

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
TRABALHO DE CONCLUSÃO DE CURSO

**CHEMSTRUCT: UM PACOTE EM PYTHON PARA ANÁLISE
ESTRUTURAL DE SISTEMAS ATÔMICOS E AUXÍLIO EM
SIMULAÇÕES DE DINÂMICA MOLECULAR**

TRABALHO DE CONCLUSÃO DE CURSO

PEDRO GUERRA DEMINGOS

**PORTO ALEGRE, RS
2019**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
TRABALHO DE CONCLUSÃO DE CURSO

**CHEMSTRUCT: UM PACOTE EM PYTHON PARA ANÁLISE
ESTRUTURAL DE SISTEMAS ATÔMICOS E AUXÍLIO EM
SIMULAÇÕES DE DINÂMICA MOLECULAR**

PEDRO GUERRA DEMINGOS

Trabalho de Conclusão de curso apresentado como requisito parcial para obtenção do título de Engenheiro Químico.

Orientador:
Prof. André Rodrigues Muniz, D.Sc.

**PORTO ALEGRE, RS
2019**

Demingos, P. G.

ChemStruct: um pacote em Python para análise estrutural de sistemas atômicos e auxílio em simulações de dinâmica molecular / Pedro Guerra Demingos. -- 2019. 51 f.

Orientador: André Rodrigues Muniz, D.Sc.

Trabalho (Conclusão de curso) - Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Departamento de Engenharia Química, Porto Alegre, BR-RS, 2019.

Python, estrutura química, dinâmica molecular, potenciais interatômicos I. Muniz, A. R., orient.
II. Título.

Agradecimentos

Agradeço aos dois orientadores que tive ao longo dos cinco anos de graduação: o professor André Muniz, do Departamento de Engenharia Química, e a professora Jacqueline Ferreira, do Instituto de Química. Um trabalho de mil páginas não faria jus ao quanto aprendi com cada um deles.

Resumo

Ferramentas computacionais de análise estrutural são de grande interesse em simulação molecular, uma vez que a predição da dinâmica das partículas depende das interações entre elas, que por sua vez são regidas pela configuração atômica local. Em vista disso, aqui é apresentado o ChemStruct, um pacote desenvolvido em Python para classificação atômica e identificação de grupos funcionais em moléculas, bem como para a geração de arquivos de entrada adequados para simulações de dinâmica molecular no pacote LAMMPS. Neste trabalho são descritas e avaliadas as partes centrais do código, como o algoritmo classificador e os dois métodos implementados para o cálculo das ligações químicas. As estruturas identificadas pelo pacote são validadas, e exemplos de auxílio em simulações de dinâmica molecular são apresentados, demonstrando que o pacote cumpre bem seus objetivos. Em suma, o ChemStruct ajuda a elucidar as características estruturais de uma grande gama de moléculas, auxiliando na compreensão de suas propriedades geométricas e vibracionais e possibilitando simulações de dinâmica molecular com parametrização adequada. Essa ferramenta pode ser utilizada por químicos teóricos, cientistas de materiais e engenheiros do âmbito computacional, e também como instrumento didático.

Palavras-chave: Python, estrutura química, dinâmica molecular, potenciais interatômicos

Sumário

Lista de Figuras	viii
Lista de Tabelas	ix
Lista de Símbolos	xii
1 Introdução	1
2 Revisão Bibliográfica	3
2.1 Python	3
2.2 Dinâmica Molecular	4
2.3 LAMMPS	7
2.4 CGenFF: CHARMM General Force Field	8
3 Metodologia	13
3.1 Estrutura do Pacote	13
3.2 Cálculo das Ligações Químicas	15
3.3 Outras Estruturas Topológicas	15
3.4 Classificação dos Átomos	17
3.5 Funções para Utilização do Usuário	21
4 Resultados e Discussão	23
4.1 Eficiência do Cálculo das Ligações Químicas	23
4.2 Validação da Classificação dos Átomos	24
4.3 Utilização do Pacote para Auxílio em Simulações de Dinâmica Molecular	26
5 Conclusões e Trabalhos Futuros	33
Referências Bibliográficas	35

Lista de Figuras

Figura 2.1	Numeração de átomos ligados, ilustrando os pares 1-2, 1-3 e 1-4. . . .	6
Figura 2.2	Molécula de etano, ilustrando geometria de ligação, ângulo e diedro próprio.	10
Figura 3.1	Divisão em regiões realizada para otimizar o cálculo das ligações. . .	16
Figura 4.1	Eficiência comparada dos dois métodos de cálculo das ligações. . . .	23
Figura 4.2	Hidrocarboneto com classificação atômica.	24
Figura 4.3	Ácido carboxílico com classificação atômica.	25
Figura 4.4	Aldeído com classificação atômica.	25
Figura 4.5	Moléculas cíclicas com classificação atômica.	25
Figura 4.6	Arquivo de texto em formato <i>.xyz</i> de molécula de etano.	27
Figura 4.7	Arquivo de texto em formato LAMMPS Data File (<i>.lmp</i>) para uma molécula de etano, gerado pelo ChemStruct.	28
Figura 4.8	Sistema com 125 moléculas de etano.	30

Lista de Tabelas

Tabela 3.1	Classificações de carbonos alifáticos segundo o ChemStruct e o CGenFF.	20
------------	--	----

Lista de Símbolos

χ	Ângulo de diedro impróprio	rad
χ_0	Ângulo mais estável de diedro impróprio	rad
Δt	Intervalo de Tempo	s
ϵ	Energia característica de Lennard-Jones	kCal mol ⁻¹
$\mathcal{O}(x)$	Valor da mesma ordem de grandeza que x	[x]
ϕ	Ângulo de torção	rad
σ	Distância característica de Lennard-Jones	Å
θ	Ângulo entre duas ligações	rad
θ_0	Valor mais estável de ângulo	rad
θ_{UB}	Ângulo de Urey-Bradley	rad
$\vec{\nabla}$	Operador Gradiente	m ⁻¹
\vec{a}	Aceleração de uma Partícula	m s ⁻²
\vec{F}	Força	N
\vec{v}	Velocidade de uma Partícula	m s ⁻¹
\vec{x}	Posição de uma Partícula	m
d	Deslocamento de termo torsional da energia	–
E_e	Energia de Coulomb	kCal mol ⁻¹
E_r	Energia de Ligação	kCal mol ⁻¹
E_χ	Energia de Diedro Impróprio	kCal mol ⁻¹
E_ϕ	Energia (Torsional) de Diedro Próprio	kCal mol ⁻¹
E_θ	Energia Angular	kCal mol ⁻¹

E_{LJ}	Energia de Lennard-Jones	kCal mol ⁻¹
k_e	Constante de Coulomb	kCal mol ⁻¹ Å e ⁻²
k_r	Constante de energia de ligação	kCal mol ⁻¹ Å ⁻²
k_χ	Constante de energia de diedro impróprio	kCal mol ⁻¹ rad ⁻²
k_ϕ	Constante de energia de torsão	kCal mol ⁻¹ rad ⁻²
k_θ	Constante de energia de ângulo	kCal mol ⁻¹ rad ⁻²
k_{UB}	Constante de energia de Urey-Bradley	kCal mol ⁻¹ rad ⁻²
m	Massa	kg
n	Valor inteiro positivo	rad ⁻¹
q_1, q_2	Carga de partículas	e
r	Posição ou Distância	Å
r_0	Distância mais estável de ligação	Å
t	Tempo	s
U	Energia Potencial	J

Siglas

AIREBO	Adaptive Intermolecular Reactive Empirical Bond Order
ASE	Atomistic Simulation Environment
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at HARvard Macromolecular Mechanics
DM	Dinâmica Molecular
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
ReaxFF	Reactive Force Field

Capítulo 1

Introdução

Ferramentas computacionais são cada vez mais comuns no cotidiano de cientistas e engenheiros, uma vez que permitem a análise sistemática de grandes quantidades de dados e a rápida manipulação de arquivos com enormes volumes de informação, além de possibilitar a simulação de processos de interesse acadêmico e industrial. No ramo da engenharia química destacam-se a simulação de processos, a fluidodinâmica computacional e a dinâmica molecular (MD, de *Molecular Dynamics*), que também é amplamente utilizada na bioquímica e na ciência de materiais.

A utilização dessas ferramentas para a simulação de determinados compostos exige o conhecimento de suas propriedades físico-químicas, que por sua vez dependem de suas características moleculares, podendo ser obtidas através de medições experimentais ou, ainda, de cálculos teóricos – caso em que é necessário conhecer detalhadamente a estrutura química das espécies envolvidas. Em simulações de MD, em particular, não são usados dados macroscópicos das substâncias, mas suas características a nível atômico, como a geometria molecular e a natureza e intensidade das ligações químicas e forças intermoleculares, o que torna a análise estrutural ainda mais importante nesse caso.

Uma vez que diversos métodos de simulação computacional relevantes às ciências e engenharias se beneficiam de ferramentas capazes de elucidar estruturas moleculares, foi desenvolvido o ChemStruct (DEMINGOS, 2019), um pacote de Python para análise estrutural de sistemas atômicos. Essa ferramenta tem como objetivo primário a identificação de ligações químicas e a classificação de tipos atômicos, partindo de uma lista contendo apenas as coordenadas espaciais de cada átomo, identificando

para isso as hibridizações e grupos funcionais. Uma funcionalidade secundária é a preparação de arquivos de entrada em formato adequado para simulações MD, contendo não apenas os átomos devidamente classificados, mas também uma lista de estruturas topológicas (ligações, ângulos e diedros), bem como os correspondentes parâmetros de potenciais intermoleculares necessários para a simulação do sistema.

