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

Lopez Ayme, Alvaro Jhovaldo, 1988-

L881e LopEstudo computacional da fluorescência do triptofano em proteínas por meio de simulações de dinâmica molecular / Alvaro Jhovaldo Lopez Ayme. –

Campinas, SP : [s.n.], 2019.

LopOrientador: Leandro Martínez.

LopTese (doutorado) – Universidade Estadual de Campinas, Instituto de

Química.

Lop1. 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

Powered by TCPDF (www.tcpdf.org)

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 (*λem*) do triptofano em proteínas é muito va riável. Existem proteínas com *λem* baixos (*e.g.* Azurina, 308 nm), e outras com *λem* altos (*e.g.* Glucagon, 355 nm). Essa diversidade de *λem* 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 pre dizer os *λem*. 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 desenvol vidos modelos clássicos parametrizados baseados na área acessível ao solvente e baseados nas interações eletrostáticas para a predição dos *λem* do triptofano. Esses modelos clássicos, permi tem 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 *λem* experimentais e calculados, es tes 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 *λem* foram calculados para as estruturas amostradas nessas simulações, sendo similares aos observados nos experimentos. Como era previsto, a dinâmica reorientacional cal culada 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 (*λem*) of tryptophan in proteins is very variable. There are proteins with low *λem* (*e.g.* Azurine, 308 nm), and others with high *λem* (*e.g.* Glucagon, 355 nm). The diversity of *λem* 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 *λem*. 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 correla tion 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 *λem*. These classical models allow better predictions compared to the model based on first principles of Vivian and Cal lis (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 incon sistency 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 prefera bly 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 ex perimental and calculated *λem*, 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 *λem* were calculated for the structures sampled

Abstract

in these simulations, being similar to those observed in experiments. As expected, the calcula ted 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.

**Lista 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) amos tra, 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) fenila lanina, 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 foto multiplicador, 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 (*µa*) e emissão (*µe*). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 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 *trp* de *E.coli*. 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 *trp* de *E.coli* com respeito às condições desnaturantes do sol vente. A) Apo-represor W99F, e B) Apo-represor 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 tripto fano experimental (*λem*) 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 (*λexp*

*em* ) versus

o calculado (*λcalc*

*em* ) usando os modelos baseados na área acessível ao solvente.

A) Modelo 1: Baseado na *hSASAi* do grupo indol do triptofano. B) Modelo 2: Baseado na decomposição da *hSASAi* dos grupos benzeno incompleto e pirrol completo do indol. C) Modelo 3: Baseado na decomposição da *hSASAi* dos grupos benzeno completo e pirrol incompleto do indol. D) Modelo 4: Baseado na decomposição da *hSASAi* 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 *La* do indol. B) Correlação entre os *λem* 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 experi mental 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 (*λexp*

*em* ) versus

o calculado (*λcalc*

*em* ) usando os modelos baseados nas interações eletrostáticas.

A) Modelo 1: Baseado na interação eletrostática do grupo indol com o ambi ente (água + proteína). B) Modelo 2: Baseado nas interações eletrostáticas do grupo indol com a água e a proteína. C) Modelo 3: Baseado nas intera ções eletrostáticas do benzeno incompleto-ambiente (água + proteína) e pirrol completo-ambiente (água + proteína). D) Modelo 4: Baseado na decomposi ção 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 (*λexp*

*em* ) versus

o calculado (*λcalc*

*em* ) 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 ben zeno 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 re portado 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). . . . . . . . . . . . . . . . . . . . . 109

6.2 Fluorescência do Trp livre em água. As distribuições dos *λem* 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 re síduo 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 (*∼*5 Å). . . . . . . . . . . . . . . . . . . . . . . . . . . 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 expe rimentais (azul) e distribuições dos *λem* calculados (vermelho) do Trp113 da sC em água. Os cálculos dos *λem* foram feitos usando o modelo parametri zado baseado na SASA do benzeno/pirrol (em A) ou o modelo baseado nas interações eletrostáticas do indol com a água/proteína (em B). C) Anisotro pia de fluorescência resolvida no tempo do Trp113 da sC em água. As curvas azuis contínuas e tracejadas correspondem aos dados experimentais reporta dos 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 (*∼*12 Å). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 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 *λem* calculados (vermelho) do Trp113 da sC em água. Os cál culos dos *λem* 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 traceja das 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 ini ciais 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

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. . . . . . . . . . . . . . . . . . . . . 143

**Lista de Tabelas**

1.1 Rendimento quântico (Φ*F* ), e comprimentos de onda máximo de absorção (*λabs*) e emissão (*λem*) dos aminoácidos fenilalanina (Phe), tirosina (Tyr) e triptofano (Trp) em solução aquosa. . . . . . . . . . . . . . . . . . . . . . . . 38

1.2 Diversidade dos comprimentos de onda de emissão de fluorescência (*λem*) 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 (*λexp*

*em* ), e áreas acessíveis ao solvente cal

culadas a partir das simulações de dinâmica molecular. *SASAind*, *SASAben,i*, *SASApyr,i*, *SASAben,c*, *SASApyr,c* são as áreas acessíveis ao solvente do indol, benzeno incompleto, pirrol incompleto, benzeno completo e pirrol completo, respectivamente. *hi* representa a média sobre a simulações. O desvio padrão é mostrado. Os *λexp*

*em* estão em nm, e as SASAs em Å2. . . . . . . . . . . . . . 76

4.2 Comprimentos de onda de emissão de fluorescência experimental e calculado do triptofano: *λexp*

*em* é o comprimento de onda de emissão de fluorescência ex

perimental. *λ*(1*−*4)

*em* e ∆*λ*(1*−*4)

*em* são o comprimento de onda de emissão de fluo

rescência 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*2é o coeficiente de correlação, SD é o desvio padrão das predições com respeito aos dados experimentais. *SASAind*, *SASAben,c*, *SASAben,i*, *SASApyr,c*, *SASApyr,i* são as áreas acessíveis ao solvente do indol, benzeno completo, benzeno incom pleto, pirrol completo e pirrol incompleto, respectivamente. As unidades das SASAs são Å2. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

Lista de Tabelas

4.4 Robustez do modelo baseado na SASA do indol (Modelo 1) em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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 (*λexp*

*em* ), e interações eletros

táticas calculadas a partir das simulações de dinâmica molecular. *QT* , *Qw* e *Qp* 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 *hi* representa a média sobre as simulações. O desvio padrão é mostrado. Os *λexp*

*em* estão em

nm, e as interações eletrostáticas em *kJ × mol−*1. . . . . . . . . . . . . . . . . 90 5.2 Proteínas usadas, códigos de acesso das estruturas (PDB id.), comprimentos de onda de emissão de fluorescência experimentais (*λexp*

*em* ), e interações eletros

táticas calculadas a partir das simulações de dinâmica molecular. *Qben−i,w*, *Qben−i,p*, *Qpyr−c,w* e *Qpyr−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 *Qben−i* = *Qben−i,w* +*Qben−i,p*. Da mesma maneira, as interações eletrostáticas entre o pirrol completo e o am biente (água + proteína) são computadas como *Qpyr−c* = *Qpyr−c,w* + *Qpyr−c,p*. O símbolo *hi* representa a média sobre as simulações. O desvio padrão é mos trado. Os *λexp*

*em* estão em nm, e as interações eletrostáticas em *kJ × 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 (*λexp*

*em* ), e interações eletros

táticas calculadas a partir das simulações de dinâmica molecular. *Qben−c,w*, *Qben−c,p*, *Qpyr−i,w* e *Qpyr−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. respec tivamente. As interações eletrostáticas entre o benzeno completo e o ambiente (água + proteína) são computados como *Qben−c* = *Qben−c,w* + *Qben−c,p*. Da mesma maneira, as interações eletrostáticas entre o pirrol incompleto e o am biente (água + proteína) são computados como *Qpyr−i* = *Qpyr−i,w* + *Qpyr−i,p*. O símbolo *hi* representa a média sobre as simulações. O desvio padrão é mos trado. Os *λexp*

