luz (região ultravioleta) a solução de quinina tornava-se azul. Isso chamou a atenção de Stokes e utilizando as palavras dele, “*parecia que a luz se originava da escuridão*”. Uma das primeiras conclusões que Stokes obteve a partir dos seus experimentos foi que a emissão é provocada pela absorção prévia de luz de um comprimento de onda determinado. Adicionalmente, Stokes notou que o comprimento de onda de absorção era menor que o da emissão de luz, nascendo assim o famoso termo chamado desvio de Stokes [2].

Até esse momento, Stokes chamava o fenômeno de emissão de luz de *reflexão dispersiva*. Em 1853, Stokes resolveu utilizar definitivamente o nome de *fluorescência* para se referir ao processo de emissão de luz [3]. Esse nome foi derivado da pedra *fluorspar*, um mineral composto de fluoreto de cálcio e que também apresentava emissão de luz [1].

Durante a segunda guerra mundial, aparelhos práticos para medir a fluorescência foram desenvolvidos com a finalidade de monitorar drogas antimaláricas [4]. A partir desse

momento a fluorescência ganhou muita importância e aplicabilidade na área científica e tecnológica. Entre as aplicações que encontramos, hoje em dia, temos: explorar estrutura e dinâmica das proteínas [5, 6, 7], cinética enzimática [8, 9], observar micro- e nano-estruturas celulares [10, 11, 12], entre outras.

1.2 Conceitos básicos de fluorescência

1.2.1 Diagrama de Jablonski

Também conhecido como o diagrama de Perrin-Jablonski, este é usado para representar esquematicamente os processos de absorção de fótons, conversões internas, fluorescência, cruzamento intersistemas, fosforescência e transições tripleto-tripleto (Figura 1.1) [1, 4].

A continuação vamos a descrever brevemente cada um desses processos.

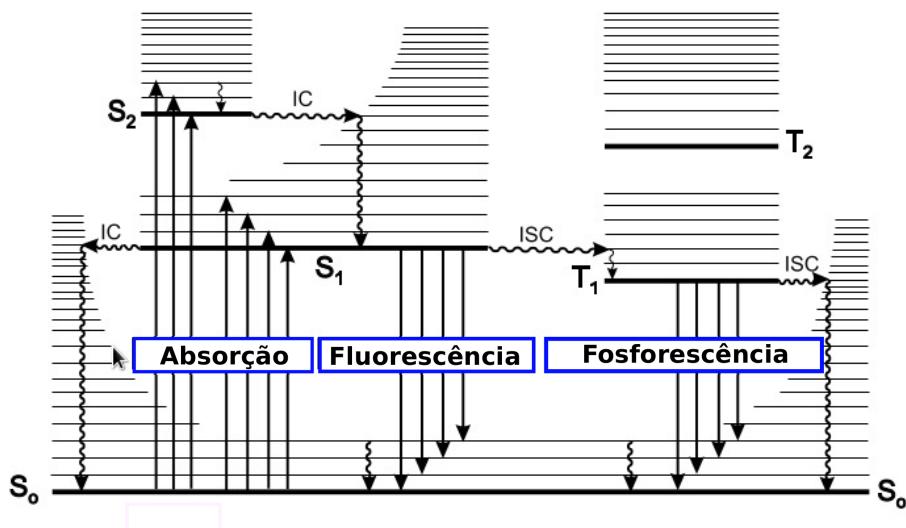


Figura 1.1: O diagrama de Jablonski. S_0 é o estado fundamental, S_1 e S_2 são os estados excitados singletos, T_1 e T_2 são os estados tripletos, IC é a conversão interna, e ISC é a conversão intersistema. As linhas verticais contínuas e as linhas onduladas indicam processos radiativos e não-radiativos, respectivamente.

- **Conversão interna:** É uma transição não-radiativa entre dois estados eletrônicos que possuem a mesma multiplicidade do spin.
- **Fluorescência:** É o processo mediante o qual há emissão de fótons devido a uma relaxação do estado excitado ao fundamental, sendo que estes estados possuem a mesma a

multiplicidade do spin. (i.e. $S_1 \rightarrow S_0$)

- **Cruzamento intersistema:** É um processo não-radiativo entre dois níveis vibracionais da mesma energia, mas pertencentes a estados eletrônicos de diferente multiplicidade. Em princípio estas transições são proibidas, mas o acoplamento órbita-spin pode ser suficientemente grande para permiti-las.
- **Fosforescência:** As transições a partir do estado T_1 podem ser por meio de de-excitações não-radiativas ou por de-excitações conhecidas como fosforescência. As primeiras acontecem em meios solventes, onde o número de colisões favorece o cruzamento intersistema e o posterior relaxamento vibracional ao estado S_0 . Por outro lado, a fosforescência acontece a baixas temperaturas e em meios rígidos, sendo o tempo de vida deste processo longo e pode ser observado na escala de tempo dos segundos até minutos.

1.2.2 Propriedades da fluorescência

Ao observar o espectro de fluorescência, é possível de notar duas características importantes. A primeira delas é a intensidade máxima do espectro, e a segunda é o comprimento de onda no qual aparece o máximo de emissão. Tecnicamente, essas características estão relacionadas ao rendimento quântico e ao desvio de Stokes, respectivamente. Outras propriedades da fluorescência só foram mais recentemente medidas graças ao avanço das técnicas experimentais. Essas características são o tempo de vida e o tempo de reorientação dos fluoróforos.

Rendimento quântico

Imaginemos que um determinado fluoróforo é excitado por um curto pulso de luz. Certo número de moléculas do fluoróforo passarão para o estado excitado S_1 e após certo tempo voltarão ao estado fundamental S_0 mediante processos radiativos ou não-radiativos. Se modelarmos esse processo, podemos ver que a taxa de decaimento do fluoróforo no estado excitado (A^*) depende das constantes de decaimento radiativo (k_r^S) e não-radiativo (k_{nr}^S).

$$-\frac{d[A^*]}{dt} = (k_r^S + k_{nr}^S)[A^*] \quad (1.2.1)$$

Por outro lado, a intensidade de fluorescência (i_F) é a quantidade de fótons emitidos na passagem do estado excitado ao fundamental ($A^* \rightarrow A + \text{fóton}$), sendo proporcional à concentração de moléculas que ainda estão no estado excitado ($i_F \propto k_r^S[A^*]$). A intensidade de fluorescência observada experimentalmente (I_F) é proporcional a i_F , e depende das condições instrumentais. Também é importante dizer que I_F é obtido normalmente usando alguma escala arbitrária escolhida no experimento.

O rendimento quântico (Φ_F) é uma medida da eficiência da conversão dos fótons absorvidos durante a excitação em fótons de emissão, e pode ser expressado mediante a Equação 1.2.2.

$$\Phi_F = \frac{Fótons_{em}}{Fótons_{ex}} = \frac{k_r^S}{k_r^S + k_{nr}^S} \quad (1.2.2)$$

onde $Fótons_{ex}$ e $Fótons_{em}$ são os números de fótons absorvidos durante a excitação e os liberados durante a emissão de fluorescência, respectivamente. As constantes de velocidade k_r^S e k_{nr}^S despopulam o estado excitado, e uma forma alternativa, para o cálculo do rendimento quântico, baseada na fração dessas constantes é possível (Equação 1.2.2).

Experimentalmente existem duas formas populares de medir o rendimento quântico de moléculas fluorescentes. Essas metodologias são conhecidas como o método pontual e o método comparativo [4, 13]. Ambas precisam de alguma amostra referência, que possua rendimento quântico conhecido (Φ_R).

No método pontual, medidas de absorbância e fluorescência numa concentração fixa do fluoróforo problema e referência são necessários para calcular o rendimento quântico do fluoróforo, usando:

$$\Phi_F = \Phi_R \frac{I_F}{I_{FR}} \frac{I_{AR}}{I_A} \quad (1.2.3)$$

sendo I_F e I_{FR} as intensidades de fluorescência, e I_A e I_{AR} as intensidades de absorção do fluoróforo problema e referência, respectivamente.

No método comparativo, é necessário de realizar medidas de absorbância e fluo-

rescência do fluoróforo problema e referência usando diferentes concentrações. Isso permitirá construir gráficas que mostrem a dependência da intensidade de fluorescência em função da intensidade de absorção numa faixa de concentrações do fluoróforo. Os coeficientes angulares dessas relações serão calculados para o fluoróforo problema (M) e referência (m_R). Assim, o rendimento quântico será calculado usando:

$$\Phi_F = \Phi_R \frac{M}{m_R} \quad (1.2.4)$$

O rendimento quântico de diversos fluoróforos foram estimados usando esses métodos. O triptofano livre em água, por exemplo, apresenta um rendimento quântico de 0,14, enquanto que em proteínas este resíduo apresenta uma variabilidade de rendimentos que vão de $\sim 0,0$ até $\sim 0,3$. Essa variabilidade é principalmente decorrente de eventos de transferência de carga entre o grupo indol e a ligação peptídica presente nas proteínas [14, 15]. As cadeias laterais dos aminoácidos lisina, glutamina, asparagina, histidina, aspartate, glutamate, tirosina e cisteína também estão envolvidos na supressão da fluorescência do triptofano [16].

Tempo de vida

O tempo de vida do fluoróforo (τ_S) está definido como o tempo médio que a molécula fica no estado excitado antes de voltar ao estado fundamental. Matematicamente, podemos expressar o tempo de vida pela Equação 1.2.5.

$$\tau_S = \frac{1}{k_r^S + k_{nr}^S} \quad (1.2.5)$$

Se compararmos as Equações 1.2.2 e 1.2.5, notamos que existe uma relação entre o rendimento quântico e o tempo de vida da fluorescência (Equação 1.2.6).

$$\tau_S = \Phi_F / k_r^S \quad (1.2.6)$$

ou seja, o rendimento quântico é o tempo de vida multiplicado pela taxa de decaimento radiativa.

Desvio de Stokes

Observando no diagrama de Jablonski (Figura 1.1) podemos notar que a energia de excitação é maior comparada à de emissão de fluorescência. Experimentalmente, essa diferença é visualizada entre os máximos dos espectros de absorção e emissão. Essa diferença de energia é conhecida como desvio de Stokes [2]. O desvio de Stokes é um importante parâmetro que pode ser utilizado para obter informações sobre os estados excitados [1, 4]. Essa perda de energia é devida a vários processos dinâmicos que acontecem durante a excitação do fluoróforo. Um desses processos está relacionado à polaridade do solvente. Assim temos que quando o solvente for polar (por exemplo, a água), o processo de relaxamento ao redor do fluoróforo será favorável e portanto a energia do estado excitado será menor [4].

Essa dependência do desvio de Stokes com a polarizabilidade do solvente é considerado no modelo de Lippert-Mataga (Equação 1.2.7) [4].

$$\nu_A - \nu_E = \frac{2}{hc} \left[\frac{\varepsilon - 1}{2\varepsilon - 1} - \frac{n^2 - 1}{2n^2 + 1} \right] \frac{(\mu_E - \mu_G)^2}{a^3} \quad (1.2.7)$$

sendo ν_A e ν_E as energias de absorção e emissão, respectivamente. μ_E e μ_G são os dipolos de emissão e absorção respectivamente. ε e n são a constante dielétrica e o índice de refração do solvente, respectivamente.

1.3 Fluorescência em proteínas

A fluorescência tem sido amplamente utilizada para investigar a estrutura e a dinâmica das proteínas. Os métodos usados podem ser comumente divididos em dois: fluorescência estacionária e resolvida no tempo. Nesta seção descreveremos estas modalidades com a finalidade de conhecer detalhes, como a instrumentação, e assim entender como são obtidos os dados experimentais.

1.3.1 Fluorescência estacionária

Arranjo experimental

O espectrofluorímetro, ou simplesmente fluorímetro, é o instrumento usado para realizar as medidas de fluorescência. Os componentes básicos de um fluorímetro são a lâmpada de excitação, os monocromadores de excitação e emissão e o sistema de aquisição de dados (Figura 1.2).

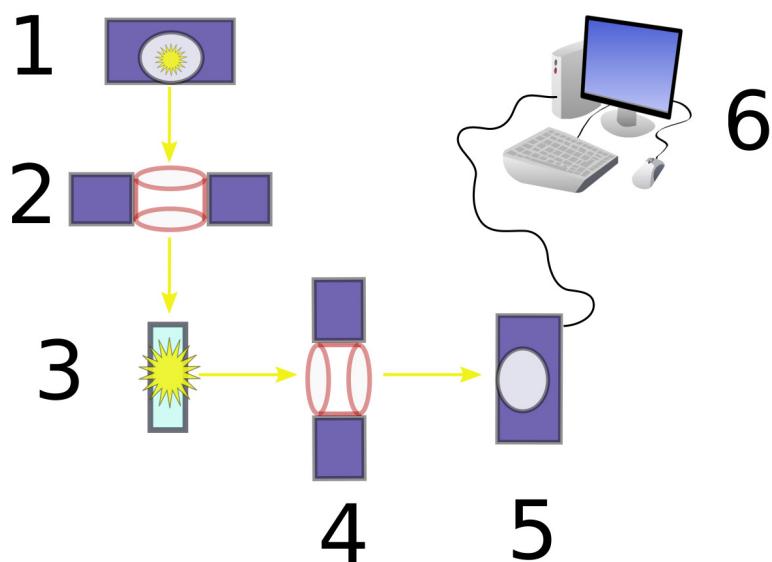


Figura 1.2: Diagrama esquemático do fluorímetro, onde são mostrados os componentes básicos do aparelho. 1) Fonte de luz, 2) monocromador de excitação, 3) amostra, 4) monocromador de emissão, 5) sensor de detecção de luz e 6) sistema de aquisição dos dados.

A fonte de luz de excitação é geralmente uma lâmpada de arco de xenônio de alta pressão. Esta lâmpada tem a capacidade de fornecer luz contínua na faixa de 250 nm até o espectro infravermelho. Os monocromadores servem para selecionar os comprimentos de onda de excitação e emissão. Estes permitem o escaneamento automático dos comprimentos de onda. O sistema de detecção da fluorescência está composto por tubos fotomultiplicadores e aparelhos eletrônicos específicos de detecção. Como é observado na Figura 1.2, a fluorescência é registrada num ângulo de 90° com respeito à luz de excitação. Essa geometria na aquisição de dados é necessária para prevenir que a luz de excitação transmitida seja confundida com a fluorescência. Os dados coletados pelo sistema de detecção são transferidos ao computador e guardados digitalmente.

Características espectrais dos aminoácidos

A fenilalanina (F), tirosina (Y) e o triptofano (W) são os aminoácidos naturais com capacidade de emitir fluorescência. Os espectros de absorção e emissão destes aminoácidos são distintos e são caracterizados, principalmente, pelo seus comprimentos de onda máximos (Figura 1.3).

O espectro de absorção da fenilalanina apresenta três picos característicos localizados em ~ 252 , 258 e 263 nm (Figura 1.3A). A maior absorção acontece em 258 nm (chamaremos λ_{abs} ao comprimento de onda de máxima absorção). Os espectros de absorção da tirosina e do triptofano estão mostrados nas Figuras 1.3B e C, respectivamente. O λ_{abs} destes aminoácidos são próximos de 280 nm.

Nos experimentos de fluorescência as proteínas são geralmente excitadas com comprimentos de onda maiores que 280 nm. Considerando que o rendimento quântico da fenilalanina é baixo ($\Phi_F = 0,022$), a fluorescência deste aminoácido é quase nula nesses experimentos. Por outro lado, os rendimentos quânticos da tirosina e do triptofano em solução aquosa são muito próximos, sendo estes de $0,13$ e $0,12$, respectivamente. Assim, experimentos que têm por finalidade estudar a fluorescência unicamente do triptofano, devem usar comprimentos de onda de excitação na faixa dos 295 - 305 nm, pois nessa região a absorção da tirosina é nula [1, 4, 17].

O espectro de emissão de fluorescência da fenilalanina, apresenta um comprimento de onda máximo de emissão (λ_{em}) em ~ 280 nm. O λ_{em} da tirosina em água é localizado em 303 nm e é pouco sensível à polaridade do meio. O triptofano livre em água apresenta λ_{em} em 354 nm, e é altamente sensível à polaridade e ao ambiente local, motivo pelo qual é a sonda natural mais amplamente utilizada nos experimentos de fluorescência de proteínas. As características espectrais dos aminoácidos naturais fenilalanina (F), tirosina (Y) e o triptofano (W), tais como rendimento quântico, comprimento de onda de absorção e emissão, estão resumidas na Tabela 1.1.

Tabela 1.1: Rendimento quântico (Φ_F), e comprimentos de onda máximo de absorção (λ_{abs}) e emissão (λ_{em}) dos aminoácidos fenilalanina (Phe), tirosina (Tyr) e triptofano (Trp) em solução aquosa.

Aminoácido	Φ_F	λ_{abs} (nm)	λ_{em} (nm)
Phe	0,022 [18]	258 [19]	280 [18]
Tyr	0,13 [18]	274 [19]	303 [18]
Trp	0,14 [18]	278 [19]	354 [18]

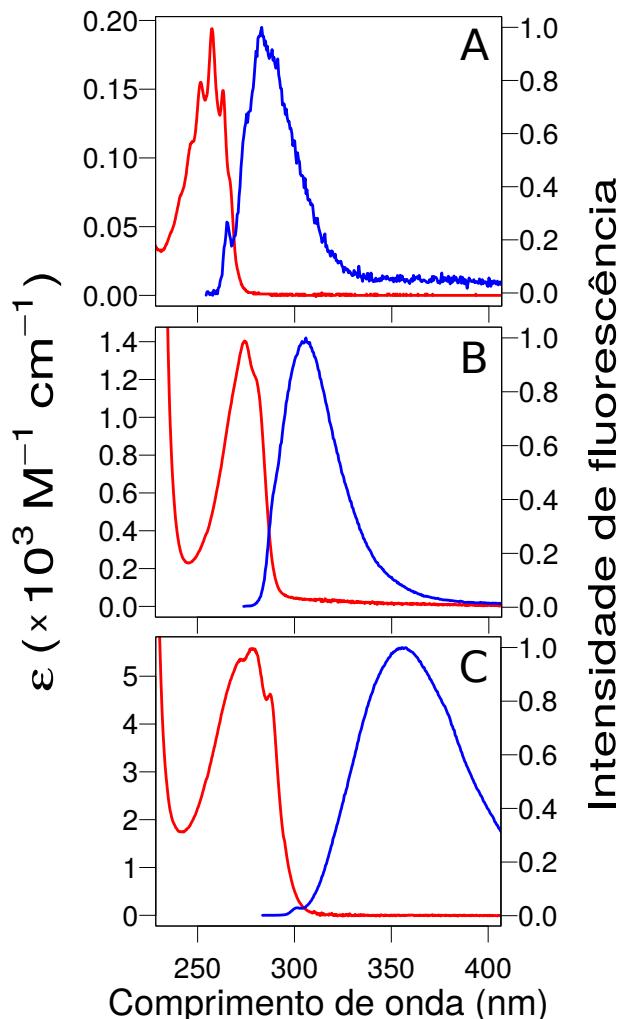


Figura 1.3: Espectros de absorção (vermelho) e emissão (azul) dos aminoácidos A) fenilalanina, B) tirosina e C) triptofano em solução aquosa. Os dados experimentais dos espectros de absorção e emissão correspondem a Fasman et al. (1976) [19] e Chen et al. (1972) [18], respectivamente. Os espectros estão disponíveis no site da PhotochemCAD [20].

Variabilidade da fluorescência do triptofano

As medidas da fluorescência do triptofano em proteínas são hoje em dia comuns nos laboratórios, sendo uma técnica basicamente de rotina. Apesar da sua simplicidade, a

fluorescência nos fornece informações muito importantes sobre a estrutura e estabilidade das proteínas, assim como também sobre o ambiente local do triptofano.

Devido a fluorescência do triptofano ser altamente sensível à polaridade do meio é que seu comprimento de onda de emissão (λ_{em}) é tão variável nas proteínas. Assim temos proteínas que apresentam baixos (e.g. Azurina), intermediários (e.g. Nuclease) e altos (e.g. Glucagon) valores de λ_{em} . Na Tabela 1.2 estão apresentados os λ_{em} de algumas proteínas. Essa diversidade de comprimentos de onda de emissão tem sido motivo de intensa pesquisa na área, com a finalidade de esclarecer suas bases moleculares [15, 21].

Tabela 1.2: Diversidade dos comprimentos de onda de emissão de fluorescência (λ_{em}) do triptofano em proteínas. Todas as proteínas possuem um único resíduo de triptofano. Os números de acesso das estruturas no Protein Data Bank (**PDB id.**) são mostrados.

Proteína	PDB id.	λ_{em}
Azurina	4AZU[22]	308[23]
Parvalbumina	1B8R[24]	316[6]
Mioglobina	1MYT[25]	321[6]
Ribonuclease T1	9RNT[26]	328[27]
FKBP12	1D6O[28]	330[29]
T4-lisozima W126Y W138Y	1LYD[30]	330[31]
T4-lisozima W126Y W158Y	1LYD[30]	330[31]
T4-lisozima W138Y W158Y	1LYD[30]	330[31]
subtilisina Carlsberg	1SBC[32]	331[33]
Nuclease de estafilococo	1STN[34]	334[5]
Che-Y	1CHN[35]	335[6]
α -Cobratoxina	1CTX[36]	340[6]
HSA	1BM0[37]	340[38]
Monelina	1MOL[39]	342[27]
Toxina colérica B (monômero)	1CHP[40]	345[6]
Tioredoxina W31A	2TRX[41]	345[21]
Melitina	2MLT[42]	346[6]
Fosfolipase A2	2BPP[43]	348[44]
Glucagon	1GCN[45]	352[27]

1.3.2 Anisotropia de fluorescência resolvida no tempo

As medidas de fluorescência resolvida no tempo são importantes devido a que fornecem detalhes sobre os processos fotofísicos, fotoquímicos e fotobiológicos. Com a resolução temporal foi possível determinar os tempos de vida dos fluoróforos no estado excitado e a di-

nâmica reorientacional após excitação. As técnicas usadas para estudar esses processos são conhecidos como *tempo de vida da fluorescência resolvida no tempo* e *anisotropia de fluorescência resolvida no tempo*, respectivamente. Nesta seção descreveremos a segunda técnica que é fundamental para o trabalho aqui apresentado.

Arranjo experimental

As técnicas de fluorescência resolvida no tempo usam, basicamente, dois tipos de arranjos experimentais, sendo conhecidos como fluorimetria por pulso e fluorimetria por modulação de fase. A primeira delas utiliza um curto pulso de luz de excitação e registra a resposta de um pulso de fluorescência da amostra. O segundo usa uma luz modulada em frequências variáveis e registra uma resposta harmônica da amostra. Centraremos nossa atenção apenas na fluorimetria de pulso, que é o arranjo mais popular [1, 4].

A implementação da fluorescência resolvida no tempo com o método de fluorimetria de pulso, em princípio necessitaria fazer uso de um único pulso de luz de excitação, após o qual registraria-se temporalmente o sinal de fluorescência. Isso significa que a partir de um único ciclo de excitação-emissão seria possível obter as medidas. Na prática, isso não é possível devido a que o decaimento do sinal de fluorescência é muito rápido e os detectores possuem um tempo morto (*dead time*) após receber um sinal de fluorescência, o que impossibilita receber o próximo sinal se o tempo morto não for respeitado. Para vencer esses limites, a maioria dos equipamentos operando com fluorimetria por pulso funcionam baixo o método de contagem de fóton individual tempo correlacionado (*time-correlated single-photon counting*, TCSPC). Esse método é baseado no uso de múltiplos ciclos de excitação-emissão com registros precisos de fóttons individuais [1, 4, 46].

O arranjo experimental do fluorímetro para realizar medidas dos decaimentos de anisotropia é mostrado na Figura 1.4. Este consiste de uma fonte de luz laser com capacidade de emitir pulsos de luz a alta frequência. Um pulso elétrico é associado ao pulso de luz gerado com a finalidade de criar uma referência para o tempo de início da excitação. O pulso de luz de excitação passa por um polarizador que seleciona uma determinada orientação da luz (no caso a polarização é vertical). A luz de excitação verticalmente polarizada incide sobre a amostra, e o sinal de fluorescência passa por um divisor de feixe, o qual divide a luz em duas polarizações (I_{\parallel}

e I_{\perp}). Os monocromadores de emissão selecionam um determinado comprimento de onda. A sinal é logo amplificada pelos fotomultiplicadores. A informação da fluorescência, assim como também os tempos de início da excitação e do registro da fluorescência, é interpretado pelo sistema de detecção integrada. Os histogramas dos fótons de emissão registrados durante o tempo do experimento são usados para graficar o comportamento das intensidades de luz de polarização paralela (I_{\parallel}) e perpendicular (I_{\perp}) em função do tempo.

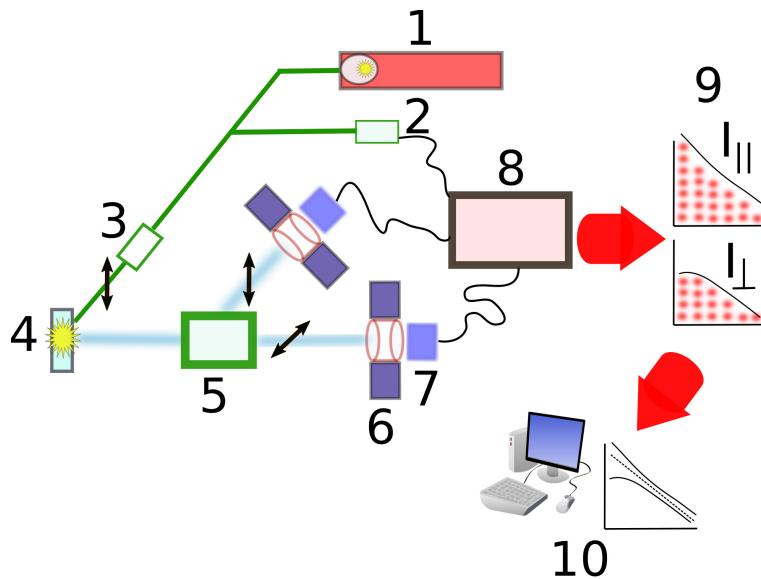


Figura 1.4: Diagrama esquemático do fluorímetro usado nas medidas de anisotropia de fluorescência resolvida no tempo. 1) Fonte de luz laser, 2) discriminador, 3) polarizador, 4) amostra, 5) divisor de feixe, 6) monocromador, 7) tubo fotomultiplicador, 8) instrumento de detecção integrada, 9) sinal de fluorescência, e 10) processamento do sinal de fluorescência

Fundamentos e análise de dados

O estudo da anisotropia de fluorescência resolvida no tempo tem como objetivo monitorar a dinâmica reorientacional do fluoróforo. Para entender esta técnica, dois conceitos fundamentais devem ser citados; i) a natureza do campo elétrico da luz e ii) a orientação dos momentos de dipolo de absorção e emissão do fluoróforo [47]. A luz polarizada é utilizada como fonte de excitação nos experimentos de anisotropia de fluorescência, sendo a orientação do seu vetor elétrico o padrão sob a qual as orientações das emissões de fluorescência do fluoróforo serão medidos.

Experimentalmente a anisotropia de fluorescência resolvida no tempo é definida

como:

$$r(t) = \frac{I_{\parallel}(t) - I_{\perp}(t)}{I_{\parallel}(t) + 2I_{\perp}(t)} \quad (1.3.1)$$

sendo $I_{\parallel}(t)$ e $I_{\perp}(t)$ as intensidades de fluorescência de polarizações paralela e perpendicular à polarização da luz incidente.

A probabilidade do fluoróforo ser excitado depende da orientação relativa do campo elétrico da luz de excitação e do seu dipolo de absorção. Quando a orientação do campo elétrico é paralela ao vetor momento de dipolo de absorção ($\theta = 0^\circ$), o fluoróforo terá probabilidade máxima de ser excitado; enquanto que quando a orientação do campo elétrico é perpendicular ao dipolo de absorção ($\theta = 90^\circ$), o fluoróforo tem probabilidade mínima de excitação. Quantitativamente, a probabilidade de excitação de um fluoróforo pela luz polarizada é proporcional a $\cos^2 \theta_A$, sendo θ_A o ângulo entre o plano de polarização da luz incidente e o momento de dipolo de absorção (Figura 1.5). Essa preferência pela orientação é conhecida como *fotoseleção* [1, 4, 47].

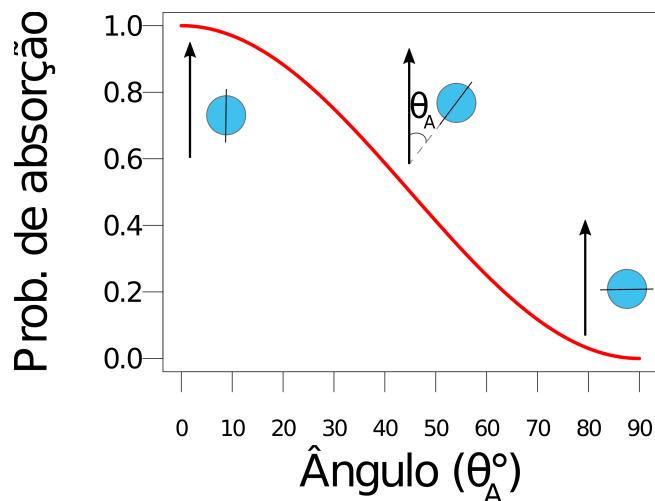


Figura 1.5: Probabilidade de excitação do fluoróforo em função do ângulo entre o vetor de campo elétrico da luz polarizada incidente e o dipolo de absorção do fluoróforo.

Suponhamos que temos uma população de N moléculas orientadas aleatoriamente e estas são excitadas no instante de tempo $t = 0$, por um curto pulso de luz polarizada. Após um intervalo de tempo Δt , os momentos de transição de emissão, μ_E , das moléculas excitadas possuirão uma distribuição angular caracterizada pelo ângulo θ_E , o qual é definido em relação ao

eixo da luz polarizada de excitação (Figura 1.6A). A relação entre a anisotropia de fluorescência e essa distribuição angular é descrita pela Equação 1.3.2 [1, 4].

$$r(t) = \frac{3\overline{\cos^2 \theta_E(t)} - 1}{2} \quad (1.3.2)$$

Algumas informações são possíveis de serem obtidas analisando a equação acima. Por exemplo, quando os momentos de transição de absorção (μ_A) e emissão (μ_E) são paralelos (Figura 1.6B), $\theta_A = \theta_E = \theta$, portanto $\cos^2 \theta_A = \cos^2 \theta_E = \cos^2 \theta$, e considerando que o número de moléculas com momentos de transição orientados entre θ e $\theta + \Delta\theta$ antes da excitação seja isotropicamente distribuído, é possível determinar o valor de $\overline{\cos^2 \theta}$ como sendo de 3/5. Substituindo esse valor, do $\overline{\cos^2 \theta}$, na Equação 1.3.2, chegamos num valor de anisotropia de 0,4.

$$r_0 = \frac{3\overline{\cos^2 \theta} - 1}{2} = \frac{2}{5} = 0,4 \quad (1.3.3)$$

Esse valor é conhecido como a anisotropia fundamental (r_0), e representa o valor teórico da anisotropia na ausência de movimento por parte do fluoróforo. Experimentalmente esse valor é menor, sendo entre 0,32 - 0,39, e é conhecido como anisotropia limite [1, 4]. A razão dessa diferença não é completamente entendida, e a sua causa está associada a: i) artefatos instrumentais (i.e. efeitos do alinhamento do polarizador), ii) efeitos dependentes da matriz (i.e. despolarização por luz espalhada), iii) efeitos intramoleculares (i.e. despolarização associada às geometrias dos estados fundamental e excitado) ou iv) efeitos intermoleculares (i.e. despolarização por fenômenos de transferência de energia) [48].

Por outro lado, geralmente os momentos de transição de absorção (μ_A) e emissão (μ_E) não são paralelos, senão estão orientados com um ângulo α um respeito do outro (Figura 1.6C). Nesse caso, é possível de expressar a anisotropia em função dos ângulos θ_A e α (Equação 1.3.4).

$$r_0 = \frac{3\overline{\cos^2 \theta_E} - 1}{2} = \frac{3\overline{\cos^2 \theta_A} - 1}{2} \times \frac{3\overline{\cos^2 \alpha} - 1}{2} \quad (1.3.4)$$

Como foi mencionado anteriormente, $\overline{\cos^2 \theta_A} = 3/5$, então a anisotropia toma a

seguinte forma.

$$r_0 = \frac{2}{5} \frac{3\cos^2 \alpha - 1}{2} \quad (1.3.5)$$

Assim de acordo com a Equação 1.3.5, a anisotropia de fluorescência assumirá valores entre -0,2 ($\alpha = 90^\circ$, momentos de transição perpendiculares) e 0,4 ($\alpha = 0$, momentos de transição paralelos). Valores negativos de anisotropia foram encontrados em algumas moléculas aromáticas, como, por exemplo, o perileno [1].

Quando o fluoróforo já excitado tem liberdade para rotar, a fluorescência emitida será parcial ou totalmente despolarizada. A orientação do fluoróforo foto-selecionado no tempo $t = 0$ será afetada pela sua dinâmica reorientacional. Desse modo, analisando a despolarização da fluorescência no tempo, é possível de obter informações sobre o movimento molecular e sobre a fluidez do meio [1, 4, 47]. Assim, teremos fluoróforos com liberdade completa e outros com movimentação restrita. Quando o meio é isotrópico e o fluoróforo pode ser reorientar no espaço, a anisotropia de fluorescência decaíra de um valor inicial r_0 até um valor final de 0, indicando orientações dos fluoróforos completamente aleatórias.

Por outro lado, quando a dinâmica reorientacional do fluoróforo for restrita ou quando o meio não é isotrópico (por exemplo, como em um sólido ou uma membrana, respectivamente), a aleatoriedade das orientações do fluoróforo não será atingida, observando-se uma ausência de decaimento total da anisotropia de fluorescência.

A dinâmica reorientacional do momento de dipolo de transição será caracterizado pelo ângulo $\omega(t)$, que é o ângulo formado pelo momento de transição de emissão no tempo t ($\boldsymbol{\mu}_e(t)$) com respeito ao tempo $t = 0$ ($\boldsymbol{\mu}_e(0)$) (Figura 1.6D).

Considerando a Equação 1.3.2 e assumindo que os processos que governam os graus de liberdade de rotação são independentes, podemos escrever a despolarização total como um produto da despolarização decorrente de cada movimento. Assim podemos expressar a aniso-

tropia em função dos ângulos $\theta_A(0)$, α e $\omega(t)$ (Equação 1.3.6).

$$r(t) = \frac{\overline{3\cos^2 \theta_E(t)} - 1}{2} = \frac{\overline{3\cos^2 \theta_A(0)} - 1}{2} \times \frac{\overline{3\cos^2 \alpha} - 1}{2} \times \frac{\overline{3\cos^2 \omega(t)} - 1}{2} \quad (1.3.6)$$

O termo $\frac{\overline{3\cos^2 \theta_A(0)} - 1}{2} \times \frac{\overline{3\cos^2 \alpha} - 1}{2}$ é constante e igual, segundo a Equação 1.3.4, a r_0 .

Portanto, substituindo o valor de r_0 da Equação 1.3.4 na Equação 1.3.6 obtemos a seguinte expressão para a anisotropia:

$$r(t) = r_0 \frac{\overline{3\cos^2 \omega(t)} - 1}{2} \quad (1.3.7)$$

O termo $\frac{\overline{3\cos^2 \omega(t)} - 1}{2}$ é conhecido como a função de autocorrelação das orientações do momento de transição da emissão. Considerando que a quantidade $\frac{3x^2 - 1}{2}$ é o polinômio de Legendre de segunda ordem ($P_2(x)$), podemos expressar alternativamente a anisotropia como [49, 50, 51]:

$$r(t) = r_0 \langle P_2[\cos \omega(t)] \rangle \quad (1.3.8)$$

sendo $\langle \rangle$ a média sobre todas as moléculas excitadas. Para fluoróforos caracterizados com decaimentos do tipo monoexponencial, podemos expressar o polinômio de Legendre da Equação 1.3.8 em função do coeficiente de difusão rotacional do fluoróforo (D_r), usando $\langle P_2[\cos \omega(t)] \rangle = \exp(6D_r t)$ [1].

