

UNESP/IQ/AQA

Programa de Pós-Graduação em Biotecnologia

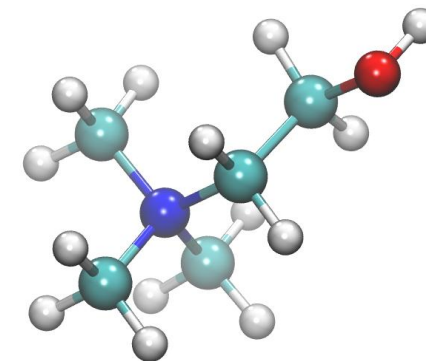
Simulação de Biomoléculas: dinâmica molecular básica

Patrick Allan dos Santos Faustino

Prof. Dr. Fernando Masarin

Prof. Dr. Elias de Souza Monteiro Filho

04 de setembro de 2025



Dinâmica molecular (MD)

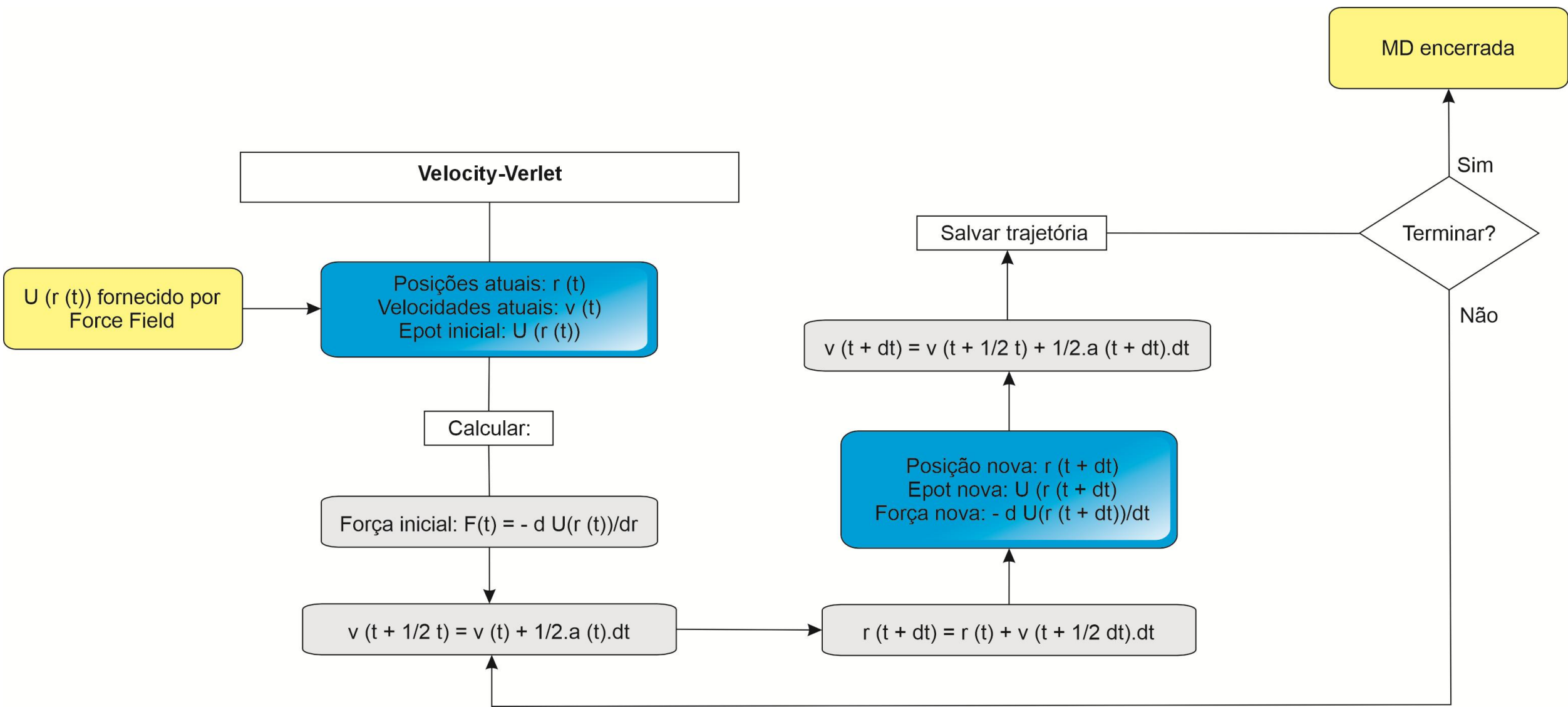
- A dinâmica molecular é uma ferramenta extremamente poderosa por diversas razões: versátil, é possível ajustar parâmetros e meios de simulação (1).
- Método **PREDITIVO**; ‘Caixa de ferramenta’; *in silico*.
 - Mudança conformacional, ligação de ligantes, dobramento de proteínas, protonação, umbrella sampling etc.
- Método baseado na aplicação da 2ª Lei de Newton, Lei de Coulomb e Lei de Hooke. Robusto.
- Softwares: GROMACS, NAMD, LAMMPS, OPENMM etc.

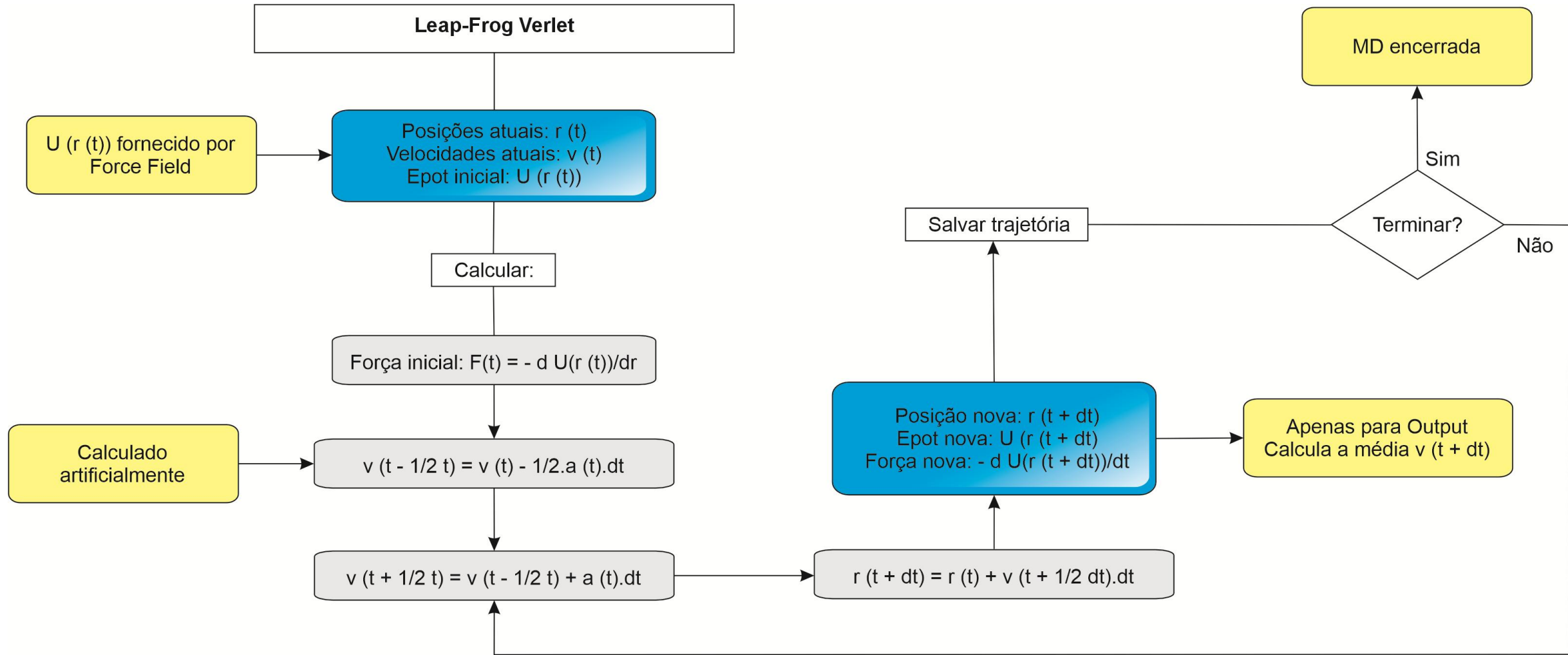
Dinâmica molecular (MD)

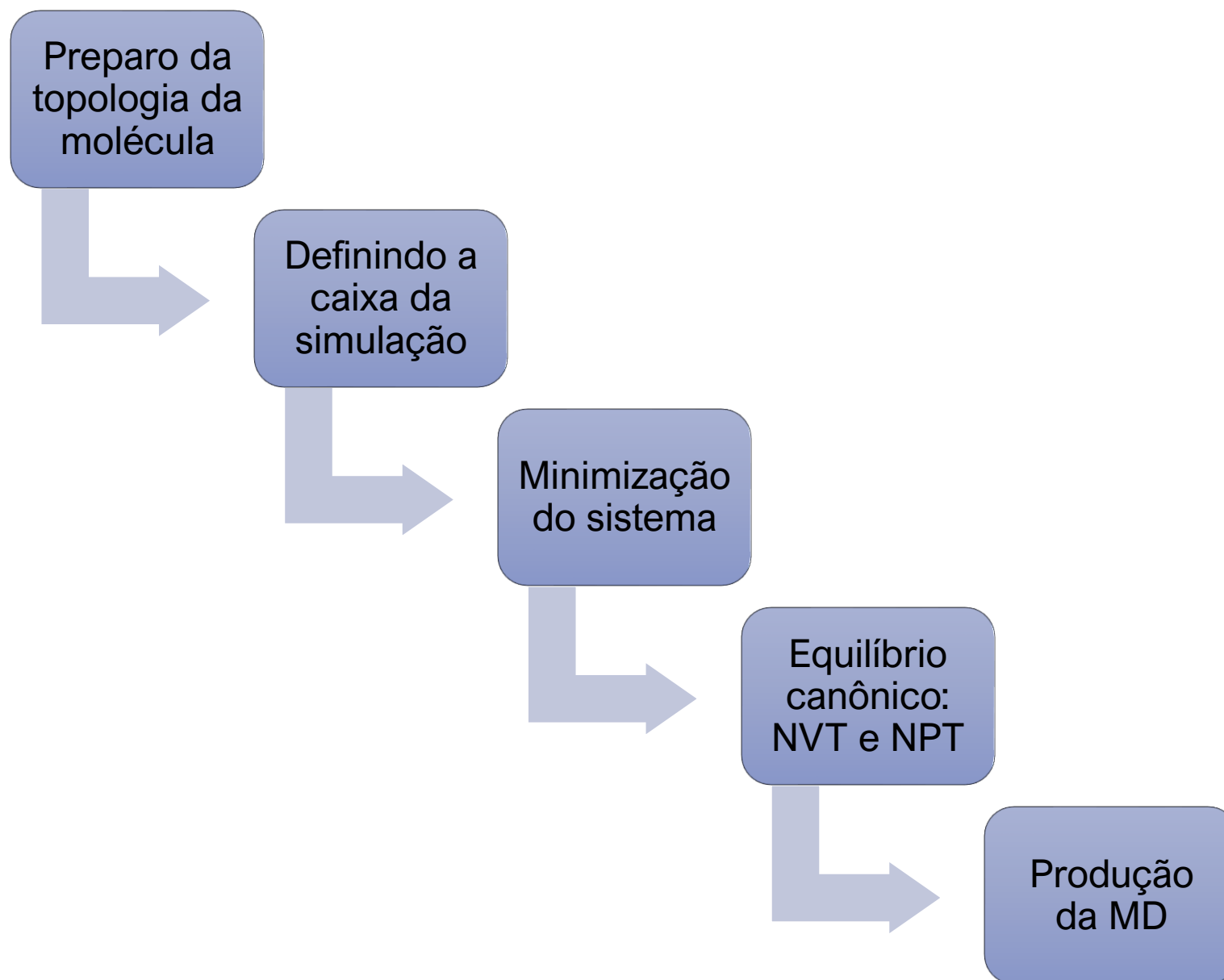
- 1957: gases simples com Alder and Wainwright (após a liberação do computador MANIAC I, Los Alamos, EUA)
- 1977: primeira simulação com proteína por McCammon et. al.
- 2013: Prêmio Nobel de Química, baseado nos trabalhos de Lifson de 1968.
- 2013: medicamentos para distúrbios neurodegenerativos com Wu e Shea.
- 2015: design de ferramentas oticogenéticas por Kato et al.
- 2016: sinalização neural com Dawe et al.
- 2018: desenvolvimento de medicamentos por McCorvy.
- 2020: Covid-19, proteína Spike e receptores ACE2, com desenvolvimento da vacina.
- 2022... Avanços em ML, IA, NNPOt e GPUs.