*em* estão em nm, e as interações eletrostáticas em *kJ × mol−*1. . . 92 5.4 Comprimentos de onda de emissão de fluorescência experimentais e calcula dos do triptofano: *λexp*

*em* é o comprimento de onda de emissão de fluorescência

experimental. *λ*(1*−*4)

*em* e ∆*λ*(1*−*4)

*em* são os comprimentos de onda de emissão de

fluorescência calculados e os desvios relativos aos valores experimentais, res pectivamente, 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 calcula dos do triptofano: *λexp*

*em* é o comprimento de onda de emissão de fluorescência

experimental. *λ*(5*−*8)

*em* e ∆*λ*(5*−*8)

*em* são o comprimento de onda de emissão de

fluorescência calculado e o desvio relativo ao valor experimental, respectiva mente, 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 compri mentos de onda de emissão de fluorescência (*λem*) do triptofano em proteínas por meio de simulações de dinâmica molecular. *R*2é o coeficiente de correlação quadrático de Pearson e SD é o desvio padrão das predições com respeito aos dados experimentais. *QT* , *Qw*, *Qp*, *Qben−i*, *Qben−c*, *Qpyr−i*, *Qpyr−c*, *Qben−i,w*, *Qben−i,p*, *Qpyr−i,w*, *Qpyr−i,p*, *Qben−c,w*, *Qben−c,p*, *Qpyr−c,w* e *Qpyr−c,p* são as inte rações 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 incom pleto - proteína, benzeno completo - água, benzeno completo - proteína, pirrol completo - água e pirrol completo - proteína, respectivamente. As unidades das interações eletrostáticas são kJ*×*mol*−*1. . . . . . . . . . . . . . . . . . . . 103

5.7 Robustez do Modelo 1 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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 (Croma tografia líquida de alta eficiência). FPLC (Cromatografia líquida de proteína rápida). *λex* e *λem* são os comprimentos de onda de excitação e emissão, res pectivamente. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 110

Lista de Tabelas

6.2 Parâmetros experimentais e calculados da anisotropia de fluorescência resol vida no tempo do triptofano livre em água. *λem* é o comprimento de onda usado para registrar os decaimentos de anisotropia. O *λem* foi calculado usando (*a*) o modelo baseado na SASA do benzeno/pirrol e (*b*) nas interações eletrostá ticas 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 *× e*(*−t/τ*1), sendo *τ*1 o tempo de reorientação do fluoróforo. Os *λem* calculados correspon dem 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 resol vida no tempo do Trp113 da subtilisina Carlsberg em água. *λem* é o com primento de onda usado para registrar os decaimentos de anisotropia. O *λem* 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 *×*(*β*1*e−t/τ*1 + *β*2*e−t/τ*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, respec tivamente. Os *λem* 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 com pleto (Modelo 2) em relação aos dados de treinamento usados. *λcalc*

*em* é o com

primento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. ∆*λcalc*

*em* (1) é o desvio dos comprimen

tos de onda quando é removida uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) representam a média dos desvios quando duas e três pro

teínas 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 incom pleto (Modelo 3) em relação aos dados de treinamento usados. *λcalc*

*em* é o com

primento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. ∆*λcalc*

*em* (1) é o desvio dos comprimen

tos de onda quando é removida uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) representam a média dos desvios quando duas e três pro

teínas 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. *λcalc*

*em* é o compri

mento de onda de emissão de fluorescência calculado usando o conjunto de 19 proteínas como dados de treinamento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos

de onda quando é removida uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) representam a média dos desvios quando duas e três pro

teínas 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

C.1 Robustez do Modelo 2 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 149

C.2 Robustez do Modelo 3 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 150

C.3 Robustez do Modelo 4 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 151

Lista de Tabelas

C.4 Robustez do Modelo 5 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 152

C.5 Robustez do Modelo 6 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 153

C.6 Robustez do Modelo 7 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 154

Lista de Tabelas

C.7 Robustez do Modelo 8 baseado nas interações eletrotáticas em relação aos dados de treinamento usados. *λcalc*

*em* é o comprimento de onda de emissão de

fluorescência calculado usando o conjunto de 19 proteínas como dados de trei namento. ∆*λcalc*

*em* (1) é o desvio dos comprimentos de onda quando é removida

uma proteína à vez dos dados de treinamento. ∆*λcalc*

*em* (2) e ∆*λcalc*

*em* (3) represen

tam a média dos desvios quando duas e três proteínas são removidas dos dados de treinamento, respectivamente. Os valores máximos dos desvios são mos trados 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. . . . . . . . . . . . . . . . 155

**Lista de Abreviaturas**

*λem* Comprimento de onda emissão de fluorescência *λexp*

*em* Comprimento de onda emissão de fluorescência experimental *λcalc*

*em* Comprimento de onda emissão de fluorescência calculado **ABF** *Adaptive Biasing Force*

**DNA** *Deoxyribonucleic acid*

**MD** *Molecular dynamics*

**NAMD** *Not (just) Another Molecular Dynamics program* **PDB** *Protein Data Bank*

*QT* Interação eletrostática entre o indol e o ambiente (água + pro teína)

*Qp* Interação eletrostática entre o indol e a proteína *Qw* Interação eletrostática entre o indol e a água **QM-MM** Simulações híbridas mecânica quântica-mecânica molecular **RMSD** *Root-mean-square deviation*

**SASA** *Solvent acessible surface area*

**sC** subtilisina Carlsberg

**Trp** Triptofano

**Trp113** Triptofano da posição 113

**TCSPC** *Time-correlated single-photon counting*

**Sum´ario**

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 aces sível ao solvente do triptofano . . . . . . . . . . . . . . . . . . . 68

4.3.2 Construção dos modelos . . . . . . . . . . . . . . . . . . . . . . 70 4.3.3 Robustez dos modelos . . . . . . . . . . . . . . . . . . . . . . . 79

Sum´ario

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 *λem* 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 *λem* . . . . . . . . . . . . . . . . . . . . . . . . . . . 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 subti lisina Carlsberg parcialmente protegido . . . . . . . . . . . . . . 117

Sum´ario

6.3.3 Decaimento da anisotropia de fluorescência do Trp113 da subti lisina 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

30

**Cap´ıtulo 1**

**Fluorescˆencia**

**1.1 Uma breve introdu¸c˜ao `a fluorescˆencia**

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 disper siva*. 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

CAP´ITULO 1. FLUORESCENCIA ˆ 31

momento a fluorescência ganhou muita importância e aplicabilidade na área científica e tecno lógica. 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 b´asicos de fluorescˆencia**

**1.2.1 Diagrama de Jablonski**

Também conhecido como o diagrama de Perrin-Jablonski, este é usado para repre sentar 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 a multiplicidade do spin.

• Fluorescência: É o processo mediante o qual há emissão de fótons devido a uma rela xação do estado excitado ao fundamental, sendo que estes estados possuem a mesma a

CAP´ITULO 1. FLUORESCENCIA ˆ 32 multiplicidade do spin. (i.e. *S*1 *→ 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 aconte cem 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 fluorescˆencia**

Ao observar o espectro de fluorescência, é possível de notar duas características im portantes. 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 relaci onadas 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 quˆantico**

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 (*kSr*) e não-radiativo (*kSnr*).

*−d*[*A∗*]

*dt* = (*kSr* + *kSnr*)[*A∗*] (1.2.1)

CAP´ITULO 1. FLUORESCENCIA ˆ 33

Por outro lado, a intensidade de fluorescência (*iF* ) é a quantidade de fótons emitidos na passagem do estado excitado ao fundamental (*A∗ → A* + *foton* ´ ), sendo proporcional à concentração de moléculas que ainda estão no estado excitado (*iF ∝ kSr*[*A∗*]). A intensidade de fluorescência observada experimentalmente (*IF* ) é proporcional a *iF* , e depende das condições instrumentais. Também é importante dizer que *IF* 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 Equa ção 1.2.2.