Em função do momento de dipolo de absorção (μ_a) e emissão (μ_e), a Equação 1.3.8 pode ser reescrita como:

$$r(t) = \frac{2}{5} \langle P_2[\mu_a(0)\mu_e(t)] \rangle \quad (1.3.9)$$

Para uma molécula esférica com rotação livre, o decaimento da anisotropia pode ser

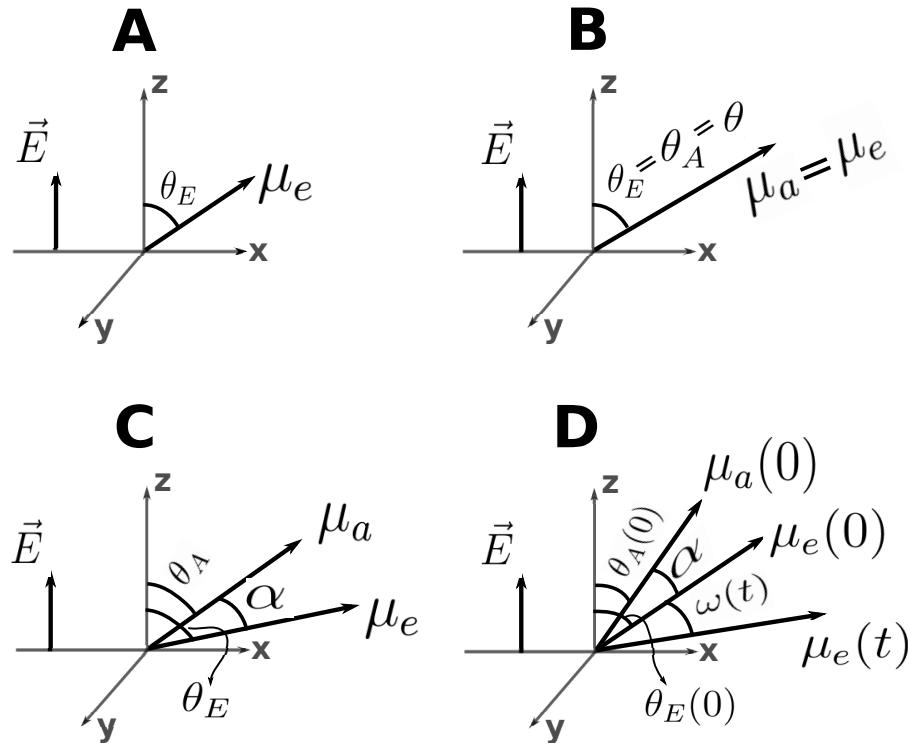


Figura 1.6: Diagramas das orientações dos momentos de dipolo de absorção (μ_a) e emissão (μ_e).

expressado como um decaimento monoexponencial:

$$r(t) = r_0 e^{-t/\tau_{rot}} \quad (1.3.10)$$

sendo r_0 o valor da anisotropia medida no instante da fotoseleção (anisotropia fundamental) e τ_{rot} o tempo característico da rotação, sendo este último o parâmetro indicador da velocidade reorientacional do fluoróforo.

Moléculas complexas, como por exemplo proteínas, apresentam um decaimento de anisotropia complexo. Esse decaimento é, normalmente, modelado como uma soma de vários decaimentos exponenciais:

$$r(t) = r_0 \sum_{i=1} \beta_i e^{-t/\tau_{rot,i}} \quad (1.3.11)$$

sendo β_i as contribuições populacionais do decaimento de anisotropia com tempos de correlação $\tau_{rot,i}$.

Capítulo 2

Simulações de dinâmica molecular

O estudo dos processos (bio)químicos envolvem uma série de fenômenos moleculares que devem de ser tratados por meio de métodos baseados na mecânica quântica de núcleos e elétrons. Diversos métodos, tais como o Hartree-Fock e a teoria do funcional da densidade já foram desenvolvidos e sua implementação encontra-se disponível em programas computacionais. Devido ao alto custo computacional destes métodos, sua aplicação está geralmente restrita ao estudo de sistemas relativamente pequenos, envolvendo poucos átomos ($< 10^2$ átomos). No entanto, as moléculas de interesse para as áreas da biologia molecular e bioquímica, tais como as proteínas e o DNA, são complexas e possuem muitos átomos ($> 10^3$ átomos). É de ressaltar que a maioria dessas moléculas realizam suas funções em soluções ou na presença de outras moléculas, de tal forma que uma representação realista de seu ambiente molecular envolva um número de partículas muito maior. Assim, a aplicação dos métodos baseados nos primeiros princípios resulta inviável na maior parte dos casos. Para estudar as (bio)moléculas é necessário simplificar a representação do sistema. Essa simplificação se dá fazendo com o que os átomos sejam entidades pontuais, com massa e carga adequadas, e suas interações são parametrizadas através de campos de força clássicos. O estudo da evolução temporal destes sistemas clássicos são as simulações de dinâmica molecular. Nesta tese essas simulações serão amplamente usadas, por isso há a necessidade de dedicar este capítulo para explicar seus fundamentos. A maioria dos conceitos apresentados aqui foram baseados nas referências de Allen e Tildesley (1987) [52], e Allen (2004) [53].

2.1 Simulações de dinâmica molecular

Simulações de dinâmica molecular consistem em resolver as equações de movimento de um sistema, e tem por finalidade gerar trajetórias. No caso puramente clássico, para uma partícula i de massa m_i , que está baixo a influência de uma força \vec{f}_i , a dinâmica molecular tem

como objetivo resolver a equação de Newton,

$$\frac{d^2\vec{r}_i}{dt^2} = \frac{\vec{f}_i(\vec{r}_1, \vec{r}_2, \dots)}{m_i} \quad (2.1.1)$$

sendo a força dependente das posições (\vec{r}_i).

A implementação das simulações de dinâmica molecular seguem uma série de passos. O quadro seguinte resume esses passos, os quais serão detalhados ao longo deste capítulo.

Passos da simulação de dinâmica molecular

1. No início da simulação, no tempo $t = 0$, são definidas as posições e velocidades dos átomos. O passo do tempo (Δt) também é definido no início da simulação.
2. Calcular as forças sobre todos os átomos do sistema.
3. Integrar as equações de Newton. Logo, atualizar as posições e velocidades dos átomos para o instante de tempo $t + \Delta t$.
4. Repetir os passos 2 ao 4 até alcançar o tempo total desejado .

2.1.1 Condições iniciais

Para iniciar uma simulação de dinâmica molecular é necessário definir o conjunto de posições iniciais ($\vec{r}_1(0), \vec{r}_2(0), \dots$) e velocidades iniciais ($\vec{v}_1(0), \vec{v}_2(0), \dots$) dos átomos do sistema.

As posições iniciais podem ser geradas a partir da disposição dos átomos numa rede regular de pontos ou com programas especializados, como o Packmol [54, 55]. Para o caso das proteínas, as posições iniciais dos átomos são obtidos por meio de experimentos, como a cristalografia de raios X ou ressonância magnética nuclear [56, 57]. Essas estruturas encontram-se depositadas no *Protein Data Bank* [58]. Quando as estruturas experimentais não estão disponíveis, ainda é possível estimar a disposição dos átomos por meio de ferramentas teóricas como são a determinação de estruturas por meio de homologia ou por previsões *ab-initio* [59].

Com respeito às velocidades iniciais, estas podem ser escolhidas de maneira aleatória,

segundo uma distribuição de Maxwell-Boltzmann (Equação 2.1.2).

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right] \quad (2.1.2)$$

sendo $p(v_{ix})$ a probabilidade de um determinado átomo i de massa m_i numa temperatura T possua uma velocidade v_{ix} .

2.1.2 Potenciais de interação

O passo seguinte, após definidas as condições iniciais, é calcular as forças sobre cada átomo do sistema. A força pode ser obtida mediante a derivação do potencial de interação (Equação 2.1.3).

$$\vec{f}_i = -\frac{\partial}{\partial r_i} U(\vec{r}_1, \vec{r}_2, \dots) \quad (2.1.3)$$

Para isso é necessário especificar o potencial de interação (U). Esse potencial costuma ser representado como uma soma de contribuições não-ligantes (U^{NLIG}) e ligantes (U^{LIG}) (Equação 2.1.4).

$$U = U^{NLIG} + U^{LIG} \quad (2.1.4)$$

A continuação detalharemos a forma como são comumente apresentados esses potenciais.

Potenciais não-ligantes

1. Potencial de Lennard-Jones

Este potencial é caracterizado pela presença de um termo repulsivo ($\sim 1/r^{12}$) que atua a curtas distâncias e outro atrativo que possui maior alcance ($\sim 1/r^6$). A forma

deste potencial é:

$$V_{LJ} = \sum_{ij} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.1.5)$$

sendo ϵ_{ij} , σ_{ij} e r_{ij} a profundidade do potencial, a distância na qual o potencial é zero e a distância entre os átomos i e j , respectivamente. O fator $\sim 1/r^6$ pretende representar as interações de van der Waals, enquanto que o fator $\sim 1/r^{12}$ impede a sobreposição dos átomos, e tem uma forma funcional conveniente por ser simplesmente o quadrado do termo atrativo [60].

Usualmente, é necessário truncar os potenciais de Lennard-Jones nas simulações de dinâmica molecular, e a forma como o programa NAMD o implementa é mediante a introdução de uma função que suavemente realize o truncamento a uma distância **cutoff**. O raio de truncamento, o raio no qual a função de truncamento começa a ser aplicado e a função de truncamento são definidos pelos parâmetros **cutoff**, **switchdist** e **switching**, respectivamente [61].

2. Potencial de Coulomb

Para um sistema constituído de N átomos localizados nas posições $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$ e que possuem cargas q_1, q_2, \dots, q_N , o potencial de Coulomb (ou eletrostático) desse sistema é descrito por:

$$V_{elec} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.1.6)$$

sendo q_i e q_j as cargas parciais dos átomos i e j , r_{ij} a distância entre os átomos i e j , e ϵ a constante de permissividade no vácuo.

O cômputo deste potencial pode gerar problemas porque decai muito lentamente com a distância ($\sim 1/r$). Devido a isso não é possível truncar o potencial eletrostático da mesma forma como é feito para o de Lennard-Jones. Métodos de baseados nas somas de Ewald foram desenvolvidos para computar os potenciais de Coulomb (ou outros potenciais de longo alcance). Basicamente, consiste em dividir o potencial eletrostático em duas

contribuições: potencial de curto alcance (E^S) e potencial de longo alcance (E^L). As contribuições de curto alcance são calculados no espaço real, enquanto as de longo alcance são calculadas usando transformadas de Fourier [62]. Assim, o potencial total de Coulomb pode ser expressado como uma soma das suas contribuições de curto e longo alcance (Equação 2.1.7).

$$V_{elec} = E^S + E^L \quad (2.1.7)$$

Nas nossas simulações de dinâmica molecular o potencial eletrostático é calculado usando o algoritmo *Particle Mesh Ewald Sum* (PME), o qual está implementado no programa NAMD [61].

Finalmente, a expressão do potencial não-ligante é a soma dos potenciais de Lennard-Jones e Coulomb para todos os átomos do sistema (Equação 2.1.8).

$$U^{NLIG} = \underbrace{\sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{V_{LJ}} + \underbrace{\sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}}_{V_{elec}} \quad (2.1.8)$$

Potenciais ligantes

Nas moléculas, as interações entre átomos próximos de acordo com a conectividade covalente, são descritas por três contribuições principais: Potencial de estiramento, potencial angular e potencial diedral. A continuação detalharemos esses potenciais.

1. Potencial de estiramento

O potencial que descreve a ligação covalente entre dois átomos i e j é conhecido como potencial de estiramento (Figura 2.1A). Este potencial é do tipo harmônico e possui a seguinte forma:

$$V_{est} = \frac{1}{2} \sum_{ligações} k_{ij}^b (r_{ij} - r_0)^2 \quad (2.1.9)$$

sendo k_{ij}^b , r_{ij} , e r_0 a constante de força da ligação, a distância entre os átomos i e j , e a distância de equilíbrio da ligação respectivamente. Os parâmetros necessários para construir este potencial são k_{ij}^b e r_0 . Este potencial descreve bem o comportamento das ligações na vizinhança da posição de equilíbrio.

2. Potencial angular

As configurações que adotam três átomos no espaço podem ser descritas pelo ângulo que estes formam (Figura 2.1B). Assim, para incluir as penalidades angulares, é necessário de introduzir o potencial angular. Este potencial pode ser descrito também por uma função harmônica:

$$V_{ang} = \frac{1}{2} \sum_{ângulos} k_{ijk}^\theta (\theta_{ijk} - \theta_0)^2 \quad (2.1.10)$$

sendo k_{ijk}^θ , θ_{ijk} e θ_0 a constante da força da deformação angular, o ângulo formado pelos átomos i , j k , e o ângulo de equilíbrio dos átomos i , j k , respectivamente. Os parâmetros k_{ijk}^θ e θ_{ijk} são definidos para cada grupo de átomos ijk .

3. Potencial diedral

Três pontos no espaço definem um plano. Analogamente três átomos também definem um plano. Se considerarmos quatro átomos contíguos e ligados covalentemente em pares como mostrado na Figura 2.1C, é possível de criar dois planos formados pelos átomos ijk e klj . O potencial dihedral está relacionado à flexibilidade (torsão) do ângulo formado pelos planos dos átomos ijk e klj .

$$V_{dih} = \frac{1}{2} \sum_{dihedros} k_\varphi [1 + \cos(n_{ijkl}\varphi_{ijkl} - \delta)] \quad (2.1.11)$$

sendo k_φ , a constante da força; n_{ijkl} , a multiplicidade (a qual indica o número de mínimos quando a ligação rota 360°). A multiplicidade é um número inteiro positivo. φ_{ijkl} , é o ângulo entre os planos formados pelos átomos ijk e klj ; e δ é a fase do ângulo.

Assim, a expressão do potencial ligante é composta de três termos principais, os quais

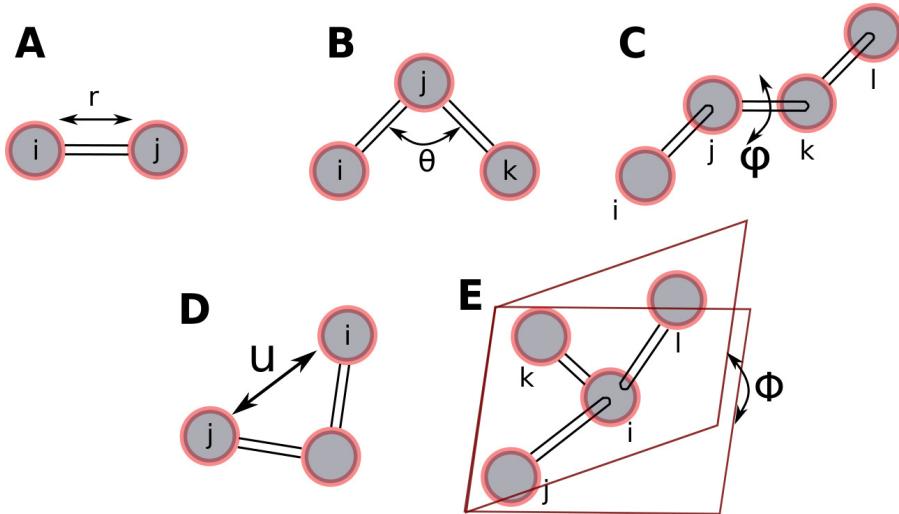


Figura 2.1: Parâmetros usados na construção dos potenciais ligantes. A) Estiramento das ligações, B) deformação angular, C) deformação diedral, D) potencial de Urey-Bradley e E) deformação de diedros impróprios

envolvem o estiramento das ligações, a deformação angular e a rotação diedral (Equação 2.1.12).

$$U^{LIG} = \underbrace{\frac{1}{2} \sum_{\text{ligações}} k_b(r - r_0)^2}_{V_{est}} + \underbrace{\frac{1}{2} \sum_{\text{ângulos}} k_\theta(\theta - \theta_0)^2}_{V_{ang}} + \underbrace{\frac{1}{2} \sum_{\text{dihedros}} k_\varphi[1 + \cos(n\varphi - \delta)]}_{V_{dih}} \quad (2.1.12)$$

Existem ainda outros potenciais ligantes, além dos citados acima, que são: o potencial de Urey-Bradley e de diedros impróprios. O potencial de Urey-Bradley tem por finalidade descrever a ligação não-covalente entre dois átomos separados por duas ligações covalentes [63]. Este potencial é harmônico e tem como finalidade restringir o movimento dos átomos (Equação 2.1.13).

$$V_{Urey-Bradley} = k_{UB}(u - u_0)^2 \quad (2.1.13)$$

sendo k_{UB} a constante da força, u a distância entre os átomos i e j e u_0 a distância de equilíbrio (Figura 2.1D)

O potencial de diedros impróprios tem por finalidade manter a planaridade das

moléculas, e também é descrito por uma função harmônica [64] (Equação 2.1.14).

$$V_{imprópios} = k_{imp}(\phi - \phi_0)^2 \quad (2.1.14)$$

sendo k_{imp} a constante da força, ϕ o ângulo entre os planos formados pelos átomos ijk e klj , e ϕ_0 o ângulo de equilíbrio (Figura 2.1D).

Esses potenciais ligantes e não-ligantes, descritos acima, estão definidos no campo de força CHARMM [65, 66, 67], o qual é usado neste trabalho.

2.1.3 Integrando as equações de movimento

Até o momento temos as condições iniciais dos átomos (definidas pelas suas posições e velocidades) e a forma dos potenciais. Agora é momento de movimentar o sistema, e para isso será necessário calcular a equação diferencial de segunda ordem apresentada ao início deste capítulo (Equação 2.1.1). Para resolver essa equação, normalmente são usadas aproximações e diferenças finitas, sendo as implementações mais populares nas simulações de dinâmica molecular os algoritmos de Verlet e Verlet-velocidade [53].

O algoritmo de Verlet pode ser derivado usando expansões de Taylor da função posição. Assim, expandindo essa função no tempo $t + \Delta t$ em função do instante de tempo t e mostrando apenas até a quarta ordem da expansão temos:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i(t)}{dt} \Delta t + \frac{1}{2!} \frac{d^2\vec{r}_i(t)}{dt^2} \Delta t^2 + \frac{1}{3!} \frac{d^3\vec{r}_i(t)}{dt^3} \Delta t^3 + O(\Delta t^4) \quad (2.1.15)$$

Analogamente, podemos expandir a função posição no instante de tempo anterior, $t - \Delta t$, em função da posição atual, t .

$$\vec{r}_i(t - \Delta t) = \vec{r}_i(t) - \frac{d\vec{r}_i(t)}{dt} \Delta t + \frac{1}{2!} \frac{d^2\vec{r}_i(t)}{dt^2} \Delta t^2 - \frac{1}{3!} \frac{d^3\vec{r}_i(t)}{dt^3} \Delta t^3 + O(\Delta t^4) \quad (2.1.16)$$

Observando as Equações 2.1.15 e 2.1.16 é possível notar uma reciprocidade nos termos que contém derivadas de ordem ímpar. Assim, se somarmos essas equações é possível eliminar essas derivadas e obter uma expressão simplificada da função posição. O algoritmo de Verlet nasce da soma das Equações 2.1.15 e 2.1.16 (Equação 2.1.17).

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{d^2\vec{r}_i(t)}{dt^2}\Delta t^2 + O(\Delta t^4) \quad (2.1.17)$$

Segundo a Equação 2.1.17 é possível de calcular as posições com uma precisão da ordem de Δt^3 , desprezando os termos de ordem maior que Δt^4 .

Sabendo que $\frac{d^2\vec{r}_i(t)}{dt^2} = -\frac{\vec{f}(r_i(t))}{m_i}$ e desprezando os termos de quarta (ou maiores) ordem, temos o algoritmo de Verlet (Equação 2.1.18).

O algoritmo Verlet

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{f}(r_i(t))}{m_i}\Delta t^2 \quad (2.1.18)$$

Se analizarmos a Equação 2.1.18 no instante de tempo $t = 0$, podemos observar que o algoritmo de Verlet apresenta um problema, pois o instante de tempo anterior ao $t = 0$ não é definido. Para evitar esse problema durante a inicialização das simulações de dinâmica molecular ($t = 0$) é necessário usar uma aproximação dadas as condições iniciais $\vec{r}_i(0)$ e $\vec{v}_i(0)$ [60]. Para isso vamos a substituir $t = 0$ na Equação 2.1.15 e truncar a expressão até a derivada de segunda ordem. Assim teremos que a posição no tempo $t = \Delta t$ é:

$$\vec{r}_i(\Delta t) \approx 2\vec{r}_i(0) + \frac{d\vec{r}_i(0)}{dt}\Delta t + \frac{1}{2!} \frac{d^2\vec{r}_i(0)}{dt^2}\Delta t^2 \quad (2.1.19)$$

Com as posições $\vec{r}_i(\Delta t)$ é possível computar as forças $\vec{f}(r_i(\Delta t))$. Após isso o cálculo pode ser feito iterativamente usando a Equação 2.1.18.

Implementação do algoritmo de Verlet

1. Definidos as condições iniciais, $\vec{r}_i(0)$, $\vec{v}_i(0)$ e $\vec{a}_i(0)$, calculamos o $\vec{r}_i(\Delta t)$:

$$\vec{r}_i(\Delta t) \approx 2\vec{r}_i(0) + \frac{d\vec{r}_i(0)}{dt}\Delta t + \frac{1}{2} \frac{d^2\vec{r}_i(0)}{dt^2}\Delta t^2$$

2. Calcular as novas forças, e assim $\vec{a}_i(\Delta t)$.

$$\vec{a}_i(\Delta t) = -\frac{1}{m_i} \nabla V_i(\vec{r}_i(\Delta t))$$

3. Calcular as novas posições iterativamente usando o algoritmo de Verlet.

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{f}(r_i(t))}{m_i} \Delta t^2$$

Alternativamente, outra metodologia de integração consegue computar simultaneamente as posições e velocidades. Esta metodologia é uma sofisticação da anterior e é conhecida como algoritmo Verlet-velocidade [53].

Como vimos anteriormente, a função $\vec{r}_i(t + \Delta t)$ pode ser expandida em função de $\vec{r}_i(t)$ (Equação 2.1.15). A expansão considerando as derivadas de segunda ordem é:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i(t)}{dt}\Delta t + \frac{1}{2!} \frac{d^2\vec{r}_i(t)}{dt^2}\Delta t^2 \quad (2.1.20)$$

sendo $v(t) = d\vec{r}_i(t)/dt$ e $a(t) = d^2\vec{r}_i(t)/dt^2$, podemos reescrever a Equação 2.1.20 como:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + v(t)\Delta t + \frac{1}{2!}a(t)\Delta t^2 \quad (2.1.21)$$

Por outro lado, a função velocidade $v(t + \Delta t)$ também pode ser aproximada por meio da expansão de Taylor de segunda ordem (Equação 2.1.22).

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{d\vec{v}_i(t)}{dt}\Delta t + \frac{1}{2!} \frac{d^2\vec{v}_i(t)}{dt^2}\Delta t^2 \quad (2.1.22)$$

O termo $d^2\vec{v}_i(t)/dt^2$ pode ser expressado em função de derivadas de ordem inferior.

$$\frac{d\vec{v}_i(t + \Delta t)}{dt} = \frac{d\vec{v}_i(t)}{dt} + \frac{d^2\vec{v}_i(t)}{dt^2}\Delta t + O\Delta t^2 \quad (2.1.23)$$

Isolando o termo $d^2\vec{v}_i(t)/dt^2$ da Equação 2.1.23, e substituindo este na Equação 2.1.22 temos:

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{1}{2} \frac{d\vec{v}_i(t + \Delta t)}{dt} \Delta t + \frac{1}{2} \frac{d\vec{v}_i(t)}{dt} \Delta t \quad (2.1.24)$$

Assim, as posições e velocidades podem ser computadas simultaneamente no algoritmo Verlet-velocidade.

Substituindo os termos $d^2\vec{r}_i/dt^2$ e $d\vec{v}_i/dt$ por $\vec{f}(r_i(t))/m_i$ na Equação 2.1.20, e o termo $d\vec{v}_i(t + \Delta t)/dt$ por $\vec{f}(r_i(t + \Delta t))/m_i$ na Equação 2.1.24 obtemos as equações básicas do algoritmo Verlet-velocidade (Equação 2.1.25 e 2.1.26)

O algoritmo Verlet-velocidade

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{\vec{f}(r_i(t))}{m_i}\Delta t^2 \quad (2.1.25)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{1}{2} \frac{\vec{f}(r_i(t + \Delta t))}{m_i}\Delta t + \frac{1}{2} \frac{\vec{f}(r_i(t))}{m_i}\Delta t \quad (2.1.26)$$

No seguinte quadro mostra-se a implementação do algoritmo Verlet-velocidade.

Implementação do algoritmo de Verlet-velocidade

1. Definir as condições iniciais das posições ($\vec{r}_i(0)$), velocidades ($\vec{v}_i(0)$) e acelerações ($\vec{a}_i(0)$) (passo 1 da Figura 2.2).
2. Calcular as novas posições no tempo $t + \Delta t$ (passo 2 da Figura 2.2).

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\vec{a}_i(t)\Delta t^2$$

3. Calcular as velocidades num intervalo de tempo intermediário (passo 3 da Figura 2.2).

$$\vec{v}_i(t + \Delta t/2) = \vec{r}_i(t) + \frac{1}{2}\vec{a}_i(t)\Delta t$$

4. Com as posições obtidas no tempo $t + \Delta t$ é possível de calcular as forças e, portanto, as acelerações (passo 4 da Figura 2.2).

$$\vec{a}_i(t + \Delta t) = -\frac{1}{m}\nabla V_i(\vec{r}_1(t + \Delta t) \dots \vec{r}_N(t + \Delta t))$$

5. Calcular as velocidades no tempo $t + \Delta t$ (passo 5 da Figura 2.2).

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t + \Delta t/2) + \frac{1}{2}\vec{a}_i(t + \Delta t)\Delta t$$

6. Repetir iterativamente os passos 2 ao 5.

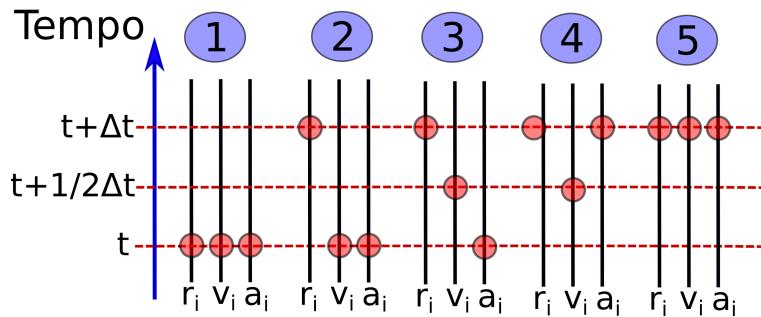


Figura 2.2: Representação esquemática do funcionamento do algoritmo Verlet-velocidade.

2.2 Simulações ABF

Por meio das simulações de dinâmica molecular convencionais é possível de explorar diversas conformações das proteínas. No entanto, muitas vezes, as conformações exploradas se restringem a uma determinada parte da superfície de energia potencial, devido a presença de barreiras energéticas, sendo assim a amostragem ineficiente.

Imagine que temos uma proteína, para a qual desejamos investigar a superfície de energia potencial. Essa superfície está composta de dois mínimos locais de energia (Figura 2.3). A região do mínimo I representa a estrutura cristalográfica da proteína, enquanto o mínimo II é uma conformação alternativa. Normalmente as simulações convencionais tomam a estrutura cristalográfica como ponto de partida, assim regiões próximas ao mínimo I poderão ser bem amostradas. Inclusive a rugosidade do mínimo I será revelada devido a que as flutuações térmicas do sistema o permitirão. Essas flutuações serão ineficientes para vencer a barreira energética e amostrar o mínimo II. Por outro lado, se as simulações começarem com uma conformação da proteína representativa do mínimo II, pode acontecer da proteína transitar para o mínimo I rapidamente, limitando a amostragem das vizinhanças de II. Isso seria uma série dificuldade se estivéssemos interessados em amostrar únicamente o mínimo II.

Uma forma de conseguir amostrar as regiões “difícies” da superfície de energia potencial é mediante a introdução de forças que guiem a amostragem de novas configurações [68, 69]. Segundo o formalismo das simulações ABF (*Adaptive Biasing Force*), as forças introduzidas correspondem à força média calculada num intervalo da coordenada de reação. Essa força é adicionada ao potencial, fazendo com que a superfície de energia potencial seja aplanada e assim novas regiões sejam amostradas com apenas as flutuações térmicas [70].

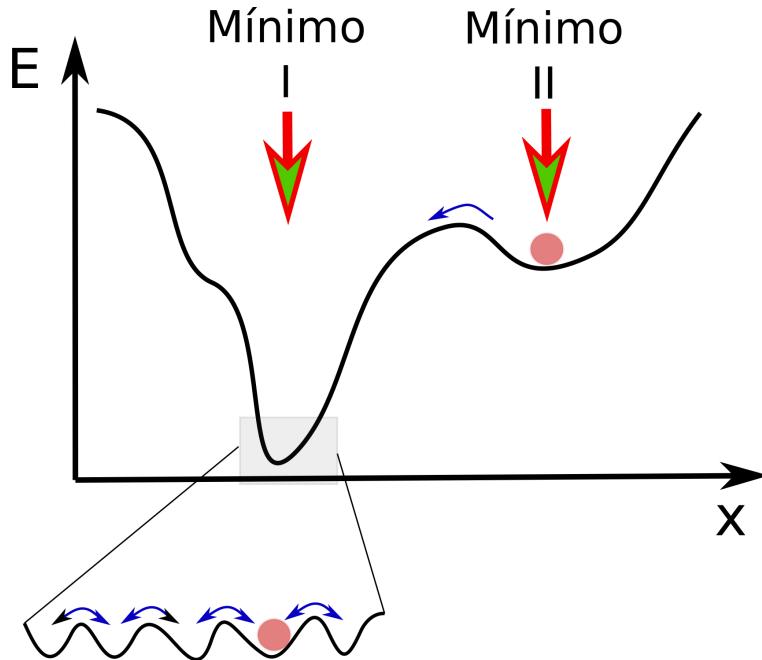


Figura 2.3: Representação esquemática da superfície de energia potencial de uma proteína. Dois mínimos locais são mostrados. Uma ampliação do mínimo I revela alguns detalhes da superfície de energia potencial.

Para implementar as simulações ABF é necessário definir uma coordenada de reação (ξ), a qual é discretizada em M porções ($\delta\xi$). A força média na direção da coordenada de reação na porção k depois de N_{passos} passos de simulação de dinâmica molecular será estimada simplesmente como a média simples das forças amostradas (Equação 2.2.1).

$$\bar{F}_\xi(N_{passos}, k) = \frac{1}{N_{passos}^k} \sum_{\mu=1}^{N_{passos}^k} F_\mu^k \quad (2.2.1)$$

sendo N_{passos}^k o número de amostras acumuladas na porção k depois de N_{passos} passos e F_μ^k a força instantânea registrada na porção k . O valor de N_{passos}^k não pode ser muito baixo devido a que pode causar falhas na estimativa da força média [70]. Considerando que o valor de N_{passos}^k é o suficiente, a força $\bar{F}_\xi(N_{passos}, k)$ poderá ser computada satisfatoriamente para cada porção k .

Essa força média é a que impede amostrar corretamente a superfície de energia potencial. Portanto, se removemos essa força do sistema, as barreiras energéticas que impedem a amostragem correta da superfície de energia potencial serão vencidas. As simulações ABF tem como objetivo introduzir forças externas opostas à força que atua sobre a coordenada ξ ,

possibilitando que as flutuações térmicas explorem as regiões que antes eram proibidas [71].

A partir dessas forças é possível de computar a diferença de energia livre entre os estados iniciais e finais (Equação 2.2.2).

$$\Delta A_\xi = - \sum_{i=1}^M \bar{F}_\xi(N_{passos}, k) \delta \xi \quad (2.2.2)$$

O método ABF está implementado no programa NAMD, sendo os parâmetros mais importantes os seguintes [72]:

- **fullSamples**: Determina o número amostras que são registradas na porção k (N_{passos}^k) antes da aplicação da força $\bar{F}_\xi(N_{passos}, k)$.
- **width**: Determina o tamanho de cada porção da coordenada de reação ($\delta \xi$).
- **lowerboundary**: Define o limite inferior da coordenada de reação.
- **upperboundary**: Define o limite superior da coordenada de reação.
- **lowerwallconstant**: É o valor da constante de força (Kcal/mol/Å²) de um potencial semi-harmônico aplicado para garantir o limite inferior da coordenada de reação.
- **upperwallconstant**: É o valor da constante de força (Kcal/mol/Å²) de um potencial semi-harmônico para garantir o limite superior da coordenada de reação.

No presente trabalho usa-se o método ABF, não para calcular as energias livres senão, com a finalidade de amostrar preferencialmente uma conformação alternativa da enzima subtilisina Carlsberg. Tecnicamente, nossas simulações ABF nunca aplicam força ao sistema devido a que definimos o parâmetro **fullSamples** como sendo muito maior ao número total de amostras da simulação.

Capítulo 3

Objetivos

Os objetivos desta tese de doutorado são:

- **Desenvolver métodos clássicos paramétricos com capacidade de predizer os comprimentos de onda de emissão do triptofano em proteínas.** Nesse sentido foram desenvolvidos os modelos baseados na área acessível ao solvente e os baseados nas interações eletrostáticas.
- **Avaliar a robustez dos modelos clássicos paramétricos desenvolvidos.** A análise feita garante que os modelos desenvolvidos nesta tese não são dependentes dos dados de treinamento usados para a construção dos modelos, portanto são robustos.
- **Reproduzir por meio de simulações de dinâmica molecular os experimentos de anisotropia de fluorescência resolvido no tempo do Trp113 da subtilisina Carlsberg.** Os cálculos realizados demonstraram que existe uma conformação preferencial da enzima usada nos experimentos. A amostragem correta das estruturas da enzima garante uma consistência entre as dinâmicas de reorientação obtidas pelo experimento e pelas simulações.

Capítulo 4

Modelos paramétricos baseados na área acessível ao solvente para computar os comprimentos de onda de emissão do triptofano

Neste capítulo serão desenvolvidos modelos clássicos parametrizados baseados na área acessível ao solvente do indol com capacidade preditiva dos comprimentos de onda de emissão do triptofano em proteínas. Os modelos serão construídos considerando um grupo de 19 proteínas. Todas estas proteínas contém apenas um único resíduo de triptofano e seus comprimentos de onda de emissão de fluorescência foram já determinados experimentalmente. Observaremos que estes modelos possuem uma boa capacidade preditiva para comprimentos de onda maiores de 314 nm.

4.1 Área acessível ao solvente

O grupo indol do triptofano tem natureza aromática e hidrofóbica. Assim, é de se esperar que seja encontrado parcialmente ou totalmente inserido no interior dos bolsos hidrofóbicos das proteínas. Experimentos de enovelamento e desenovelamento de proteínas têm aproveitado amplamente essa propriedade do indol. Usualmente agentes desnaturantes, tais como ureia ou guanidina, são empregados e as mudanças estruturais da proteína são seguidas

por meio das variações na fluorescência do triptofano.

Para exemplificar como a fluorescência do triptofano é usada, descreveremos brevemente o resultado dos experimentos de desenovelamento da proteína aporepressor da *trp* de *E.coli*, a qual contém dois triptofanos na sua estrutura: Trp19 e Trp99 (Figura 4.1) [7]. O resíduo Trp19 localiza-se no interior de um bolso hidrofóbico e está praticamente inacessível ao solvente. Por outro lado, o resíduo Trp99 encontra-se parcialmente acessível ao solvente.

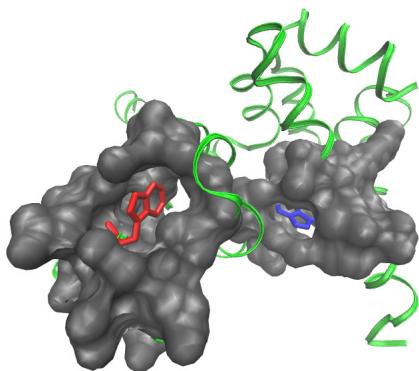


Figura 4.1: Estrutura da proteína apo-repressora do *trp* de *E.coli*. Os triptofanos Trp19 (em azul) e Trp99 (em vermelho) são destacados.

Estudar mediante fluorescência proteínas com mais de um triptofano é um desafio, devido a que a presença de multiples fluoróforos dificulta a interpretação dos espectros. Graças ao avanço das técnicas de biología molecular, agora é possível de realizar mutações em resíduos específicos. Assim o estudo das mudanças conformacionais de cada triptofano (ou região vizinha ao triptofano) durante o processo de desenovelamento são possíveis.

Para o caso da proteína apo-repressora do *trp* de *E.coli* duas mutações foram desenvolvidas [7]. A primeira delas, conhecida como W99F, consiste numa substituição do Trp99 por uma fenilalanina. Esta mutação permitiu investigar exclusivamente a fluorescência do Trp19. A outra mutação chamada de W19F, também uma substituição por fenilalanina, permitiu estudar a dinâmica do Trp99.

Os espectros de fluorescência das formas nativas das proteínas W99F e W19F estão mostrados na Figura 4.2A e B, respectivamente. O valor máximo do espectro é normalmente chamado de comprimento de onda de emissão de fluorescência (λ_{em}). Os λ_{em} dos Trp19 e Trp99

na estrutura nativa da proteína apo-repressora do *trp* de *E.coli* são de ~ 320 nm e ~ 340 nm, respectivamente. Esses resultados indicam que o triptofano com menor exposição ao solvente possui menor λ_{em} , e viceversa.

