



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA FUNDAMENTAL
BACHARELADO EM QUÍMICA

JOSÉ GUILHERME VIEIRA MOURA

**OTIMIZAÇÃO DE PARÂMETROS INTERATÔMICOS DE ÁLCOOIS
ALIFÁTICOS DIRECIONADA POR APRENDIZADO DE MÁQUINA:
REPRODUÇÃO DE PROPRIEDADES FÍSICO-QUÍMICAS EM
SIMULAÇÕES DE DINÂMICA MOLECULAR**

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Moura, José Guilherme Vieira.

Otimização de parâmetros interatômicos de álcoois alifáticos direcionada por aprendizado de máquina: Reprodução de propriedades físico-químicas em simulações de dinâmica molecular / José Guilherme Vieira Moura. - Recife, 2025. 83 p. : il., tab.

Orientador(a): Filipe da Silva Lima

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Química - Bacharelado, 2025.

Inclui referências, apêndices.

1. Dinâmica Molecular. 2. Química Computacional. 3. Aprendizado de Máquina. 4. Redes Neurais. I. Lima, Filipe da Silva. (Orientação). II. Título.

540 CDD (22.ed.)

JOSÉ GUILHERME VIEIRA MOURA

**OTIMIZAÇÃO DE PARÂMETROS INTERATÔMICOS DE
ÁLCOOIS ALIFÁTICOS DIRECIONADA POR APRENDIZADO DE
MÁQUINA: REPRODUÇÃO DE PROPRIEDADES FÍSICO-
QUÍMICAS EM SIMULAÇÕES DE DINÂMICA MOLECULAR**

Trabalho de Conclusão de Curso apresentado
ao Curso de Bacharelado em Química da
Universidade Federal de Pernambuco, como
requisito parcial para obtenção do título de
Bacharel em Química.

Aprovado em: 04/08/2025

BANCA EXAMINADORA

Prof. Dr. Filipe da Silva Lima (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Roberto Dias Lins Neto
Universidade Federal de Pernambuco

Prof. Dr. Flamarion Borges Diniz
Universidade Federal de Pernambuco

AGRADECIMENTOS

Gostaria de iniciar agradecendo, principalmente, à Laisla, minha psicóloga, que foi essencial para minha sobrevivência nesta jornada; obrigado por todas as lições e conselhos nestes anos em que você me atendeu!

Não menos importante, gostaria de agradecer ao PET-Química, grupo que me acolheu desde 2019 e não só me ensinou muito, como também contribuiu enormemente para a minha formação. Especialmente às minhas tutora e ex-tutora, Profa. Dra. Giovannia e Profa. Dra. Janaína, e a todos os meus colegas que passaram pelo grupo.

Ao meu orientador, Prof. Dr. Filipe, sou extremamente grato por ter me ensinado — e continuar me ensinando — a ser um pesquisador de verdade. Obrigado pela liberdade dada a mim neste trabalho, de explorar minha curiosidade acerca dos problemas encontrados, e pelas boas músicas apresentadas!

A todos os meus colegas e heróis do departamento, obrigado pelos momentos, histórias e cafés compartilhados. A Anthony e a João, pelo nosso sofrimento e apoio mútuo nesta fase final do início de nossa formação. A Gabriel, à Adri e à Hisla, irmãos de laboratório que eu espero poder levar para a vida.

Aos meus amigos da resistência — Pedro, Renato, Clécio, Arthur e todos os outros —, obrigado pela presença em mais de 10 anos da minha vida, e espero que continue por muitos mais.

À minha mãe, Giselle, e à minha tia, Isabela, obrigado pelo apoio incondicional que vocês me deram nesta vida e nesta reta final. Ao meu pai, Guilherme, obrigado por todos os conselhos, puxões de orelha e cafés da manhã feitos nesta vida.

À minha querida futura esposa, Juliana, obrigado por estar sempre ao meu lado, me ajudando e deixando meus dias difíceis menos ásperos; que eu consiga sempre lhe retribuir o amor que você me dá!

Aos professores do Departamento de Química Fundamental, obrigado por contribuírem de forma indispensável para a minha formação. E a todas as outras pessoas que foram necessárias nesta minha jornada, ofereço os meus mais sinceros “Muito obrigado!!!”.

“I want AI to do my laundry and dishes
so that I can do art and writing, not for
AI to do my art and writing so that I can
do my laundry and dishes.”

- Maciejewska. Joanna, 2024.

RESUMO

Álcoois alifáticos compõem uma ampla família de solventes e reagentes, sendo indispensáveis nos setores científicos e industriais. No nível molecular, a combinação do grupo hidroxila hidrofílico e cadeia carbônica hidrofóbica confere caráter anfifílico capaz de perturbar membranas, estabilizar proteínas e facilitar extrações verdes de biomoléculas. Técnicas de dinâmica molecular permitiram investigar esses efeitos com resolução atômica, mas sua acurácia depende criticamente da qualidade dos parâmetros do potencial efetivo empregado. Os campos de força aditivos frequentemente não representam de maneira adequada propriedades de álcoois alifáticos lineares, em especial a função de distribuição radial $g_{oo}(r)$, o coeficiente de autodifusão e a constante dielétrica, comprometendo a representação de processos moleculares dependentes desses solventes. Neste trabalho, empregou-se um protocolo de aprendizado de máquina ativo baseado em ciclos iterativos de predição–simulação–retreinamento para otimizar, no CHARMM36, os parâmetros de Lennard-Jones e a diferença de carga dos átomos da hidroxila em cinco álcoois lineares (etanol a 1-hexanol). Conjuntos de 100 modelos de perceptrons de múltiplas camadas (MLPs) avaliaram mais de 10^6 combinações de parâmetros, minimizando uma função-alvo multivariada que captura os desvios entre simulação e dados experimentais. O protocolo de aprendizagem de máquina gerou novos parâmetros que aumentaram o coeficiente de determinação R^2 entre os $g_{oo}(r)$ simulados e suas contrapartes experimentais de valores inferiores a 0,2 para valores superiores a 0,95, além de reduzir significativamente o desvio relativo médio em relação aos dados experimentais, resultando em uma representação muito mais fiel dos álcoois trabalhados. Além disso, a transferibilidade dos parâmetros foi testada em (i) uma mistura DIPE/água/etanol, (ii) a solução etanólica do peptídeo β -amiloide e (iii) uma micela reversa de DTAB em 1-hexanol. Nos três cenários, os álcoois reparametrizados reproduziram melhor propriedades experimentais de interesse, resultados que os parâmetros originais não alcançaram. Os resultados demonstram que ajustes direcionados por aprendizado de máquina melhoram a descrição de propriedades experimentais e revelam a importância da descrição correta da estrutura dos solventes para os sistemas trabalhados. O protocolo, de fácil extensão a outros solventes e campos de força, valida uma rota prática e precisa para a reparametrização de potenciais clássicos.

Palavras-chave: dinâmica molecular; aprendizado de máquina; redes neurais; função de distribuição radial; ajustes NBFIX; solventes orgânicos.

ABSTRACT

Aliphatic alcohols constitute a ubiquitous class of solvents and reagents whose amphiphilic character—hydrophilic hydroxyl group plus hydrophobic alkyl chain—modulates membranes, proteins and green extraction processes. Atomistic insight into these phenomena relies on molecular-dynamics (MD) simulations, yet additive force-fields such as CHARMM36 systematically misrepresent key liquid properties of linear alcohols, notably the oxygen–oxygen radial distribution function $g_{\text{oo}}(r)$, the self-diffusion coefficient D and the dielectric constant ϵ . Here we implement an active-learning protocol that iteratively couples prediction, MD simulation and retraining to re-optimize Lennard-Jones parameters and the hydroxyl charge split for five alcohols (ethanol–1-hexanol). One hundred multilayer perceptrons explored 10^6 parameter combinations, minimizing a multivariate target function against experimental values. The resulting parameters sets raised the coefficient of determination between simulated and experimental $g_{\text{oo}}(r)$ from 0.2 to 0.95 and reduced the mean relative deviation across all targets, delivering a markedly more faithful description of the liquids. Transferability tests on (i) a DIPE/water/ethanol interface, (ii) β -amyloid in ethanol, and (iii) a DTAB reverse micelle in 1-hexanol confirmed superior reproduction of experimental observables that the original force-field failed to capture. These results show that machine-learning-guided, local corrections restore consistency between microscopic structure and macroscopic properties, providing a robust parameter set for complex multicomponent simulations. The workflow is readily extendable to other solvents and force-fields, offering a practical and accurate route for the re-parameterization of classical potentials.

Keywords: molecular dynamics; machine learning; neural networks; NBFIX parameterization; organic solvents; radial distribution function.

LISTA DE GRÁFICOS

Gráfico 1 - Representação arbitrária do potencial de Lennard-Jones.....	20
Gráfico 2 - Perfis experimentais da função de distribuição radial $g(r)$ para álcoois lineares (etanol – 1-hexanol) a 298 K.....	36
Gráfico 3 - Perfil de densidade ao longo do eixo z dos componentes do sistema.....	40
Gráfico 4 - Desvio relativo por propriedade e campo de força.....	46
Gráfico 5 - Função $g(r)$ entre átomos de oxigênio do 1-propanol.....	47
Gráfico 6 - Potencial entre os pares de átomos de oxigênio do 1-propanol.....	48
Gráfico 7 - Espaço de parâmetros amostrado nas cinco iterações de reparametrização.....	51
Gráfico 8 - Comparação do desvio relativo por propriedade.....	55
Gráfico 9 - Influência da correção NBFIX na estrutura do 1-propanol e potencial $v(r)$	56
Gráfico 10 - Perfil de densidade para os componentes do sistema ternário.....	58
Gráfico 11 - Distribuição de probabilidade do número de moléculas de etanol vizinhas nas fases orgânica e aquosa.....	60
Gráfico 12 - Evolução temporal da estrutura secundária da β -amiloide (15 – 27) em solução etanólica.....	61
Gráfico 13 - Evolução do número de moléculas de água no núcleo micelar de DTAB.....	63
Gráfico B1 – Densidade em função do tamanho da cadeia dos álcoois.....	76
Gráfico B2 – Coeficiente de autodifusão em função do tamanho da cadeia dos álcoois.....	76
Gráfico B3 – Constante dielétrica em função do tamanho da cadeia dos álcoois.....	77
Gráfico B4 – Entalpia de vaporização em função do tamanho da cadeia dos álcoois.....	77
Gráfico B5 – Coeficiente de determinação em função do tamanho da cadeia dos álcoois simulados.....	78
Gráfico C1 – Comparação da densidade em função do tamanho da cadeia dos álcoois.....	78
Gráfico C2 – Comparação da coeficiente de autodifusão em função do tamanho da cadeia dos álcoois.....	80
Gráfico C3 – Comparação da constante dielétrica em função do tamanho da cadeia dos álcoois.....	80
Gráfico C4 – Comparação da entalpia de vaporização em função do tamanho da cadeia dos álcoois.....	81
Gráfico C5 – Comparação do coeficiente de determinação em função do tamanho da cadeia dos álcoois.....	81
Gráfico D1 - Evolução temporal da densidade de etanol no <i>bulk</i> das fases orgânica (c1) e aquosa (c2).....	82

Gráfico D2 – Ajuste polinomial para o coeficiente de partição do etanol em função da fração molar em DIPE.....	82
Gráfico D3 – Coeficiente de partição do etanol em função do tempo nas múltiplas simulações.....	83

LISTA DE TABELAS

Tabela 1 - Densidades experimentais, densidades moleculares e número de moléculas simuladas para álcoois puros.....	36
Tabela 2 - Dados experimentais de densidade (ρ), coeficiente de autodifusão (D), constante dielétrica (ϵ) e entalpia de vaporização molar (ΔH) dos álcoois lineares a 298 K e 1 bar.....	37
Tabela 3 - Parâmetros de Lennard-Jones e cargas parciais do oxigênio (σ , ϵ , q_o) e do hidrogênio (q_H) da hidroxila dos álcoois em cada campo de força.....	46
Tabela 4 - Desvio Relativo Médio (DRM) das propriedades simuladas para cada álcool em cada campo de força.....	46
Tabela 5 - Número de coordenação para os diferentes álcoois.....	50
Tabela 6 - Parâmetros finais obtidos para cada um dos álcoois.....	54
Tabela 7 - Desvio relativo entre as previsões dos modelos de aprendizado de máquina e suas respectivas simulações.....	54
Tabela 8 - Desvio Relativo Médio (DRM) das propriedades simuladas para cada álcool com os diferentes conjuntos de parâmetros.....	56
Tabela 9 - Comparação do número de coordenação para os diferentes álcoois com os diferentes parâmetros.....	58
Tabela 10 - Comparação entre os coeficientes de partição do etanol experimental e computacional.....	60
Tabela 11 – Tempo na estrutura de α -hélice por resíduo, para os diferentes conjuntos de parâmetros.....	62
Tabela A1 - Distância do primeiro mínimo da função de distribuição radial $g(r)$ para álcoois lineares.....	73
Tabela A2 – Parâmetros dos sistemas simulados empregados no treinamento inicial do modelo de aprendizado de máquina (Iteração 1).....	74
Tabela A3 – Parâmetros dos sistemas simulados na segunda iteração do aprendizado de máquina.....	75
Tabela A4 – Parâmetros dos sistemas simulados na terceira iteração do aprendizado de máquina.....	75
Tabela A5 – Parâmetros dos sistemas simulados na quarta iteração do aprendizado de máquina.....	75
Tabela A6 – Parâmetros dos sistemas simulados na quinta iteração do aprendizado de máquina.....	75

Tabela C1 – Propriedades preditas pelos modelos de aprendizado de máquina.....	79
Tabela C2 – Propriedades obtidas através das simulações dos álcoois utilizando os conjuntos finais de parâmetros.....	79

LISTA DE FIGURAS

Figura 1 - Esquema do cálculo da função distribuição radial $g(r)$	26
Figura 2 - Exemplo da arquitetura de um MLP.....	29
Figura 3 - Representação da caixa de simulação utilizada como estrutura inicial.....	40
Figura 4 - Estrutura do fragmento A β 15-27 funcionalizado utilizada nas simulações em etanol.....	42
Figura 5 - Estrutura inicial da micela reversa de DTAB.....	43
Figura 6 - Estruturas finais das micelas reversas de DTAB simuladas.....	64

LISTA DE EQUAÇÕES

(1).....	19
(2).....	20
(3).....	20
(4.1).....	22
(5).....	23
(6).....	23
(7.1).....	23
(8).....	24
(9).....	24
(10).....	24
(11).....	24
(12).....	25
(13).....	25
(14).....	27
(15).....	27
(16).....	27
(17).....	27
(18).....	28
(19).....	29
(20).....	36
(21).....	38
(22.1) e (22.2).....	39
(23).....	39
(24).....	41

LISTA DE ABREVIATURAS E SIGLAS

Ala — Alanina

Asn — Asparagina

Asp — Ácido aspártico

ACE — Acetil]

DIPE — *Di-isopropyl Ether* (Éter di-isopropílico)

DM — Dinâmica Molecular

DSSP — *Dictionary of Secondary Structure of Proteins* (Dicionário de Estruturas Secundárias de Proteínas)

DTAB — *Dodecyltrimethylammonium Bromide* (Brometo de dodeciltrimetilamônio)

Gln — Glutamina

Glu — Ácido glutâmico

Gly — Glicina

Leu — Leucina

Lys — Lisina

MLP — *Multi-Layer Perceptron* (Perceptron de múltiplas camadas)

MSD — *Mean Square Displacement* (Deslocamento quadrático médio)

NBFI — *Non-Bonded FI* (Correção de interações não-ligadas)

NME — N-Metilamida

NPT — Ensemble isotérmico-isobárico

NVT — Ensemble canônico

PBC — *Periodic Boundary Conditions* (Condições de contorno periódicas)

RDF — *Radial Distribution Function* (Função de Distribuição Radial)

RMSD — *Root Mean Square Deviation* (Desvio Quadrático Médio)

Ser — Serina

Val — Valina

SUMÁRIO

1	INTRODUÇÃO.....	17
2	FUNDAMENTAÇÃO TEÓRICA.....	19
2.1	FUNDAMENTOS DA DINÂMICA MOLECULAR.....	19
2.1.1	Modelagem do potencial empírico	
2.1.2	Equações de movimento	
2.1.3	Controle da temperatura e pressão de uma simulação	
2.2	CÁLCULO DE PROPRIEDADES EM SIMULAÇÕES DE DINÂMICA MOLECULAR.....	24
2.2.1	Coeficiente de autodifusão	
2.2.2	Constante dielétrica	
2.2.3	Entalpia de vaporização	
2.3	A FUNÇÃO DE DISTRIBUIÇÃO RADIAL.....	26
2.4	FUNDAMENTOS DA APRENDIZAGEM DE MÁQUINA.....	28
3	OBJETIVOS.....	31
3.1	Objetivos específicos.....	31
4	METODOLOGIA.....	32
4.1	Estruturas e topologias.....	32
4.2	Parâmetros das simulações de dinâmica molecular.....	32
4.2.1	Parâmetros dos sistemas de álcoois puros	
4.2.2	Parâmetros do sistema ternário	
4.2.3	Parâmetros do sistema contendo um fragmento da proteína β-amiloide	
4.2.4	Parâmetros do sistema contendo uma micela reversa	
4.3	Sistemas de álcoois puros.....	35
4.4	Modelo de aprendizado de máquina.....	37
4.5	Sistema ternário DIPE/Água/Etanol.....	40
4.6	Solução etanólica da β -amiloide.....	41
4.7	Sistema micelar reverso.....	43
5	RESULTADOS E DISCUSSÃO.....	45
5.1	Sistemas de álcoois puros com parâmetros originais.....	43
5.1.1	Divergências na função de distribuição radial	
5.1.2	Divergências na constante dielétrica	
5.1.3	Divergências no coeficiente de autodifusão	
5.2	Desempenho dos modelos de aprendizado de máquina.....	51

5.3	Comparação entre álcoois com os parâmetros originais e com as correções NBFIX.....	55
5.4	Influência da reparametrização no sistema ternário DIPE/água/etanol.....	58
5.5	Influência da reparametrização na solução etanólica da β -amiloide.....	61
5.6	Influência da reparametrização no sistema micelar reverso.....	63
6	CONCLUSÕES E PERSPECTIVAS.....	65
	REFERÊNCIAS.....	67
	APÊNDICE A – Parâmetros adicionais.....	73
	APÊNDICE B – Propriedades individuais para os campos de força originais.....	76
	APÊNDICE C – Propriedades preditas e simuladas dos álcoois reparametrizados.....	79
	APÊNDICE D – Informações adicionais para os sistemas multicomponentes.....	82

1 INTRODUÇÃO

Os álcoois alifáticos constituem uma extensa família de compostos orgânicos que desempenham papéis cruciais como reagentes, solventes e agentes extratores em diversos segmentos industriais, entre eles o têxtil, o farmacêutico, o alimentício, além de servirem como matéria-prima para estruturas químicas de maior complexidade (Gao *et al.*, 2019; Periyasamy; Dutta; Tehrani-Bagha, 2024; Sagandira *et al.*, 2020). Particularmente, o etanol atua como principal biocombustível líquido do planeta, e especificamente no Brasil, responde por 17 % de todo o consumo de combustível ou energia utilizado para transporte (Pelkmans; Vieira; Leal, 2024). Do ponto de vista molecular, a presença do grupo hidroxila confere a essas moléculas a capacidade de formar pontes de hidrogênio, tanto entre si quanto com solutos ou macromoléculas, enquanto a cadeia carbônica acoplada introduz um caráter hidrofóbico. Essa anfifilicidade é crucial em processos bioquímicos: álcoois de cadeia curta penetram e fluidificam bicamadas fosfolipídicas, regulando a elasticidade e a função de proteínas de membrana (Galindo *et al.*, 2024; Ingólfsson; Andersen, 2011); atuam como cossolventes capazes de estabilizar ou desnaturar proteínas, modulando dobramento e atividade enzimática (Gazi; Maity; Jana, 2023; Nick Pace *et al.*, 2004); e figuram como solventes verdes na extração seletiva de lipídios, fenóis e outras biomoléculas de interesse industrial (Carré *et al.*, 2024; Lee *et al.*, 2024).

Nas últimas décadas, técnicas de simulação computacional, como a dinâmica molecular (DM), tornaram-se ferramentas centrais para o estudo de sistemas em fase condensada, fornecendo resolução atômica a processos inacessíveis por métodos puramente experimentais (Allen; Tildesley, 2017). Diversos campos de força, como o OPLS-AA e o TraPPE, incorporam séries homólogas de álcoois durante a etapa de ajuste de parâmetros do potencial empírico, com o objetivo de reproduzir propriedades como densidade e entalpia de vaporização (Jorgensen; Maxwell; Tirado-Rives, 1996; Obeidat; Al-Salman; Abu-Ghazleh, 2018). Esses modelos têm sido aplicados diretamente em investigações que abrangem desde interfaces vapor-líquido de etanol, passando por misturas água-álcool, até a interação de cadeias curtas de álcool com bicamadas fosfolipídicas, evidenciando utilidade e aplicabilidade das simulações de DM em contextos biológicos e tecnológicos (Patra *et al.*, 2006; Pérez; González-Salgado; Lomba, 2021; Yang *et al.*, 2023).

Contudo, avaliações sistemáticas, incluindo as conduzidas neste trabalho, mostram que propriedades como a constante dielétrica, o coeficiente de autodifusão e a estrutura do solvente, podem divergir substancialmente dos valores experimentais a depender do campo de força empregado (Kohns, 2020; Zangi, 2018). Para reduzir essas divergências, é comum refinar

pontualmente os parâmetros do potencial empírico, gerando variantes do campo de força que reproduzem com maior fidelidade o sistema estudado (Yoo; Aksimentiev, 2018; Zangi, 2018; Zêzere *et al.*, 2023). Entretanto, mesmo em líquidos puros, o número de parâmetros ajustáveis é elevado, e eles estão fortemente acoplados entre si, o que torna a otimização um desafio não trivial.

Nesse contexto, técnicas de aprendizado de máquina têm emergido como ferramentas poderosas na físico-química computacional. Em particular, redes neurais profundas e outros algoritmos de aprendizagem de máquina vêm sendo empregados na exploração e otimização do espaço de parâmetros de campos de força clássicos (Befort *et al.*, 2021; Noé *et al.*, 2020). Estas novas estratégias de aprendizado ativo permitem a criação de modelos capazes de avaliar milhões de combinações de parâmetros por uma fração do custo computacional, guiando a otimização para regiões que reproduzem simultaneamente propriedades experimentais de interesse de forma muito mais eficiente (Befort *et al.*, 2021). O presente trabalho demonstrou que um protocolo de aprendizado ativo, baseado em ciclos iterativos de predição, simulação e retreinamento, é altamente eficaz para mapear o espaço de parâmetros e otimizar o potencial empírico utilizado nas simulações de DM. Aplicado aos álcoois alifáticos, o método corrige as discrepâncias de D e $g(r)$ relatadas na literatura e confirmadas aqui, elevando a precisão dos campos de força clássicos para simulações dos álcoois puros e, por fim, verificando a aplicabilidade dos novos conjuntos de parâmetros para sistemas multicomponentes.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 FUNDAMENTOS DA DINÂMICA MOLECULAR

A modelagem molecular reúne técnicas destinadas a descrever o comportamento de átomos, moléculas e agregados moleculares por meio da representação explícita de suas estruturas e interações. Entre essas técnicas, destaca-se a Dinâmica Molecular (DM), na qual a energia potencial de um sistema é calculada exclusivamente a partir das coordenadas nucleares, ignorando o movimento eletrônico, graças a aproximação de Born-Oppenheimer (Leach, 2001). A evolução temporal dessas partículas é obtida através da mecânica clássica, integrando numericamente as equações de movimento de Newton, mantendo assim um baixo custo computacional que possibilita o estudo de milhares a milhões de partículas em escalas de tempo compatíveis com fenômenos físico-químicos inviáveis de se tratar por cálculos *ab initio* (Leach, 2001).

