

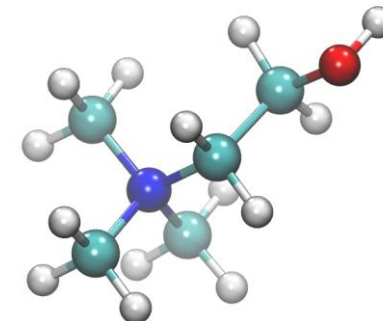
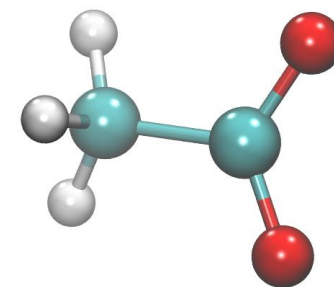
UNESP/IQ/AQA – Programa de Pós-  
Graduação em Biotecnologia



# **Simulação de Biomoléculas: dinâmica molecular básica**

Doutorando: Patrick Allan dos Santos Faustino

Orientadores: Prof. Dr. Fernando Masarin e 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, fosforilação etc.
- Método baseado na aplicação da 2ª Lei de Newton, Lei de Coulomb e Lei de Hooke.
- Softwares: GROMACS, NAMD, LAMMPS, OPENMM etc.

# Dinâmica molecular (MD)

- 1957: gases simples com Alder and Wainwright.
- 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. 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  $r_1(x_1, y_1, z_1)$ .
- É determinada a posição futura ( $t + \Delta t$ ), a velocidade e aceleração para cada átomo através de algoritmos (2).

$$v_i(t) = \frac{dr}{dt} \quad a_i(t) = \frac{d^2r}{dt^2}$$



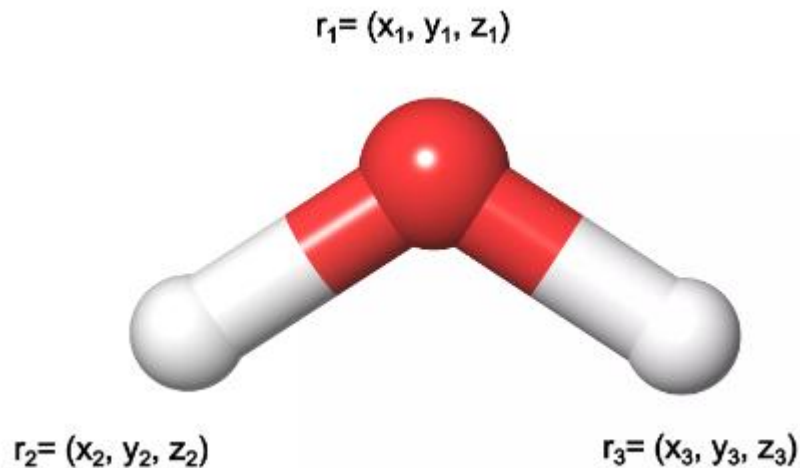
$$F_i(t) = m \cdot a_i = m \cdot \frac{d^2r}{dt^2} = - \frac{dU(r)}{dr}$$



$$a_i(t) = \frac{F_i}{m} = \frac{- \frac{dU(r)}{dr}}{m} = - \frac{dV}{dr} \cdot \frac{1}{m}$$