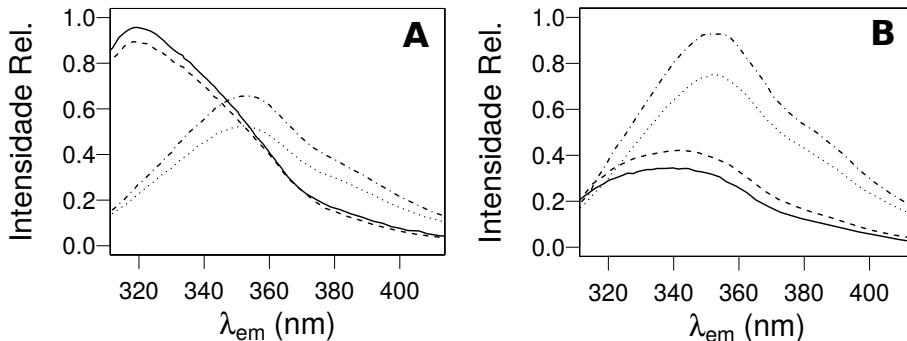


Figura 4.2: Dependência dos espectros de fluorescência dos triptofanos da proteína apo-repressora do *trp* de *E.coli* com respeito às condições desnaturantes do solvente. A) Apo-repressor W99F, e B) Apo-repressor W19F. O desnaturante usado é ureia e as concentrações são as seguintes: 0 M (—), 3 M (---), 5 M (.....), e 9 M (- - -). Adaptado de Royer et al. (1993) [7].

Os espectros de fluorescência das proteínas W99F e W19F na presença de concentrações crescentes de agente desnaturante estão mostrados na Figura 4.2A e B, respectivamente. Duas mudanças são observadas nos espectros devido à presença de agente desnaturante:

A primeira delas é a variação na intensidade dos espectros. A intensidade de fluorescência está relacionada ao rendimento quântico do fluoróforo [1, 13]. As mutantes W99F e W19F ao desnaturar, exibem um comportamento distinto na sua intensidade de fluorescência. Assim, nas condições desnaturantes as mutantes W99F e W19F apresentam diminuição e aumento da intensidade de fluorescência, respectivamente. Apesar das bases moleculares que guiam esse comportamento específico para essas mutantes não são conhecidas, é possível dizer que eventos que estabilizam ou desestabilizam a transferência de elétrons do grupo indol para um acceptor (que podem ser, por exemplo, grupos amina) estejam envolvidos [15]. Observamos que o comportamento da intensidade do triptofano, e portanto do rendimento quântico, é complexa e encontrar algum padrão parece não ser uma tarefa trivial.

A segunda mudança no espectro de fluorescência está relacionada ao comprimento de onda da intensidade máxima, ou seja ao comprimento de onda de emissão (λ_{em}). Os mutantes ao desnaturar, exibem um aumento dos λ_{em} (Figura 4.2A e B). Os mutantes W99F e W19F nas suas formas completamente desnaturadas apresentam λ_{em} de ~ 355 nm e ~ 351 nm,

respectivamente. O fenômeno que determina o λ_{em} é produto de uma estabilização da transferência de densidade eletrônica do anel pirrólico para o benzênico do indol, sendo assim também de natureza quântica [73, 74]. Apesar disso é possível de dizer que a simples exposição do triptofano ao solvente favorece maiores λ_{em} e vice-versa.

A variação do λ_{em} que acabamos de observar nos experimentos de desenovelamento da proteína apo-repressor da *trp* de *E. coli* é também encontrada em outras proteínas. A partir dessa informação, poderíamos pensar que o desenovelamento das proteínas expõe o triptofano, ou em outras palavras o faz mais acessível ao solvente, e que essa maior acessibilidade está correlacionada ao aumento no comprimento de onda. Essa ideia será a base para desenvolver os modelos parametrizados baseados na área acessível ao solvente, os quais serão apresentados neste Capítulo.

4.2 Metodologia

4.2.1 Simulações de dinâmica molecular

As estruturas cristalográficas das proteínas presentes na Tabela 4.1 foram obtidas da base de dados do PDB [58]. As águas cristalográficas das proteínas foram mantidas. As configurações iniciais, contendo a proteína, água, e íons de sódio e cloro foram construídas com ajuda do software Packmol [54, 55]. As proteínas foram solvatadas com uma camada de 15 Å de água e íons de sódio e cloro adicionados aos sistemas para deixá-los neutros, sendo a concentração de 0.16 M. Essa necessidade de deixar os sistemas neutros radica em que é fundamental para as somas de Ewald convergirem. O modelo de água usado foi o TIP3P [75]. A proteína, água e íons foram simulados usando o campo de força CHARMM22 [65, 66, 67]. Condições periódicas de contorno foram usadas nas simulações. Os sistemas foram equilibrados usando os seguintes passos:

- 1000 passos de minimização usando o método de gradientes conjugados, seguidos de uma simulação dinâmica molecular de 200 ps com todos os átomos da proteína fixos;
- 500 passos de minimização usando o método de gradientes conjugados, seguidos de uma simulação dinâmica molecular de 200 ps considerando os átomos C α da proteína fixos; e

- 2 ns de uma simulação dinâmica molecular sem nenhuma restrição.

As estruturas finais da última simulação foram consideradas configurações de início das simulações de dinâmica molecular de produção, a partir das quais as análises foram feitas. Todas as simulações foram executadas nas seguintes condições: ensemble NPT a 1 atm de pressão e 298.15 K de temperatura, com passos de tempo de 2 fs. O controle de temperatura foi mantido pela dinâmica de Langevin usando os seguintes parâmetros: temperatura de Langevin de 298.15 K e coeficiente de fricção de 10/ps. As pressões das simulações foram controladas com o método de Nosé-Hoover, sendo os parâmetros: tempo de oscilação do barostato de 200 fs, taxa de decaimento de 100 fs e temperatura de Langevin de 298.15 K. O algoritmo de Verlet-velocidade foi empleado com passos de tempo de 2 fs. Restrições implementadas pelo algoritmo SHAKE foram aplicadas a todos os átomos contendo hidrogênios. As forças não-ligantes e ligantes foram calculadas cada 1 e 2 passos de tempo, respectivamente. As interações eletrostáticas foram calculadas com o método de *Particle Mesh Ewald Sum* (PME), sendo o raio de truncamento de 12 Å. As configurações foram salvas a cada 1 ps. O tempo de produção de todas as simulações foi de 20 ns. O software usado para todas as simulações de dinâmica molecular foi o NAMD [76].

Os RMSDs (*Root mean square deviation*) foram calculados para a etapa de produção das simulações, usando a estrutura inicial minimizada como a estrutura de referência. O programa utilizado para realizar esses cálculos foi o VMD[77]. A média dos RMSDs para todas as proteínas, com exceção da Mioglobina e Glucagon, foram entre \sim 1-2 Å. Essas duas proteínas em particular, apresentaram os maiores valores de médios do RMSD, sendo de \sim 4 Å(Figura A.1 e A.2). Os cálculos do RMSD indicam que não houve uma mudança estrutural significativa para nossas proteínas durante as simulações.

4.2.2 Cálculo da área acessível ao solvente

A área acessível ao solvente do grupo indol do triptofano nas proteínas foi calculado usando o método de Shrake-Rupley [78], o qual está implementado no programa VMD [77]. Nessa metodologia, cada átomo da proteína é considerado como uma esfera cujo raio é definido pelo seu raio de van der Waals [79]. O solvente, água, é considerada como uma esfera de raio

igual a 1.4 Å[79]. A lista de átomos vizinhos j ao átomo i do indol é obtido se $d(C_iC_j) < r_i + r_j + 2r$, sendo $d(C_iC_j)$ a distância entre os centros dos átomos i e j , e r_i , r_j e r os raios dos átomo i , j e do solvente. Em cada átomo do indol é gerado e distribuído de maneira uniforme N pontos (valor padrão é $N = 500$). Cada ponto é avaliado com a finalidade de conhecer se dito ponto está dentro de outra esfera vizinha. Se o ponto está dentro de outra esfera, então o ponto representará um lugar não acessível ao solvente. Caso contrário, será considerado acessível ao solvente. Assim a área acessível ao solvente do átomo i do indol pode ser calculada como a relação entre o número de pontos P não localizados dentro de uma outra esfera vizinha e o número N de pontos totais,

$$SASA_i = 4\pi r_i^2 \times \left(\frac{P}{N} \right) \quad (4.2.1)$$

sendo r_i o raio de van der Waals do átomo i do indol.

A área acessível ao solvente do indol será calculada como a soma das áreas acessíveis de cada átomo i ,

$$SASA_{indol} = \sum_{i=1} SASA_i \quad (4.2.2)$$

4.3 Resultados e discussões

4.3.1 Relação entre o comprimento de onda de emissão e a área acessível ao solvente do triptofano

Na Seção 4.1 observou-se que é razoável pensar que existam correlações quantitativas entre o comprimento de onda de emissão de fluorescência (λ_{em}) e a área acessível ao solvente (SASA) do triptofano em proteínas. Para ilustrar essa ideia, a área acessível ao solvente do indol de três proteínas, as quais possuem um só resíduo de triptofano, foram calculadas. As proteínas escolhidas foram a Parvalbumina [6], Nuclease de estafilococo [5], e Glucagon [27]. Essas proteínas são representativas da diversidade de λ_{em} , devido a que possuem valores baixo (Parvalbumina, $\lambda_{em} = 316$ nm [6]), intermediário (Nuclease de estafilococo, $\lambda_{em} = 334$ nm [5]),

e alto (Glucagon, $\lambda_{em} = 352$ nm [27]) de λ_{em} do triptofano. Assim, é de esperar-se que quanto maior seja a área acessível ao solvente do indol, maior será o λ_{em} .

Na Figura 4.3 é mostrada a evolução temporal da área acessível ao solvente do grupo indol do triptofano para as três proteínas. Observa-se que a SASA do indol da Parvalbumina é a menor de todas, sendo de $\sim 6 \text{ \AA}^2$. A SASA do indol da Nuclease de estafilococo é de $\sim 50 \text{ \AA}^2$, enquanto que a SASA do indol do Glucagon é maior das três, sendo aproximadamente de 150 \AA^2 . Observamos que existe claramente uma correspondência entre o λ_{em} e a SASA do indol dessas três proteínas.

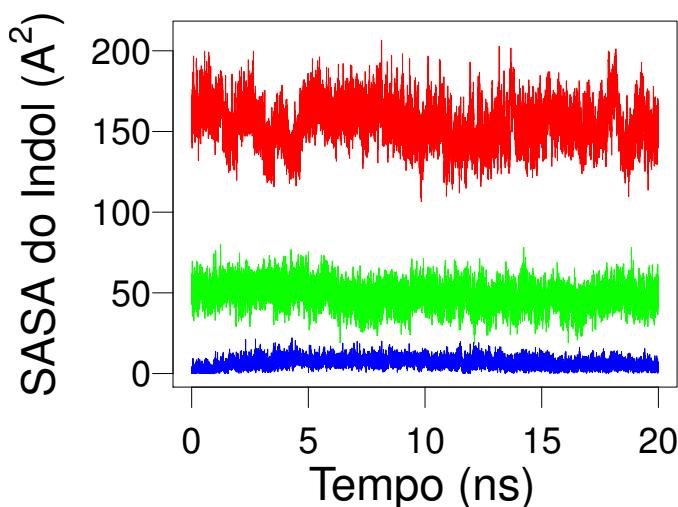


Figura 4.3: Evolução temporal da SASA do indol na Parvalbumina (azul), Nuclease de estafilococo (verde) e Glucagon (vermelho). O aumento da área acessível ao solvente está associado ao incremento no comprimento de onda de emissão de fluorescência.

Com a finalidade de estender nossa ideia, cálculos similares foram realizados para um grupo de proteínas que cumprem as seguintes condições:

- A estrutura cristalográfica está disponível.
- Possuem um só resíduo de triptofano.
- Os λ_{em} estão disponíveis na literatura.

Baseado nas condições anteriores, um grupo de 19 proteínas foi selecionado (Tabela 4.1). Os λ_{em} experimentais mínimo e máximo desse grupo de proteínas selecionadas foram de 308 e 352 nm, respectivamente, sendo assim representativos dos λ_{em} encontrados na natureza.

As SASAs do indol das 19 proteínas foram calculadas para cada estrutura amostrada nas simulações de dinâmica molecular, e os valores médios foram computados (Tabela 4.1). A Azurina, a proteína com menor λ_{em} , apresentou o menor valor para a SASA do indol, de $0,45 \text{ \AA}^2$. Por outro lado, a proteína com maior λ_{em} , o Glucagon, apresentou a maior SASA, de 154 \AA^2 .

A correlação entre o λ_{em} experimental e a SASA do indol para as 19 proteínas é mostrada na Figura 4.4, sendo o coeficiente de correlação de $R^2 = 0,64$. As proteínas que parecem escapar dessa relação linear são as que possuem baixos valores de λ_{em} (Azurina, Parvalbumina e Mioglobina). O coeficiente de correlação aumenta para $R^2 = 0,71$, se desconsiderarmos essas proteínas. Portanto, a exposição ao solvente do indol está correlacionada ao comprimento de onda de emissão nas proteínas, e modelos baseados nessas correlações podem ser desenvolvidos.

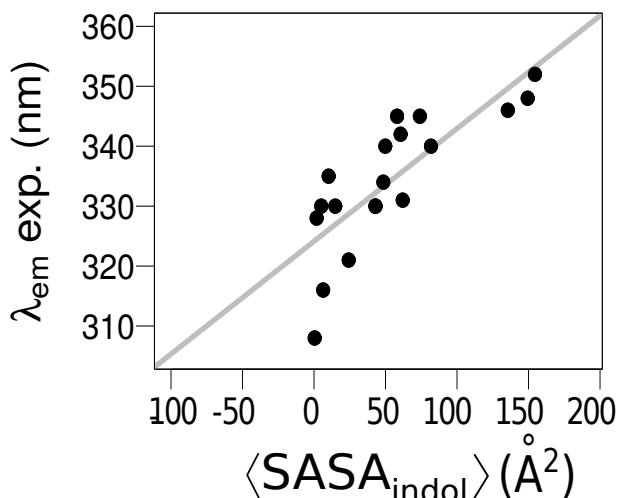


Figura 4.4: Correlação entre o comprimento de onda de emissão de fluorescência do triptofano experimental (λ_{em}) e o valor médio da SASA do indol em 19 proteínas. As proteínas correspondem aos da Tabela 4.1. A linha cinza mostra a tendência dos dados.

4.3.2 Construção dos modelos

Na subseção anterior descrevemos a existência de uma correlação linear entre a SASA do indol e o λ_{em} do Trp em proteínas. Assim é de se esperar que modelos baseados na SASA do indol possam ter capacidade preditiva dos λ_{em} do triptofano. Nesta subseção usaremos essa correlação para construir os modelos baseados na SASA do indol.

O primeiro modelo (Modelo 1) que construiremos considera a existência de uma

relação linear entre o λ_{em} experimental e o valor médio da SASA do indol (Equação 4.3.1).

$$\lambda_{em} = A_0 + A_1 \times \langle SASA_{ind} \rangle \quad (4.3.1)$$

Este primeiro modelo possui apenas dois parâmetros. O primeiro parâmetro, A_0 , é o valor mínimo do λ_{em} que o modelo consegue predizer. O segundo parâmetro, A_1 , é o fator que relaciona a SASA do indol ao λ_{em} . Usando os valores médios das SASAs do indol calculados para as 19 proteínas, da Tabela 4.1, o modelo foi construído. Os valores dos parâmetros deste modelo estão mostrados na Tabela 4.3. O valor de A_0 é de 324 nm, o que significa que proteínas que possuem λ_{em} menores a este valor serão sobreestimados. Das proteínas que usamos para construir o modelo, só três possuem λ_{em} experimentais menores que 324 nm, sendo estas a Azurina ($\lambda_{em} = 308$ nm), Parvalbumina ($\lambda_{em} = 316$ nm), e a Mioglobina ($\lambda_{em} = 321$ nm). O λ_{em} predito com este modelo é sobre-estimado em 16, 9 e 7 nm para a Azurina, Parvalbumina e Mioglobina, respectivamente.

A Figura 4.5A mostra a correlação e o ajuste linear entre os λ_{em} preditos, usando o Modelo 1, versus os λ_{em} experimentais. O Modelo 1 tem uma capacidade preditiva dos λ_{em} com um desvio padrão de 6,80 nm e um coeficiente de correlação de $R^2 = 0,64$. O coeficiente de correlação aumenta para $R^2 = 0,71$ se desconsiderarmos as proteínas Azurina, Parvalbumina e a Mioglobina.

O grupo indol consiste num anel benzênico fundido a um anel pirrólico (Figura 4.6). Os átomos que pertencem unicamente ao anel benzênico são os carbonos de numeração 4, 5, 6 e 7; além dos hidrogênios ligados a estes. Os átomos que fazem parte unicamente do anel pirrólico são o nitrogênio e os carbonos de numeração 2 e 3, e os respectivos hidrogênios. Os átomos de carbono de numeração 8 e 9 do indol estão compartilhados pelo anel benzênico e pirrólico.

Assim, para dividir as contribuições da SASA do indol nas suas componentes benzênicas e pirrólicas é necessário definir os átomos que farão parte de cada componente. Assim definiremos as seguintes nomenclaturas:

- Benzeno incompleto: Formado pelos átomos de carbono de numeração 4, 5, 6 e 7 e seus hidrogênios.

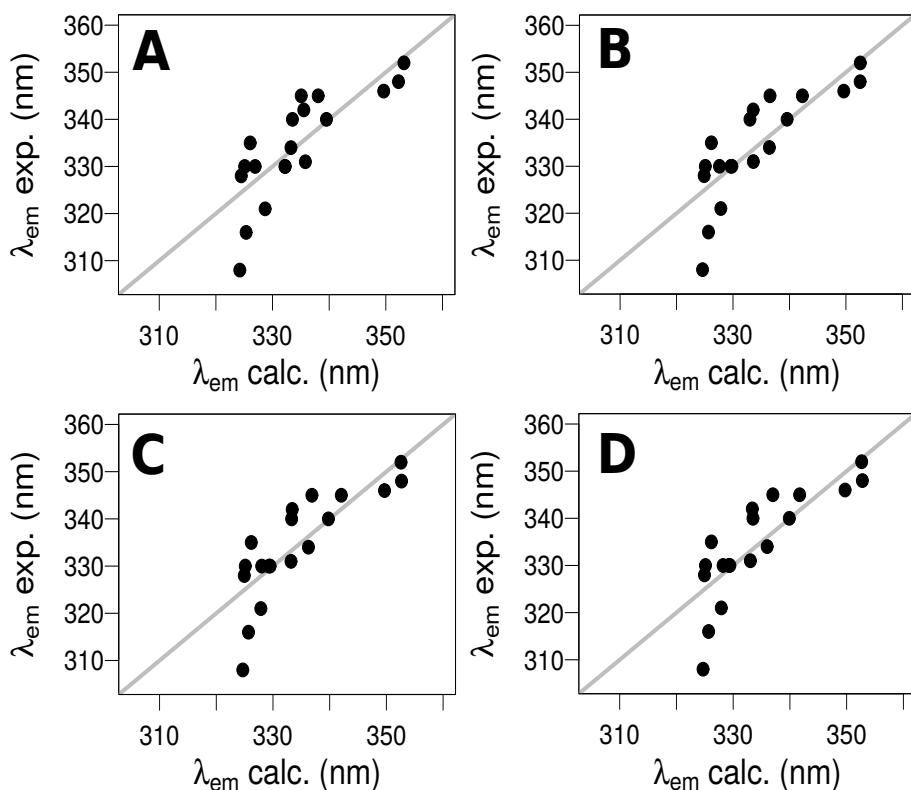


Figura 4.5: Comprimento de onda de emissão de fluorescência experimental (λ_{em}^{exp}) versus o calculado ($\lambda_{em}^{calc.}$) usando os modelos baseados na área acessível ao solvente. A) Modelo 1: Baseado na $\langle SASA \rangle$ do grupo indol do triptofano. B) Modelo 2: Baseado na decomposição da $\langle SASA \rangle$ dos grupos benzeno incompleto e pirrol completo do indol. C) Modelo 3: Baseado na decomposição da $\langle SASA \rangle$ dos grupos benzeno completo e pirrol incompleto do indol. D) Modelo 4: Baseado na decomposição da $\langle SASA \rangle$ dos grupos benzeno completo e pirrol completo do indol.

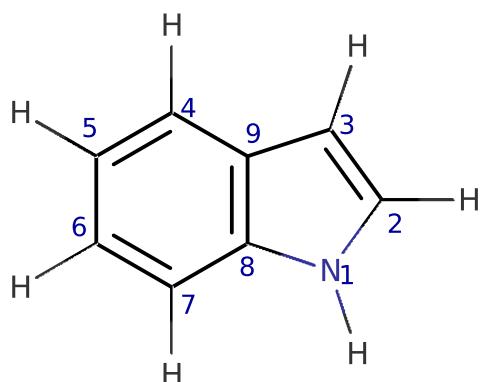


Figura 4.6: Representação esquemática do indol. Os átomos de carbono e nitrôgenio estão numerados para identificar os átomos que farão parte do anel benzênico e pirrólico durante os cálculos da SASA ou interações eletrostáticas.

- Benzeno completo: Formado pelos átomos de carbono de numeração 4, 5, 6, 7, 8 e 9 e seus hidrogênios.
- Pirrol incompleto: Formado pelos átomos de numeração 1, 2 e 3 e seus hidrogênios.
- Pirrol completo: Formado pelos átomos de numeração 1, 2, 3, 8 e 9 e seus hidrogênios.

Com base nessas definições é possível de dividir a SASA do indol em seus componentes benzênico e pirrólico. Assim, baseados nas contribuições da SASA do indol serão construídos os seguintes modelos alternativos.

- Baseado na SASA do benzeno incompleto e do pirrol completo.
- Baseado na SASA do benzeno completo e do pirrol incompleto.
- Baseado na SASA do benzeno completo e do pirrol completo.

O Modelo 2, o qual é baseado na SASA do benzeno incompleto e do pirrol completo, possui três parâmetros (Equação 4.3.2).

$$\lambda_{em} = B_0 + B_1 \times \langle SASA_{ben,i} \rangle + B_2 \times \langle SASA_{pyr,c} \rangle \quad (4.3.2)$$

sendo B_0 o comprimento de onda mínimo que o modelo consegue predizer e B_1 e B_2 são os fatores que associam a SASA do benzeno incompleto ($SASA_{ben,i}$) e pirrol completo ($SASA_{pyr,c}$) ao λ_{em} , respectivamente. Os valores da SASA do benzeno incompleto e do pirrol completo para o grupo de 19 proteínas foram calculados e os seus valores médios estão reportados na Tabela 4.1. Os parâmetros do Modelo 2 foram otimizados com esse grupo de dados e seus valores estão na Tabela 4.3. O parâmetro B_0 é de 324 nm, o qual indica que proteínas com baixos valores de SASA ainda terão seus λ_{em} sobreestimados. Este modelo possui uma capacidade preditiva com um desvio padrão de 6,58 nm com respeito ao λ_{em} experimental e um coeficiente de correlação $R^2 = 0,66$. Os valores dos λ_{em} preditos com o Modelo 2 estão mostrados na Tabela 4.2. O maior e o menor desvios com respeito aos dados experimentais correspondem às proteínas Azurina (~ 17 nm) e T4-lisozima W126Y W138Y ($\sim -0,02$ nm), respectivamente. A correlação e o ajuste linear dos λ_{em} preditos com o Modelo 2 versus os λ_{em} experimentais estão na Figura 4.5B.

O terceiro modelo que iremos a construir, o Modelo 3, será baseado nas contribuições

da SASA do benzeno completo e pirrol incompleto (Equação 4.3.3).

$$\lambda_{em} = C_0 + C_1 \times \langle SASA_{ben,c} \rangle + C_2 \times \langle SASA_{pyr,i} \rangle \quad (4.3.3)$$

sendo C_0 o valor mínimo de λ_{em} que o modelo consegue predizer, e C_1 e C_2 são os fatores que associam o SASA do benzeno completo ($SASA_{ben,c}$) e pirrol incompleto ($SASA_{pyr,i}$) ao λ_{em} , respectivamente. Os valores dos $SASA_{ben,c}$ e $SASA_{pyr,i}$ foram calculados para cada estrutura amostrada nas simulações das 19 proteínas, e seus valores médios foram computados e estão na Tabela 4.1. Os parâmetros foram otimizados com esse grupo de dados e estão mostrados na Tabela 4.3. Este modelo, igual que os anteriores, tende a sobreestimar os λ_{em} das proteínas com triptofanos totalmente enterrados no bolsão hidrofóbico. A capacidade preditiva deste modelo tem uma performance com um desvio padrão de 6,56 nm com respeito aos dados experimentais. O coeficiente de correlação do modelo é de $R^2 = 0,66$. A predição do λ_{em} da proteína Azurina é a apresenta o maior desvio com respeito aos dados experimentais (~ 17 nm), enquanto a proteína da α -Cobratoxina representa a melhor predição do Modelo 3 ($\sim -0,2$ nm). Os valores dos λ_{em} preditos com o Modelo 3 e o gráfico das predições versus os dados experimentais estão na Tabela 4.2 e na Figura 4.5C, respectivamente.

O último modelo baseado na SASA que construiremos, Modelo 4, considerará os grupos benzeno e pirrol do indol como entidades separadas, ou seja, ambos serão completos (Equação 4.3.4).

$$\lambda_{em} = D_0 + D_1 \times \langle SASA_{ben,c} \rangle + D_2 \times \langle SASA_{pyr,c} \rangle \quad (4.3.4)$$

sendo D_0 o valor mínimo de λ_{em} que o modelo consegue predizer, e D_1 e D_2 são os fatores que associam a SASA do benzeno completo ($SASA_{ben,c}$) e pirrol completo ($SASA_{pyr,c}$) ao λ_{em} , respectivamente. Os valores médios da $SASA_{ben,c}$ e $SASA_{pyr,c}$ estão mostrados na Tabela 4.1. Este modelo contém três parâmetros (D_0 , D_1 e D_2) e seus valores foram otimizados com os dados da SASA das 19 proteínas da Tabela 4.1. A capacidade preditiva deste modelo possui um desvio padrão de 6,55 nm e um coeficiente de correlação de $R^2 = 0,67$. O gráfico das predições dos λ_{em} com o Modelo 4 versus os λ_{em} experimentais para as 19 proteínas usadas na construção

do modelo estão mostradas na Figura 4.5D. Nessa figura é possível ver a tendência comum dos modelos baseados na SASA, a de sobreestimar os λ_{em} de proteínas com triptofanos com baixa acessibilidade ao solvente. Os valores das previsões dos λ_{em} do Modelo 4 estão na Tabela 4.2.

Tabela 4.1: Proteínas usadas, números de acesso das estruturas (**PDB id.**), comprimentos de ondas de emissão experimentais (λ_{em}^{exp}), e áreas acessíveis ao solvente calculadas a partir das simulações de dinâmica molecular. $SASA_{ind}$, $SASA_{ben,i}$, $SASA_{pyr,i}$, $SASA_{ben,c}$, $SASA_{pyr,c}$ são as áreas acessíveis ao solvente do indol, benzeno incompleto, pirrol incompleto, benzeno completo e pirrol completo, respectivamente. $\langle \rangle$ representa a média sobre a simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as SASAs em Å².

Proteína	PDB id.	λ_{em}^{exp}	$\langle SASA_{ind} \rangle$	$\langle SASA_{ben,i} \rangle$	$\langle SASA_{pyr,c} \rangle$	$\langle SASA_{ben,c} \rangle$	$\langle SASA_{pyr,i} \rangle$
Azurina	4AZU[22]	308[23]	0,45±0,78	0,21±0,48	0,24±0,50	0,26±0,52	0,19±0,45
Parvalbumina	1B8R[24]	316[6]	6,39±3,03	3,22±2,44	3,17±2,59	3,28±2,46	3,11±2,56
Mioglobina	1MYT[25]	321[6]	24,23±13,89	5,63±5,48	18,60±12,46	6,68±5,99	17,55±11,56
Ribonuclease T1	9RNT[26]	328[27]	1,79±1,20	1,19±1,07	0,60±0,58	1,19±1,07	0,60±0,58
FKBP12	1D6O[28]	330[29]	14,87±5,56	10,74±5,07	4,13±2,66	13,73±5,29	1,14±1,63
T4-lisozima W126Y W138Y	1LYD[30]	330[31]	43,04±9,74	5,68±4,46	37,36±6,05	5,68±4,48	37,36±6,05
T4-lisozima W126Y W158Y	1LYD[30]	330[31]	5,01±2,42	0,32±0,64	4,69±2,28	0,32±0,64	4,69±2,28
T4-lisozima W138Y W158Y	1LYD[30]	330[31]	42,88±6,52	4,60±3,19	38,28±4,75	4,60±3,19	38,28±4,75
subtilisina Carlsberg	1SBC[32]	331[33]	61,98±8,86	18,59±5,88	43,39±4,45	18,70±5,98	43,28±4,36
Nuclease de estafilococo	1STN[34]	334[5]	48,46±7,51	48,34±7,45	0,12±0,27	48,36±7,46	0,10±0,23
Che-Y	1CHN[35]	335[6]	10,20±4,52	4,05±3,22	6,15±4,17	4,15±3,22	6,05±4,13
α -Cobratoxina	1CTX[36]	340[6]	81,75±7,85	46,13±6,18	35,62±5,25	50,96±6,64	30,79±4,89
HSA	1BM0[37]	340[38]	49,83±7,93	23,21±6,64	26,62±5,76	27,25±7,16	22,58±6,04
Monelina	1MOL[39]	342[27]	60,51±15,26	19,44±10,72	41,07±8,67	21,06±12,65	39,45±7,79
Toxina colérica B (monômero)	1CHP[40]	345[6]	58,07±14,91	42,09±13,50	15,98±4,09	46,35±14,07	11,72±3,60
Tioredoxina W31A	2TRX[41]	345[21]	73,99±14,08	70,78±12,40	3,21±2,84	71,48±12,92	2,51±2,02
Melitina	2MLT[42]	346[6]	135,49±14,03	77,51±13,66	57,98±11,70	83,89±14,61	51,60±11,59
Fosfolipase A2	2BPP[43]	348[44]	149,38±30,20	87,75±31,09	61,63±9,49	95,43±33,11	53,95±9,84
Glucagon	1GCN[45]	352[27]	154,47±13,28	84,27±11,82	70,20±7,59	91,53±13,39	62,94±7,42

Tabela 4.2: Comprimentos de onda de emissão de fluorescência experimental e calculado do triptofano: λ_{em}^{exp} é o comprimento de onda de emissão de fluorescência experimental. $\lambda_{em}^{(1-4)}$ e $\Delta\lambda_{em}^{(1-4)}$ são o comprimento de onda de emissão de fluorescência calculado e o desvio relativo ao valor experimental, respectivamente, usando os modelos baseados na área acessível ao solvente 1 ao 4. As unidades dos comprimentos de onda e os desvios são os nm.

Proteína	λ_{em}^{exp}	$\lambda_{em}^{(1)}$	$\Delta\lambda_{em}^{(1)}$	$\lambda_{em}^{(2)}$	$\Delta\lambda_{em}^{(2)}$	$\lambda_{em}^{(3)}$	$\Delta\lambda_{em}^{(3)}$	$\lambda_{em}^{(4)}$	$\Delta\lambda_{em}^{(4)}$
Azurina	308[23]	324,22	16,22	324,61	16,61	324,68	16,68	324,69	16,69
Parvalbumina	316[6]	325,34	9,34	325,65	9,65	325,68	9,68	325,67	9,67
Mioglobina	321[6]	328,69	7,69	327,84	6,84	327,86	6,86	327,91	6,91
Ribonuclease T1	328[27]	324,47	-3,53	324,89	-3,11	324,94	-3,06	324,94	-3,06
FKBP12	330[29]	326,93	-3,07	327,60	-2,40	328,01	-1,99	328,22	-1,78
T4-lisozima W126Y W138Y	330[31]	332,23	2,23	329,80	-0,20	329,49	-0,51	329,41	-0,59
T4-lisozima W126Y W158Y	330[31]	325,08	-4,92	325,10	-4,90	325,12	-4,88	325,12	-4,88
T4-lisozima W138Y W158Y	330[31]	332,20	2,20	329,63	-0,37	329,32	-0,68	329,24	-0,76
subtilisina Calsberg	331[33]	335,80	4,80	333,60	2,60	333,18	2,18	333,03	2,03
Nuclease de estafilococo	334[5]	333,25	-0,75	336,44	2,44	336,25	2,25	335,99	1,99
Che-Y	335[6]	326,05	-8,95	326,16	-8,84	326,17	-8,83	326,15	-8,85
α -Cobratoxina	340[6]	339,52	-0,48	339,58	-0,42	339,78	-0,22	339,90	-0,10
HSA	340[38]	333,51	-6,49	333,00	-7,00	333,29	-6,71	333,49	-6,51
Monelina	342[27]	335,52	-6,48	333,57	-8,43	333,39	-8,61	333,37	-8,63
Toxina colérica B (monômero)	345[6]	335,06	-9,94	336,55	-8,45	336,87	-8,13	336,99	-8,01
Tioredoxina W31A	345[21]	338,06	-6,94	342,29	-2,71	342,05	-2,95	341,72	-3,28
Melitina	346[6]	349,63	3,63	349,62	3,62	349,67	3,67	349,72	3,72
Fosfolipase A2	348[44]	352,24	4,24	352,52	4,52	352,67	4,67	352,77	4,77
Glucagon	352[27]	353,20	1,20	352,55	0,55	352,58	0,58	352,65	0,65

Tabela 4.3: Modelos baseados na área acessível para a predição de comprimentos de onda de emissão de fluorescência por simulações de dinâmica molecular. R^2 é o coeficiente de correlação, SD é o desvio padrão das previsões com respeito aos dados experimentais. $SASA_{ind}$, $SASA_{ben,c}$, $SASA_{ben,i}$, $SASA_{pyr,c}$, $SASA_{pyr,i}$ são as áreas acessíveis ao solvente do indol, benzeno completo, benzeno incompleto, pirrol completo e pirrol incompleto, respectivamente. As unidades das SASAs são Å².

Modelo	Descrição	Equação	R^2	SD
1	SASA do indol	$\lambda_{em} = 324 + 0,188 \langle SASA_{ind} \rangle$	0,64	6,80
2	SASA do benzeno(i)/pirrol(c)	$\lambda_{em} = 324 + 0,246 \langle SASA_{ben,i} \rangle + 0,103 \langle SASA_{pyr,c} \rangle$	0,66	6,58
3	SASA do benzeno(c)/pirrol(i)	$\lambda_{em} = 324 + 0,241 \langle SASA_{ben,c} \rangle + 0,094 \langle SASA_{pyr,i} \rangle$	0,66	6,56
4	SASA do benzeno(c)/pirrol(c)	$\lambda_{em} = 324 + 0,235 \langle SASA_{ben,c} \rangle + 0,093 \langle SASA_{pyr,c} \rangle$	0,67	6,55

A inclusão de um parâmetro adicional não melhorou significativamente a capacidade preditiva dos modelos baseados em SASA. Apesar disso algumas informações são possíveis de ser extraídas. A decomposição da SASA nos permite determinar as contribuições das partes benzênicas e pirrólicas para a determinação dos λ_{em} . A relação entre os parâmetros B_1/B_2 é 2,39; o qual contrasta com a relação entre $\overline{\langle SASA_{ben,i} \rangle} / \overline{\langle SASA_{pyr,c} \rangle} = 1,18$. Resultados similares são encontrados na relação entre os parâmetros C_1/C_2 e D_1/D_2 , sendo de 2,56 e 2,53; respectivamente. Portanto, podemos dizer que o benzeno é o grupo determinante do λ_{em} em proteínas. A importância do benzeno também foi reportada por Vivian e Callis (2004) [74]. Segundo eles, o benzeno é estabilizado através de uma aproximação dos prótons da água aos átomos de carbono do benzeno, enquanto os átomos do pirrol não interagem fortemente com a água. Assim, eles concluem que o benzeno é o maior responsável pelo desvio para o vermelho dos λ_{em} .

As proteínas podem ser divididas de acordo com Burstein et al. em 5 classes discretas [80, 81]:

- Classe A: Caracterizada por um $\lambda_{em} = 308$ nm, e o Trp completamente protegido no bolsão hidrofóbico da proteína.
- Classe S: Caracterizada por um $\lambda_{em} = 316$ nm, e o Trp quase completamente protegido no bolsão hidrofóbico da proteína.
- Classe I: Caracterizada por um λ_{em} entre 330 e 332 nm, e o Trp parcialmente exposto ao

solvente.

- Classe II: Caracterizada por um λ_{em} entre 340, e 342 nm e o Trp exposto ao solvente.
- Classe III: Caracterizada por um λ_{em} entre 350, e 353 nm e o Trp completamente exposto ao solvente.

De acordo aos valores dos λ_{em} experimentais das 19 proteínas, reportados na Tabela 4.1, é possível classificá-las seguindo a nomenclatura dos estados discretos propostos por Burstein. Assim temos:

- Classe A: Azurina.
- Classe S: Parvalbumina, Mioglobina e Ribonuclease T1.
- Classe I: FKBP12, T4 lisozima W126Y W138Y, T4 Lisozima W126Y W158Y, T4 lisozima W138Y W158Y, S. Carlsberg, Nucleasa de estafilococo e Che-Y.
- Classe II: α -Cobratoxina, HSA, Monelina, Toxina colérica B e Tioredoxina W31A.
- Classe III: Melitina, Fosfolipase A2 e Glucagon.