2.1.1 Modelagem do potencial empírico

De um ponto de vista atomístico, qualquer sistema contendo N partículas pode ter sua energia potencial $\mathcal{V}$ decomposta em contribuições $v(r_i, r_j, \dots)$ que dependem das coordenadas de pares, trios, quartetos e, de maneira geral, de agrupamentos sucessivamente maiores de átomos, conforme a Equação 1.

$$\mathcal{V} = \sum_i^N \sum_{j>i}^{N-1} v_2(r_i, r_j) + \sum_i^N \sum_{j>i}^{N-1} \sum_{k>j}^{N-2} v_3(r_i, r_j, r_k) + \dots \quad (1)$$

Embora as contribuições de três corpos ou mais corpos possam ser significativas, elas quase nunca são incluídas em simulações porque qualquer soma sobre trios ou grupos maiores de partículas torna a simulação computacionalmente proibitiva (Allen; Tildesley, 2017). Em vez disso, adota-se a aproximação por pares, na qual todos os efeitos de ordem superior são incorporados em um potencial efetivo de dois corpos descrito pela Equação 2. Por incorporar o efeito médio das interações de múltiplos corpos, esse potencial irá exibir dependência de variáveis termodinâmicas do sistema, como temperatura (T), pressão (P) e densidade (ρ), embora o verdadeiro potencial de dois corpos $v_2(r_{ij})$ seja fundamentalmente independente dessas grandezas (Allen; Tildesley, 2017).

$$\mathcal{V} \approx \sum_i^N \sum_{j>i}^{N-1} v_2^{eff}(r_{ij}, \{T, P, \rho, \dots\}) \quad (2)$$

Na dinâmica molecular, um campo de força corresponde ao conjunto de funções e parâmetros utilizadas para descrever o potencial da Equação 2, englobando todas as contribuições inter- e intramoleculares. A escolha das funções e dos parâmetros numéricos varia segundo a filosofia de construção e o propósito do modelo, mas o objetivo comum é reproduzir, com boa fidelidade, observáveis obtidas experimentalmente ou por cálculos *ab initio*. Quando bem calibrados, tais parâmetros tornam-se transferíveis, permitindo descrever famílias inteiras de compostos sem necessidade de reajustes (Allen; Tildesley, 2017; Leach, 2001). Os campos de força agrupam-se em classes, definidas pela complexidade de sua forma funcional, como a inclusão de termos cruzados ou de polarização explícita a aproximação de pares. O presente trabalho concentra-se nos campos de força de Classe I (GAFF, OPLS e CHARMM36), cuja energia potencial segue a forma mostrada na Equação 3.

$$\begin{aligned} \mathcal{U}(r^N) = & \sum_{\text{ligações}} \frac{k_{l,i}}{2} (l_i - l_{i,0})^2 + \sum_{\text{ângulos}} \frac{k_{\theta,i}}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{torções}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\ & + \sum_{i=1}^N \sum_{j>i}^{N-1} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i=1}^N \sum_{j>i}^{N-1} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \end{aligned} \quad (3)$$

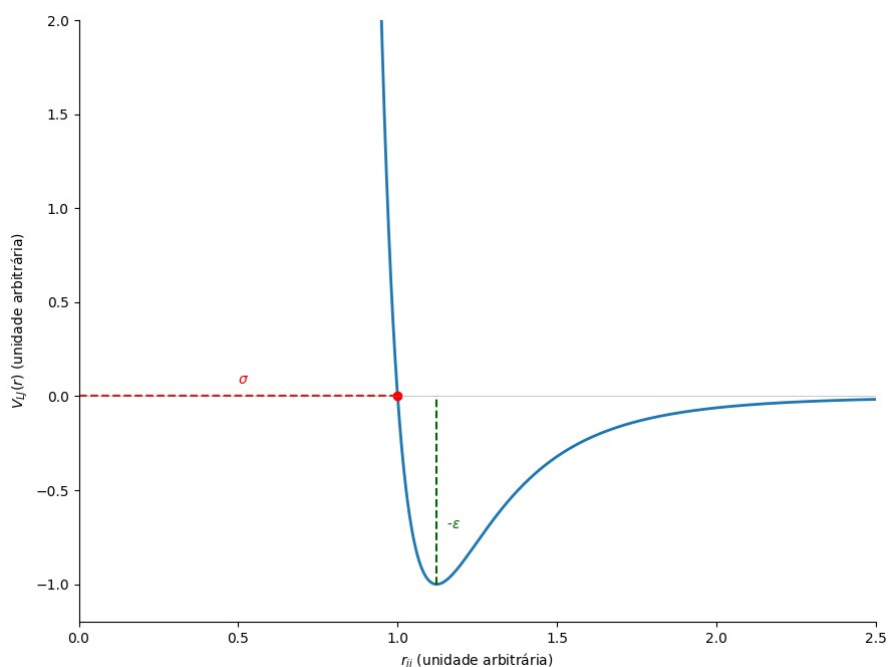
No primeiro e segundo termo da equação, um potencial harmônico é utilizado para modelar o alongamento de ligações covalentes entre pares de átomos e as variações nos ângulos formados entre três átomos consecutivos numa mesma molécula, respectivamente. Assim, no primeiro termo, l_i é comprimento da i -ésima ligação e $l_{i,0}$ é a posição de repouso da ligação. De maneira análoga, no segundo termo, θ_i é o valor do i -ésimo ângulo e $\theta_{i,0}$ corresponde a posição de repouso ângulo. Ambos termos possuem uma constante de força independente atrelada ($k_{l,i}$ ou $k_{\theta,i}$). O terceiro termo descreve a energia torsional associada à rotação em torno de uma ligação simples; V_n é a magnitude da barreira de potencial, n a multiplicidade, definida como o número de mínimos em uma rotação completa, ω o ângulo diedro e γ a fase que desloca esses mínimos. Esse pote

A todos meus colegas e heróis do departamento, obrigado pelos momentos, hncial é essencial para representar a flexibilidade conformacional de moléculas, especialmente em cadeias orgânicas, onde variações do ângulo diedro governam o acesso a diferentes conformações de baixa energia (Leach, 2001). Esses três termos formam a componente de ligação do potencial, englobando

exclusivamente interações intramoleculares que descrevem os graus de liberdade das ligações covalentes (estiramento, flexão e torção).

As interações entre átomos não ligados são tratadas pelos dois somatórios duplos restantes. O primeiro corresponde ao potencial de Coulomb, que expressa a energia eletrostática entre pares de cargas pontuais onde: q_i e q_j são as cargas dos átomos i e j , r_{ij} é a distância que os separa e ϵ_0 a permissividade elétrica do vácuo. O segundo somatório introduz o potencial de Lennard-Jones, empregado para modelar forças de van der Waals. Esse termo depende de dois parâmetros empíricos: σ_{ij} , o diâmetro de colisão (distância na qual $V_{LJ}=0$), e ϵ_{ij} , a profundidade do poço de potencial, como indicado no Gráfico 1. Na equação, a parte proporcional a r_{ij}^{-6} representa o componente atrativo decorrente das forças de dispersão de London, enquanto o termo repulsivo em r_{ij}^{-12} é introduzido para capturar, de forma empírica, a repulsão dos elétrons devido ao princípio da exclusão de Pauli, dominante em curtas distâncias (Leach, 2001).

Gráfico 1 - Representação arbitrária do potencial de Lennard-Jones



Fonte: elaborado pelo próprio autor

Um dos principais desafios na parametrização de um campo de força é atribuir valores adequados de Lennard-Jones (σ, ϵ) a cada tipo atômico. O campo de força CHARMM36, por exemplo, distingue hoje mais de 500 tipos de átomo, cada um com seu conjunto (σ, ϵ) distinto. Ajustar explicitamente todas as interações cruzadas em um sistema com múltiplos tipos atômicos seria impraticável: um sistema com n tipos distintos exigiria $n(n-1)/2$ pares adicionais de (σ, ϵ) (Leach, 2001). Para contornar essa questão, os campos de força utilizam regras de combinação que

estimam os parâmetros de um par de átomos distintos a partir dos valores definidos para cada espécie isolada (σ_{ii} e ε_{ii}). Embora as expressões matemáticas para as regras de combinação variem, sendo adaptadas às diferentes metodologias de parametrização, a formulação mais comumente utilizada é a de Lorentz–Berthelot, mostrada nas Equação 4.1 e 4.2.

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad (4.1)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii} \varepsilon_{jj}} \quad (4.2)$$

As regras de combinação são uma ferramenta útil, porém não deixam de ser expressões aproximadas, não havendo garantia de que forneçam valores acurados para todas as combinações de átomos. De fato, estudos mostram que especialmente a combinação de ε por média geométrica pode ser inadequada para a descrição de líquidos na maioria dos casos (Haslam; Galindo; Jackson, 2008). Para contornar essas limitações, muitos campos de força permitem a definição de parâmetros específicos de pares atômicos, que substituem os valores obtidos pelas Equações 4.1 e 4.2. Essa estratégia, denominada NBFIX (*Non-Bonded FIX*), ajusta diretamente os valores σ_{ij} e ε_{ij} para os pares atômicos desejados, garantindo melhor reprodução de propriedades experimentais (Polêto; Lemkul, 2022; Yoo; Aksimentiev, 2018).

2.1.2 Equações de movimento

A dinâmica molecular clássica consiste em integrar iterativamente as equações de movimento de Newton para todas as partículas do sistema. Para cada átomo i , de massa m_i , a aceleração obedece a Equação 5, onde f_i é a força resultante. Em sistemas conservativos, como os das simulações de DM, essa força se relaciona com a energia potencial definida pelo campo de força através da Equação 6. Conhecido o conjunto de forças em um dado instante t , algoritmos de integração, como *leap-frog* ou *velocity-Verlet*, resolvem a Equação 5, atualizando posições e velocidades das partículas constituintes do sistema. A repetição desse ciclo por milhares a milhões de passos forma a trajetória temporal do sistema, da qual se extraem propriedades estatísticas e termodinâmicas (Leach, 2001).

$$\frac{d^2 \vec{r}_i}{dt^2} = \frac{\vec{f}_i}{m_i} \quad (5)$$

$$\vec{f}_i = -\nabla_{r_i} \mathcal{V} \quad (6)$$

O algoritmo *leap-frog*, empregado em todas as simulações deste trabalho, é um dos integradores mais utilizados em dinâmica molecular por conciliar implementação simples com boa precisão numérica (Leach, 2001). Seu nome decorre do método de cálculo utilizado, onde posições e velocidades são computadas em instantes defasados de meio passo, “saltando” um sobre o outro. Inicialmente atribuem-se velocidades $v(t-\delta t/2)$ obtidas da distribuição de Maxwell–Boltzmann; em seguida, conhecendo-se as posições $r(t)$ e as forças correspondentes, calcula-se a velocidade no meio passo seguinte e, a partir dela, a posição no próximo passo completo, conforme as Equações 7.1 e 7.2.

$$v_i(t + \frac{1}{2}\delta t) = v_i(t - \frac{1}{2}\delta t) + \frac{\delta t}{m_i} f_i \quad (7.1)$$

$$r(t + \delta t) = r(t) + \delta t v(t + \frac{1}{2}\delta t) \quad (7.2)$$

2.1.3 Controle da temperatura e pressão de uma simulação

Em muitas aplicações de dinâmica molecular se deseja amostrar um *ensemble* termodinâmico bem definido, como o canônico (NVT) ou o isobárico-isotérmico (NPT). Para manter a temperatura da simulação constante utiliza-se um termostato, isto é, um algoritmo que acopla o sistema a um “banho térmico” virtual. Entre as opções mais utilizadas, destaca-se o *Velocity Rescale (V-rescale)* descrito por Bussi, Donadio e Parrinello; (2007). Nesse método, a cada n passos de integração, todas as velocidades são multiplicadas por um fator α (Equação 8) calculado para que a energia cinética instantânea K flutue em torno de K_t , um valor sorteado da distribuição canônica de energias cinéticas. A transferência de energia entre o sistema e o banho térmico é expressa pela Equação 9, na qual τ_T determina a intensidade do acoplamento e dW corresponde ao incremento de Wiener, termo de ruído introduzido pelo tratamento estocástico utilizado pelos autores (Bussi; Donadio; Parrinello, 2007).

$$\alpha = \sqrt{\frac{K_t}{K}} \quad (8)$$

$$\frac{dK}{dt} = \frac{(\langle K \rangle - K)}{\tau_T} + 2\sqrt{\frac{\langle K \rangle K}{3N\tau_T}} \frac{dW}{dt} \quad (9)$$

Para controlar a pressão em simulações de dinâmica molecular empregam-se barostatos, algoritmos que ajustam o volume da caixa simulação até que a pressão interna se equilibre com o valor alvo p_0 . O algoritmo mais simples, ainda amplamente usado para etapas de equilíbrio, é o barostato de Berendsen (Berendsen *et al.*, 1984). Nele a taxa de mudança na pressão é dada pela Equação 10, onde p é a pressão instantânea e τ_p é a constante de acoplamento. A equação produz um redimensionamento isotrópico suave do volume, eficiente para levar o sistema rapidamente à região de pressão desejada, mas suprime flutuações termodinâmicas e, portanto, não amostra rigorosamente o ensemble NPT (Ke *et al.*, 2022; Mark Abraham *et al.*, 2024). Para fases de produção, que exigem estatística correta, recorre-se a barostatos mais sofisticados, como o Parrinello–Rahman (Parrinello; Rahman, 1981) ou o *C-rescale* (Bernetti; Bussi, 2020), adotados em algumas etapas deste trabalho.

$$\frac{dp}{dt} = \frac{(p_0 - p)}{\tau_p} \quad (10)$$

2.2 CÁLCULO DE PROPRIEDADES EM SIMULAÇÕES DE DINÂMICA MOLECULAR

2.2.1 Coeficiente de autodifusão

O coeficiente de autodifusão D é obtido a partir da equação de Einstein, que relaciona D ao deslocamento quadrático médio (*Mean Square Displacement* - MSD) das moléculas ao longo do tempo. Em três dimensões, a expressão é dada pela Equação 11:

$$D = \lim_{t \rightarrow \infty} \frac{\langle |r(t) - r(0)|^2 \rangle}{6t} \quad (11)$$

Onde t é o tempo da simulação e $r(0)$ é a posição do centro de massa da molécula ou átomo no início do tempo considerado. Assim, D é determinado pela inclinação da curva de MSD versus tempo no regime difusivo (linear em t). A abordagem de Einstein é robusta e amplamente utilizada na literatura para calcular difusão em MD, apesar de não ser a única (Maginn *et al.*, 2020). É importante notar que o valor de D obtido em simulações depende do tamanho da caixa periódica, de modo que os coeficientes calculados aumentem com o tamanho do sistema simulado, idealmente

exigindo a aplicação de correções analíticas (Maginn *et al.*, 2020; Yeh; Hummer, 2004). Na prática, o tamanho das caixas de simulação utilizada nos dias atuais são grandes o bastante que tornam o efeito da PBC no valor de D , na maioria dos casos, irrelevante.

2.2.2 Constante dielétrica

Em sistemas líquidos isotrópicos e em equilíbrio, a constante dielétrica estática ϵ pode ser calculada a partir das flutuações do momento dipolar total M do sistema. A expressão empregada é:

$$\epsilon = 1 + \frac{\langle M^2 \rangle - \langle M \rangle^2}{3 \epsilon_0 k_b T V} \quad (12)$$

Onde ϵ_0 é a permissividade elétrica do vácuo; k_b é a constante de Boltzmann; V é o volume da caixa de simulação e T a temperatura do sistema. O fator $1/3$ expressa a média sobre as três direções cartesianas exigida pela isotropia (Neumann, 1983).

2.2.3 Entalpia de vaporização

A entalpia de vaporização molar ΔH_{vap} é estimada pela diferença entre a energia do sistema nos estados gasoso e líquido, incluindo o termo pV_m . Em simulações de DM, avalia-se a energia potencial média (Alguns *softwares* já fornecem diretamente a entalpia média) no líquido e no gás — uma única molécula isolada, para aproximar o sistema a um gás ideal. Considerando que para um gás ideal, $pV_m = RT$, tem-se a Equação 13:

$$\Delta H_{\text{vap}} = (E_g + RT) - \frac{H_l}{N} \quad (13)$$

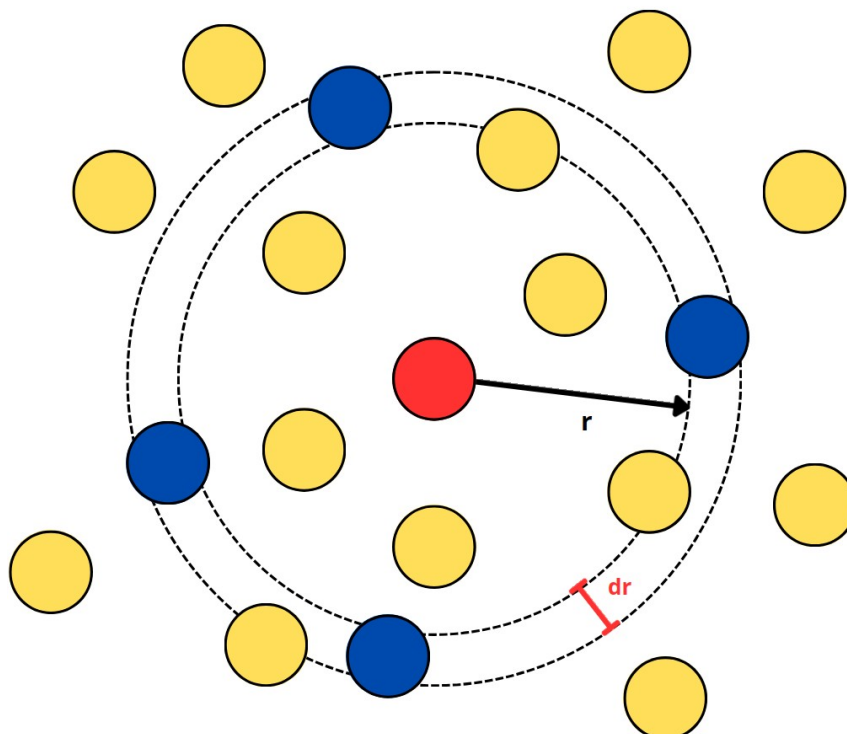
Nesta equação, E_g corresponde à energia molar total do sistema na fases gasosa; H_l corresponde à entalpia molar total do sistema na fase líquida; N é o número de moléculas na fase líquida; R é a constante universal dos gases; e T é a temperatura de ambos sistemas.

2.3 A FUNÇÃO DE DISTRIBUIÇÃO RADIAL

A função de distribuição radial (*Radial Distribution Function* - RDF), usualmente denotada por $g(r)$, é a principal grandeza estatística usada para descrever a estrutura local de líquidos (Allen;

Tildesley, 2017; Madeira; Vitiello, 2012). Ela mede como a densidade local de partículas varia em torno de uma partícula de referência em comparação com a densidade média do sistema. Na prática, escolhe-se arbitrariamente uma partícula como origem e calcula-se quantas outras partículas cujos centros encontram-se numa casca esférica entre r e $r+dr$, como representado na Figura 1.

Figura 1 - Esquema do cálculo da função distribuição radial $g(r)$



Fonte: elaborado pelo próprio autor

Nota: Representação 2D da contagem realizada para o cálculo de $g(r)$. A partícula de referência (círculo vermelho) encontra-se no centro de uma série de anéis concêntricos (linhas tracejadas). A seta preta indica a distância radial r até a borda interna da região considerada, enquanto dr representa sua espessura. As esferas azuis ilustram moléculas contadas dentro dessa casca. As esferas amarelas simbolizam demais partículas do fluido, distribuídas fora do intervalo radial analisado.

Por definição, a densidade local de partículas região é dado pelo produto $\rho_0 \times g(r)$, em que ρ_0 é a densidade *bulk* do fluido. Logo, a função de distribuição radial pode ser dada pela Equação 14:

$$g(r) = \frac{\rho(r)}{\rho_0} \quad (14)$$

Assim, $g(r)$ representa uma densidade local normalizada: quando $g(r) = 1$ a densidade local coincide com a densidade média, enquanto $g(r) > 1$ indica densidade local acima da média e $g(r) <$

1, abaixo da média — por isso que $g(r)$ esta associada a estrutura local de um sistema. O fato de $g(r)$ estar relacionada à densidade local implica que o número médio de partículas entre r e $r+dr$ a partir de uma partícula de referência corresponde a uma casca concêntrica de espessura dr .

$$N(dr) = 4\pi r^2 \rho_0 g(r) dr \quad (15)$$

Considerando que a região do primeiro pico da função de distribuição radial reflete os vizinhos de primeira coordenação, o segundo pico os vizinhos de segunda coordenação, etc. A Equação 16 fornece o número médio de vizinhos de primeira coordenação de cada partícula, ou seja, o número de coordenação N_c (Hansen; McDonald, 2013).

$$N_c = 4\pi \rho_0 \int_0^{r_m} r^2 g(r) dr \quad (16)$$

Onde r_m é a posição do primeiro mínimo de $g(r)$.

Além de descrever a estrutura local do fluido, a função distribuição radial também serve de ponte para propriedades termodinâmicas, como o potencial médio de interação entre duas partículas $v(r)$ descrito pela Equação 17 (Hansen; McDonald, 2013). Essa igualdade decorre de que, no ensemble canônico, a probabilidade relativa de encontrar uma partícula a uma distância r da referência é proporcional a $g(r)$, sendo então possível estabelecer a relação com o potencial de interação (Hansen; McDonald, 2013).

$$v(r) = -k_b T \ln(g(r)) \quad (17)$$

2.4 FUNDAMENTOS DA APRENDIZAGEM DE MÁQUINA

Os algoritmos de aprendizado de máquina são categorizados, em linhas gerais, em duas classes: modelos de aprendizado supervisionado e modelos de aprendizado não supervisionado. Para problemas de classificação e regressão, que constituem o escopo deste trabalho, recorre-se predominantemente ao aprendizado supervisionado (Goodfellow; Bengio; Courville, 2016). Nessa

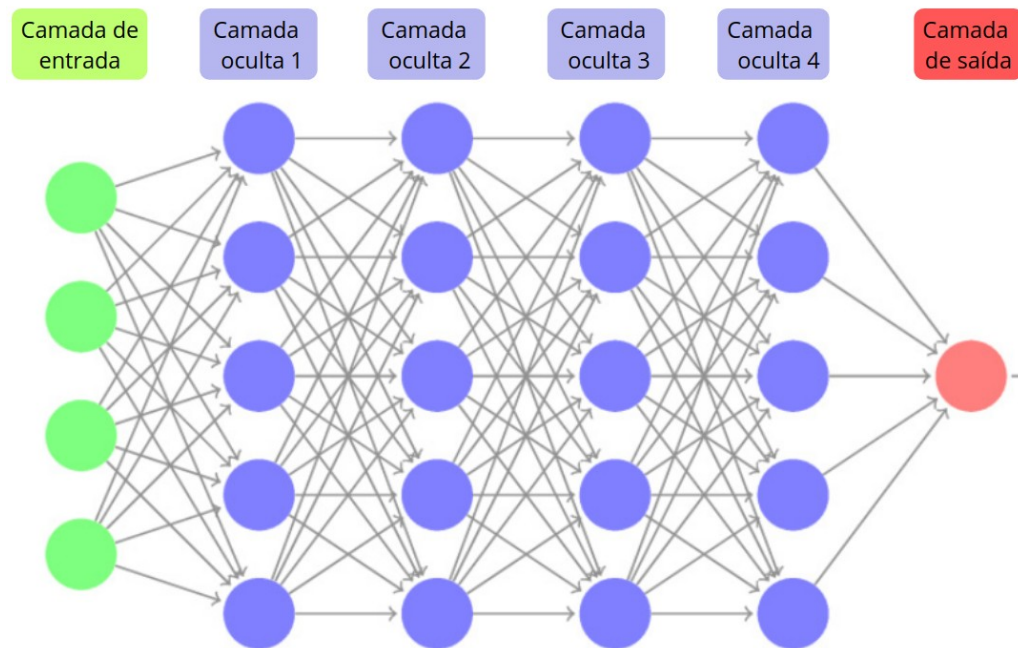
abordagem, busca-se ajustar uma função capaz de associar, com o menor erro possível, um vetor de entrada $\{x\}$ ao respectivo vetor de saída $\{y\}$. Durante o treinamento, cada exemplo é apresentado ao modelo como um par rotulado $(\{x\}, \{y\})$. Embora, historicamente, os vetores de saída fossem fornecidos manualmente por pessoas, originando o termo “supervisionado”, a nomenclatura aplica-se hoje a qualquer técnica que utiliza conjuntos de dados rotulados, independentemente da forma como as saídas $\{y\}$ foram obtidas (Goodfellow; Bengio; Courville, 2016). O perceptron de múltiplas camadas (*Multi-Layer Perceptron* - MLP) é uma arquitetura de rede neural frequentemente usada em tarefas supervisionadas, devido a alta capacidade de aproximar qualquer função contínua, independente da dimensionalidade, tornando-os extremamente expressivos para problemas complexos (Goodfellow; Bengio; Courville, 2016).

A estrutura básica de um MLP consiste em camadas totalmente conectadas de neurônios artificiais, onde há uma camada de entrada (que recebe o vetor entrada $\{x\}$), uma ou mais camadas ocultas e uma camada de saída (que fornece a predição final $\{y\}$), conforme ilustrado na Figura 2. Cada neurônio em uma camada k recebe sinais x_i das camadas anteriores, calcula uma combinação linear ponderada (usando pesos w_i e um termo de viés b) e aplica uma função de ativação não linear $g(\cdot)$, de modo que seu valor de saída pode ser descrito pela Equação 18. Todos os neurônios de uma camada transmitem suas saídas para todos os neurônios da camada seguinte, formando uma rede totalmente conectada.

$$y = g\left(\sum_i w_i x_i + b\right) \quad (18)$$

A função de ativação g serve para introduzir não linearidade nos MLPs, com o tipo específico sendo definida pelo usuário. Exemplos comuns incluem a sigmoide e a função ReLU (*Rectified Linear Unit*).

Figura 2 - Exemplo da arquitetura de um MLP



Fonte: adaptado de Pérez-Enciso e Zingaretti; (2019)

Nota: exemplo de um MLP contendo quatro camadas ocultas: cada seta entre neurônios de camadas adjacentes representa um peso distinto w_i . Para cada neurônio, o conjunto desses pesos multiplica as entradas correspondentes; a soma ponderada resultante é então passada pela função de ativação, gerando a saída conforme a Equação 18.

Para se realizar o treinamento do modelo, define-se uma função perda que mede o erro entre as saídas previstas e as corretas. Um exemplo é o erro quadrático médio (MSE), normalmente usado em problemas de regressão (Goodfellow; Bengio; Courville, 2016), dado pela Equação 19.

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N (y_i(\theta) - t_i)^2 \quad (19)$$

Onde N é o número total de exemplos do conjunto de treino, e θ representa vetor de parâmetros treináveis do modelo, neste caso, θ inclui todos os pesos w e vieses b . $y_i(\theta)$ é a saída (predição) que o modelo, com parâmetros θ , gera para a entrada i ; e t_i é o valor (rótulo) correto para esse mesmo exemplo.