*Fotons* ´ *ex*=*kSr*

Φ*F* =*Fotons* ´ *em*

*kSr* + *kSnr*(1.2.2)

onde *Fotons* ´ *ex* e *Fotons* ´ *em* são os números de fótons absorvidos durante a excitação e os li berados durante a emissão de fluorescência, respectivamente. As constantes de velocidade *kSr* e *kSnr* 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 quân tico 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:

Φ*F* = Φ*RIF IF R*

*IAR*

*IA*(1.2.3)

sendo *IF* e *IF R* as intensidades de fluorescência, e *IA* e *IAR* as intensidades de absorção do fluoróforo problema e referência, respectivamente.

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

CAP´ITULO 1. FLUORESCENCIA ˆ 34

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 (*mR*). Assim, o rendimento quântico será calculado usando:

Φ*F* = Φ*RM*

*mR*(1.2.4)

O rendimento quântico de diversos fluoróforos foram estimados usando esses mé todos. 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 *∼* 0,0 até *∼* 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 (*τS*) 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.

*τS* =1

*kSr* + *kSnr*(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).

*τS* = Φ*F /kSr*(1.2.6)

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

CAP´ITULO 1. FLUORESCENCIA ˆ 35 **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 conside rado no modelo de Lippert-Mataga (Equação 1.2.7) [4].

*νA − νE* =2*hc* "*ε −* 1

2*ε −* 1*−n*2 *−* 1

2*n*2 + 1

#(*µE − µG*)2

*a*3(1.2.7)

sendo *νA* e *νE* as energias de absorção e emissão, respectivamente. *µE* e *µG* 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 Fluorescˆencia em prote´ınas**

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

CAP´ITULO 1. FLUORESCENCIA ˆ 36 **1.3.1 Fluorescˆencia estacion´aria**

**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.

CAP´ITULO 1. FLUORESCENCIA ˆ 37 **Caracter´ısticas espectrais dos amino´acidos**

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 *λabs* 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 *λabs* destes aminoácidos são próximos de 280 nm.

Nos experimentos de fluorescência as proteínas são geralmente excitadas com com primentos de onda maiores que 280 nm. Considerando que o rendimento quântico da fenilala nina é baixo (Φ*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 (*λem*) em *∼*280 nm. O *λem* da tirosina em água é localizado em 303 nm e é pouco sensível à polaridade do meio. O triptofano livre em água apresenta *λem* 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 carac terísticas 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.

CAP´ITULO 1. FLUORESCENCIA ˆ 38

Tabela 1.1: Rendimento quântico (Φ*F* ), e comprimentos de onda máximo de absorção (*λabs*) e emissão (*λem*) dos aminoácidos fenilalanina (Phe), tirosina (Tyr) e triptofano (Trp) em solução aquosa.

Aminoácido Φ*F λabs* (nm) *λem* (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]

)

1 −

0.20 0.15 0.10 0.05

0.00 1.4

1.2

A B

1.0 0.8 0.6 0.4 0.2 0.0 1.0 0.8

a i

c

n

ê

c

s

e

m

r

c

1 −

1.0 0.8 0.6

0.6 0.4

o

u

lf

M

0.4

3

0.2

0

0.2

e

d

e

1

x

( ε

0.0

5

4

3

2

1

0

C

0.0 1.0

0.8 0.6 0.4 0.2 0.0

da

di

s

n

e

t

n

I

250 300 350 400

Comprimento de onda (nm)

Figura 1.3: Espectros de absorção (vermelho) e emissão (azul) dos aminoácidos A) fenilala nina, 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 fluorescˆencia 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

CAP´ITULO 1. FLUORESCENCIA ˆ 39

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 (*λem*) 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 *λem*. Na Tabela 1.2 estão apresentados os *λem* 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 (*λem*) do trip tofano 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.

Proteina PDB id. *λem*

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 fluorescˆencia resolvida no tempo**

As medidas de fluorescência resolvida no tempo são importantes devido a que for necem 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-

CAP´ITULO 1. FLUORESCENCIA ˆ 40

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 fluores cência 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 modu lação 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 fótons 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 (*Ik*

CAP´ITULO 1. FLUORESCENCIA ˆ 41

e *I⊥*). 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 (*Ik*) e perpendicular (*I⊥*) em função do tempo.

Figura 1.4: Diagrama esquemático do fluorímetro usado nas medidas de anisotropia de fluores cência 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 an´alise 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

CAP´ITULO 1. FLUORESCENCIA ˆ 42 como:

*r*(*t*) = *Ik*(*t*) *− I⊥*(*t*)

*Ik*(*t*) + 2*I⊥*(*t*)(1.3.1)

sendo *Ik*(*t*) e *I⊥*(*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 (*θ* = 0*◦*), o fluoróforo terá probabilidade máxima de ser excitado; enquanto que quando a orientação do campo elétrico é perpendicular ao dipolo de absorção (*θ* = 90*◦*), o fluoróforo tem probabilidade mínima de excitação. Quanti tativamente, a probabilidade de excitação de um fluoróforo pela luz polarizada é proporcional a cos2*θA*, sendo *θA* 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].

1.0

o

ã

çr

o

s

b

a

e

d

.

b

o

r

P

0.8 0.6 0.4 0.2 0.0

A

0 10 20 30 40 50 60 70 80 90

Ângulo ( °)

A

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 ∆*t*, os momentos de transição de emissão, ***µ****E*, das moléculas excitadas possuirão uma distribuição angular caracterizada pelo ângulo *θE*, o qual é definido em relação ao

CAP´ITULO 1. FLUORESCENCIA ˆ 43

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*) = 3cos2 *θE*(*t*) *−* 1

2(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), *θA* = *θE* = *θ*, portanto cos2*θA* = cos2*θE* = cos2*θ*, e considerando que o nú mero de moléculas com momentos de transição orientados entre *θ* e *θ* + ∆*θ* antes da excitação seja isotropicamente distribuído, é possível determinar o valor de cos2 *θ* como sendo de 3/5. Substituindo esse valor, do cos2 *θ*, na Equação 1.3.2, chegamos num valor de anisotropia de 0,4.

*r*0 =3cos2 *θ −* 1

2=25= 0*,*4 (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 asso ciada 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 (Fi gura 1.6C). Nesse caso, é possível de expressar a anisotropia em função dos ângulos *θA* e *α* (Equação 1.3.4).

*r*0 =3cos2 *θE −* 1

2=3cos2 *θA −* 1

2*×*3cos2 *α −* 1

2(1.3.4)

Como foi mencionado anteriormente, cos2 *θA* = 3*/*5, então a anisotropia toma a

CAP´ITULO 1. FLUORESCENCIA ˆ 44 seguinte forma.

*r*0 =253cos2 *α −* 1

2(1.3.5)

Assim de acordo com a Equação 1.3.5, a anisotropia de fluorescência assumirá valores entre -0,2 (*α* = 90*◦*, momentos de transição perpendiculares) e 0,4 (*α* = 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, res pectivamente), 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 *ω*(*t*), que é o ângulo formado pelo momento de transição de emissão no tempo *t* (***µ****e*(*t*)) com respeito ao tempo *t* = 0 (***µ****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-

CAP´ITULO 1. FLUORESCENCIA ˆ 45 tropia em função dos ângulos *θA*(0), *α* e *ω*(*t*) (Equação 1.3.6).

*r*(*t*) = 3cos2 *θE*(*t*) *−* 1

2=3cos2 *θA*(0) *−* 1

2*×*3cos2 *α −* 1

2*×*3cos2 *ω*(*t*) *−* 1

2(1.3.6)

O termo 3cos2 *θA*(0)*−*1

2 *×*3cos2 *α−*1

2é 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*03cos2 *ω*(*t*) *−* 1

2(1.3.7)

O termo 3cos2 *ω*(*t*)*−*1

2é conhecido como a função de autocorrelação das orientações

do momento de transição da emissão. Considerando que a quantidade 3*x*2*−*1

2é o polinômio de

Legendre de segunda ordem (*P*2(*x*)), podemos expressar alternativamente a anisotropia como [49, 50, 51]:

*r*(*t*) = *r*0 *hP*2[cos *ω*(*t*)]*i* (1.3.8)

sendo *hi* a média sobre todas as moléculas excitadas. Para fluoróforos caracterizados com decai mentos do tipo monoexponencial, podemos expressar o polinômio de Legendre da Equação 1.3.8 em função do coeficiente de difusão rotacional do fluoróforo (*Dr*), usando *hP*2[cos *ω*(*t*)]*i* = exp(6*Drt*) [1].