Os valores médios da SASA do indol das proteínas das classes A, S, I, II e III são 0,45; 10,80; 32,35; 64,83 e 146,45 Å, mostrando uma correspondência com as classes discretas definidas por Burstein et al. Uma exceção é a proteína Mioglobina, a qual possui uma SASA do indol muito maior ao esperado para o seu λ_{em} . O λ_{em} calculado para esta proteína a partir da SASA ficaria em ~ 330 nm (Tabela 4.2), sendo sua classificação como Classe I.

4.3.3 Robustez dos modelos

Durante o processo de construção dos modelos a determinação dos parâmetros é um passo fundamental e necessário. Geralmente os parâmetros são otimizados considerando um conjunto de dados que servem para treinar o modelo. Esse conjunto de dados é conhecido como dados de treinamento. Para evitar que o modelo criado seja muito dependente dos dados do treinamento, é necessário de avaliar a capacidade preditiva do modelo, e para isso é necessário usar outro conjunto de dados, os quais chamaremos de dados de teste [82]. O uso de dados de treinamento e teste significa que uma considerável quantidade de dados é disponível, mas

isso não sempre é possível. Assim, métodos alternativos para avaliar a capacidade preditiva dos modelos, quando a quantidade de dados é limitada, são adotados. Vamos agora explicar como funciona esse método alternativo. Imaginemos que temos N dados. Desse conjunto, extraímos um número inteiro de k dados. Assim, usaremos $N - k$ dados como os dados de treinamento e os k dados restantes serão usados como dados teste. Devido a que o número de vezes que treinamos o modelo é uma combinatoria do tipo C_k^N , este método pode demandar um alto custo computacional quando o número N de dados for muito grande. Para exemplificar, usaremos um caso particular com $N = 4$ e $k = 1$. Assim, o modelo será construído com $N - k = 3$ dados e a robustez do modelo será avaliado a cada $k = 1$ dado. A quantidade de modelos construídos será de $C_1^4 = 4$ (Figura 4.7).

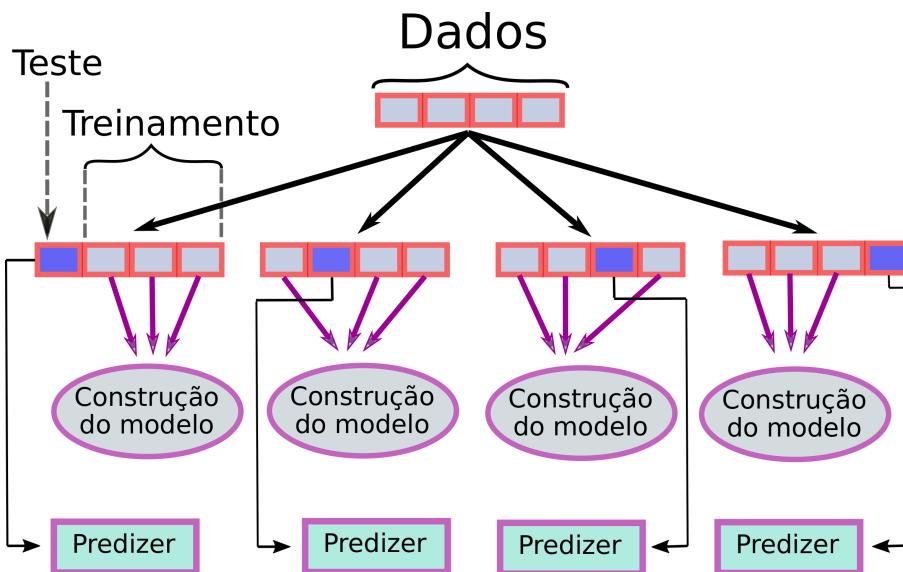


Figura 4.7: Representação esquemática do método de validação da robustez dos modelos. O número total de dados é $N = 4$. Para este caso em particular, o processo de validação considera deixar um dado fora ($k = 1$) para construir os modelos, assim o conjunto de dados de treinamento serão de $N - k = 3$ dados. O dado deixado fora será usado para avaliar a robustez do modelo em relação aos dados de treinamento usados para a parametrização.

Nesta Subseção vamos a determinar a robustez dos modelos lineares baseados na SASA do indol em relação aos dados de treinamento usados para a parametrização. Os modelos apresentados na Subseção 4.3.2 foram parametrizados usando um conjunto de $N = 19$ dados. Para avaliar a robustez desses modelos consideraremos valores de $k = 1, 2$ e 3 . Para $k = 1$, teremos $C_1^{19} = 19$ combinações diferentes de dados a partir das quais vamos a construir os modelos, cada um com $N - k = 18$ dados. O dado restante será avaliado com os modelos

construídos. Para $k = 2$ teremos $C_2^{19} = 171$ combinações diferentes de construir os modelos, cada um com $N - k = 17$ dados. Devido a que quando $k > 1$ as previsões para cada modelo serão feitas mais de uma vez, é necessário registrar as médias de cada predição. Para $k = 3$ teremos um similar cenário, mostrado anteriormente para com $k = 2$, com a diferença que o número de combinações diferentes de modelos será muito maior.

Vamos analisar agora a robustez dos modelos, baseados na SASA, desenvolvidos na Subseção 4.3.2. O primeiro modelo (Modelo 1) é aquele que depende unicamente da SASA do indol. Para $k = 1$, o desvio dos λ_{em} preditos com o novo conjunto de $N - k = 18$ dados de treinamento são avaliados com respeito ao λ_{em} calculado com o conjunto $N = 19$ de dados. Esse desvio está reportado como $\Delta\lambda_{em}^{calc}(1)$ na Tabela 4.4. Para valores de $k = 2$ e $k = 3$ teremos os desvios médios, os quais estão reportados na Tabela 4.4 como $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$, respectivamente. Os desvios e desvios médios são menores que 3 nm, indicando que o modelo é muito robusto em relação aos dados de treinamento usados. Os maiores desvios observados correspondem às proteínas com λ_{em} altos, como são a Melitina, Fosfolipase A2 e Glucagon. Desconsiderando estas proteínas, o maior desvio não supera os 6 nm. Portanto, os erros nas previsões não estão associados aos dados usados senão a limitações dos modelos.

Para os modelos baseados na SASA dos grupos benzênico e pirrólico do indol (Modelos 2 ao 4) os resultados de robustez foram similares aos reportados para ao caso da SASA do indol. Estes resultados encontram-se nas seguintes tabelas: Tabela B.1 (robustez do Modelo 2), Tabela B.2 (robustez do Modelo 3), e Tabela B.3 (robustez do Modelo 4).

4.4 Conclusões

O triptofano é a sonda natural mais amplamente usada nos experimentos de fluorescência de proteínas, para por exemplo avaliar a estrutura das proteínas. Um dos principais motivos para essa popularidade é a sensibilidade ao ambiente dos seus comprimentos de onda máximo de emissão (λ_{em}). O estado excitado responsável pela fluorescência do triptofano é o estado L_a , o qual apresenta um dipolo muito distinto do estado fundamental, sendo o principal motivo da variabilidade dos λ_{em} .

O triptofano livre em água é caracterizado por um λ_{em} de 360 nm. Nas proteínas,

Tabela 4.4: Robustez do modelo baseado na SASA do indol (Modelo 1) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 1 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	324,22	2,22	2,22(3,89)	2,21(5,26)
Parvalbumina	325,34	1,11	1,11(3,65)	1,11(4,95)
Mioglobina	328,69	0,61	0,62(2,51)	0,62(4,03)
Ribonuclease T1	324,47	-0,46	-0,46(1,99)	-0,47(3,89)
FKBP12	326,93	-0,29	-0,29(1,75)	-0,29(3,34)
T4-lisozima W126Y W138Y	332,23	0,13	0,14(1,45)	0,15(2,49)
T4-lisozima W126Y W158Y	325,08	-0,60	-0,61(-1,85)	-0,61(3,50)
T4-lisozima W138Y W158Y	332,20	0,13	0,14(1,45)	0,15(2,49)
subtilisina Calsberg	335,80	0,27	0,28(1,11)	0,30(1,79)
Nuclease de estafilococo	333,25	-0,04	-0,03(1,12)	-0,02(2,04)
Che-Y	326,05	-0,97	-0,97(-1,74)	-0,97(2,71)
α -Cobratoxina	339,52	-0,04	-0,02(0,69)	-0,01(1,85)
HSA	333,51	-0,36	-0,35(-0,98)	-0,34(-1,66)
Monelina	335,52	-0,37	-0,36(-0,98)	-0,35(-1,55)
Toxina da cólera B (monômero)	335,06	-0,56	-0,55(-1,09)	-0,54(-1,59)
Tioredoxina W31A	338,06	-0,47	-0,45(-1,10)	-0,44(-1,61)
Melitina	349,63	0,97	1,03(3,42)	1,09(9,16)
Fosfolipasa A2	352,24	1,56	1,63(3,86)	1,70(10,38)
Glucagon	353,20	0,49	0,57(3,53)	0,67(10,82)

o triptofano apresenta uma diversidade dos λ_{em} , o qual varia entre 308 e 355 nm. As proteínas com menor e maior λ_{em} , do conjunto de proteínas da Tabela 4.1, são a Azurina e Glucagon, respectivamente.

A Azurina apresenta o triptofano imerso numa cavidade da proteína que é inacessível ao solvente ($SASA = 0,45 \text{ \AA}^2$), enquanto o triptofano do Glucagon está completamente exposto ($SASA = 154 \text{ \AA}^2$), possibilitando o contato com as moléculas de água. Assim, apesar da natureza quântica da fluorescência, as características estruturais da proteína podem ser utilizadas para o desenvolvimento de modelos parametrizados. Embora muitos trabalhos tenham usado a ideia da acessibilidade ao solvente do indol para explicar a fluorescência em sistemas específicos, nenhum até o momento estabeleceu alguma relação quantitativa entre dita acessibilidade e os λ_{em} do triptofano. Nossa primeiro objetivo foi estabelecer e determinar essa relação.

Assim, quatro modelos clássicos parametrizados baseados na área acessível ao solvente do indol foram propostos para predizer os λ_{em} do triptofano em proteínas a partir de simulações de dinâmica molecular. Estes modelos apresentaram um desempenho bastante satisfatório, sendo o desvio padrão (SD) das previsões com respeito aos dados experimentais de ~ 7 nm, e o quadrado do coeficiente de correlação de Pearson de $R^2 \sim 0,6$. Apesar de o desempenho destes modelos clássicos ser similar ao obtido pelo modelo desenvolvido por Vivian e Callis (2004), os modelos baseados na *SASA* possuem um erro sistemático. Esse erro consiste em sobre-estimar os λ_{em} menores de 324 nm.

Capítulo 5

Modelos paramétricos baseados na interação eletrostática para computar os comprimentos de onda de emissão do triptofano

Neste capítulo serão desenvolvidos modelos clássicos parametrizados baseados nas interações eletrostáticas do indol com o ambiente (água + proteína). Observaremos quão importantes são as contribuições da água para a determinação dos comprimentos de onda. Estes modelos apresentaram melhor capacidade preditiva comparada aos modelos baseados na SASA. Estes modelos, junto aos apresentados no Capítulo 4, proporcionam uma maneira rápida e prática para interpretar os resultados experimentais.

5.1 Introdução

5.1.1 O papel das interações eletrostáticas nos λ_{em} do Trp

As características espectrais da fluorescência do triptofano, tais como o rendimento quântico (Φ_F) e o comprimento de onda de emissão (λ_{em}), são muito sensíveis à polaridade do ambiente. Por isso, essas características do triptofano são muito utilizadas para sondar a eletrostática interna e local das proteínas [15, 73, 74]. Essa sensibilidade está geralmente associada ao dipolo gerado pela excitação do indol ao estado L_a (Figura 5.1A) [49, 83].

A variabilidade dos λ_{em} do triptofano em proteínas, entre ~ 308 nm (Azurina) e ~ 355 nm (Glucagon), é um indicador da polaridade do meio e, como mostramos no Capítulo 4, também da acessibilidade ao solvente. Recentemente tem-se demonstrado que o estado excitado L_a , sem exceção, é o único estado responsável pela emissão de fluorescência do triptofano em proteínas [23].

Cálculos de estrutura eletrônica têm mostrado que a excitação do triptofano ao estado L_a é acompanhada de uma mudança de densidade eletrônica do anel pirrólico para o anel benzênico do indol [73, 74]. Assim, é de se esperar que as cargas na vizinhança dos anéis pirrólico e benzênico sejam importantes na determinação dos λ_{em} . Por exemplo, e citando as palavras de P.Callis (2004) [73, 74], cargas positivas (ou resíduos carregados positivamente) perto do anel benzênico ou cargas negativas (ou resíduos carregados negativamente) perto do anel pirrólico estabilizam o processo de transferência eletrônica e, portanto, contribuem para maiores comprimentos de onda (*red shift*). O processo contrário contribue para menores comprimentos de onda (*blue shift*).

Baseado nessas ideias, o grupo do Prof. Patrick Callis foi o pioneiro em desenvolver uma metodologia híbrida QM-MM para calcular os λ_{em} do triptofano em proteínas. Essa metodologia consiste, basicamente, em calcular as transições eletrônicas do grupo indol do triptofano por meio de cálculos de estrutura eletrônica considerando o resto do sistema (água + proteína) como cargas pontuais. Um requisito importante na determinação do λ_{em} é ter a geometria e cargas adequadas do estado L_a do indol, os quais podem ser determinados por cálculos quânticos. Simulações clássicas de dinâmica molecular de 19 proteínas contendo um único resíduo de triptofano foram realizadas, sendo a geometria e cargas dos átomos do indol correspondentes ao estado L_a . Portanto, essas simulações correspondem a simulações com a distribuição de cargas do estado excitado L_a do indol. As estruturas amostradas pela dinâmica molecular serviram como estruturas de entrada para calcular as transições eletrônicas e determinar os λ_{em} instantâneos. Os valores experimentais versus a média dos valores calculados dos λ_{em} do triptofano nas 19 proteínas estão mostrados na Figura 5.1B. As previsões do modelo híbrido QM-MM teve um desvio padrão com respeito aos dados experimentais de 6,87 nm e o quadrado do coeficiente de correlação de Pearson foi de $R^2 = 0,69$ nm.

Assim, as interações eletrostáticas entre os átomos de indol e os átomos restantes do

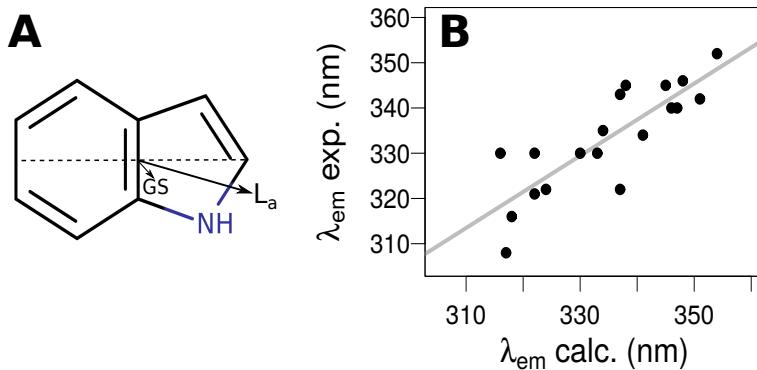


Figura 5.1: A) Orientação dos dipolos do estado fundamental (GS) e L_a do indol. B) Correlação entre os λ_{em} experimentais e calculados, segundo o modelo híbrido QM-MM desenvolvido por Vivian e Callis (2004) [74].

sistema (água + proteína) são determinantes do λ_{em} do triptofano. Esses resultados sugerem que as interações eletrostáticas podem ser usadas como potenciais preditores para o desenvolvimento de modelos clássicos parametrizados com capacidade de predizer o λ_{em} . Neste capítulo, descreveremos o desenvolvimento de modelos clássicos parametrizados baseados nas interações eletrostáticas do indol com o ambiente (água e proteína).

5.2 Metodologia

5.2.1 Simulações de dinâmica molecular

As simulações utilizadas nesta parte do trabalho são as mesmas descritas no capítulo anterior, portanto a metodologia está apresentada na Subseção 4.2.1.

5.2.2 Cálculo das interações eletrostáticas

As interações eletrostáticas entre o grupo indol e o ambiente (água + proteína) foram calculadas para cada *frame* das simulações usando a Equação 5.2.1.

$$E(t) = \frac{1}{4\pi\epsilon_0} \sum_i \sum_j \frac{q_i q_j}{d_{ij}} \quad (5.2.1)$$

sendo ϵ_0 a permissividade do vácuo, d_{ij} a distância entre os átomos i e j , e q_i e q_j as cargas dos átomos i e j , respectivamente. Os átomos i correspondem aos átomos do grupo indol do

triptofano e os átomos j correspondem aos demais átomos do sistema, excluindo os átomos ligados covalentemente ao indol. As cargas dos átomos correspondem às definidas pelo campo de força CHARMM22 [65, 66, 67], as mesmas que foram utilizadas nas simulações de dinâmica molecular. Todos os cálculos das interações eletrostáticas foram truncados num raio de corte de 15 Å. Esse valor de raio de corte é considerado efetivo para o cálculo das interações eletrostáticas devido a que é esperada uma convergência nos cálculos [84]. Os íons de sódio e cloro, usados para manter a neutralidade do sistema, não foram incluídos no cálculo das interações eletrostáticas. A decomposição do $E(t)$ nas suas contribuições atômicas foram realizadas de maneira simples devido à característica aditiva das interações eletrostáticas. O cálculo das interações eletrostáticas foi realizado com softwares do pacote MDAnalysis [85].

5.3 Resultados e discussões

5.3.1 Interações eletrostáticas do indol

Nos trabalhos de P. Callis, a introdução das interações eletrostáticas entre o grupo indol e o ambiente nos cálculos de estrutura eletrônica foram fundamentais para a determinação dos λ_{em} do Trp em proteínas [73, 74]. Portanto, é de se esperar que essas interações possam ser usadas como descritores para a construção de modelos clássicos parametrizados.

Para observar uma possível correlação entre as interações eletrostáticas e os λ_{em} experimentais nas proteínas, as interações eletrostáticas entre o indol e o ambiente (água + proteína) necessitam ser calculados para cada estrutura amostrada por dinâmica molecular. Assim, a evolução temporal dessas interações são calculadas para três proteínas: Parvalbumina ($\lambda_{em} = 316$ nm [6]), Nuclease de estafilococo ($\lambda_{em} = 334$ nm [5]), e o Glucagon ($\lambda_{em} = 352$ nm [27]). Na Figura 5.2 mostra-se a evolução temporal das interações eletrostáticas do indol com o ambiente (água + proteína) dessas proteínas. As interações eletrostáticas são muito variáveis ao longo das simulações, o que gera uma certa dificuldade para estabelecer alguma correlação a simples vista. Se calcularmos a média das interações eletrostáticas da Figura 5.2 obtemos valores de -3,29; -35,87 e -37,30 $kJ \times mol^{-1}$ para as proteínas Parvalbumina, Nuclease de estafilococo e Glucagon, respectivamente. Assim proteínas com maiores λ_{em} tem interações eletrostáticas mais negativas.

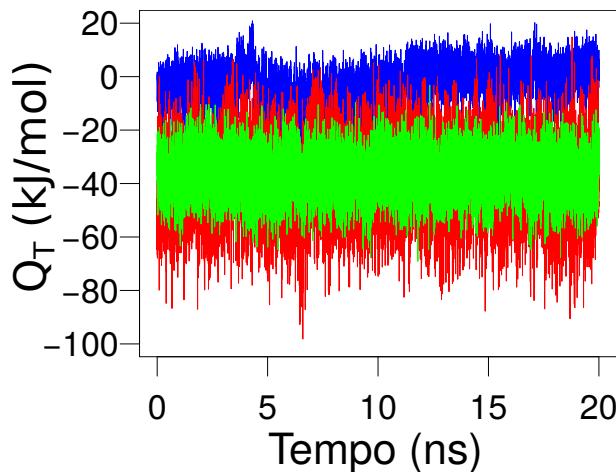


Figura 5.2: Evolução temporal das interações eletrostáticas do indol com o ambiente (água + proteína) nas proteínas: Parvalbumina (azul), Nuclease de Estafilococo (verde) e Glucagon (vermelho).

Os cálculos das interações eletrostáticas do indol com o ambiente (água + proteína) foram ampliados para um conjunto de 19 proteínas, as quais possuem um único resíduo de triptofano (Tabela 5.1). Com esses dados será possível construir um gráfico que mostre a correlação entre o valor médio das interações eletrostáticas e o λ_{em} experimental. Esse gráfico está na Figura 5.3. Nessa figura observa-se uma correlação negativa entre as interações eletrostáticas e os λ_{em} experimentais. O coeficiente de correlação é de $R^2 = 0,69$. A correlação entre as interações eletrostáticas e os λ_{em} experimentais parece ser razoavelmente linear, apesar de que algumas previsões apresentam desvios significativos. A proteína que apresenta um maior desvio da linearidade é a Mioglobina (Figura 5.3). Ainda assim, este resultado abre a possibilidade de formular modelos clássicos baseados nas interações eletrostáticas.

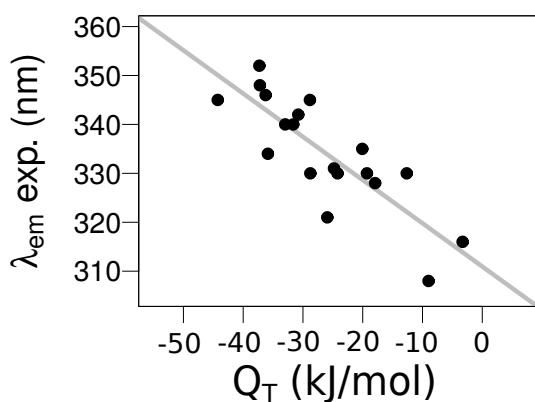


Figura 5.3: Correlação entre o comprimento de onda de emissão de fluorescência experimental e o valor médio das interações eletrostáticas do indol com o ambiente (água + proteína) para 19 proteínas que contém um único resíduo de triptofano.

As interações eletrostáticas do indol com o ambiente (Q_T) podem ser divididas em função das suas contribuições individuais, da água e da proteína. Assim as interações eletrostáticas do indol com a água (Q_w) e do indol com a proteína (Q_p) foram calculadas, e seus valores médios estão apresentados na Tabela 5.1. Devido ao comportamento aditivo dessas interações, cumpre-se o seguinte: $Q_T = Q_w + Q_p$. Adicionalmente, o indol pode ser dividido em suas componentes benzênicas e pirrólicas (como foi descrito na Subseção 4.3.2), e as interações eletrostáticas dessas componentes podem ser avaliadas. Assim temos que as interações eletrostáticas entre o benzeno incompleto do indol e a água ($Q_{ben-i,w}$), o benzeno incompleto do indol e a proteína ($Q_{ben-i,p}$), o benzeno completo do indol e a água ($Q_{ben-c,w}$), o benzeno completo do indol e a proteína ($Q_{ben-c,p}$), o pirrol incompleto do indol e a água ($Q_{pyr-i,w}$), o pirrol incompleto do indol e a proteína ($Q_{pyr-i,p}$), o pirrol completo do indol e a água ($Q_{pyr-c,w}$) e o pirrol completo do indol e a proteína ($Q_{pyr-c,p}$) foram calculadas e seus valores médios estão apresentados na Tabela 5.2 e 5.3.

Com esses dados podemos definir o seguinte:

- A interação eletrostática entre o benzeno incompleto e o ambiente (água + proteína) será computada como: $Q_{ben-i} = Q_{ben-i,w} + Q_{ben-i,p}$.
- A interação eletrostática entre o benzeno completo e o ambiente (água + proteína) será computada como: $Q_{ben-c} = Q_{ben-c,w} + Q_{ben-c,p}$.
- A interação eletrostática entre o pirrol incompleto e o ambiente (água + proteína) será computada como: $Q_{pyr-i} = Q_{pyr-i,w} + Q_{pyr-i,p}$.
- A interação eletrostática entre o pirrol completo e o ambiente (água + proteína) será computada como: $Q_{pyr-c} = Q_{pyr-c,w} + Q_{pyr-c,p}$.

Tabela 5.1: Proteínas usadas, códigos de acesso das estruturas (**PDB id.**), comprimentos de onda de emissão de fluorescência experimentais (λ_{em}^{exp}), e interações eletrostáticas calculadas a partir das simulações de dinâmica molecular. Q_T , Q_w e Q_p são as interações eletrostáticas entre o indol e o ambiente (água + proteína), o indol e a água, e o indol e a proteína, respectivamente. O símbolo $\langle \rangle$ representa a média sobre as simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as interações eletrostáticas em $kJ \times mol^{-1}$.

Proteína	PDB id.	λ_{em}^{exp}	$\langle Q_T \rangle$	$\langle Q_w \rangle$	$\langle Q_p \rangle$
Azurina	4AZU[22]	308[23]	-8,99±5,05	-3,05±4,31	-5,94±3,06
Parvalbumina	1B8R[24]	316[6]	-3,29±8,67	-6,90±8,28	3,61±4,13
Mioglobina	1MYT[25]	321[6]	-25,93±13,81	-3,62±16,30	-22,31±20,75
Ribonuclease T1	9RNT[26]	328[27]	-17,93±8,05	-8,37±7,11	-9,56±4,75
FKBP12	1D6O[28]	330[29]	-12,64±7,65	-12,21±7,43	-0,43±3,95
T4-lisozima W126Y W138Y	1LYD[30]	330[31]	-28,77±15,97	-13,48±10,72	-15,29±14,12
T4-lisozima W126Y W158Y	1LYD[30]	330[31]	-24,22±6,75	4,88±4,69	-29,10±6,08
T4-lisozima W138Y W158Y	1LYD[30]	330[31]	-19,30±11,35	-20,04±11,59	0,74±6,47
subtilisina Carlsberg	1SBC[32]	331[33]	-24,81±10,79	-18,91±10,64	-5,90±4,07
Nucleasa de estafilococo	1STN[34]	334[5]	-35,87±8,02	-17,90±8,96	-17,97±6,41
Che-Y	1CHN[35]	335[6]	-20,08±11,47	-12,84±11,27	-7,24±5,58
α -Cobratoxina	1CTX[36]	340[6]	-32,97±12,16	-21,11±13,12	-11,86±6,17
HSA	1BM0[37]	340[38]	-31,63±11,61	-25,23±11,46	-6,40±5,36
Monelina	1MOL[39]	342[27]	-30,79±13,18	-23,54±15,70	-7,25±12,17
Toxina colérica B (monômero)	1CHP[40]	345[6]	-28,84±11,69	-19,43±11,76	-9,41±5,83
Tioredoxina W31A	2TRX[41]	345[21]	-44,26±14,05	-16,35±11,70	-27,91±12,16
Melitina	2MLT[42]	346[6]	-36,25±13,82	-38,03±15,67	1,78±6,88
Fosfolipase A2	2BPP[43]	348[44]	-37,21±13,92	-31,00±14,42	-6,21±5,07
Glucagon	1GCN[45]	352[27]	-37,30±13,77	-48,40±14,13	11,10±4,13

Tabela 5.2: Proteínas usadas, códigos de acesso das estruturas (**PDB id.**), comprimentos de onda de emissão de fluorescência experimentais (λ_{em}^{exp}), e interações eletrostáticas calculadas a partir das simulações de dinâmica molecular. $Q_{ben-i,w}$, $Q_{ben-i,p}$, $Q_{pyr-c,w}$ e $Q_{pyr-c,p}$ são as interações eletrostáticas entre o benzeno incompleto do indol e a água, o benzeno incompleto do indol e a proteína, o pirrol completo do indol e a água, o pirrol completo do indol e a proteína e respectivamente. As interações eletrostáticas entre o benzeno incompleto e o ambiente (água + proteína) são computadas como $Q_{ben-i} = Q_{ben-i,w} + Q_{ben-i,p}$. Da mesma maneira, as interações eletrostáticas entre o pirrol completo e o ambiente (água + proteína) são computadas como $Q_{pyr-c} = Q_{pyr-c,w} + Q_{pyr-c,p}$. O símbolo $\langle \rangle$ representa a média sobre as simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as interações eletrostáticas em $kJ \times mol^{-1}$.

Proteína	PDB id.	λ_{em}^{exp}	$\langle Q_{ben-i,w} \rangle$	$\langle Q_{ben-i,p} \rangle$	$\langle Q_{pyr-c,w} \rangle$	$\langle Q_{pyr-c,p} \rangle$
Azurina	4AZU[22]	308[23]	1,01±1,96	-2,13±1,90	-4,06±3,88	-3,80±2,66
Parvalbumina	1B8R[24]	316[6]	1,38±2,27	-4,25±2,98	-8,28±7,87	7,86±4,47
Mioglobina	1MYT[25]	321[6]	-2,77±5,45	-3,17±6,04	-0,86±14,59	-19,14±17,87
Ribonuclease T1	9RNT[26]	328[27]	3,08±2,80	-9,76±2,50	-11,45±6,79	0,20±3,76
FKBP12	1D6O[28]	330[29]	-3,12±3,87	-1,94±2,50	-9,09±5,70	1,52±3,29
T4-lisozima W126Y W138Y	1LYD[30]	330[31]	2,39±3,64	-9,62±4,25	-15,88±10,67	-5,67±10,79
T4-lisozima W126Y W158Y	1LYD[30]	330[31]	-2,68±2,22	0,061±2,03	7,56±4,31	-29,17±5,89
T4-lisozima W138Y W158Y	1LYD[30]	330[31]	-4,10±4,21	0,43±4,13	-15,94±10,81	0,31±4,72
subtilisina Carlsberg	1SBC[32]	331[33]	-2,73±3,22	-1,13±2,11	-16,19±10,00	-4,76±3,17
Nucleasa de estafilococo	1STN[34]	334[5]	-1,86±5,10	-6,68±4,37	-16,03±7,13	-11,29±3,34
Che-Y	1CHN[35]	335[6]	-6,81±4,85	2,19±2,72	-6,03±10,24	-9,44±5,39
α -Cobratoxina	1CTX[36]	340[6]	-0,25±5,97	-10,09±3,88	-20,86±11,28	-1,77±3,73
HSA	1BM0[37]	340[38]	-5,71±5,30	0,79±3,13	-19,52±10,01	-7,19±4,51
Monelina	1MOL[39]	342[27]	1,07±5,54	-9,17±5,74	-24,60±13,86	1,91±9,31
Toxina colérica B (monômero)	1CHP[40]	345[6]	-6,53±5,12	-3,41±3,01	-12,89±10,29	-6,01±4,57
Tioredoxina W31A	2TRX[41]	345[21]	-1,53±6,31	-12,18±5,05	-14,82±9,97	-15,72±9,01
Melitina	2MLT[42]	346[6]	-9,02±6,91	-1,75±4,73	-29,01±13,02	3,52±4,89
Fosfolipase A2	2BPP[43]	348[44]	-7,68±6,56	-5,52±3,60	-23,32±12,04	-0,69±3,20
Glucagon	1GCN[45]	352[27]	-17,54±6,50	6,11±1,96	-30,86±12,23	4,98±3,48

Tabela 5.3: Proteínas usadas, códigos de acesso das estruturas (**PDB id.**), comprimentos de onda de emissão de fluorescência experimentais (λ_{em}^{exp}), e interações eletrostáticas calculadas a partir das simulações de dinâmica molecular. $Q_{ben-c,w}$, $Q_{ben-c,p}$, $Q_{pyr-i,w}$ e $Q_{pyr-i,p}$ são as interações eletrostáticas entre o benzeno completo do indol e a água, benzeno completo do indol e a proteína, pirrol incompleto do indol e a água e pirrol incompleto do indol e a proteína, respectivamente. As interações eletrostáticas entre o benzeno completo e o ambiente (água + proteína) são computados como $Q_{ben-c} = Q_{ben-c,w} + Q_{ben-c,p}$. Da mesma maneira, as interações eletrostáticas entre o pirrol incompleto e o ambiente (água + proteína) são computados como $Q_{pyr-i} = Q_{pyr-i,w} + Q_{pyr-i,p}$. O símbolo $\langle \rangle$ representa a média sobre as simulações. O desvio padrão é mostrado. Os λ_{em}^{exp} estão em nm, e as interações eletrostáticas em $kJ \times mol^{-1}$.

Proteína	PDB id.	λ_{em}^{exp}	$\langle Q_{ben-c,w} \rangle$	$\langle Q_{ben-c,p} \rangle$	$\langle Q_{pyr-i,w} \rangle$	$\langle Q_{pyr-i,p} \rangle$
Azurina	4AZU[22]	308[23]	7,73±2,78	-4,33±2,70	-10,78±4,36	-1,61±3,58
Parvalbumina	1B8R[24]	316[6]	9,39±3,28	-7,25±4,21	-16,29±7,45	10,86±6,03
Mioglobina	1MYT[25]	321[6]	5,92±7,59	-7,08±7,97	-9,54±14,81	-15,23±16,32
Ribonuclease T1	9RNT[26]	328[27]	12,24±4,55	-13,88±3,48	-20,61±7,09	4,33±5,08
FKBP12	1D6O[28]	330[29]	6,76±4,19	-3,38±3,34	-18,97±8,39	2,96±3,94
T4-lisozima W126Y W138Y	1LYD[30]	330[31]	-13,57±6,78	19,59±5,04	0,09±11,15	-34,88±17,22
T4-lisozima W126Y W158Y	1LYD[30]	330[31]	-3,92±3,27	5,20±2,71	8,80±5,03	-34,31±6,12
T4-lisozima W138Y W158Y	1LYD[30]	330[31]	-8,45±8,26	14,11±5,83	-11,59±10,28	-13,37±6,53
subtilisina Calsberg	1SBC[32]	331[33]	-2,87±6,20	4,55±3,18	-16,04±11,14	-10,45±5,25
Nucleasa de estafilococo	1STN[34]	334[5]	-5,13±7,67	2,25±4,80	-12,77±10,51	-20,22±7,36
Che-Y	1CHN[35]	335[6]	-3,99±5,89	2,30±4,35	-8,85±10,99	-9,54±6,79
α -Cobratoxina	1CTX[36]	340[6]	-6,88±8,74	1,34±5,67	-14,23±13,86	-13,20±7,33
HSA	1BM0[37]	340[38]	-18,83±10,91	30,66±4,92	-6,40±12,85	-37,06±6,30
Monelina	1MOL[39]	342[27]	1,51±10,35	-1,19±11,41	-25,05±14,55	-6,07±8,82
Toxina colérica B (monômero)	1CHP[40]	345[6]	6,87±6,42	-9,10±4,35	-26,29±12,16	-0,32±6,56
Tioredoxina W31A	2TRX[41]	345[21]	-10,52±9,10	6,05±6,76	-5,83±12,78	-33,96±14,27
Melitina	2MLT[42]	346[6]	-21,48±9,18	25,35±4,82	-16,55±16,51	-23,57±7,27
Fosfolipase A2	2BPP[43]	348[44]	-8,20±9,53	3,52±4,61	-22,79±15,62	-9,73±5,11
Glucagon	1GCN[45]	352[27]	5,87±9,84	-13,17±2,73	-54,27±16,21	24,27±5,11

5.3.2 Construção dos modelos

Uma relação linear entre as interações eletrostáticas do indol com o ambiente (água + proteína) e os λ_{em} do Trp em proteínas foi encontrada e descrita na subseção anterior. Nesta subseção descreveremos a construção de modelos clássicos parametrizados baseados nas interações eletrostáticas com capacidade preditiva dos λ_{em} do Trp em proteínas.

O primeiro modelo (Modelo 1) explora a relação linear entre as interações eletrostáticas do indol com o ambiente (água + proteína) e os λ_{em} . Este modelo considera apenas a presença de dois parâmetros, sendo estes o A_0^{pot} e A_1^{pot} . A forma do modelo é descrita pela Equação 5.3.1.

$$\lambda_{em} = A_0^{pot} + A_1^{pot} \times \langle Q_T \rangle \quad (5.3.1)$$

sendo Q_T as interações eletrostáticas entre o indol e o ambiente (água + proteína) e o termo $\langle \rangle$ indica a média ao longo da simulação de dinâmica molecular. O parâmetro A_0^{pot} pode ser interpretado como o λ_{em} quando Q_T é nulo, e A_1^{pot} é o parâmetro que associa o Q_T ao λ_{em} . Os parâmetros do modelo foram otimizados com os dados de $\langle Q_T \rangle$ de 19 proteínas (Tabela 5.1). Esses parâmetros estão mostrados na Tabela 5.6. As previsões do λ_{em} das 19 proteínas do Modelo 1, estão mostradas na Tabela 5.4. As piores previsões do modelo correspondem às proteínas Mioglobina e Azurina, com desvios de ~ 13 e ~ 11 nm, respectivamente. Entre as proteínas que mostraram as melhores previsões estão a α -Cobratoxina e HSA, com desvios de $\sim 0,09$ e $\sim -1,00$ nm, respectivamente. A Figura 5.4A mostra a correlação entre o λ_{em} experimental e o calculado com o Modelo 1. O coeficiente de correlação e o desvio padrão das previsões deste modelo com respeito ao λ_{em} experimental foram de $R^2 = 0,69$ e $6,36$ nm, respectivamente. O desempenho deste modelo não parece ser melhor comparado aos modelos baseado na SASA. Apesar disso, o modelo baseado nas interações eletrostáticas tem capacidade preditiva para λ_{em} baixos, sendo assim uma vantagem em relação aos modelos baseados em SASA.