Mecânica molecular (MM)

- A trajetória de uma molécula é obtida através da integração das equações de movimento de Newton e expansões de Taylor.
- Para cada molécula, os átomos possuem coordenadas de posição $r_i(x_1, y_1, z_1)$.
- É determinada a posição futura ($t + \Delta t$), a velocidade e aceleração para cada átomo através de algoritmos (2).
- Todas as interações são descritas pela energia total potencial do sistema, $U(r)$.

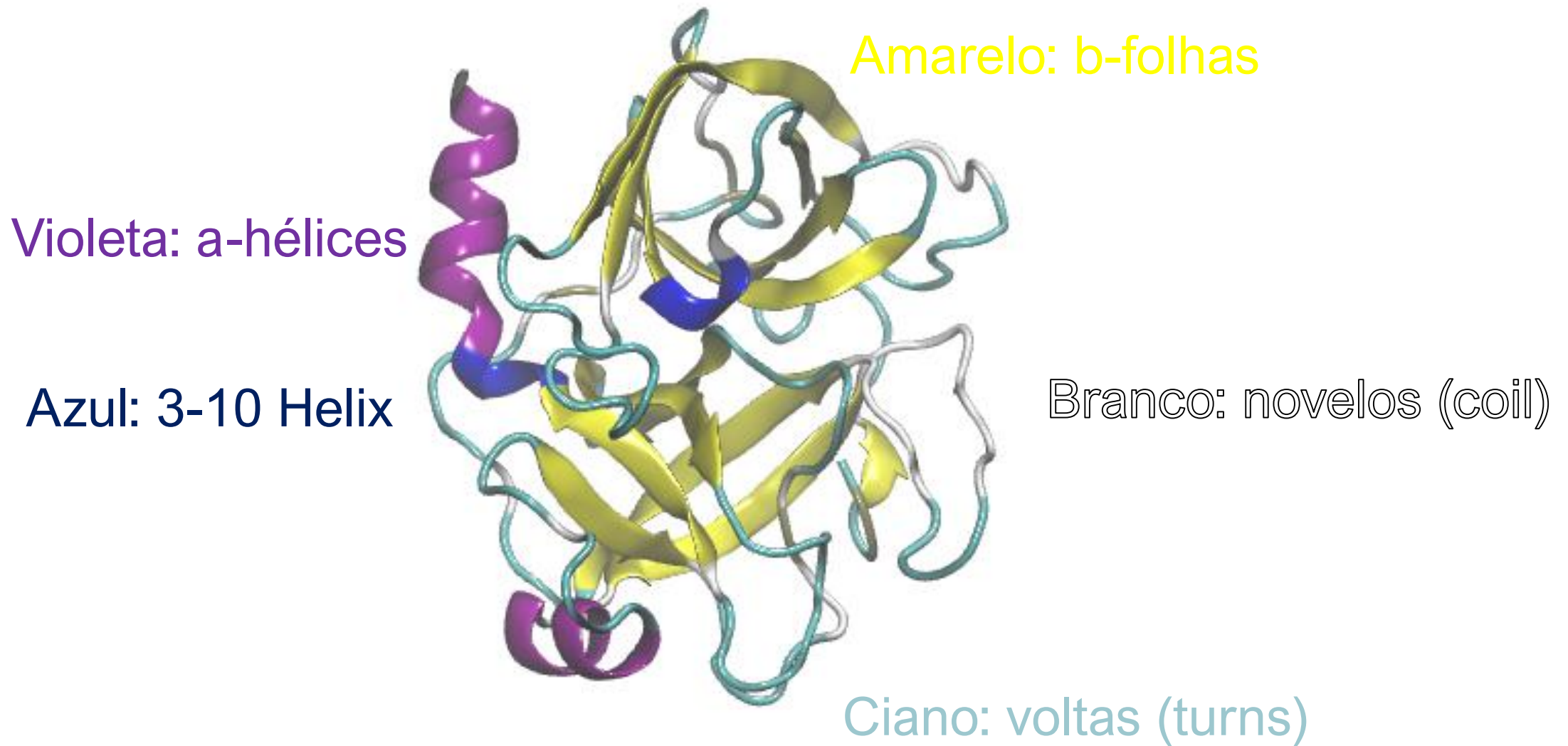






Preparo da topologia da biomolécula

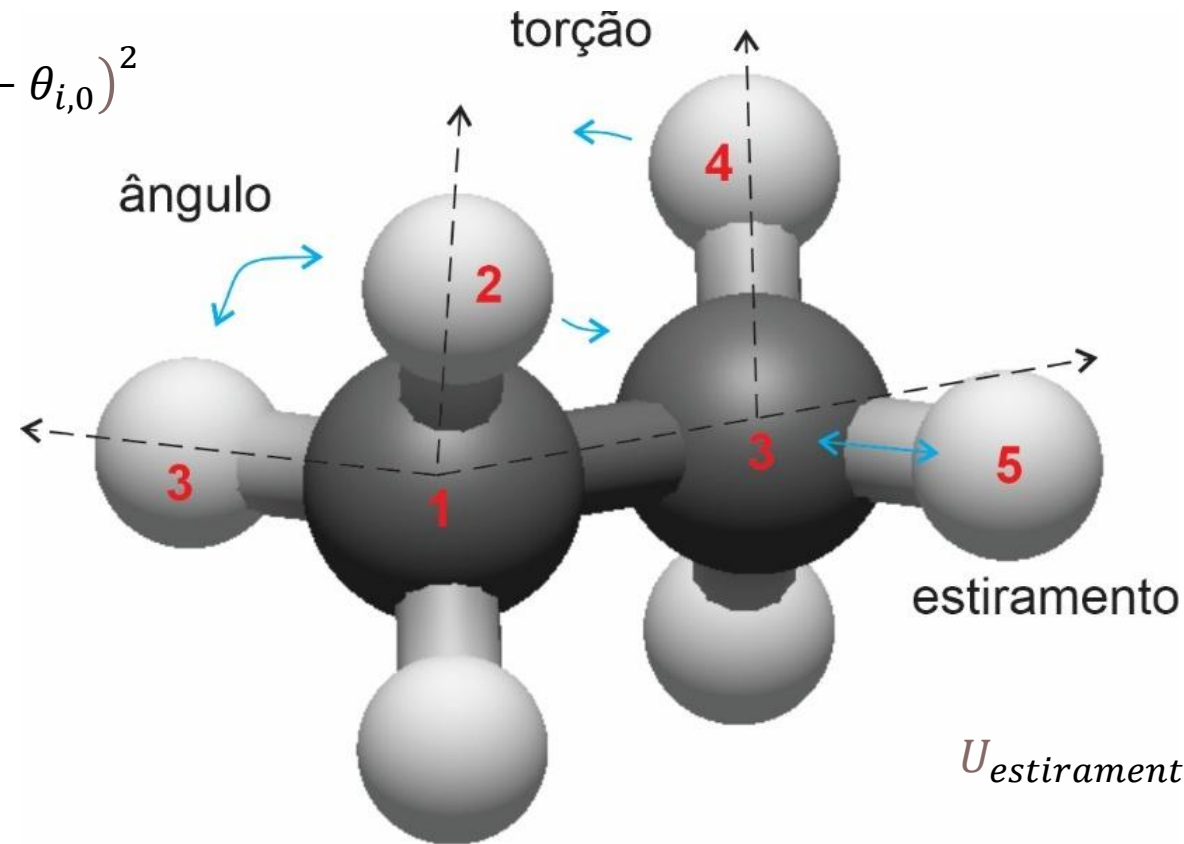
- Obtenção das coordenadas iniciais das moléculas de interesse (exemplo: PDB, PUBChem).
- Dê preferência a estruturas com resolução cristalográfica inferior a 2,5 Å.
- Campos de forças (4): CHARMM, AMBER, GROMOS e OPLS.
- Modelos de água (4): SPC, TIP3P, TIP4P, TIP5P e OPC.



Tripsina Pancreática Bovina, PDB ID 1S0Q (5)

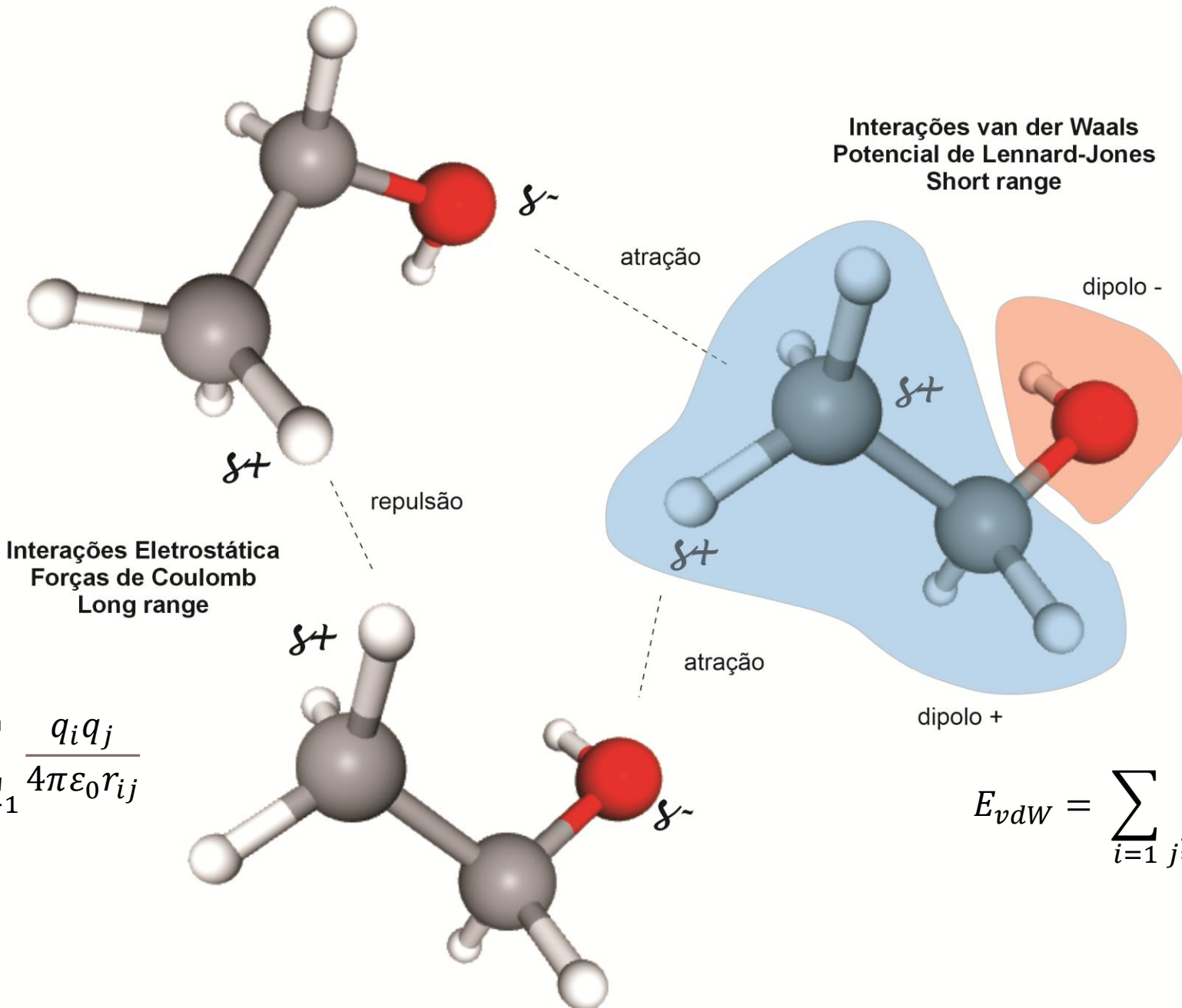
$$U_{torção} = \sum_i \frac{V_i}{2} (1 + \cos(i\omega - \gamma))$$