O presente trabalho traz dados relativos à primeira versão do ChemStruct (2019.06). O pacote segue em desenvolvimento, e sua versão mais recente pode ser encontrada no link <<https://github.com/pdemingos/chemstruct>>.

Capítulo 2

Revisão Bibliográfica

2.1 Python

Python (PYTHON-SOFTWARE-FOUNDATION, 2001) é uma linguagem de programação orientada a objetos, cuja filosofia é diminuir o esforço do programador, priorizando a legibilidade do código antes da sua velocidade. Lançada em 1991, é atualmente uma das linguagens mais utilizadas por programadores, possuindo amplos pacotes para ciência e engenharia, como os módulos Numpy, Scipy e Matplotlib (NUMPY-TEAM, 2019; SCIPY-TEAM, 2019; HUNTER et al., 2002).

Apesar de ser uma linguagem orientada a objetos, Python permite a utilização de programação funcional, isto é, da definição de funções em vez de classes de objetos. Isso dá à linguagem uma maior flexibilidade, uma vez que a abordagem funcional é similar ao raciocínio numérico-matemático, enquanto classes customizadas permitem maior reutilização de código, além de torná-lo mais intuitivo e de fácil manutenção.

Essas características tornam Python uma escolha adequada para projetos como o ChemStruct (DEMINGOS, 2019), que conta tanto com programação orientada a objetos quanto com funções criadas para facilitar a utilização por parte do usuário. A filosofia da linguagem também garante fácil manutenção do código, bem como a sua ampliação por parte de cientistas e engenheiros, não havendo para isso a necessidade de conhecimentos avançados de programação.

2.2 Dinâmica Molecular

Dinâmica molecular clássica (MD) é um tipo de simulação computacional para estudo de sistemas atômicos. Esse método é baseado na integração da equação do movimento (Segunda Lei de Newton), tratando os átomos como corpos pontuais clássicos que interagem uns com os outros de acordo com um potencial interatômico pré-definido, parametrizado empiricamente (FRENKEL; SMIT, 2002; RAPAPORT, 2004). Como resultado tem-se as trajetórias dinâmicas dos átomos, que permitem a observação direta de fenômenos na escala de tempo e comprimento correspondente, além do cálculo de propriedades macroscópicas mensuráveis usando relações da termodinâmica estatística. É um método amplamente empregado por químicos, engenheiros, biólogos e cientistas de materiais. Pode ser utilizado para o estudo de fenômenos de transporte (MOULTOS et al., 2014), das propriedades de nanomateriais (MUNIZ et al., 2015; DEMINGOS; MUNIZ, 2018), de processos de separação por membranas (WU et al., 2014), de reações químicas (LIU; GUO, 2017), entre diversas outras aplicações de interesse científico e industrial, além de ser uma excelente ferramenta didática para o ensino de princípios de físico-química e teoria cinética.

Simulações MD têm dois passos fundamentais: o cálculo das forças que agem sobre as partículas e a integração temporal de seus efeitos na dinâmica do sistema. A primeira parte faz uso da relação fundamental que existe entre força e energia potencial (equação 2.1). Portanto, conhecendo-se esse potencial, podem ser calculadas as interações entre os átomos.

$$\vec{F} = m \frac{d^2 r}{dt^2} = -\vec{\nabla}U \quad (2.1)$$

A segunda parte, por sua vez, consiste na obtenção da posição e da velocidade de cada partícula ao longo do tempo, o que pode ser realizado por qualquer algoritmo numérico de integração temporal, como o integrador de Verlet (equações 2.2 e 2.3), para atualizar as posições e velocidades dos átomos com base nas forças previamente

calculadas.

$$\vec{x}(t + \Delta t) = 2\vec{x}(t) - \vec{x}(t - \Delta t) + \vec{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (2.2)$$

$$\vec{v}(t + \Delta t) = \frac{\vec{x}(t + \Delta t) - \vec{x}(t)}{\Delta t} + \mathcal{O}(\Delta t) \quad (2.3)$$

Além disso, utilizando as relações da mecânica estatística, é possível calcular propriedades macroscópicas do sistema (temperatura, pressão, etc) a partir da média de propriedades calculadas em uma sequência de microestados (posição e velocidade de cada átomo).

A modelagem da energia potencial entre os átomos do sistema é o coração da DM, e chama-se Potencial Interatômico qualquer conjunto de equações proposto para esse fim. Os potenciais mais simples são aqueles que modelam as interações entre pares; um exemplo típico é o de Lennard-Jones (equação 2.4), que modela a atração das forças de van der Waals e a repulsão de Pauli (FRENKEL; SMIT, 2002; RAPAPORT, 2004). Sozinho, esse potencial tem aplicações limitadas, podendo ser utilizado, por exemplo, para o estudo de gases nobres e moléculas esféricas (como o metano), sendo cada partícula tratada como uma esfera maciça. As forças de Coulomb (equação 2.5) são outro exemplo de potencial de pares; quando usadas em conjunto com as de Lennard-Jones, descrevem de forma adequada as interações intermoleculares em potenciais de muitos corpos.

$$E_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (2.4)$$

$$E_e = k_e \frac{q_1 q_2}{r} \quad (2.5)$$

A simulação de estruturas complexas, por outro lado, trata cada átomo individualmente, utilizando termos de interação intramolecular que capturam a geometria e a rigidez das vibrações das moléculas (JR et al., 1998a; VANOMMESLAEGHE et al.,

2010b; WEINER et al., 1984; SUN, 1998); esses são os chamados potenciais de muitos corpos. Para as ligações, uma aproximação harmônica é suficiente na maioria dos casos. Esse termo é muitas vezes chamado de 1-2, uma vez que modela a interação entre vizinhos imediatos (como ilustrado na figura 2.1). Também são necessários termos dependentes dos ângulos entre átomos não-ligados. Termos 1-3 impedem que dois átomos ligados a um terceiro se aproximem demais um do outro, enquanto termos 1-4 (diedros) capturam a energia torsional, representando a diferença entre conformações alternadas e eclipsadas. No caso de ligações que devem permanecer no mesmo plano (hibridização sp^2 , por exemplo), há ainda a necessidade de um termo chamado de diedro impróprio, que emprega uma função harmônica cujo mínimo de energia é a estrutura plana. Há algumas variações possíveis para a modelagem desses termos; a seção 2.4 traz um exemplo completo de potencial interatômico para moléculas flexíveis.

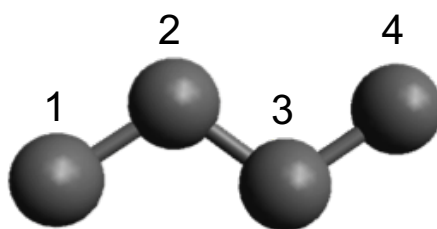


FIGURA 2.1. Numeração de átomos ligados, ilustrando os pares 1-2, 1-3 e 1-4.

Naturalmente, vizinhos próximos (1-2, 1-3 e, em alguns casos, 1-4) não exercem forças de Lennard-Jones e de Coulomb um sobre o outro. Todos os efeitos de suas interações são capturadas pelos termos de energia intramolecular. Esse é um dos motivos pelos quais todos os parâmetros de um potencial são otimizados simultaneamente, de maneira que, em seu conjunto, capturem a realidade vibracional das estruturas de interesse. Portanto, não é possível misturar parametrizações diferentes (a menos, é claro, que elas tenham sido calculadas de maneira propositalmente compatível).

Dentre os potenciais de muitos corpos, cabe citar também os potenciais reativos, como o AIREBO (STUART et al., 2000) e o ReaxFF (SENFLE et al., 2016), que permitem descrever a quebra e a criação de ligações químicas. Essas ferramentas não são tratadas em detalhes neste trabalho, mas é importante salientar que modelos re-

ativos também podem fazer um bom uso de ferramentas de análise estrutural como o ChemStruct, uma vez que os grupos funcionais presentes no sistema podem mudar durante a simulação, precisando ser identificadas a partir das novas coordenadas atômicas. Esses métodos de simulação são de grande interesse na indústria química, uma vez que podem fornecer dados importantes sobre processos reativos, como pirólise e combustão (LIU; GUO, 2017).

2.3 LAMMPS

Simuladores de DM trabalham com arquivos de diversos tipos. Uma vez que o pacote ChemStruct tem por objetivo secundário fornecer as informações atômicas necessárias no formato apropriado para simulação, é preciso escolher um software para esse fim, adequando as funcionalidades do algoritmo aos arquivos específicos do simulador. O pacote LAMMPS (*Large-scale Atomic/Molecular Massively Parallel Simulator*) é uma opção *open-source* amplamente utilizada por cientistas e engenheiros, uma vez que fornece grande flexibilidade e altos níveis de paralelização, possibilitando simulações de dinâmica molecular de grande tamanho e duração (PLIMPTON, 1995; PLIMPTON et al., 1995). Além disso, o LAMMPS é um simulador de ampla versatilidade, contando com diversos ensembles estatísticos, opções *coarse-grained* e *united-atom*, partículas de dipolo e rígidas, múltiplos modelos de polarização e equilíbrio de cargas, diversos potenciais de água e solvente implícito, criação e manipulação de átomos, simulações de Monte Carlo, criação de diferentes tipos de barreiras, dezenas de tipos de computações automáticas e arquivos de saída, entre outras funcionalidades. Esses fatores fazem do LAMMPS a escolha ideal para muitos estudos, como aqueles que envolvem simulações fora do equilíbrio e testes de deformação de nanomateriais (MUNIZ et al., 2015; MOULTOS et al., 2014; WU et al., 2014). Além das vantagens já citadas, o LAMMPS conta com um amplo manual online, onde podem ser encontradas todas as informações necessárias para a sua utilização (PLIMPTON, 1995).