As contribuições das interações eletrostáticas do indol com o sistema podem ser divididas nas componentes que envolvem a interação do indol com a água e com a proteína.

Baseados nessa decomposição é que propomos o Modelo 2, o qual é composto de três parâmetros (B_0^{pot} , B_1^{pot} e B_2^{pot}),

$$\lambda_{em} = B_0^{pot} + B_1^{pot} \times \langle Q_w \rangle + B_2^{pot} \times \langle Q_p \rangle \quad (5.3.2)$$

sendo Q_w e Q_p as interações eletrostáticas do indol com a água e a proteína, respectivamente. O termo $\langle \rangle$ indica a média ao longo da simulação de dinâmica molecular. B_0^{pot} é o intercepto do modelo. B_1^{pot} e B_2^{pot} são os parâmetros que associam as interações eletrostáticas do indol - água e indol - proteína com o λ_{em} , respectivamente. Os dados de Q_w e Q_p das 19 proteínas, as quais estão mostrados na Tabela 5.1, foram usados para otimizar os parâmetros do Modelo 2 (Tabela 5.6). As previsões dos λ_{em} das 19 proteínas deste modelo estão mostradas na Tabela 5.4. O maior e menor desvios das previsões do λ_{em} com respeito aos dados experimentais foram de 10,04 e 0,24 nm e correspondem às proteínas Azurina e Parvalbumina, respectivamente. Na Figura 5.4B mostra-se a correlação entre os λ_{em} experimentais e os preditos com o Modelo 2 para 19 proteínas. O coeficiente de correlação e o desvio padrão das previsões do Modelo 2 com respeito aos dados experimentais são de $R^2 = 0,81$ e 4,89 nm, respectivamente. A divisão das contribuições eletrostáticas melhorou significativamente o desempenho do modelo. O modelo híbrido QM-MM, desenvolvido por Vivian e Callis (2004), teve um desvio padrão com respeito aos dados experimentais de 6,87 nm e um coeficiente de correlação $R^2 = 0,69$ [74]. Assim, nosso modelo clássico parametrizado, apesar da sua simplicidade, mostra-se promissor para a previsão dos λ_{em} do triptofano em proteínas, e assim se apresenta como um método rápido e prático para interpretar os resultados experimentais.

Como foi explicado na Subseção 4.3.2 do Capítulo 4, o grupo indol pode ser dividido nas suas componentes benzênica e pirrólica. Os modelos que desenvolveremos a continuação (Modelos 3 ao 8), serão baseados nas interações eletrostáticas da parte benzênica e pirrólica do indol.

O Modelo 3 possui três parâmetros e é baseado nas interações eletrostáticas das partes benzênicas e pirrólicas do indol com o ambiente (água + proteína). Este modelo é

mostrado na Equação 5.3.3,

$$\lambda_{em} = C_0^{pot} + C_1^{pot} \times \langle Q_{ben,i} \rangle + C_2^{pot} \times \langle Q_{pyr,c} \rangle \quad (5.3.3)$$

sendo C_0^{pot} , C_1^{pot} e C_2^{pot} os parâmetros do modelo. $Q_{ben,i}$ e $Q_{pyr,c}$ são as interações eletrostáticas do benzeno incompleto com o ambiente (água + proteína) e do pirrol completo com o ambiente (água + proteína), respectivamente. O termo $\langle \rangle$ indica a média ao longo da simulação de dinâmica molecular. Os parâmetros do modelo 3 foram ajustados com os dados das interações eletrostáticas $Q_{ben,i}$ e $Q_{pyr,c}$ das mesmas 19 proteínas (Tabela 5.2). Os valores dos parâmetros estão na Tabela 5.6. Os valores dos λ_{em} preditos com o Modelo 3 estão na Tabela 5.4. O maior e menor desvios dos λ_{em} preditos com respeito aos dados experimentais são de ~ 12 e $\sim -0,2$ nm e correspondem às proteínas Mioglobina e Fosfolipase A2, respectivamente. A Figura 5.4C mostra a correlação entre os λ_{em} preditos pelo modelo e os dados experimentais. O coeficiente de correlação e o desvio padrão dos λ_{em} preditos com respeito aos dados experimentais deste modelo são de $R^2 = 0,74$ e 5,77 nm, respectivamente.

A interação eletrostática do benzeno incompleto com o ambiente (água + proteína) pode ser decomposta em função das contribuições da água e proteína ($Q_{ben,i} = Q_{ben-w,i} + Q_{ben-p,i}$). Decomposição similar pode ser feita com a interação eletrostática do pirrol completo com o ambiente (água + proteína). Essas decomposições serão consideradas para a construção de um novo modelo. Assim, o Modelo 4 vai possuir 5 parâmetros (Equação 5.3.4) e portanto, apresentará maior complexidade que os anteriores. A equação do modelo é,

$$\lambda_{em} = D_0^{pot} + D_1^{pot} \times \langle Q_{ben-w,i} \rangle + D_2^{pot} \times \langle Q_{ben-p,i} \rangle + D_3^{pot} \times \langle Q_{pyr-w,c} \rangle + D_4^{pot} \times \langle Q_{pyr-p,c} \rangle \quad (5.3.4)$$

sendo D_0^{pot} , D_1^{pot} , D_2^{pot} , D_3^{pot} e D_4^{pot} os parâmetros do Modelo 4. $Q_{ben-w,i}$, $Q_{ben-p,i}$, $Q_{pyr-w,c}$ e $Q_{pyr-p,c}$ são as interações eletrostáticas do benzeno incompleto - água, benzeno incompleto - proteína, pirrol completo - água e pirrol completo - proteína, respectivamente. O termo $\langle \rangle$ indica a média ao longo da simulação de dinâmica molecular. Este modelo foi construído com os dados das interações eletrostáticas das 19 proteínas da Tabela 5.2, e os parâmetros otimizados do

modelo estão disponíveis na Tabela 5.6. A introdução de mais parâmetros no modelo conseguiu melhorar a capacidade preditiva do modelo, sendo o coeficiente de correlação e o desvio padrão dos λ_{em} preditos com respeito aos dados experimentais de $R^2 = 0,85$ e 4,45 nm. As previsões dos λ_{em} das 19 proteínas usando o Modelo 4 estão na Tabela 5.4. Os desvios dos λ_{em} preditos com o Modelo 4 em relação aos dados experimentais para as proteínas Azurina e subtilisina Carlsberg foram os maiores (~ 9 nm) e menores ($\sim 0,2$ nm), respectivamente. A correlação entre os λ_{em} preditos pelo modelo e os dados experimentais está mostrada na Figura 5.4D.

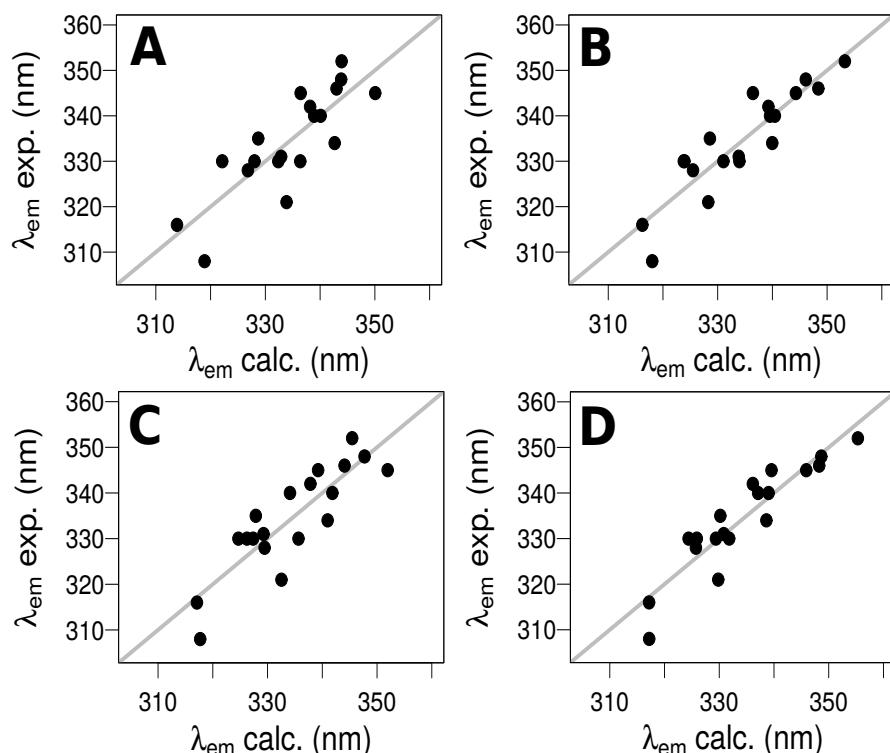


Figura 5.4: Comprimento de onda de emissão de fluorescência experimental (λ_{em}^{exp}) versus o calculado (λ_{em}^{calc}) usando os modelos baseados nas interações eletrostáticas. A) Modelo 1: Baseado na interação eletrostática do grupo indol com o ambiente (água + proteína). B) Modelo 2: Baseado nas interações eletrostáticas do grupo indol com a água e a proteína. C) Modelo 3: Baseado nas interações eletrostáticas do benzeno incompleto-ambiente (água + proteína) e pirrol completo-ambiente (água + proteína). D) Modelo 4: Baseado na decomposição das interações eletrostáticas do benzeno incompleto e pirrol completo com a água e a proteína.

Tabela 5.4: Comprimentos de onda de emissão de fluorescência experimentais e calculados do triptofano: λ_{em}^{exp} é o comprimento de onda de emissão de fluorescência experimental. $\lambda_{em}^{(1-4)}$ e $\Delta\lambda_{em}^{(1-4)}$ são os comprimentos de onda de emissão de fluorescência calculados e os desvios relativos aos valores experimentais, respectivamente, usando os modelos baseados nas interações eletrostáticas 1 ao 4. Os comprimentos de onda e os desvios estão em nm.

Proteína	λ_{em}^{exp}	$\lambda_{em}^{(1)}$	$\Delta\lambda_{em}^{(1)}$	$\lambda_{em}^{(2)}$	$\Delta\lambda_{em}^{(2)}$	$\lambda_{em}^{(3)}$	$\Delta\lambda_{em}^{(3)}$	$\lambda_{em}^{(4)}$	$\Delta\lambda_{em}^{(4)}$
Azurina	308[23]	318,90	10,89	318,04	10,04	317,70	9,69	317,20	9,19
Parvalbumina	316[6]	313,86	-2,14	316,24	0,24	317,10	1,10	317,18	1,18
Mioglobina	321[6]	333,87	12,87	328,28	7,28	332,54	11,54	329,82	8,82
Ribonuclease T1	328[27]	326,80	-1,19	325,49	-2,51	329,46	1,46	325,73	-2,27
FKBP12	330[29]	322,12	-7,88	323,93	-6,07	324,68	-5,32	325,90	-4,10
T4-lisozima W126Y W138Y	330[31]	336,38	6,38	333,97	3,97	335,65	5,65	331,82	1,82
T4-lisozima W126Y W158Y	330[31]	332,36	2,36	323,81	-6,19	327,36	-2,64	324,40	-5,60
T4-lisozima W138Y W158Y	330[31]	328,02	-1,98	331,06	1,05	326,24	-3,76	329,39	-0,61
subtilisina Carlsberg	331[33]	332,88	1,88	333,85	2,85	329,26	-1,74	330,82	-0,18
Nuclease de estafilococo	334[5]	342,66	8,66	339,97	5,97	340,97	6,97	338,61	4,61
Che-Y	335[6]	328,70	-6,29	328,58	-6,42	327,86	-7,14	330,20	-4,80
α -Cobratoxina	340[6]	340,09	0,09	339,57	-0,43	341,85	1,85	339,02	-0,98
HSA	340[38]	338,91	-1,09	340,46	0,46	334,10	-5,90	337,10	-2,89
Monelina	342[27]	338,17	-3,83	339,27	-2,73	337,83	-4,17	336,14	-5,86
Toxina colérica B	345[6]	336,44	-8,56	336,44	-8,56	339,23	-5,76	339,55	-5,45
Tioredoxina W31A	345[21]	350,07	5,07	344,30	-0,70	351,94	6,94	345,90	0,89
Melitina	346[6]	343,00	-3,00	348,40	2,40	344,06	-1,94	348,26	2,26
Fosfolipase A2	348[44]	343,84	-4,16	346,10	-1,89	347,71	-0,29	348,65	0,65
Glucagon	352[27]	343,92	-8,08	353,25	1,25	345,45	-6,55	355,32	3,32

À continuação apresenta-se o Modelo 5, baseado nas interações eletrostáticas do benzeno completo e do pirrol incompleto com o ambiente (água + proteína). Este modelo contém 3 parâmetros e tem a seguinte forma (Equação 5.3.5):

$$\lambda_{em} = E_0^{pot} + E_1^{pot} \times \langle Q_{ben,c} \rangle + E_2^{pot} \times \langle Q_{pyr,i} \rangle \quad (5.3.5)$$

sendo E_0^{pot} , E_1^{pot} e E_2^{pot} os parâmetros do modelo. Os termos $\langle Q_{ben,c} \rangle$ e $\langle Q_{pyr,i} \rangle$ são os preditores do modelo e representam as interações eletrostáticas do benzeno completo - sistema (água + proteína) e pirrol incompleto - sistema (água + proteína), respectivamente. A média tomada ao longo de uma simulação de dinâmica molecular é representada pelo termo $\langle \rangle$. As $\langle Q_{ben,c} \rangle$ e $\langle Q_{pyr,i} \rangle$ para 19 proteínas foram calculadas e os valores estão na Tabela 5.3. Com esse conjunto de dados se ajustaram os parâmetros do Modelo 5 (Tabela 5.6). Este modelo apresentou um coeficiente de correlação e um desvio padrão dos λ_{em} preditos com respeito aos dados experimentais de $R^2 = 0,69$ e $6,31$ nm, respectivamente. Entre as melhores e piores previsões do modelo temos as proteínas Ribonuclease T1 e Mioglobina, as quais apresentaram desvios da previsão dos λ_{em} com respeito aos dados experimentais de $\sim 0,6$ e ~ 13 nm, respectivamente. As previsões dos λ_{em} e os desvios com respeito aos dados experimentais das 19 proteínas usando o Modelo 5 estão descritos na Tabela 5.5. Na Figura 5.5A mostra-se a correlação entre os λ_{em} preditos pelo modelo e os dados experimentais.

A decomposição do Modelo 5 tomando em consideração diferentes pesos para as interações eletrostáticas com a água e a proteína serão considerados no Modelo 6 (Equação 5.3.6). Essa decomposição aumenta o número de parâmetros para 5,

$$\lambda_{em} = F_0^{pot} + F_1^{pot} \times \langle Q_{ben-w,c} \rangle + F_2^{pot} \times \langle Q_{ben-p,c} \rangle + F_3^{pot} \times \langle Q_{pyr-w,i} \rangle + F_4^{pot} \times \langle Q_{pyr-p,i} \rangle \quad (5.3.6)$$

sendo F_0^{pot} , F_1^{pot} , F_2^{pot} , F_3^{pot} e F_4^{pot} os parâmetros do modelo. Os termos $\langle Q_{ben-w,c} \rangle$, $\langle Q_{ben-p,c} \rangle$, $\langle Q_{pyr-w,i} \rangle$ e $\langle Q_{pyr-p,i} \rangle$ representam as interações eletrostáticas do benzeno completo - água, benzeno completo - proteína, pirrol incompleto - água e pirrol incompleto - proteína, respectivamente. A média tomada ao longo da simulação de dinâmica molecular está representada pelo $\langle \rangle$. Os

descritores $Q_{ben-w,c}$, $Q_{ben-p,c}$, $Q_{pyr-w,i}$ e $Q_{pyr-p,i}$ foram calculados para um conjunto de 19 proteínas e os seus valores estão na Tabela 5.3. Esse conjunto de dados foi usado para treinar o Modelo 6 (Equação 5.3.6), e os parâmetros estão descritos na Tabela 5.6. A capacidade preditiva do modelo foi melhorada com a introdução de mais parâmetros, com respeito ao Modelo 5, sendo agora o coeficiente de correlação e o desvio padrão dos λ_{em} preditos com respeito aos dados experimentais de $R^2 = 0,82$ e $4,81$ nm (Tabela 5.6). Na Tabela 5.5 estão mostradas as previsões dos λ_{em} das 19 proteínas do Modelo 6. Os λ_{em} das proteínas T4 Lisozima W138Y W158Y e Azurina foram as melhores e piores previsões do Modelo 6, respectivamente (Tabela 5.5). Na Figura 5.5B está mostrada a correlação entre os λ_{em} preditos pelo Modelo 6 e os λ_{em} experimentais.

O modelo 7 baseia-se nas interações eletrostáticas do benzeno completo e pirrol completo com o ambiente (água + proteína). A forma deste modelo está detalhada na Equação 5.3.7,

$$\lambda_{em} = G_0^{pot} + G_1^{pot} \times \langle Q_{ben,c} \rangle + G_2^{pot} \times \langle Q_{pyr,c} \rangle \quad (5.3.7)$$

sendo G_0^{pot} , G_1^{pot} e G_2^{pot} os parâmetros deste modelo e os termos $Q_{ben,c}$ e $Q_{pyr,c}$ as interações eletrostáticas do benzeno completo - sistema (água + proteína) e pirrol completo - sistema (água + proteína), respectivamente. $\langle \rangle$ representa a média de uma simulação de dinâmica molecular. Os valores médios das $Q_{ben,c}$ e $Q_{pyr,c}$ para 19 proteínas foram calculados e estão mostrados nas Tabelas 5.3 e 5.2, respectivamente. Esse conjunto de dados é usado para treinar o Modelo 7 e os seus parâmetros encontram-se na Tabela 5.6. O coeficiente de correlação e o desvio padrão dos λ_{em} preditos com respeito aos dados experimentais do modelo são $R^2 = 0,59$ e $7,27$ nm, respectivamente. As previsões dos λ_{em} das 19 proteínas encontram-se na Tabela 5.5. A melhor previsão corresponde ao λ_{em} da proteína Ribonuclease T1, sendo o desvio com respeito ao λ_{em} experimental de $\sim 0,7$ nm. Por outro lado, os maiores desvios, os quais foram da ordem dos ~ 13 nm, correspondem às previsões dos λ_{em} das proteínas Azurina e Mioglobina. A Figura 5.5C revela a correlação entre as previsões dos λ_{em} usando o Modelo 7 e os dados experimentais.

O Modelo 8 é uma variação do Modelo 7, no qual as interações eletrostáticas com a água e a proteína são tratadas com pesos diferentes. Isso fará com que a complexidade do

modelo seja maior com respeito ao anterior (Equação 5.3.8),

$$\lambda_{em} = H_0^{pot} + H_1^{pot} \times \langle Q_{ben-w,c} \rangle + H_2^{pot} \times \langle Q_{ben-p,c} \rangle + H_3^{pot} \times \langle Q_{pyr-w,c} \rangle + H_4^{pot} \times \langle Q_{pyr-p,c} \rangle \quad (5.3.8)$$

sendo H_0^{pot} , H_1^{pot} , H_2^{pot} , H_3^{pot} e H_4^{pot} os parâmetros do Modelo 8. As interações eletrostáticas entre o benzeno completo e a água ($Q_{ben-w,c}$), o benzeno completo e a proteína ($Q_{ben-p,c}$), o pirrol completo e a água ($Q_{pyr-w,c}$) e o pirrol completo e a proteína ($Q_{pyr-p,c}$) foram calculados para um conjunto de 19 proteínas e os valores médios estão mostrados nas Tabelas 5.2 e 5.3. Com esse conjunto de dados, o Modelo 8 é treinado e os seus parâmetros estão na Tabela 5.6. O coeficiente de correlação e o desvio padrão dos λ_{em} preditos com o Modelo 8 com respeito aos dados experimentais são $R^2 = 0,75$ e $5,68$ nm, respectivamente. Na Tabela 5.5 estão as previsões dos λ_{em} e os desvios com respeito aos dados experimentais das 19 proteínas usadas para construir o Modelo 8. O desvio das previsões dos λ_{em} para as proteínas Azurina e Toxina colérica B é de ~ 11 nm, sendo estes as piores previsões do Modelo 8. Por outro lado, um desvio de 0.08 nm foi observado para a proteína Tioredoxina W31A. Na Figura 5.5D mostra-se a correlação entre os λ_{em} preditos pelo modelo e os dados experimentais.

A decomposição das interações eletrostáticas em seus componentes baseados nas interações indol - proteína e indol - água demonstra são as interações com as moléculas de água as que determinam os λ_{em} do triptofano. Similarmente, P.Callis demonstrou a importância das moléculas de água para a determinação dos λ_{em} [74]. Por outro lado, os modelos 3, 5 e 7 realizam a decomposição das interações eletrostáticas baseados nas contribuições benzênicas e pirrólicas. Destes modelos é o modelo 3, o qual considera ao benzeno incompleto e pirrol completo, o que apresenta melhor capacidade preditiva ($R^2 = 0.74$, SD = 5.77). A relação dos parâmetros C_1/C_2 é 3.58, sendo assim que o maior peso está relacionado ao benzeno.

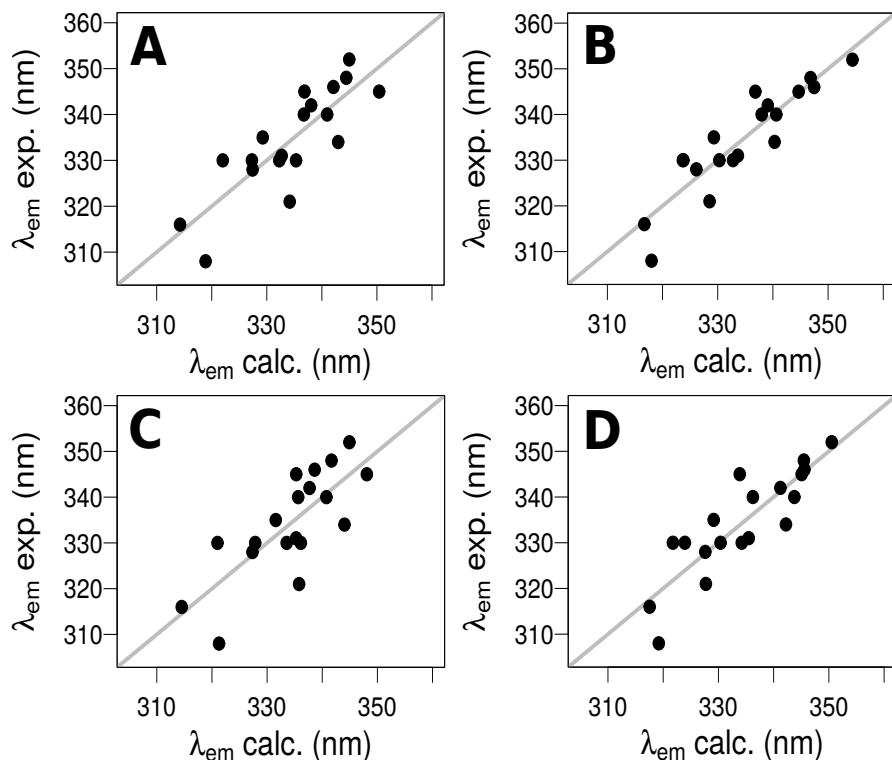


Figura 5.5: Comprimento de onda de emissão de fluorescência experimental (λ_{em}^{exp}) versus o calculado (λ_{em}^{calc}) usando os modelos baseados nas interações eletrostáticas. A) Modelo 5: Baseado nas interações eletrostáticas do benzeno completo - ambiente (água + proteína) e pirrol incompleto - ambiente (água + proteína). B) Modelo 6: Baseado na decomposição das interações eletrostáticas do benzeno completo e pirrol incompleto com a água e a proteína. C) Modelo 7: Baseado nas interações eletrostáticas do benzeno completo - ambiente (água + proteína) e pirrol completo - ambiente (água + proteína). D) Modelo 8: Baseado na decomposição das interações eletrostáticas do benzeno completo e pirrol completo com a água e a proteína.

Tabela 5.5: Comprimentos de onda de emissão de fluorescência experimentais e calculados do triptofano: λ_{em}^{exp} é o comprimento de onda de emissão de fluorescência experimental. $\lambda_{em}^{(5-8)}$ e $\Delta\lambda_{em}^{(5-8)}$ são o comprimento de onda de emissão de fluorescência calculado e o desvio relativo ao valor experimental, respectivamente, usando os modelos baseados nas interações eletrostáticas 5 ao 8. Os comprimentos de onda e os desvios estão em nm.

Proteína	λ_{em}^{exp}	$\lambda_{em}^{(5)}$	$\Delta\lambda_{em}^{(5)}$	$\lambda_{em}^{(6)}$	$\Delta\lambda_{em}^{(6)}$	$\lambda_{em}^{(7)}$	$\Delta\lambda_{em}^{(7)}$	$\lambda_{em}^{(8)}$	$\Delta\lambda_{em}^{(8)}$
Azurina	308[23]	318,88	10,88	318,00	10,00	321,29	13,29	319,21	11,21
Parvalbumina	316[6]	314,24	-1,76	316,71	0,71	314,52	-1,48	317,55	1,55
Mioglobina	321[6]	334,16	13,16	328,52	7,52	335,80	14,80	327,73	6,73
Ribonuclease T1	328[27]	327,42	-0,58	326,13	-1,87	327,34	-0,66	327,62	-0,38
FKBP12	330[29]	322,00	-7,99	323,77	-6,23	321,01	-8,99	321,76	-8,24
T4-lisozima W126Y W138Y	330[31]	335,31	5,31	332,77	2,77	333,55	3,55	334,23	4,23
T4-lisozima W126Y W158Y	330[31]	332,27	2,27	323,68	-6,32	336,11	6,11	323,92	-6,08
T4-lisozima W138Y W158Y	330[31]	327,29	-2,71	330,30	0,29	327,84	-2,16	330,39	0,39
subtilisina Carlsberg	331[33]	332,70	1,70	333,65	2,65	335,26	4,26	335,49	4,49
Nuclease de estafilococo	334[5]	342,96	8,96	340,30	6,30	344,03	10,03	342,23	8,23
Che-Y	335[6]	329,26	-5,74	329,30	-5,69	331,57	-3,43	329,14	-5,86
α -Cobratoxina	340[6]	340,96	0,96	340,62	0,62	340,75	0,75	343,75	3,75
HSA	340[38]	336,73	-3,27	337,97	-2,03	335,63	-4,37	336,24	-3,76
Monelina	342[27]	338,05	-3,95	339,07	-2,93	337,70	-4,29	341,23	-0,77
Toxina colérica B	345[6]	336,84	-8,16	336,82	-8,18	335,28	-9,72	333,87	-11,13
Tioredoxina W31A	345[21]	350,41	5,41	344,66	-0,34	348,09	3,09	345,08	0,08
Melitina	346[6]	342,09	-3,91	347,47	1,47	338,62	-7,38	345,57	-0,43
Fosfolipase A2	348[44]	344,43	-3,57	346,82	-1,18	341,68	-6,32	345,47	-2,53
Glucagon	352[27]	344,97	-7,03	354,42	2,42	344,93	-7,07	350,53	-1,47

Tabela 5.6: Modelos baseados nas interações eletrostáticas para a predição dos comprimentos de onda de emissão de fluorescência (λ_{em}) do triptofano em proteínas por meio de simulações de dinâmica molecular. R^2 é o coeficiente de correlação quadrático de Pearson e SD é o desvio padrão das previsões com respeito aos dados experimentais. Q_T , Q_w , Q_p , Q_{ben-i} , Q_{ben-c} , Q_{pyr-i} , Q_{pyr-c} , $Q_{ben-i,w}$, $Q_{ben-i,p}$, $Q_{pyr-i,w}$, $Q_{pyr-i,p}$, $Q_{ben-c,w}$, $Q_{ben-c,p}$, $Q_{pyr-c,w}$ e $Q_{pyr-c,p}$ são as interações eletrostáticas entre o indol - ambiente (água + proteína), indol - água, indol - proteína, benzeno incompleto - ambiente (água + proteína), benzeno completo - ambiente (água + proteína), pirrol incompleto - ambiente (água + proteína), pirrol completo - ambiente (água + proteína), benzeno incompleto - água, benzeno incompleto - proteína, pirrol incompleto - água, pirrol incompleto - proteína, benzeno completo - água, benzeno completo - proteína, pirrol completo - água e pirrol completo - proteína, respectivamente. As unidades das interações eletrostáticas são $\text{kJ} \times \text{mol}^{-1}$.

Modelo	Descrição	Equação	R^2	SD
1	Eletrostático total	$\lambda_{em} = 311 - 0,884 \langle Q_T \rangle$	0,69	6,36
2	Eletrostático água/proteína	$\lambda_{em} = 311 - 0,998 \langle Q_w \rangle +$ - 0,591 $\langle Q_p \rangle$	0,81	4,89
3	Eletrostático benzeno(i)/pirrol(c)	$\lambda_{em} = 311 - 1,809 \langle Q_{ben,i} \rangle +$ - 0,506 $\langle Q_{pyr,c} \rangle$	0,74	5,77
		$\lambda_{em} = 312 - 1,930 \langle Q_{ben-w,i} \rangle +$ - 1,263 $\langle Q_{ben-p,i} \rangle +$ - 0,602 $\langle Q_{pyr-w,c} \rangle +$ - 0,388 $\langle Q_{pyr-p,c} \rangle$		
4	Eletrostático benzeno(i)/pirrol(c)	$\lambda_{em} = 312 - 1,031 \langle Q_{ben,c} \rangle +$ - 0,854 $\langle Q_{pyr,i} \rangle$	0,85	4,45
5	Eletrostático benzeno(c)/pirrol(i)	$\lambda_{em} = 312 - 1,182 \langle Q_{ben-w,c} \rangle +$ - 0,763 $\langle Q_{ben-p,c} \rangle +$ - 0,958 $\langle Q_{pyr-w,i} \rangle +$ - 0,549 $\langle Q_{pyr-p,i} \rangle$	0,69	6,31
6	Eletrostático benzeno(c)/pirrol(i)	$\lambda_{em} = 315 - 0,529 \langle Q_{ben,c} \rangle +$ - 0,998 $\langle Q_{pyr,c} \rangle$	0,82	4,81
7	Eletrostático benzeno(c)/pirrol(c)	$\lambda_{em} = 315 - 0,719 \langle Q_{ben-w,c} \rangle +$ - 0,608 $\langle Q_{ben-p,c} \rangle +$ - 1,113 $\langle Q_{pyr-w,c} \rangle +$ - 0,595 $\langle Q_{pyr-p,c} \rangle$	0,59	7,27
8	Eletrostático benzeno(c)/pirrol(c)		0,75	5,68

5.3.3 Robustez dos modelos

Todos os modelos desenvolvidos na subseção anterior foram parametrizados usando um conjunto de 19 proteínas. Com a finalidade de conhecer a capacidade preditiva dos modelos com respeito a um dado desconhecido, ou simplesmente um dado não usado na construção do modelo, é necessário de realizar testes de robustez. Na Subseção 4.3.3 esses testes foram realizados para os modelos parametrizados baseados na SASA do indol. Nesta subseção analisaremos quanto sensível é a capacidade preditiva dos modelos clássicos parametrizados baseados nas interações eletrostáticas em função dos dados de treinamento.

A robustez do primeiro modelo baseado nas interações eletrostáticas (Modelo 1) é analisada a continuação. Os desvios, desvios médios e desvios máximos quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento estão disponíveis na Tabela 5.7. Os desvios e desvios médios são menores a 3 nm. Os maiores desvios correspondem às proteínas que possuem λ_{em} baixos, como são a Azurina e a Parvalbumina, sendo de no máximo ~ 6 nm. Estes resultados indicam que os desvios das predições com respeito aos dados experimentais observados para o Modelo 1 baseado nas interações eletrostáticas não estão associados ao conjunto de dados usados, senão que é próprio da limitação do modelo.

A análise de robustez também são aplicadas aos demais modelos baseados nas interações eletrostáticas (Modelos 2 ao 8). A continuação mostra-se um resumo desses resultados.

- **Modelo 2:** Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 3 nm. Os maiores desvios são de ~ 7 (Parvalbumina, T4-lisozima W126Y W158Y e Melitina) e 10 nm (Glucagon). Excetuando essas proteínas, os maiores desvios são menores que 6 nm. Na Tabela C.1 encontram-se os desvios, desvios médios e desvios máximos da análise de robustez do Modelo 2.
- **Modelo 3:** Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 3 nm. Os maiores desvios estão entre ~ 7 e 9 nm (HSA, T4-lisozima W126Y W158Y, Parvalbumina). Excetuando essas proteínas, os maiores desvios foram menores que 6 nm. Na Tabela C.2 encontram-se os desvios, desvios médios e desvios máximos da análise de robustez do

Modelo 3.

- Modelo 4: Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 6 nm. Os maiores desvios estão entre ~ 9 e 13 nm (T4-lisozima W126Y W158Y e Glucagon). Na Tabela C.3 mostra-se os desvios, desvios médios e desvios máximos da análise de robustez do Modelo 4.
- Modelo 5: Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 3 nm. Os maiores desvios são de ~ 7 e 11 nm (Parvalbumina e HSA). Excetuando essas proteínas, os maiores desvios foram menores que 6 nm. Na Tabela C.4 mostra-se os desvios, desvios médios e desvios máximos da análise de robustez do Modelo 5.
- Modelo 6: Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 4 nm. O maior desvio é de ~ 16 nm e corresponde ao Glucagon. Na Tabela C.5 mostra-se os desvios, desvios médios e desvios máximos da análise de robustez do Modelo 6.
- Modelo 7: Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 5 nm. Os maiores desvios são de ~ 9 e 14 nm, os quais correspondem às proteínas Parvalbumina e HSA, respectivamente. Excetuando essas proteínas, os maiores desvios foram menores que 7 nm. Na Tabela C.6 encontram-se os desvios, desvios médios e desvios máximos da análise de robustez do Modelo 7.
- Modelo 8: Os desvios e desvios médios quando são removidos uma ($k = 1$), duas ($k = 2$) ou três ($k = 3$) proteínas dos dados de treinamento são menores que 6 nm. Os maiores desvios estão entre ~ 9 e 12 nm (Parvalbumina, Glucagon e T4-lisozima W126Y W158Y). Na Tabela C.7 encontram-se os desvios, desvios médios e desvios máximos da análise de robustez do Modelo 8.

Tabela 5.7: Robustez do Modelo 1 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 1 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	318,90	2,73	2,74(3,76)	2,73(5,00)
Parvalbumina	313,86	-0,97	-0,95(-4,24)	-0,93(-6,62)
Mioglobina	333,87	0,72	0,72(1,59)	0,73(2,10)
Ribonuclease T1	326,80	-0,11	-0,11(1,76)	-0,11(2,79)
FKBP12	322,12	-1,34	-1,35(-2,89)	-1,35(-4,43)
T4-lisozima W126Y W138Y	336,38	0,37	0,38(1,14)	0,38(1,92)
T4-lisozima W126Y W158Y	332,36	0,14	0,14(1,19)	0,14(2,09)
T4-lisozima W138Y W158Y	328,02	-0,17	-0,17(1,49)	-0,17(-2,50)
subtilisina Carlsberg	332,88	0,11	0,11(1,08)	0,11(1,98)
Nuclease de estafilococo	342,66	0,93	0,94(2,17)	0,94(3,16)
Che-Y	328,70	-0,49	-0,49(-1,52)	-0,49(-2,88)
α -Cobratoxina	340,09	0,01	0,02(0,88)	0,02(2,04)
HSA	338,91	-0,08	-0,07(-0,88)	-0,07(1,71)
Monelina	338,17	-0,25	-0,25(-1,03)	-0,25(-1,72)
Toxina colérica B	336,44	-0,49	-0,49(-1,18)	-0,49(-1,72)
Tioredoxina W31A	350,07	1,34	1,34(3,20)	1,34(4,46)
Melitina	343,00	-0,34	-0,34(-1,47)	-0,33(-2,35)
Fosfolipase A2	343,84	-0,51	-0,51(-1,73)	-0,50(-2,60)
Glucagon	343,92	-1,01	-1,00(-1,74)	-0,99(-2,61)

5.4 Conclusões

A variabilidade dos λ_{em} do triptofano foi investigada por Vivian e Callis (2004), os quais usando cálculos de estrutura eletrônica, determinaram que a introdução das cargas do ambiente (água + proteína) nos cálculos das transições eletrônicas do indol são necessárias para computar corretamente os λ_{em} . Esse modelo teve um bom desempenho, sendo o desvio padrão das previsões com respeito aos dados experimentais de 6,87 nm, e o quadrado do coeficiente de correlação de Pearson obtido foi de $R^2 = 0,69$. No entanto, a aplicação desse método requer conhecimento e experiência para a determinação correta das transições eletrônicas responsáveis pela fluorescência. Adicionalmente, os cálculos de estrutura eletrônica requerem uso de programas específicos e recursos computacionais que possam garantir o funcionamento desses

programas. Essas limitações fazem com que este método não possa ser usado amplamente pela comunidade experimental.