$$U_{angulo} = \sum_i \frac{k_i}{2} (\theta_i - \theta_{i,0})^2$$



$$U_{estiramento} = \sum_i \frac{k_i}{2} (l_i - l_{i,0})^2$$

$$U_{Ligação} = \sum U_{angulo} + \sum U_{torção} + \sum U_{estiramento}$$



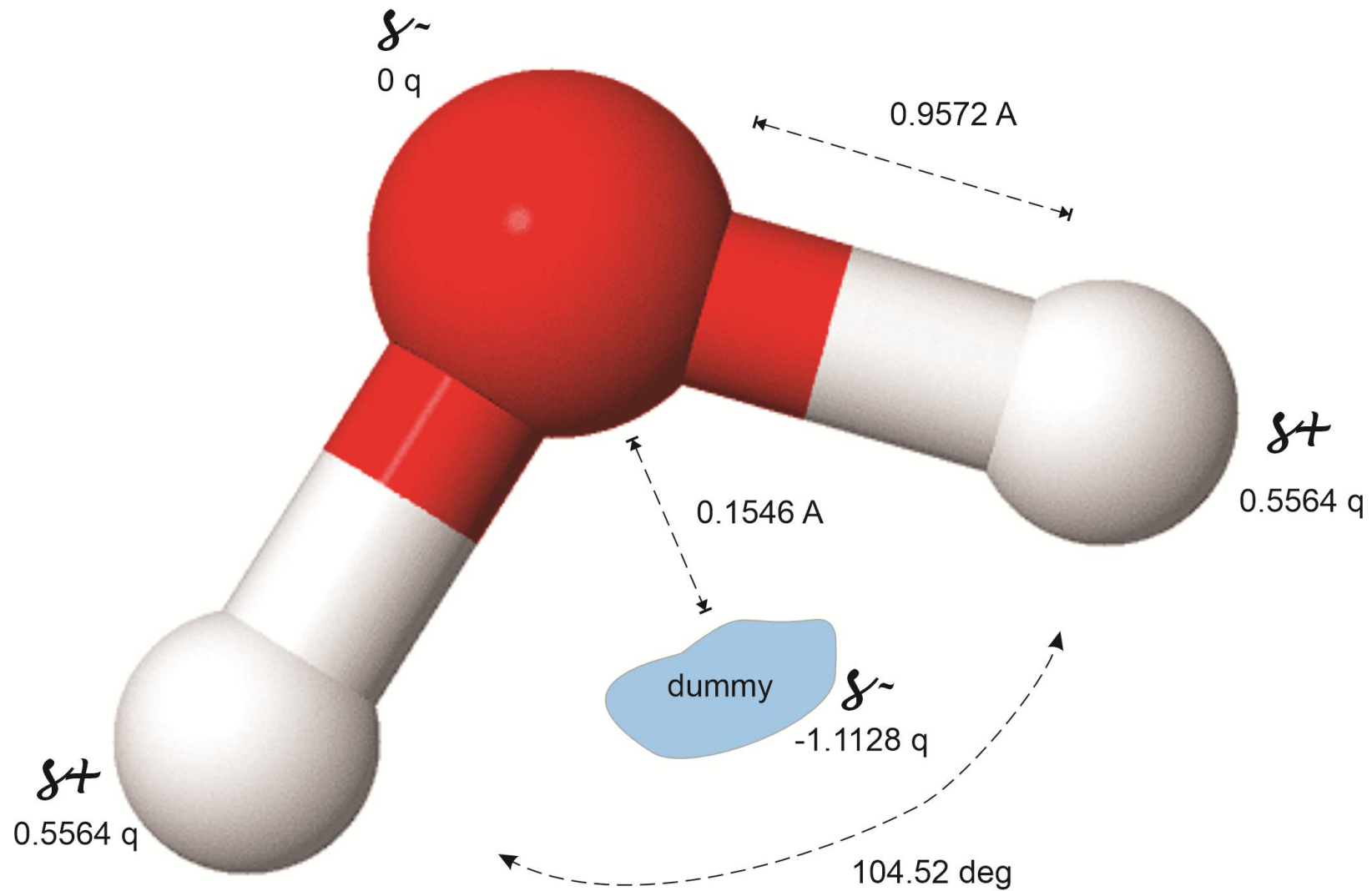
$$U_{elet} = \sum_{i=1} \sum_{j=i+1} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

$$E_{vdW} = \sum_{i=1} \sum_{j=i+1} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

Campo de Força	Informações	Modelo de água	cut-off
OPLS	O campo de força OPLS-AA (como OPLS-AA 2022) é amplamente usado para simulações de proteínas, pequenas moléculas, solventes, lipídios, dentre outros.	TIP4P recomendado, mas pode usar TIP3P. Não recomendado SPC.	1.0~1.2 nm
AMBER	A família de campos de força AMBER (como ff19sb, ff14sb etc.) é amplamente usada para proteínas, DNA/RNA e simulações biomoleculares.	TIP3P ou OPC, não recomendado TIP4P e SPC.	1.0~1.2 nm
CHARMM	O campo de força CHARMM (como charmm36-jul2022.ff) é extremamente detalhado, especialmente para lipídios, proteínas e açúcares, e foi parametrizado com switching functions, o que o diferencia das abordagens anteriores.	TIP3P modificado, não substituir por TIP3P comum.	1.2 nm
GROMOS	O campo de força GROMOS96 (como gromos54a7.ff) é uma escolha clássica para simulações de proteínas, sistemas aquosos e alguns tipos de estudos de bioenergia. Ele é o único desta lista a usar potencial truncado sem PME.	SPC	1.4 nm

Modelo	Tipo	Descrição
SPC	3 pontos	Modelo rígido, ângulo fixo de 109.47° , parametrizado para propriedades macroscópicas.
SPC/E	3 pontos	Versão estendida do SPC, com correção de energia de polarização. Melhor densidade e constante dielétrica.
TIP3P	3 pontos	Muito usado com AMBER e CHARMM. Simples e compatível com muitos campos de força.
TIP4P	4 pontos	Inclui ponto virtual (M-site) para carga negativa fora do oxigênio, melhorando propriedades de fase.
OPC	4 pontos	É o "novo padrão de ouro" para simulações com campos de força modernos (como AMBER), equilibrando precisão estrutural e custo.
TIP5P	5 pontos	Dois pontos extra para os pares de elétrons do oxigênio. Mais preciso para estrutura tetraédrica, porém mais custoso.

TIP4P 2005



Preparo da topologia da biomolécula



```
gmx pdb2gmx -v -f 1S0Q_clean.pdb -o tripsina.gro
```

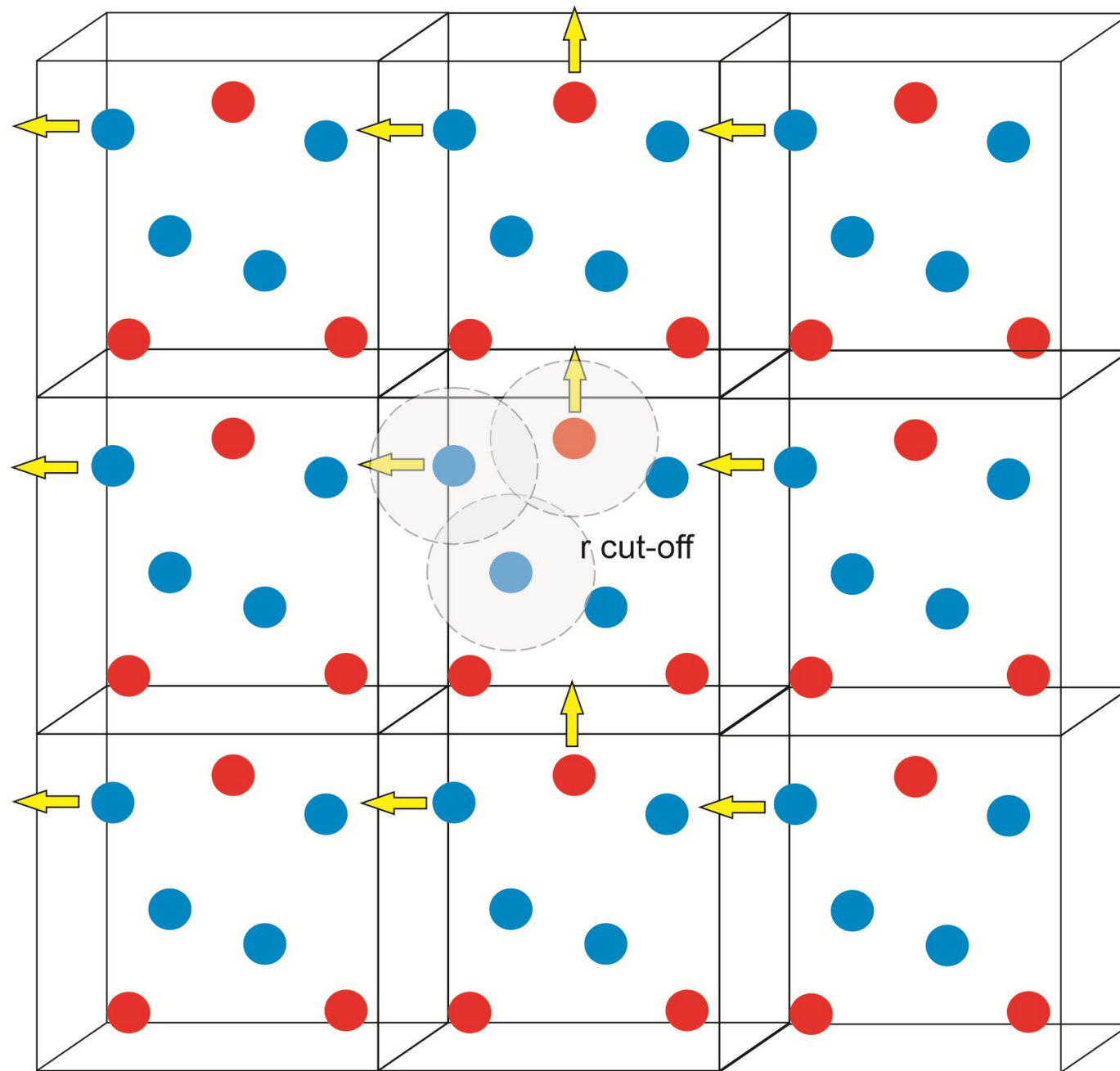
```
# -v = verbose, para visualizar o processo.
```

```
# -f = file input, arquivo de entrada das coordenadas.
```

```
# -o = file output, arquivo de saída das coordenadas.
```

Definindo a caixa de simulação (PBC)

- Modelo da caixa de simulação: cubic, triclinic, octahedron ou dodecahedron.
- Distância borda da caixa: 1,0 a 2,5 nm.
- Adicionamos o modelo de água selecionado anteriormente.
- Neutralizamos o sistema: condições fisiológicas.
- Condições Periódicas de Contorno e Convenção de Imagem Mínima.