No entanto, há um contraponto inerente à versatilidade do LAMMPS. O simulador permite que o usuário escolha dentre uma ampla gama de *constraints* e parâmetros, mas essas especificações devem ser fornecidas manualmente. Esse baixo grau de automaticidade se deve à grande liberdade garantida pelo LAMMPS. Por exemplo, o

simulador não faz a classificação atômica para todos os tipos de potenciais, uma vez que o usuário pode utilizar tipos customizados de átomos, com qualquer parametrização da literatura, sem que o LAMMPS tenha sido preparado especificamente para ela. Logo, muitas vezes é necessária uma ferramenta como o ChemStruct, que classifica os átomos presentes em um sistema atômico e gera um arquivo de entrada adequado, o LAMMPS Data File (um arquivo de texto ao qual se adiciona a extensão *.lmp* para clareza), que contém a lista e classificação de todas as estruturas topológicas presentes, como ligações, ângulos e diedros (ver seção 2.4 para a definição desses termos). Um exemplo completo de transformação de *.xyz* para *.lmp* é mostrado na seção 4.3, onde as figuras 4.6 e 4.7 ilustram, respectivamente, a entrada e a saída do ChemStruct para uma molécula de etano (onde fica evidente que a geração manual de um LAMMPS Data File para sistemas estendidos, com milhares de átomos, é inviável).

2.4 CGenFF: CHARMM General Force Field

CHARMM (*Chemistry at HARvard Macromolecular Mechanics*) pode referir-se tanto a um pacote para simulações MD quanto ao Potencial Interatômico por ele utilizado (JR et al., 1998a; JR et al., 1998b); no presente trabalho, a sigla CHARMM é empregada sempre como sinônimo do Potencial, que pode ser utilizado em diversos outros softwares, incluindo o já citado LAMMPS.

O CHARMM foi desenvolvido originalmente para a simulação de sistemas biológicos, como proteínas e carboidratos, mas os parâmetros presentes no CGenFF (CHARMM *General Force Field*) foram otimizados para modelar moléculas orgânicas em geral (VANOMMESLAEGHE et al., 2010b; VANOMMESLAEGHE et al., 2010a). Uma vez que o ChemStruct tem como objetivo secundário a geração de arquivos de MD, o CGenFF foi escolhido como fonte de parâmetros para as simulações. Versões futuras do pacote poderão incluir outras parametrizações, como AMBER (WEINER et al., 1984), COMPASS (WEINER et al., 1984; SUN, 1998) e OPLS (JORGENSEN et al., 1996), que, assim como o CHARMM, são Potenciais de muitos corpos capazes de capturar as interações Coulombicas e de van der Waals, bem como as de ligação, ângulo e torsão.

A energia potencial modelada segundo o CGenFF possui os termos indicados na equação 2.6. Todos eles aparecem dentro de um somatório porque são relativos a uma interação específica, havendo várias de cada tipo em um sistema atômico.

$$U = \sum E_{LJ} + \sum E_e + \sum E_r + \sum E_\theta + \sum E_\phi + \sum E_\chi \quad (2.6)$$

A cada um dos termos da equação 2.6 cabem alguns comentários. Os dois primeiros são relativos às interações intermoleculares, isto é, as energias de Lennard-Jones e de Coulomb, anteriormente mencionadas e descritas pelas equações 2.4 e 2.5, respectivamente.

O terceiro termo, E_r , é do tipo 1-2 e representa a energia de uma ligação química através de um potencial harmônico, que depende do quadrado da distância r entre os dois átomos, conforme a equação 2.7. O parâmetro r_0 é a distância de menor energia, enquanto k_r é o ganho de energia (por deslocamento ao quadrado) em relação à posição mais estável, relacionado diretamente à intensidade da força da ligação. Esse potencial simples descreve de forma adequada a energia em pequenas deformações.

$$E_r = k_r(r - r_0)^2 \quad (2.7)$$

O quarto termo, E_θ , é do tipo 1-3 e representa a energia entre ligações adjacentes, que depende do ângulo θ entre elas, conforme a equação 2.8, que contém dois termos quadráticos, às vezes necessários para uma boa representação do perfil de energia. Cada termo possui um mínimo próprio, bem como um ganho de energia (por deslocamento angular ao quadrado). Os parâmetros com subscrito UR são chamados de termos de Urey-Bradley, sendo iguais a zero nos casos em que um perfil harmônico simples é suficiente para capturar as características vibracionais angulares da estrutura.

$$E_\theta = k_\theta(\theta - \theta_0)^2 + k_{UB}(\theta - \theta_{UB})^2 \quad (2.8)$$

O quinto termo, E_ϕ , é do tipo 1-4 e representa a energia torsional de um diedro,

que depende do ângulo ϕ entre duas ligações separadas por uma terceira, conforme a equação 2.9. Esse termo do Potencial é uma função periódica de ϕ , uma vez que duas ligações não-adjacentes são livres para girar uma em relação à outra, havendo diversos mínimos e máximos de energia, relativos a conformações alternadas e eclipsadas, respectivamente. Podem ser necessários dois ou mais termos como o da equação 2.9 para capturar o perfil energético de uma torsão.

$$E_{\phi} = k_{\phi}(1 + \cos(n\phi - d)) \quad (2.9)$$

As variáveis r , θ e ϕ das equações acima estão ilustradas na figura 2.2, em que uma molécula de etano é utilizada para exemplificar as estruturas topológicas mais comuns.

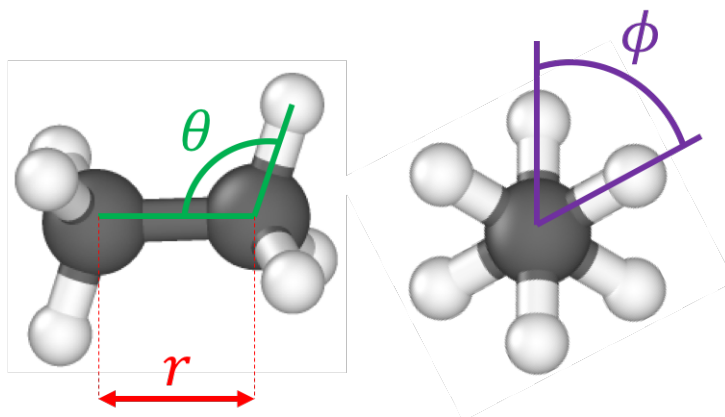


FIGURA 2.2. Molécula de etano, ilustrando geometria de ligação, ângulo e diedro próprio.

O sexto e último termo, E_{χ} , é a energia de um diedro impróprio, isto é, o potencial existente entre quatro átomos não alinhados, regido pela equação 2.10. A utilização mais comum desse termo são as hibridizações planas, como átomos de carbono sp^2 . Considerando um átomo central i e seus três ligantes j , k e l , a variável χ é o ângulo entre os planos jik e kil , por exemplo; isto é, E_{χ} é a energia necessária para tirar os átomos de sua configuração plana. Na maioria dos casos, portanto, tem-se o parâmetro $\chi_0 = 0$, uma vez que o mínimo de energia corresponde ao estado em que os

planos estão paralelos.

$$E_{\chi} = k_{\chi}(\chi - \chi_0)^2 \quad (2.10)$$

Uma vez que tenham sido identificadas e classificadas todas as estruturas presentes em um sistema atômico, os parâmetros apropriados podem ser obtidos de arquivos fornecidos pelos desenvolvedores do Potencial Interatômico, e simulações de MD podem ser realizadas. No entanto, a classificação dessas estruturas não é trivial, uma vez que depende não apenas do elemento, mas também da hibridização e da natureza funcional (cetona, amida, etc) do átomo. Para simulação de um sistema atômico no LAMMPS, é necessário fornecer uma lista contendo todas as suas ligações, ângulos e diedros, bem como suas classificações de acordo com o Potencial que se deseja utilizar. O pacote ChemStruct identifica e classifica todas essas estruturas, além de gerar o arquivo necessário para a simulação (o LAMMPS Data File). O CHARMM é um dos Potenciais que podem ser utilizados, e mesmo para essa alternativa há muitas parametrizações diferentes (JR et al., 1998b); a primeira versão do ChemStruct está automatizada para a geração de arquivos contendo parâmetros do CGenFF (VANOM-MESLAEGHE et al., 2010b).

Capítulo 3

Metodologia

3.1 Estrutura do Pacote

ChemStruct é um pacote de Python para análise e manipulação de estruturas atômicas (DEMINGOS, 2019). Logo, sua classe mais fundamental é *Atom*, que possui atributos como *position* para a posição do átomo, *type* para sua espécie atômica, entre outros. Cada átomo do sistema é uma instância dessa classe.

Adicionalmente, há informações que não pertencem a átomos em particular, como a caixa ou célula que define o sistema. Portanto, a segunda classe que fundamenta o pacote é *Atoms*, que representa o sistema atômico inteiro. Seus atributos mais importantes são *atoms*, uma lista dos objetos de tipo *Atom*, e *cell*, uma matriz 3x3 cujos vetores definem o sistema espacialmente (dimensões e formato da caixa – cúbica, ortorrômbica, triclínica, etc).