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

*r*(*t*) = 25*hP*2[*µa*(0)*µe*(*t*)]*i* (1.3.9)

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

CAP´ITULO 1. FLUORESCENCIA ˆ 46 **A B**

**z**

**x**

**y**

**z**

=

==

**x**

**y**

**C D z**

**x**

**z**

**x**

**y**

**y**

Figura 1.6: Diagramas das orientações dos momentos de dipolo de absorção (*µa*) e emissão (*µe*).

expressado como um decaimento monoexponencial:

*r*(*t*) = *r*0*e−t/τrot* (1.3.10)

sendo *r*0 o valor da anisotropia medido no instante da fotoseleção (anisotropia fundamental) e *τrot* 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*0X *i*=1

*βie−t/τrot,i* (1.3.11)

sendo *βi* as contribuições populacionais do decaimento de anisotropia com tempos de correlação *τrot,i*.

47

**Cap´ıtulo 2**

**Simula¸c˜oes de dinˆamica 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 (*<* 102 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 (*>* 103 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 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 Tidesley (1987) [52], e Allen (2004) [53].

**2.1 Simula¸c˜oes de dinˆamica 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 *mi*, que está baixo a influência de uma força *~fi*, a dinâmica molecular tem

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 48 como objetivo resolver a equação de Newton,

*dt*2=*~fi*(*~r*1*, ~r*2*, ...*)

*d*2*~ri*

sendo a força dependente das posições (*~ri*).

*mi*(2.1.1)

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 (∆*t*) 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 veloci dades dos átomos para o instante de tempo *t* + ∆*t*.

4. Repetir os passos 2 ao 4 até alcançar o tempo total desejado .

**2.1.1 Condi¸c˜oes iniciais**

Para iniciar uma simulação de dinâmica molecular é necessário definir o conjunto de posições iniciais (*~r*1(0)*, ~r*2(0)*, ...*) e velocidades iniciais (*~v*1(0)*, ~v*2(0)*, ...*) 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 predições *ab-initio* [59].

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

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 49 seguindo uma distribuição de Maxwell-Boltzmann (Equação 2.1.2).

*p*(*vix*) =

*mi*

2*πkBT*

1*/*2exp "*−*12*miv*2*ix kBT*

#

(2.1.2)

sendo *p*(*vix*) a probabilidade de um determinado átomo *i* de massa *mi* numa temperatura *T* possua uma velocidade *vix*.

**2.1.2 Potenciais de intera¸c˜ao**

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).

*~fi* = *−∂∂riU*(*~r*1*, ~r*2*, ...*) (2.1.3)

Para isso é necessário especificar o potencial de interação (*U*). Esse potencial cos tuma ser representado como uma soma de contribuições não-ligantes (*UNLIG*) e ligantes (*ULIG*) (Equação 2.1.4).

*U* = *UNLIG* + *ULIG* (2.1.4)

A continuação detalharemos a forma como são comumente apresentados esses po tenciais.

**Potenciais n˜ao-ligantes**

1. Potencial de Lennard-Jones

Este potencial é caracterizado pela presença de um termo repulsivo (*∼* 1*/r*12) que atua a curtas distâncias e outro atrativo que possui maior alcance (*∼* 1*/r*6). A forma

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 50 deste potencial é:

*VLJ* =X *ij*

4*εij*

 

*σij rij*

!12

*−*

*σij rij*

!6

 (2.1.5)

sendo *εij* , *σij* e *rij* a profundidade do potencial, a distância na qual o potencial é zero e a distância entre os átomos *i* e *j*, respectivamente. O fator *∼* 1*/r*6 pretende representar as interações de van der Waals, enquanto que o fator *∼* 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 *~r*1*, ~r*2*, ...~rN* e que possuem cargas *q*1*, q*2*, ...qN* , o potencial de Coulomb (ou eletrostático) desse sistema é descrito por:

*Velec* =X *i,j*

*qiqj*

4*πε*0*rij*(2.1.6)

sendo *qi* e *qj* as cargas parciais dos átomos *i* e *j*, *rij* 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 (*∼* 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

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 51

contribuições: potencial de curto alcance (*ES*) e potencial de longo alcance (*EL*). 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).

*Velec* = *ES* + *EL*(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).

*UNLIG* =X *i,j*

4*εij*

 

*σij rij*

!12

*−*

*σij rij*

!6 

+X *i,j*

*qiqj*

4*πε*0*rij*

(2.1.8)

**Potenciais ligantes**

| {z } *VLJ*

| {z } *Velec*

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:

*Vest* =12X *liga*çõ*es*

*kbij* (*rij − r*0)2(2.1.9)

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 52

sendo *kbij* , *rij* , 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 *kbij* 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 descritos 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:

*Vang* =12X â*ngulos*

*kθijk*(*θijk − θ*0)2(2.1.10)

sendo *kθijk*, *θijk* 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 *kθijk* e *θijk* são definidos para cada grupo de átomos *ijk*.

3. Potencial diedral

Três pontos no espaço definem um plano. Analogamente três átomos também de finem 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 *jkl*. O potencial dihedral está relacionado à flexibilidade (torsão) do ângulo formado pelos planos dos átomos *ijk* e *jkl*.

*Vdih* =12X *dihedros*

*kϕ*[1 + cos(*nijklϕijkl − δ*)] (2.1.11)

sendo *kϕ*, a constante da força; *nijkl*, a multiplicidade (a qual indica o número de mínimos quando a ligação rota 360*o*). A multiplicidade é um número inteiro positivo. *ϕijkl*, é o ângulo entre os planos formados pelos átomos *ijk* e *jkl*; e *δ* é a fase do ângulo.

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

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 53

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).

*ULIG* =12X *liga*çõ*es*

*kb*(*r − r*0)2

+12X â*ngulos*

*kθ*(*θ − θ*0)2

+12X *dihedros*

*kϕ*[1 + *cos*(*nϕ − δ*)]

(2.1.12)

| {z } *Vest*

| {z } *Vang*

| {z } *Vdih*

Existem ainda outros potenciais ligantes, além dos citados acima, que são: o poten cial 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 harmónico e tem como finalidade restringir o movimento dos átomos (Equa ção 2.1.13).

*VUrey−Bradley* = *kUB*(*u − u*0)2(2.1.13)

sendo *kUB* 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

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 54 moléculas, e também é descrito por uma função harmónica [64] (Equação 2.1.14).

*Vimpr*ó*pios* = *kimp*(*φ − φ*0)2(2.1.14)

sendo *kimp* a constante da força, *φ* o ângulo entre os planos formados pelos átomos *ijk* e *jkl*, 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 equa¸c˜oes de movimento**

Até o momento temos as condições iniciais dos átomos (definidas pelas suas posi ções 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 aproxima ções 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* + ∆*t* em função do instante de tempo *t* e mostrando apenas até a quarta ordem da expansão temos:

*dt* ∆*t* +12!*d*2*~ri*(*t*)

*~ri*(*t* + ∆*t*) = *~ri*(*t*) + *d~ri*(*t*)

*dt*2 ∆*t*2 +13!*d*3*~ri*(*t*)

*dt*3 ∆*t*3 + *O*(∆*t*4) (2.1.15)

Analogamente, podemos expandir a função posição no instante de tempo anterior, *t −* ∆*t*, em função da posição atual, *t*.

*dt* ∆*t* +12!*d*2*~ri*(*t*)

*~ri*(*t −* ∆*t*) = *~ri*(*t*) *−d~ri*(*t*)

*dt*2 ∆*t*2 *−*13!*d*3*~ri*(*t*)

*dt*3 ∆*t*3 + *O*(∆*t*4) (2.1.16)

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 55

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).