O processo de treinamento do MLP envolve duas fases principais em cada iteração: propagação direta (*forward pass*) e retropropagação (*backpropagation*) (Goodfellow; Bengio; Courville, 2016; Pérez-Enciso; Zingaretti, 2019). Na propagação direta, os sinais de entrada percorrem a rede camada a camada até a saída, onde se calcula o erro em relação ao valor esperado pela Equação 19. Na fase de retropropagação, são calculadas as derivadas parciais do erro em

relação a cada peso e viés, propagando o sinal de erro de volta através das camadas. Os gradientes de erro obtidos permitem atualizar os parâmetros por meio de métodos baseados em gradiente, movendo os pesos numa direção que reduz o valor de L . O processo é repetido iterativamente até que θ tenha convergido.

Ao final do treinamento, o MLP treinado é representado por um conjunto fixo de pesos e vieses otimizados, $\theta = (\{w_i\}, \{b_i\})$. Em outras palavras, o modelo resultante consiste apenas nesses parâmetros numéricos, que representam o aprendizado obtido. Após a convergência, novos padrões de entrada são processados por meio desse conjunto estático de parâmetros para produzir as previsões, sem mais ajustes.

3 OBJETIVOS

O presente estudo visou desenvolver, por meio de aprendizado de máquina, um conjunto otimizado de parâmetros de Lennard-Jones (σ , ϵ) e de cargas parciais do oxigênio da hidroxila capaz de reproduzir, de forma simultânea, a estrutura local e propriedades termodinâmicas e de transporte de álcoois lineares (etanol a 1-hexanol), bem como avaliar a transferibilidade desses parâmetros para sistemas multicomponentes de maior complexidade.

3.1 OBJETIVOS ESPECÍFICOS

Especificamente, este trabalho buscou:

- I) Caracterizar o desempenho dos campos de força de referência (CHARMM36, GAFF e OPLS-AA) quantificando, para os 5 álcoois, os desvios simulados em $g(r)$, densidade (ρ), coeficiente de autodifusão (D), constante dielétrica (ϵ) e entalpia de vaporização (ΔH_{vap}) em relação aos dados experimentais;
- II) Treinar e aplicar um conjunto de redes neurais para mapear o espaço multidimensional dos parâmetros (σ , ϵ , dq) do oxigênio e minimizar a função custo, que captura o desvio entre as propriedades experimentais acima e as propriedades simuladas;
- III) Selecionar o conjunto de parâmetros ótimos para cada álcool, por meio de um ciclo iterativo de “predição, simulação e retreinamento”, nos quais as propriedades experimentais de interesse sejam bem representadas;
- IV) Avaliar o impacto dos parâmetros otimizados em três sistemas multicomponentes: (i) mistura ternária etanol/DIPE/água, (ii) fragmento da proteína β -amiloide em solução etanólica e (iii) micela reversa de DTAB em 1-hexanol.

4 METODOLOGIA

4.1 ESTRUTURAS E TOPOLOGIAS

Para cada campo de força empregado, as topologias e os parâmetros das moléculas foram obtidos por meio de *webservers* específicos. No caso do campo de força GAFF (Wang *et al.*, 2004), utilizou-se o servidor ACPYPE (Kagami *et al.*, 2023) para gerar as topologias e os parâmetros dos álcoois etanol, 1-propanol, 1-butanol, 1-pentanol e 1-hexanol, atribuindo-se as cargas parciais pelo método semi-empírico AM1-BCC (Jakalian; Jack; Bayly, 2002).

Para o campo de força OPLS-AA (Jorgensen; Maxwell; Tirado-Rives, 1996), empregou-se o servidor LigParGen (Dodda *et al.*, 2017b) para parametrizar os mesmos cinco álcoois, adotando-se o modelo de cargas 1.14*CM1A-LBCC (Dodda *et al.*, 2017a). Por fim, para o campo de força CHARMM36 (Best *et al.*, 2012), as estruturas e os parâmetros dos álcoois foram gerados via CharmmGUI (Kim *et al.*, 2017), utilizando um esquema de incrementos “carga-ligação” baseado no banco de dados do próprio campo de força (Vanommeslaeghe; Raman; MacKerell, 2012).

Adicionalmente, para os sistemas multicomponentes também simulados com CHARMM36, as topologias do éter di-isopropílico (DIPE), do dodeciltrimetilamônio (DTA) e dos resíduos 15–27 da β -amiloide cristalizada por Coles *et al.* (1998) foram igualmente obtidos no CharmmGUI. Para os sistemas que incluem fase aquosa, empregou-se o modelo de água TIPS3P (MacKerell *et al.*, 1998). Nas simulações das micelas reversas, os parâmetros do íon brometo, utilizado como contra-íon do surfactante, foram obtidos do estudo de Orabi *et al.* (2021).

4.2 PARÂMETROS DAS SIMULAÇÕES DE DINÂMICA MOLECULAR

Foram conduzidas simulações em quatro tipos de sistemas:

- I) Sistemas de álcoois puros (etanol, 1-propanol, 1-butanol, 1-pentanol e 1-hexanol).
- II) Sistema ternário composto por DIPE, etanol e água.
- III) Sistema contendo um fragmento da proteína β -amiloide em solução etanólica.
- IV) Sistema contendo uma micela reversa de brometo de dodeciltrimetilamônio (DTAB) imersa em 1-hexanol.

Todas as simulações foram executadas com o GROMACS 2023.5 (Mark Abraham *et al.*, 2024). As equações de movimento foram integradas pelo método *Leap-Frog*, utilizando um passo de integração de 2 fs. As interações eletrostáticas de longo alcance foram tratadas pelo método *Particle-Mesh Ewald* (PME) (Darden; York; Pedersen, 1993), com espaçamento de grade de 0,12 nm. A velocidade translacional do centro de massa foi removida a cada 10 passos das simulações. Para manter as ligações envolvendo átomos de hidrogênio rígidas, aplicou-se o algoritmo LINCS (Hess *et al.*, 1997). Condições de contorno periódicas foram impostas em todas as direções.

Antes de qualquer simulação, as geometrias foram otimizadas por minimização de energia utilizando o algoritmo *steepest descent*, com passo de 0,001 nm. O protocolo de equilíbrio e produção variou de acordo com o sistema, mas, salvo indicação em contrário, aplicou-se um raio de corte de 1,5 nm para o PME, e as interações de Lennard-Jones de curto alcance foram suavemente zeradas no intervalo de 1,4 a 1,5 nm. Todas as simulações realizadas com acoplamento de pressão tiveram a compressibilidade do sistema definida para $4,5 \times 10^{-5} \text{ bar}^{-1}$.

4.2.1 Parâmetros dos sistemas de álcoois puros

Os procedimentos de equilíbrio para os sistemas de álcoois puros, utilizando tanto os parâmetros originais dos campos de força quanto o campo CHARMM36 com modificações, iniciaram-se com uma etapa de equilíbrio em *ensemble* NVT de 1 ns a 298 K, mantidos pelo termostato *V-rescale* (Bussi; Donadio; Parrinello, 2007), com constante de acoplamento (τ_T) de 0,1 ps, e com acoplamento térmico a cada 100 passos da simulação. Na sequência, procedeu-se ao equilíbrio em *ensemble* NPT também de 1 ns a 1 bar, controlado pelo barostato de Berendsen (Berendsen *et al.*, 1984) com uma constante de acoplamento (τ_P) de 1 ps, e acoplamento a cada 100 passos, mantendo-se a temperatura em 298 K pelo mesmo termostato *V-rescale*.

Para a fase de produção, os sistemas com parâmetros originais foram submetidos a 12 ns em *ensemble* NPT a 1 bar, utilizando o barostato de Parrinello–Rahman (Parrinello; Rahman, 1981) ($\tau_P = 10 \text{ ps}$) com acoplamento a cada 50 passos, enquanto o conjunto utilizando CHARMM36 modificado seguiu por pelo menos 8 ns sob as mesmas condições de temperatura (*V-rescale*, $\tau_T = 0,1 \text{ ps}$, acoplamento a cada 100 passos) e condições similares de pressão (Parrinello–Rahman, $\tau_P = 8 \text{ ps}$, acoplamento a cada 20 passos).

Além das simulações em fase condensada, uma molécula isolada de cada álcool foi simulada em vácuo para representar a fase gasosa. Essas simulações foram conduzidas em *ensemble* NVT por pelo menos 100 ns, utilizando os mesmos parâmetros de controle de temperatura aplicados nas simulações em fase líquida.

4.2.2 Parâmetros do sistema ternário

No sistema ternário de etanol, DIPE e água, o equilíbrio NVT de 1 ns foi conduzido a 298,15 K com o termostato *V-rescale* ($\tau_T = 0,1$ ps) e acoplamento a cada 100 passos, seguido de 1 ns em *ensemble* NPT a 1 bar empregando o barostato *C-rescale* (Bernetti; Bussi, 2020) ($\tau_P = 8$ ps) com acoplamento a cada 20 passos, mantendo-se a temperatura pelos mesmos parâmetros do NVT. A fase de produção consistiu em 150 ns em *ensemble* NPT, sob as mesmas condições de termostato e barostato aplicadas na etapa de equilíbrio.

4.2.3 Parâmetros do sistema contendo um fragmento da proteína β -amiloide

Para o fragmento de β -amiloide em solução etanólica, foi adotado um *cutoff* do PME de 1,2 nm e suavização das interações de Lennard-Jones entre 1,0 e 1,2 nm. O equilíbrio inicial em NPT durou 1 ns a 298,15 K e 1 bar, com *V-rescale* e *C-rescale* utilizando os parâmetros descritos na Seção 4.2.2, mantendo os átomos pesados do peptídeo restritos por um potencial quadrático com uma constante de força de $2092 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$. Após a etapa de equilibração, as restrições foram removidas e a produção transcorreu por 300 ns em *ensemble* NPT sob as mesmas condições de termostato e barostato.

4.2.4 Parâmetros do sistema contendo uma micela reversa

A micela de brometo de dodeciltrimetilamônio (DTAB + H_2O) foi inicialmente equilibrada no vácuo em *ensemble* NVT por 2 ns, a 200 K, empregando o termostato *V-rescale* ($\tau_T = 0,1$ ps; acoplamento a cada 100 passos). Concluída essa etapa, adicionaram-se moléculas de 1-hexanol ao sistema, que então passou por minimização de energia (vide Seção 4.2) seguida de 1 ns de equilibração NVT sob as mesmas condições de temperatura e termostato. Em seguida, realizou-se 1

ns em ensemble NPT a 200 K e 1 bar, mantendo-se o termostato V-rescale e aplicando o barostato de Berendsen ($\tau_P = 1$ ps; acoplamento a cada 100 passos).

Para aquecer o sistema de forma controlada, executou-se uma rampa de temperatura composta de três etapas sucessivas de 5 ns em ensemble NPT a 200, 230 e 260 K, respectivamente, todas a 1 bar. Cada etapa utilizou o termostato V-rescale ($\tau_T = 0,1$ ps; acoplamento a cada 100 passos) em conjunto com o barostato de Parrinello–Rahman ($\tau_P = 8$ ps; acoplamento a cada 20 passos). Após o aquecimento gradual, a fase de produção foi conduzida em ensemble NPT a 298 K e 1 bar pelo mesmo par de termostato e barostato, durante até 150 ns.

Também simulou-se um sistema contendo apenas água TIPS3P, que foi empregado posteriormente na construção da micela (Seção 4.7). O protocolo incluiu 1 ns de equilíbrio em ensemble NPT a 298,15 K e 1bar, ambos controlados pelo termostato *V-rescale* ($\tau_T = 0,1$ ps; acoplamento a cada 100 passos) e pelo barostato *C-rescale* ($\tau_P = 8$ ps; acoplamento a cada 20 passos) já adotados nos demais sistemas. A etapa de produção transcorreu por mais 10 ns em ensemble NPT, mantendo-se as mesmas configurações de termostato e barostato.

4.3 SISTEMAS DE ÁLCOOIS PUROS

Caixas cúbicas de simulação, com arestas de 7,5 nm, foram construídas utilizando o software Packmol (Martínez *et al.*, 2009). A Tabela 1 apresenta os cinco álcoois investigados, listando suas densidades experimentais (Haynes, 2014) (em $\text{kg}\cdot\text{m}^{-3}$), as correspondentes densidades moleculares (em nm^{-3}) e o número total de moléculas incluídas nas simulações de cada composto.

Tabela 1 - Densidades experimentais, densidades moleculares e número de moléculas simuladas para álcoois puros

Álcool	Densidade / $\text{kg}\cdot\text{m}^{-3}$	Densidade / nm^{-3}	Nº de moléculas simuladas
Etanol	789.3	10.3	4351
1-Propanol	799.5	8.01	3394
1-Butanol	809.5	6.58	2775
1-Pentanol	814.4	5.56	2347

1-Hexanol	813.6	4.80	2038
------------------	-------	------	------

Fonte: elaborado pelo próprio autor

Durante a análise dos resultados, um intervalo inicial de 1 a 2 ns da fase de produção foi desconsiderado, a fim de garantir que os sistemas estivessem plenamente equilibrados. As propriedades calculadas incluíram a energia total do sistema, a função de distribuição radial entre os átomos de oxigênio, as densidades médias das caixas de simulação, o coeficiente de autodifusão e a constante dielétrica. Todos esses parâmetros foram obtidos por meio das ferramentas disponíveis no pacote GROMACS (Mark Abraham *et al.*, 2024). Adicionalmente, os valores da entalpia de vaporização molar, ΔH_{vap} , foram estimados com base nas energias e entalpias médias das fases gasosa e líquida, respectivamente, utilizando a Equação 13 descrita na Seção 2.

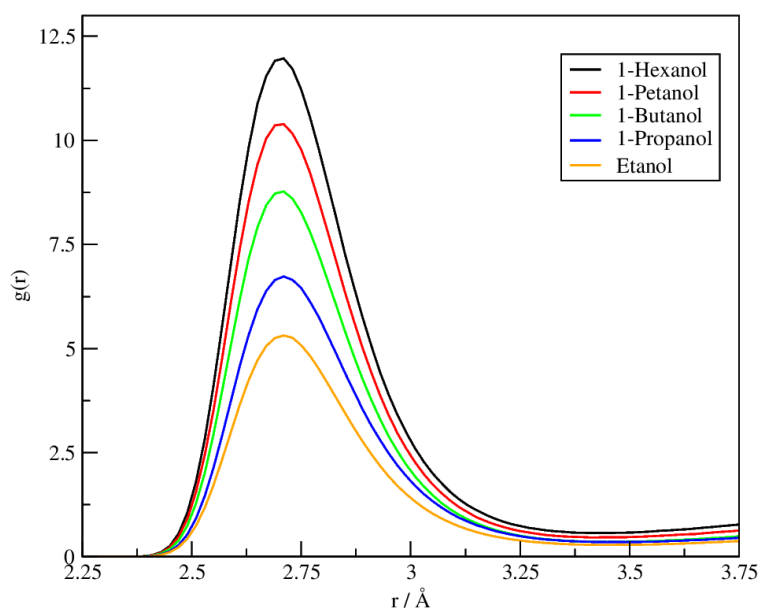
Todos os resultados dos sistemas de álcoois puros calculados neste trabalho também foram avaliados em termos do desvio relativo médio, DRM, das propriedades de interesse calculado pela Equação 20.

$$DRM = \frac{1 - R^2 + \sum_{i=1}^4 |X_i / X_{i,0} - 1|}{5} \quad (20)$$

Onde R^2 é o coeficiente de determinação entre a função de distribuição radial simulada, $g_{\text{sim}}(r)$, e a experimental de Tomšič *et al.* (2007), $g_{\text{exp}}(r)$. A comparação foi feita no intervalo de 2,3 Å até o primeiro mínimo de $g_{\text{exp}}(r)$ para cada álcool (Apêndice A, Tabela A1); X_i representa o valor simulado da i -ésima propriedade físico-química, enquanto $X_{i,0}$ corresponde ao respectivo valor experimental. Os índices $i=1$ a 4 referem-se, nesta ordem, à densidade, coeficiente de autodifusão, constante dielétrica e entalpia de vaporização molar.

As propriedades experimentais utilizadas de referencia para este cálculo encontram-se no Gráfico 2 e na Tabela 2 abaixo.

Gráfico 2 - Perfis experimentais da função de distribuição radial $g(r)$ para álcoois lineares (etanol – 1-hexanol) a 298 K



Fonte: modificado pelo autor de Tomšič *et al.* (2007)

Tabela 2 - Dados experimentais de densidade (ρ), coeficiente de autodifusão (D), constante dielétrica (ϵ) e entalpia de vaporização molar (ΔH) dos álcoois lineares a 298 K e 1 bar

	Etanol	1-Propanol	1-Butanol	1-Pentanol	1-Hexanol
Densidade bulk / kg.m^{-3}	789,3	799,7	809,5	814,4	813,6
Coeficiente de autodifusão / $10^{-9} \text{ m}^2.\text{s}^{-1}$	1,01	0,552	0,416	0,296	0,218
Constante dielétrica	24,33	20,44	17,54	15,03	13,32
Entalpia de vaporização / kJ.mol^{-1}	42,32	47,45	52,35	57,02	61,61

Fonte: Bezman, Casassa, Kay, (1997); Haynes, (2014); Iwahashi *et al.* (1986); Sastry, Valand, (1997).

4.4 MODELO DE APRENDIZADO DE MÁQUINA

Para prever a relação entre os parâmetros do oxigênio e as propriedades físico-químicas do sistema, foi empregado um conjunto de 100 perceptrons de múltiplas camadas (MLP), implementados em Python por meio da biblioteca *Scikit-learn* (Pedregosa *et al.*, 2011). A arquitetura dos modelos consistiu em três camadas ocultas, cada uma com 100 neurônios, utilizando a função de ativação ReLU (Rectified Linear Unit). A otimização dos pesos foi realizada pelo

algoritmo L-BFGS, com regularização L2 ($\alpha = 0,1$) e critério de tolerância de 10^{-4} para convergência. Cada um dos 100 MLPs individuais foi inicializado com uma *seed* pseudoaleatória distinta, garantindo variação nos pesos iniciais.

O conjunto de treinamento foi inicialmente gerado a partir de 54 simulações de álcoois puros. Como variáveis de entrada foram escolhidos: o tipo de álcool (codificado por *one-hot encoding*), os parâmetros de Lennard-Jones σ e ϵ relativos à interação oxigênio–oxigênio e a diferença de carga pontual entre o oxigênio da hidroxila e seu hidrogênio (dq). Para ampliar a amostragem do espaço de fase, σ foi escalado pelos fatores 0,975 e 0,95¹; ϵ pelos fatores 0,9 e 1,1¹; e dq pelos fatores 0,95 e 1,05, preservando-se a carga total da hidroxila (−0,23 em etanol e −0,231 nos demais álcoois, conforme gerado originalmente pelo CharmmGUI). Cada combinação de parâmetros foi simulada duas vezes em álcoois distintos (Apêndice A , Tabela A2), totalizando 54 simulações para o treinamento do modelo.

Como variáveis resposta do modelo, foi utilizado o seguinte conjunto de propriedades:

- I) Coeficiente de determinação, R^2 ;
- II) Densidade média da caixa de simulação, ρ ;
- III) Coeficiente de autodifusão, D ;
- IV) Constante dielétrica, ϵ ;
- V) Entalpia de vaporização molar, ΔH_{vap} , calculada pela Equação 13.

E, em alguns casos, o valor máximo de $g_{sim}(r)$ foi explicitamente considerado. Todas as propriedades foram calculadas de acordo com a seção 4.3.

Após o treinamento do conjunto de 100 MLPs, cada um com pesos iniciais distintos, as predições individuais foram agregadas pela média aritmética para obter um único valor de saída por propriedade desejada.

Para determinar a combinação de parâmetros (σ, ϵ, dq) que melhor reproduz este conjunto de propriedades, definiu-se uma função alvo f dada pela Equação 21:

$$f = a(1 - R^2) + b|\rho/\rho_0 - 1| + c|D/D_0 - 1| + d|\epsilon/\epsilon_0 - 1| + e|\Delta H_{vap}/\Delta H_{vap,0} - 1| \quad (21)$$

1. Em relação aos parâmetros atômicos do tipo OG311 referentes aos átomos de oxigênio de todos os álcoois, exceto o etanol.

onde ρ_0 , D_0 , ϵ_0 e $\Delta H_{\text{vap},0}$ são os valores experimentais correspondentes, e os coeficientes a , b , c , d , e foram determinados como o inverso da raiz quadrada do desvio quadrático médio de cada propriedade no conjunto de treinamento, conforme ilustrado nas Equações 22.1 e 22.2.

$$a = \langle 1 - R^2 \rangle^{-\frac{1}{2}} \text{ ou } x = \langle |X/X_0 - 1| \rangle^{-\frac{1}{2}} \quad (22.1) \text{ e } (22.2)$$

Na Equação 22.2, x corresponde a qualquer um dos coeficientes “b” a “e”, e X é a propriedade experimental associada. As propriedades experimentais empregadas na definição da função alvo e em seus coeficientes encontram-se resumidas no Gráfico 2 e na Tabela 2.

Por fim, o conjunto de 100 MLPs foi utilizado, de forma iterativa, para determinar - individualmente para cada um dos cinco álcoois - os trio de parâmetros (σ , ϵ , dq) que minimizavam a função-alvo f . A cada ciclo, os parâmetros preditos foram utilizados em novas simulações de dinâmica molecular, cujas propriedades obtidas eram utilizados para atualizar o conjunto de dados de treinamento (Apêndice A, Tabelas A3 - A5), permitindo o retreinamento do modelo. Em todas as iterações, a função-alvo incorporou uma penalização unitária quando o valor predito de R^2 decaía abaixo de 0,95.

Especificamente para 1-butanol, 1-pentanol e 1-hexanol, o valor do primeiro máximo de $g_{\text{sim}}(r)$, foi incluído tanto nas variáveis respostas do modelo quanto no cálculo de uma nova função alvo, como mostra a Equação 23.

$$f' = f + \gamma |g_{\text{sim}}(r_{\text{max}}) / g_{\text{exp}}(r_{\text{max}}) - 1| \quad (23)$$

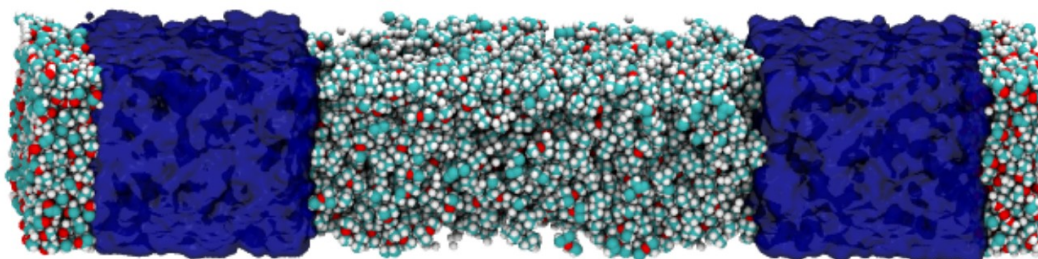
Onde f é a função alvo definida na Equação 21; $g(r_{\text{max}})$ corresponde ao valor máximo de $g(r)$, e γ é um coeficiente calculado pela Equação 22.2. Para estes mesmos 3 álcoois, um termo adicional de penalidade igual a 2 é somado a função alvo f' sempre que a discrepância entre os coeficientes de autodifusão experimental e computacional excedia 10 %.

O ciclo de “predição, simulação e retreinamento” foi repetido até que os resultados simulados apresentassem boa concordância com os dados experimentais, isto é, baixos desvios nas propriedades-alvo utilizadas nas funções f ou f' . Ao final desse processo, o conjunto de parâmetros que resultou no menor valor da função alvo na última iteração foi selecionado, para cada álcool, e empregado nas simulações de produção com o campo de força CHARMM36 modificado.

4.5 SISTEMA TERNÁRIO DIPE/ÁGUA/ETANOL

Dois sistemas ternários compostos por DIPE, etanol e água foram simulados com o campo de força CHARMM36. Ambos foram construídos no Packmol em caixas retangulares de dimensões $5 \times 5 \times 27$ nm (eixos x, y e z). As moléculas foram distribuídas ao longo do eixo z da seguinte forma: 950 moléculas de etanol ocupando os primeiros 3,5 nm; 1248 moléculas de DIPE distribuídas ao longo de 12 nm centrais; e 9405 moléculas de água preenchendo os intervalos restantes. (Figura 3). No primeiro sistema, empregaram-se os parâmetros originais de etanol de CHARMM36; no segundo, aplicaram-se os parâmetros ajustados pelo protocolo de aprendizado de máquina descrito na Seção 4.4.

Figura 3 - Representação da caixa de simulação utilizada como estrutura inicial

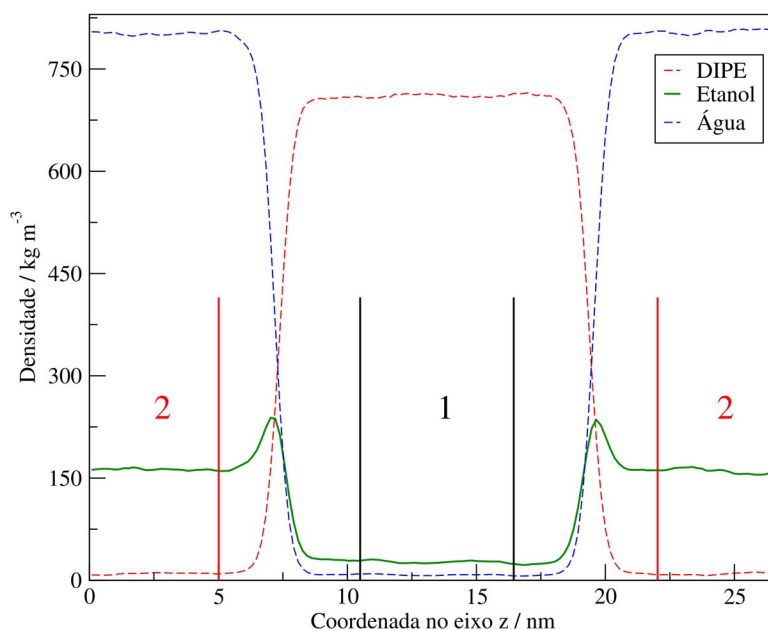


Fonte: elaborado pelo próprio autor

Nota: visualização após centralizar o sistema em DIPE. A água é apresentada como um volume contínuo (azul semitransparente), enquanto DIPE e etanol estão representados explicitamente; átomos de carbono em ciano, hidrogênio em branco e oxigênio em vermelho.