Um pacote bem estabelecido de Python que faz uso de uma estrutura parecida é o *Atomic Simulation Environment* (ASE), que possui recursos para simulações de dinâmica molecular (LARSEN et al., 2017a; LARSEN et al., 2017b) em simuladores equivalentes ao LAMMPS.

Além dos átomos e da célula, a classe *Atoms* do ChemStruct possui como atributos listas de ligações, ângulos e diedros próprios e impróprios, cujo conhecimento é necessário para simulações no LAMMPS. Uma vez que o objetivo secundário do pacote é gerar um arquivo com todas essas informações a partir das posições atômicas, boa parte do código tem como finalidade o cálculo dessas estruturas.

O método *compute_topology* da classe *Atoms* faz a análise estrutural completa, chamando outras partes do algoritmo para realizar funções específicas. A primeira etapa é o cálculo das ligações químicas; depois, é possível computar ciclos, ângulos, diedros, hibridizações e moléculas; por fim, com base nas informações anteriores, é realizada a classificação dos átomos, que é o objetivo primário do ChemStruct. Essas etapas estão detalhadas nas seções seguintes (3.2, 3.3 e 3.4).

Algumas funcionalidades do ChemStruct não fazem parte do cálculo topológico. Por exemplo, há métodos em *Atoms* para transladar e rotacionar o sistema, bem como para extendê-lo (duplicá-lo, triplicá-lo, etc) e para encher seus espaços vazios com uma molécula da escolha do usuário (de solvente, por exemplo). Todos esses métodos estão documentados no próprio código.

Há duas partes do pacote que não estão inclusas nas classes *Atom* e *Atoms*. Uma delas é o conjunto de classes que representam estruturas topológicas, como *Bond* para ligação, bem como o conjunto de classes para tipos topológicos, como *BondType* para um tipo específico de ligação; todos esses objetos estão definidos no arquivo *atoms.py*. Cada estrutura possui um atributo *type*, que retorna seu tipo topológico. Por exemplo, a ligação existente entre os átomos de número 101 e 102 de um dado sistema pode ser do tipo carbono-carbono; se a ligação entre os átomos 103 e 104 for do mesmo tipo, haverá dois objetos *Bond* com atributo *type* apontando para o mesmo objeto *BondType*. Na estrutura do ChemStruct, cada classe possui as informações relativas ao que representa: a ligação conhece os átomos que a compõem, enquanto seu tipo guarda os parâmetros para a simulação de dinâmica molecular.

A última parte do pacote são os códigos auxiliares que não estão presentes em *atoms.py*. Em *quick.py* estão as funções que procuram automatizar, na medida do possível, o processo para o usuário – elas estão documentadas em sua própria definição e são exemplificadas na pasta *examples*; também estão descritas na seção 3.5 do presente trabalho. Por fim, a pasta *files* possui classes voltadas a tipos específicos de arquivos, como *XYZ* para objetos de extensão *xyz* e *LmpDat* para LAMMPS Data Files.

3.2 Cálculo das Ligações Químicas

Dada a posição de cada átomo no sistema, encontrar as ligações químicas significa calcular a distância de cada par de átomos e verificar se ela é inferior a um valor de corte (que depende das espécies atômicas). Caso o sistema seja periódico, também é preciso verificar se átomos de lados diferentes da caixa estão ligados. Tudo isso é realizado pelos algoritmos implementados no ChemStruct, sem limitações quanto ao tamanho ou formato da célula (caixas triclinicas estão contempladas, bem como sistemas periódicos).

O grande problema desse método é que o tempo computacional de processamento aumenta com o quadrado do número de átomos, uma vez que é preciso verificar todos os pares de partículas. Caso o sistema seja pequeno ou o usuário especifique manualmente (utilizando o parâmetro *simple=True* nas funções), esse método simples e pouco otimizado será utilizado para o cálculo das ligações.

Há, no entanto, uma segunda opção. Caso o sistema seja grande, o método *compute_smart* irá dividi-lo em regiões menores para o cálculo das ligações, como ilustrado na figura 3.1. O código passa de região em região e, dentro de cada uma, de átomo em átomo. Para cada átomo, verifica se ele está ligado a outros da mesma região, bem como a átomos das regiões adjacentes (26 ao todo, nas três direções, incluindo diagonais). Esse algoritmo tem um tempo computacional diretamente proporcional ao número de átomos, sendo fortemente recomendado para sistemas grandes (da ordem de dez mil átomos ou mais). Uma comparação entre os dois tratamentos é mostrada na seção 4.1.

3.3 Outras Estruturas Topológicas

Um ângulo é formado entre quaisquer duas ligações que tenham um átomo em comum, e um diedro próprio é formado entre quaisquer dois ângulos que tenham dois átomos em comum. Esse cálculo exige, para cada tipo de estrutura topológica, a análise de todos os pares da estrutura computada anteriormente. Logo, assim como o método simples das ligações, esse processamento exige um tempo quadraticamente

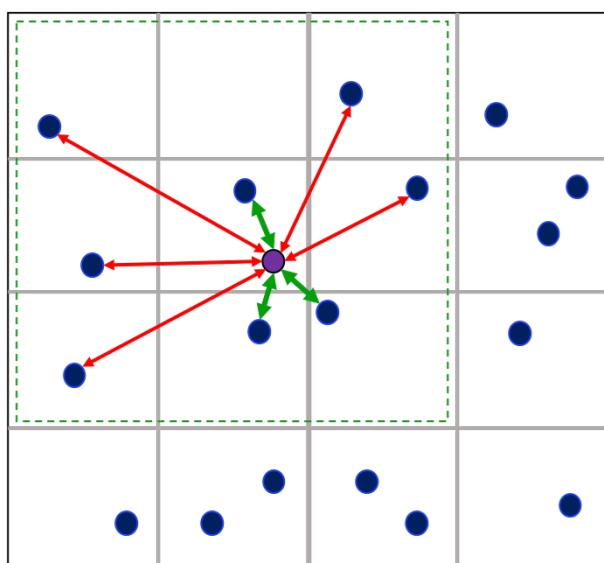


FIGURA 3.1. Divisão em regiões realizada para otimizar o cálculo das ligações. O ponto roxo é o átomo em análise; todos os outros estão representados em azul. A região tracejada de verde é o espaço dentro do qual os pares são verificados. As setas vermelhas representam pares não-ligados; as verdes, aqueles que estão perto o bastante para caracterizar uma ligação química.

proporcional ao tamanho do sistema.

Uma vez que o problema é o mesmo, é utilizada a mesma solução: às ligações e aos ângulos é dada uma posição aproximada, e a seguir eles são divididos em regiões do espaço, possibilitando que apenas os pares razoavelmente próximos sejam checados. Esse método otimizado do cálculo de outras estruturas topológicas exige um tempo de processamento linearmente proporcional ao tamanho do sistema.

Os diedros impróprios, por sua vez, são identificados depois da hibridização de cada átomo, uma vez que a sua existência depende de estruturas planas.

As hibridizações das ligações são deduzidas a partir do número de ligantes, uma vez que questões de ressonância não podem ser rigorosamente tratadas sem considerações energéticas. Estruturas cíclicas, depois de identificadas, são considerados aromáticos se todos os seus átomos tiverem hibridização sp^2 ; essa definição abrange todos os casos de interesse. Da mesma maneira, ligações são consideradas duplas quando ambos os átomos estão subcoordenados, isto é, exigem mais de uma ligação com algum de seus vizinhos; em casos evidentes de ressonância, todas as ligações envolvidas são tomadas como duplas por conveniência, e diferenciações são realizadas apenas na

classificação, quando forem necessárias para fins de parametrização de MD.

3.4 Classificação dos Átomos

Assim que todas as ligações tenham sido identificadas, é possível realizar a classificação dos átomos, o que é feito pelo método `compute_atom_classification` do objeto `Atoms`. A maior parte do processo é de uso geral, enquanto alguns detalhes refletem as especificações do CGenFF (como a identificação do íon AlF_4). Uma vez que o objeto `Atoms` é o fundamento do pacote, seus métodos foram implementados de maneira flexível, isto é, não há nada neles que limite as classificações a um potencial interatômico específico; situações de possível conflito estão implementadas no próprio objeto `CharmmGeneral` do arquivo `charmm.py`, onde alguns átomos são reclassificados conforme o CGenFF.

Na primeira etapa do método `compute_atom_classification` os carbonos são analisados em busca de grupos funcionais, e todos os átomos desses grupos são classificados. Isso é feito via análise da lista de vizinhos (isto é, outros átomos aos quais ele está ligado). Por exemplo, se um átomo de carbono está ligado a três átomos de oxigênio, ele é classificado como um *carbono de carbonato*, e cada um dos três oxigênios é classificado como um *oxigênio de carbonato*. A procura dos grupos funcionais se dá na seguinte ordem, sendo que cada item só é testado caso todos os anteriores tenham falhado:

Carbonato Carbono ligado a três oxigênios.

Guanidina Carbono ligado a três nitrogênios. Se nenhum dos nitrogênios é sp^2 , o grupo é um íon **guanidínio**.

Amidina Carbono ligado a dois nitrogênios. Se nenhum dos nitrogênios é sp^2 , o grupo é um íon **amidínio**.

Amida Carbono ligado a um nitrogênio e com ligação dupla com oxigênio. Se o carbono também estiver ligado a um hidrogênio, trata-se de **formamida**. O grau de substituição do nitrogênio também é considerado na subclassificação da amida.

Nitro Carbono com ligação dupla com nitrogênio, sendo o nitrogênio ligado a dois oxigênios.