*~ri*(*t* + ∆*t*) = 2*~ri*(*t*) *− ~ri*(*t −* ∆*t*) + *d*2*~ri*(*t*)

*dt*2 ∆*t*2 + *O*(∆*t*4) (2.1.17)

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

*dt*~~2~~ = *−f~*(*ri*(*t*))

Sabendo que *d*2*~ri*(*t*)

*mi*e desprezando os termos de quarta (ou maiores)

ordem, temos o algoritmo de Verlet (Equação 2.1.18).

O algoritmo Verlet

*~ri*(*t* + ∆*t*) = 2*~ri*(*t*) *− ~ri*(*t −* ∆*t*) +*~f*(*ri*(*t*))

*mi*∆*t*2(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 *~ri*(0) e *~vi*(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* = ∆*t* é:

*dt* ∆*t* +12!*d*2*~ri*(0)

*~ri*(∆*t*) *≈* 2*~ri*(0) + *d~ri*(0)

*dt*2 ∆*t*2(2.1.19)

Com as posições *~ri*(∆*t*) é possível computar as forças *~f*(*ri*(∆*t*)). Após isso o cálculo pode ser feito iterativamente usando a Equação 2.1.18.

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 56

Implementação do algoritmo de Verlet

1. Definidos as condições iniciais, *~ri*(0), *~vi*(0) e *~ai*(0), calculamos o *~ri*(∆*t*): *dt* ∆*t* +12*d*2*~ri*(0)

*~ri*(∆*t*) *≈* 2*~ri*(0) + *d~ri*(0)

2. Calcular as novas forças, e assim *~ai*(∆*t*).

*dt*2 ∆*t*2

*~ai*(∆*t*) = *−*1*mi∇Vi*(*~ri*(∆*t*))

3. Calcular as novas posições iterativamente usando o algoritmo de Verlet. *~ri*(*t* + ∆*t*) = 2*~ri*(*t*) *− ~ri*(*t −* ∆*t*) +*~f*(*ri*(*t*))

*mi*∆*t*2

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

Como vimos anteriormente, a função *~ri*(*t* + ∆*t*) pode ser expandida em função de *~ri*(*t*) (Equação 2.1.15). A expansão considerando as derivadas de segunda ordem é:

*dt* ∆*t* +12!*d*2*~ri*(*t*)

*~ri*(*t* + ∆*t*) = *~ri*(*t*) + *d~ri*(*t*)

*dt*2 ∆*t*2(2.1.20)

sendo *v*(*t*) = *d~ri*(*t*)*/dt* e *a*(*t*) = *d*2*~ri*(*t*)*/dt*2, podemos reescrever a Equação 2.1.20 como:

*~ri*(*t* + ∆*t*) = *~ri*(*t*) + *v*(*t*)∆*t* +12!*a*(*t*)∆*t*2(2.1.21)

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

*dt* ∆*t* +12!*d*2*~vi*(*t*)

*~vi*(*t* + ∆*t*) = *~vi*(*t*) + *d~vi*(*t*)

*dt*2 ∆*t*2(2.1.22)

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 57 O termo *d*2*~vi*(*t*)*/dt*2 pode ser expressado em função de derivadas de ordem inferior.

*dt* +*d*2*~vi*(*t*)

*dt* =*d~vi*(*t*)

*d~vi*(*t* + ∆*t*)

*dt*2 ∆*t* + *O*∆*t*2(2.1.23)

Isolando o termo *d*2*~vi*(*t*)*/dt*2 da Equação 2.1.23, e substituindo este na Equação 2.1.22 temos:

*~vi*(*t* + ∆*t*) = *~vi*(*t*) + 12*d~vi*(*t* + ∆*t*)

*dt* ∆*t* +12*d~vi*(*t*)

*dt* ∆*t* (2.1.24)

Assim, as posições e velocidades podem ser computadas simultaneamenete no algo ritmo Verlet-velocidade.

Substituindo os termos *d*2*~ri/dt*2e *d~vi/dt* por *~f*(*ri*(*t*))*/mi* na Equação 2.1.20, e o termo *d~ vi*(*t* + ∆*t*)*/dt* por *~f*(*ri*(*t* + ∆*t*))*/mi* 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

*~ri*(*t* + ∆*t*) = *~ri*(*t*) + *~vi*(*t*)∆*t* +*~f*(*ri*(*t*))

*mi*∆*t*2(2.1.25)

*~vi*(*t* + ∆*t*) = *~vi*(*t*) + 12*~f*(*ri*(*t* + ∆*t*))

*mi*∆*t* +12*~f*(*ri*(*t*))

*mi*∆*t* (2.1.26)

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

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 58

Implementação do algoritmo de Verlet-velocidade

1. Definir as condições iniciais das posições (*~ri*(0)), velocidades (*~vi*(0)) e acelerações (*~ai*(0)) (passo 1 da Figura 2.2).

2. Calcular as novas posições no tempo *t* + ∆*t* (passo 2 da Figura 2.2). *~ri*(*t* + ∆*t*) = *~ri*(*t*) + *~vi*(*t*)∆*t* +12*~ai*(*t*)∆*t*2

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

*~vi*(*t* + ∆*t/*2) = *~ri*(*t*) + 12*~ai*(*t*)∆*t*

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

*~ai*(*t* + ∆*t*) = *−*1*m∇Vi*(*~r*1(*t* + ∆*t*)*...~rN* (*t* + ∆*t*))

5. Calcular as velocidades no tempo *t* + ∆*t* (passo 5 da Figura 2.2).

*~vi*(*t* + ∆*t*) = *~vi*(*t* + ∆*t/*2) + 12*~ai*(*t* + ∆*t*)∆*t*

6. Repetir iterativamente os passos 2 ao 5.

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 59 Tempo

ri vi ai

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

**2.2 Simula¸c˜oes 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.

Imaginemos 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 séria dificuldade se estivéssemos interessados em amostrar únicamente o mínimo II.

Uma forma de conseguir amostrar as regiões “difíceis” da superfície de energia po tencial 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 introduzi das 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].

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 60

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 (*δξ*). A força média na direção da coordenada de reação na porção *k* depois de *Npassos* passos de simulação de dinâmica molecular será estimada simplesmente como a média simples das forças amostradas (Equação 2.2.1).

*Fξ*(*Npassos, k*) = 1 *Nkpassos*

*Nkpassos* X *µ*=1

*Fkµ*(2.2.1)

sendo *Nkpassos* o número de amostras acumuladas na porção *k* depois de *Npassos* passos e *Fkµ*a força instantânea registrada na porção *k*. O valor de *Nkpassos* não pode ser muito baixo devido a que pode causar falhas na estimativa da força média [70]. Considerando que o valor de *Nkpassos* é o suficiente, a força *Fξ*(*Npassos, 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*,

CAP´ITULO 2. SIMULAC¸OES DE DIN ˜ AMICA MOLECULAR ˆ 61 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).

*M*

∆*Aξ* = *−*X *i*=1

*Fξ*(*Npassos, k*)*δξ* (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* (*Nkpassos*) antes da aplicação da força *Fξ*(*Npassos, k*).

• width: Determina o tamanho de cada porção da coordenada de reação (*δξ*). • 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.

62

**Cap´ıtulo 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 preferen cial 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.

63

**Cap´ıtulo 4**

**Modelos param´etricos baseados**

**na ´area acess´ıvel ao solvente para computar os comprimentos de onda de emiss˜ao 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 Area acess´ıvel 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 hi drofóbicos 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

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 64 por meio das variações na fluorescência do triptofano.

Para exemplificar como a fluorescência do triptofano é usada, descreveremos breve mente o resultado dos experimentos de desenovelamento da proteína aporepressora do *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 biologia 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 desen volvidas [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 (*λem*). Os *λem* dos Trp19 e Trp99

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 65

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 *λem*, e viceversa.