Para quantificar a densidade de etanol nas fases orgânica e aquosa, extraiu-se o perfil de densidade dos três componentes ao longo do eixo z usando as ferramentas de análise do GROMACS. As regiões de *bulk* foram definidas a partir do perfil de densidade ao longo do eixo z: a região de *bulk* orgânico (Região 1) abrange 10,5–16,5 nm, enquanto a região de *bulk* aquoso (Região 2) corresponde aos intervalos 0–5 nm e 22–27 nm (ver Gráfico 3).

Gráfico 3 - Perfil de densidade ao longo do eixo z dos componentes do sistema



Fonte: elaborado pelo próprio autor

Nota: densidades calculadas como médias no intervalo de 100–150 ns para o sistema 1, utilizando os parâmetros originais do campo de força.

As densidades médias de etanol nessas duas regiões (c_1 e c_2 , respectivamente) foram calculadas em janelas de 2 ns ao longo de toda a etapa de produção das simulações. O coeficiente de partição k do etanol foi então obtido pela Equação 24:

$$k = \frac{c_1}{c_2} \quad (24)$$

Após o equilíbrio dos sistemas, investigou-se o número e o tamanho dos agregados de etanol em ambas as fases utilizando o *software Cluster Analysis* (Montenegro, 2025).

4.6 SOLUÇÃO ETANÓLICA DA β -AMILOIDE

O arquivo PDB contendo a estrutura completa do peptídeo β -amilóide (Coles *et al.*, 1998) foi editado para isolar o fragmento correspondente aos resíduos Gln¹⁵–Asn²⁷. As extremidades N-terminal e C-terminal foram funcionalizadas com grupos acetil (ACE) e N-metilamida (NME), respectivamente, conforme descrito por Lemkul, Huang e MacKerell (2015). Em seguida, foi realizada a minimização de energia seguindo o protocolo apresentado na Seção 4.2, aplicando-se

um potencial harmônico (constante de força de $2092 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$) para restringir a posição dos átomos pesados originalmente presentes no peptídeo.

A conformação final da A β 15-27, Ilustrada na Figura 4, foi solvatada em uma caixa cúbica de 6 nm de aresta contendo 2207 moléculas de etanol. Também foi adicionado um íon Na^+ para neutralizar a carga negativa do sistema. Partindo dessa estrutura inicial, foram realizadas duas simulações de dinâmica molecular: uma empregando diretamente os parâmetros originais gerados pelo CharmmGUI e outra utilizando os valores modificados para etanol conforme o protocolo descrito na Seção 4.4. A evolução da estrutura secundária dos resíduos ao longo das simulações foi analisada por meio das ferramentas integradas do GROMACS e do *software* DSSP (Kabsch; Sander, 1983).

Figura 4 - Estrutura do fragmento A β 15-27 funcionalizado utilizada nas simulações em etanol.



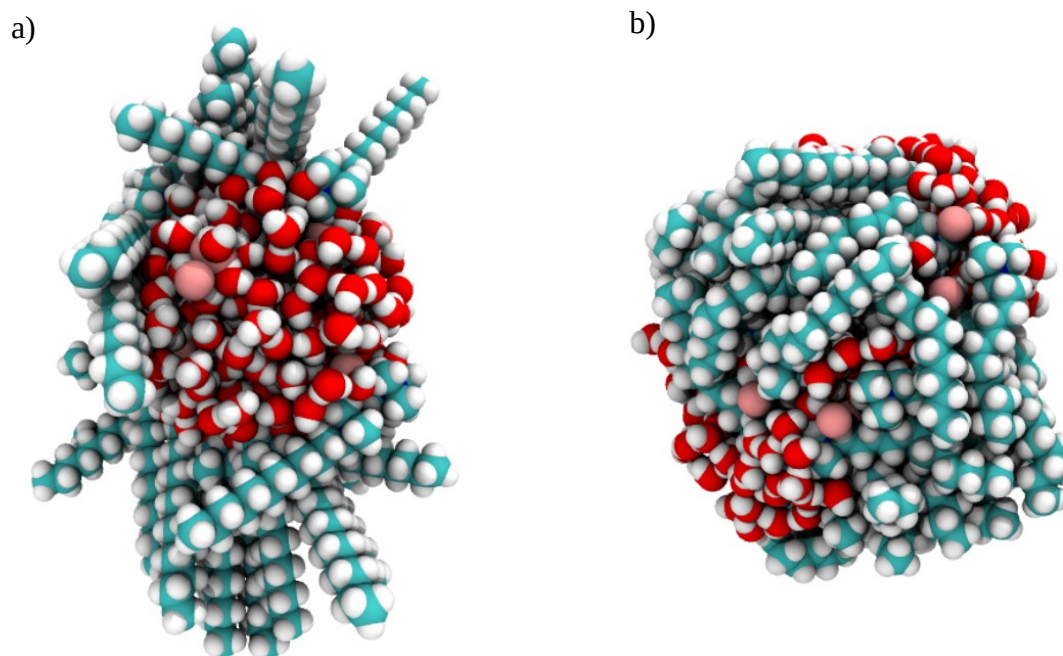
Fonte: elaborado pelo próprio autor

Nota: (a) Sequência primária do fragmento A β 15-27, com as extremidades funcionalizadas pelos grupos ACE e NME; cada resíduo está destacado por uma cor específica. (b) Conformação após a minimização de energia, mostrada em representação *cartoon* com as mesmas cores de resíduos da parte (a). Os átomos explícitos do peptídeo aparecem em modelo bastão-esfera semitransparente (C em ciano, O em vermelho, H em branco, N em azul).

4.7 SISTEMA MICELAR REVERSO

Para construir a micela reversa de brometo de dodeciltrimetilamônio (DTAB), simulou-se inicialmente uma caixa cúbica de 6 nm contendo 7 201 moléculas de água TIPS3P durante 10 ns em ensemble NPT. A partir do último quadro, recortou-se uma esfera de 1,3 nm de raio, correspondendo a 312 moléculas de água. Em seguida, com o Packmol, dispuseram-se 26 moléculas de DTA de modo que seus átomos de nitrogênio apontassem para o núcleo aquoso, adicionando-se também 26 íons Br^- distribuídos isotropicamente. Esse sistema passou então por uma breve simulação no vácuo, permitindo o colapso das caudas hidrofóbicas sobre o núcleo micelar (Figura 5).

Figura 5 - Estrutura inicial da micela reversa de DTAB



Fonte: elaborado pelo próprio autor

Nota: (a) Arranjo gerado no Packmol; Moléculas de DTAB dispostas com as cabeças quaternárias apontadas para o núcleo esférico de 312 moléculas de água (O em vermelho, H em branco, C em ciano e Br^- em rosa). (b) Estrutura após minimização de energia e 2 ns de equilíbrio a vácuo, exibindo o colapso das caudas hidrofóbicas sobre o núcleo aquoso.

A micela enovelada foi solvatada em uma caixa cúbica de 8 nm contendo 2 284 moléculas de 1-hexanol e submetida ao protocolo de equilíbrio e aquecimento descrito na Seção 4.2.4. Empregaram-se dois conjuntos de parâmetros para o solvente: o original do campo de força

CHARMM36 e o conjunto modificado obtido na Seção 4.4. Após o aquecimento controlado até 260 K, realizaram-se três produções independentes em ensemble NPT a 298,15 K e 1 bar:

- I) continuação direta com os parâmetros originais,
- II) continuação direta com os parâmetros modificados,
- III) produção iniciada a partir da configuração aquecida com parâmetros modificados, porém revertendo aos parâmetros originais antes da etapa de produção.

A influência das correções NBFIX foi avaliada monitorando-se a estabilidade micelar ao longo das trajetórias. Essa avaliação incluiu inspeção visual das estruturas e quantificação do número de moléculas de água retidas no núcleo micelar, através do *software Cluster Analysis*.

5 RESULTADOS E DISCUSSÃO

Os resultados estão organizados em seis subseções: Na Seção 5.1, as propriedades obtidas a partir das simulações dos cinco álcoois utilizando os parâmetros originais dos campos de força CHARMM36, GAFF e OPLS-AA, são comparadas aos respectivos valores experimentais;. A Seção 5.2 descreve o protocolo de aprendizado ativo empregado para reparametrizar o campo de força CHARMM36, apresentando a convergência dos parâmetros σ , ϵ e dq , bem como a interpretação física dessas tendências; Em seguida, a Seção 5.3 avalia o impacto das correções NBFIX nos sistemas de álcoois puros, contrastando os novos resultados tanto com os dados experimentais quanto com aqueles obtidos com CHARMM36 original; por fim, as Seções 5.4 a 5.6 testam a transferibilidade dos parâmetros otimizados em três cenários multicomponentes—(i) a mistura ternária DIPE/etanol/água, (ii) a solução etanólica do fragmento A β 15-27 e (iii) a micela reversa de DTAB em 1-hexanol—examinando, respectivamente, a redistribuição de etanol entre as fases, a evolução da estrutura secundária do peptídeo e a estabilidade do agregado micelar, sendo discutindo como o a reparametrização altera o balanço solvente–solvente e solvente–soluto em cada contexto.

5.1 SISTEMAS DE ÁLCOOIS PUROS COM PARÂMETROS ORIGINAIS

Dentro de cada campo de força, os cinco álcoois compartilham exatamente os mesmos parâmetros de Lennard-Jones e as mesmas cargas para os átomos constituintes do grupo hidroxila; a única exceção é o etanol no CHARMM36, cujos valores diferem ligeiramente (Tabela 3). A padronização entre as molécula pertencentes ao mesmo campo de força decorre do princípio de transferibilidade, segundo o qual parâmetros definidos para uma função química devem ser aplicáveis a moléculas estruturalmente semelhantes (Allen; Tildesley, 2017; Leach, 2001)—prática bem estabelecida no desenvolvimento de campos de força (Jorgensen; Maxwell; Tirado-Rives, 1996; Wang *et al.*, 2004).

Tabela 3 - Parâmetros de Lennard-Jones e cargas parciais do oxigênio (σ , ϵ , q_o) e do hidrogênio (q_H) da hidroxila dos álcoois em cada campo de força

Campo de força	σ / nm	ϵ / kJ.mol ⁻¹	q_o	q_H	dq
CHARMM36	0,315	0,636; 0,804	-0,660; -0,650	+0,430; +0,419	1,09; 1,069
GAFF	0,307	0,880	-0,598	+0,396	0,994
OPLS-AA	0,312	0,711	-0,688	+0,406	1,094

Fonte: elaborado pelo próprio autor

Nota: Os valores de q_o e q_H , para os campos de força GAFF e OPLS-AA, possuem uma variação na última casa decimal para os diferentes álcoois, em função do método empregado pelos *web servers* para gerar as cargas, conforme descrito na Seção 4.1. .

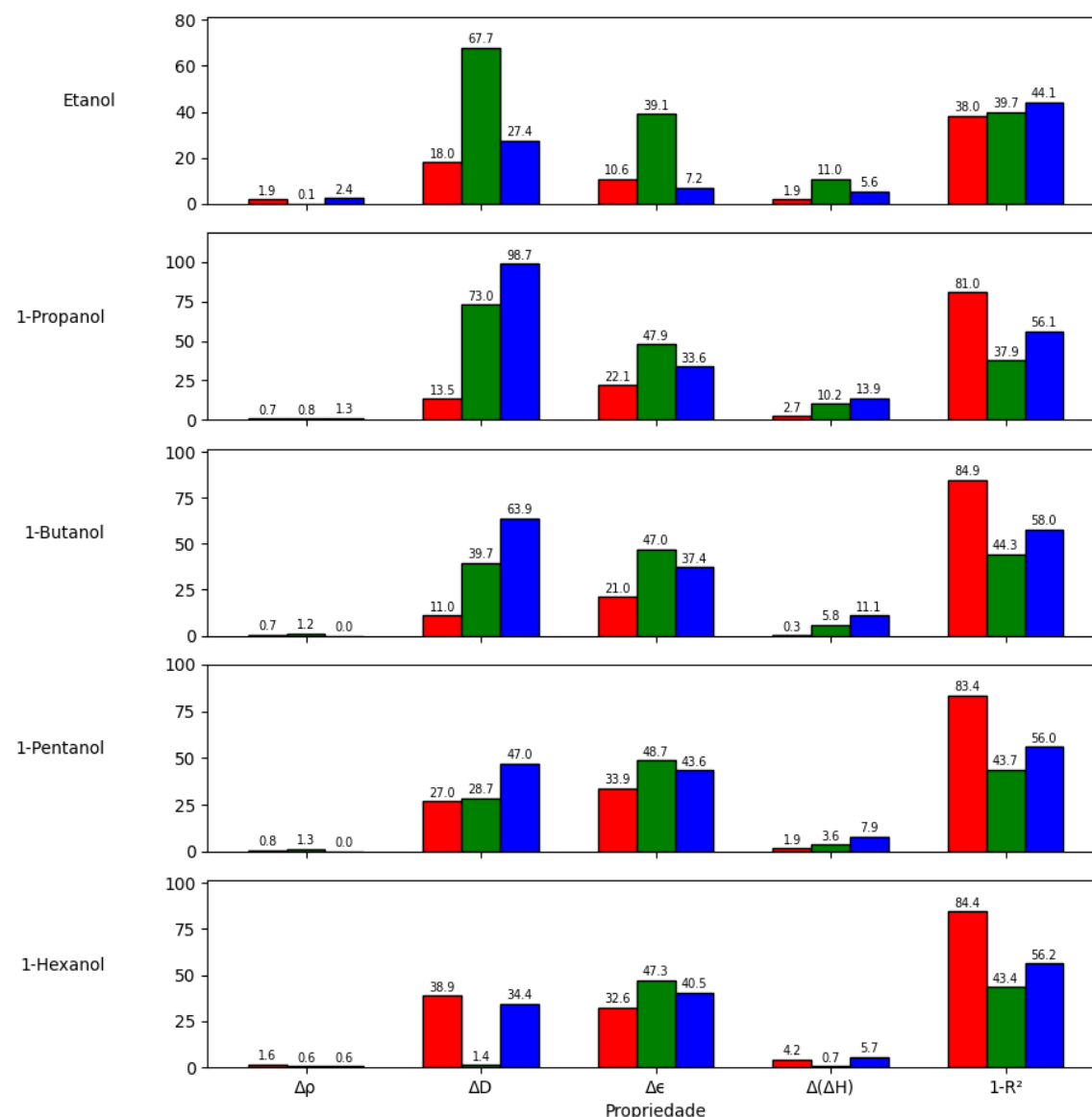
Para avaliar o quão bem cada campo de força reproduz as propriedades experimentais, calculou-se, a partir das propriedades listadas na Seção 4.3, o desvio relativo médio (DRM) via Equação 20. Os valores obtidos estão na Tabela 4, enquanto o Gráfico 4 detalha a contribuição do desvio de cada propriedade ao DRM de cada álcool. A inspeção desse gráfico deixa claro que a maior contribuição em todos os casos provém do baixo coeficiente de determinação associado à função de distribuição radial $g(r)$ simulada. Isso reflete o fato de que todos esses modelos subestimaram sistematicamente a estrutura e intensidade do primeiro pico de $g(r)$ (Gráfico 5 e Apêndice B, Gráfico B5) – correspondente à primeira camada de solvatação – em comparação com as medidas experimentais, indicando uma divergência considerável na descrição da mesma. A constante dielétrica e o coeficiente de autodifusão também se mostraram uma fonte relevante de desvio em todos os álcoois simulados com os 3 campos de força, exceto o 1-hexanol modelo pelo GAFF, que reproduziu o valor experimental de D com uma precisão de 1,4%. Em contraste, propriedades macroscópicas como densidade líquida e entalpia de vaporização foram reproduzidas com boa precisão pelos três campos de força, contribuindo muito pouco para o DRM.

Tabela 4 - Desvio Relativo Médio (DRM) das propriedades simuladas para cada álcool em cada campo de força

Álcool	DRM _{CHARMM36}	DRM _{GAFF}	DRM _{OPLS}
Etanol	14,10%	31,35%	17,34%
1-Propanol	23,98%	33,95%	40,73%
1-Butanol	23,57%	27,62%	34,11%
1-Pentanol	29,37%	25,20%	30,88%
1-Hexanol	32,35%	18,67%	27,49%

Fonte: elaborado pelo próprio autor

Gráfico 4 - Desvio relativo por propriedade e campo de força



Fonte: elaborado pelo próprio autor

Nota: as barras vermelha, verde e azul correspondem, respectivamente, aos valores dos desvios obtidos com os campos de força CHARMM36, GAFF e OPLS-AA.

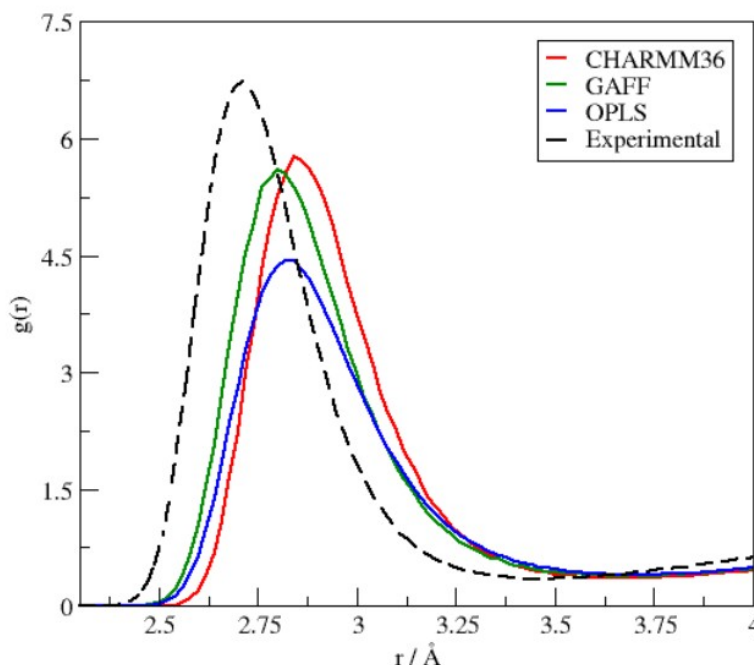
5.1.1 Divergências na função de distribuição radial

As divergências observadas na função de distribuição radial obtida a partir das simulações também são reportadas na literatura para sistemas similares. Obeidat, Al-Salman e Abu-Ghazleh (2018), ao investigarem misturas binárias etanol/água, mostraram que os campos de força TraPPE-UA (Martin; Siepmann, 1998) e OPLS-AA subestimam a altura do primeiro pico de $g_{OO}(r)$ em

sistemas de etanol puro, sugerindo uma representação inadequada da primeira camada de solvatação. Essa subestimação pode ser atribuída principalmente a incapacidade do potencial efetivo de reproduzir a energia de interação entre os átomos oxigênicos, descrita pela Equação 17 e ilustrado no Gráfico 6.

Silberstein e Makov (2022) relatam divergências análogas entre valores experimentais e simulados de $g(r)$ para diversos pares atômicos em soluções de metanol, atribuindo-as à simplificação inerente ao modelo *united-atom* do OPLS utilizado, no qual o grupo metil é representado por um único sítio efetivo. Em conjunto com a análise de Obeidat, Al-Salman e Abu-Ghazleh (2018), os resultados deste trabalho reforçam que a subestimação do primeiro pico de $g_{OO}(r)$ decorre, fundamentalmente, da incapacidade do potencial efetivo em reproduzir o perfil da energia de interação entre oxigênios.

Gráfico 5 - Função $g(r)$ entre átomos de oxigênio do 1-propanol



Fonte: elaborado pelo próprio autor

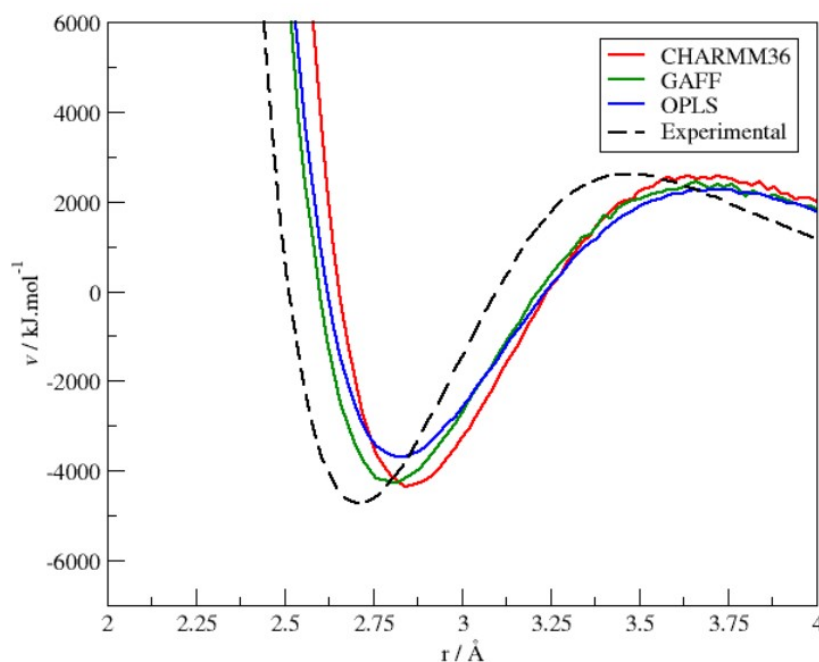
Nota: Perfil obtido similar para todos álcoois. Dados experimentais foram retirados de Tomšič *et al.* (2007).

Embora as simulações subestimem a intensidade primeiro pico de $g_{OO}(r)$, o número de coordenação N_c , descrito pela Equação 16, permanece bem reproduzido pelo modelo do CHARMM36 e GAFF, diferindo menos de 2 % dos valores experimentais, conforme representado na Tabela 5. Isso ocorre porque N_c não depende só da intensidade da função, mas também de sua largura e da densidade do sistema, de modo que variações nessas variáveis podem se

contrabalançar, preservando o valor integrado do número de coordenação. Em contrapartida, O OPLS-AA apresenta um maior desvio, que se acentua com o alongamento da cadeia.

Em conjunto, esses resultados indicam que a forma detalhada de $g(r)$ e o valor integrado de N_c capturam aspectos complementares da estrutura do líquido. Alcançar o valor correto de N_c não garante uma descrição fiel da primeira camada de solvatação, pois pode-se reproduzir o número de vizinhos mesmo que a posição ou a altura do pico de $g(r)$ divirja dos dados experimentais. Ajustes voltados apenas ao número de coordenação podem, portanto, ocultar falhas nos parâmetros do campo de força. De forma análoga, modificar somente a intensidade de $g(r)$ sem levar em conta o eventual aumento de densidade decorrente de interações O–O mais fortes pode acabar inflando N_c . Esses cenários evidenciam a natureza multidimensional—e nada trivial—do ajuste consistente dos parâmetros de potenciais efetivos.

Gráfico 6 - Potencial entre os pares de átomos de oxigênio do 1-propanol



Fonte: elaborado pelo próprio autor

Nota: Perfil obtido similar para todos álcoois. Dados experimentais foram retirados de Tomšič *et al.* (2007).

Tabela 5 - Número de coordenação para os diferentes álcoois

	$N_{c,exp}$	$N_{c,CHARMM36}$	$N_{c,GAFF}$	$N_{c,OPLS}$
Etanol	1,98	2,01	2,02	1,99
1-Propanol	1,95	1,97	1,98	1,82
1-Butanol	1,98	1,99	1,96	1,85
1-Pentanol	1,98	2,01	1,95	1,84
1-Hexanol	1,97	1,98	1,95	1,78

Fonte: elaborado pelo próprio autor

Nota: Para o cálculo de N_c , a função $g(r)$ e a densidade respectiva de cada sistema foi utilizada. Para o cálculo do N_c experimental, foram utilizados os dados de Haynes (2014) e Tomšič *et al.* (2007)

5.1.2 Divergências na constante dielétrica

Os valores de ϵ obtidos (Apêndice B, Gráfico B3) ficaram sistematicamente abaixo dos experimentais, e o desvio relativo cresce à medida que a cadeia carbônica do álcool se alonga. Essa tendência reflete uma limitação fundamental dos campos de força aditivos empregados: as cargas atômicas são fixas e, portanto, incapazes de responder a mudanças no ambiente químico, reduzindo as flutuações do momento de dipolo e, por consequência, a constante dielétrica calculada pela Equação 12 (Jorge, 2024; Leontyev; Stuchebrukhov, 2009). Essa limitação não significa que as interações de polarização sejam totalmente negligenciadas; elas são incorporadas de forma indireta nos parâmetros do potencial efetivo, juntamente com outros efeitos de muitos corpos, conforme discutido na Seção 2.1.1. A literatura relata a mesma subestimação para diversas famílias de solventes orgânicos, particularmente nos campos de força GAFF (Beauchamp *et al.*, 2015; Coleman *et al.*, 2012) e OPLS (Coleman *et al.*, 2012), reforçando que o viés observado aqui é intrínseco a modelos de carga fixa.