As cargas dos átomos na vizinhança do triptofano, como foi apontado por Vivian e Callis (2004), jogam um papel fundamental na estabilidade do estado excitado L_a , e portanto na determinação dos λ_{em} do triptofano. As interações eletrostáticas são facilmente computadas a partir das simulações de dinâmica molecular, assim podem ser usadas no desenvolvimento de modelos clássicos parametrizados. O primeiro modelo, baseado nas interações eletrostáticas entre o indol e o ambiente (água + proteína), teve um desempenho com um SD = 6,36 nm e R^2 = 0,69, portanto teve uma capacidade preditiva similar à do modelo desenvolvido por Vivian e Callis (2004). Esse resultado é satisfatório porque com um modelo clássico conseguimos obter estimativas de λ_{em} tão boas quanto as feitas com os métodos baseados em primeiros princípios. Neste primeiro modelo, os pesos das contribuições das cargas dos átomos da água e dos aminoácidos são iguais. Devido a que o comportamento das moléculas de água e proteína são diferentes durante o processo de relaxamento do estado excitado, é necessário tratar essas duas componentes de maneira diferente. Assim, o segundo modelo baseado nas interações eletrostáticas propõe diferentes pesos para as interações eletrostáticas entre o indol - água e indol - proteína. O desempenho deste modelo foi de SD = 4,89 nm e R^2 = 0,81; sendo superior ao desempenho do modelo desenvolvido por Vivian e Callis (2004). Este modelo clássico, com apenas três parâmetros, conseguiu ter uma ótima capacidade preditiva dos λ_{em} do triptofano em proteínas. Adicionalmente seis modelos alternativos baseados nas interações eletrostáticas foram desenvolvidos. Estes últimos modelos tiveram entre 3 e 5 parâmetros, sendo o melhor desempenho associado a um SD = 4,45 nm e R^2 = 0,85.

Com os modelos parametrizados apresentados e desenvolvidos neste trabalho, é possível de computar os comprimentos de onda de emissão de fluorescência do triptofano em proteínas a partir de simulações de dinâmica molecular. Devido à simplicidade dos modelos e do seu uso, estes resultados apresentam-se como uma metodologia rápida e prática, que pode auxiliar na interpretação dos dados experimentais.

Capítulo 6

Simulações de dinâmica molecular da anisotropia de fluorescência resolvida no tempo da subtilisina Carlsberg

6.1 Introdução

6.1.1 Estrutura e função da enzima subtilisina Carlsberg

A subtilisina Carlsberg (sC), também conhecida como subtilisina A, subtilopeptidase A ou alcalase, pertence ao grupo de enzimas conhecidas como subtilases [86]. A maioria destas enzimas são sintetizadas como pré-pró-enzimas, as quais são transportadas extracelularmente via peptídeo sinal, para logo serem clivadas e formarem o peptídeo maduro. O tamanho do peptídeo maduro varia de ~226 a 1775 aminoácidos [87]. Este grupo de enzimas é caracterizado pela sua função de serina-proteases, e pela presença de um sítio catalítico composto dos aminoácidos de histidina (His), asparagina (Asp) e serina (Ser) [88]. Particularmente, a sC está composta de 274 resíduos [89].

A estrutura tridimensional da sC foi determinada a 2,5 Å de resolução e encontra-se depositada na base de dados do *Protein Data Bank* com o código de acesso 1SBC [32]. Na sua estrutura secundária possui elementos de α hélices, folhas β e uma tríade catalítica composta dos aminoácidos Asp32, His64 e Ser221 (Figura 6.1).

As características da subtilisina Carlsberg, como a sua alta estabilidade em diversos

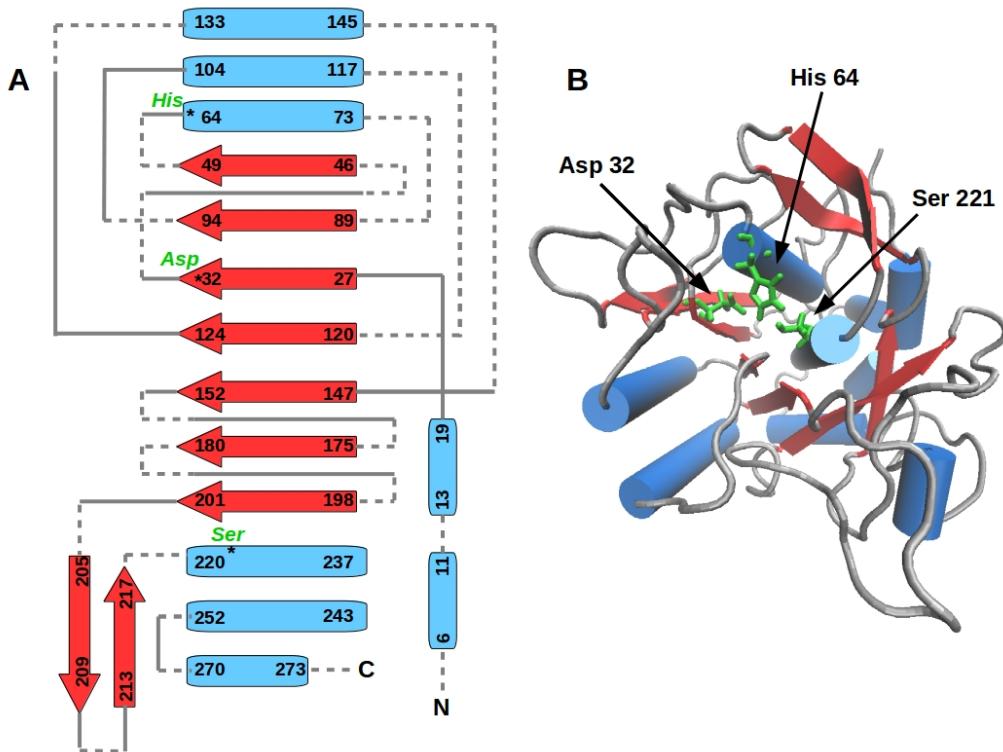


Figura 6.1: A) Representação esquemática da estrutura secundária da enzima subtilisina Carlsberg (sC). As α -hélices estão representadas por cilindros azuis, as folhas- β como flechas vermelhas. As regiões conservadas (linhas contínuas em cinza) e variáveis (linhas descontínuas em cinza) estão mostradas de acordo ao reportado por Siezen, *et al.* para a família das subtilases [86]. B) Estrutura cristalográfica da enzima subtilisina Carlsberg (PDB id. 1SBC) mostrando as α -hélices (cilindros azuis) e folhas- β (flechas vermelhas), assim como a tríade catalítica, Asp32, His64 e Ser221 (verde).

solventes, condições de pH e na presença de detergentes, assim como sua capacidade de reconhecer diversos substratos, fazem que seja amplamente usada na indústria. Entre as aplicações, por citar algumas, temos produtos limpadores de lentes de contato, produtos de limpeza de superfícies, tratamento de água, suplementos digestivos e processamento de alimentos [90].

6.1.2 Fluorescência da subtilisina Carlsberg

A subtilisina Carlsberg possui na sua estrutura um único resíduo de triptofano, o qual está localizado na posição 113 (Trp113) [32]. A presença desse resíduo faz com que os estudos por fluorescência sejam possíveis. Os espectros de fluorescência do Trp113 da sC foram reportados por diversos autores, sendo que o comprimento de onda máximo de emissão de fluorescência (λ_{em}) mostrou-se bastante variável:

- 322 nm, segundo Willis e Szabo (1989) [17],
- 331 nm, segundo Bayley et al. (1989) [33],
- 351 nm, segundo Pal et al. (2002) [91],
- 355 nm, segundo Lakshmikanth e Krishnamoorthy (1999) [92] e
- 360 nm, segundo Shaw e Pal (2007) [93].

Essa variabilidade observada pode ser entendida como sendo produto de conformações preferenciais da proteína nas proximidades do Trp113, as quais estariam sendo amostradas de forma distinta em cada condição experimental específica. Assim, baixos valores de λ_{em} correspondem a conformações nas quais o Trp113 da sC está localizado preferencialmente em um bolsão hidrofóbico, enquanto que maiores valores de λ_{em} indicam uma exposição completa do Trp113 ao solvente [74, 94]. Agora vamos a discutir, brevemente, sobre os experimentos de fluorescência do Trp da sC obtido por Willis e Szabo (1989) [17], Bayley et al. (1989) [33], Pal et al. (2002) [91], Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93].

Tabela 6.1: Condições experimentais das medidas de fluorescência do Trp113 da subtilisina Carlsberg. A proteína sC foi adquirida comercialmente e passos de purificação foram realizados antes de realizar as medidas de fluorescência. HPLC (Cromatografia líquida de alta eficiência). FPLC (Cromatografia líquida de proteína rápida). λ_{ex} e λ_{em} são os comprimentos de onda de excitação e emissão, respectivamente.

Autor	Purificação	Solução	λ_{ex} (nm)	λ_{em} (nm)
Willis e Szabo (1989) [17]	HPLC	< 50 μ M de sC em	295	311
		10mM tampão fosfato, pH 6.2, 63 mM NaCl		
Bayley et al. (1989) [33]	FPLC	10 μ M de sC em 100mM tampão fosfato, pH 7.0	300	331
Pal et al. (2002) [91]	—	400 μ M de sC em 100mM tampão fosfato, pH 7.0	296	351
Lakshmikanth e Krishnamoorthy (1999) [92]	—	20-200 μ M de sC em 20mM tampão fosfato, pH 7.0	295	355
Shaw e Pal (2007) [93]	—	100mM buffer fosfato, pH 7.0	299	360

Segundo Willis e Szabo (1989) [17], a sC é uma proteína que possui um triptofano com rendimento quântico baixo, devido a isso as tirosinas teriam uma maior contribuição nos

espectros de fluorescência quando os comprimentos de onda de excitação (λ_{ex}) são menores que 300 nm. Na Tabela 6.1 observa-se como varia o λ_{em} em função do comprimento de onda de excitação. Uma excitação com $\lambda_{ex} = 300$ nm garante que o espectro de fluorescência tenha unicamente a contribuição do triptofano [17]. Outro fato adicional, que Willis e Szabo (1989) reportaram, foi a presença de autólise da sC nas seguintes condições:

- Quando a proteína após purificada é armazenada por um período maior a 24 horas, e
- em soluções com concentrações de sC maiores a $50\mu\text{M}$.

Essa autólise seria a responsável pelo aumento do comprimento de onda de emissão, passando de 322 nm para 340 nm [17].

Bayley et al. (1989) também alertaram para a possibilidade de autólise da sC, sendo que concentrações de $10\mu\text{M}$ da enzima mostraram pequenas mudanças na fluorescência num período de 20 horas a 20°C [33]. Adicionalmente, analisando espectros de dicroísmo circular (CD), Bayley et al. (1989), determinaram que concentrações elevadas de sC ($\sim 100\mu\text{M}$) provocam autólise da proteína, fazendo com que 30-35% da intensidade de CD seja perdida num intervalo de tempo de 5 horas.

Os outros três trabalhos, Pal et al. (2002) [91], Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93], reportaram valores elevados do λ_{em} do Trp113 da sC, os quais indicariam que o triptofano encontra-se exposto ao solvente (Tabela 6.1). Estes autores usaram a proteína sC dissolvendo-a na solução tampão diretamente sem passar por nenhum passo de purificação prévio. Pal et al. (2002) prepararam soluções estoques de $400\mu\text{M}$ e usaram comprimentos de onda de excitação de 296 nm [91]. Portanto, é possível que as medidas de fluorescência reportadas por Pal et al. (2002) tenham sido afetadas pela autólise da enzima, além da forte contribuição das tirosinas ao espectro de fluorescência. Nos experimentos de Lakshmikanth e Krishnamoorthy (1999), a proteína foi estabilizada na solução tampão por várias horas antes do início das medidas de fluorescência, indicando também uma possível autólise da sC. Por outro lado, Shaw e Pal (2007) [93] realizaram experimentos de dicroísmo circular e teste de atividade catalítica da enzima sC demonstrando que a proteína tem estrutura secundária e é ativa. Isso faz pensar que a estrutura da proteína está preservada, ao menos em certo grau, e que a emissão elevada do Trp113 é devida a uma conformação preferencial

amostrada no experimento e não à presença de autólise. Adicionalmente, Shaw e Pal (2007) [93], usaram um λ_{ex} de 299 nm, o qual excitará especificamente o resíduo de triptofano.

Na Tabela 6.1 mostram-se resumidamente as condições experimentais, assim como os comprimentos de onda de emissão reportados para a sC.

6.1.3 Anisotropia de fluorescência resolvida no tempo da subtilisina Carlsberg

A dinâmica reorientacional do resíduo Trp113 da enzima subtilisina Carlsberg (SC) foi reportada por dois grupos de pesquisa, de maneira independente, sendo o primeiro deles Lakshmikanth e Krishnamoorthy (1999) [92] e ou segundo Shaw e Pal (2007) [93]. A continuação vamos descrever, brevemente, os resultados desses experimentos. Maiores detalhes, sobre esses resultados experimentais serão fornecidos na Seção 6.3.

Como foi mencionado anteriormente, na Subseção 6.1.2, ambos grupos reportaram valores de λ_{em} entre 355 - 360 nm para o espectro de fluorescência do Trp113 da sC. Esse valor de λ_{em} foi usado como comprimento de onda fixo para seguir a dinâmica reorientacional do Trp113 da sC. Apesar de ambos grupos terem reportado o decaimento de anisotropia como um decaimento bi-exponencial, a forma dos decaimentos foram muito diferentes.

Segundo Lakshmikanth e Krishnamoorthy (1999) [92] os tempos de reorientação da componente rápida e lenta do decaimento são de 0,17 e 3,50 ns, respectivamente. A contribuição da componente rápida no decaimento de anisotropia correspondeu a 90%, motivo pelo qual Lakshmikanth e Krishnamoorthy (1999) observam uma despolarização completa da fluorescência nos primeiros nanosegundos.

Por outro lado, Shaw e Pal (2007) [93] reportaram um decaimento rápido de 0,469 ns, enquanto que o decaimento lento não é bem resolvido (em torno de 10-100 ns). A despolarização da fluorescência não chega a ser completa na escala de tempo do experimento.

Essa divergência entre os dados experimentais gera confusão e dúvidas sobre o que realmente está acontecendo o ponto de vista molecular. Com a finalidade de ajudar na compreensão da dinâmica reorientacional do Trp113 da sC, simulações de dinâmica molecular foram

realizadas. Com as trajetórias geradas, decaimentos de anisotropia de fluorescência foram calculados e comparações aos dados experimentais foram possíveis.

6.2 Metodologia

6.2.1 Simulações de dinâmica molecular

A estrutura cristalográfica da enzima subtilisina Carlsberg (sC) encontra-se depositada na base de dados PDB, com o código de acesso 1SBC [32]. As águas presentes na estrutura cristalográfica foram mantidas. Esta estrutura será considerada como ponto de partida das simulações de dinâmica molecular da sC com o Trp113 parcialmente protegido ao solvente. A solvatação da sC foi realizada com o software Packmol [54, 55], considerando como modelo de água o TIP3P [75]. Íons de sódio e cloro foram adicionados ao sistema com a finalidade de deixá-los neutros. A proteína, água e íons foram simulados usando o campo de força CHARMM22 [65, 66, 67]. Condições periódicas de contorno foram usadas nas simulações.

Todas as simulações foram executadas nas seguintes condições: ensemble NPT a 1 atm de pressão e 298.15 K de temperatura, com um passo de tempo de 2 fs. Para controlar a temperatura e a pressão, foi acoplado um banho de Langevin com coeficiente de fricção de 10 ps^{-1} e o algoritmo de Nosé-Hoover com período de oscilação do pistão de 200 fs e taxa de decaimento de 100 fs. O método de *Particle Mesh Ewald Sum* foi usado para calcular as interações eletrostáticas de longo alcance.

Os sistemas foram equilibrados como segue:

- 1000 passos de minimização usando o método de gradientes conjugados, seguidos de uma simulação dinâmica molecular de 200 ps considerando todos os átomos da proteína fixos;
- 500 passos de minimização usando o método de gradientes conjugados, seguidos de uma simulação dinâmica molecular de 200 ps considerando os átomos C α da proteína fixos; e
- 2 ns de uma simulação dinâmica molecular sem nenhuma restrição.

O primeiro passo da equilibração permite relaxar apenas o solvente em torno da proteína. O segundo passo, permitir relaxar adicionalmente as cadeias laterais da proteína.

No último passo, o sistema inteiro (água + proteína) são relaxados. As estruturas finais do processo de equilibração foram consideradas como as configurações de início das simulações de dinâmica molecular de produção, isto é a partir das quais as análises foram feitas. O software usado para as simulações de dinâmica molecular foi o NAMD [76]. O tempo de produção das simulações foi de 20 ns.

A estrutura subtilisina Carlsberg com o Trp113 completamente exposto ao solvente foi obtida pela realização de uma simulação de 10 ns da proteína cristalográfica no vácuo, mantendo todos os átomos fixos exceto os do resíduo Trp113, a uma temperatura de 815K. Esta simulação amostrou de maneira essencialmente aleatória as conformações da cadeia lateral do Trp, enquanto que a proteína como um todo permaneceu estática. Uma estrutura com o resíduo Trp113 exposto ao solvente foi selecionada como estrutura de inicio para as simulações de dinâmica molecular. Essas simulações amostraram, inevitavelmente após os primeiros ns de simulação, conformações com o Trp113 parcialmente protegido. Isso é devido a que a estabilidade da conformação da sC com o Trp113 parcialmente protegido é maior comparada à estabilidade da conformação com o Trp113 exposto ao solvente. Para manter o Trp113 da sC exposto ao solvente durante todo o tempo das simulações, foi necessário definir uma restrição conformacional. A distância entre os átomos Ala29-C β e Trp113-CZ3 foi restringida, de forma a manter uma distância mínima de 12 Å. A implementação das simulações com restrições conformacionais serão feitos usando as simulações ABF, o qual será explicado na Subseção seguinte (Subseção 6.2.2).

6.2.2 Simulações ABF

A metodologia ABF (*Adaptive Biasing Forces molecular dynamics*) é uma técnica que permite explorar regiões de difícil acesso da superfície de energia potencial. Esta metodologia está disponível no programa NAMD [76]. A metodologia permite a determinação de perfis de energia livre ao longo de coordenadas de reação definidas pelo investigador. No entanto, no nosso caso, apenas usamos sua implementação para introduzir as restrições que mantém a cadeia lateral do Trp exposta ao solvente. Nas simulações ABF, a coordenada de reação introduzida consistiu nas distâncias mínima e máxima entre o átomo C β do resíduo Ala29 e o átomo CZ3 do Trp113 como sendo de 12 Å e 100 Å, respectivamente. Com a finalidade

de garantir que os limites mínimos e máximos sejam respeitados, nas simulações ABF foram definidos potenciais semi-harmônicos constantes com uma constante de força de 50 kcal mol⁻¹ Å⁻². Normalmente numa simulação ABF uma força externa é aplicada na coordenada de reação com a finalidade de facilitar a amostragem. No nosso caso, nenhuma força foi introduzida nas simulações ABF, devido a que o tempo de amostragem antes de começar aplicar a força foi muito maior ao tempo total da simulação, ja que nossa intenção é de observar a dinâmica do Trp113 da sC exposto ao solvente determinado unicamente pelas interações deste com o sistema. Os parâmetros das simulações ABF foram similares aos descritos anteriormente para as simulações de dinâmica molecular.

6.2.3 Cálculos do λ_{em}

Os λ_{em} do Trp113 da sC foram calculados a partir das simulações de dinâmica molecular usando os modelos parametrizados desenvolvidos nos Capítulos 4 e 5. Os modelos baseados na SASA do benzeno/pirrol (Modelo 2 da Tabela 4.3) e nas interações eletrostáticas do indol com a água/proteína (Modelo 2 da Tabela 5.6) foram usados. O cômputo da área acessível ao solvente (SASA) do benzeno incompleto e do pirrol completo foram realizados no programa VMD [77]. As interações eletrostáticas entre o indol e a água/proteína foram computados com programa MDAnalysis [85].

6.2.4 Cálculos dos decaimentos da anisotropia de fluorescência

Como foi explicado no Capítulo 1, a dependência temporal do decaimento da anisotropia de fluorescência pode ser escrita em termos da função de correlação temporal entre os momentos e dipolo de absorção e emissão do fluoróforo (Equação 1.3.9). Isto faz com que seja possível reproduzir os decaimentos de anisotropia de fluorescência experimentais usando as simulações de dinâmica molecular totalmente clássicas.

O termo que aparece no lado direito da Equação 1.3.9 pode ser estimado a partir

da média temporal calculada ao longo da simulação para intervalos de tempo t_m .

$$\langle P_2[\mu_a(0) \cdot \mu_e(t_m)] \rangle \approx \frac{1}{N-m} \sum_{n=1}^{N-m} P_2[\mu_a(t_n) \cdot \mu_e(t_n + t_m)] \quad (6.2.1)$$

onde μ_a e μ_e , definidos para os instantes de tempo t_n e $t_n + t_m$, respectivamente, são calculados a partir das trajetórias de dinâmica molecular. N representa o número total de estruturas amostradas na simulação de dinâmica molecular. Os termos t_n e t_m correspondem aos tempos da n -ésima estrutura e a intervalo de tempo após m estruturas, dentro do conjunto de estruturas amostradas por uma única simulação de dinâmica molecular. O momento de dipolo de absorção do indol usado no presente trabalho corresponde ao do estado L_a [49]. O vector do estado L_a foi definido como o vector que tem como início o ponto médio entre os carbonos C δ 2 e C ξ 2, e fim o nitrogênio da grupo indol do triptofano. O ângulo formado entre os dipolos de absorção e emissão é dado pela Equação 1.3.5. O valor experimental de r_0 foi extraído dos experimentos de Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93].

6.3 Resultados e discussões

6.3.1 Decaimento da anisotropia do triptofano livre em água

Nos experimentos de anisotropia de fluorescência resolvida no tempo, a emissão é medida em um comprimento de onda previamente definido, que geralmente corresponde ao máximo de intensidade do espectro de fluorescência (λ_{em}). Isso permite observar a dinâmica reorientacional da conformação de fluoróforos de maior população ou de maior rendimento quântico. Naturalmente, esta escolha é feita porque aumenta a sensibilidade do experimento.

O comprimento de onda de emissão usado para registrar experimentalmente os decaimentos de anisotropia de fluorescência do triptofano livre em água foi de 360 nm [92]. Esse valor correspondeu ao λ_{em} do espectros de fluorescência do Trp livre em água reportado por Lakshmikanth e Krishnamoorthy (1999) (Figura 6.2A) [92]. Resultados similares foram observados por outros autores [95, 96].

Os λ_{em} do Trp podem ser calculados a partir das simulações de dinâmica molecular, graças ao desenvolvimento dos modelos parametrizados descritos nos Capítulos 4 e 5. O λ_{em} do

Trp livre em água foi computado para cada passo das simulações de dinâmica molecular usando os modelos parametrizados baseados na SASA do benzeno/pirrol e nas interações eletrostáticas do indol com a água/proteína. A distribuição dos λ_{em} de cada passo é mostrada na Figura 6.2A e B. O valor máximo dessas distribuições corresponde ao λ_{em} calculado. Os λ_{em} calculados foram de 364 e 342 nm para os modelos baseados na SASA do benzeno/pirrol e nas interações eletrostáticas do indol com a água/proteína, respectivamente (Tabela 6.2). Os valores são muito próximos dos valores usados no experimento de anisotropia de fluorescência resolvida no tempo, e portanto podemos dizer que as simulações de dinâmica molecular do Trp em água estão amostrando as conformações usadas no experimento.

O decaimento da anisotropia de fluorescência do Trp livre em água é de natureza mono-exponencial, ou seja não apresenta heterogeneidade, com tempos de reorientação da ordem dos 46 [97] ou 60 ps [92], dependendo das condições experimentais. O decaimento de anisotropia experimental do Trp livre em água reportado por Lakshmikanth e Krishnamoorthy (1999) [92], é mostrado na Figura 6.2C.

A partir das simulações de dinâmica molecular do Trp livre em água, calculamos os decaimentos de anisotropia de fluorescência (Figura 6.2C). O decaimento calculado do Trp em água foi ligeiramente mais rápido que o experimental. Segundo o modelo de decaimento mono-exponencial, o tempo de reorientação calculado do Trp em água é de 30 ps. Esse resultado é próximo aos reportados pelos experimentos, e portanto valida nossa metodologia para calcular os decaimentos de anisotropia por meio de simulações de dinâmica molecular. Os decaimentos calculados por dinâmica molecular são geralmente mais rápidos comparados aos experimentais devido a que o modelo de água TIP3P usado nas simulações tem um comportamento difusivo acelerado [51, 75, 98]

6.3.2 Decaimento da anisotropia de fluorescência do Trp113 da subtilisina Carlsberg parcialmente protegido

A proteína sC possui um único resíduo de triptofano na sua estrutura, e isso faz com que seja um alvo a ser estudado pelas técnicas de fluorescência. O espectro de fluorescência experimental do Trp113 da sC reportado por Lakshmikanth e Krishnamoorthy (1999) [92] e

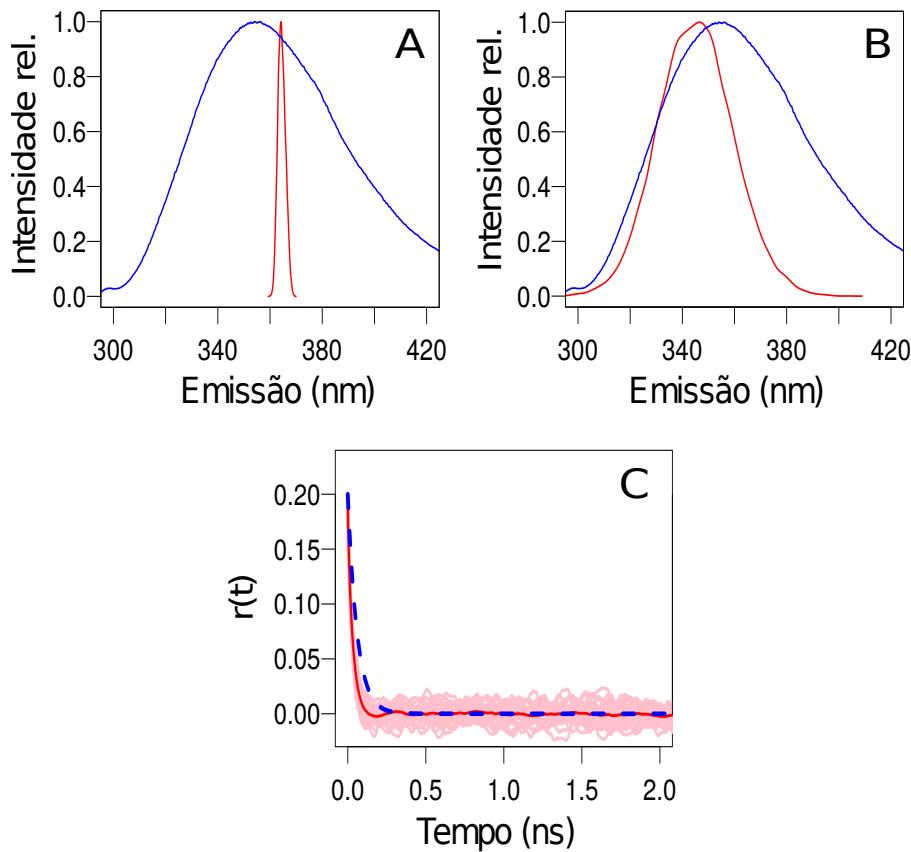


Figura 6.2: Fluorescência do Trp livre em água. As distribuições dos λ_{em} do Trp em água foram calculados segundo os modelos parametrizados baseados na A) SASA do benzeno/pirrol e B) interação eletrostática do indol com o sistema (água + proteína). C) Anisotropia de fluorescência resolvida no tempo do Trp livre em água Os dados experimentais e calculados estão desenhados em azul e vermelho, respectivamente.

Shaw e Pal (2007) [93] determinaram o λ_{em} sendo de 355 e 360 nm, respectivamente (Figura 6.4 e Tabela 6.3). O valor do λ_{em} indica que, na conformação mais provável, o Trp113 encontra-se completamente exposto ao solvente.

Observando a estrutura cristalográfica da sC, nota-se que o Trp113 encontra-se parcialmente protegido do solvente (Figura 6.3A e B) [32]. A distância entre o átomo CZ3 do Trp113 e o átomo C β da Ala29 é de ~ 5 Å na estrutura cristalográfica. Simulações de dinâmica molecular tomando como ponto de partida a estrutura cristalográfica só foram capazes de amostrar as conformações do Trp113 parcialmente protegido do solvente, e durante todo o tempo das simulações as distâncias entre os átomos Ala29-C β e Trp113-CZ3 se mantiveram em torno de ~ 5 Å. Essas simulações foram usadas para os cálculos dos λ_{em} e da dinâmica reorientacional do Trp113 em sua conformação cristalográfica.

Tabela 6.2: Parâmetros experimentais e calculados da anisotropia de fluorescência resolvida no tempo do triptofano livre em água. λ_{em} é o comprimento de onda usado para registrar os decaimentos de anisotropia. O λ_{em} foi calculado usando ^(a) o modelo baseado na SASA do benzeno/pirrol e ^(b) nas interações eletrostáticas do indol com a água/proteína. O modelo mono-exponencial usado para o ajuste do decaimento da anisotropia de fluorescência é: $r(t) = r_0 \times e^{(-t/\tau_1)}$, sendo τ_1 o tempo de reorientação do fluoróforo. Os λ_{em} calculados correspondem a média dos comprimentos de onda calculados a partir de 30 simulações de dinâmica molecular. O ajuste exponencial da anisotropia foi aplicado à média das 30 curvas de decaimento de anisotropia de fluorescência calculadas a partir das simulações de dinâmica molecular.

Molécula	λ_{em} (nm)	r_0	τ_1 (ns)
Experimental	360 [92]	0,20	0,06
	340 [97]	0,19	0,046
Calculado	$364,44 \pm 0,30$ ^a	0,19	0,03
	$343,68 \pm 2,03$ ^b		

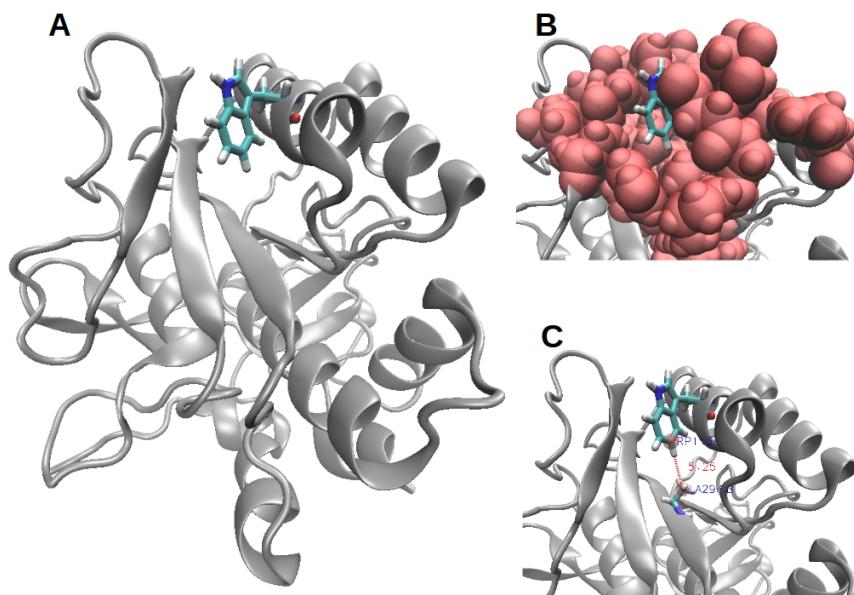


Figura 6.3: A) Conformação da enzima subtilisina Carlsberg com o Trp113 parcialmente protegido do solvente. B) Representação da acessibilidade ao solvente do resíduo Trp113. C) Representação da distância característica entre os átomos Ala29-C β e Trp113-CZ3 na conformação da sC com o Trp113 parcialmente protegido do solvente ($\sim 5 \text{ \AA}$).

Usando os modelos parametrizados baseados na SASA do benzeno/pirrol e nas interações eletrostáticas do indol com a água/proteína foram computados os λ_{em} instantâneos para cada estrutura gerada nas trajetórias da sC, e suas distribuições são mostrados na Figura 6.4A e B, respectivamente. Observa-se uma dissimilaridade entre os λ_{em} calculados e experimentais (Tabela 6.3). Os valores de λ_{em} do Trp113 da sC calculados usando os modelos baseados na

SASA do benzeno/pirrol e nas interações eletrostáticas do indol com água/proteína foram de 332,93 nm e 331,35 nm, respectivamente. Esses valores são ~ 25 nm menores com respeito aos λ_{em} experimentais (Tabela 6.3).

Vivian e Callis (2004), usando também a estrutura cristalográfica da sC como ponto de partida das simulações híbridas QM-MM, determinaram o λ_{em} sendo de 337 nm [74]. Portanto, podemos dizer que as simulações, usando as estruturas cristalográficas como ponto de partida, estão amostrando conformações do Trp113 diferentes das amostradas no experimento.

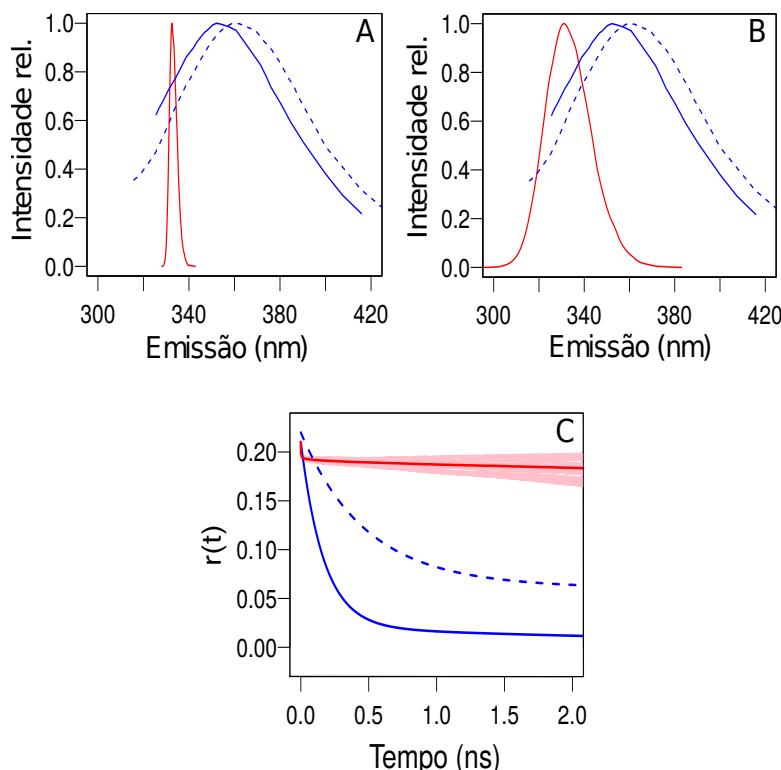


Figura 6.4: Fluorescência da conformação da subtilisina Carlsberg com o resíduo Trp113 parcialmente protegido do solvente. A) e B) Espectros de fluorescência experimentais (azul) e distribuições dos λ_{em} calculados (vermelho) do Trp113 da sC em água. Os cálculos dos λ_{em} foram feitos usando o modelo parametrizado baseado na SASA do benzeno/pirrol (em A) ou o modelo baseado nas interações eletrostáticas do indol com a água/proteína (em B). C) Anisotropia de fluorescência resolvida no tempo do Trp113 da sC em água. As curvas azuis contínuas e tracejadas correspondem aos dados experimentais reportados por Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93], respectivamente.

Os decaimentos de anisotropia de fluorescência do Trp113 da sC foram computados a partir das simulações tomando como ponto de partida a sua estrutura cristalográfica. O objetivo destes cálculos é comparar esses decaimentos com os dados experimentais reportados por Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93]. Segundo, Lakshmikanth

e Krishnamoorthy (1999), o decaimento da anisotropia do Trp113 da sC está caracterizado por um comportamento bi-exponencial, com tempos de reorientação de 0,17 e 3,50 ns para a componente rápida e lenta, respectivamente (Tabela 6.3) [92]. Por outro lado, Shaw e Pal (2007), reporta tempos de reorientação de 0,469 ns para a componente rápida e em torno de 10-100 ns para a componente lenta (Tabela 6.3) [93]. Os decaimentos experimentais estão mostrados na Figura 6.4C.

A anisotropia de fluorescência resolvida no tempo calculada a partir das simulações de dinâmica molecular da conformação com o Trp113 parcialmente protegido do solvente da sC é mostrada na Figura 6.4C. A comparação visual com respeito aos dados experimentais, nos indica que os decaimentos calculados foram muito lentos comparados aos decaimentos experimentais. Essa discordância entre dados experimentais e calculados pode ser explicada pela diferença de conformações que estariam sendo amostradas no experimento e nas simulações. As conformações amostradas nos experimentos de anisotropia de fluorescência correspondem a um λ_{em} entre 355 e 360 nm, enquanto que as conformações amostradas nas simulações de dinâmica molecular correspondem a um $\lambda_{em} \approx 331$ nm.