$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a_i(t)\Delta t^2$$



$$v_i(t) = \frac{dr}{dt}$$

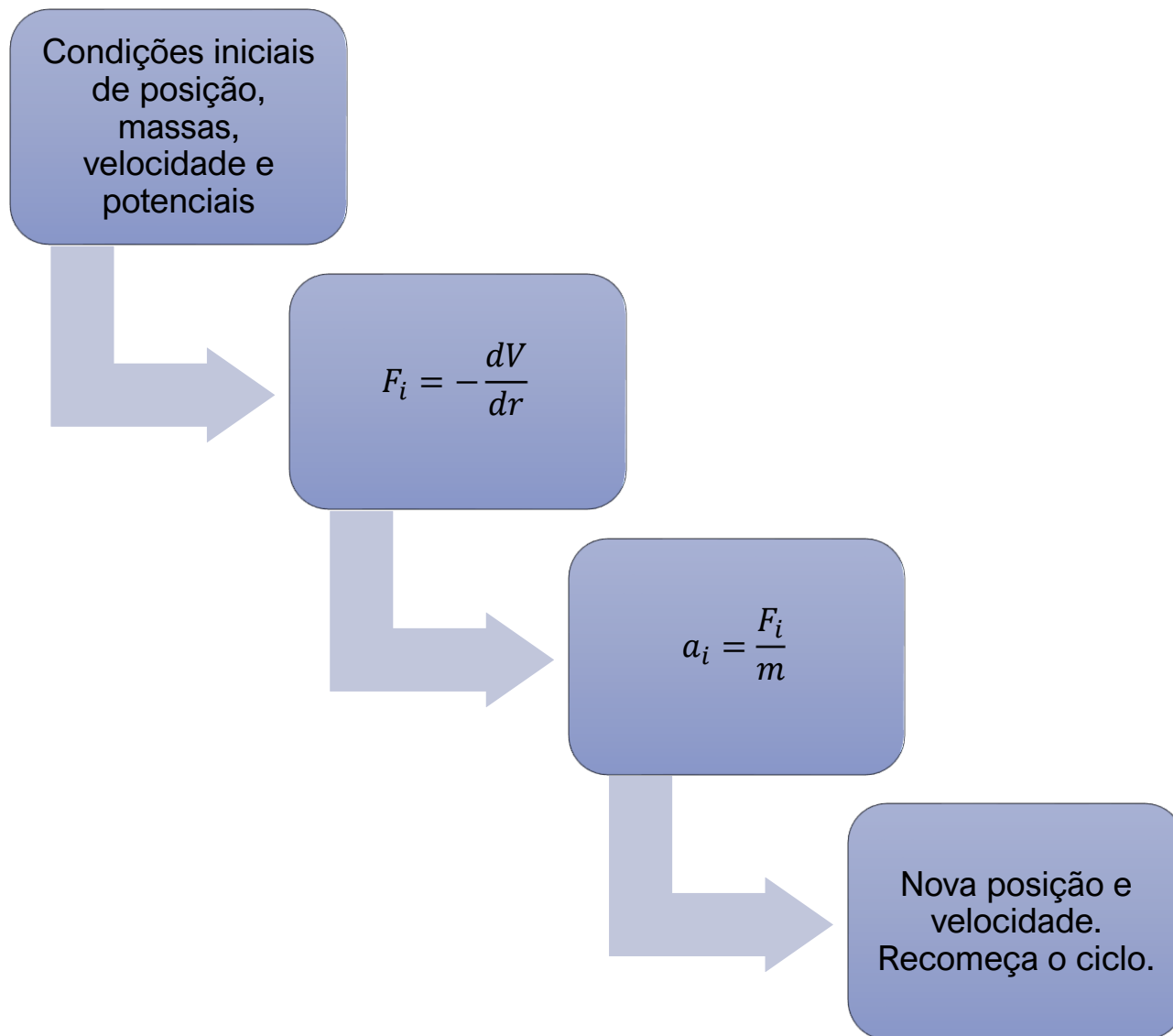
$$a_i(t) = \frac{d^2r}{dt^2}$$