Para lidar com este desvio, o método mais adotado na literatura é aplicar um escalonamento das cargas do sistema, ou inserir o mesmo em um meio com constante dielétrica continua (Duboué-Dijon *et al.*, 2020; Leontyev; Stuchebrukhov, 2010). Cardona, Jorge e Lue (2020) demonstraram que multiplicar todas as cargas de um líquido por um único fator para que este melhor reproduza os dipolos experimentais permite recuperar, com alta precisão, a constante dielétrica de líquidos puros e misturas numa ampla faixa de temperaturas. O Trabalho de Duboué-Dijon *et al.* (2020) revelou que, quando esse escalonamento é aplicado apenas aos grupos fortemente carregados, obtém-se correção semelhante: o sobre-acoplamento iônico é reduzido e a descrição de solventes melhora para os modelos GAFF e OPLS.

5.1.3 Divergências no coeficiente de autodifusão

Entre os três campos de força avaliados, o CHARMM36 apresentou, em média, a melhor descrição do coeficiente de autodifusão (Apêndice B, Gráfico B2). Apesar disso, o desvio desse parâmetro ainda contribuiu de modo significativo para o DRM global. GAFF e OPLS-AA exibiram desempenho inferior, sobretudo para os álcoois de cadeia curta, faixa em que o valor experimental de D é mais elevado. Uma tendência análoga foi relatada por Anker *et al.* (2024) ao comparar múltiplas variantes dos campos de força GAFF e OPLS para soluções aquosas de ureia. No trabalho citado, a maior parte dos modelos superestimou a mobilidade da ureia, efeito mais pronunciado em concentrações diluídas, exatamente onde D atinge maior magnitude.

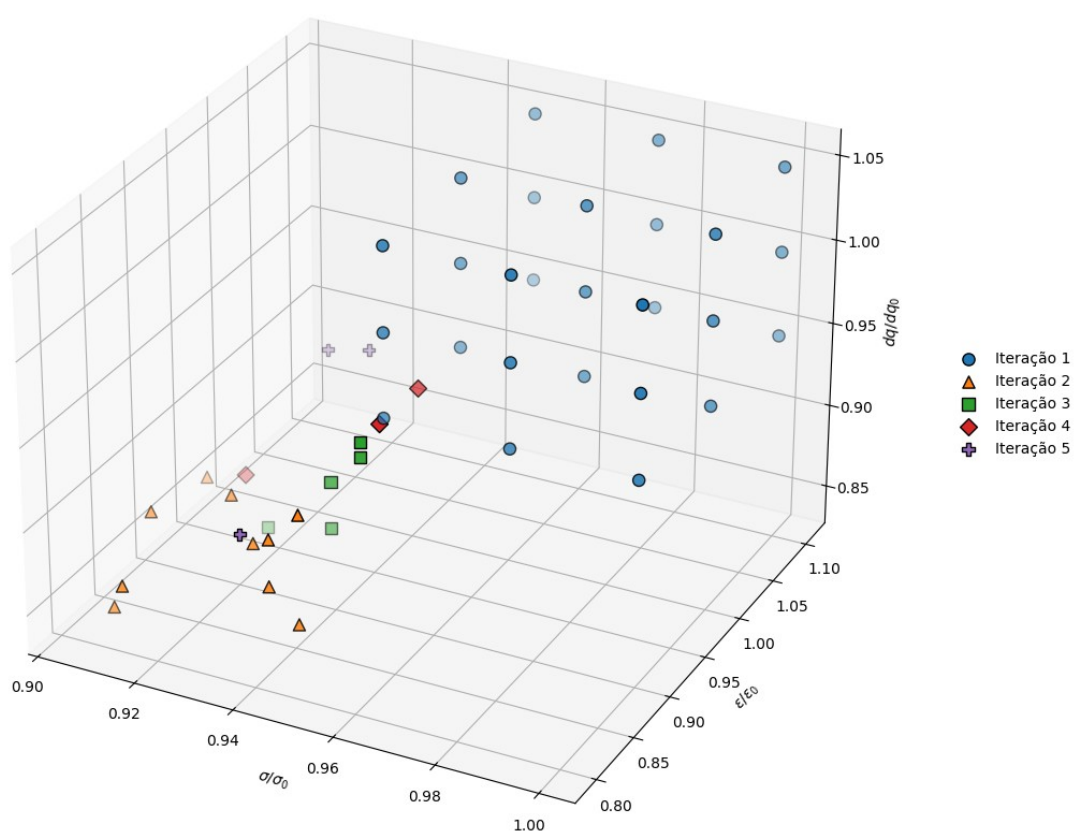
No contexto de líquidos orgânicos, Zangi (2018) identificou desvios sistemáticos em D para uma série de álcoois primários, secundários e dióis, enquanto Zêzere *et al.* (2023) quantificaram erros tanto na autodifusão do etanol (D_{11}) quanto na difusão de diversos solutos nesse solvente (D_{12}). Ambos os estudos recorreram a correções NBFIX para melhorar a descrição do coeficiente de autodifusão e de outras propriedades experimentais, porém por rotas distintas. Zangi (2018) substituiu as cargas do grupo hidroxila por valores alternativos disponíveis na literatura, e Zêzere *et al.* (2023) reduziram em $\approx 2\%$ o raio de Lennard-Jones (σ) do oxigênio do etanol. Neste último, também foi mostrado que ajustes simultâneos em no parâmetro de energia de Lennard-Jones (ϵ) produzem respostas diferentes no valor de D , a depender do raio σ adotado. Esses resultados reforçam o ponto central deste trabalho: a elevada dimensionalidade do espaço de parâmetros, aliada à forte correlação entre eles, torna a busca por combinações que minimizem simultaneamente as discrepâncias experimentais um problema nada trivial, justificando o emprego de abordagens sistemáticas, como a otimização guiada por aprendizado de máquina.

5.2 DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA

Para validar a estratégia de reparametrização, adotou-se o CHARMM36, pois ele apresentou os menores valores de R^2 nas comparações com os dados experimentais discutidas na Seção 5.1. Seguindo o protocolo descrito na Seção 4.3, o treinamento começou com um conjunto inicial de 54 simulações (Apêndice A, Tabela A2) que serviram de base para a primeira geração de redes neurais. A cada ciclo iterativo — denotado por $i_1, i_2, i_3, \dots$ — um subconjunto dos parâmetros mais promissores para cada álcool era selecionado, sendo utilizado em novas simulações de dinâmica molecular e então incorporado ao banco de dados de treinamento. O conjunto de treinamento correspondente a cada etapa é indicado como $C_1, C_2, C_3, \dots$; por definição, $C_1 = i_1$, $C_2 = i_1 + i_2$, $C_3 =$

$i_1 + i_2 + i_3$, e assim por diante, pois os resultados de cada iteração somam-se cumulativamente aos já existentes. Os parâmetros utilizados em todas as simulações — tanto na etapa inicial quanto nas iterações subsequentes — encontram-se detalhados nas Tabelas A2 a A6 do Apêndice A. Por fim, o Gráfico 7 oferece uma visão global do espaço de parâmetros efetivamente amostrado ao longo de todo o processo de otimização, evidenciando como as novas amostras expandiram progressivamente as regiões de interesse.

Gráfico 7 - Espaço de parâmetros amostrado nas cinco iterações de reparametrização



Fonte: elaborado pelo próprio autor

Nota: cada ponto representa um conjunto de parâmetros nas simulações com o CHARMM36; cores e marcadores distinguem as iterações i1–i5 (vide legenda). O primeiro ciclo (i1) corresponde aos pontos utilizados no treinamento inicial; as iterações subsequentes restringem-se às regiões mais promissoras indicadas pela rede neural

A evolução do modelo de otimização evidenciou um padrão claro já a partir da segunda iteração; os pontos convergiram para uma região contendo valores menores de σ/σ_0 e dq/dq_0 , permanecendo nesse mínimo local até a última iteração, enquanto ϵ/ϵ_0 oscilou dentro de um intervalo maior. Como resultado, apenas a primeira iteração ocupa o quadrante de valores mais elevados de σ e dq no Gráfico 7. Esse comportamento decorre de dois efeitos principais. Primeiro, a redução de σ aproxima os átomos de oxigênio, deslocando o primeiro pico de $g_{OO}(r)$ para distâncias

menores que se aproximam mais dos valores experimentais. Esta aproximação fortalece a atração entre os oxigênios, aumentando a altura do pico de $g(r)$ e consequentemente a profundidade do poço de energia. Segundo, embora as correções de Lennard-Jones sejam aplicadas apenas ao par oxigênio-oxigênio — sobrescrevendo as regras de combinação exclusivamente para esses átomos (Seção 2.2) — a alteração é suficiente para modificar o balanço de forças em todo o líquido. O modelo, portanto, privilegia combinações que reduzem o momento de dipolo da hidroxila (menor dq), atenuando as interações eletrostáticas intensificadas pelo ajuste de σ e ϵ .

Para testar essa hipótese, foram realizadas simulações adicionais nos cinco álcoois reduzindo σ em 5 % e aumentando ϵ em 5 %, sem alterar as cargas parciais. O ajuste elevou o coeficiente de determinação entre $g_{\text{sim}}(r)$ e $g_{\text{exp}}(r)$ acima de 0,90 para todos os compostos, mas piorou sistematicamente a representação da densidade, a entalpia de vaporização e, sobretudo, o coeficiente de autodifusão. Enquanto ρ e ΔH_{vap} passaram a superestimar os valores experimentais, D sofreu queda acentuada. Conclui-se, portanto, que o fortalecimento excessivo da interação O-O supervaloriza a coesão dos álcoois, dando uma justificativa física para a preferencia dos modelos de aprendizado de máquina utilizados levarem os parâmetros para uma região de menor interação eletrostática.

Concluídos os cinco ciclos de otimização, foram obtidos cinco conjuntos distintos de parâmetros, resumidos na Tabela 6. Para chegar nesses resultados, dois tipos de modelo foram empregados: (i) Modelo **F** — considera apenas as propriedades envolvidas no cálculo do DRM como variáveis resposta e minimiza a função alvo da Equação 21. (ii) Modelo **G** — inclui, além dessas propriedades, a altura máxima da função de distribuição radial, $g(r_{\text{max}})$, como variável adicional; a otimização segue a função alvo da Equação 23. Cada modelo foi treinado utilizando um conjunto C_i , também especificado na Tabela 6.

Tabela 6 - Parâmetros finais obtidos para cada um dos álcoois

Álcool	Modelo	σ/σ_0	ϵ/ϵ_0	dq/dq_0
Etanol	F (C_1)	0,946	0,895	0,935
1-Propanol	F (C_1)	0,946	0,895	0,926
1-Butanol	G (C_3)	0,905	1,107	0,873
1-Pentanol	G (C_3)	0,914	1,107	0,879
1-Hexanol	G (C_2)	0,914	0,945	0,870

Fonte: elaborado pelo próprio autor

Nota: como comentado previamente, todos os conjuntos finais levaram a estados com menores valores de σ e dq .

Os valores previstos por cada modelo (Apêndice C, Tabela C1) foram comparados com as propriedades obtidas nas simulações de dinâmica molecular (Apêndice C, Tabela C2); os desvios relativos encontram-se na Tabela 21. Em praticamente todas as propriedades avaliadas — densidade, entalpia de vaporização (ΔH_{vap}), coeficiente de autodifusão (D) e R^2 da função de distribuição radial — os desvios médios permaneceram abaixo de 5 %, indicando que a abordagem de treinamento adotada foi capaz em explorar o espaço paramétrico com alta eficiência. Resultados semelhantes foram obtidos por Befort *et al.* (2021), que utilizaram aprendizado de máquina para otimizar campos de força aplicados a sistemas contendo hidrofluorocarbonetos, com o objetivo de reproduzir propriedades termodinâmicas em equilíbrio líquido-vapor, como densidades de fases, pressão de vapor e ΔH_{vap} . Os modelos selecionados por esse estudo apresentaram erros percentuais médios inferiores a 2,5 %, demonstrando a elevada precisão da metodologia orientada por aprendizado utilizada.

Tabela 7 - Desvio relativo entre as predições dos modelos de aprendizado de máquina e suas respectivas simulações

Álcool	$\Delta\rho$	ΔD	$\Delta\epsilon$	$\Delta(\Delta H)$	ΔR^2	Desvio Médio
Etanol	0,04%	3,22%	11,57%	1,19%	2,49%	3,70%
1-Propanol	0,02%	4,80%	12,03%	0,29%	3,23%	4,07%
1-Butanol	0,08%	10,82%	7,26%	0,86%	1,60%	4,12%
1-Pentanol	0,02%	0,63%	20,25%	0,27%	1,22%	4,48%
1-Hexanol	0,06%	4,37%	15,58%	0,34%	0,00%	4,07%

Fonte: elaborado pelo próprio autor

Nota: propriedades com desvio relativo maior que 5% estão destacadas em vermelho.

A propriedade que foi pior descrita pelos modelos foi a constante dielétrica, principalmente para os álcoois 1-pentanol e 1-hexanol. Essa discrepância pode ser associada da baixa magnitude de ϵ observada nestes sistemas. Por exemplo, segundo o modelo **G**(C_3), a constante dielétrica prevista para o 1-pentanol foi 6,433, enquanto o valor obtido a partir da simulação foi de 8,066—diferença absoluta de 1,633. Contudo, por se tratar de um valor pequeno, mesmo essa variação modesta gera um desvio relativo elevado, de 20,25 %.

5.3 COMPARAÇÃO ENTRE ÁLCOOIS COM OS PARÂMETROS ORIGINAIS E COM AS CORREÇÕES NBFIX

A Tabela 8 resume o DRM obtido para cada álcool quando se utilizam os parâmetros originais do CHARMM36 e, em seguida, as correções NBFIX determinadas pelo protocolo de aprendizado de máquina. Já o Gráfico 8 mostra a contribuição das cinco propriedades que o compõem. Do ponto de vista geral, a reparametrização foi extramente eficaz, o DRM caiu $45 \pm 9 \%$, em média, e a razão $\text{DRM}_{\text{Original}}/\text{DRM}_{\text{NBFIX}}$ põe em perspectiva o quão substancial é a diferença entre os desvios ao se utilizar os parâmetros originais e os desvios ao se utilizar os parâmetros modificados. A maior contribuição para a redução do DRM provém da correção da primeira camada de solvatação. Para todos os sistemas, exceto etanol, o termo $1-R^2$ era originalmente superior a 80 %, refletindo a sub-representação do primeiro pico de $g_{\text{oo}}(r)$, porem ao se utilizar os novos parâmetros, esse desvio cai abaixo de 3,5 %, indicando uma melhor representação do primeira camada de solvatação e, conseqüentemente, da energia de interação entre os pares de oxigênio (Gráfico 9a e b).

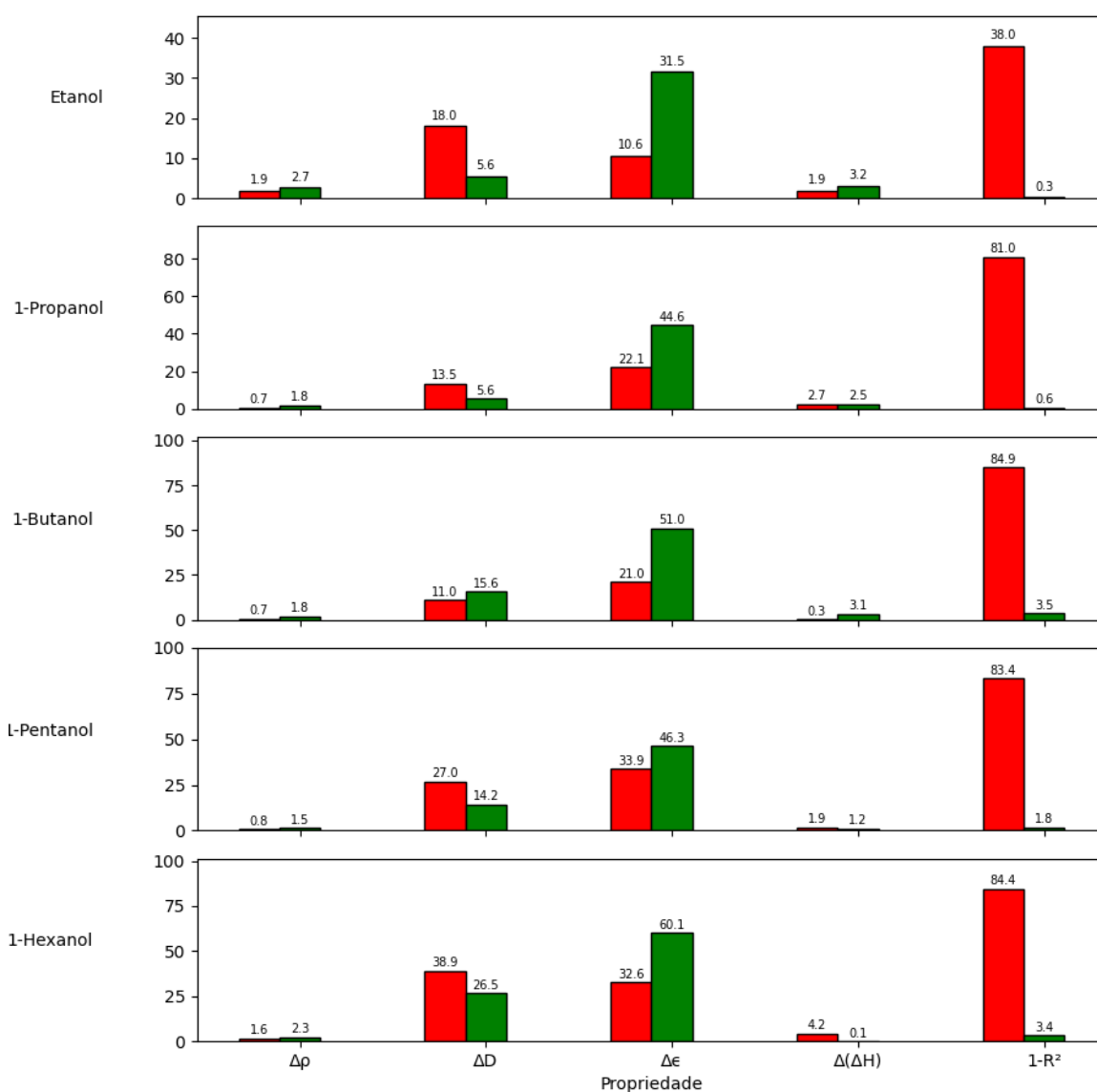
A densidade e a entalpia de vaporização—propriedades já reproduzidas com alta precisão pelas simulações que utilizaram os parâmetros originais—permaneceram dentro da margem experimental após o ajuste, evidenciando que o reescalonamento do dipolo evita a superestimação das interações álcool álcool, um contraste direto com o teste da Seção 5.2, em que se manteve d_q inalterado. O coeficiente de autodifusão melhorou para todos os álcoois, à exceção do 1-butanol, cujo desvio cresceu moderadamente de 11,0 % para 15,6 %. Nas demais moléculas o ganho foi expressivo, principalmente no etanol, onde o erro caiu de 18 % para 5 %. Considerando que o etanol possui o maior valor absoluto de difusão dessa série de álcoois ($1,01 \times 10^{-9} \text{m}^2 \text{s}^{-1}$; Tabela 2), esta variação no desvio corresponde a uma correção absoluta de $\approx 1,3 \times 10^{-9} \text{m}^2 \text{s}^{-1}$ —diferença significativa para processos de transporte (Nguyen; Bouchaudy; Salmon, 2022).

Tabela 8 - Desvio Relativo Médio (DRM) das propriedades simuladas para cada álcool com os diferentes conjuntos de parâmetros

Álcool	DRM _{Original}	DRM _{NBFX}	DRM _{Original} / DRM _{NBFX}
Etanol	14,10%	8,65%	1,63
1-Propanol	23,98%	11,02%	2,18
1-Butanol	23,57%	15,00%	1,57
1-Pentanol	29,37%	13,01%	2,26
1-Hexanol	32,35%	18,49%	1,75

Fonte: elaborado pelo próprio autor

Gráfico 8 - Comparação do desvio relativo por propriedade

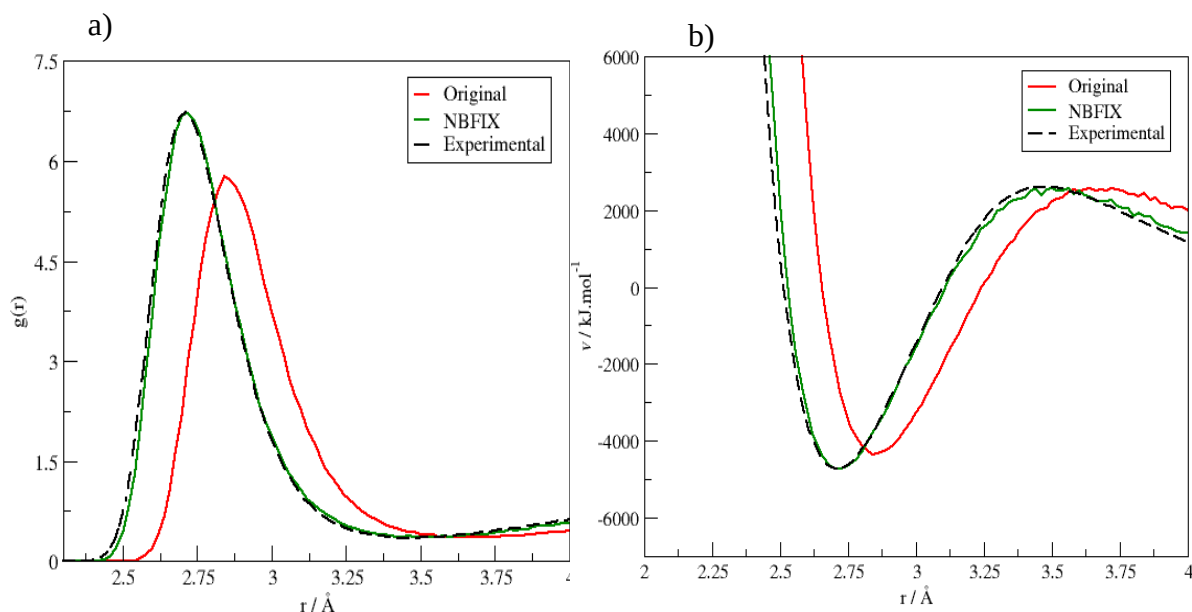


Fonte: elaborado pelo próprio autor

Nota: as barras vermelha e verde correspondem aos valores dos desvios obtidos com o campos de força CHARMM36 utilizando os parâmetros originais e correções NBFX, respectivamente.

A constante dielétrica foi a única propriedade cujo desvio aumentou sistematicamente. Tal resultado era esperado: os modelos de carga fixa já subestimam ϵ (Seção 5.1.2) e, ao reduzir o dipolo da hidroxila para garantir que a maior energia de interação entre os pares de oxigênio (Gráfico 8b) não afete as demais propriedades, as flutuações do momento de dipolo sobre as quais ϵ depende diminuem ainda mais (Equação 4). Porém, mesmo que não seja possível otimizar simultaneamente todas as cinco propriedades considerando a região dos parâmetros amostrada, manter ϵ na função alvo (Equações 7 e 9) é fundamental. Esse termo impede que o algoritmo converja para soluções nas quais $dq_0 \ll dq$ apenas para acomodar melhor ρ , ΔH_{vap} ou D , preservando assim um equilíbrio físico coerente entre estrutura, termodinâmica e polaridade do sistema.

Gráfico 9 - Influência da correção NBFIX na estrutura do 1-propanol e potencial $v(r)$



Fonte: elaborado pelo próprio autor

Nota: a) apresenta a função distribuição radial $g_{\text{oo}}(r)$; b) mostra o potencial calculado através da Equação 17. As curvas experimentais foram extraídas de Tošić *et al.* (2007). Perfis equivalentes foram obtidos para todos os álcoois analisados,

O impacto das correções NBFIX também foi examinado sobre o número de coordenação N_c (Tabela 9). Embora todos os sistemas reparametrizados exibam um primeiro pico de $g_{\text{sim}}(r)$ mais intenso (Gráfico 9a) e a densidade permaneça praticamente inalterada, N_c diminui ligeiramente após o ajuste. Isso ocorre porque N_c depende de uma integral de $g(r)$ entre $r = 0$ e o primeiro mínimo r_{min} (Equação 16). Ao contrair a primeira camada de solvatação, r_{min} se desloca para distâncias menores —no caso do 1-propanol, de 3,64 Å (parâmetros originais) para 3,45 Å (NBFIX)—reduzindo o limite superior de integração e, por consequência, o valor integrado de N_c . Mesmo com essa leve

queda, o modelo ajustado reproduz adequadamente o número de vizinhos, com as diferenças entre $N_{c,exp}$ e $N_{c,NBFX}$ permanecendo à segunda casa decimal.