.

l

e

R

e

d

a

d

i

s

n

e

t

n

I

1.0 0.8 0.6 0.4 0.2 0.0

1.0

**A B**

.

l

0.8

e

R

0.6

e

d

a

0.4

d

i

s

0.2

n

e

t

0.0

n

I

320 340 360 380 400 λem (nm)

320 340 360 380 400 λem (nm)

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 represor W99F, e B) Apo-represor 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 concentra ções 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 flu orescência está relacionada ao rendimento quântico do fluoróforo [1, 13]. As mutantes W99F e W19F ao desnaturar, exibem um comportamento distinto na sua intensidade de fluorescên cia. 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 aceptor (que podem ser, por exemplo, grupos amina) estejam envolvidos [15]. Observa mos 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 (*λem*). Os mu tantes ao desnaturar, exibem um aumento dos *λem* (Figura 4.2A e B). Os mutantes W99F e W19F nas suas formas completamente desnaturadas apresentam *λem* de *∼*355 nm e *∼*351 nm,

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 66

respectivamente. O fenômeno que determina o *λem* é produto de uma estabilização da transfe rência 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 *λem* e vice-versa.

A variação do *λem* que acabamos de observar nos experimentos de desenovelamento da proteína apo-repressora do *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 Simula¸c˜oes de dinˆamica 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 foram adicionados aos sistemas para deixá-los neutros, sendo a concentração de 0.16 M. Essa necessidade de deixar aos sistemas neutrais 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

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 67 • 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 trucamento 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 pro dução 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 *∼* 1-2 Å. Essas duas proteínas em particular, apresentaram os maiores valores de médios do RMSD, sendo de *∼* 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 C´alculo da ´area acess´ıvel 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

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 68

igual a 1.4 Å[79]. A lista de átomos vizinhos *j* ao átomo *i* do indol é obtido se *d*(*CiCj* ) *< ri* + *rj* + 2*r*, sendo *d*(*CiCj* ) a distância entre os centros dos átomos *i* e *j*, e *ri*, *rj* 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,

*SASAi* = 4*πr*2*i ×*

*P N*

(4.2.1)

sendo *ri* 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*,

*SASAindol* =X

*i*=1

**4.3 Resultados e discuss˜oes**

*SASAi* (4.2.2)

**4.3.1 Rela¸c˜ao entre o comprimento de onda de emiss˜ao e a ´area acess´ıvel 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 (*λem*) 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 *λem*, devido a que possuem valores baixo (Parvalbumina, *λem* = 316 nm [6]), intermediário (Nuclease de estafilococo, *λem* = 334 nm [5]),

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 69

e alto (Glucagon, *λem* = 352 nm [27]) de *λem* do triptofano. Assim, é de esperar-se que quanto maior seja a área acessível ao solvente do indol, maior será o *λem*.

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 *∼*6 Å2. A SASA do indol da Nuclease de estafilococo é de *∼*50 Å2, enquanto que a SASA do indol do Glucagon é maior das três, sendo aproximadamente de 150 Å2. Observamos que existe claramente uma correspondência entre o *λem* e a SASA do indol dessas três proteínas.

)

2

° A

(

l

o

d

n

I

o

d

A

S

A

S

200

150

100

50

0

0 5 10 15 20 Tempo (ns)

Figura 4.3: Evolução temporal da SASA do indol na Parvalbumina (azul), Nuclease de estafi lococo (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 *λem* estão disponíveis na literatura.

Baseado nas condições anteriores, um grupo de 19 proteínas foi selecionado (Ta bela 4.1). Os *λem* experimentais mínimo e máximo desse grupo de proteínas selecionadas foram de 308 e 352 nm, respectivamente, sendo assim representativos dos *λem* encontrados na natureza.

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 70

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 *λem*, apresentou o menor valor para a SASA do indol, de 0,45 Å2. Por outro lado, a proteína com maior *λem*, o Glucagon, apresentou a maior SASA, de 154 Å2.

A correlação entre o *λem* experimental e a SASA do indol para as 19 proteínas é mos trada 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 *λem* (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. 360

)

m

n

(

.

p

x

e

350 340 330 320

m e

λ

310

100 -50 0 50 100 150 200 -° SASAindol (A2)

Figura 4.4: Correlação entre o comprimento de onda de emissão de fluorescência do tripto fano experimental (*λem*) 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 Constru¸c˜ao dos modelos**

Na subseção anterior descrevemos a existência de uma correlação linear entre a SASA do indol e o *λem* do Trp em proteínas. Assim é de se esperar que modelos baseados na SASA do indol possam ter capacidade preditiva dos *λem* 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

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 71 relação linear entre o *λem* experimental e o valor médio da SASA do indol (Equação 4.3.1).

*λem* = *A*0 + *A*1 *× hSASAindi* (4.3.1)

Este primeiro modelo possui apenas dois parâmetros. O primeiro parâmetro, *A*0, é o valor mínimo do *λem* que o modelo consegue predizer. O segundo parâmetro, *A*1, é o fator que relaciona a SASA do indol ao *λem*. 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 *λem* menores a este valor serão sobrestimados. Das proteínas que usamos para construir o modelo, só três possuem *λem* experimentais menores que 324 nm, sendo estas a Azurina (*λem* = 308 nm), Parvalbumina (*λem* = 316 nm), e a Mioglobina (*λem* = 321 nm). O *λem* 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 *λem* preditos, usando o Modelo 1, versus os *λem* experimentais. O Modelo 1 tem uma capacidade preditiva dos *λem* 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 ben zênicas 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.

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 72

360 350

360

**A B**

)

m

n

(

.

p

x

e

340 330 320

)

m

n

(

.

p

x

e

350 340 330 320

m e

m e

λ

310 360

λ

310 330 350 λem calc. (nm)

310 360

310 330 350 λem calc. (nm)

**C D**

)

m

n

(

.

p

x

e

350 340 330 320

)

m

n

(

.

p

x

e

350 340 330 320

m e

m e

λ

310

310 330 350 λem calc. (nm)

λ

310

310 330 350 λem calc. (nm)

Figura 4.5: Comprimento de onda de emissão de fluorescência experimental (*λexp*

calculado (*λcalc*

*em* ) versus o

*em* ) usando os modelos baseados na área acessível ao solvente. A) Modelo 1: Baseado na *hSASAi* do grupo indol do triptofano. B) Modelo 2: Baseado na decomposição da *hSASAi* dos grupos benzeno incompleto e pirrol completo do indol. C) Modelo 3: Baseado na decomposição da *hSASAi* dos grupos benzeno completo e pirrol incompleto do indol. D) Modelo 4: Baseado na decomposição da *hSASAi* 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.

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 73

• 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 com ponentes 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).

*λem* = *B*0 + *B*1 *× hSASAben,ii* + *B*2 *× hSASApyr,ci* (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 (*SASAben,i*) e pirrol completo (*SASApyr,c*) ao *λem*, 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 *λem* sobrestimados. Este modelo possui uma capacidade preditiva com um desvio padrão de 6,58 nm com respeito ao *λem* experimental e um coeficiente de correlação *R*2 = 0,66. Os valores dos *λem* 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 (*∼*-0,02 nm), respectivamente. A correlação e o ajuste linear dos *λem* preditos com o Modelo 2 versus os *λem* experimentais estão na Figura 4.5B.

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

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 74 da SASA do benzeno completo e pirrol incompleto (Equação 4.3.3).

*λem* = *C*0 + *C*1 *× hSASAben,ci* + *C*2 *× hSASApyr,ii* (4.3.3)

sendo *C*0 o valor mínimo de *λem* que o modelo consegue predizer, e *C*1 e *C*2 são os fatores que associam o SASA do benzeno completo (*SASAben,c*) e pirrol incompleto (*SASApyr,i*) ao *λem*, respectivamente. Os valores dos *SASAben,c* e *SASApyr,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 sobrestimar os *λem* 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 *λem* 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 (*∼*-0,2 nm). Os valores dos *λem* 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).