Definindo a caixa de simulação (PBC)

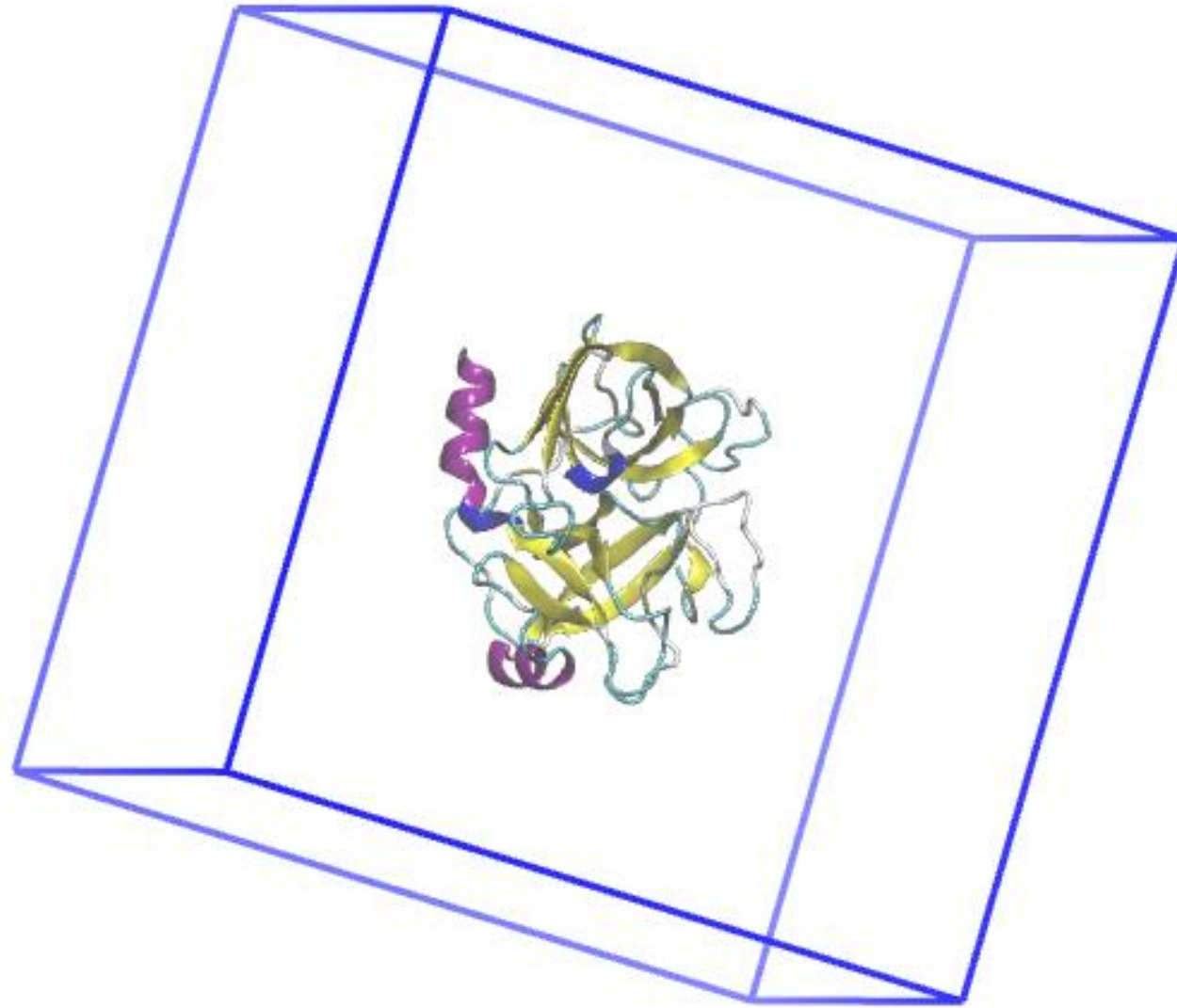


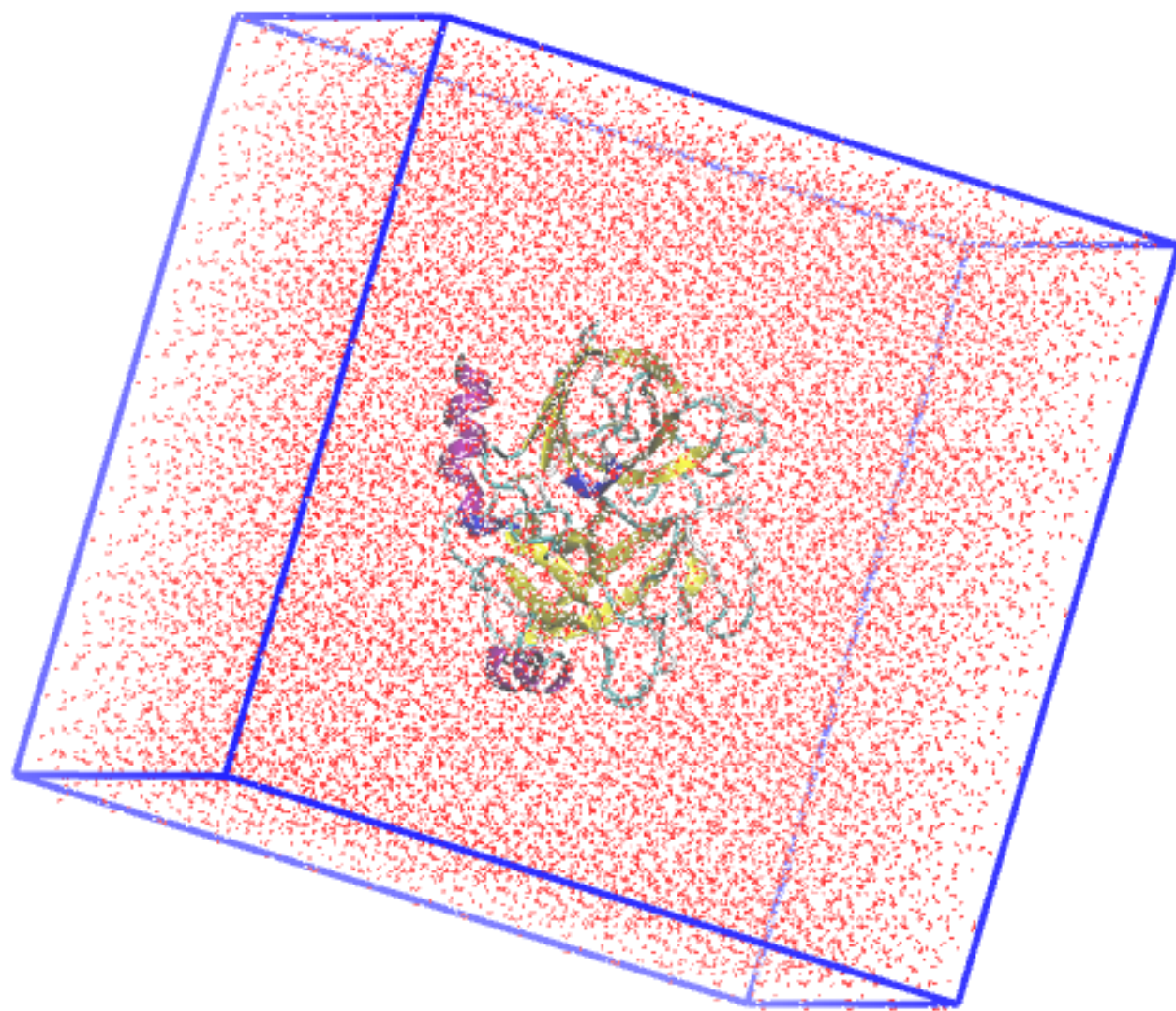
```
gmx editconf -f tripsina.gro -o box.gro -c -d 2.0 -bt cubic
```

```
# -c = center, para centralizar a biomolécula na caixa.  
# -d = distance, distância em nm entre todas moléculas e a borda.  
# -bt = box type, formato da caixa.
```

```
gmx solvate -cp box.gro -cs spc216.gro -o solvated.gro -p topol.top
```

```
# -cp = coordinates protein, coordenadas do nosso soluto (geralmente,  
proteína).  
# -cs = coordinates solvent, coordenadas da molécula que será usada como  
solvente.  
# -p = processing, para processar o arquivo de topologia do sistema.
```





Definindo a caixa de simulação (PBC)



```
gmx grompp -v -f inputs/ions.mdp -o ions.tpr -c solvated.gro -p topol.top
```

```
# -c = coordenates, arquivo com as coordenadas do sistema.
```

```
gmx genion -s ions.tpr -o solvated_ions.gro -p topol.top -pname NA -nname CL -  
neutral -conc 0.15
```

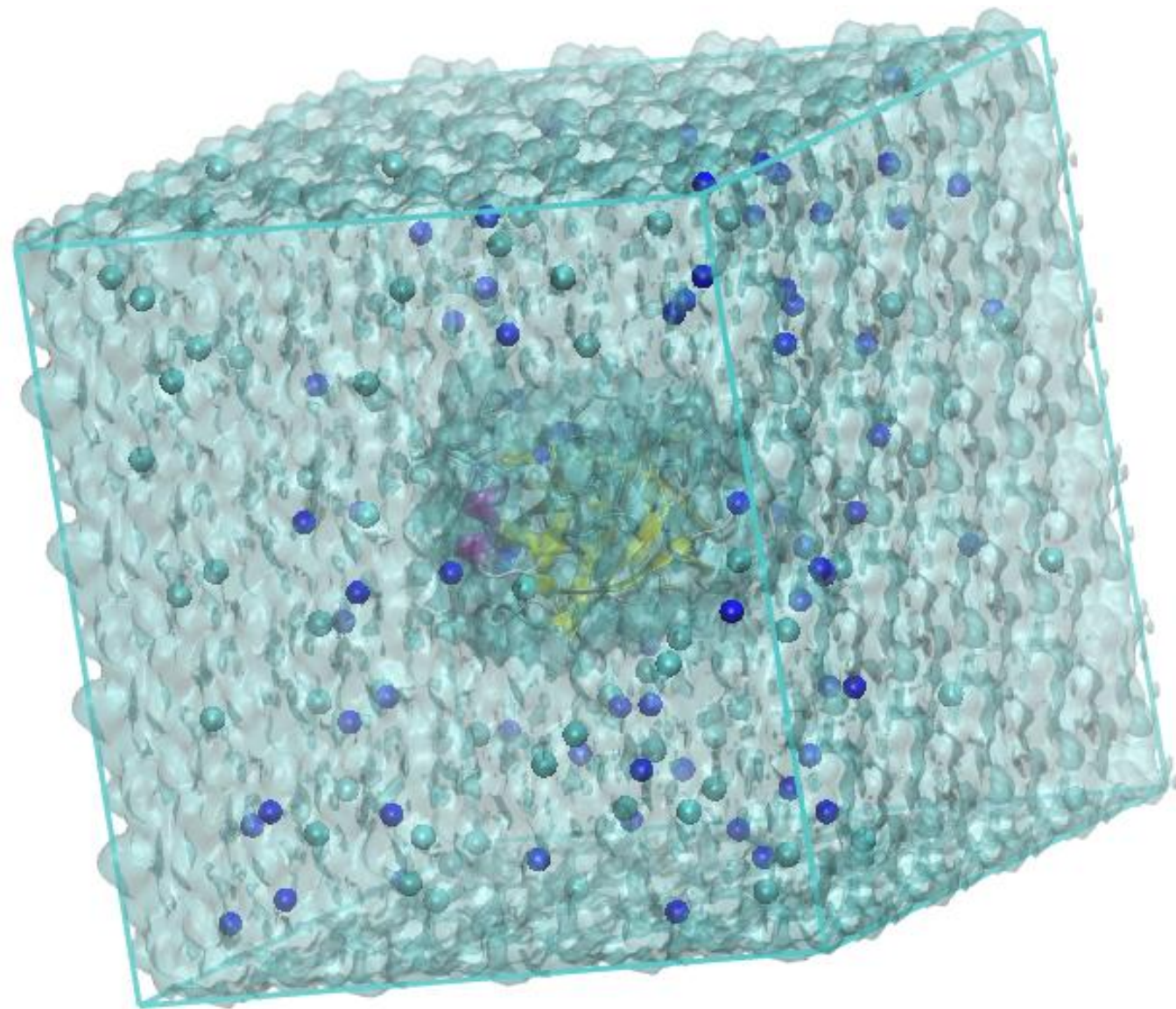
```
# -s = submit binary, arquivo binário criado anteriormente com todas  
informações do sistema.
```

```
# -pname = nome do cátion(+), nesse caso NA Sódio.
```

```
# -nname = nome do ânion(-), nesse caso CL Cloro.
```

```
# -neutral = para neutralizar completamente o sistema, às vezes desnecessário.
```

```
# -conc 0.15 = concentration, define a concentração em mol/L.
```

Minimização do sistema

- O objetivo da minimização é reduzir a energia potencial do sistema ajustando as coordenadas dos átomos para eliminar as colisões atômicas e sobreposições.