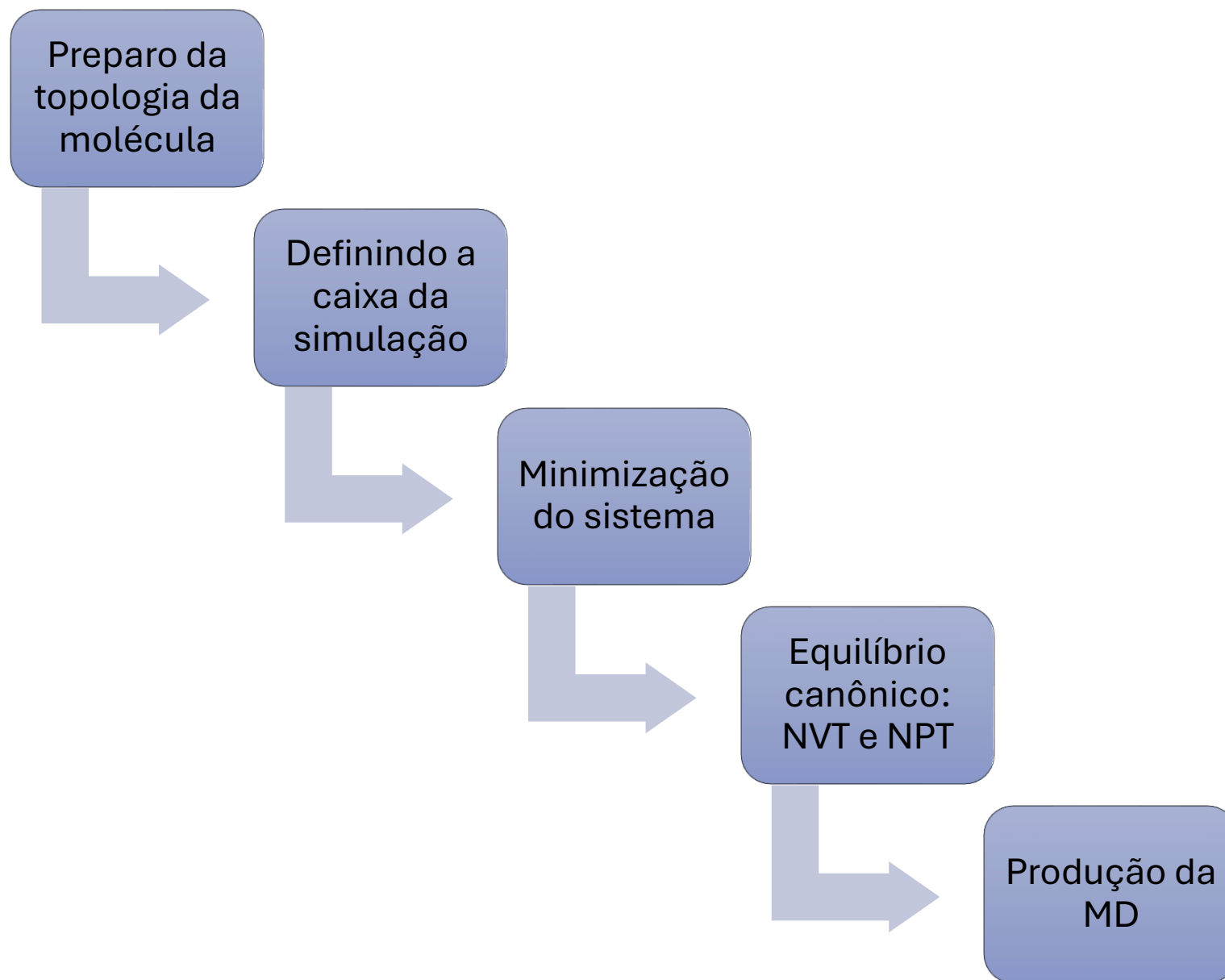


$$F_i(t) = m \cdot a_i = m \cdot \frac{d^2r}{dt^2} = -\frac{dU(r)}{dr}$$



$$a_i(t) = \frac{F_i}{m} = \frac{-\frac{dU(r)}{dr}}{m} = -\frac{dV}{dr} \cdot \frac{1}{m}$$