Imina Carbono com ligação dupla com nitrogênio.

Cianeto Carbono ligado a um nitrogênio *sp*.

Amina Carbono ligado a um nitrogênio. O grau de substituição do nitrogênio também é considerado na subclassificação da amina.

Ácido Carboxílico Carbono ligado a dois oxigênios, tendo com um deles uma ligação dupla, e o outro oxigênio estando ligado a um hidrogênio. Se o carbono também estiver ligado a um hidrogênio, trata-se de **ácido fórmico**.

Éster Carbono ligado a dois oxigênios, tendo com um deles uma ligação dupla, e o outro oxigênio estando ligado a mais um carbono.

Carboxilato Carbono ligado a dois oxigênios. Se o carbono for *sp*, trata-se de **dióxido de carbono**.

Aldeído Carbono com ligação dupla com um oxigênio e ligação com um hidrogênio.

Cetona Carbono com ligação dupla com um oxigênio.

Hidroxila Carbono com ligação simples com um oxigênio, estando esse oxigênio ligado a um hidrogênio.

Éter Carbono com ligação simples com um oxigênio, estando esse oxigênio ligado a mais um carbono.

Haloalcano Carbono ligado a um ou mais átomos de flúor, cloro, bromo ou iodo. O número de ligações é levado em consideração na subclassificação dos átomos.

Hidrocarboneto Outros carbonos. São classificados entre *sp³*, *sp²* e *sp*. Além disso, são subclassificados de acordo com o número de hidrogênios aos quais estão ligados.

Na segunda etapa são analisados átomos de outros elementos, procurando estruturas na seguinte ordem:

Hidrogênio Molecular Hidrogênio ligado a outro hidrogênio.

Amônio Nitrogênio ligado a quatro hidrogênios.

Amônia Nitrogênio ligado a três hidrogênios.

Hidrazina Nitrogênio ligado a dois hidrogênios e um outro nitrogênio.

Água Oxigênio ligado a dois hidrogênios.

Sulfato Enxofre ligado a quatro oxigênios.

Dissulfeto Enxofre ligado a um outro enxofre.

Tioéter Enxofre ligado a dois carbonos.

Tiol Enxofre ligado a um hidrogênio.

Fosfato Fósforo ligado a quatro oxigênios. Se algum desses oxigênios estiver ligado a dois fósforos, classifica-se o grupo como **pirofosfato**.

Tetrafloroaluminato Alumínio ligado a quatro átomos de flúor (AlF_4).

Na terceira etapa são realizadas re-classificações em moléculas cíclicas (detalhes que visam apenas contemplar a estrutura fina do CGenFF estão omitidos):

Carbono-Ponte Carbono pertencente a dois ciclos diferentes, de maneira a ligá-los.

Ciclopropil Ciclo de 3 membros.

Ciclopentil Ciclo, aromático ou não, de 5 membros. A classificação dos átomos depende da aromaticidade e dos átomos ligantes.

Cicloheptil Ciclo aromático de 7 membros.

Anilina Nitrogênio ligado a um ciclo aromático de 6 membros.

Amida Aromática Grupo amida em anel aromático de 5 ou 6 membros.

Carboxila Aromática Grupo carboxila em anel aromático de 6 membros.

Aromático Carbono em anel aromático de 6 membros.

Furano Oxigênio em anel de 5 membros, ligado apenas a átomos sp^2 . Do contrário, trata-se apenas de um éter cíclico de 5 membros.

Pirano Oxigênio em anel de 6 membros, ligado apenas a átomos sp^2 . Do contrário, trata-se apenas de um éter cíclico de 6 membros.

Tiofeno Enxofre em anel aromático de 5 membros.

Uma vez realizadas as classificações, elas podem ser traduzidas para as especificações de qualquer potencial, a fim de escrever um arquivo para simulações no LAMMPS. Na primeira versão do ChemStruct só está implementado o CGenFF. O processo de tradução é relativamente simples, e a variável *FROM_CLASSIFICATION_TO_CGENFF*, definida no arquivo *charm.py* do pacote, é o dicionário utilizado para isso. A título de exemplo, são mostradas as classificações de carbonos alifáticos na tabela 3.1.

TABELA 3.1. Classificações de carbonos alifáticos segundo o ChemStruct e o CGenFF.

ChemStruct	CGenFF	Descrição
"sp C"	CG1T1	Carbono sp
"CH2 sp2 C"	CG2D2	Carbono sp^2 ligado a dois hidrogênios
"CHR sp2 C"	CG2D1	Carbono sp^2 ligado a um hidrogênio
"CRR' sp2 C"	CG2D1	Carbono sp^2 ligado e nenhum hidrogênio
"conjugated CH2 sp2 C"	CG2DC3	Carbono sp^2 conjugado, dois hidrogênios
"conjugated CHR sp2 C"	CG2DC2	Carbono sp^2 conjugado, um hidrogênio
"conjugated CRR' sp2 C"	CG2DC1	Carbono sp^2 conjugado, nenhum hidrogênio
"CH3 sp3 C"	CG331	Carbono sp^3 ligado a três hidrogênios
"CH2 sp3 C"	CG321	Carbono sp^3 ligado a dois hidrogênios
"CH sp3 C"	CG311	Carbono sp^3 ligado a um hidrogênio
"CC4 C"	CG301	Carbono sp^3 ligado a nenhum hidrogênio

Os arquivos gerados com os parâmetros do Potencial foram utilizados com sucesso em simulações MD no LAMMPS, fornecendo resultados consistentes (a seção 4.3 exemplifica essa funcionalidade).

3.5 Funções para Utilização do Usuário

O arquivo *quick.py* do ChemStruct possui duas funções especiais para utilização por parte dos usuários. Ambas estão documentadas no próprio código, e essa documentação é mostrada aqui em português.

write_impdat Escreve um LAMMPS Data File com *atom_style='full'* (especificação do LAMMPS). O parâmetro *xyz_file* é o caminho para o arquivo de coordenadas atômicas em formato *.xyz* que deve ser convertido em LAMMPS Data File, contendo também os vetores que definem a caixa de simulação (o software Ovito, por exemplo, escreve arquivos *.xyz* no formato apropriado). O parâmetro *impdat_file* é o caminho para o LAMMPS Data File que deve ser escrito. O parâmetro opcional *par_file* é o caminho para o arquivo contendo os parâmetros para simulações de dinâmica molecular (arquivo *.par*), caso haja algum; se nenhum for dado, um arquivo do formato certo será escrito pela própria função. O parâmetro opcional *periodic* informa em quais dimensões o sistema é periódico (por exemplo, *'xyz'*, *'xy'* ou *'z'*; ainda, usar *' '* para um sistema não periódico); se esse parâmetro não for dado, trata-se o sistema como não periódico. O parâmetro opcional *simple* informa se o método utilizado para calcular as ligações, ângulos e diedros deve ser o algoritmo simples, menos otimizado, porém mais robusto para sistemas pequenos, onde deve ser usado; se não for informado, é tomado como *False*, e o próprio código decide qual algoritmo usar (ver seção 3.2 para a descrição dos dois algoritmos).

finish_impdat Escreve um LAMMPS Data File com *atom_style='full'* (especificação do LAMMPS) de um arquivo *.xyz* com os átomos já classificados pela função *write_impdat*. Pode ser usada quando a função *write_impdat* é levada até o fim sem que o arquivo de parâmetros gerado tenha sido corretamente editado pelo usuário, ou se já há arquivos *.xyz* e *.par* com as classificações e os parâmetros desejados. Os parâmetros são os mesmos que aqueles da função *write_impdat*, com a diferença de que, aqui, o *par_file* não é facultativo.

Além disso, a pasta *examples* traz exemplos da utilização do pacote. O *ex01.py*

mostra como utilizar a função *write_impdat*; o *ex02.py* ilustra o uso de *finish_impdat*; e o *ex03.py* expande a utilidade de *write_impdat* para um sistema periódico. A seção 4.3 do presente trabalho traz um exemplo de utilização do pacote com os arquivos de entrada (.xyz) e saída (LAMMPS Data File).

Mais exemplos podem ser adicionados nas versões futuras, e dúvidas podem estar respondidas no link: <<https://github.com/pdemingos/chemstruct/>>.

Capítulo 4

Resultados e Discussão

4.1 Eficiência do Cálculo das Ligações Químicas

Como mencionado na seção 3.2, o ChemStruct conta com dois métodos para o cálculo das ligações químicas. Quando há um número pequeno de átomos, todos os pares são verificados, mas sistemas grandes são previamente divididos em regiões, otimizando o tempo de processamento. Com base nos algoritmos implementados, é razoável prever que o custo computacional do primeiro método aumenta com o quadrado do número de átomos, enquanto o segundo tem uma relação linear. A figura 4.1 exemplifica essas relações, destacando a vantagem de se utilizar o algoritmo otimizado para analisar sistemas muito grandes.