minim.mdp

Code

Blame

11 lines (11 loc) · 807 Bytes



Raw



```
1  integrator = steep      ; Algorithm (steep = steepest descent minimization)
2  emtol      = 1000.0     ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
3  emstep     = 0.01      ; Minimization step size (nm)
4  nsteps     = 50000     ; Maximum number of (minimization) steps to perform
5  nstlist    = 20        ; Frequency to update the neighbor list and long range forces
6  cutoff-scheme = Verlet  ; Buffered neighbor searching
7  coulombtype = PME       ; Treatment of long range electrostatic interactions
8  rcoulomb    = 1.0       ; Short-range electrostatic cut-off
9  rvdw       = 1.0       ; Short-range Van der Waals cut-off
10 rlist      = 1.0       ; Cut-off for making neighbor list
11 pbc        = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```


Minimização do sistema

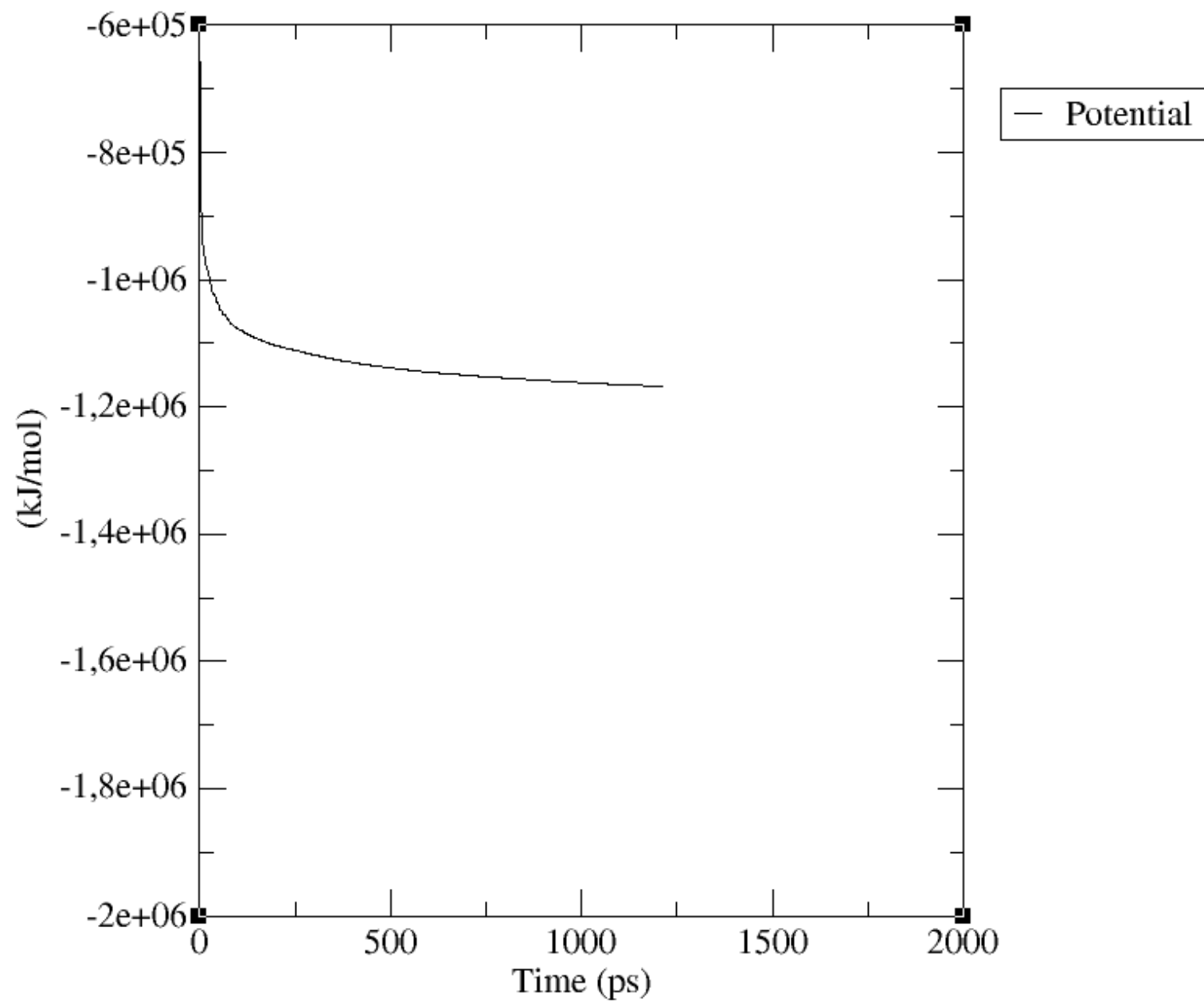


```
gmx grompp -v -f inputs/minim.mdp -c solvated_ions.gro -o em.tpr -p topol.top
```

```
gmx mdrun -v -deffnm em
```

```
# -deffnm = define o nome padrão de todos arquivos de entrada e saída.
```

GROMACS Energies



Equilíbrio canônico: NVT e NPT

- Equilíbrio canônico NVT (6): número de partículas, volume e temperatura são mantidos constantes.
 - Berendsen, Nosé-Hoover, Andersen e V-rescale*.
- Equilíbrio canônico NPT (6): número de partículas, pressão e temperatura são mantidos constantes.
 - Berendsen, Parrinello-Rahman e C-rescale*.
- Restrição de posições para átomos da biomolécula para equilibrar a vizinhança.

Equilíbrio canônico: NVT



```
gmx grompp -v -f inputs/nvt.mdp -c em.gro -r em.gro -o nvt.tpr -p topol.top
```

-r = restrain file, arquivo de coordenadas com as restrições iniciais (geralmente mesmo arquivo).

```
gmx mdrun -v -deffnm nvt
```

Termostato	Características	Vantagens	Limitações
Berendsen	Rápido para equilibrar temperatura	Simples e eficiente para equilíbrios	Não reproduz corretamente as flutuações canônicas
V-rescale*	Mantém temperatura média correta e flutuações realistas	Estável e mais preciso que Berendsen	Ligeiramente mais complexo
Nose-Hoover	Mantém distribuição canônica (NVT)	Correto estatisticamente	Pode ter acoplamento mais lento

Equilíbrio canônico: NPT



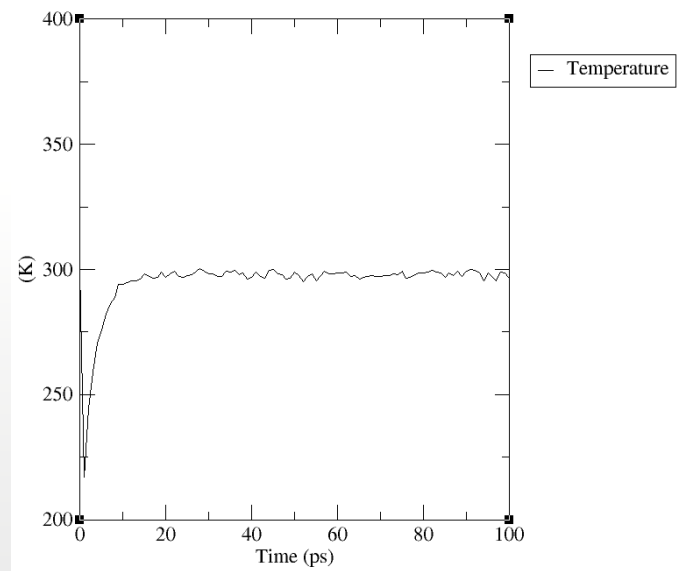
```
gmx grompp -v -f inputs/npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -o npt.tpr -p  
topol.top
```

-t = time file, arquivo com checkpoint anterior (geralmente utilizado para indicar o ponto de partida com relação a dinâmica anterior)

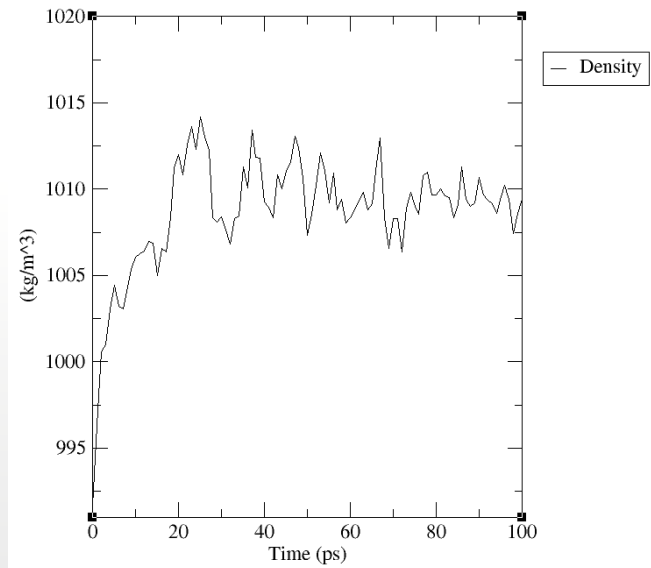
```
gmx mdrun -v -deffnm npt
```

Barostato	Características	Vantagens	Limitações
Berendsen	Ajusta pressão rapidamente durante o equilíbrio	Simples, ideal para pré-produção	Não reproduz corretamente as flutuações canônicas
Parrinello-Rahman	Permite flutuações de volume e forma da caixa (NPT)	Correto para simulações de produção	Pode ser instável sem bom equilíbrio inicial
C-rescale*	Versão estocástica rigorosa de controle de pressão. Mantém flutuações canônicas corretas no ensemble NPT	Produce NPT canônico exato, mais robusto e estável que Parrinello-Rahman em algumas situações	Disponível a partir do GROMACS 2023, pouco testado em comparação com Parrinello-Rahman

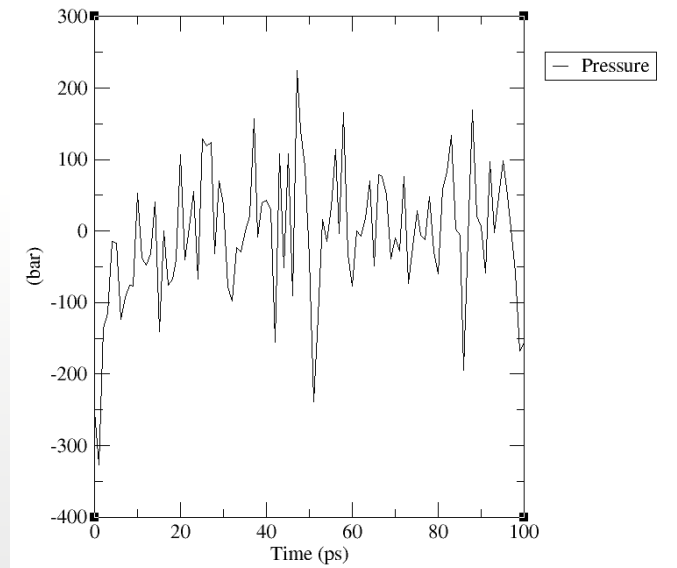
GROMACS Energies



GROMACS Energies



GROMACS Energies



Produção da DM

- Etapa longa, com custo computacional.
- GPU vs. CPU.
- Monitorar o processo: temperatura e carga de uso dos hardwares.
- Escolha do integrador.
 - Verlet, leap-frog, velocity-verlet, estocástico etc.
- Sem restrições.

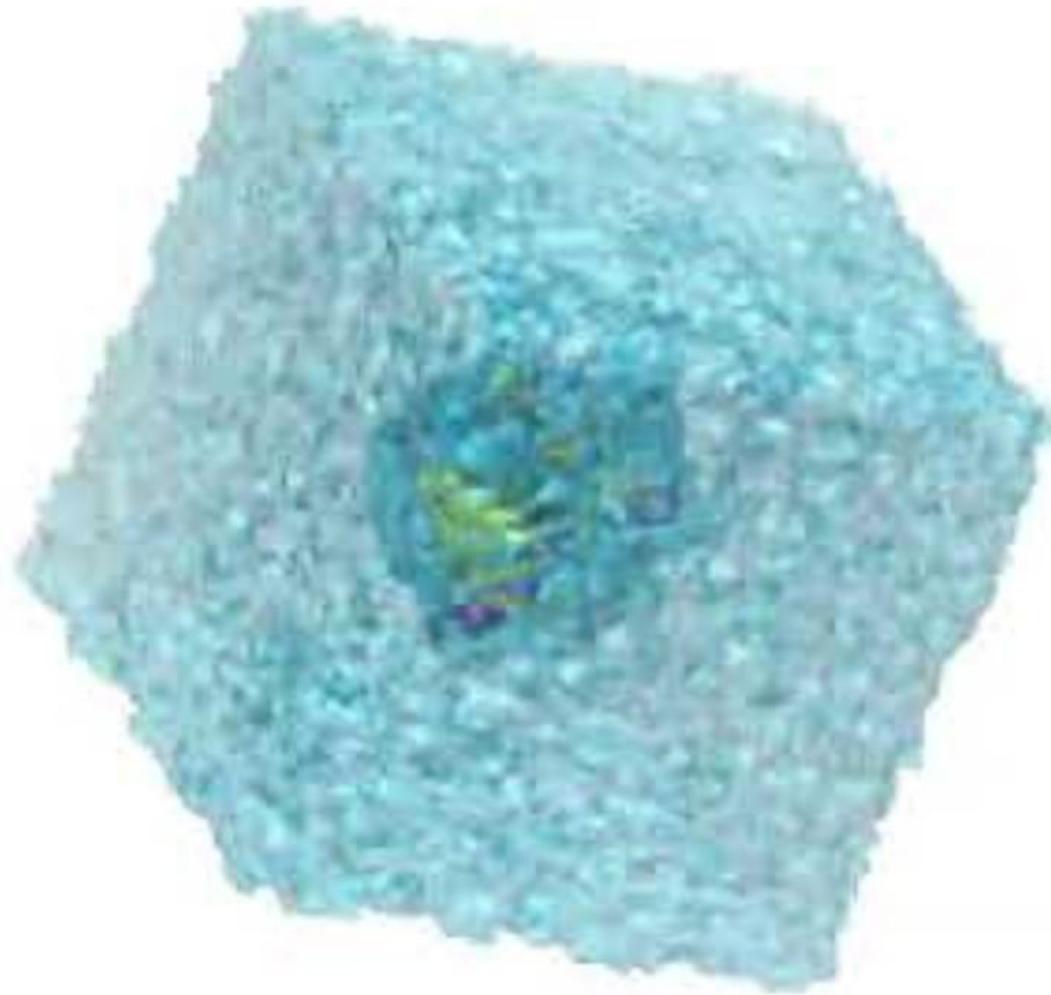
Integrador	Características	Vantagens	Limitações / Quando evitar	Uso típico
md*	Leap-frog Verlet. Passo de tempo curto (1–2 fs). Conserva bem energia e momento.	Robusto, padrão, eficiente.	Velocidades não coincidem com posições.	Produção em proteínas, membranas, solventes.
md-vv	Velocity-Verlet. Calcula velocidades no mesmo ponto que posições.	Melhora cálculo de velocidades.	Pouco ganho em muitos casos.	Transporte, difusão, análise energética.
md-vv-avek	Velocity-Verlet com controle de energia cinética média (AVEK).	Temperatura estável sem termostato.	Mais pesado; pouco usado.	Equilíbrios longos sensíveis a flutuações.
sd	Dinâmica de Langevin (stochastic). Força de fricção + ruído gaussiano..	Excelente controle térmico.	Distorce dinâmica real em excesso.	Sistemas viscosos, líquidos iônicos, membranas.
bd	Dinâmica Browniana (Langevin overdamped). Ignora momento, apenas difusão.	Simple e estável.	Perde informação de movimento rápido.	Difusão lenta, modelos grosseiros (CG).
steep	Minimização por gradiente descendente. Desce na direção de maior inclinação.	Rápido para remover contatos ruins.	Convergência lenta perto do mínimo.	Pré-MD, relaxamento inicial.
cg	Minimização gradiente conjugado.	Mais eficiente que steep.	Menos robusto no início.	Refinar após steep.
l-bfgs	Minimização quasi-Newton.	Muito rápido em sistemas pequenos.	Ineficiente em sistemas grandes.	Clusters ou moléculas pequenas.

Produção da DM