De acordo com a acessibilidade ao solvente do Trp, podemos estabelecer o seguinte: As conformações do Trp113 da sC amostradas nos experimentos de anisotropia de fluorescência resolvida no tempo correspondem a conformações onde o triptofano está completamente exposto ao solvente; enquanto que as amostradas nas simulações correspondem a conformações do triptofano parcialmente protegido do solvente. Portanto, é de se esperar que cálculos dos decaimentos de anisotropia de fluorescência partindo das conformações corretas do Trp113 da sC possam ter maior concordância com dados experimentais.

6.3.3 Decaimento da anisotropia de fluorescência do Trp113 da subtilisina Carlsberg completamente exposto ao solvente

Em vista dos tempos de decaimento obtidos para o Trp na conformação cristalográfica, a reprodução dos resultados experimentais depende de amostrar conformações distintas do Trp113 da sC. As conformações que são amostradas nos experimentos correspondem àquelas na qual o Trp113 está exposto ao solvente ($\lambda_{em} \sim 355 - 360$ nm). Com a finalidade de obter uma

estrutura da sC com o resíduo Trp113 exposto ao solvente, uma simulação de dinâmica molecular de 10 ns no vácuo numa temperatura de 815K, considerando a proteína fixa com exceção do resíduo Trp113, foi realizada. Dita simulação permitiu amostrar as conformações possíveis do Trp113 sem alterar a estrutura da proteína. Uma estrutura com o resíduo Trp113 exposto ao solvente foi selecionada para dar início às simulações de dinâmica molecular (Figura 6.5A).

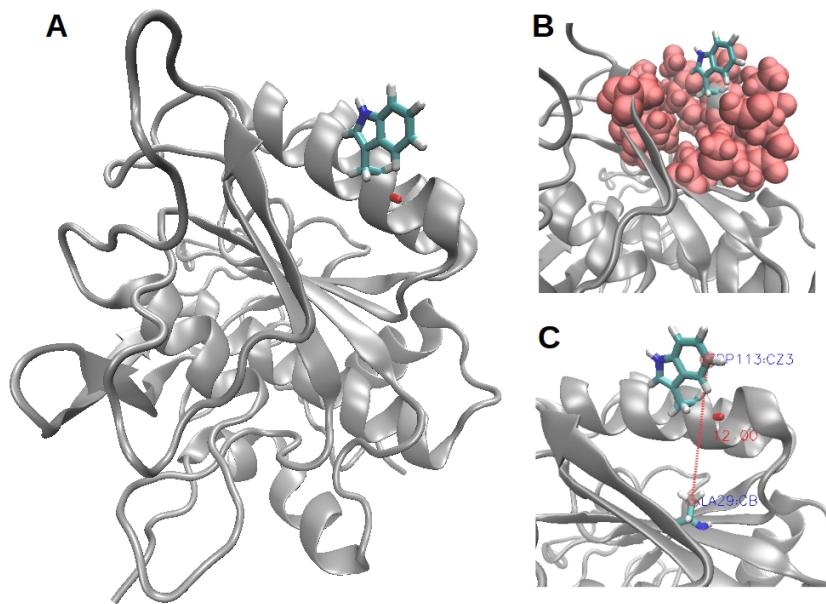


Figura 6.5: A) Conformação da enzima subtilisina Carlsberg com o Trp113 completamente exposto ao solvente. B) Representação da acessibilidade ao solvente do resíduo Trp113. C) Representação da distância característica entre os átomos Ala29-C β e Trp113-CZ3 na conformação da sC com o Trp113 completamente exposto ao solvente ($\sim 12 \text{ \AA}$).

A estrutura do Trp113 no bolsão hidrofóbico é mais estável comparada à conformação com o resíduo exposto ao solvente. Devido a isso, as simulações usando a conformação do Trp113 completamente exposto ao solvente como ponto de partida amostraram quase inevitavelmente as conformações do Trp113 no bolsão hidrofóbico. Para amostrar as conformações do Trp113 exposto ao solvente, é necessário criar restrições que impeçam ao Trp113 entrar no bolsão hidrofóbico durante as simulações de dinâmica molecular.

Fazendo uma análise visual nas simulações da proteína sC, encontramos que a distância entre os átomos Ala29-C β e Trp113-CZ3 está relacionado ao grau de exposição do resíduo Trp113 ao solvente (Figura 6.5C). Assim, distâncias curtas entre os átomos Ala29-C β e Trp113-CZ3 são características das conformações com o Trp113 parcialmente protegido do

solvente, enquanto que distâncias maiores entre estes átomos caracterizam conformações com o Trp113 completamente exposto ao solvente. Distâncias mínimas de 12 Å entre os átomos Ala29-C β e Trp113-CZ3 foram escolhidas como sendo necessárias para amostrar conformações da sC com o Trp113 completamente exposto ao solvente durante todo o tempo das simulações. Assim, simulações tomado como ponto de partida a estrutura da sC com o Trp113 completamente exposto ao solvente e com restrições da distância entre os átomos Ala29-C β e Trp113-CZ3 foram realizadas.

O comprimento de onda de emissão de fluorescência foi calculado para cada estrutura amostrada. A distribuição dos λ_{em} instantâneos calculados com os modelos parametrizados baseados na SASA do benzeno/pirrol e nas interações eletrostáticas do indol com a água/proteína estão mostrados na Figura 6.6A e B, respectivamente. Analisando a localização dos picos dos espectros experimentais e das distribuições calculadas (Figura 6.6A e B) é possível de observar a existência de uma concordância nos valores dos λ_{em} .

Segundo o modelo baseado na SASA do benzeno/pirrol o λ_{em} calculado é de ~ 355 nm, sendo muito próximo ao dado experimental. Por outro lado, o modelo baseado nas interações eletrostáticas do indol com a água/proteína estima um λ_{em} de ~ 347 nm. Estes valores calculados, segundo a classificação de Burstein et al. apresentada no Capítulo 4, correspondem a conformações da proteína com o Trp completamente exposto ao solvente [80, 81]. Os valores dos λ_{em} calculados da conformação da subtilisina Carlsberg com o Trp113 exposto ao solvente são mostrados na Tabela 6.3. A similaridade dos λ_{em} calculados e experimentais nos indicam que a inclusão das restrições das distâncias entre os átomos Ala29-C β e Trp113-CZ3 é suficiente para amostrar o Trp113 exposto ao solvente, as quais foram amostradas preferencialmente nos experimentos.

A anisotropia de fluorescência resolvida no tempo foi calculada para a conformação da subtilisina Carlsberg com o Trp113 exposto ao solvente (Figura 6.6C). Os decaimentos da anisotropia calculados dessas conformações foram mais rápidos comparados aos decaimentos da conformação cristalográfica. As curvas calculadas mostram um rápido decaimento durante os primeiros nanosegundos sem alcançar uma despolarização completa (Figura 6.6C).

Um ajuste bi-exponencial aos decaimentos de anisotropia calculados permite estimar os tempos de reorientação das componentes rápida e lenta da dinâmica da conformação

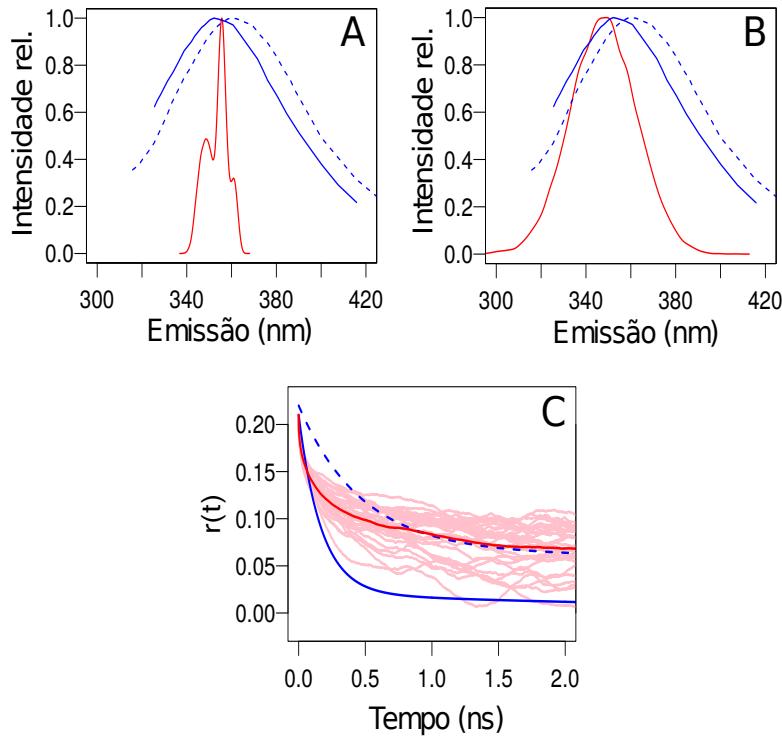


Figura 6.6: Fluorescência da conformação da subtilisina Carlsberg com o resíduo Trp113 exposto ao solvente. A) e B) Espectro de fluorescência experimental (azul) e distribuição dos λ_{em} calculados (vermelho) do Trp113 da sC em água. Os cálculos dos λ_{em} foram feitos usando o modelo parametrizado baseado na SASA do benzeno/pirrol (em A) ou o modelo baseado nas interações eletrostáticas do indol com a água/proteína (em B). C) Anisotropia de fluorescência resolvida no tempo do Trp113 da sC em água. As curvas azuis contínuas e tracejadas correspondem aos dados experimentais reportados por Lakshmikanth e Krishnamoorthy (1999) [92] e Shaw e Pal (2007) [93], respectivamente.

exposta ao solvente do Trp113 da subtilisina Carlsberg. As componentes rápida e lenta foram caracterizadas com tempos de reorientação de 0,18 e \sim 10 ns, respectivamente (Tabela 6.3).

O tempo de reorientação calculado da componente rápida é muito próximo ao reportado por Lakshmikanth e Krishnamoorthy (1999) [92], o qual foi de 0,17 ns. Cabe ressaltar que, apesar dessa similaridade, os decaimentos obtidos pelas simulações deveriam ser mais rápidos que os experimentais. Portanto, não é possível associar diretamente a componente rápida do tempo de reorientação do Trp113 calculado pelas simulações aos resultados experimentais. Com respeito à componente lenta dos decaimentos de anisotropia calculados, estes são da ordem de \sim 3 a 15 ns. Os dados experimentais da componente lenta mostram valores de 3,5 ns [92] e 10-100 ns [93].

O maior determinante da forma das curvas de decaimento de anisotropia não são

os tempos de reorientação, senão as contribuições das componentes rápidas e lentas. No experimento de Lakshmikanth e Krishnamoorthy (1999), a contribuição da componente rápida é de 90%, responsável pelo decaimento rápido e completo após os primeiros 2 ns [92]. Por outro lado, nos experimentos de Shaw e Pal (2007), a contribuição da componente rápida é de apenas 70%, além do que é mais lenta (4,7 ns), gerando assim um decaimento no qual após 2 ns é possível observar ainda sinal de luz polarizada [93].

As contribuições das componentes rápidas e lentas do decaimento de anisotropia calculada para a conformação da subtilisina Carlsberg com o Trp113 no bolsão hidrofóbico são de 3% e 97%, respectivamente (Tabela 6.3). A contribuição da componente rápida sobe para 43% quando a conformação com o Trp113 completamente exposto ao solvente é amostrada (Tabela 6.3). Esse aumento da contribuição da componente rápida faz com que o decaimento da anisotropia seja mais rápido, sendo assim parecido ao resultado experimental reportado por Shaw e Pal (2007) [93] (Figura 6.6C). Assim, podemos estabelecer que a curva de decaimento da anisotropia da subtilisina Carlsberg é produto da população de conformações amostradas, sendo a conformação com o Trp113 exposto ao solvente a mais provável nos experimentos.

Assim, para reproduzir adequadamente os decaimentos experimentais da anisotropia de fluorescência por meio de simulações de dinâmica molecular, é necessário amostrar conformações similares às usadas nos experimentos. Uma forma de lograr isso, e que é exemplificada neste Capítulo usando a enzima subtilina Carlsberg, é mediante a comparação dos comprimentos de onda de emissão de fluorescência experimental e calculado, sendo este último possível de obter de maneira prática graças aos modelos clássicos parametrizados, desenvolvidos também nesta tese.

6.4 Conclusões

A subtilisina Carlsberg (sC) é uma enzima que possui um único resíduo de triptofano (Trp113), e como foi discutido anteriormente, apresenta uma variabilidade nos λ_{em} dependendo das condições experimentais. A dinâmica reorientacional do Trp113 da sC foi estudada, usando a técnica de anisotropia de fluorescência resolvida no tempo por dois grupos independentes. O primeiro deles, Lakshmikanth e Krishnamoorthy (1999), reportou uma dinâmica rápida com perda completa de polarização após os primeiros 2 ns. Já segundo Shaw e Pal (2007), o

Tabela 6.3: Parâmetros experimentais e calculados da anisotropia de fluorescência resolvida no tempo do Trp113 da subtilisina Carlsberg em água. λ_{em} é o comprimento de onda usado para registrar os decaimentos de anisotropia. O λ_{em} foi calculado usando ^(a) o modelo baseado na SASA do benzeno/pirrol e ^(b) nas interações eletrostáticas do indol com a água/proteína. O modelo bi-exponencial usado para o ajuste do decaimento da anisotropia de fluorescência foi: $r(t) = r_0 \times (\beta_1 e^{-t/\tau_1} + \beta_2 e^{-t/\tau_2})$, sendo β_1 e β_2 os termos pré-exponenciais, e τ_1 e τ_2 os tempos de reorientação das componentes rápidas e lentas, respectivamente. Os λ_{em} calculados correspondem a média dos comprimentos de onda calculados a partir de 30 simulações de dinâmica molecular. O ajuste exponencial da anisotropia foi aplicado à média das 30 curvas de decaimento de anisotropia de fluorescência calculadas a partir das simulações de dinâmica molecular. ^(†) Conformação da sC com o Trp113 parcialmente protegido do solvente. ^(‡) Conformação da sC com o Trp113 completamente exposto ao solvente.

	$\lambda_{em}(nm)$	r_0	β_1	τ_1 (ns)	β_2	τ_2 (ns)
Experimental [92]	355	0,21	0,90	0,17	0,10	3,50
Experimental [93]	360	0,22	0,70	0,469	0,30	10-100
Calculado [†]	$332,93 \pm 0,47$ ^a	0,20	0,03	1,29	0,97	86
	$331,35 \pm 1,21$ ^b					
Calculado [‡]	$354,52 \pm 2,08$ ^a	0,18	0,43	0,18	0,57	3-15
	$346,78 \pm 2,25$ ^b					

decaimento da anisotropia não chega a ser completo após os primeiros 2 ns. Essa divergência nos dados experimentais motivou nosso grupo de pesquisa a estudar a dinâmica reorientacional do Trp113 da sC usando simulações de dinâmica molecular. Numa primeira tentativa as simulações foram incapazes de predizer de maneira satisfatória os decaimentos de anisotropia [99]. Essas simulações tomavam como ponto de partida a estrutura cristalográfica da sC, a qual possui o resíduo Trp113 orientado para dentro do bolso hidrofóbico e portanto com pouca acessibilidade ao solvente. Os λ_{em} foram calculados com os modelos parametrizados desenvolvidos nesta tese para essas conformações, sendo de ~ 331 nm. Isso contrasta com os valores de λ_{em} usados nos experimentos de anisotropia de fluorescência, os quais eram de ~ 355 nm. Essa comparação entre os λ_{em} calculados e experimentais nos indicava que diferentes conformações estariam sendo amostradas pela dinâmica molecular e pelas condições experimentais. Assim a dinâmica explorava preferencialmente conformações com o Trp113 parcialmente protegido do solvente, e os experimentos amostravam conformações do Trp113 completamente exposto ao solvente. Portanto, era de se esperar que a dinâmica reorientacional do Trp113 obtida nas simulações seja muito lenta comparada ao experimental.

Para lograr uma concordância entre os decaimentos de anisotropia de fluorescência calculados e experimentais era necessário amostrar unicamente as conformações com o Trp113

da sC exposto ao solvente. Assim simulações com Trp113 exposto ao solvente foram realizadas, e restrições conformacionais foram aplicadas nessas simulações para garantir que o Trp113 fique exposto durante todo o tempo das simulações. Os λ_{em} calculados foram, então, de ~ 350 nm, sendo similares aos experimentais. A partir de uma comparação dos λ_{em} podemos dizer que as estruturas amostradas pelas simulações e no experimento são similares. Com essas simulações procedemos a calcular os decaimentos de anisotropia de fluorescência do Trp113 exposto ao solvente. Os decaimentos foram muito mais rápidos comparados aos obtidos com a estrutura cristalográfica da sC, conciliando o resultado com as medidas experimentais.

Assim, é necessário que exista uma similaridade na amostragem das conformações no experimento e nas simulações, para obter uma correspondência nas dinâmicas reorientacionais experimentais e calculadas. Uma forma de avaliar essa similaridade de conformações, e que foi aplicada para o caso da subtilisina Carlsberg, é mediante a comparação dos comprimentos de onda de fluorescência usados no experimento e os obtidos pelas simulações (estes últimos podendo ser computados com os nossos modelos parametrizados). A metodologia apresentada serve como guia para obter decaimentos de anisotropia de fluorescência por meio de simulações de dinâmica molecular de proteínas com grupos fluorescentes que apresentam múltiplas conformações.

Capítulo 7

Conclusões finais

A fluorescência do triptofano em proteínas é altamente sensível ao ambiente, gerando assim uma variabilidade nos comprimentos de onda de emissão (λ_{em}) e nos rendimentos quânticos (Φ_F). Essas propriedades foram estudadas usando métodos baseados em primeiros princípios, os quais permitiram desenvolver metodologias híbridas com capacidade para prever os λ_{em} e Φ_F do triptofano. Apesar da boa capacidade preditiva dos modelos híbridos, o seu uso está restrito ao pessoal teórico pois é uma metodologia que requer experiência, além de programas computacionais específicos. Por isso houve a necessidade de construir modelos clássicos parametrizados com capacidade de prever os comprimentos de onda de emissão usando informações estruturais das proteínas por meio de simulações de dinâmica molecular. As informações usadas para a construção dos modelos foram a área acessível ao solvente do indol e as interações eletrostáticas do indol com o ambiente (água + proteína). Este modelos, apesar da sua simplicidade, mostraram boa capacidade preditiva, possibilitando assim a compreensão dos dados experimentais de uma maneira simples e rápida.

A inconsistência encontrada entre os decaimentos anisotropia de fluorescência experimentais e calculados do Trp113 da subtilisina Carlsberg (sC) também foi explorada no presente trabalho. Os experimentos mostraram que os decaimentos de anisotropia do Trp113 são rápidos, e junto aos dados de fluorescência sugerem que o Trp113 está completamente exposto ao solvente. Por outro lado, os decaimentos de anisotropia calculados a partir da estrutura cristalográfica da sC revelam que movimentação é restrita e acoplada a movimento da proteína. Adicionalmente, os λ_{em} calculados com os modelos parametrizados para essas estruturas indicam que o Trp113 está inserido no bolsão hidrofóbico. Simulações de dinâmica molecular considerando a cadeia lateral do resíduo Trp113 completamente exposto ao solvente foram realizadas com a finalidade de encontrar uma concordância com os dados experimentais.

Os λ_{em} calculados com essas estruturas foram similares aos experimentais e sugerem assim que o Trp113 exposto ao solvente é a conformação que está sendo amostrada nos experimentos. Os decaimentos de anisotropia calculados com as estruturas do Trp113 exposto ao solvente foram mais rápidas comparadas aos decaimentos calculados a partir da estrutura cristalográfica, e similares aos decaimentos experimentais. Esse caso da dinâmica reorientacional da sC é um exemplo de como os modelos parametrizados podem auxiliar a interpretação dos dados experimentais.

Capítulo 8

Perspectivas

- **O desenvolvimento de modelos clássicos para a predição do rendimento quântico do triptofano em proteínas.** O grupo amida próximo ao indol tem um papel importante no processo supressão da fluorescência do triptofano. Adicionalmente, os grupos laterais dos aminoácidos lisina, glutamina, asparagina, histidina, ácido glutâmico, ácido aspártico, tirosina e cisteína atuam também como supressores da fluorescência do triptofano. O modelo parametrizado para a predição dos rendimentos quânticos deve considerar a proximidade grupo indol a esses supressores.
- **Estender nosso modelos parametrizados para estudar a fluorescência de proteínas com múltiplos triptofanos.** Esse é um fenômeno complexo, pois é necessário considerar tanto o comprimento de onda, como o rendimento quântico de cada triptofano presente na proteína. Esse modelo pode abrir as portas para uma compreensão estrutural muito mais completo dos experimentos de fluorescência.
- **Desenvolver modelos parametrizados para a predição do λ_{em} do triptofano em solventes distintos da água.** As interações eletrostáticas são importantes para a determinação dos λ_{em} , a pergunta é se esse modelo pode ser aplicado para sistemas onde o solvente não é a água.
- **Realizar experimentos de fluorescência e anisotropia de fluorescência resolvido no tempo do Trp113 da subtilisina Carlsberg.** Devido a inconsistência encontrada na literatura sobre os dados experimentais, é necessário refazer esses experimentos considerando a possibilidade de proteólise e contaminação da fluorescência por tirosinas. Esses novos resultados experimentais, acompanhados dos resultados das simulações podem ajudar a melhor compreensão do problema.

Bibliografia

- [1] Bernard Valeur e Mário Nuno Berberan-Santos. *Molecular Fluorescence: Principles and Applications*. Ed. por Wiley-VCH Verlag. Second Edition. 2012.
- [2] G.G. Stokes. “On the change of refrangibility of light”. Em: *Philosophical Transactions of the Royal Society of London* 142 (1852), pp. 463–562.
- [3] G.G. Stokes. “On the change of refrangibility of light - No. II.” Em: *Philosophical Transactions of the Royal Society of London* 143 (1853), pp. 385–396. URL: <https://archive.org/details/philtrans07674256>.
- [4] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Ed. por Springer. Third Edition. 2006.
- [5] Gordon Wright e Robert B. Freedman. “Structure and dynamics of staphylococcal nucleic acid mutants as studied by fluorescence quenching techniques”. Em: *Protein Engineering, Design and Selection* 2 (1989), pp. 583–588. URL: <https://doi.org/10.1093/protein/2.8.583>.
- [6] Maurice R. Eftink. “Fluorescence techniques for studying protein structure”. Em: *Methods of Biochemical Analysis* 35 (1991), pp. 127–205. DOI: 10.1002/9780470110560. URL: <https://doi.org/10.1002/9780470110560.ch3>.
- [7] Catherine A. Royer, Craig J. Mann e C. Robert Matthews. “Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants”. Em: *Protein science* 2.11 (1993), pp. 1844–1852. DOI: 10.1002/pro.5560021106. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2142281/>.
- [8] Ulrich Kettling et al. “Real-time enzyme kinetics monitored by dual-color fluorescence cross-correlation spectroscopy”. Em: *Proceedings of the National Academy of Sciences of the United States of America* 95 (1998), pp. 1416–1420. URL: <https://doi.org/10.1073/pnas.95.4.1416>.
- [9] Mark O. Palmier e Steven R. Van Doren. “Rapid Determination of Enzyme Kinetics from Fluorescence: Overcoming the Inner Filter Effect”. Em: *Analytical Biochemistry* 371 (2007), pp. 43–51. URL: <https://doi.org/10.1016/j.ab.2007.07.008>.

- [10] Sally A. Kim, Katrin G. Heinze e Petra Schwille. “Fluorescence correlation spectroscopy in living cells”. Em: *Nature Methods* 4 (2007), pp. 963–973. DOI: 10.1038/nmeth1104.
- [11] Ronan Chéreau, Jan Tønnesen e U. Valentin Nägerl. “STED microscopy for nanoscale imaging in living brain slices”. Em: *Methods* 88 (2015), pp. 57–66. URL: <https://doi.org/10.1016/jymeth.2015.06.006>.
- [12] C. Roobala, I.P. Ilanila e J.K. Basu. “Applications of STED fluorescence nanoscopy in unravelling nanoscale structure and dynamics of biological systems”. Em: *Journal of Biosciences* 43 (2018), pp. 471–484. URL: <https://doi.org/10.1007/s12038-018-9764-3>.
- [13] Kathryn P. Wall, Rebecca Dillon e Michelle K. Knowles. “Fluorescence quantum yield measurements of fluorescent proteins: a laboratory experiment for a biochemistry or molecular biophysics laboratory course”. Em: *Biochemistry and Molecular Biology Education* 43 (2015), pp. 52–59. URL: <https://doi.org/10.1002/bmb.20837>.
- [14] Paul D. Adams et al. “Intramolecular quenching of tryptophan fluorescence by the peptide bond in cyclic hexapeptides”. Em: *Journal of the American Chemical Society* 7 (2002), pp. 9278–9286. DOI: 10.1021/ja0167710.
- [15] Patrik R. Callis e Tiqing Liu. “Quantitative Prediction of Fluorescence Quantum Yields for Tryptophan in Proteins.” Em: *The Journal of Physical Chemistry B* 108 (2004), pp. 4248–4259. DOI: 10.1021/jp0310551.
- [16] Y. Chen e M.D. Barkley. “Toward Understanding Tryptophan Fluorescence in Proteins”. Em: *Biochemistry* 37 (1998), pp. 9976–9982. URL: <https://doi.org/10.1021/bi980274n>.
- [17] Kevin J. Willis e Arthur G. Szabo. “Resolution of tyrosyl and tryptophyl fluorescence emission from subtilisins”. Em: *Biochemistry* 28 (1989), pp. 4902–4908. DOI: 0.1021/bi00437a056.
- [18] Raymond F. Chen. “Measurements of absolute values in biochemical fluorescence spectroscopy”. Em: *Journal of Research of the National Bureau of Standards* 76 (1972), pp. 593–606.
- [19] G.D. Fasman. *Handbook of Biochemistry and Molecular Biology*. Ed. por Ohio CRC Press Cleveland. Third Edition. 1976.
- [20] PhotochemCAD. Em: (2018). URL: <https://omlc.org/spectra/PhotochemCAD/>.

- [21] Patrik R. Callis, W.R. Kirk e F.G. Prendergast. “pH-Induced fluorescence shifts in E.coli thioredoxin mutants. Comparison of experiment and theory”. Em: *Biophysical Journal Abstract* 78 (2000), 127a.
- [22] Nar Herbert et al. “Crystal structure analysis of oxidized Pseudomonas aeruginosa azurin at pH 5.5 and pH 9.0. A pH-induced conformational transition involves a peptide bond flip.” Em: *Journal of Molecular Biology* 221 (1991), pp. 765–772. URL: [https://doi.org/10.1016/0022-2836\(91\)80173-R](https://doi.org/10.1016/0022-2836(91)80173-R).
- [23] Jaap Broos et al. “The emitting state of tryptophan in proteins with highly blue-shifted fluorescence”. Em: *Angewandte Chemie* 46 (2007), pp. 5137–5139. URL: <https://doi.org/10.1002/anie.200700839>.
- [24] M. Susan Cates et al. “Metal-ion affinity and specificity in EF-hand proteins: coordination geometry and domain plasticity in parvalbumin”. Em: *Structure* 7 (1999), pp. 1269–1278. URL: [https://doi.org/10.1016/S0969-2126\(00\)80060-X](https://doi.org/10.1016/S0969-2126(00)80060-X).
- [25] George I. Birnbaum et al. “1.70 Å resolution structure of myoglobin from yellowfin tuna. An example of a myoglobin lacking the D helix”. Em: *Acta crystallographica. Section D, Biological crystallography* 50 (1994), pp. 283–289. URL: <https://doi.org/10.1107/S0907444993014271>.
- [26] José Martinez-Oyanedel et al. “Ribonuclease T1 with free recognition and catalytic site: crystal structure analysis at 1.5 Å resolution”. Em: *Journal of Molecular Biology* 222 (1991), pp. 335–352. URL: [https://doi.org/10.1016/0022-2836\(91\)90215-R](https://doi.org/10.1016/0022-2836(91)90215-R).
- [27] R. Swaminathan, G. Krishnamoorthy e N. Periasamy. “Similarity of fluorescence lifetime distributions for single tryptophan proteins in the random coil state”. Em: *Biophysical Journal* 67 (1994), pp. 2013–2023. DOI: [10.1016/S0006-3495\(94\)80685-X](https://doi.org/10.1016/S0006-3495(94)80685-X).
- [28] P. Burkhard, P. Taylor e M.D. Walkinshaw. “X-ray structures of small ligand-FKBP complexes provide an estimate for hydrophobic interaction energies”. Em: *Journal of Molecular Biology* 295 (2000), pp. 953–962. DOI: [10.1006/jmbi.1999.3411](https://doi.org/10.1006/jmbi.1999.3411).
- [29] D.A. Egan et al. “Equilibrium denaturation of recombinant human FK binding protein in urea”. Em: *Biochemistry* 32 (1993), pp. 1920–1927. DOI: [10.1021/bi00059a006](https://doi.org/10.1021/bi00059a006).
- [30] D.R. Rose et al. “Crystal structure of T4-lysozyme generated from synthetic coding DNA expressed in Escherichia coli”. Em: *Protein Engineering, Design and Selection* 2 (1988), pp. 277–282. URL: <https://doi.org/10.1093/protein/2.4.277>.

- [31] D.L. Harris e B.S. Hudson. “Fluorescence and molecular dynamics study of the internal motion of the buried tryptophan in bacteriophage T4 lysozyme: Effects of temperature and alteration of nonbonded networks”. Em: *Chemical Physics* 158 (1991), pp. 353–382. URL: [https://doi.org/10.1016/0301-0104\(91\)87077-9](https://doi.org/10.1016/0301-0104(91)87077-9).
- [32] David J. Neidhart e Gregory A. Petsko. “The refined crystal structure of subtilisin Carlsberg at 2.5 Å resolution”. Em: *Protein Engineering, Design and Selection* 2 (1988), pp. 271–276. URL: <https://doi.org/10.1093/protein/2.4.271>.
- [33] P.M. Bayley, J.M. Janot e S.R. Martin. “Subtilisin enzymes: A note on time-resolved fluorescence and circular dichroism properties”. Em: *FEBS Letters* 250 (1989), pp. 389–394. URL: [https://doi.org/10.1016/0014-5793\(89\)80762-8](https://doi.org/10.1016/0014-5793(89)80762-8).
- [34] Thomas R. Hynes e Robert O. Fox. “The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution”. Em: *Proteins* 10 (1991), pp. 92–105. URL: <https://doi.org/10.1002/prot.340100203>.
- [35] Lluís Bellsolell et al. “Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface”. Em: *Journal of Molecular Biology* 238 (1994), pp. 489–495. URL: <https://doi.org/10.1006/jmbi.1994.1308>.
- [36] M.D. Walkinshaw, W. Saenger e A. Maelicke. “Three-dimensional structure of the long neurotoxin from cobra venom”. Em: *Proceedings of the National Academy of Sciences of the United States of America* 77 (1980), pp. 2400–2404. URL: <https://doi.org/10.1073/pnas.77.5.2400>.
- [37] S. Sugio et al. “Crystal structure of human serum albumin at 2.5 Å resolution”. Em: *Protein Engineering, Design and Selection* 12 (1999), pp. 439–446. URL: <https://doi.org/10.1093/protein/12.6.439>.
- [38] Michael Dockal et al. “Five recombinant fragments of human serum albumin-tools for the characterization of the warfarin binding site”. Em: *Protein science* 9 (2000), pp. 1455–1465. URL: <https://doi.org/10.1110/ps.9.8.1455>.
- [39] John R. Somoza et al. “Two crystal structures of a potently sweet protein. Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution”. Em: *Journal of Molecular Biology* 234 (1993), pp. 390–404. URL: <https://doi.org/10.1006/jmbi.1993.1594>.

- [40] Ethan A. Merritt et al. “Surprising leads for a cholera toxin receptor-binding antagonist: crystallographic studies of CTB mutants”. Em: *Structure* 3 (1995), pp. 561–570. URL: [https://doi.org/10.1016/S0969-2126\(01\)00190-3](https://doi.org/10.1016/S0969-2126(01)00190-3).
- [41] Suresh K. Katti, David M. LeMaster e Hans Eklund. “Crystal structure of thioredoxin from Escherichia coli at 1.68 Å resolution”. Em: *Journal of Molecular Biology* 212 (1990), pp. 167–184. URL: [https://doi.org/10.1016/0022-2836\(90\)90313-B](https://doi.org/10.1016/0022-2836(90)90313-B).
- [42] Thomas C. Terwilliger e David S. Eisenberg. “The structure of melittin. I. Structure determination and partial refinement”. Em: *Journal of Biological Chemistry* 257 (1982), pp. 6010–6015. URL: <http://www.jbc.org/content/257/11/6010.full.pdf>.
- [43] Joseph P. Noel et al. “Phospholipase A2 engineering. X-ray structural and functional evidence for the interaction of lysine-56 with substrates”. Em: *Biochemistry* 30 (1991), pp. 11801–11811. DOI: 10.1021/bi00115a010.
- [44] Liang Zhao et al. “Dynamics of ordered water in interfacial enzyme recognition: bovine pancreatic phospholipase A2”. Em: *Angewandte Chemie* 43 (2003), pp. 60–63. URL: <https://doi.org/10.1002/anie.200352961>.
- [45] Kyoyu Sasaki et al. “X-ray analysis of glucagon and its relationship to receptor binding”. Em: *Nature* 257 (1995), pp. 751–757. URL: <https://doi.org/10.1038/257751a0>.
- [46] Michael Wahl. “Time-Correlated Single Photon Counting”. Em: *Technical note PicoQuant GmbH* (2014). URL: https://www.picoquant.com/images/uploads/page/files/7253/technote_tcspc.pdf.
- [47] S. Vogel et al. *Chapter 10: Time-Resolved Fluorescence Anisotropy*. In book: *FLIM Microscopy in Biology and Medicine*. Ed. por Chapman e Hall CRC. First Edition. 2009. DOI: 10.1201/9781420078916-11.
- [48] Telmo J.V. Prazeres et al. “Accurate Determination of the Limiting Anisotropy of Rhodamine 101. Implications for Its Use as a Fluorescence Polarization Standard”. Em: *The Journal of Physical Chemistry A* 112 (2008), pp. 5034–5039. DOI: 10.1021/jp710625j.
- [49] Toshiko Ichiye e Martin Karplus. “Fluorescence depolarization of tryptophan residues in proteins: a molecular dynamics study”. Em: *Biochemistry* 7 (1983), pp. 2884–2893. DOI: 10.1021/bi00281a017.

- [50] Gunnar F. Schröder, Ulrike Alexiev e Helmut Grubmüller. “Simulation of Fluorescence Anisotropy Experiments: Probing Protein Dynamics”. Em: *Biophysical Journal* 89 (2005), pp. 3757–3770. DOI: 10.1529/biophysj.105.069500.
- [51] Mariana R.B. Batista e Leandro Martínez. “Dynamics of nuclear receptor Helix-12 switch of transcription activation by modeling time-resolved fluorescence anisotropy decays”. Em: *Biophysical Journal* 105 (2013), pp. 1670–1680. DOI: 10.1016/j.bpj.2013.07.032.
- [52] M.P. Allen, D.J. Tildesley e Lois Allen. *Computer simulation of liquids*. Ed. por USA Oxford University Press. First Edition. 1989. ISBN: 9780198556459.
- [53] M.P. Allen. *Introduction to Molecular Dynamics Simulation*. In book: *Computational Soft Matter: From Synthetic Polymers to Proteins*. Ed. por Julich NIC Series. First Edition. 2004. ISBN: 3-00-012641-4.
- [54] José Mario Martínez e Leandro Martínez. “Packing Optimization for Automated Generation of Complex System’s Initial Configurations for Molecular Dynamics and Docking”. Em: *Journal of Computational Chemistry* 24 (2003), pp. 819–825. DOI: 10.1002/jcc.10216.
- [55] Leandro Martínez et al. “Packmol: A package for building initial configurations for molecular dynamics simulations”. Em: *Journal of Computational Chemistry* 30 (2009), pp. 2157–2164. DOI: 10.1002/jcc.21224.
- [56] A. Ilari e C. Savino. “Protein structure determination by x-ray crystallography”. Em: *Methods in Molecular Biology* 452 (2008), pp. 63–87. DOI: 10.1007/978-1-60327-159-2_3.
- [57] Andrea Cavalli et al. “Protein structure determination from NMR chemical shifts”. Em: *Proceedings of the National Academy of Sciences of the United States of America* 104 (2007), pp. 9615–9620. DOI: 10.1073/pnas.0610313104.
- [58] H.M. Berman et al. “The protein data bank”. Em: *Nucleic acids research* 28 (2000), pp. 235–242. URL: <https://doi.org/10.1093/nar/28.1.235>.
- [59] Yang Zhang. “Progress and challenges in protein structure prediction”. Em: *Current Opinion in Structural Biology* 18 (2008), pp. 342–348. DOI: 10.1016/j.sbi.2008.02.004.
- [60] E. Carlon, M. Laleman e S. Nomidis. *Molecular Dynamics Simulations*. URL: http://itf.fys.kuleuven.be/~enrico/Teaching/molecular_dynamics_2015.pdf (acesso em 25/01/2018).