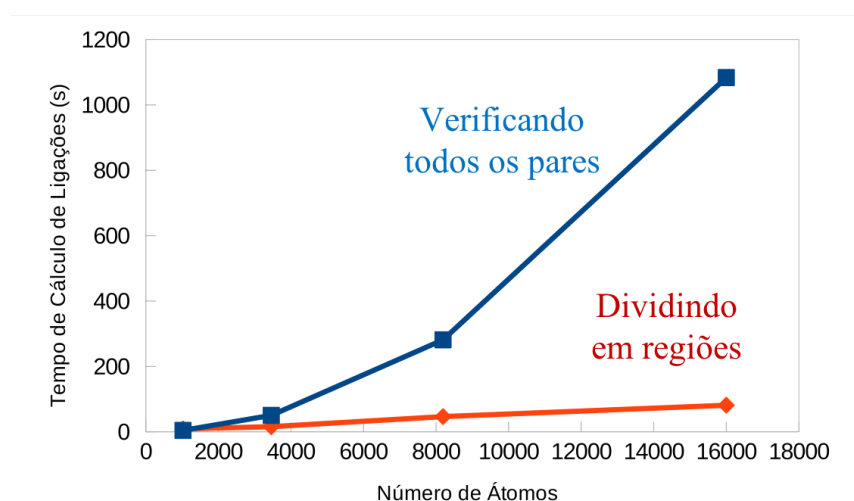


FIGURA 4.1. Eficiência comparada dos dois métodos de cálculo das ligações.

4.2 Validação da Classificação dos Átomos

Resultados contendo classificações do CGenFF foram comparados com arquivos gerados por uma ferramenta online desenvolvida pelos criadores do CGenFF, para fins de validação. Esse classificador online possui diversas limitações, como um limite máximo de 384 átomos no sistema e a exigência de algum pré-processamento; no entanto, pode ser utilizado em paralelo ao ChemStruct com sistemas reduzidos, com a finalidade de garantir que nenhuma das classificações realizadas pelo pacote extrapola os limites do Potencial.

As figuras 4.2, 4.3, 4.4 e 4.5 ilustram alguns dos exemplos utilizados para a verificação; essas e todas as outras estruturas de validação estão na pasta *files/charmm_proof* do pacote ChemStruct. A função *charmm_proof* definida no arquivo *charmm.py* pode ser utilizada para revalidar o algoritmo sempre que necessário, analisando os arquivos da pasta de mesmo nome e comparando os resultados com a classificação correta, informando o usuário sobre quaisquer inadequações.

Como mencionado na seção 3.4, o ChemStruct classifica hidrocarbonetos (Fig. 4.2), ácidos carboxílicos (Fig.4.3), aldeídos (Fig.4.4), moléculas aromáticas (Fig.4.5), além de muitos outros grupos funcionais, a maioria dos quais estão exemplificados nos arquivos da pasta *files/charmm_proof*.

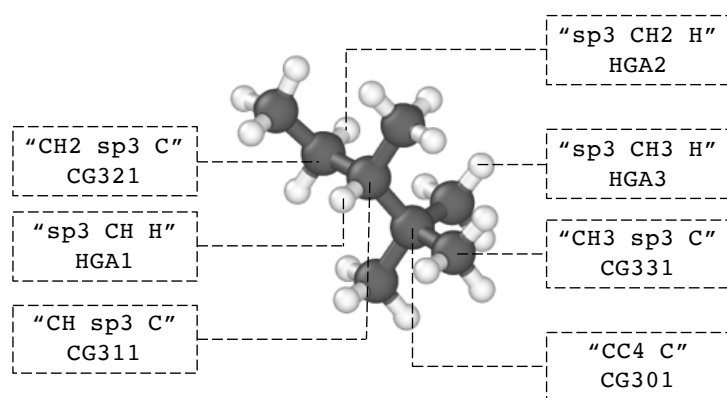


FIGURA 4.2. Hidrocarboneto com classificação atômica. Átomos classificados segundo o ChemStruct (entre aspas) e o CGenFF.

Cabem comentários sobre as moléculas apresentadas acima, visando explicar

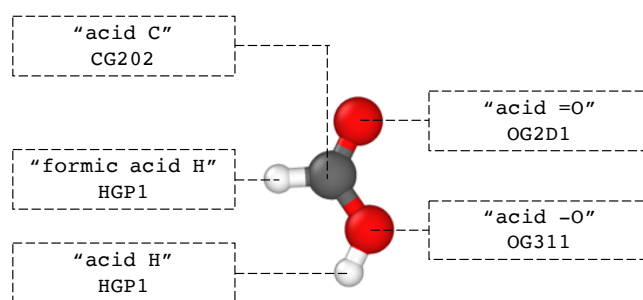


FIGURA 4.3. Ácido carboxílico com classificação atômica. Átomos classificados segundo o ChemStruct (entre aspas) e o CGenFF.

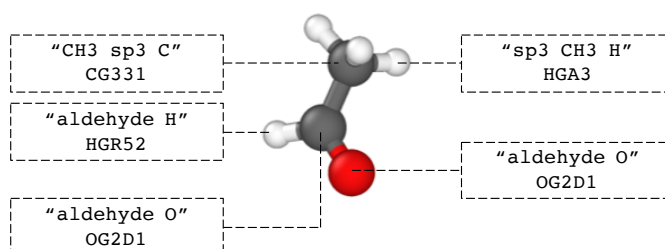


FIGURA 4.4. Aldeído com classificação atômica. Átomos classificados segundo o ChemStruct (entre aspas) e o CGenFF.

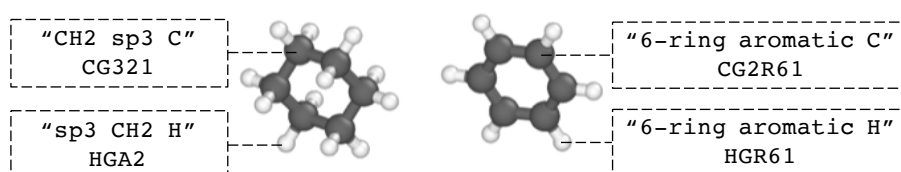


FIGURA 4.5. Moléculas cíclicas com classificação atômica. Átomos classificados segundo o ChemStruct (entre aspas) e o CGenFF.

alguns detalhes sobre as classificações do pacote e do CGenFF. As classes de átomos de carbono indicados para hidrocarbonetos (figura 4.2) são algumas das apresentadas na tabela 3.1. Na análise do ácido fórmico (figura 4.3), o ChemStruct tem um tipo atômico específico para o hidrogênio ligado ao carbono (*formic acid H*), enquanto o CGenFF classifica tanto esse átomo como aquele presente no grupo funcional como hidrogênios polares (*HGP1*). Por outro lado, tanto o pacote quanto o Potencial têm uma classificação à parte para um hidrogênio de aldeído (na figura 4.4, *aldehyde H* e *HGR52*,

respectivamente). Nas moléculas cíclicas (figura 4.5), os átomos aromáticos possuem tipos específicos, enquanto os átomos de carbono sp^3 e de hidrogênio ligados a eles são tratados independentemente da existência do ciclo, uma vez que possuem comportamentos geométricos e vibracionais semelhantes aos átomos de estruturas lineares; no entanto, o ChemStruct identifica a existência do anel de seis membros em ambos os compostos, sendo fácil a implementação de funções para análises adicionais.

4.3 Utilização do Pacote para Auxílio em Simulações de Dinâmica Molecular

Um dos principais objetivos do ChemStruct é a geração de arquivos para simulações MD no LAMMPS. Para ilustrar essa funcionalidade, esta seção apresenta a entrada (arquivo *.xyz*) e a saída (LAMMPS Data File, *.imp*) para uma molécula de etano, bem como para uma simulação MD contendo N moléculas desta espécie em fase líquida (a uma temperatura coerente com esse estado físico). A entrada é fornecida à função *write_impdat* (descrita na seção 3.5), e não é necessária nenhuma edição do arquivo contendo os parâmetros da simulação de dinâmica molecular.

Como pode ser visto na figura 4.6, o arquivo de entrada possui apenas os elementos e as posições atômicas (além da caixa que contém a molécula, que é um cubo de 4Å de aresta, informação que é relevante nos casos em que o sistema é periódico, como uma estrutura cristalina ou nanoestrutura). As colunas apresentadas correspondem ao elemento e às coordenadas do átomo em x , y e z , respectivamente. A caixa está indicada na linha da *Lattice*, na forma usual de uma matriz 3x3.

A saída, mostrada na figura 4.7, é o LAMMPS Data File gerado a partir das coordenadas atômicas pela função *write_impdat*, e possui todos os dados necessários para a simulação no LAMMPS. Primeiramente está apresentada a contagem de cada um dos itens que serão detalhados a seguir (isto é, as estruturas e seus tipos, discutidos abaixo no texto). Depois é mostrada a caixa na qual o sistema está contido; no LAMMPS, isso é feito através dos pontos que delimitam o sistema nas três direções, calculados pelo pacote de maneira que o átomo mais abaixo (nos três eixos) coincida com a origem. A seguir, em *Masses*, são listadas as massas dos átomos presentes, retiradas de


```
8
Lattice="4.0 0.0 0.0 0.0 4.0 0.0 0.0 0.0 4.0"
C      -4.82269      -0.70304      -0.00000
C      -3.37213      -0.25125       0.00000
H      -5.44446       0.02666      -0.56025
H      -5.19551      -0.77264       1.04360
H      -4.90695      -1.69913      -0.48334
H      -3.28788       0.74485       0.48334
H      -2.99932      -0.18164      -1.04360
H      -2.75037      -0.98094       0.56025
```

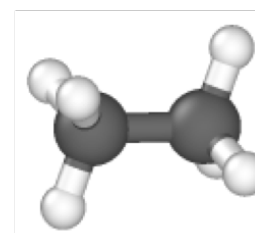


FIGURA 4.6. Arquivo *.xyz* de molécula de etano. À esquerda, o arquivo de texto em formato *.xyz* para a molécula de etano. À direita, a sua representação espacial, observada em software de visualização gráfica.

um arquivo do pacote.