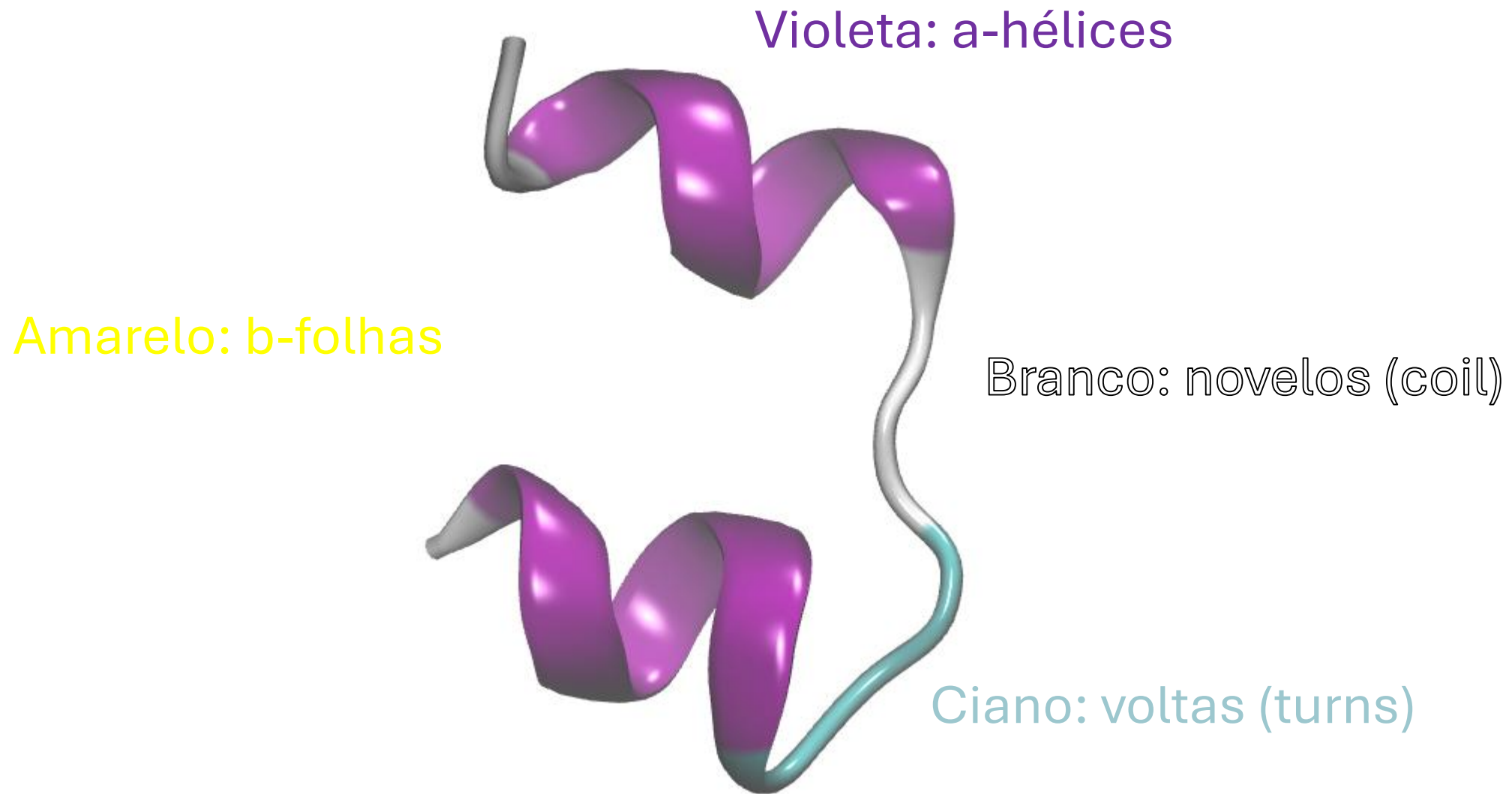




# Preparo da topologia da biomolécula

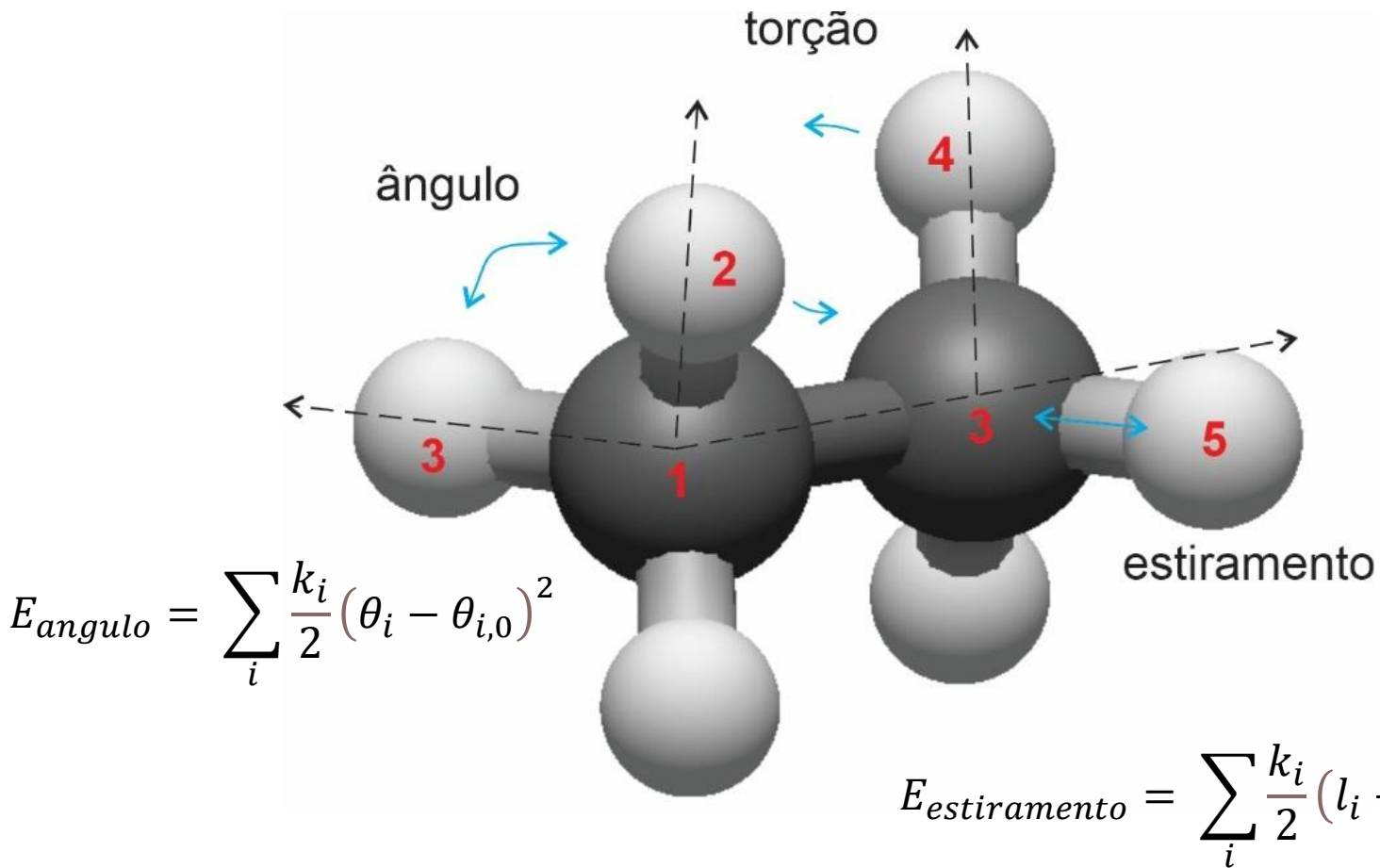
- Obtenção das coordenadas iniciais das moléculas de interesse (exemplo: PDB, PUBChem).
- Dê preferência a estruturas com resolução cristalográfica inferior a 2,5 Å.
- Campos de forças (4): CHARMM, AMBER, GROMOS e OPLS.
- Modelos de água (4): SPC, TIP3P, TIP4P e TIP5P.





Insulina humana, PDB ID 3i40 (5)

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



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

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

$$E = \sum_{\text{ligações}} E_{estiramento} + \sum_{\text{ângulos}} E_{angulo} + \sum_{\text{diedros}} E_{torção} + \sum_{\text{pares}} E_{não-ligação}$$

<b>Campo de Força</b>	<b>Informações</b>	<b>Modelo de água</b>	<b>cut-off</b>
<b>OPLS</b>	O campo de força OPLS-AA (Optimized Potentials for Liquid Simulations – All Atom) 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
<b>AMBER</b>	A família de campos de força AMBER (como amber99sb, amber99sb-ildn, amber14, etc.) é amplamente usada para proteínas, DNA/RNA e simulações biomoleculares.	TIP3P, não recomendado TIP4P e SPC.	1.0~1.2 nm
<b>CHARMM</b>	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
<b>GROMOS</b>	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

---

<b>Modelo</b>	<b>Tipo</b>	<b>Descrição</b>
<b>SPC</b>	3 pontos	Modelo rígido, ângulo fixo de $109.47^\circ$ , parametrizado para propriedades macroscópicas.
<b>SPC/E</b>	3 pontos	Versão estendida do SPC, com correção de energia de polarização. Melhor densidade e constante dielétrica.
<b>TIP3P</b>	3 pontos	Muito usado com AMBER e CHARMM. Simples e compatível com muitos campos de força.
<b>TIP4P</b>	4 pontos	Inclui ponto virtual (M-site) para carga negativa fora do oxigênio, melhorando propriedades de fase.
<b>TIP5P</b>	5 pontos	Dois pontos extra para os pares de elétrons do oxigênio. Mais preciso para estrutura tetraédrica, porém mais custoso.