*λem* = *D*0 + *D*1 *× hSASAben,ci* + *D*2 *× hSASApyr,ci* (4.3.4)

sendo *D*0 o valor mínimo de *λem* que o modelo consegue predizer, e *D*1 e *D*2 são os fatores que associam a SASA do benzeno completo (*SASAben,c*) e pirrol completo (*SASApyr,c*) ao *λem*, respectivamente. Os valores médios da *SASAben,c* e *SASApyr,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 *λem* com o Modelo 4 versus os *λem* experimentais para as 19 proteínas usadas na construção

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 75

do modelo estão mostradas na Figura 4.5D. Nessa figura é possível ver a tendência comum dos modelos baseados na SASA, a de sobrestimar os *λem* de proteínas com triptofanos com baixa acessibilidade ao solvente. Os valores das predições dos *λem* do Modelo 4 estão na Tabela 4.2.

s

a

d on

*S*

,

*d*

z

n

be

e

*p*

*m*

*x*

*i*

*c*

*,*

*n*

*e*

2

5,

6

4,

9

9,

7

0,

9

2,

5

8

4,

4

6,

9

1,

8

9,

5

6

4,

7

2

2,

4

6,

6

6

1,

7

5

6,

2

7

0,

4

2

9,

2

1

6,

4

1

1,

3

9

3,

3

*n*

*i*

,

*e*

*e*

*b*

0

2

5

1

*±*

4

0

3

*±*

*±*

3

*±*

*±*

1

1

1

1

3

1

de

s

o

t

n

e

m

i

r

p

com

,

) id.

B

D

(P

s

a

r

u

t

u

r

est

s

da

o

s

*A*

*S*

*A*

*S*

.

r

a

l

u

c

e

l

mo

a

c

i

m

â

din

de

s

e

õ

ç

a

l

u

sim

s

da

r

i

o

t

e

lp

m

o

inc

l

o

r

pir

,

o

t

e

lp

m

o

inc

o

n

e

z

n

be

,

l

o

ind do

*λ*

Os

.

o

d

a

r

t

s

é mo

o

ã

r

d

pa

o

i

v

des

O

.

s

e

õ

ç

a

l

u

a sim

*A*

*S*

*A*

*hS*

*i*

*c*

*,*

*r*

*y*

*p*

*A*

*S*

*A*

*hS*

*i*

*i*

*,*

*n*

*e*

*b*

*A*

*S*

*A*

*hS*

*id*

*n*

*i*

*A*

*S*

*A*

*S*

*h*

*±*

6  0,2

0

5

,

0

*±*

4  0,2

8

4

,

0

*±*

1  0,2

8

7

,

0

*±*

5  0,4

*±*

8  3,2

9

5

,

2

*±*

7  3,1

4

4

,

2

*±*

2  3,2

3

0

,

3

*±*

9  6,3

*±*

8  6,6

6

4

,

2

1

*±*

0

6

18,

8

4

,

5

*±*

3  5,6

9

8

,

3

1

*±*

3

2

24,

*±*

9  1,1

8

5

,

0

*±*

0  0,6

7

0

,

1

*±*

9  1,1

0

2

,

1

*±*

9  1,7

3

7

13,

6

6

,

2

*±*

3  4,1

7

0

,

5

*±*

4

7

10,

6

5

,

5

*±*

7

8

14,

*±*

8  5,6

5

0

,

6

*±*

6

3

37,

6

4

,

4

*±*

8  5,6

4

7

,

9

*±*

4

0

43,

*±*

2  0,3

8

2

,

2

*±*

9  4,6

4

6

,

0

*±*

2  0,3

2

4

,

2

*±*

1  5,0

*±*

0  4,6

5

7

,

4

*±*

8

2

38,

9

1

,

3

*±*

0  4,6

2

5

,

6

*±*

8

8

42,

0

7

18,

5

4

,

4

*±*

9

3

43,

8

8

,

5

*±*

9

5

18,

6

8

,

8

*±*

8

9

61,

6

3

48,

7

2

,

0

*±*

2  0,1

5

4

,

7

*±*

4

3

48,

1

5

,

7

*±*

6

4

48,

*±*

5  4,1

7

1

,

4

*±*

5  6,1

2

2

,

3

*±*

5  4,0

2

5

,

4

*±*

0

2

10,

6

9

50,

5

2

,

5

*±*

2

6

35,

8

1

,

6

*±*

3

1

46,

5

8

,

7

*±*

5

7

81,

5

2

27,

6

7

,

5

*±*

2

6

26,

4

6

,

6

*±*

1

2

23,

3

9

,

7

*±*

3

8

49,

*±*

6

0

21,

7

6

,

8

*±*

7

0

41,

2

7

,

0

1

*±*

4

4

19,

6

2

,

5

1

*±*

1

5

60,

*±*

5

3

46,

9

0

,

4

*±*

8

9

15,

0

5

,

3

1

*±*

9

0

42,

1

9

,

4

1

*±*

7

0

58,

*±*

8

4

71,

4

8

,

2

*±*

1  3,2

0

4

,

2

1

*±*

8

7

70,

8

0

,

4

1

*±*

9

9

73,

*±*

9

8

83,

0

7

,

1

1

*±*

8

9

57,

6

6

,

3

1

*±*

1

5

77,

3

0

,

4

1

*±*

9

4

,

135

*±*

3

4

95,

9

4

,

9

*±*

3

6

61,

9

0

,

1

3

*±*

5

7

87,

0

2

,

0

3

*±*

8

3

,

149

*±*

3

5

91,

9

5

,

7

*±*

0

2

70,

2

8

,

1

1

*±*

7

2

84,

8

2

,

3

1

*±*

7

4

,

154

s ace

t

r

a pa

e

e

*p*

]3

2

]

6

[

]

6

[

]

7

2

]9

2

]1

3

]1

3

]1

3

]3

3

]

5

[

]

6

[

]

6

[

]8

3

]7

2

]

6

[

]1

2

]

6

[

]

4

4

]

7

2

tn

r

sob

*xe*

*m*

[

*e*

308

316

321

[ 328

[ 330

[ 330

[ 330

[ 330

[ 331

334

335

340

[ 340

[ 342

345

[ 345

346

[ 348

[ 352

de

s

o

r

e

m

nú

,

s

a

d usa

s

a

n

í

e

t

o

Pr

:

4.1

a

le

ba

T

s

a

d

a

l

u

c

cal

e

t

n

e

v

sol ao

s

ie

v

ís

s

ace

s

a

áree

e

v

sol

ao

s

i

e

v

í

s

s

ace

s

a

áre

as

o

ã

s

*c*

*,*

*r*

*y*

*p*

*A*

*S*

*A*

*S*

a

i

d

a mé

a

t

n

e

s

e

r

p

e

r

*i h*

.

e

t

n

e

m

a

v

i

t

c

e

p

s

e

r

*λ*

id.

B

PD

a

n

í

e

t

o

r

P

]

2

2

[

U

Z

4A

a

n

i

r

u

z

A

]4

2

[

R

8

1B

a

n

i

m

u

b

l

a

v

r

a

P

]5

2

[T

Y

1M

a

n

i

b

o

l

g

o

i

M

]6

2

[T

N

9R

T1

e

s

a

e

l

c

u

n

o

b

i

R

]8

2

[

O

6

1D

2

1

P

B

K

F

]0

3

[

D

Y

1L

Y

8

3

1

W

Y

6

2

1

W

a

m

i

z

o

s

i

l

-

4

T

]0

3

[

D

Y

1L

Y

8

5

1

W

Y

6

2

1

W

a

m

i

z

o

s

i

l

-

4

T

]0

3

[

D

Y

1L

Y

8

5

1

W

Y

8

3

1

W

a

m

i

z

o

s

i

l

-

4

T

]2

3

[

C

B

1S

g

r

e

b

s

l

r

Ca

a

n

i

s

i

l

i

t

b

u

s

]

4

3

[

N

T

1S

o

c

o

c

o

l

fi

a

est

de

e

s

a

e

l

c

u

N

]5

3

[N

H

1C

Y

-

e

h

C

]6

3

[

X

T

1C

a

n

i

x

o

t

a

r

b

o

C

-

*α*

]7

3

[0

M

1B

A

S

H

]9

3

[L

O

1M

a

n

i

l

e

n

o

M

]0

4

[P

H

1C

)

o

r

e

m

ô

n

o

(m B

a

c

i

r

é

col

a

n

i

x

o

T

]