O próximo item é uma lista dos oito átomos presentes na molécula de etano, cujas posições e elementos são replicados do arquivo de entrada. A primeira coluna é a numeração dos átomos. A segunda é a numeração das moléculas às quais os átomos pertencem; neste caso há apenas uma, mas, havendo mais, o ChemStruct as identifica e numera automaticamente. A terceira coluna é o tipo dos átomos (carbonos *CG331* e hidrogênios *HGA3*, conforme a numeração mostrada em *Masses*; a classificação deste carbono foi descrita na tabela 3.1, referindo-se a um átomo de C alifático ligado a três hidrogênios). A quarta coluna corresponde à carga de cada átomo, cujos valores são obtidos de padrões do CGenFF ou fornecidos pelo usuário durante a pausa da função *write_impdat*. As últimas três colunas são as coordenadas de cada átomo. Tudo o que está escrito depois do símbolo # é um comentário, ignorado pelo LAMMPS (aqui a função escreve os tipos atômicos por extenso, para clareza).

Os próximos itens contêm as ligações, ângulos e diedros presentes na molécula (*Bonds*, *Angles* e *Dihedrals*, respectivamente). Em cada item, a primeira coluna é a numeração, a segunda informa o tipo de estrutura (uma ligação C-C ou C-H, por exemplo), e as outras numeram os átomos envolvidos (dois nas ligações, três nos ângulos, e quatro nos diedros). Uma vez que os carbonos do etano têm hibridização sp^3 , não há diedros impróprios (como informado no topo do arquivo, na linha que diz: *0 improper*).

Os próximos quatro itens contêm os parâmetros de MD relativos às estruturas que o ChemStruct encontrou na molécula. Os *Pair Coeffs*, ou coeficientes de pares,

```

LAMMPS data file
8 atoms
7 bonds
12 angles
18 dihedrals
0 impropers

2 atom types
2 bond types
2 angle types
2 dihedral types
0 improper types

-5.44446 -1.44446 xlo xhi
-1.69913 2.30087 ylo yhi
-1.0436 2.9564 zlo zhi

Masses
1 12.011 # CG331
2 1.008 # HGA3

Atoms
1 1 1 -0.27 -4.82269 -0.70304 -0.0 # CG331
2 1 1 -0.27 -3.37213 -0.25125 0.0 # CG331
3 1 2 0.09 -5.44446 0.02666 -0.56025 # HGA3
4 1 2 0.09 -5.19551 -0.77264 1.0436 # HGA3
5 1 2 0.09 -4.90695 -1.69913 -0.48334 # HGA3
6 1 2 0.09 -3.28788 0.74485 0.48334 # HGA3
7 1 2 0.09 -2.99932 -0.18164 -1.0436 # HGA3
8 1 2 0.09 -2.75037 -0.98094 0.56025 # HGA3

Bonds
1 1 1 2
2 2 1 3
3 2 1 4
4 2 1 5
5 2 2 6
6 2 2 7
7 2 2 8

Angles
1 1 3 1 2
2 1 4 1 2
3 1 5 1 2
4 1 6 2 1
5 1 7 2 1
6 1 8 2 1

7 2 4 1 3
8 2 5 1 3
9 2 5 1 4
10 2 6 2 7
11 2 6 2 8
12 2 8 2 7

Dihedrals
1 1 3 1 2 6
2 2 3 1 2 6
3 1 3 1 2 7
4 2 3 1 2 7
5 1 3 1 2 8
6 2 3 1 2 8
7 1 4 1 2 6
8 2 4 1 2 6
9 1 4 1 2 7
10 2 4 1 2 7
11 1 4 1 2 8
12 2 4 1 2 8
13 1 5 1 2 6
14 2 5 1 2 6
15 1 5 1 2 7
16 2 5 1 2 7
17 1 5 1 2 8
18 2 5 1 2 8

Pair Coeffs
1 0.078 3.65268 0.01 3.38542 # CG331
2 0.024 2.38761 0.024 2.38761 # HGA3

Bond Coeffs
1 222.5 1.53 # CG331:CG331
2 322.0 1.111 # CG331:HGA3

Angle Coeffs
1 37.5 110.1 22.53 2.179 # HGA3:CG331:CG331
2 35.5 108.4 5.4 1.802 # HGA3:CG331:HGA3

Dihedral Coeffs
1 0.155 3 0 1.0 # HGA3:CG331:CG331:HGA3
2 0.155 3 0 1.0 # HGA3:CG331:CG331:HGA3

```

FIGURA 4.7. Arquivo de texto em formato LAMMPS Data File (.lmp) para uma molécula de etano, gerado pelo ChemStruct.

são relativos às interações de Lennard-Jones. Os outros itens trazem os parâmetros para ligações, ângulos e diedros (*Bond Coeffs*, *Angle Coeffs* e *Dihedral Coeffs*, respectivamente). Esses são os coeficientes apresentados na seção 2.4, que são obtidos do arquivo *par_all36_cgenff.prm*, fornecido pelos desenvolvedores do CGenFF (VANOMMESLA-EGHE et al., 2010b) e presente na distribuição do ChemStruct (a sua ordenação está de acordo com o padrão imposto pelo LAMMPS, que pode ser estudado no manual do simulador (PLIMPTON et al., 1995)). Nesses itens, os comentários mostram os tipos estruturais de cada conjunto de parâmetros. Uma rápida inspeção da molécula de

etano (figura 4.6) mostrará ao leitor que todos os tipos topológicos presentes nela estão contemplados no LAMMPS Data File mostrado na figura 4.7. A geração desse arquivo não exige nada além da utilização da função *write_impdat* e da existência da entrada em formato *.xyz*.

O arquivo da figura 4.7 é simples o bastante para ser escrito manualmente. No entanto, a maior parte das simulações MD utilizam milhares (ou até milhões) de átomos, o que exige a automação da análise estrutural. Portanto, para demonstrar uma utilização mais extensiva do ChemStruct, foi feita no LAMMPS a simulação do etano a 184 K, temperatura em que essa substância se encontra no estado líquido. Para este caso, deve-se empregar um número N de moléculas que represente o comportamento de um sistema infinito (neste caso, $N = 125$). O LAMMPS Data File utilizado para isso foi gerado pelo ChemStruct, partindo-se do arquivo original (figura 4.6) e fazendo uso dos seguintes comandos:

from files.xyz import Xyz Há uma pasta chamada *files* no ChemStruct, e dentro dela há um arquivo *xyz.py*, onde está definida a classe *Xyz* para utilização de arquivos nesse formato. Esta linha de comando importa essa classe.

*ethane = Xyz(*path to input xyz file*)* Esta linha de comando instancia (cria) um objeto da classe *Xyz* para ler o arquivo contendo as coordenadas dos átomos da molécula de etano. Um dos atributos do objeto *Xyz* é um objeto *Atoms*, onde são guardadas essas informações atômicas.

bulk = ethane.atoms.extended(5,5,5) Esta linha utiliza um dos métodos do objeto *Atoms* acessado pelo atributo *ethane.atoms*: o método *extended(nx,ny,nz)*, que retorna um novo objeto *Atoms* com o sistema replicado $nx \times ny \times nz$ vezes. Esse líquido, contendo 125 moléculas de etano, é então guardado na variável *bulk*. (Nesta etapa é necessária a definição da *Lattice* citada anteriormente: o cubo de aresta 4 dá origem a um sistema de $20 \times 20 \times 20$ angstroms.)

*bulk.write_xyz(*path to output xyz file*)* A este ponto, o objeto *bulk*, já definido, poderia ser utilizado manualmente para a produção do LAMMPS Data File. No entanto, a função *write_impdat* faz isso automaticamente, exigindo apenas um arquivo *.xyz*. Esta linha de comando o escreve no formato certo.

from quick import write_impdat Importa a função *write_impdat*.

*write_impdat(*path to bulk xyz file*, *path to output imp file*)* Esta linha escreve um LAMMPS Data File (*.imp*) a partir do arquivo com as coordenadas (*.xyz*) das 125 moléculas de etano. A função faz sua pausa usual para edição dos parâmetros, mas neste caso não é necessária nenhuma alteração manual por parte do usuário.

O arquivo de saída está pronto para simulação no LAMMPS. Esse LAMMPS Data File possui 1000 átomos, 875 ligações químicas corretamente computadas, 1500 ângulos e 1125 diedros. A representação espacial atômica está mostrada na figura 4.8, bem como a posição dos átomos durante a simulação de dinâmica molecular realizada pelo pacote LAMMPS (observe-se que o sistema é periódico, isto é, átomos que saem da caixa pela direita, por exemplo, retornam pela esquerda). Especificando uma temperatura de 184 K e pressão atmosférica, o simulador conduziu o líquido para uma densidade de aproximadamente 500 kg m^{-3} , próximo do valor experimental (a precisão pode ser aumentada utilizando-se um sistema maior e/ou realizando simulações mais longas).