Tabela 9 - Comparação do número de coordenação para os diferentes álcoois com os diferentes parâmetros

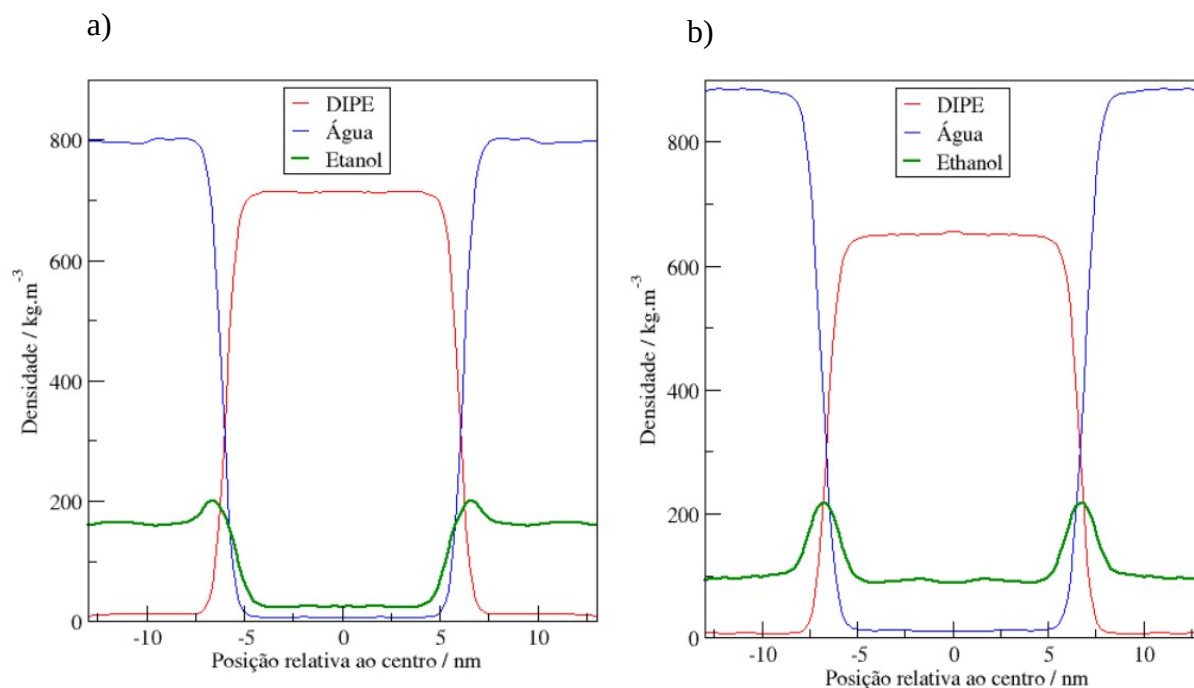
Álcool	$N_{c,exp}$	$N_{c,Original}$	$N_{c,NBFX}$
Etanol	1,98	2,01	1,95
1-Propanol	1,95	1,97	1,97
1-Butanol	1,98	1,99	1,92
1-Pentanol	1,98	2,01	1,95
1-Hexanol	1,97	1,98	1,93

Fonte: elaborado pelo próprio autor

5.4 INFLUÊNCIA DA REPARAMETRIZAÇÃO NO SISTEMA TERNÁRIO DIPE/ÁGUA/ETANOL

Foram realizadas duas simulações do sistema ternário, que diferem exclusivamente pelos parâmetros atribuídos ao etanol. A primeira simulação emprega o conjunto original do CHARMM36, enquanto a segunda incorpora as correções NBFX determinadas na Seção 5.2. Para aferir o impacto dessas modificações, adotou-se como propriedade-alvo o coeficiente de partição do etanol k , grandeza de interesse em diversos sistemas líquido-líquido (Kashurin *et al.*, 2024; Poole, 2024). Os perfis de densidade médios dos três componentes, calculados no intervalo de 100–150 ns em ambas simulações, encontram-se no Gráfico 9. A evolução temporal da densidade do etanol nas regiões de *bulk* orgânico (Região 1) e aquoso (Região 2), definidas na Seção 4.5 esta descrita no Apêndice D, Gráfico D1. Ao comparar os perfis de densidade obtidos, observa-se que a versão reparametrizada desloca parte significativa do etanol da fase aquosa para a fase orgânica. Kashurin *et al.* (2024) examinaram o mesmo sistema com o campo de força CHARMM36; diferiram apenas nas dimensões da caixa de simulação e no *ensemble* adotado. O perfil exibido no Gráfico 10a reproduz fielmente o relatado pelos autores, reforçando a consistência dos dados obtidos.

Gráfico 10 -Perfil de densidade para os componentes do sistema ternário



Fonte: elaborado pelo próprio autor.

Nota: gráfico a) obtido utilizando os parâmetros originais do CHARM36; gráfico b) obtido com etanol reparametrizado.

Por consequência do deslocamento do etanol, a fração molar na região *bulk* aquosa caiu de 0,193 para 0,115, enquanto no *bulk* orgânico subiu de 0,030 para 0,113, revelando maior afinidade do etanol reparametrizado pela fase etérea. Como o coeficiente de partição experimental k_{exp} depende da fração de etanol dissolvida em DIPE (Arce *et al.*, 2002), cada simulação requer um valor de referência específico. Para compará-las de forma coerente, os pontos experimentais foram ajustados a um polinômio de 4.º grau (Apêndice D, Gráfico D2; $R^2=0,995$). Desse ajuste extraíram-se os pares de fração molar de etanol na fase orgânica, X_{etOH} , e o coeficiente de partição k_{exp} correspondentes às composições simuladas. Usando as densidades instantâneas em cada fase (Gráfico D1) e a Equação 24 obteve-se $k(t)$ (Apêndice D, Gráfico D3); a média no intervalo 100–150 ns definiu k_{sim} . Tabela 10 sintetiza os resultados obtidos, comparando k_{sim} aos valores de referência extraídos do ajuste polinomial dos dados experimentais. O desvio relativo, indicado na coluna Δk , corresponde à diferença percentual entre o coeficiente de partição obtido por simulação e o experimental ajustado. Observa-se que, com os parâmetros originais do CHARMM36, Δk atinge quase 79 %, revelando grande subestimação da partição do etanol em DIPE. Já com a inclusão das correções NBFIX, esse desvio cai para aproximadamente 23 %, configurando uma melhor representatividade do valor experimental.

Tabela 10 - Comparação entre os coeficientes de partição do etanol experimental e computacional

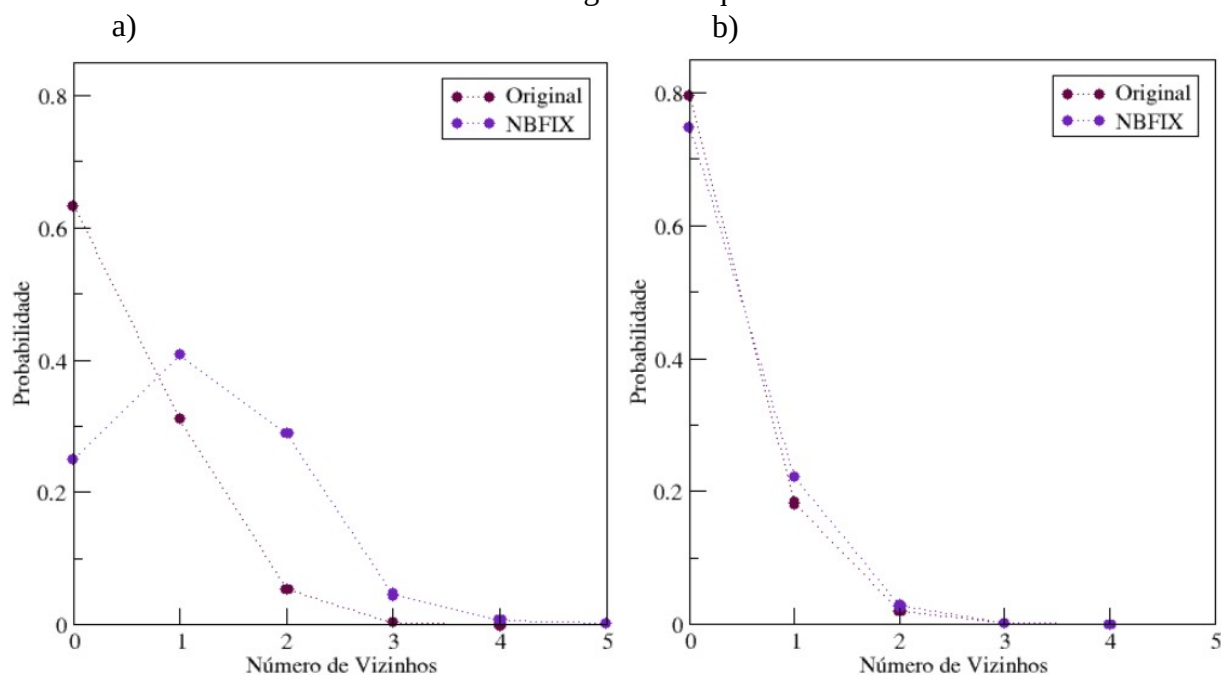
X_{EtOH}	k_{exp}	k_{sim}	$\Delta k\%$
0,030	0,685	0,146	78,69%
0,113	1,182	0,909	23,10%

Fonte: elaborado pelo próprio autor.

Nota: Os valores de k_{exp} foram obtidos pelo ajuste polinomial aplicado aos dados experimentais de Arce *et al.* (2002).

Para verificar o efeito da reparametrização em escala atômica, a distribuição do número de etanóis vizinhos em cada molécula de etanol $P(n)$ foi avaliada em ambas regiões bulk (Gráfico 10). Duas moléculas de etanol foram consideradas vizinhas quando a distância oxigênio-oxigênio fosse menor ou igual ao primeiro mínimo de $g_{\text{sim}}(r)$. Com o conjunto original de parâmetros do CHARMM36, o etanol apresenta-se predominantemente monomérico em ambas as regiões *bulk*: na fase orgânica cerca de 63 % das moléculas não possuem vizinhos de mesma espécie, enquanto na fase aquosa essa fração sobe para quase 80 %. Pares ($P(1)$) e trios ($P(2)$) são, portanto, eventos pouco prováveis e a distribuição das moléculas de etanol permanece esparsa. A introdução das correções NBFIX altera substancialmente esse perfil, pois na fase orgânica $P(0)$ cai para ~0,25, ao passo que $P(1)$ e $P(2)$ elevam-se para ~0,40 e ~0,28, respectivamente, indicando a formação de micro-agregados estáveis impulsionados pelo reforço da interação álcool-álcool. Mesmo na água, onde a densidade do álcool diminui de forma pronunciada devido ao aumento do coeficiente de partição, a probabilidade de se encontrar pares de etanol cresce de 18 % para 22 %, acompanhada de leve queda em $P(0)$. Assim, pode-se afirmar que a reparametrização NBFIX intensifica a afinidade mútua entre moléculas de etanol em ambas as fases, contribuindo tanto para a migração do etanol, quanto para a reconfiguração da própria estrutura do mesmo em solução.

Gráfico 11 - Distribuição de probabilidade do número de moléculas de etanol vizinhas nas fases orgânica e aquosa



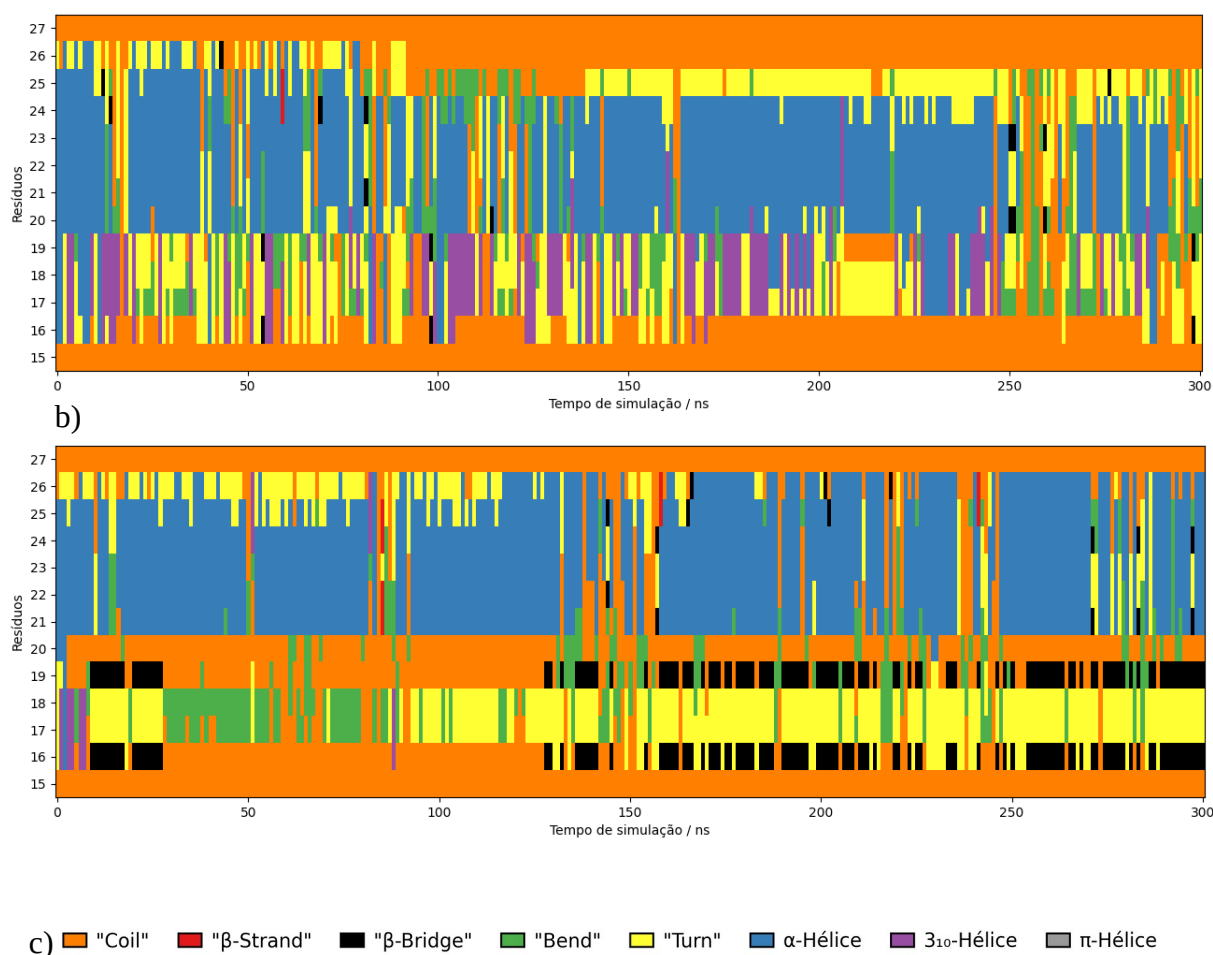
Fonte: elaborado pelo próprio autor

Nota: probabilidades foram obtidas a partir da janela de 100–150 ns. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados.

5.5 INFLUÊNCIA DA REPARAMETRIZAÇÃO NA SOLUÇÃO ETANÓLICA DA β -AMILOIDE

As análises de DSSP ao longo de 300 ns de trajetória de ambas as simulações, representadas nos Gráficos 17a e 17b, indicam que a reparametrização NBFIX para o etanol altera de modo mensurável o perfil de estrutura secundária da proteína. A distribuição dos resíduos que permanecem em conformação α -hélice variou e os tempos médios de retenção listados na Tabela 11 variaram em consequência. Com o conjunto original do CHARMM36, o peptídeo conserva uma α -hélice entre Phe²⁰ e Val²⁴ por mais de 190 ns, enquanto os demais resíduos alternam-se em outras estruturas, principalmente entre “turns” e “coils”. Este resultado é coerente com o estudo de Lemkul, Huang e MacKerell; (2015) que registrou mais de 50 % de helicidade para uma sequência similar de aminoácidos da β -amiloide em etanol usando o mesmo campo de força. No mesmo trabalho, a substituição pelo campo força polarizável baseado no oscilador de Drude alterou esse valor para 10,2 %, implicando que à uma dependência entre a estrutura secundaria da proteína e o modelo do potencial empírico utilizado.

Gráfico 12 - Evolução temporal da estrutura secundária da β -amiloide (15 – 27) em solução etanólica



Fonte: elaborado pelo próprio autor.

Nota: a) Simulação com os parâmetros originais do CHARMM36. b) Simulação com o etanol reparametrizado. c) Legenda das estruturas secundárias atribuídas pelo algoritmo DSSP.

Tabela 11 – Tempo na estrutura de α -hélice por resíduo, para os diferentes conjuntos de parâmetros

	Gln ¹⁵	Lys ¹⁶	Leu ¹⁷	Val ¹⁸	Phe ¹⁹	Phe ²⁰	Ala ²¹	Glu ²²	Asp ²³	Val ²⁴	Gly ²⁵	Ser ²⁶	Asn ²⁷
t ₀ / ns	0	15,6	25,3	32,9	39,7	204,2	223	226,5	227,4	191,0	53,0	27,3	0
t _N / ns	0	1,1	1,1	1,1	1,1	4,4	228,6	242,2	249,8	249,8	213,4	142,8	0

Fonte: elaborado pelo próprio autor.

Nota: t₀ representa os valores obtidos com os parâmetros originais, e t_N representa os valores obtidos com os ajustes NBFIX.

Após a introdução dos termos NBFIX, o núcleo helicoidal desloca-se para a região entre os resíduos Ala²¹–Ser²⁶ e formam-se pontes- β persistentes em Lys¹⁶ e Phe¹⁹, sobretudo depois de 150 ns. Transições semelhantes já foram descritas para outros fragmentos da β -amiloide em misturas hidro-orgânicas ou em solventes de menor constante dielétrica, nas quais a predominância de α -hélice decorre do maior favorecimento das ligações intrapeptídicas quando ϵ é baixo (Soto *et al.*, 1995; Wei; Shea, 2006). A reparametrização, além da modular diretamente as interações etanol–

etanol, reduziu a constante dielétrica efetiva do meio de $\epsilon = 21,7$ para $\epsilon = 16,7$ (Apêndice C, Gráfico 3), como previamente discutido na Seção 5.3. Uma outra suposição razoável seria que o aumento da interação álcool-álcool diminui a frequência de contatos álcool-peptídeo e, indiretamente, favorece a formação de ligações intrapeptídicas. Desse modo, o rearranjo conformacional não pode ser atribuído a um único fenômeno. Junto com a literatura, estes resultados mostram a sensibilidade de biomoléculas ao ambiente, e reforçam a necessidade de validar, de forma integrada, as interações entre solvente e soluto sempre que se refinarem potenciais efetivos (Lemkul; Huang; MacKerell, 2015; Wei; Shea, 2006).

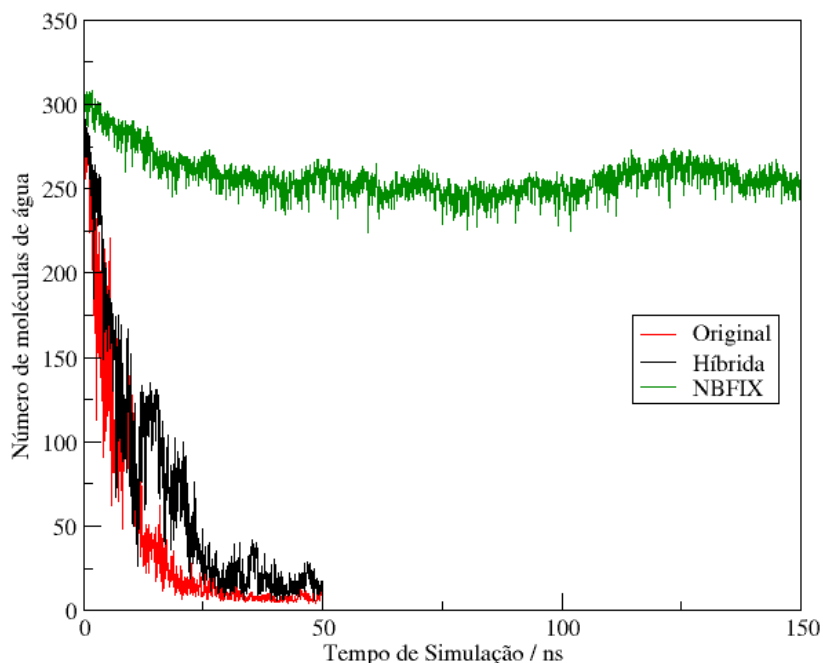
5.6 INFLUÊNCIA DA REPARAMETRIZAÇÃO NO SISTEMA MICELAR REVERSO

Foram conduzidas três simulações independentes para investigar o impacto da reparametrização nos agregados micelares reversos de DTAB: (i) parâmetros originais do CHARMM36; (ii) parâmetros NBFIX aplicados durante todo o protocolo; e (iii) abordagem híbrida, equilibrada com NBFIX, mas com fase de produção revertida aos parâmetros originais (Seção 4.7). A estabilidade do sistema foi avaliada pela evolução do maior agregado de moléculas de água, que indicam o tamanho do núcleo micelar, em função do tempo de simulação (Gráfico 13). Duas moléculas de água foram consideradas vizinhas quando a distância O-O era menor ou igual ao primeiro mínimo de $g_{\text{sim}}(r)$ obtido através de uma simulação de água pura. Estruturas das micelas obtidas ao final de cada simulação encontram-se na Figura 6.

Os resultados obtidos destacam o papel essencial da reparametrização na estabilidade do agregado micelar reverso. Nos sistemas que empregaram os parâmetros originais ou o protocolo híbrido, o número de moléculas de água no núcleo micelar despenca nos primeiros nanossegundos, sinalizando desagregação do núcleo micelar. Já com o conjunto NBFIX o núcleo permanece coeso durante toda a trajetória, mantendo em média 257 moléculas de água. A análise visual das configurações finais confirma essa tendência. Após 50 ns, as micelas simuladas com o conjunto original e com o protocolo híbrido já não existiam mais, apenas se observando surfactantes e moléculas de água espalhadas na fase orgânica formando dímeros ou trímeros de DTAB (Figura 6). Essa disposição do sistema explica a queda abrupta observada no Gráfico 13, pois o maior agregado aquoso deixa de representar o núcleo da micela, e passa a corresponder a agrupamentos menores de água dispersos no solvente. Em contraste, a micela simulada com os parâmetros NBFIX permanece coesa durante os 150 ns de simulação, liberando apenas uma fração de água de seu interior —

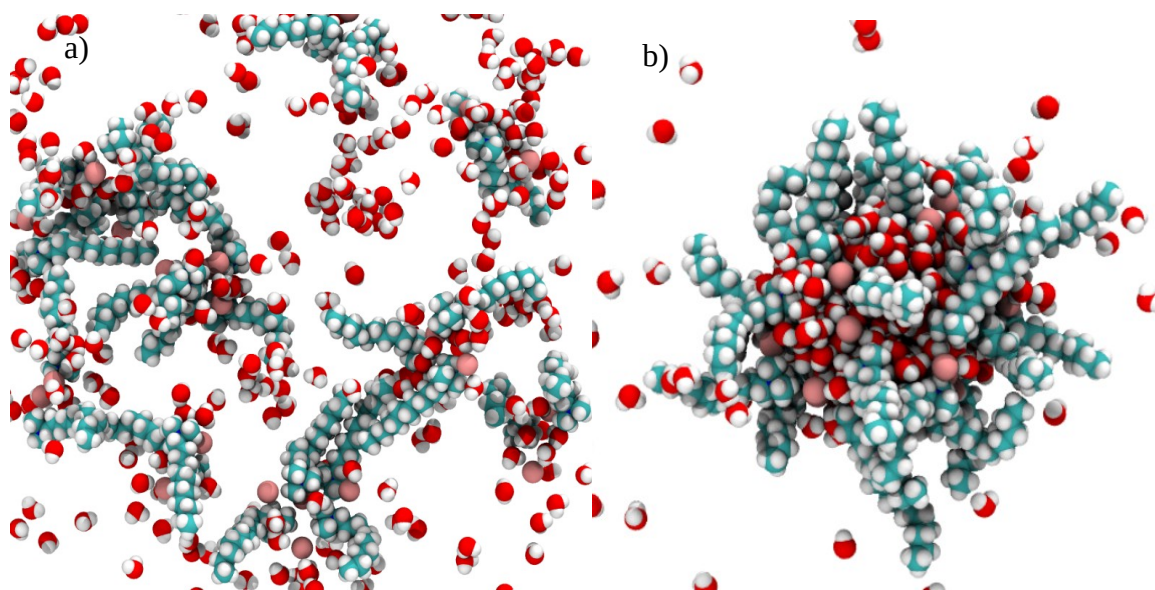
comportamento coerente com a troca e liberação de material intramicelar descrita em estudos experimentais e em simulações de dinâmica molecular (Eskici; Axelsen, 2016; Gębicki, 2004).

Gráfico 13 - Evolução do número de moléculas de água no núcleo micelar de DTAB



Fonte: elaborado pelo próprio autor.

Figura 6 - Estruturas finais das micelas reversas de DTAB simuladas



Fonte: elaborado pelo próprio autor

Nota: (a) Configuração representativa após 50 ns para os sistemas com parâmetros originais ou híbridos; (b) configuração após 150 ns para o sistema reparametrizado com NBFIX. Em ambas as imagens, oxigênios estão em vermelho, hidrogênios em branco, carbonos em ciano e brometos em rosa.

6 CONCLUSÕES E PERSPECTIVAS

A reparametrização direcionada por aprendizado de máquina do par oxigênio-oxigênio dos álcoois alifáticos realizada neste trabalho demonstrou ser essencial para restabelecer, de forma simultânea, a coerência entre estrutura local, propriedades de transporte e grandezas termodinâmicas. O ajuste individual para cada molécula do trio de parâmetros (σ , ϵ , dq) elevou o coeficiente de determinação entre a função de distribuição radial $g_{oo}(r)$ simulada e a experimental para $R^2 > 0,95$, corrigindo a discrepância na primeira camada de solvatação sem comprometer propriedades já bem descritas, como densidade, entalpia de vaporização, número de coordenação e a constante dielétrica ϵ . Este feito reforça a necessidade de considerar a multidimensionalidade do espaço paramétrico ao se realizar correções NBFIX. O fato de os parâmetros originais, refinados neste trabalho, serem aplicados indiscriminadamente a uma gama de álcoois de cadeia mais longa, dióis e polióis; aliado às evidências de Silberstein e Makov (2022), que mostram a mesma subestimação da altura do primeiro máximo e o deslocamento de $g_{oo}(r)$ para distâncias maiores em metanol, indica que esses desvios se propagam ao longo de toda a família de álcoois. Esse panorama amplia o escopo e a relevância do presente estudo, pois sugere que a abordagem de reparametrização proposta é eficaz — e potencialmente necessária — para corrigir inconsistências estruturais em número ainda mais ampla de sistemas alcoólicos.