1

4

[

X

R

2T

A

1

3

W

a

n

i

x

o

d

e

r

o

i

T

]

2

4

[

T

L

2M

a

n

i

t

i

l

e

M

]3

4

[P

P

2B

A2

e

s

a

p

i

l

o

f

s

o

F

]5

4

[

N

C

1G

n

o

g

a

c

u

l

G

t

p

tri

e

aa

)

3

(

8

*e*

6

16,

8 9,6

6 6,8

6

0 -3,

9

9 -1,

1

5 -0,

8

8 -4,

8

6 -0,

8 2,1

5 2,2

3

8 -8,

2

2 -0,

1

7 -6,

1

6 -8,

3

1 -8,

5

9 -2,

7 3,6

7 4,6

8 0,5

doo

de

a

d on

áre na

*λ*

∆

8

)

6

*m*

,

3

8

6,

6

8,

4

9,

1

0,

9

4,

2

1,

2

3,

8

1,

5

2,

7

1,

8

7,

9

2,

9

3,

7

8,

5

0,

7

6,

7

6,

8

5,

d

s

(

*e*

324

325

327

324

328

329

325

329

333

336

326

339

333

333

336

342

349

352

352

a

l

u

c

e cal

deo

o

d

a

e

s

*λ*

)

1

tn

ba

2

(

*m*

*e*

6

16,

5 9,6

4 6,8

1

1 -3,

0

4 -2,

0

2 -0,

0

9 -4,

7

3 -0,

0 2,6

4 2,4

4

8 -8,

2

4 -0,

0

0 -7,

3

4 -8,

5

4 -8,

1

7 -2,

2 3,6

2 4,5

5 0,5

l

a

t

n

e

e

m

ir

p

o com

s

o

le

d mo

*λ*

∆

1

)

6

*m*

,

2

5

6,

4

8,

9

8,

0

6,

0

8,

0

1,

3

6,

0

6,

4

4,

6

1,

8

5,

0

0,

7

5,

5

5,

9

2,

2

6,

2

5,

5

5,

(

*e*

m

ir

e

exp

o

os

*λ*

)

3242

325

327

324

327

329

325

329

333

336

326

339

333

333

336

342

349

352

352

ãs

o

d

1

(

*m*

*e*

2

16,

4 9,3

9 7,6

3

5 -3,

7

0 -3,

3 2,2

2

9 -4,

0 2,2

0 4,8

5

7 -0,

5

9 -8,

8

4 -0,

9

4 -6,

8

4 -6,

4

9 -9,

4

9 -6,

3 3,6

4 4,2

0 1,2

a

ic

) 4

*−*

n

usa

*m*

*λ*

∆

n

ê

c

s

1

(

*e*

*λ*

,

e

.

nm

) 1

2

2

*m*

,

4

3,

9

6,

7

4,

3

9,

3

2,

8

0,

0

2,

0

8,

5

2,

5

0,

2

5,

1

5,

2

5,

6

0,

6

0,

3

6,

4

2,

0

2,

e

∆ t

n

(

*e*

324

325

328

324

326

332

325

332

335

333

326

339

333

335

335

338

349

352

353

r

o

flu

e

)

4

*−*

e

ma

os

são

*λ*

]

]

]

]

]

]

]

]

]

]

]

]

]

]

]

]

]

]

]

de

*m*

1

(

*e*

v

i

*p*

32

6

[

6

[

72

92

13

13

13

33

5

[

6

[

6

[

83

72

6

[

12

6

[

44

72

*λ*

tc

s

*xe*

*m*

[

*e*

308

316

321

[ 328

[ 330

[ 330

[ 330

[ 330

[ 331

334

335

340

[ 340

[ 342

345

[ 345

346

[ 348

[ 352

o

ã

s

s

i

em

de

a

d on

de

s

o

t

n

e

m

i

r

p

m

Co

:

4.2

a

le

ba

T

.

l

a

t

n

e

m

i

r

e

exp

a

i

c

n

ê

c

s

e

r

o

flu

de

o

ã

s

s

i

em

de

a

d

n

o

e

p

res

,

l

a

t

n

e

m

i

r

e

exp

r

o

val ao

o

v

i

t

a

rel

o

i

v

o dese

o

i

v

des e os

a

d on

de

s

o

t

n

e

m

i

r

p

com

s

do

s

e

da

di

n

u

*λ*

a

n

í

e

t

o

r

P

a

n

ir

u

z

A

a

n

i

m

u

bl

a

v

r

a

P

a

n

i

b

o

lg

o

i

M

T1

e

s

a

e

lc

u

n

o

bi

R

21

PB

KF

Y8

3

1 W

Y62

1 W

a

m

iz

o

s

il

-

4T

Y8

5

1 W

Y62

1 W

a

m

iz

o

s

il

-

4T

Y8

5

1 W

Y83

1 W

a

m

iz

o

s

il

-

4T

g

r

e

b

s

l Ca

a

n

i

s

i

li

t

b

u

s

o

c

o

c

o

l

fi

a

est de

e

s

a

e

lc

u

N

Y-

e

hC

a

n

i

x

o

t

a

r

bo

C-

*α*

AS

H

a

n

il

e

n

o

M

)

o

r

e

m

ô

n

o

(m B

a

c

i

r

é

col

a

n

i

x

o

T

A

1

3 W

a

n

i

x

o

d

e

r

o

iT

a

n

i

til

e

M

A2

e

s

a

p

i

l

o

fs

o

F

n

o

g

a

c

u

lG

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 78

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*2é o coeficiente de correlação, SD é o desvio padrão das predições com respeito aos dados experimentais. *SASAind*, *SASAben,c*, *SASAben,i*, *SASApyr,c*, *SASApyr,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 *λem* = 324 + 0*,*188 *hSASAindi* 0,64 6,80 2 SASA do benzeno(i)/pirrol(c)*λem* = 324 + 0*,*246 *hSASAben,ii* +

+ 0*,*103 *hSASApyr,ci*0,66 6,58

3 SASA do benzeno(c)/pirrol(i)*λem* = 324 + 0*,*241 *hSASAben,ci* +

+ 0*,*094 *hSASApyr,ii*0,66 6,56

4 SASA do benzeno(c)/pirrol(c)*λem* = 324 + 0*,*235 *hSASAben,ci* +

+ 0*,*093 *hSASApyr,ci*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 *λem*. A relação entre os parâmetros *B*1*/B*2 é 2,39; o qual contrasta com a relação entre *hSASAben,ii /hSASApyr,ci* = 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 *λem* 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 *λem*.

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

• Classe A: Caracterizada por um *λem* = 308 nm, e o Trp completamente protegido no bolsão hidrofóbico da proteína.

• Classe S: Caracterizada por um *λem* = 316 nm, e o Trp quase completamente protegido no bolsão hidrofóbico da proteína.

• Classe I: Caracterizada por um *λem* entre 330 e 332 nm, e o Trp parcialmente exposto ao

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 79 solvente.

• Classe II: Caracterizada por um *λem* entre 340, e 342 nm e o Trp exposto ao solvente.

• Classe III: Caracterizada por um *λem* entre 350, e 353 nm e o Trp completamente exposto ao solvente.

De acordo aos valores dos *λem* experimentais das 19 proteínas, reportados na Ta bela 4.1, é possível classifica-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 exepção é a proteína Mioglobina, a qual possui uma SASA do indol muito maior ao esperado para o seu *λem*. O *λem* 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

CAP´ITULO 4. MODELOS BASEADOS NA AREA ACESS ´ ´IVEL AO SOLVENTE 80

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 a 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 combinatória do tipo *CNk*, 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*41 = 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 con sidera 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*19

1 = 19 combinações diferentes de dadosa partir das quais vamos a construir os modelos, cada um com *N − k* = 18 dados. O dado restante será avaliado com os modelos