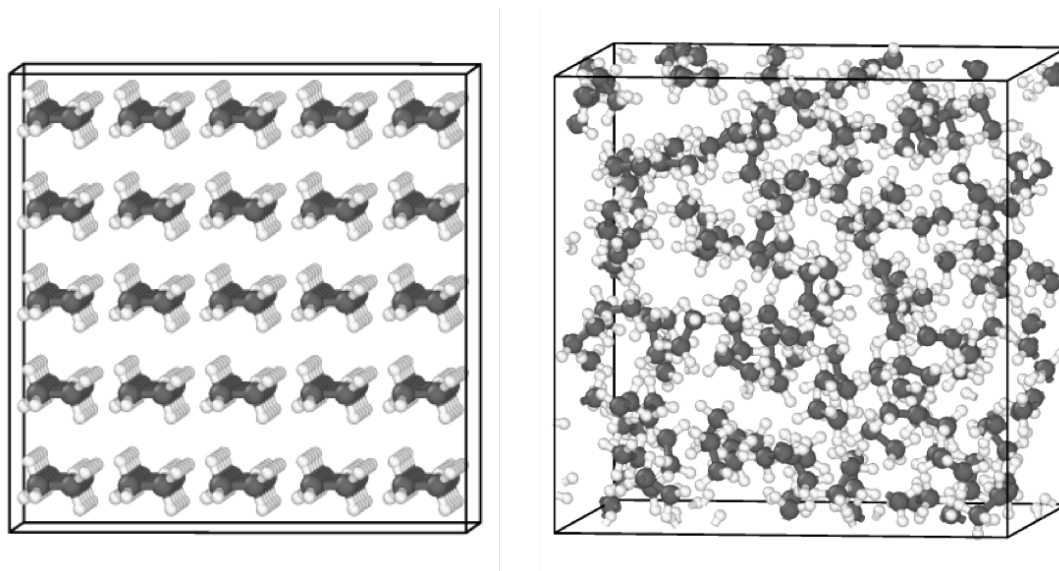


FIGURA 4.8. Sistema com 125 moléculas de etano. À esquerda, a visualização do arquivo contendo 125 moléculas, gerado pelo ChemStruct. À direita, as coordenadas desses átomos durante simulação de dinâmica molecular (observe-se que o sistema é periódico).

O ChemStruct pode ser utilizado para analisar e gerar arquivos com qualquer número de átomos (sendo limitado apenas pelo custo computacional), incluindo sistemas em diferentes fases e com diversos tipos de moléculas, como solventes, polímeros, ou qualquer estruturas cujos grupos funcionais estejam contemplados na lista da seção 3.4.

Capítulo 5

Conclusões e Trabalhos Futuros

O presente trabalho relatou a implementação do ChemStruct, um pacote em Python para análise estrutural de sistemas atômicos e auxílio em simulações de dinâmica molecular. A estrutura do código foi apresentada, dando-se ênfase a suas partes principais. Dois métodos para o cálculo de ligações químicas foram descritos e comparados, verificando-se a capacidade do algoritmo de lidar com sistemas grandes sem exigir tempos inviáveis de processamento. A ferramenta classificou corretamente átomos de moléculas orgânicas e escreveu arquivos operantes para simulações de dinâmica molecular.

O ChemStruct pode ser utilizado como ferramenta de pré-processamento para simulações de dinâmica molecular, como analisador em larga escala de arquivos de coordenadas atômicas, e também como instrumento didático. O pacote permite uma melhor compreensão das características geométricas e estruturais de moléculas de diversos tipos, facilitando a predição teórica de suas propriedades e de seu comportamento em aplicações de interesse científico e industrial.

O ChemStruct continua em desenvolvimento, e sua versão mais recente está disponível no link <<https://github.com/pdmingos/chemstruct/>>. Dentre as expansões planejadas para as próximas atualizações, destacam-se:

- Implementação de novas ferramentas de análise estrutural, como classificadores de moléculas.
- Interface com outros Potenciais Interatômicos para dinâmica molecular, como AMBER, COMPASS e OPLS.

Referências Bibliográficas

- DEMINGOS, P. *ChemStruct: A Python Package for Structural Analysis of Atomic Systems*. 2019. Disponível em: <<https://github.com/pdemingos/chemstruct>>.
- DEMINGOS, P. G.; MUNIZ, A. R. Carbon nanothreads from polycyclic aromatic hydrocarbon molecules. *Carbon*, Elsevier, v. 140, p. 644–652, 2018.
- FRENKEL, D.; SMIT, B. *Understanding Molecular Simulation*. [S.l.]: Academic Press, 2002.
- HUNTER, J.; DALE, D.; FIRING, E.; DROETTBOOM, M. *Matplotlib*. 2002. Disponível em: <<https://matplotlib.org/>>.
- JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, ACS Publications, v. 118, n. 45, p. 11225–11236, 1996.
- JR, A. D. M.; BASHFORD, D.; BELLOTT, M.; JR, R. L. D.; EVANSECK, J. D.; FIELD, M. J.; FISCHER, S.; GAO, J.; GUO, H.; HA, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, ACS Publications, v. 102, n. 18, p. 3586–3616, 1998.
- JR, A. D. M.; BASHFORD, D.; BELLOTT, M.; JR, R. L. D.; EVANSECK, J. D.; FIELD, M. J.; FISCHER, S.; GAO, J.; GUO, H.; HA, S. et al. *CHARMM: Chemistry at Harvard Macromolecular Mechanics*. 1998. Disponível em: <<https://www.charmm.org/charmm/?CFID=879393f9-8340-4a9e-977f-0acb6ac05018&CFTOKEN=0>>.
- LARSEN, A. H.; MORTENSEN, J. J.; BLOMQVIST, J.; CASTELLI, I. E.; CHRISTENSEN, R.; DUŁAK, M.; FRIIS, J.; GROVES, M. N.; HAMMER, B.; HARGUS, C. et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, IOP Publishing, v. 29, n. 27, p. 273002, 2017.
- LARSEN, A. H.; MORTENSEN, J. J.; BLOMQVIST, J.; CASTELLI, I. E.; CHRISTENSEN, R.; DUŁAK, M.; FRIIS, J.; GROVES, M. N.; HAMMER, B.; HARGUS, C. et al. *Atomistic Simulation Environment*. 2017. Disponível em: <<https://wiki.fysik.dtu.dk/ase/>>.
- LIU, J.; GUO, X. Reaxff molecular dynamics simulation of pyrolysis and combustion of pyridine. *Fuel Processing Technology*, Elsevier, v. 161, p. 107–115, 2017.

- MOULTOS, O. A.; TSIMPANOIANNIS, I. N.; PANAGIOTOPOULOS, A. Z.; ECONOMOU, I. G. Atomistic molecular dynamics simulations of co2 diffusivity in h2o for a wide range of temperatures and pressures. *The Journal of Physical Chemistry B*, ACS Publications, v. 118, n. 20, p. 5532–5541, 2014.
- MUNIZ, A. R.; MACHADO, A. S.; MAROUDAS, D. Mechanical behavior of interlayer-bonded nanostructures obtained from bilayer graphene. *Carbon*, Elsevier, v. 81, p. 663–677, 2015.
- NUMPY-TEAM. *NumPy*. 2019. Disponível em: <<https://www.numpy.org/>>.
- PLIMPTON, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, Elsevier, v. 117, n. 1, p. 1–19, 1995.
- PLIMPTON, S.; THOMPSON, A.; MOORE, S.; KOHLMAYER, A.; BERGER, R. *LAMMPS: Large-scale Atomic/Molecular Massively Parallel Simulator*. 1995. Disponível em: <<https://lammps.sandia.gov/>>.
- PYTHON-SOFTWARE-FOUNDATION. *Python*. 2001. Disponível em: <<https://www.python.org/>>.
- RAPAPORT, D. C. *The Art of Molecular Dynamics Simulation*. [S.l.]: Cambridge University Press, 2004.
- SCIPY-TEAM. *SciPy*. 2019. Disponível em: <<https://www.scipy.org/>>.
- SENFLE, T. P.; HONG, S.; ISLAM, M. M.; KYLASA, S. B.; ZHENG, Y.; SHIN, Y. K.; JUNKERMEIER, C.; ENGEL-HERBERT, R.; JANIK, M. J.; AKTULGA, H. M. et al. The reaxff reactive force-field: development, applications and future directions. *npj Computational Materials*, Nature Publishing Group, v. 2, p. 15011, 2016.
- STUART, S. J.; TUTEIN, A. B.; HARRISON, J. A. A reactive potential for hydrocarbons with intermolecular interactions. *The Journal of chemical physics*, AIP, v. 112, n. 14, p. 6472–6486, 2000.
- SUN, H. Compass: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *The Journal of Physical Chemistry B*, ACS Publications, v. 102, n. 38, p. 7338–7364, 1998.
- VANOMMESLAEGHE, K.; HATCHER, E.; ACHARYA, C.; KUNDU, S.; ZHONG, S.; SHIM, J.; DARIAN, E.; GUVENCH, O.; LOPES, P.; VOROBYOV, I. et al. *CGenFF: CHARMM General Force Field*. 2010. Disponível em: <<https://cgenff.umaryland.edu/>>.
- VANOMMESLAEGHE, K.; HATCHER, E.; ACHARYA, C.; KUNDU, S.; ZHONG, S.; SHIM, J.; DARIAN, E.; GUVENCH, O.; LOPES, P.; VOROBYOV, I. et al. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, Wiley Online Library, v. 31, n. 4, p. 671–690, 2010.

- WEINER, S. J.; KOLLMAN, P. A.; CASE, D. A.; SINGH, U. C.; GHIO, C.; ALAGONA, G.; PROFETA, S.; WEINER, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, ACS Publications, v. 106, n. 3, p. 765–784, 1984.
- WU, T.; XUE, Q.; LING, C.; SHAN, M.; LIU, Z.; TAO, Y.; LI, X. Fluorine-modified porous graphene as membrane for co₂/n₂ separation: molecular dynamic and first-principles simulations. *The Journal of Physical Chemistry C*, ACS Publications, v. 118, n. 14, p. 7369–7376, 2014.