A robustez e a transferibilidade do conjunto refinado foram validadas nos três sistemas multicomponentes analisados. Em todos eles, o aumento da interação álcool-álcool promoveu melhorias sistemáticas nas propriedades analisadas, aproximando os resultados simulados dos valores experimentais e demonstrando a capacidade do modelo em gerar informações físicoquímicas confiáveis. O efeito foi particularmente decisivo para a micela reversa de DTAB em 1-hexanol, pois sem a correção NBFIX, nenhuma micela permaneceu estável ao longo de toda trajetória da simulação, enquanto o potencial reparametrizado preservou a integridade do agregado e a retenção do núcleo aquoso. Esses achados confirmam que os ajustes propostos são indispensáveis para descrever sistemas complexos contendo álcoois e consolidam os novos modelos, e o método utilizado para obtê-los, como ferramentas robusta para estudos de dinâmica molecular.

Como perspectiva, sugere-se aplicar o mesmo protocolo de aprendizado ativo ao GAFF, cuja filosofia de atribuição de cargas e parametrização dos termos do potencial de Lennard-Jones difere substancialmente do CHARMM36, bem como estender os testes a novos sistemas multicomponentes e a outros álcoois homólogos que compartilham o mesmo conjunto original de parâmetros. Essa expansão permitirá aferir a robustez do método em superfícies de potencial

alternativas e fornecer à literatura um conjunto aprimorado de parâmetros transferíveis para uma gama ainda mais ampla de álcoois.

REFERÊNCIAS

- ALLEN, Michael Patrick; TILDESLEY, Dominic J. **Computer simulation of liquids**. 2nd ed ed. Oxford: Oxford university press, 2017.
- ANKER, Samira *et al.* Assessment of GAFF and OPLS Force Fields for Urea: Crystal and Aqueous Solution Properties. **Crystal Growth & Design**, v. 24, n. 1, p. 143–158, 3 jan. 2024.
- ARCE, Alberto *et al.* Liquid–Liquid Equilibrium of Diisopropyl Ether + Ethanol + Water System at Different Temperatures. **Journal of Chemical & Engineering Data**, v. 47, n. 3, p. 529–532, 1 maio 2002.
- BEAUCHAMP, Kyle A. *et al.* Toward Automated Benchmarking of Atomistic Force Fields: Neat Liquid Densities and Static Dielectric Constants from the ThermoML Data Archive. **The Journal of Physical Chemistry B**, v. 119, n. 40, p. 12912–12920, 8 out. 2015.
- BEFORT, Bridgette J. *et al.* Machine Learning Directed Optimization of Classical Molecular Modeling Force Fields. **Journal of Chemical Information and Modeling**, v. 61, n. 9, p. 4400–4414, 27 set. 2021.
- BERENDSEN, Herman J. C. *et al.* Molecular dynamics with coupling to an external bath. **Journal of Chemical Physics**, v. 81, n. 8, p. 3684–3690, 15 out. 1984.
- BERNETTI, Mattia; BUSSI, Giovanni. Pressure control using stochastic cell rescaling. **The Journal of Chemical Physics**, v. 153, n. 11, p. 114107, 21 set. 2020.
- BEST, Robert B. *et al.* Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. **Journal of Chemical Theory and Computation**, v. 8, n. 9, p. 3257–3273, 11 set. 2012.
- BEZMAN, Richard D.; CASASSA, Edward F.; KAY, Robert L. The temperature dependence of the dielectric constants of alkanols. **Journal of Molecular Liquids**, v. 73–74, p. 397–402, nov. 1997.
- BUSSI, Giovanni; DONADIO, Davide; PARRINELLO, Michele. Canonical sampling through velocity rescaling. **The Journal of Chemical Physics**, v. 126, n. 1, p. 014101, 7 jan. 2007.
- CALEMAN, Carl *et al.* Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. **Journal of Chemical Theory and Computation**, v. 8, n. 1, p. 61–74, 10 jan. 2012.
- CARDONA, Javier; JORGE, Miguel; LUE, Leo. Simple corrections for the static dielectric constant of liquid mixtures from model force fields. **Physical Chemistry Chemical Physics**, v. 22, n. 38, p. 21741–21749, 2020.
- CARRÉ, Patrick *et al.* Solvent solutions: comparing extraction methods for edible oils and proteins in a changing regulatory landscape. Part 1: Physical-properties. **OCL**, v. 31, p. 31, 2024.

COLES, Murray *et al.* Solution Structure of Amyloid β -Peptide(1-40) in a Water-Micelle Environment. Is the Membrane-Spanning Domain Where We Think It Is?†. **Biochemistry**, v. 37, n. 31, p. 11064–11077, 16 jul. 1998.

DARDEN, Tom; YORK, Darrin; PEDERSEN, Lee. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089–10092, 15 jun. 1993.

DODDA, Leela S. *et al.* 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. **The Journal of Physical Chemistry B**, v. 121, n. 15, p. 3864–3870, 20 abr. 2017a.

DODDA, Leela S. *et al.* LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. **Nucleic Acids Research**, v. 45, n. W1, p. W331–W336, 3 jul. 2017b.

DUBOUÉ-DIJON, E. *et al.* A practical guide to biologically relevant molecular simulations with charge scaling for electronic polarization. **The Journal of Chemical Physics**, v. 153, n. 5, p. 050901, 7 ago. 2020.

ESKICI, Gözde; AXELSEN, Paul H. The Size of AOT Reverse Micelles. **The Journal of Physical Chemistry B**, v. 120, n. 44, p. 11337–11347, 10 nov. 2016.

GALINDO, Fernanda Alvarado *et al.* Incorporation of short-chain alcohols into fluid bilayers and its effect on membrane dynamic properties as seen by neutron scattering. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v. 702, p. 135014, dez. 2024.

GAO, Pan *et al.* Comparison of solvents for extraction of walnut oils: Lipid yield, lipid compositions, minor-component content, and antioxidant capacity. **LWT**, v. 110, p. 346–352, ago. 2019.

GAZI, Rabiul; MAITY, Sankar; JANA, Madhurima. Conformational Features and Hydration Dynamics of Proteins in Cosolvents: A Perspective from Computational Approaches. **ACS Omega**, v. 8, n. 3, p. 2832–2843, 24 jan. 2023.

GEŹBICKI, Jerzy. Intermicellar material exchange in reverse micelles formed by ionic AOT and nonionic Igepal surfactants studied by means of pulse radiolysis. Influence of the temperature. **Open Chemistry**, v. 2, n. 2, p. 371–387, 1 jun. 2004.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1st. ed. Cambridge, MA: The MIT Press, 2016.

HANSEN, Jean-Pierre; MCDONALD, Ian R. **Theory of Simple Liquids: With Applications to Soft Matter**. 4th. ed. Amsterdam (NL): Academic Press / Elsevier, 2013.

HASLAM, Andrew J.; GALINDO, Amparo; JACKSON, George. Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures. **Fluid Phase Equilibria**, v. 266, n. 1–2, p. 105–128, abr. 2008.

HAYNES, William M. **CRC Handbook of Chemistry and Physics**. 95th. ed. [S.l.]: CRC Press, 2014. v. 1

HESS, Berk *et al.* LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, set. 1997.

INGÓLFSSON, Helgi I.; ANDERSEN, Olaf S. Alcohol's Effects on Lipid Bilayer Properties. **Biophysical Journal**, v. 101, n. 4, p. 847–855, ago. 2011.

IWAHASHI, Makio *et al.* The Dynamical Structure of Normal Alcohols in Their Liquids as Determined by the Viscosity and Self-Diffusion Measurements. **Bulletin of the Chemical Society of Japan**, v. 59, n. 12, dez. 1986.

JAKALIAN, Araz; JACK, David B.; BAYLY, Christopher I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. **Journal of Computational Chemistry**, v. 23, n. 16, p. 1623–1641, dez. 2002.

JORGE, Miguel. Theoretically grounded approaches to account for polarization effects in fixed-charge force fields. **The Journal of Chemical Physics**, v. 161, n. 18, p. 180901, 14 nov. 2024.

JORGENSEN, William L.; MAXWELL, David S.; TIRADO-RIVES, Julian. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. **Journal of the American Chemical Society**, v. 118, n. 45, p. 11225–11236, 13 nov. 1996.

KABSCH, Wolfgang; SANDER, Christian. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577–2637, dez. 1983.

KAGAMI, Luciano *et al.* The ACPYPE web server for small-molecule MD topology generation. **Bioinformatics**, v. 39, n. 6, p. btad350, 1 jun. 2023.

KASHURIN, Oleg V. *et al.* Force field comparison for molecular dynamics simulations of liquid membranes. **Journal of Molecular Liquids**, v. 416, p. 126347, dez. 2024.

KE, Qia *et al.* Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations. **Journal of Molecular Liquids**, v. 365, p. 120116, nov. 2022.

KIM, Seonghoon *et al.* CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules: CHARMM-GUI Ligand Reader and Modeler for CHARMM Force Field Generation of Small Molecules. **Journal of Computational Chemistry**, v. 38, n. 21, p. 1879–1886, 5 jun. 2017.

KOHNS, Maximilian. Molecular simulation study of dielectric constants of pure fluids and mixtures. **Fluid Phase Equilibria**, v. 506, p. 112393, fev. 2020.

LEACH, Andrew R. **Molecular Modelling: Principles and Applications**. 2nd. ed. Upper Saddle River, NJ: Prentice Hall / Pearson Education, 2001.

LEE, Ji-Eun *et al.* The Influence of Solvent Choice on the Extraction of Bioactive Compounds from Asteraceae: A Comparative Review. **Foods**, v. 13, n. 19, p. 3151, 2 out. 2024.

LEMKUL, Justin A.; HUANG, Jing; MACKERELL, Alexander D. Induced Dipole–Dipole Interactions Influence the Unfolding Pathways of Wild-Type and Mutant Amyloid β -Peptides. **The Journal of Physical Chemistry B**, v. 119, n. 51, p. 15574–15582, 24 dez. 2015.

LEONTYEV, I. V.; STUCHEBRUKHOV, A. A. Electronic continuum model for molecular dynamics simulations. **The Journal of Chemical Physics**, v. 130, n. 8, p. 085102, 28 fev. 2009.

LEONTYEV, I. V.; STUCHEBRUKHOV, A. A. Electronic Continuum Model for Molecular Dynamics Simulations of Biological Molecules. **Journal of Chemical Theory and Computation**, v. 6, n. 5, p. 1498–1508, 11 maio 2010.

MADEIRA, Lucas; VITIELLO, Silvio A. A função radial de distribuição de pares para sistemas Lennard-Jones bidimensionais. **Revista Brasileira de Ensino de Física**, v. 34, n. 4, dez. 2012.

MAGINN, Edward J. *et al.* Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics. **Living Journal of Computational Molecular Science**, v. 2, n. 1, 2020.

MARK ABRAHAM *et al.* GROMACS 2023.5 Manual. 3 maio 2024.

MARTIN, Marcus G.; SIEPMANN, J. Ilja. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of *n* -Alkanes. **The Journal of Physical Chemistry B**, v. 102, n. 14, p. 2569–2577, 1 abr. 1998.

MARTÍNEZ, L. *et al.* PACKMOL : A package for building initial configurations for molecular dynamics simulations. **Journal of Computational Chemistry**, v. 30, n. 13, p. 2157–2164, out. 2009.

MONTENEGRO, João Gabriel Soares. **DINÂMICA MOLECULAR PARA AVALIAÇÃO ESTRUTURAL DE MICELAS REVERSAS DE CTAB EM MISTURA DE ISOCTANO E 1-BUTANOL: EFEITO DE ÍON HIDROTRÓPICO**. Dissertação de Mestrado—Departamento de Química Fundamental: Universidade Federal de Pernambuco, 2025.

NEUMANN, Martin. Dipole moment fluctuation formulas in computer simulations of polar systems. **Molecular Physics**, v. 50, n. 4, p. 841–858, 1983.

NGUYEN, Hoang-Thanh; BOUCHAUDY, Anne; SALMON, Jean-Baptiste. Microfluidic free interface diffusion: Measurement of diffusion coefficients and evidence of interfacial-driven transport phenomena. **Physics of Fluids**, v. 34, n. 5, 1 maio 2022.

NICK PACE, C. *et al.* Protein structure, stability and solubility in water and other solvents. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, v. 359, n. 1448, p. 1225–1235, 29 ago. 2004.

NOÉ, Frank *et al.* Machine Learning for Molecular Simulation. **Annual Review of Physical Chemistry**, v. 71, n. 1, p. 361–390, 20 abr. 2020.

OBEIDAT, Abdalla; AL-SALMAN, Rakan; ABU-GHAZLEH, Hind. The validity of the potential model in predicting the structural, dynamical, thermodynamic properties of the unary and binary mixture of water-alcohol: Ethanol-water case. **AIP Advances**, v. 8, n. 7, p. 075321, 1 jul. 2018.

ORABI, Esam A. *et al.* Corrections in the CHARMM36 Parametrization of Chloride Interactions with Proteins, Lipids, and Alkali Cations, and Extension to Other Halide Anions. **Journal of Chemical Theory and Computation**, v. 17, n. 10, p. 6240–6261, 12 out. 2021.

PARRINELLO, Michele; RAHMAN, Aneesur. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182–7190, 15 dez. 1981.

PATRA, Michael *et al.* Under the Influence of Alcohol: The Effect of Ethanol and Methanol on Lipid Bilayers. **Biophysical Journal**, v. 90, n. 4, p. 1121–1135, fev. 2006.

PEDREGOSA, Fabian *et al.* Scikit-learn: Machine Learning in Python. **MACHINE LEARNING IN PYTHON**, 2011.

PELKMAN, Luc; VIEIRA, José Nilton de Souza; LEAL, Marlon Arraes Jardim. **Implementation of Bioenergy in Brazil – 2024 Update**: IEA Bioenergy Country Reports. [S.l.]: EA Bioenergy – Technology Collaboration Programme, dez. 2024.

PÉREZ, Encarnación García; GONZÁLEZ-SALGADO, Diego; LOMBA, Enrique. Molecular dynamics simulations of aqueous solutions of short chain alcohols. Excess properties and the temperature of maximum density. **Fluid Phase Equilibria**, v. 528, p. 112840, jan. 2021.

PÉREZ-ENCISO, Miguel; ZINGARETTI, Laura M. A Guide on Deep Learning for Complex Trait Genomic Prediction. **Genes**, v. 10, n. 7, p. 553, 20 jul. 2019.

PERIYASAMY, Aravin Prince; DUTTA, Shubhajit; TEHRANI-BAGHA, Ali R. Solvent-assisted salt-free reactive dyeing of cotton fabric. **Cellulose**, v. 31, n. 11, p. 7101–7118, jul. 2024.

POLÊTO, Marcelo D.; LEMKUL, Justin A. Integration of experimental data and use of automated fitting methods in developing protein force fields. **Communications Chemistry**, v. 5, n. 1, p. 38, 18 mar. 2022.

POOLE, Colin F. Assessment of liquid-liquid partition for the assignment of descriptors for the solvation parameter model. **Journal of Chromatography A**, v. 1721, p. 464850, abr. 2024.

SAGANDIRA, Clodius R. *et al.* The evolution of Tamiflu synthesis, 20 years on: Advent of enabling technologies the last piece of the puzzle? **Tetrahedron**, v. 76, n. 37, p. 131440, set. 2020.

SASTRY, N. V.; VALAND, M. K. Dielectric constants, refractive indexes and polarizations for 1-Alcohol +Heptane mixtures at 298.15 and 308.15 K. **Berichte der Bunsengesellschaft für physikalische Chemie**, v. 101, n. 2, p. 243–250, fev. 1997.

SILBERSTEIN, E.; MAKOV, G. Short range order of methanol. **Journal of Molecular Liquids**, v. 361, p. 119590, set. 2022.

SOTO, Claudio *et al.* The α -Helical to β -Strand Transition in the Amino-terminal Fragment of the Amyloid β -Peptide Modulates Amyloid Formation. **Journal of Biological Chemistry**, v. 270, n. 7, p. 3063–3067, fev. 1995.

TOMŠIČ, Matija *et al.* Structural Properties of Pure Simple Alcohols from Ethanol, Propanol, Butanol, Pentanol, to Hexanol: Comparing Monte Carlo Simulations with Experimental SAXS Data. **The Journal of Physical Chemistry B**, v. 111, n. 7, p. 1738–1751, 1 fev. 2007.

VANOMMESLAEGHE, K.; RAMAN, E. Prabhu; MACKERELL, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial

Atomic Charges. **Journal of Chemical Information and Modeling**, v. 52, n. 12, p. 3155–3168, 21 dez. 2012.

WANG, Junmei *et al.* Development and testing of a general amber force field. **Journal of Computational Chemistry**, v. 25, n. 9, p. 1157–1174, 15 jul. 2004.

WEI, Guanghong; SHEA, Joan-Emma. Effects of Solvent on the Structure of the Alzheimer Amyloid- β (25–35) Peptide. **Biophysical Journal**, v. 91, n. 5, p. 1638–1647, set. 2006.

YANG, Tao *et al.* Molecular dynamics investigation on the vapor–liquid interface behavior of long-chain alkanes, alcohols, and their mixtures. **Journal of Molecular Liquids**, v. 375, p. 121283, abr. 2023.

YEH, In-Chul; HUMMER, Gerhard. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. **The Journal of Physical Chemistry B**, v. 108, n. 40, p. 15873–15879, 1 out. 2004.

YOO, Jejoong; AKSIMENTIEV, Aleksei. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. **Physical Chemistry Chemical Physics**, v. 20, n. 13, p. 8432–8449, 2018.

ZANGI, Ronen. Refinement of the OPLSAA Force-Field for Liquid Alcohols. **ACS Omega**, v. 3, n. 12, p. 18089–18099, 31 dez. 2018.

ZÊZERE, Bruno *et al.* Influence of Ethanol Parametrization on Diffusion Coefficients Using OPLS-AA Force Field. **International Journal of Molecular Sciences**, v. 24, n. 8, p. 7316, 15 abr. 2023.

APÊNDICE A – PARÂMETROS ADICIONAIS

Tabela A1 - Distância do primeiro mínimo da função de distribuição radial $g(r)$ para álcoois lineares

Álcool	Primeiro mínimo de $g(r)$ / Å
Etanol	3,45
1-Propanol	3,45
1-Butanol	3,43
1-Pentanol	3,47
1-Hexanol	3,45

Fonte: Tomšič *et al.* (2007)

Tabela A2 – Parâmetros dos sistemas simulados empregados no treinamento inicial do modelo de aprendizado de máquina (Iteração 1)

Álcool	σ/σ_0	ϵ/ϵ_0	dq/dq_0
1-Pentanol Etanol	0,95	0,9	0,95
1-Hexanol 1-Propanol	0,95	0,9	1
1-Pentanol Etanol	0,95	0,9	1,05
1-Hexanol 1-Propanol	0,95	1	0,95
Etanol 1-Butanol	0,95	1	1
1-Pentanol Etanol	0,95	1	1,05
1-Propanol 1-Pentanol	0,95	1,1	0,95
1-Butanol 1-Hexanol	0,95	1,1	1
1-Butanol 1-Hexanol	0,95	1,1	1,05
1-Propanol 1-Pentanol	0,975	0,9	0,95
Etanol 1-Butanol	0,975	0,9	1
1-Hexanol 1-Propanol	0,975	0,9	1,05
1-Pentanol	0,975	1	0,95

Etanol			
1-Propanol	0,975	1	1
1-Pentanol			
1-Butanol	0,975	1	1,05
1-Hexanol			
1-Hexanol	0,975	1,1	0,95
1-Propanol			
Etanol	0,975	1,1	1
1-Butanol			
1-Pentanol	0,975	1,1	1,05
Etanol			
1-Propanol	1	0,9	0,95
1-Pentanol			
1-Butanol	1	0,9	1
1-Hexanol			
Etanol	1	0,9	1,05
1-Butanol			
1-Butanol	1	1	0,95
1-Hexanol			
1-Hexanol	1	1	1
1-Propanol			
Etanol	1	1	1,05
1-Butanol			
1-Propanol	1	1,1	0,95
1-Pentanol			
1-Hexanol	1	1,1	1
1-Propanol			
1-Butanol	1	1,1	1,05
1-Hexanol			

Fonte: elaborado pelo próprio autor

Tabela A3 – Parâmetros dos sistemas simulados na segunda iteração do aprendizado de máquina

Álcool	σ/σ_0	ϵ/ϵ_0	dq/dq_0
Etanol	0,908	0,818	0,842
	0,000	0,864	0,879
1-Propanol	0,946	0,818	0,926
	0,000	0,818	0,861
1-Butanol	0,940	0,818	0,879
	0,921	0,886	0,889
1-Pentanol	0,940	0,818	0,907
	0,908	0,932	0,870

1-Hexanol	0,914	0,795	0,870
	0,933	0,841	0,889

Fonte: elaborado pelo próprio autor

Tabela A4 – Parâmetros dos sistemas simulados na terceira iteração do aprendizado de máquina

Álcool	σ/σ_0	ϵ/ϵ_0	dq/dq_0
Etanol	0,946	0,895	0,935
1-Propanol	0,946	0,895	0,926
1-Butanol	0,940	0,895	0,879
1-Pentanol	0,940	0,895	0,907
1-Hexanol	0,927	0,895	0,870

Fonte: elaborado pelo próprio autor

Nota: valores em verde representam os parâmetros finais para um dado álcool

Tabela A5 – Parâmetros dos sistemas simulados na quarta iteração do aprendizado de máquina

Álcool	σ/σ_0	ϵ/ϵ_0	dq/dq_0
1-Butanol	0,946	0,970	0,935
1-Pentanol	0,946	0,920	0,935
1-Hexanol	0,914	0,945	0,870

Fonte: elaborado pelo próprio autor

Nota: valores em verde representam os parâmetros finais para um dado álcool

Tabela A6 – Parâmetros dos sistemas simulados na quinta iteração do aprendizado de máquina

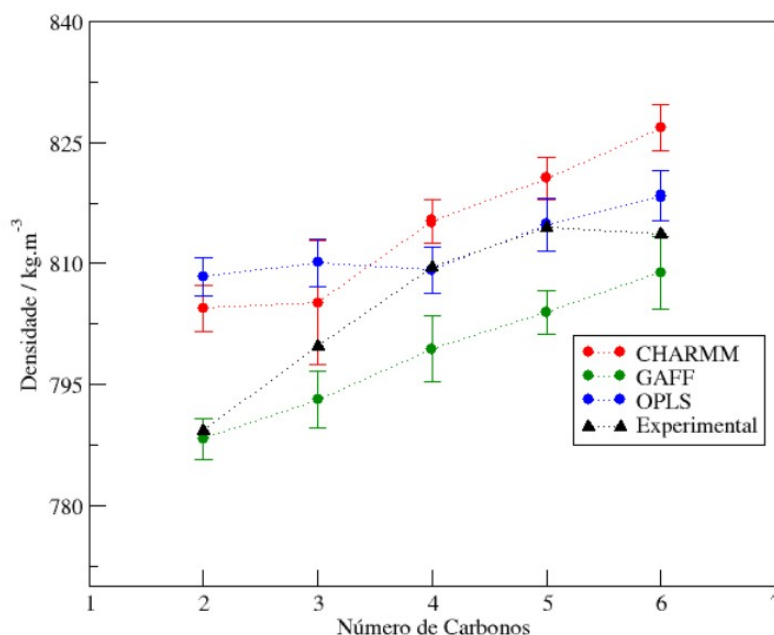
Álcool	σ/σ_0	ϵ/ϵ_0	dq/dq_0
1-Butanol	0,905	1,107	0,873
1-Pentanol	0,914	1,107	0,879
1-Hexanol	0,921	0,895	0,861

Fonte: elaborado pelo próprio autor

Nota: valores em verde representam os parâmetros finais para um dado álcool

APÊNDICE B – PROPRIEDADES INDIVIDUAIS PARA OS CAMPOS DE FORÇA ORIGINAIS

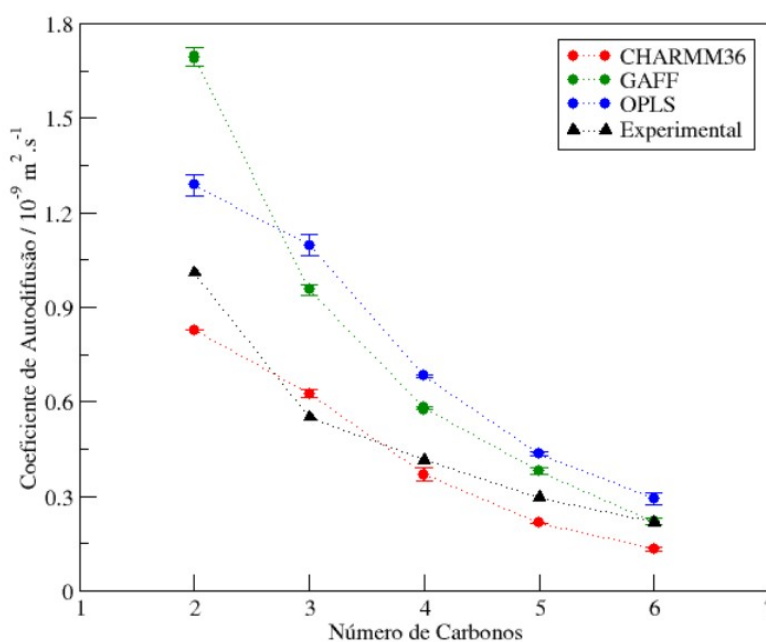
Gráfico B1 – Densidade em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As barras de erro representam o desvio padrão das medidas. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Haynes, (2014)