- [61] NAMD 2.9 User's Guide. *Non-bonded interactions*. URL: <https://www.ks.uiuc.edu/Research/namd/2.9/ug/node23.html> (acesso em 01/04/2018).
- [62] Juan M. Bello-Rivas. *Tutorial on the smooth particle-mesh Ewald algorithm*. URL: <https://superadditive.com/notes/particle-mesh-ewald/> (acesso em 01/04/2018).
- [63] EMBL Tutorials. *Potential energy function*. URL: https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part2.html (acesso em 01/04/2018).
- [64] NAMD TUTORIAL. *Parameter Files*. URL: <https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node25.html> (acesso em 01/04/2018).
- [65] A.D. MacKerell et al. “All-atom empirical potential for molecular modeling and dynamics studies of proteins”. Em: *The Journal of Physical Chemistry B* 102 (1998), pp. 3586–3616. DOI: [10.1021/jp973084f](https://doi.org/10.1021/jp973084f).
- [66] Alexander D. MacKerell Jr., Michael Feig e Charles L. Brooks. “Improved treatment of the protein backbone in empirical force fields”. Em: *Journal of Computational Chemistry* 126 (2004), pp. 698–699. DOI: [10.1021/ja036959e](https://doi.org/10.1021/ja036959e).
- [67] Alexander D. MacKerell Jr., Michael Feig e Charles L. Brooks. “Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations”. Em: *Journal of Computational Chemistry* 25 (2004), pp. 1400–1415. URL: <https://doi.org/10.1002/jcc.20065>.
- [68] J. Schlitter et al. “Targeted Molecular Dynamics Simulation of Conformational Change - Application to the T - R Transition in Insulin”. Em: *Molecular simulation* 10 (1993), pp. 291–308. URL: <https://doi.org/10.1080/08927029308022170>.
- [69] Helmut Grubmüller, Berthold Heymann e Paul Tavan. “Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force”. Em: *Science* 271 (1996), pp. 997–999. DOI: [10.1126/science.271.5251.997](https://doi.org/10.1126/science.271.5251.997).
- [70] Jeffrey Comer et al. “The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask”. Em: *The Journal of Physical Chemistry B* 119 (2015), pp. 1129–1151. DOI: [10.1021/jp506633n](https://doi.org/10.1021/jp506633n).
- [71] B. Leimkuhler et al. *Lecture Notes in Computational Science and Engineering: New Algorithms for Macromolecular Simulation*. Ed. por Springer. First Edition. 2006. ISBN: 978-3-540-31618-3.

- [72] NAMD 2.9 User's Guide. *Biasing and analysis methods*. URL: <https://www.ks.uiuc.edu/Research/namd/2.9/ug/node56.html> (acesso em 01/04/2018).
- [73] Patrik R. Callis e Berit K. Burgess. "Tryptophan Fluorescence Shifts in Proteins from Hybrid Simulations: An Electrostatic Approach". Em: *The Journal of Physical Chemistry B* 101 (1997), pp. 9429–9432. DOI: [10.1021/jp972436f](https://doi.org/10.1021/jp972436f).
- [74] James T. Vivian e Patrik R. Callis. "Mechanisms of tryptophan fluorescence shifts in proteins". Em: *Biophysical Journal* 80 (2001), pp. 9429–9432. URL: [https://doi.org/10.1016/S0006-3495\(01\)76183-8](https://doi.org/10.1016/S0006-3495(01)76183-8).
- [75] William L. Jorgensen, Jayaraman Chandrasekhar e Jeffry D. Madura. "Comparison of simple potential functions for simulating liquid water". Em: *Journal of Computational Chemistry* 79 (1983), pp. 926–935. URL: <https://doi.org/10.1063/1.445869>.
- [76] James C. Phillips et al. "Scalable molecular dynamics with NAMD". Em: *Journal of Computational Chemistry* 26 (2005), pp. 1781–1802. URL: <https://doi.org/10.1002/jcc.20289>.
- [77] William Humphrey, Andrew Dalke e Klaus Schulten. "VMD: visual molecular dynamics". Em: *Journal of Molecular Graphics* 14 (1996), pp. 33–38. URL: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [78] A. Shrake e J.A. Rupley. "Environment and exposure to solvent of protein atoms. Lysozyme and insulin". Em: *Journal of Molecular Biology* 79 (1973), pp. 351–371. URL: [https://doi.org/10.1016/0022-2836\(73\)90011-9](https://doi.org/10.1016/0022-2836(73)90011-9).
- [79] L.C. Pauling. *The Nature of the Chemical Bond*. Ed. por New York Cornell University Press Ithaca. Third Edition. 1960.
- [80] E.A. Burstein, N.S. Vedenkina e M.N. Ivkova. "Fluorescence and the location of tryptophan residues in protein molecules". Em: *Photochemistry and Photobiology* 18 (1973), pp. 263–279. URL: <https://doi.org/10.1111/j.1751-1097.1973.tb06422.x>.
- [81] Y.K. Reshetnyak, Y. Koshevnik e Burstein E.A. "Decomposition of protein tryptophan fluorescence spectra into log-normal components. III. Correlation between fluorescence and microenvironment parameters of individual tryptophan residues". Em: *Biophysical Journal* 81 (2001), pp. 1735–1758. DOI: [10.1016/S0006-3495\(01\)75825-0](https://doi.org/10.1016/S0006-3495(01)75825-0).
- [82] Max Kuhn e Kjell Johnson. *Applied Predictive Modeling*. Ed. por Springer. First Edition. 2013. ISBN: 978-1-4614-6849-3.

- [83] Daniel W. Pierce e Steven G. Boxer. “Stark effect spectroscopy of tryptophan”. Em: *Biophysical Journal* 68 (1995), pp. 1583–1591. DOI: 10.1016/S0006-3495(95)80331-0.
- [84] Radka Svobodová Vařeková, Jaroslav Koča e Chang-Guo Zhang. “Complexity and Convergence of Electrostatic and van der Waals Energies within PME and Cutoff Methods”. Em: *International Journal of Molecular Sciences* 5 (2004), pp. 154–173. URL: <https://doi.org/10.3390/i5040154>.
- [85] Leandro Martínez. *MDAnalysis. Version 17.224*. URL: <http://leandro.iqm.unicamp.br/mdanalysis>.
- [86] R.J. Siezen e J.A. Leunissen. “Subtilases: the superfamily of subtilisin-like serine proteases”. Em: *Protein science* 6 (1997), pp. 501–523. DOI: 10.1002/pro.5560060301.
- [87] R.J. Siezen, J.A.M. Leunissen e U. Shinde. *Homology analysis of the propeptides of subtilisin-like serine proteases (subtilases)*. In: *Intramolecular chaperones and folding*. Ed. por R.G. Landes Co. First Edition. 1995. ISBN: 978-1570592928.
- [88] R.J. Siezen et al. “Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases”. Em: *Protein Engineering, Design and Selection* 4 (1991), pp. 719–737. URL: <https://doi.org/10.1093/protein/4.7.719>.
- [89] Robert J. DeLange e Emil L. Smith. “Subtilisin Carlsberg. I. Amino acid composition; isolation and composition of peptides from the tryptic hydrolysate”. Em: *Journal of Biological Chemistry* 10 (1968), pp. 2134–2142. URL: <http://www.jbc.org/content/243/9/2134.long>.
- [90] Ole Kirk, Torben Vedel Borchert e Claus Crone Fuglsang. “Industrial enzyme applications”. Em: *Current Opinion in Biotechnology* 13 (2002), pp. 345–351. URL: [https://doi.org/10.1016/S0958-1669\(02\)00328-2](https://doi.org/10.1016/S0958-1669(02)00328-2).
- [91] Samir Kumar Pal, Jorge Peon e Ahmed H. Zewail. “Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution”. Em: *Proceedings of the National Academy of Sciences of the United States of America* 99 (2002), pp. 1763–1768. URL: <https://doi.org/10.1073/pnas.042697899>.
- [92] G.S. Lakshmikanth e G. Krishnamoorthy. “Solvent-Exposed Tryptophans Probe the Dynamics at Protein Surfaces”. Em: *Biophysical Journal* 77 (1999), pp. 1100–1106. URL: [https://doi.org/10.1016/S0006-3495\(99\)76960-2](https://doi.org/10.1016/S0006-3495(99)76960-2).

- [93] Ajay Kumar Shaw e Samir Kumar Pal. “Activity of Subtilisin Carlsberg in macromolecular crowding”. Em: *Journal of Photochemistry and Photobiology B: Biology* 86 (2007), pp. 199–206. DOI: <https://doi.org/10.1016/j.jphotobiol.2006.10.001>.
- [94] Alvaro J. Lopez e Leandro Martinez. “Parametric models to compute tryptophan fluorescence wavelength from classical protein simulations”. Em: *Journal of Computational Chemistry* 39 (2018), pp. 1249–1258. URL: <https://doi.org/10.1002/jcc.25188>.
- [95] Hai Du et al. “PhotochemCAD: A computer-aided design and research tool in photochemistry”. Em: *Photochemistry and Photobiology* 68 (1998), pp. 141–142. URL: <https://doi.org/10.1111/j.1751-1097.1998.tb02480.x>.
- [96] James M. Dixon Masahiko Taniguchi Jonathan S. Lindsey. “PhotochemCAD 2. A refined program with accompanying spectral databases for photochemical calculations”. Em: *Photochemistry and Photobiology* 81 (2005), pp. 212–213. URL: <https://doi.org/10.1111/j.1751-1097.2005.tb01544.x>.
- [97] Dongping Zhong et al. *Femtosecond dynamics of rubredoxin: Tryptophan solvation and resonance energy transfer in the protein*. 2002. URL: <https://doi.org/10.1073/pnas.012582399>.
- [98] Urmí Doshi e Donald Hamelberg. “Extracting realistic kinetics of rare activated processes from accelerated molecular dynamics using Kramer’s theory”. Em: *Journal of Chemical Theory and Computation* 7 (2011), pp. 575–581. DOI: [10.1021/ct1005399](https://doi.org/10.1021/ct1005399).
- [99] Emília Pécora de Barros. “Dissertação de Mestrado: Estudo computacional da solvatação e anisotropia de fluorescência resolvida no tempo de subtilisina Carlsberg em misturas de água e glicerol”. Em: (2015). URL: <http://repositorio.unicamp.br/jspui/handle/REPOSIP/248418>.

Apêndice A

RMSD

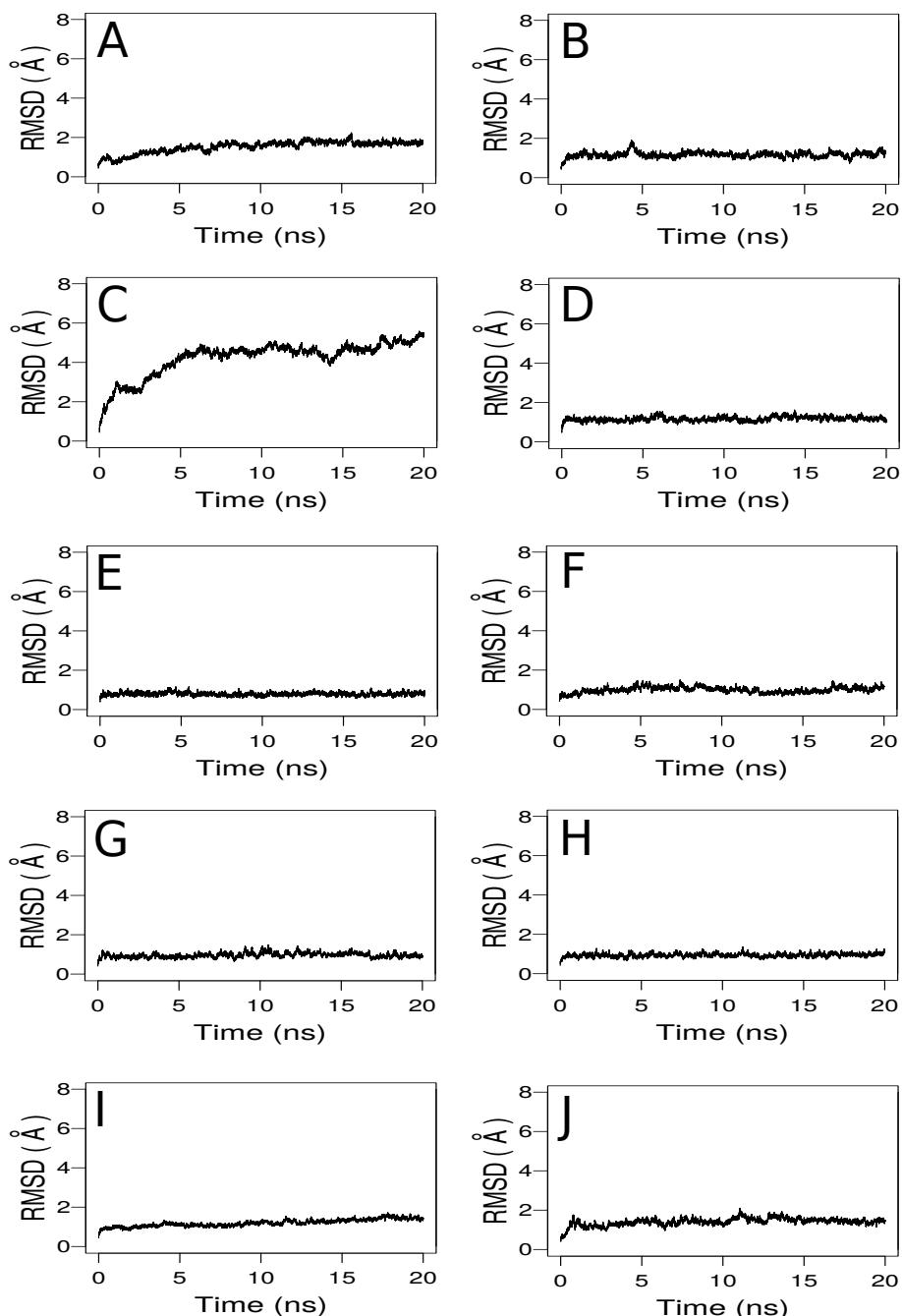


Figura A.1: Gráficas do RMSD do backbone da proteína com respeito as estructuras iniciais usadas na produção das simulações. A) Azurina, B) Parvalbumina, C) Mioglobina, D) Ribonuclease T1, E) FKBP12, F) T4-lisozima W126Y W138Y, G) T4-lisozima W126Y W158Y, H) T4-lisozima W138Y W158Y, I) subtilisina Carlsberg, J) Nucleasa de estafilococo.

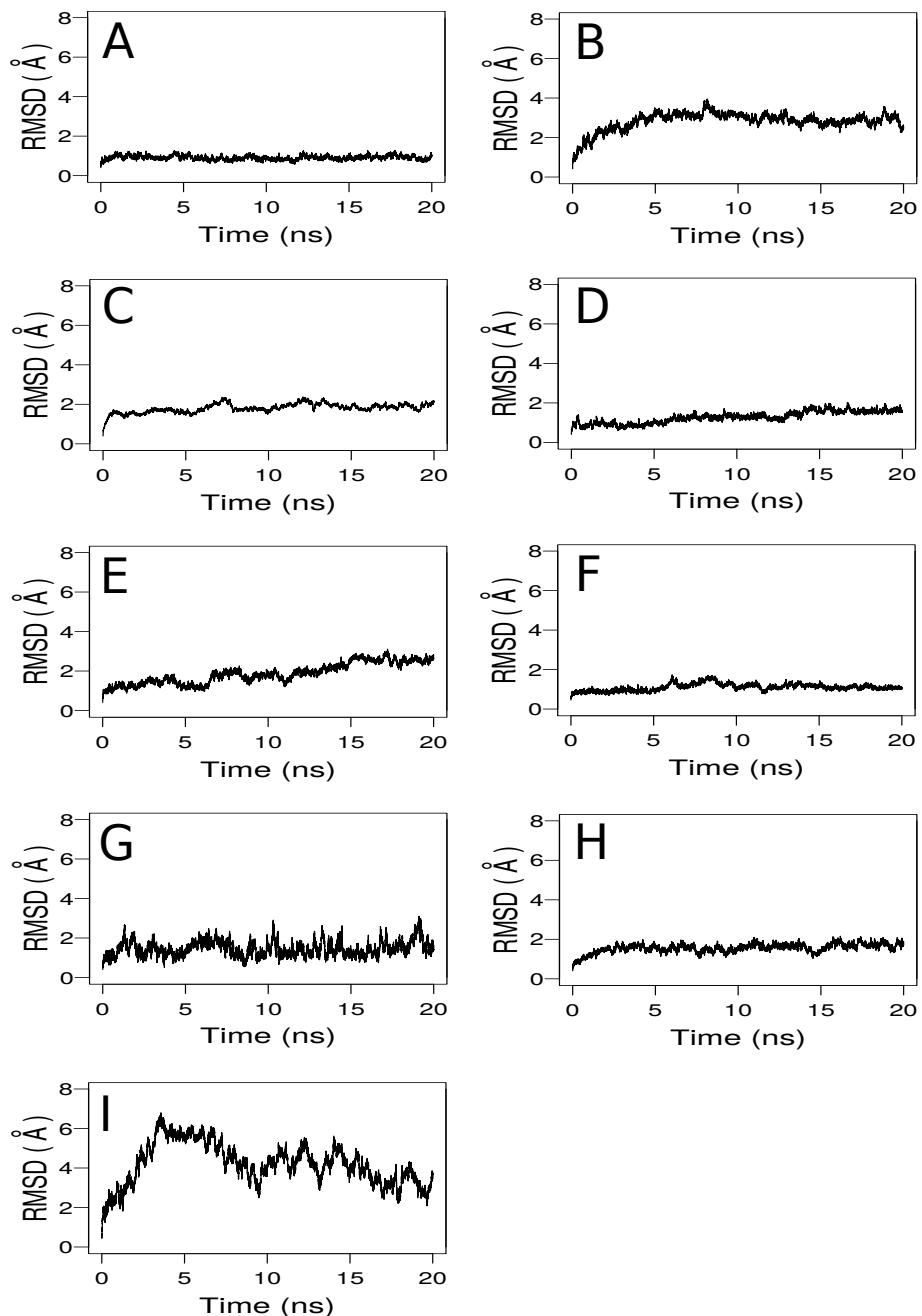


Figura A.2: Gráficas do RMSD do backbone da proteína com respeito as estructuras iniciais usadas na produção das simulações. A) Che-Y, B) α -Cobratoxina, C) HSA, D) Monelina, E) Toxina colérica B (monômero), F) Tioredoxina W31A, G) Melitina, H) Fosfolipase A2, I) Glucagon.

Apêndice B

Tabelas do Capítulo 4

Tabela B.1: Robustez do modelo baseado na SASA do benzeno incompleto e pirrol completo (Modelo 2) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 2 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	324,61	2,33	2,33(4,11)	2,33(5,29)
Parvalbumina	325,65	1,17	1,17(3,84)	1,17(4,98)
Mioglobina	327,84	0,65	0,66(2,48)	0,66(3,97)
Ribonucleasa T1	324,89	-0,43	-0,43(2,18)	-0,44(4,23)
FKBP12	327,60	-0,26	-0,26(1,98)	-0,27(3,74)
T4-lisozima W126Y W138Y	329,80	-0,03	-0,02(-1,69)	-0,008(-2,61)
T4-lisozima W126Y W158Y	325,10	-0,60	-0,61(-1,83)	-0,61(3,66)
T4-lisozima W138Y W158Y	329,63	-0,08	-0,06(-1,83)	-0,05(-2,79)
subtilisina Carlsberg	333,60	0,44	0,46(1,18)	0,48(1,99)
Nuclease de estafilococo	336,44	0,78	0,76(2,89)	0,73(-4,88)
Che-Y	326,16	-0,96	-0,97(-1,73)	-0,97(2,88)
α -Cobratoxina	339,58	-0,03	-0,02(0,75)	-0,009(1,94)
HSA	333,00	-0,43	-0,43(-1,20)	-0,42(-1,88)
Monelina	333,57	-1,20	-1,19(-1,85)	-1,18(-2,43)
Toxina colérica B	336,55	-0,88	-0,89(-2,31)	-0,91(-3,10)
Tioredoxina W31A	342,29	-1,77	-1,74(-4,83)	-1,71(-6,22)
Melitina	349,62	0,97	1,02(3,55)	1,07(8,78)
Fosfolipase A2	352,52	1,68	1,73(4,03)	1,78(9,92)
Glucagon	352,55	0,24	0,31(3,29)	0,39(10,39)

Tabela B.2: Robustez do modelo baseado na SASA do benzeno completo e pirrol incompleto (Modelo 3) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 3 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	324,68	2,36	2,36(4,15)	2,35(5,33)
Parvalbumina	325,68	1,18	1,17(3,87)	1,17(5,01)
Mioglobina	327,86	0,64	0,65(2,47)	0,65(3,96)
Ribonucleasa T1	324,94	-0,42	-0,42(2,22)	-0,43(4,29)
FKBP12	328,01	-0,24	-0,25(2,12)	-0,25(3,97)
T4-lisozima W126Y W138Y	329,49	-0,12	-0,11(-2,03)	-0,09(-3,09)
T4-lisozima W126Y W158Y	325,12	-0,60	-0,61(-1,83)	-0,61(3,69)
T4-lisozima W138Y W158Y	329,32	-0,17	-0,16(-2,19)	-0,15(-3,21)
subtilisina Carlsberg	333,18	0,45	0,47(-1,30)	0,48(2,06)
Nuclease de estafilococo	336,25	0,59	0,58(2,63)	0,57(4,21)
Che-Y	326,17	-0,96	-0,97(-1,73)	-0,98(2,90)
α -Cobratoxina	339,78	-0,02	-0,01(0,81)	0,01(2,05)
HSA	333,29	-0,38	-0,38(-1,04)	-0,38(-1,72)
Monelina	333,39	-1,29	-1,29(-1,83)	-1,29(-2,62)
Toxina colérica B	336,87	-0,99	-1,01(-2,44)	-1,02(-3,24)
Tioredoxina W31A	342,05	-1,48	-1,46(-4,31)	-1,44(-5,17)
Melitina	349,67	0,98	1,03(3,66)	1,09(8,99)
Fosfolipase A2	352,67	1,76	1,80(4,16)	1,86(10,18)
Glucagon	352,58	0,25	0,32(3,39)	0,39(10,61)

Tabela B.3: Robustez do modelo baseado na SASA do benzeno completo e pirrol completo (Modelo 4) em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 4 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	324,69	2,33	2,33(4,09)	2,33(5,28)
Parvalbumina	325,67	1,17	1,17(3,85)	1,16(4,98)
Mioglobina	327,91	0,64	0,65(2,46)	0,65(3,94)
Ribonucleasa T1	324,94	-0,41	-0,42(2,19)	-0,43(4,24)
FKBP12	328,22	-0,19	-0,19(2,02)	-0,21(3,75)
T4-lisozima W126Y W138Y	329,41	-0,12	-0,11(-1,91)	-0,09(-2,86)
T4-lisozima W126Y W158Y	325,12	-0,59	-0,60(-1,82)	-0,61(3,66)
T4-lisozima W138Y W158Y	329,24	-0,17	-0,15(-2,07)	-0,14(-2,97)
subtilisina Carlsberg	333,03	0,37	0,39(-1,23)	0,41(2,07)
Nuclease de estafilococo	335,99	0,59	0,56(2,75)	0,53(-5,05)
Che-Y	326,15	-0,96	-0,97(-1,73)	-0,97(2,89)
α -Cobratoxina	339,90	-0,01	0,01(0,86)	0,02(2,18)
HSA	333,49	-0,38	-0,38(-1,12)	-0,38(-1,78)
Monelina	333,37	-1,25	-1,24(-1,84)	-1,23(-2,49)
Toxina colérica B	336,99	-0,89	-0,91(-2,48)	-0,93(-3,27)
Tioredoxina W31A	341,72	-1,93	-1,91(-4,85)	-1,89(-6,44)
Melitina	349,72	1,01	1,05(3,78)	1,11(9,59)
Fosfolipase A2	352,77	1,82	1,87(4,32)	1,93(10,89)
Glucagon	352,65	0,28	0,36(3,59)	0,45(11,34)

Apêndice C

Tabelas do Capítulo 5

Tabela C.1: Robustez do Modelo 2 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 2 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	318,04	2,56	2,58(4,25)	2,60(5,36)
Parvalbumina	316,24	0,12	0,16(5,31)	0,21(7,20)
Mioglobina	328,28	1,36	1,35(2,53)	1,33(3,44)
Ribonuclease T1	325,49	-0,26	-0,26(1,50)	-0,26(2,62)
FKBP12	323,93	-1,13	-1,12(-1,99)	-1,11(-3,20)
T4-lisozima W126Y W138Y	333,97	0,32	0,32(1,29)	0,31(2,21)
T4-lisozima W126Y W158Y	323,81	-2,67	-2,71(-4,69)	-2,75(-7,88)
T4-lisozima W138Y W158Y	331,06	0,12	0,15(1,64)	0,17(3,11)
subtilisina Carlsberg	333,85	0,17	0,18(1,04)	0,19(1,69)
Nuclease de estafilococo	339,97	0,83	0,82(1,93)	0,82(2,84)
Che-Y	328,58	-0,49	-0,49(-1,32)	-0,49(-2,01)
α -Cobratoxina	339,57	-0,04	-0,04(-0,63)	-0,04(1,37)
HSA	340,46	0,03	0,04(-0,54)	0,05(1,24)
Monelina	339,27	-0,19	-0,19(-0,77)	-0,18(-1,11)
Toxina colérica B	336,44	-0,50	-0,51(-1,01)	-0,51(-1,52)
Tioredoxina W31A	344,30	-0,34	-0,37(-3,28)	-0,41(-4,71)
Melitina	348,40	0,59	0,64(2,01)	0,69(6,56)
Fosfolipase A2	346,10	-0,27	-0,26(-0,97)	-0,25(-1,60)
Glucagon	353,25	0,84	0,92(4,98)	1,01(10,33)

Tabela C.2: Robustez do Modelo 3 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 3 nm indicando que o Modelo 3 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	317,70	2,61	2,63(4,55)	2,65(6,06)
Parvalbumina	317,10	0,72	0,75(5,89)	0,78(8,55)
Mioglobina	332,54	0,82	0,82(1,82)	0,83(2,45)
Ribonuclease T1	329,46	0,24	0,27(1,72)	0,29(4,49)
FKBP12	324,68	-1,29	-1,29(-2,13)	-1,27(-3,39)
T4-lisozima W126Y W138Y	335,65	0,36	0,36(1,17)	0,36(1,97)
T4-lisozima W126Y W158Y	327,36	-0,88	-0,91(-3,96)	-0,95(-6,51)
T4-lisozima W138Y W158Y	326,24	-0,42	-0,43(-1,38)	-0,44(-2,86)
subtilisina Carlsberg	329,26	-0,31	-0,34(-2,23)	-0,37(-5,11)
Nuclease de estafilococo	340,97	0,94	0,94(2,27)	0,93(3,39)
Che-Y	327,86	-0,60	-0,61(-1,24)	-0,62(-2,25)
α -Cobratoxina	341,85	0,19	0,20(1,66)	0,21(2,54)
HSA	334,10	-1,93	-1,979(-3,98)	-2,00(-6,45)
Monelina	337,83	-0,28	-0,29(-0,97)	-0,29(-1,64)
Toxina colérica B	339,23	-0,75	-0,74(-1,69)	-0,74(-2,50)
Tioredoxina W31A	351,94	2,15	2,14(3,50)	2,13(4,42)
Melitina	344,06	-0,24	-0,23(1,32)	-0,23(2,48)
Fosfolipase A2	347,71	-0,08	-0,06(2,47)	-0,04(3,56)
Glucagon	345,45	-0,97	-0,97(-1,82)	-0,97(-2,63)

Tabela C.3: Robustez do Modelo 4 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 6 nm indicando que o Modelo 4 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	317,20	2,51	2,50(4,52)	2,49(5,55)
Parvalbumina	317,18	0,80	0,85(5,93)	0,89(7,18)
Mioglobina	329,82	2,49	2,43(3,49)	2,37(4,57)
Ribonuclease T1	325,73	-0,62	-0,63(-2,05)	-0,65(-3,68)
FKBP12	325,90	-1,11	-1,09(-2,28)	-1,06(-3,74)
T4-lisozima W126Y W138Y	331,82	0,39	0,35(-1,48)	0,30(-3,16)
T4-lisozima W126Y W158Y	324,40	-5,91	-5,89(-9,94)	-5,89(-12,15)
T4-lisozima W138Y W158Y	329,39	-0,12	-0,11(1,71)	-0,11(3,03)
subtilisina Carlsberg	330,82	-0,05	-0,09(1,52)	-0,14(-4,22)
Nuclease de estafilococo	338,61	0,79	0,75(1,78)	0,72(3,24)
Che-Y	330,20	-1,09	-1,15(-4,94)	-1,21(-6,69)
α -Cobratoxina	339,02	-0,19	-0,19(-1,92)	-0,18(-2,87)
HSA	337,10	-1,34	-1,38(-3,46)	-1,42(-5,02)
Monelina	336,14	-2,15	-2,13(-3,48)	-2,13(-4,63)
Toxina colérica B	339,55	-1,26	-1,24(-2,25)	-1,23(-4,16)
Tioredoxina W31A	345,90	0,62	0,56(3,73)	0,49(6,16)
Melitina	348,26	0,56	0,71(4,44)	0,86(8,59)
Fosfolipase A2	348,65	0,20	0,30(2,62)	0,41(6,49)
Glucagon	355,32	4,05	4,14(8,08)	4,24(13,41)

Tabela C.4: Robustez do Modelo 5 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 4 nm indicando que o Modelo 5 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	318,88	2,74	2,74(3,78)	2,74(5,35)
Parvalbumina	314,24	-0,85	-0,81(4,49)	-0,76(6,69)
Mioglobina	334,16	0,85	0,86(1,73)	0,86(2,37)
Ribonuclease T1	327,42	-0,08	-0,07(1,86)	-0,06(3,27)
FKBP12	322,00	-1,37	-1,39(-2,78)	-1,41(-4,26)
T4-lisozima W126Y W138Y	335,31	1,02	0,98(1,58)	0,94(-3,46)
T4-lisozima W126Y W158Y	332,27	0,14	0,13(1,19)	0,12(2,08)
T4-lisozima W138Y W158Y	327,29	-0,38	-0,43(-2,12)	-0,47(-4,28)
subtilisina Carlsberg	332,70	0,10	0,09(1,08)	0,08(1,93)
Nuclease de estafilococo	342,96	1,06	1,06(2,54)	1,06(3,78)
Che-Y	329,26	-0,64	-0,64(-1,65)	-0,64(-3,17)
α -Cobratoxina	340,96	0,16	0,17(-1,67)	0,18(3,29)
HSA	336,73	-3,37	-3,38(-7,74)	-3,38(-11,25)
Monelina	338,05	-0,26	-0,28(-0,96)	-0,30(-2,29)
Toxina colérica B	336,84	-0,61	-0,62(-1,77)	-0,62(-2,68)
Tioredoxina W31A	350,41	1,52	1,50(3,67)	1,48(5,05)
Melitina	342,09	-0,83	-0,93(-4,32)	-1,02(-6,10)
Fosfolipase A2	344,43	-0,59	-0,59(-2,59)	-0,59(-4,04)
Glucagon	344,97	-1,90	-1,88(-3,22)	-1,86(-4,97)

Tabela C.5: Robustez do Modelo 6 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 4 nm indicando que o Modelo 6 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	318,00	2,60	2,62(4,52)	2,63(5,82)
Parvalbumina	316,71	0,41	0,43(5,79)	0,45(7,48)
Mioglobina	328,52	2,34	2,31(3,91)	2,28(5,09)
Ribonuclease T1	326,13	-0,45	-0,42(-2,49)	-0,40(4,45)
FKBP12	323,77	-1,32	-1,33(-2,37)	-1,33(-3,45)
T4-lisozima W126Y W138Y	332,77	0,64	0,58(1,29)	0,51(-4,99)
T4-lisozima W126Y W158Y	323,68	-2,84	-3,06(-8,84)	-3,29(-11,51)
T4-lisozima W138Y W158Y	330,30	0,07	0,04(1,57)	0,02(3,39)
subtilisina Carlsberg	333,65	0,17	0,17(1,05)	0,18(1,74)
Nuclease de estafilococo	340,30	0,94	0,95(2,32)	0,96(3,64)
Che-Y	329,30	-2,21	-2,24(-3,55)	-2,27(-5,66)
α -Cobratoxina	340,62	0,22	0,23(-2,39)	0,23(-4,22)
HSA	337,97	-2,38	-2,38(-3,84)	-2,37(-5,64)
Monelina	339,07	-0,63	-0,62(-2,97)	-0,61(-4,47)
Toxina colérica B	336,82	-1,93	-1,93(-3,17)	-1,93(-4,42)
Tioredoxina W31A	344,66	-0,17	-0,21(-3,24)	-0,25(-5,94)
Melitina	347,47	0,85	0,82(2,06)	0,79(5,91)
Fosfolipase A2	346,82	-0,29	-0,26(-1,76)	-0,22(3,99)
Glucagon	354,42	3,48	3,65(13,26)	3,84(16,39)

Tabela C.6: Robustez do Modelo 7 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 5 nm indicando que o Modelo 7 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	321,29	2,84	2,85(4,17)	2,86(5,49)
Parvalbumina	314,52	-0,89	-0,85(-6,26)	-0,81(-8,51)
Mioglobina	335,80	0,93	0,92(1,82)	0,92(2,56)
Ribonuclease T1	327,34	-0,10	-0,09(2,24)	-0,08(-4,56)
FKBP12	321,01	-2,01	-2,04(-4,37)	-2,08(-5,87)
T4-lisozima W126Y W138Y	333,55	0,62	0,55(-1,85)	0,46(-6,85)
T4-lisozima W126Y W158Y	336,11	0,40	0,38(1,25)	0,35(2,19)
T4-lisozima W138Y W158Y	327,84	-0,31	-0,37(-2,56)	-0,44(-5,68)
subtilisina Carlsberg	335,26	0,27	0,24(1,09)	0,22(-2,49)
Nuclease de estafilococo	344,03	1,42	1,42(2,69)	1,41(4,18)
Che-Y	331,57	-0,31	-0,31(-1,36)	-0,31(-2,92)
α -Cobratoxina	340,75	0,13	0,15(-1,72)	0,16(-3,73)
HSA	335,63	-4,56	-4,59(-10,63)	-4,63(-14,45)
Monelina	337,70	-0,29	-0,32(-1,22)	-0,35(-2,98)
Toxina colérica B	335,28	-0,76	-0,76(-1,88)	-0,77(-3,09)
Tioredoxina W31A	348,09	0,76	0,74(3,17)	0,73(5,07)
Melitina	338,62	-1,16	-1,28(-5,24)	-1,39(-7,06)
Fosfolipase A2	341,68	-0,90	-0,90(-2,92)	-0,90(-4,49)
Glucagon	344,93	-1,95	-1,92(-3,85)	-1,89(-5,79)

Tabela C.7: Robustez do Modelo 8 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. λ_{em}^{calc} é o comprimento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. $\Delta\lambda_{em}^{calc}(1)$ é o desvio dos comprimentos de onda quando é removida uma proteína à vez dos dados de treinamento. $\Delta\lambda_{em}^{calc}(2)$ e $\Delta\lambda_{em}^{calc}(3)$ representam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mostrados em parênteses. Os desvios e média dos desvios são menores que 6 nm indicando que o Modelo 8 é robusto em relação aos dados de treinamento usados. A unidade do comprimento de onda é nm.

Proteína	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurina	319,21	2,71	2,73(5,06)	2,76(6,39)
Parvalbumina	317,55	1,13	1,20(7,21)	1,28(9,31)
Mioglobina	327,73	2,52	2,42(4,65)	2,32(6,28)
Ribonuclease T1	327,62	-0,08	-0,05(2,39)	-0,02(5,17)
FKBP12	321,76	-1,92	-1,93(-2,99)	-1,94(-4,11)
T4-lisozima W126Y W138Y	334,23	0,83	0,76(1,48)	0,68(-2,91)
T4-lisozima W126Y W158Y	323,92	-5,62	-5,66(-9,29)	-5,71(-12,32)
T4-lisozima W138Y W158Y	330,39	0,08	0,04(-1,64)	-0,01(3,47)
subtilisina Carlsberg	335,49	0,38	0,34(1,15)	0,31(2,11)
Nuclease de estafilococo	342,23	1,27	1,25(2,69)	1,22(3,59)
Che-Y	329,14	-1,95	-1,99(-5,42)	-2,04(-7,21)
α -Cobratoxina	343,75	1,31	1,35(3,46)	1,39(4,51)
HSA	336,24	-4,39	-4,48(-6,76)	-4,58(-8,86)
Monelina	341,23	-0,26	-0,37(-2,70)	-0,35(-7,63)
Toxina colérica B	333,87	-1,85	-1,89(-3,33)	-1,94(-5,64)
Tioredoxina W31A	345,08	0,02	-0,06(-3,22)	-0,14(-6,79)
Melitina	345,57	-0,25	-0,31(-3,22)	-0,39(-6,17)
Fosfolipase A2	345,47	-0,73	-0,71(-2,27)	-0,69(-3,07)
Glucagon	350,53	-1,19	-1,09(-5,78)	-0,98(-10,59)