# Preparo da topologia da biomolécula

```
gmx pdb2gmx -v -f 3i40_clean.pdb -o insulina.gro
```

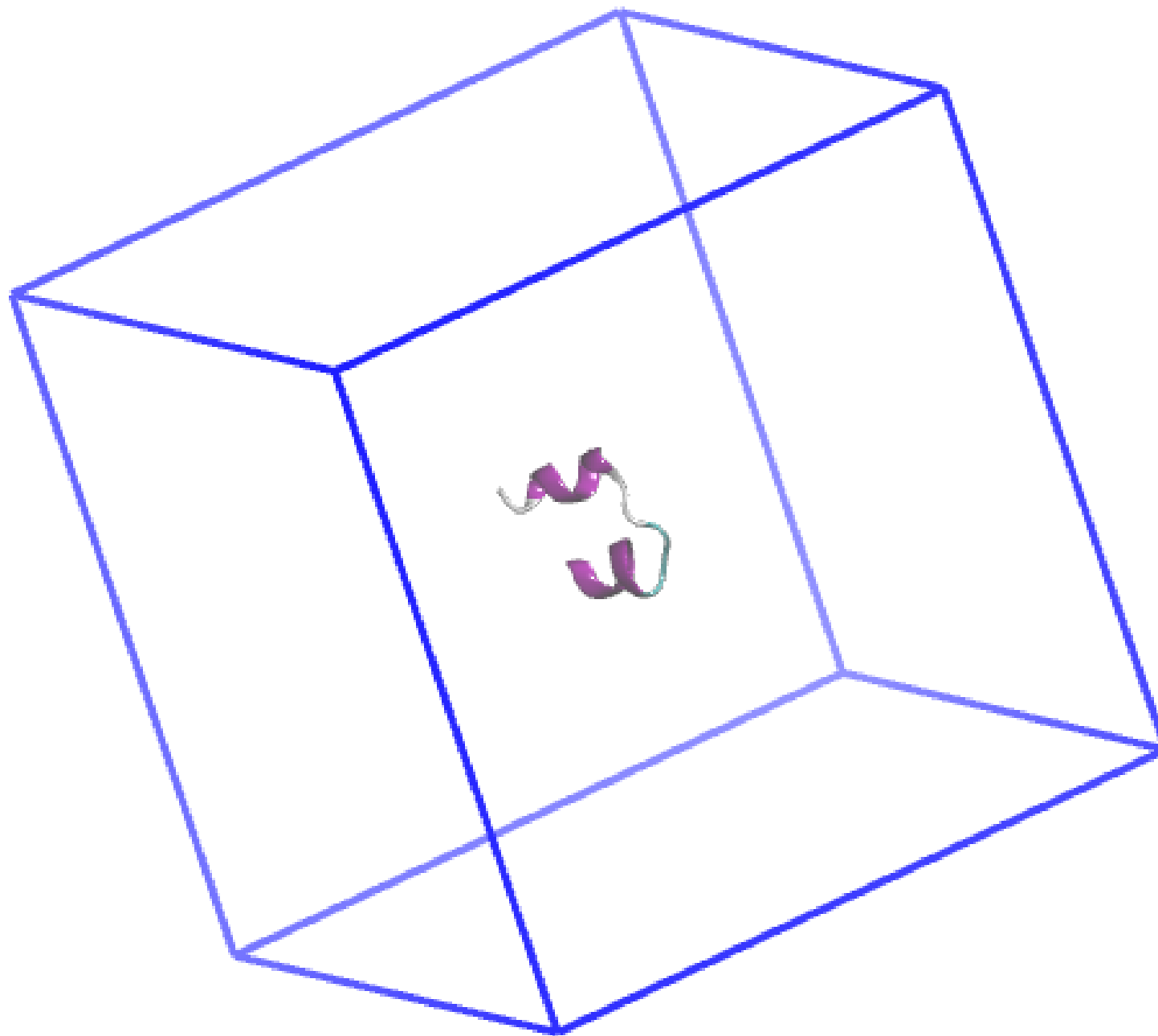
```
# -v = verbose, para visualizar o processo.
```