```
gmx grompp -v -f inputs/md.mdp -c npt.gro -t npt.cpt -o md.tpr -p topol.top
```

```
gmx mdrun -v -deffnm md
```



Análise de resultados

- Análises de propriedades físico-químicas e termodinâmicas: **capacidades caloríficas, viscosidades** etc.
- **SASA** – Área superficial acessível ao solvente.
- **Rg** – Raio de giro.
- **RMSD** – Estabilidade estrutural.
- **RDF** – Função de distribuição radial e camadas de solvatação.
- **Energia livre** de ligação e solvatação/hidratação.
- Termodinâmica de fases líquidas: teoria de Kirkwood-Buff (7, 8).

Análise de resultados



```
gmx energy -f md.edr -s md.tpr -o analysis.xvg
```

```
patrickfaustino@pck-TUF570:~/Documentos/3i40/easy$ gmx energy -f md.edr -s md.tpr -o analysis.xvg
:~) GROMACS - gmx energy, 2025.2 (-:
```

```
Executable: /home/patrickfaustino/gromacs-hip-torch_cpu/bin/gmx
Data prefix: /home/patrickfaustino/gromacs-hip-torch_cpu
Working dir: /home/patrickfaustino/Documentos/3i40/easy
Command line:
  gmx energy -f md.edr -s md.tpr -o analysis.xvg
```

Opened md.edr as single precision energy file

Select the terms you want from the following list by
selecting either (part of) the name or the number or a combination.
End your selection with an empty line or a zero.

```
-----
 1 Bond                2 Angle                3 Proper-Dih.          4 Per.-Imp.-Dih.
 5 LJ-14               6 Coulomb-14           7 LJ-(SR)              8 Disper.-corr.
 9 Coulomb-(SR)        10 Coul.-recip.        11 Potential            12 Kinetic-En.
13 Total-Energy        14 Conserved-En.       15 Temperature          16 Pres.-DC
17 Pressure            18 Constr.-rmsd        19 Box-X                20 Box-Y
21 Box-Z              22 Volume              23 Density              24 pV
25 Enthalpy            26 Vir-XX              27 Vir-XY               28 Vir-XZ
29 Vir-YX              30 Vir-YY              31 Vir-YZ               32 Vir-ZX
33 Vir-ZY              34 Vir-ZZ              35 Pres-XX              36 Pres-XY
37 Pres-XZ             38 Pres-YX             39 Pres-YY              40 Pres-YZ
41 Pres-ZX             42 Pres-ZY             43 Pres-ZZ              44 #Surf*SurfTen
45 T-Protein           46 T-non-Protein
47 Lamb-Protein        48 Lamb-non-Protein
```

```
25 0
```

Análise de resultados



```
gmx energy -f md.edr -s md.tpr -o analysis.xvg -fluct_props -nmol 15462 -driftcorr
```

```
Statistics over 5000001 steps [ 0.0000 through 10000.0000 ps ], 2 data sets
All statistics are over 50001 points
```

Energy	Average	Err.Est.	RMSD	Tot-Drift	
Temperature	298.167	0.017	1.40651	0.0236595	(K)
Enthalpy	-34.8706	0.0021	0.0650467	-0.0118601	(kJ/mol)

```
Temperature dependent fluctuation properties at T = 298.164.
```

```
Heat capacities obtained from fluctuations do *not* include
quantum corrections. If you want to get a more accurate estimate
please use the gmx dos program.
```

```
WARNING: Please verify that your simulations are converged and perform
a block-averaging error analysis (not implemented in gmx energy yet)
```

```
Enthalpy = -34.8693 kJ/mol
Heat capacity at constant pressure Cp = 87.6355 J/(mol K)
```

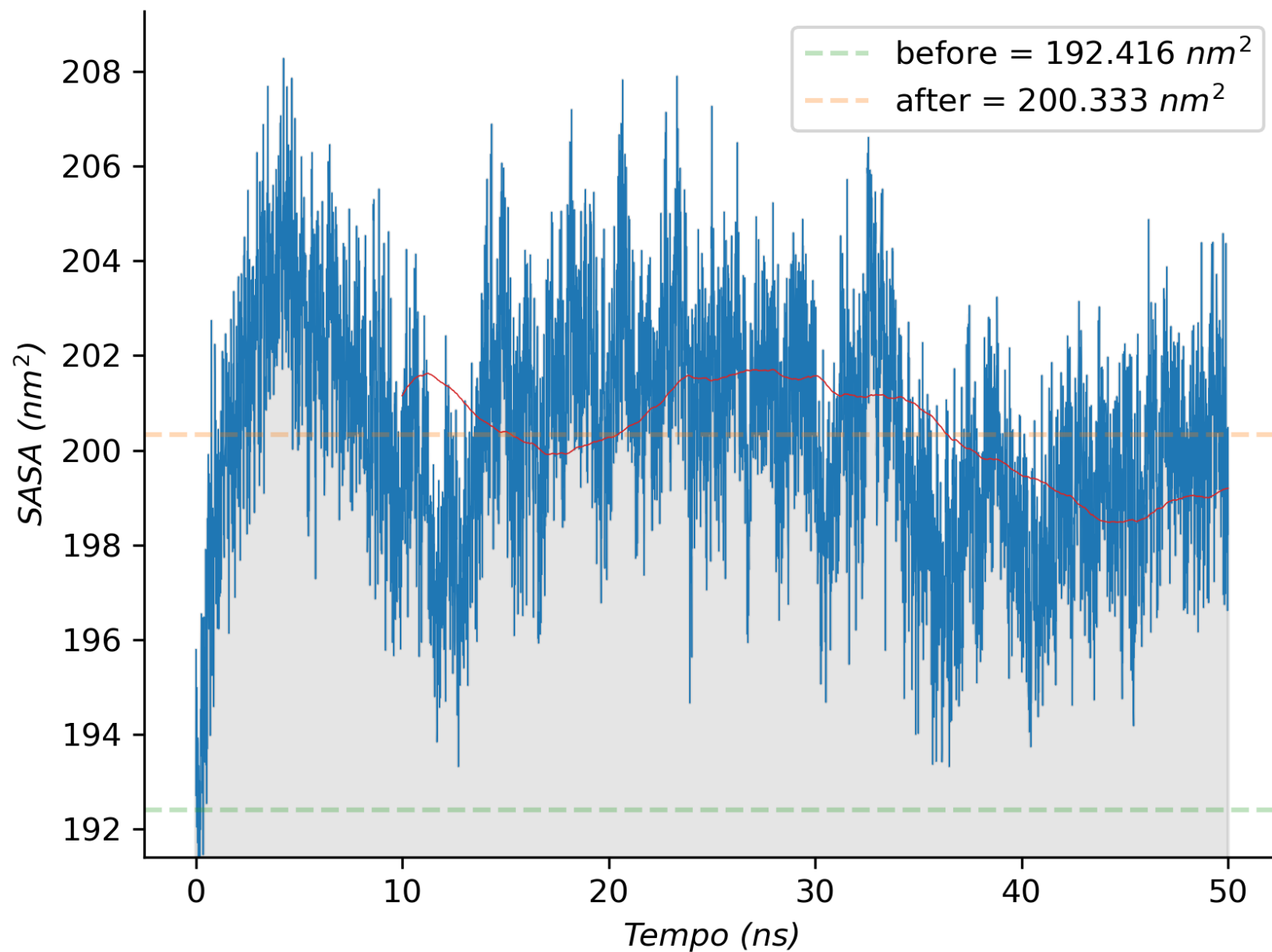
```
++++ PLEASE READ AND CITE THE FOLLOWING REFERENCE +++++
```