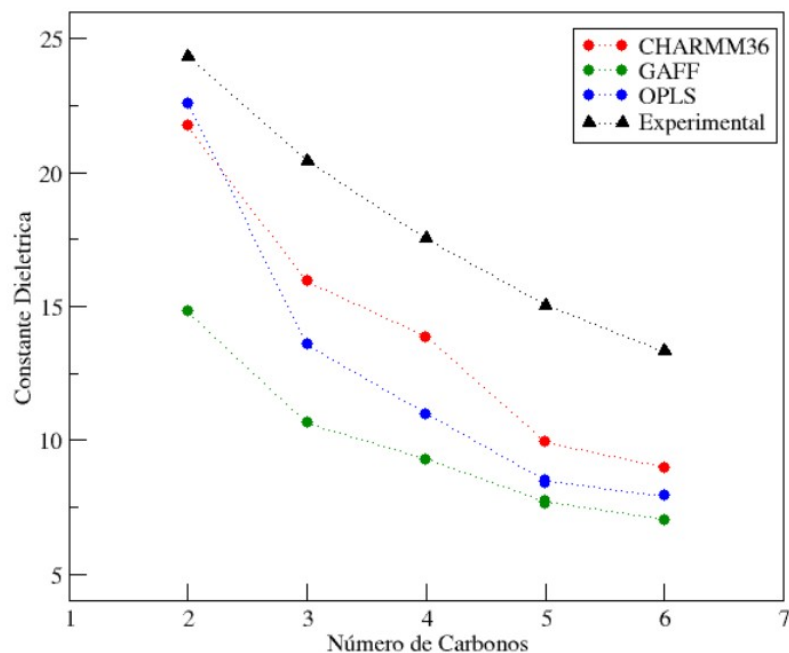
Gráfico B2 – Coeficiente de autodifusão em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As barras de erro representam o desvio padrão das medidas. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Iwahashi *et al.* (1986)

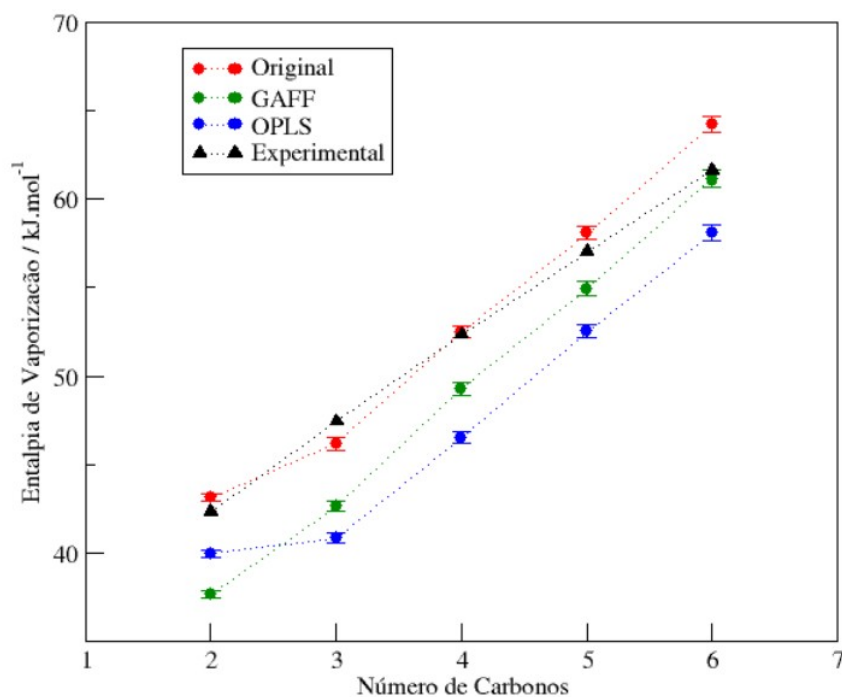
Gráfico B3 – Constante dielétrica em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Bezman, Casassa e Kay. (1997); Sastry e Valand. (1997).

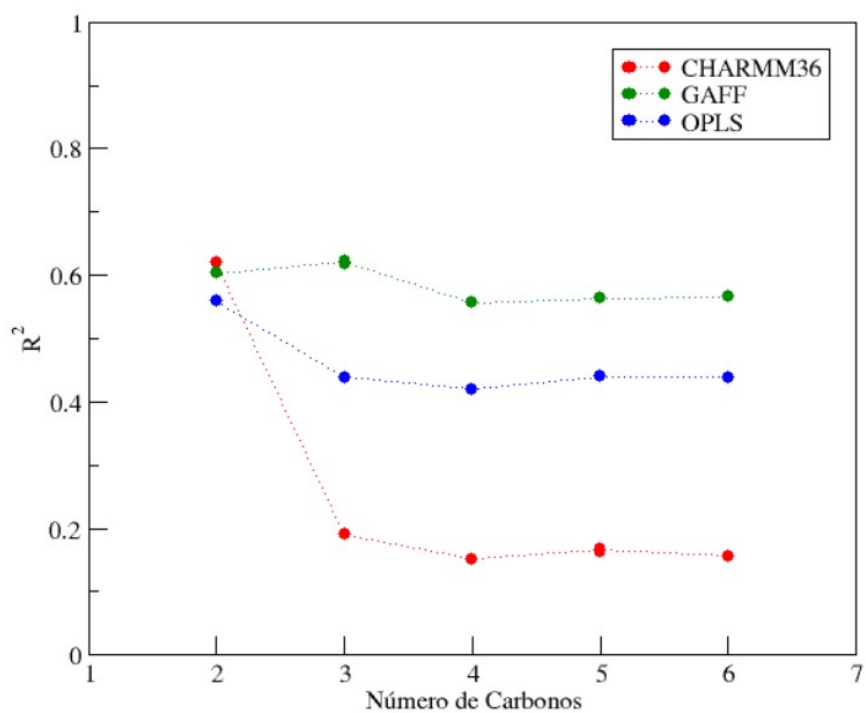
Gráfico B4 – Entalpia de vaporização em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As barras de erro representam o desvio padrão das medidas. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Haynes, (2014)

Gráfico B5 – Coeficiente de determinação em função do tamanho da cadeia dos álcoois simulados



Fonte: elaborado pelo próprio autor

Nota: As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Valores de R^2 obtidos entre os resultados das simulações e os dados experimentais retirados de Tomšič *et al.* (2007).

APÊNDICE C – PROPRIEDADES PREDITAS E SIMULADAS DOS ÁLCOOIS REPARAMETRIZADOS

Tabela C1 – Propriedades previstas pelos modelos de aprendizado de máquina

Álcool	Modelo	ρ	D	$\Delta\epsilon$	ΔH_{vap}	R^2
Etanol	$F(C_1)$	810,982	0,985	18,591	41,469	0,972
1-Propanol	$F(C_1)$	814,492	0,546	12,687	46,126	0,962
1-Butanol	$G(C_3)$	823,205	0,389	7,964	51,165	0,980
1-Pentanol	$G(C_3)$	826,999	0,256	6,433	56,190	0,970
1-Hexanol	$G(C_2)$	832,012	0,167	6,145	61,438	0,966

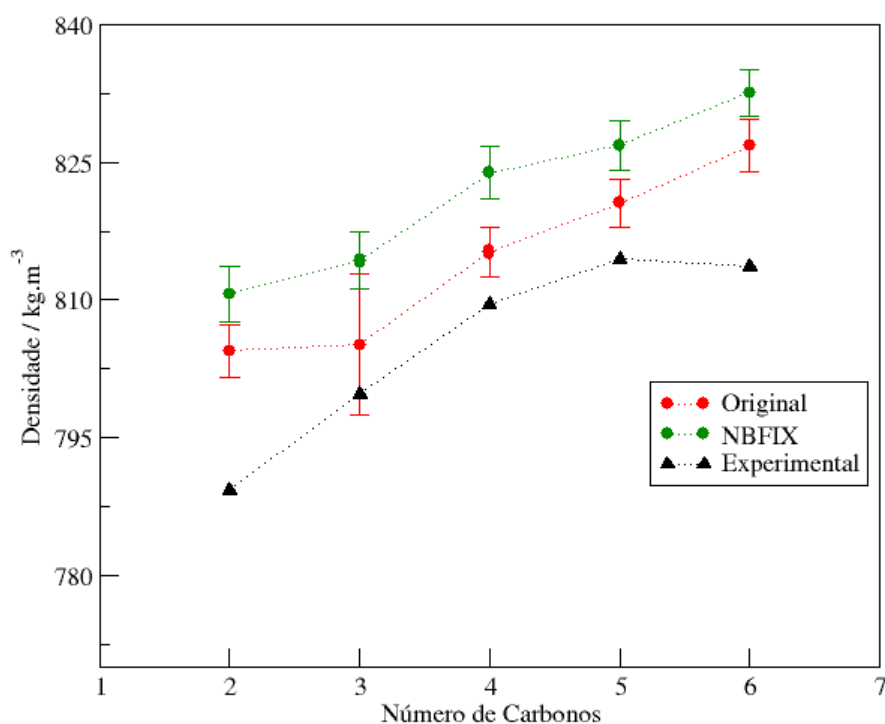
Fonte: elaborado pelo próprio autor

Tabela C2 – Propriedades obtidas através das simulações dos álcoois utilizando os conjuntos finais
de parâmetros

Álcool	ρ	D	$\Delta\epsilon$	ΔH_{vap}	R^2
Etanol	810,650	0,954	16,663	40,983	0,997
1-Propanol	814,320	0,521	11,325	46,259	0,994
1-Butanol	823,860	0,351	8,587	50,727	0,965
1-Pentanol	826,820	0,254	8,066	56,344	0,982
1-Hexanol	832,550	0,160	5,317	61,646	0,966

Fonte: elaborado pelo próprio autor

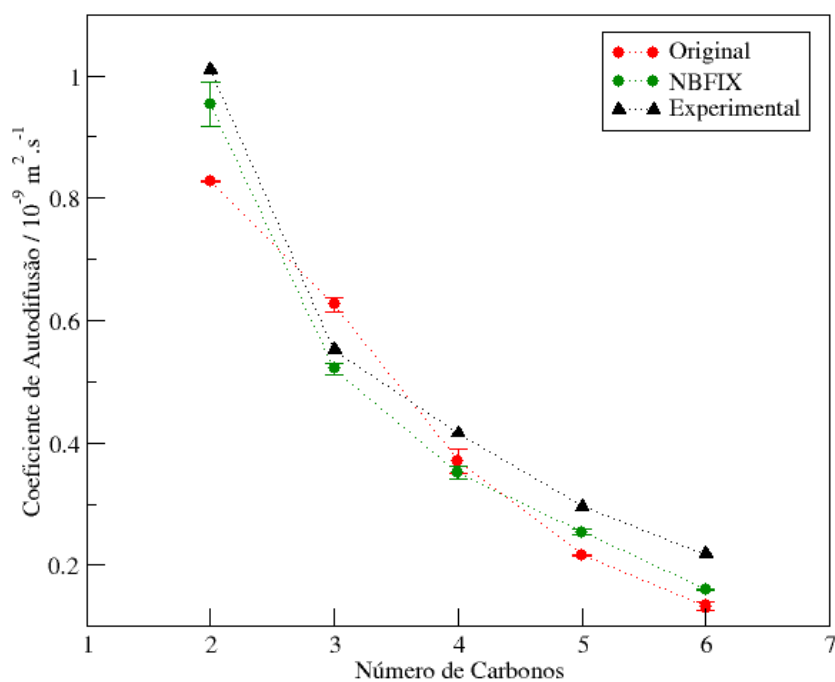
Gráfico C1 – Comparação da densidade em função do tamanho da cadeia dos
álcoois



Fonte: elaborado pelo próprio autor

Nota: As barras de erro representam o desvio padrão das medidas. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Haynes, (2014)

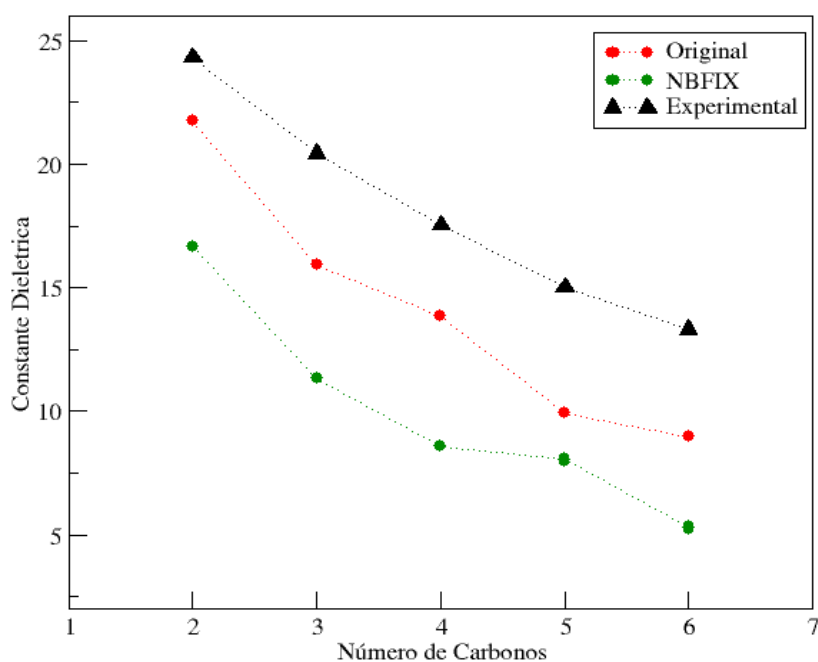
Gráfico C2 – Comparação da coeficiente de autodifusão em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As barras de erro representam o desvio padrão das medidas. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Iwahashi *et al.* (1986)

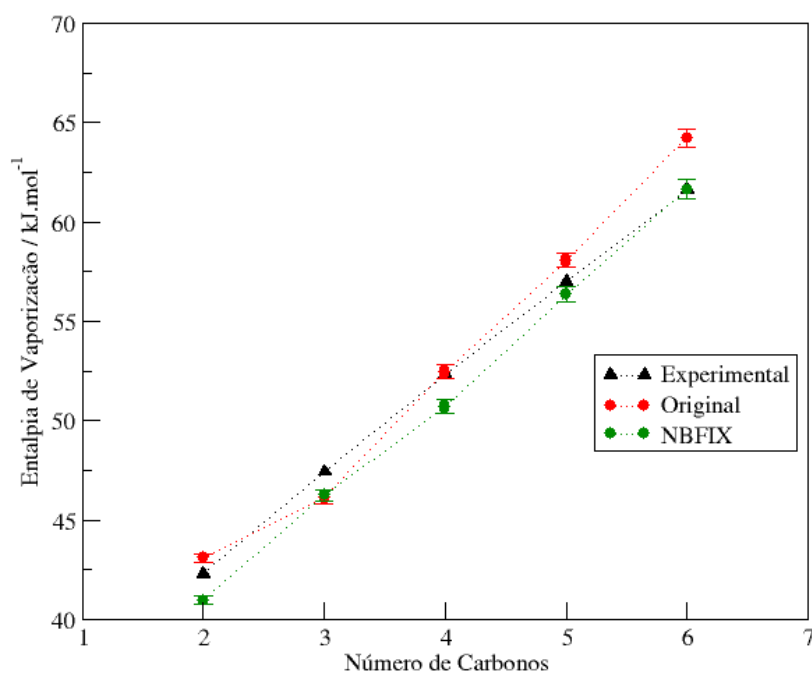
Gráfico C3 – Comparação da constante dielétrica em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Bezman, Casassa e Kay. (1997); Sastry e Valand. (1997).

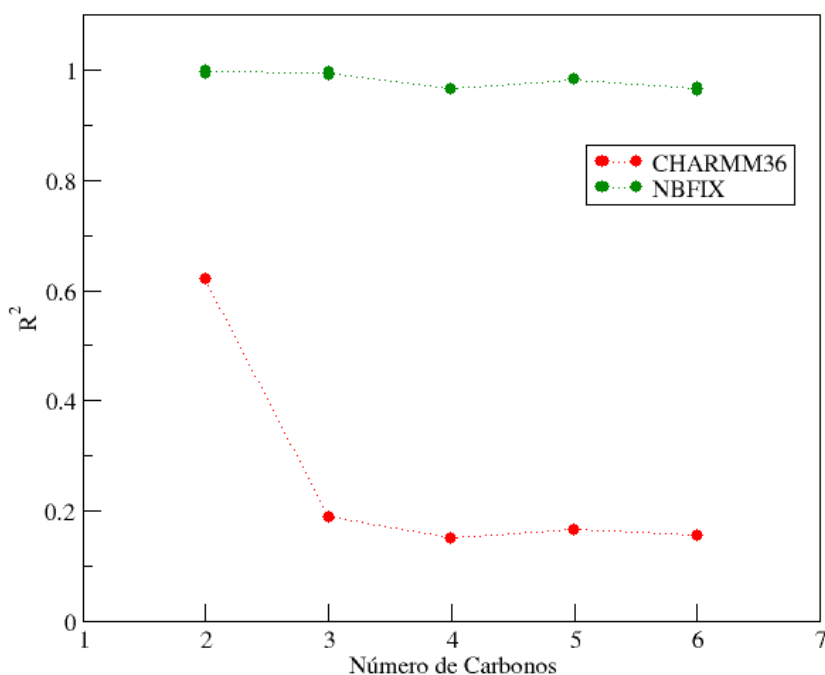
Gráfico C4 – Comparação da entalpia de vaporização em função do tamanho da cadeia dos álcoois



Fonte: elaborado pelo próprio autor

Nota: As barras de erro representam o desvio padrão das medidas. As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Dados experimentais retirados de Haynes, (2014)

Gráfico C5 – Comparação do coeficiente de determinação em função do tamanho da cadeia dos álcoois

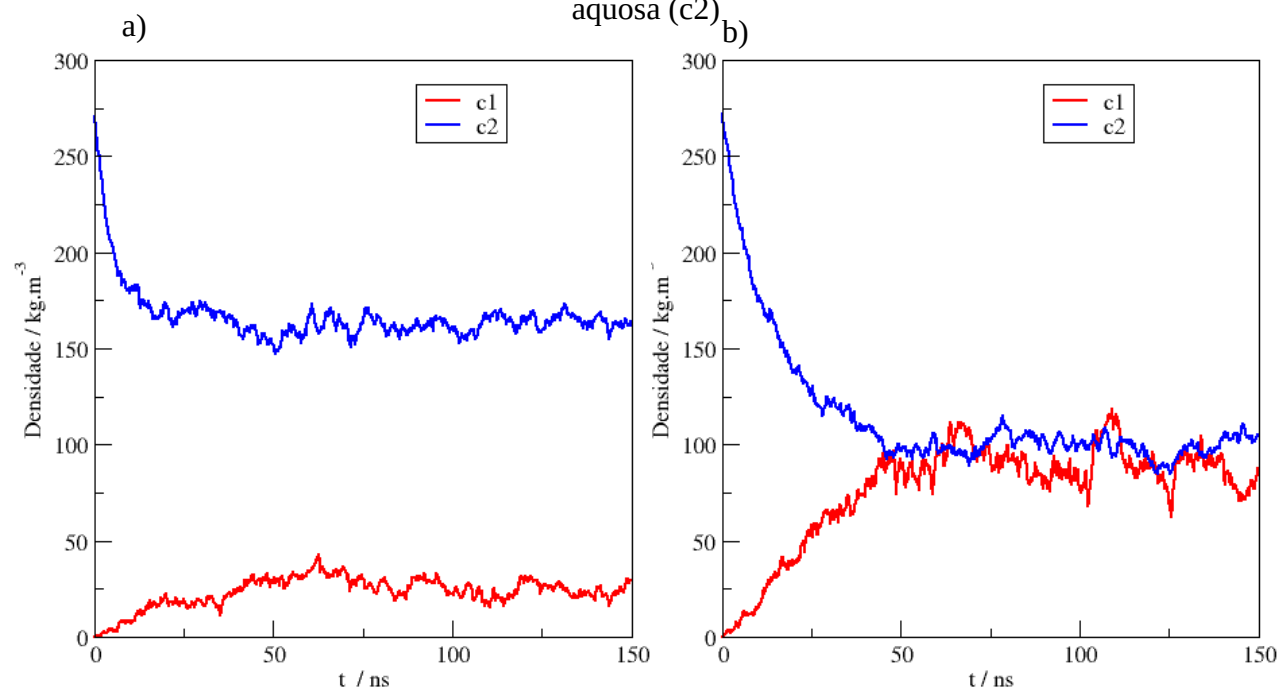


Fonte: elaborado pelo próprio autor

Nota: As linhas pontilhadas foram adicionadas apenas como guia visual para a tendência dos dados. Valores de R^2 obtidos entre os resultados das simulações e os dados experimentais retirados de Tomšič *et al.* (2007).

APÊNDICE D – INFORMAÇÕES ADICIONAIS PARA OS SISTEMAS MULTICOMPONENTES

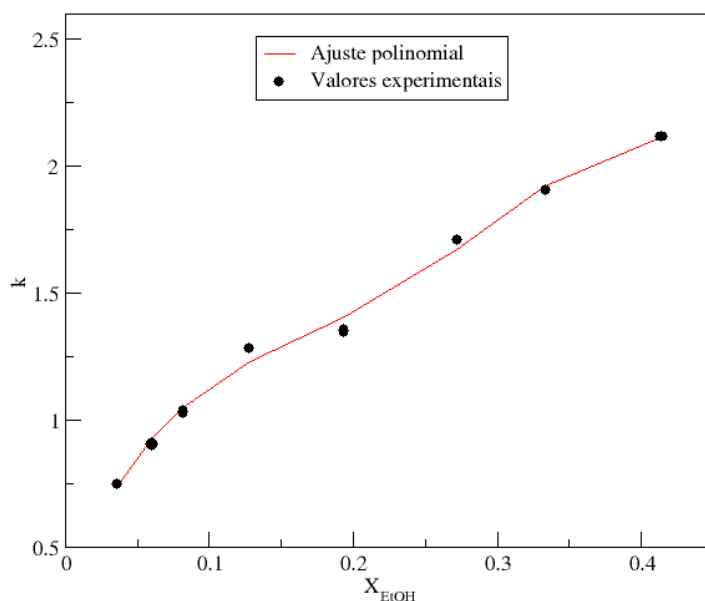
Gráfico D1 - Evolução temporal da densidade de etanol no *bulk* das fases orgânica (c1) e aquosa (c2)



Fonte: elaborado pelo autor

Nota: gráfico a) obtido utilizando os parâmetros originais do CHARM36; gráfico b) obtido com etanol reparametrizado.

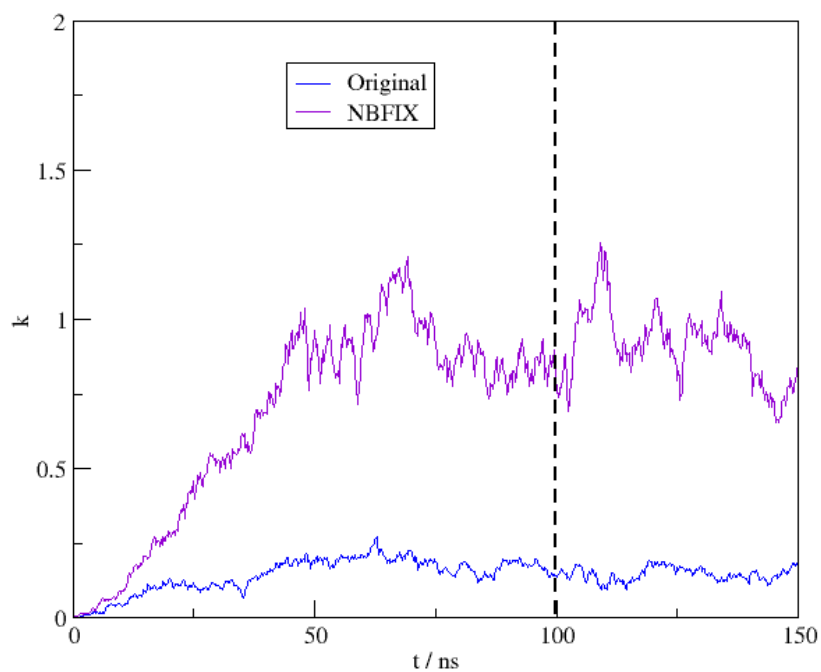
Gráfico D2 – Ajuste polinomial para o coeficiente de partição do etanol em função da fração molar em DIPE



Fonte: elaborado pelo autor a partir dos dados de Arce *et al.* (2002)

Nota: o polinômio obtido foi $y(x) = -290x^4 + 272,6x^3 - 87,66x^2 + 14,5x + 0,32$, o qual apresenta $R^2 = 0,995$ com os pontos utilizados.

Gráfico D3 – Coeficiente de partição do etanol em função do tempo nas múltiplas simulações



Fonte: elaborado pelo próprio autor

Nota: a linha vertical tracejada em $t=100$ ns marca o início do intervalo sobre o qual foi calculada a média de k .