```
M. P. Allen, D. J. Tildesley
Computer simulation of liquids
Oxford Science Publications (2017)
DOI: 10.1093/oso/9780198803195.001.0001
```

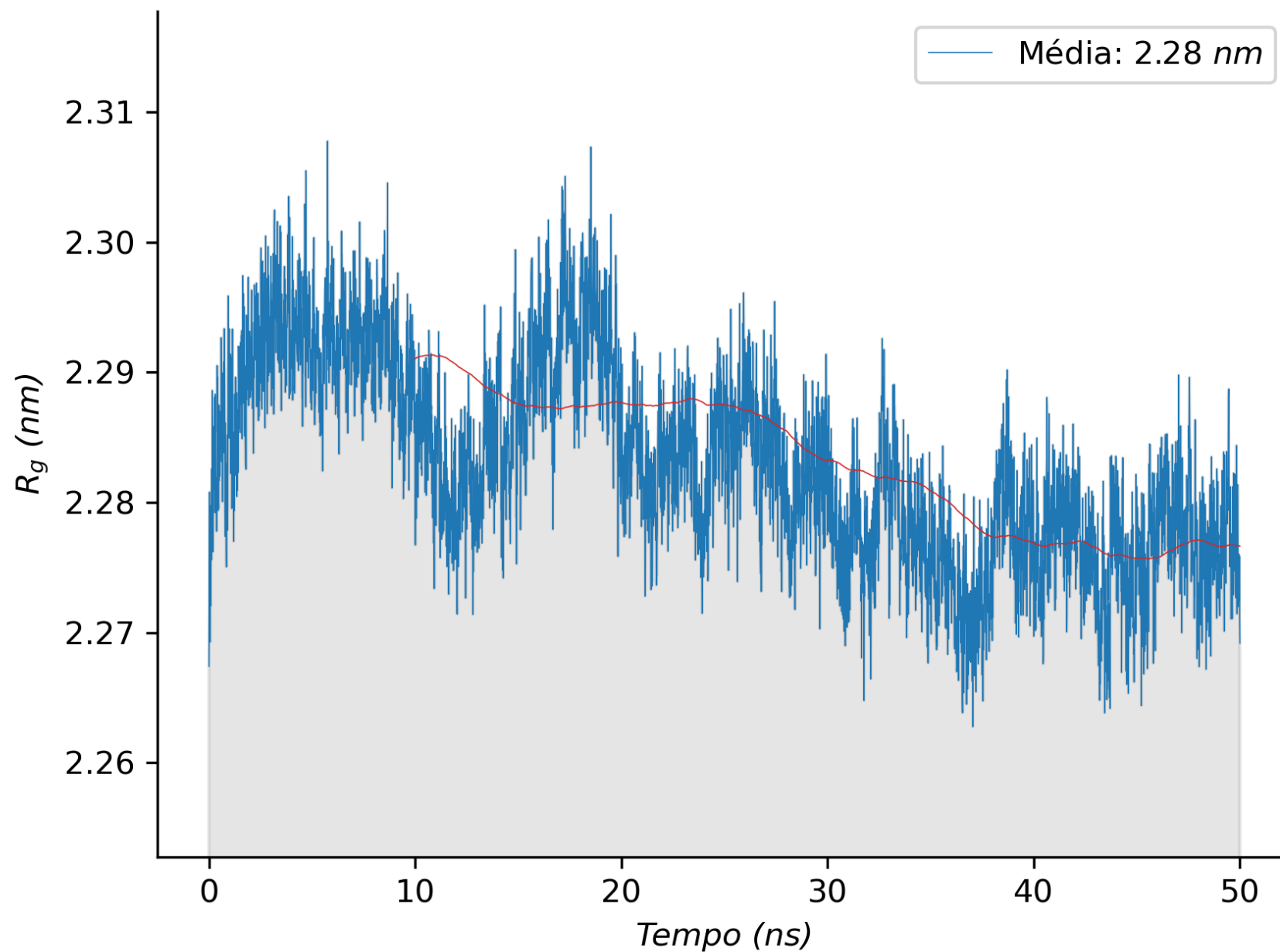
```
----- Thank You -----
```

Propriedade	Termos de Energia Necessários	Ensemble	Descrição
Capacidade Térmica a Volume Constante (C_v)	Total-Energy, Temperature	NVT	Mede como a energia do sistema muda com a temperatura a volume constante.
Capacidade Térmica a Pressão Constante (C_p)	Enthalpy, Temperature	NPT	Mede como a entalpia do sistema muda com a temperatura a pressão constante.
Coeficiente de Expansão Térmica (α)	Enthalpy, Volume, Temperature	NPT	Mede a variação fracionária no volume por unidade de aumento na temperatura.
Compressibilidade Isotérmica (β_T)	Volume, Temperature	NPT	Mede a variação fracionária no volume em resposta a uma mudança na pressão, a temperatura constante.
Módulo de Bulk Adiabático	Volume, Temperature	NPT	Mede a resistência de uma substância à compressão uniforme.

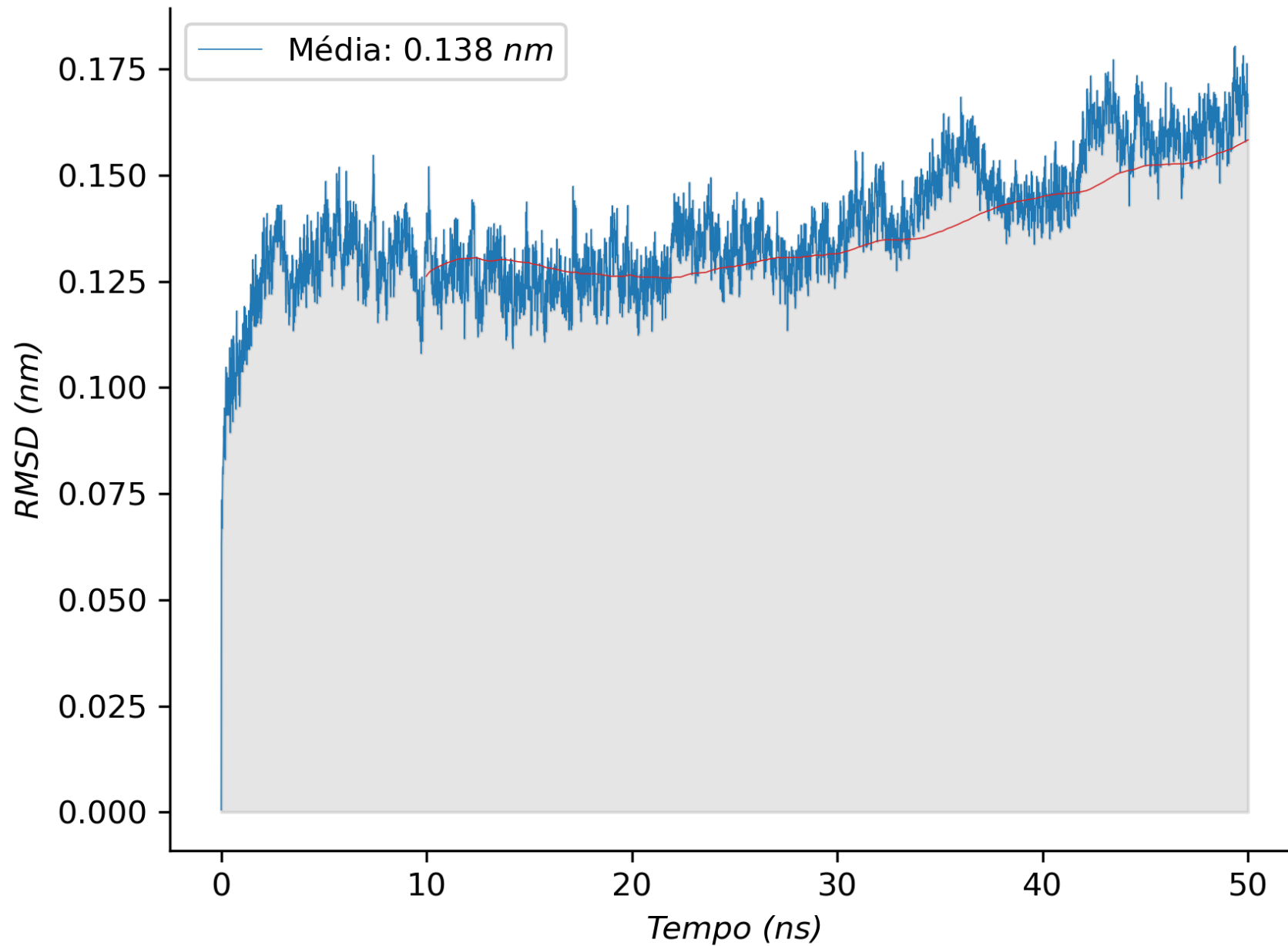
SASA da proteína 1UKC/Acetato em fase aquosa



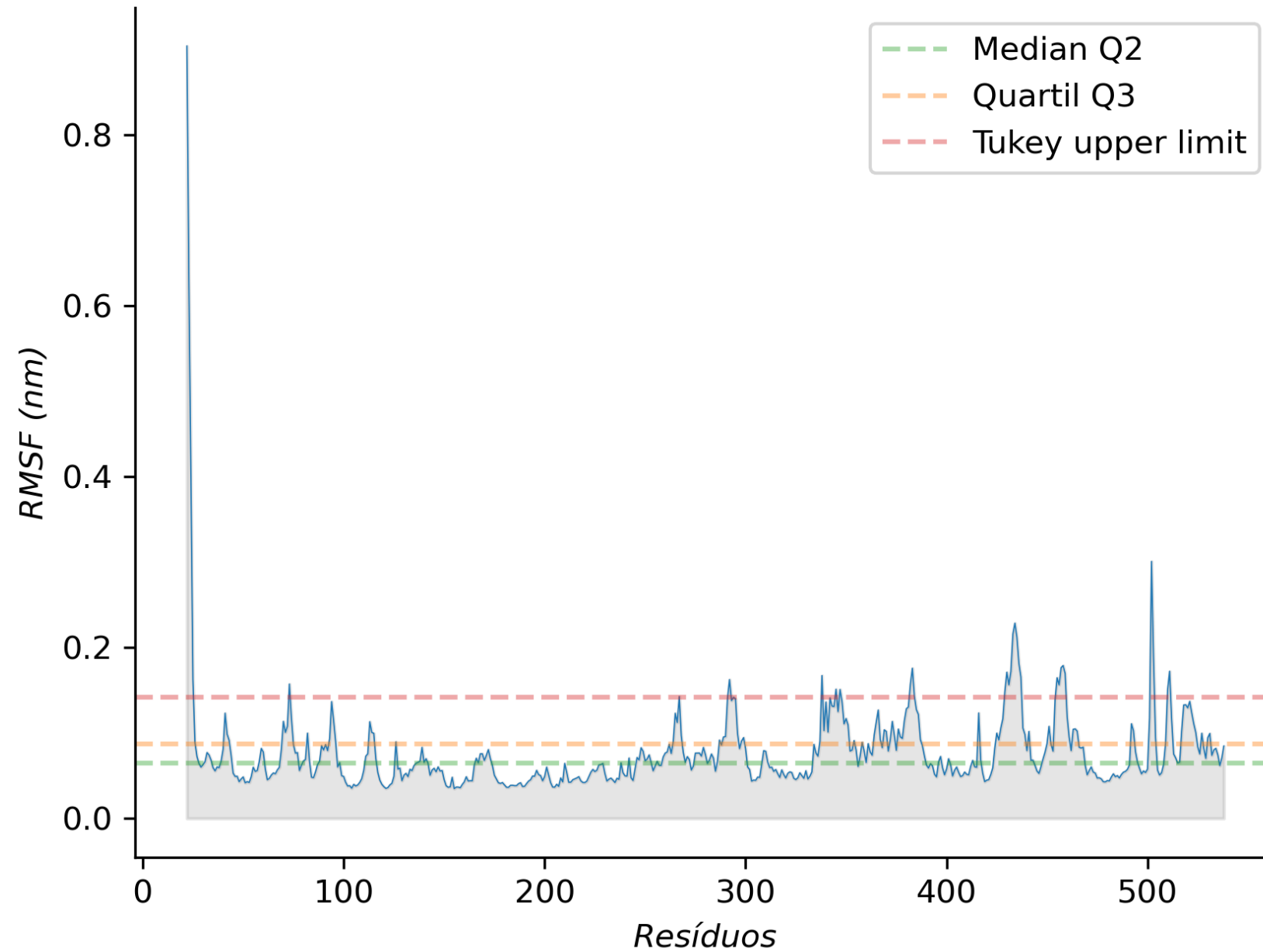
Raio de Giro da proteína 1UKC/Acetato em fase aquosa



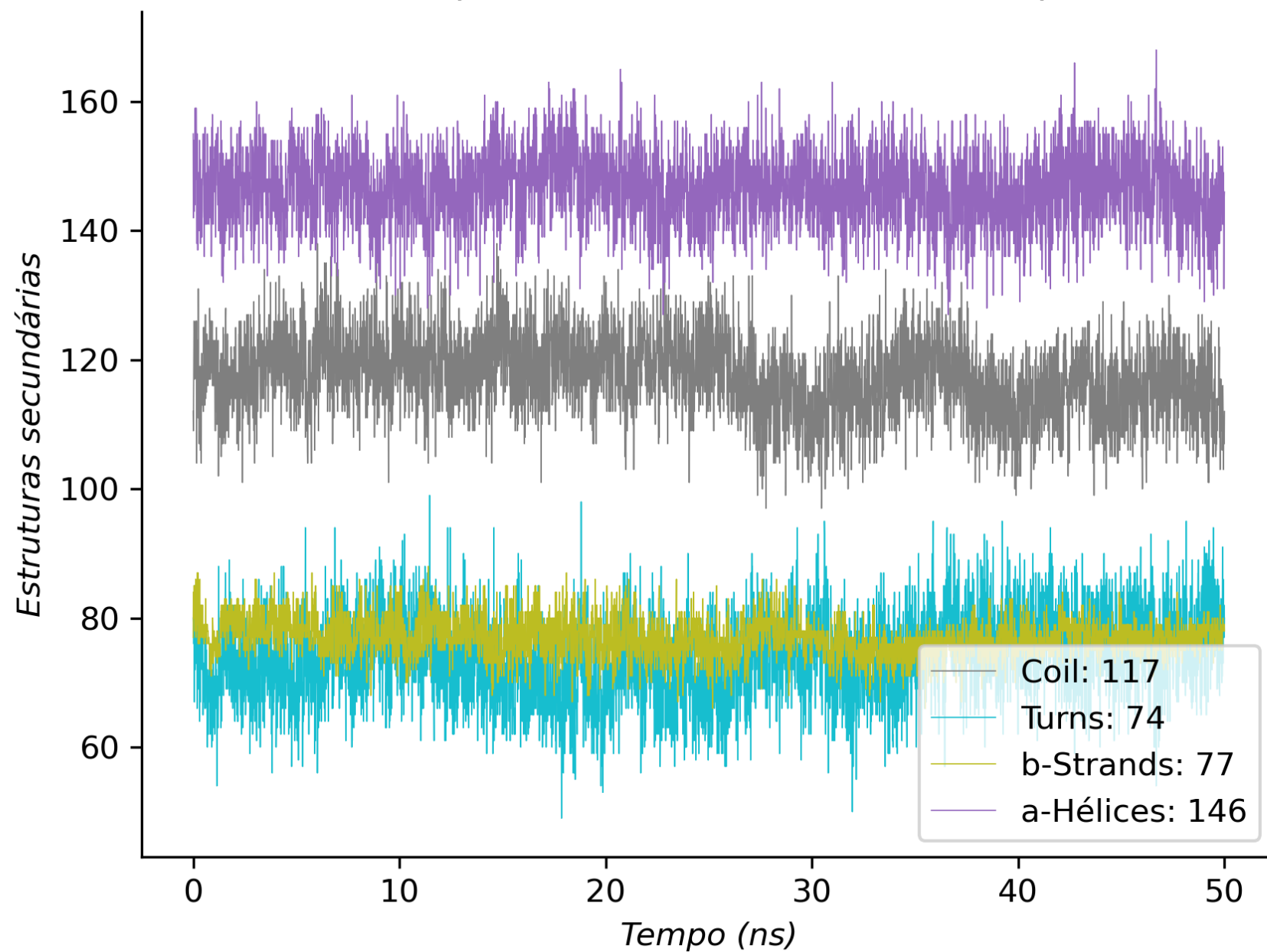
RMSD da proteína 1UKC/Acetato em fase aquosa (ref. backbone)



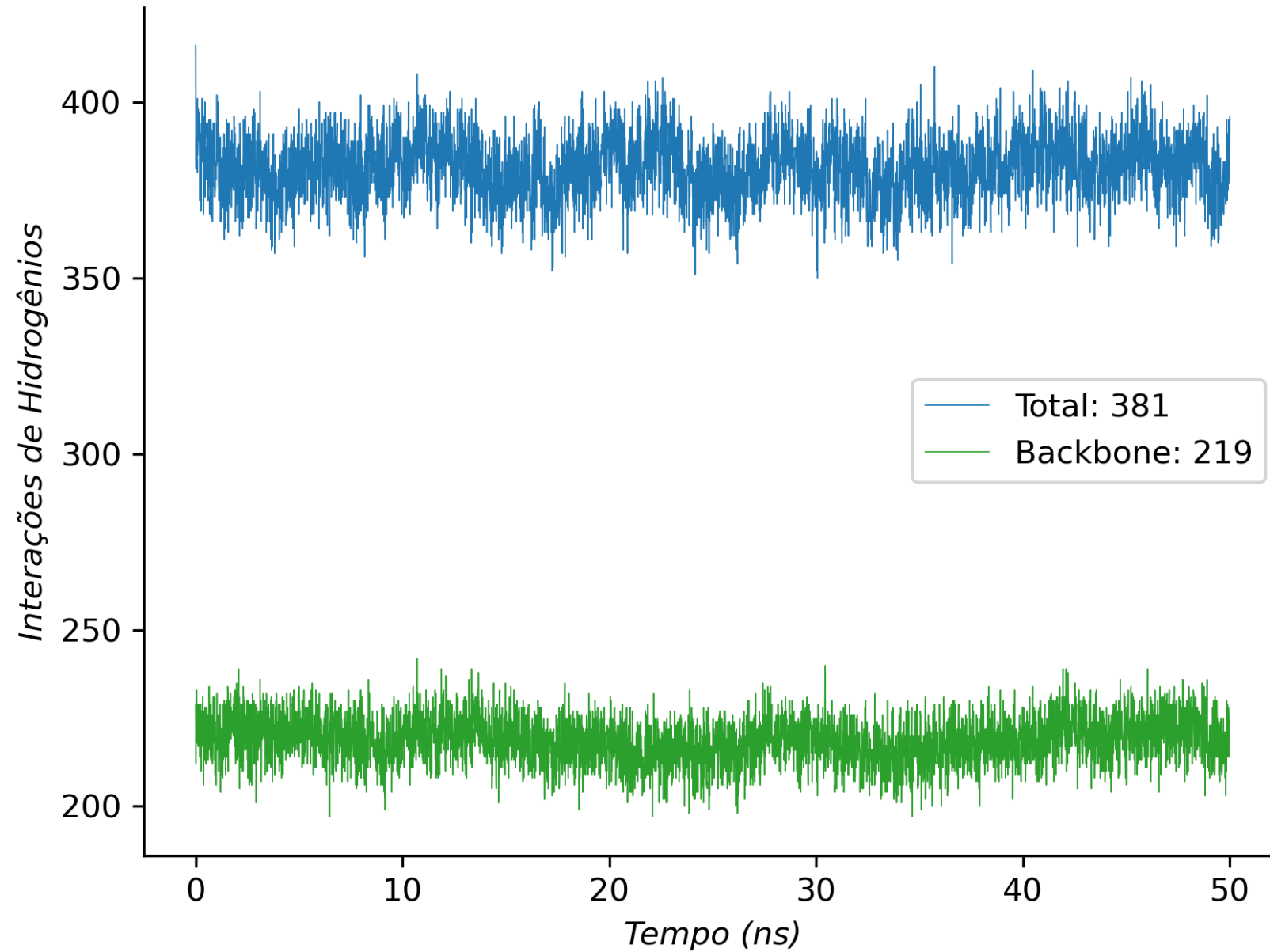
RMSF da proteína 1UKC/Acetato em fase aquosa (ref. Ca)



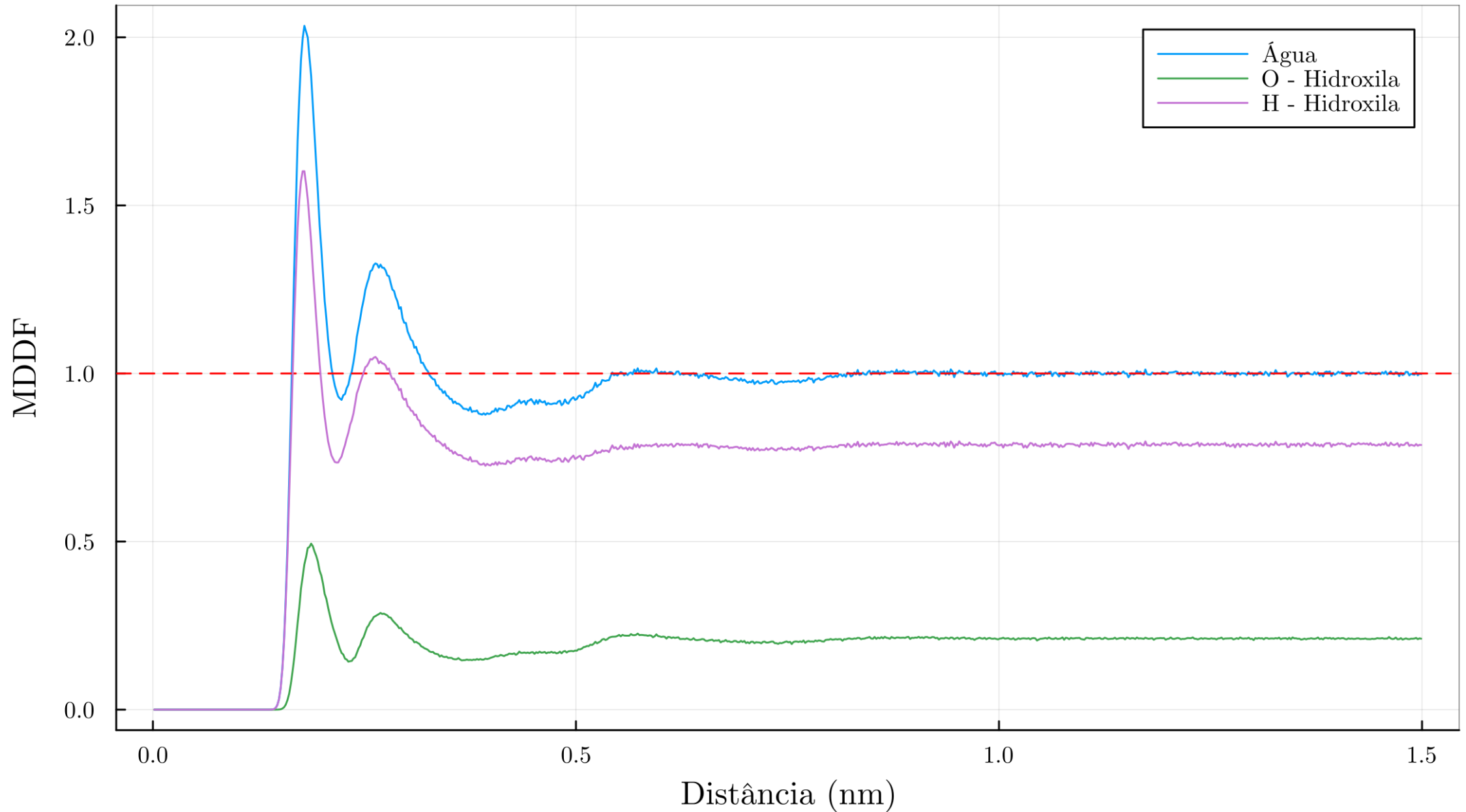
DSSP da proteína 1UKC/Acetato em fase aquosa



Interações de Hidrogênios da proteína 1UKC/Acetato em fase aquosa



1UKC/Água na Fase Aquosa com Ch+/Acetato/Triton X-114



Conclusões

- Permite investigar fenômenos em escala atômica e molecular com alto grau de detalhamento.
- Facilita o estudo de processos físicos, químicos e biológicos de forma virtual, mesmo antes de experimentos laboratoriais.
- Proporciona economias significativas de tempo e recursos em pesquisa e desenvolvimento.
- Oferece insights e informações muitas vezes inacessíveis por métodos experimentais tradicionais.
- Favorece a análise de sistemas complexos, como proteínas, membranas e materiais avançados.



github.com/patrickallanfaustino

patrick.faustino@unesp.br

<https://orcid.org/0000-0002-9323-2489>

patrickallanfaustino / README.md

PROFILE VIEWS 917 REPO SIZE 1.8 MB

Hi guys, welcome! 🍌

Patrick Allan dos Santos Faustino
patrickallanfaustino · he/him

My name is Patrick Faustino, a PhD student researching molecular dynamics simulations and applying machine learning.

Edit profile

3 followers · 15 following

📍 São Paulo State University (UNESP), Institute of Chemistry, Campus Araraquara

🌐 Brazil

🕒 21:46 (UTC -03:00)

✉ patrickfaustino@hotmail.com

🔗 <http://lattes.cnpq.br/0920550077635421>

🔗 <https://orcid.org/0000-0002-9323-2489>

🐦 [@pkfaustino](https://twitter.com/pkfaustino)

🔗 <https://www.webofscience.com/wos/author/record/LTD-8752-2024>

🔗 https://scholar.google.com.br/citations?user=Eb_n1ziAAAAJ&hl=pt-BR

About Me

- 🎓 PhD student in Biotechnology at the Paulista State University, Brazil.
- 🎓 MSc student in Biomaterials and Bioprocess Engineering at the Paulista State University, Brazil.
- 🎓 BSc student in Computer Engineering at the Virtual University of the State of São Paulo, Brazil.
- 🔬 Working on Computational Chemistry, specifically Molecular Dynamics Simulations with Machine Learning.