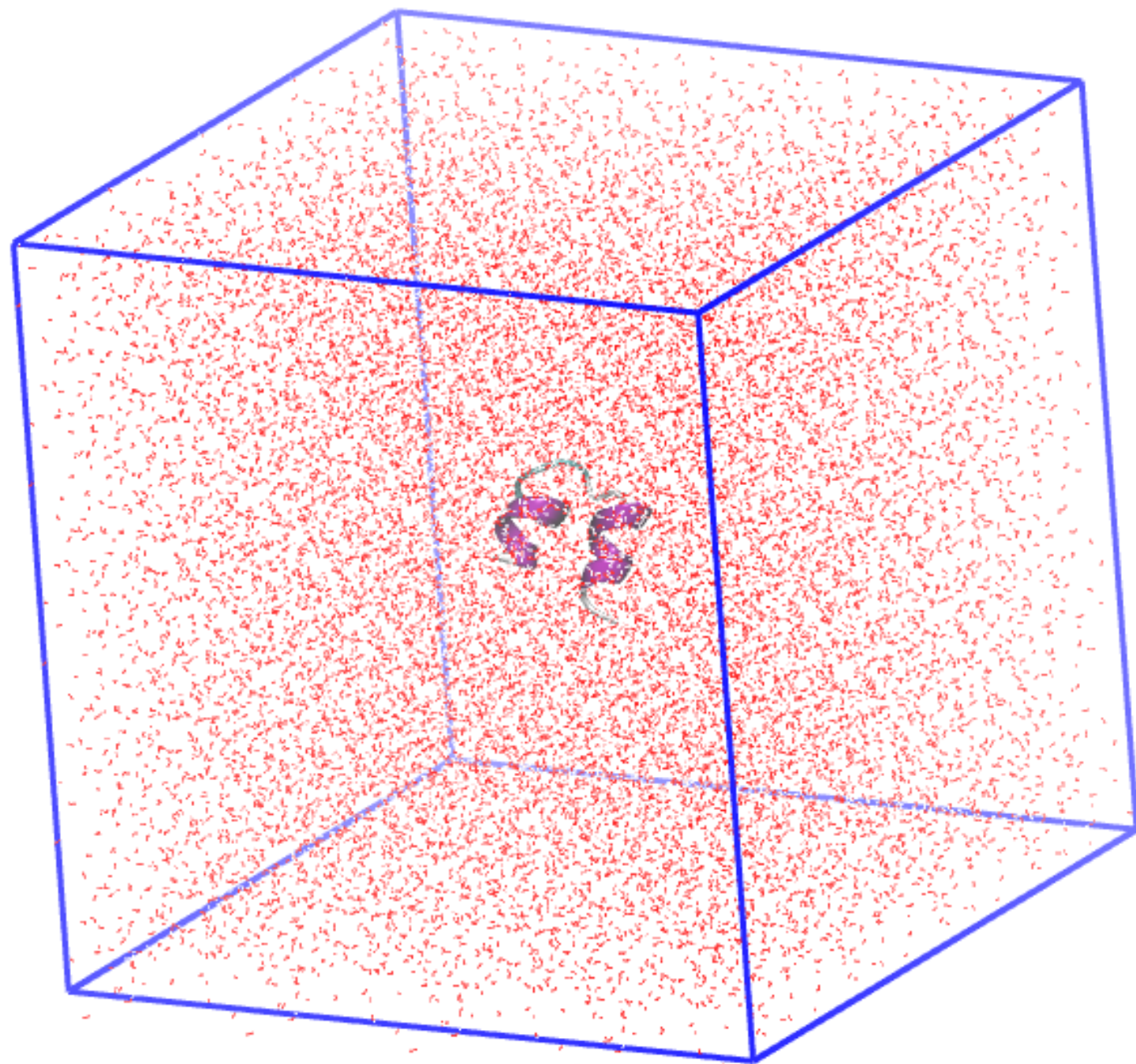
```
# -f = file input, arquivo de entrada das coordenadas.
```

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

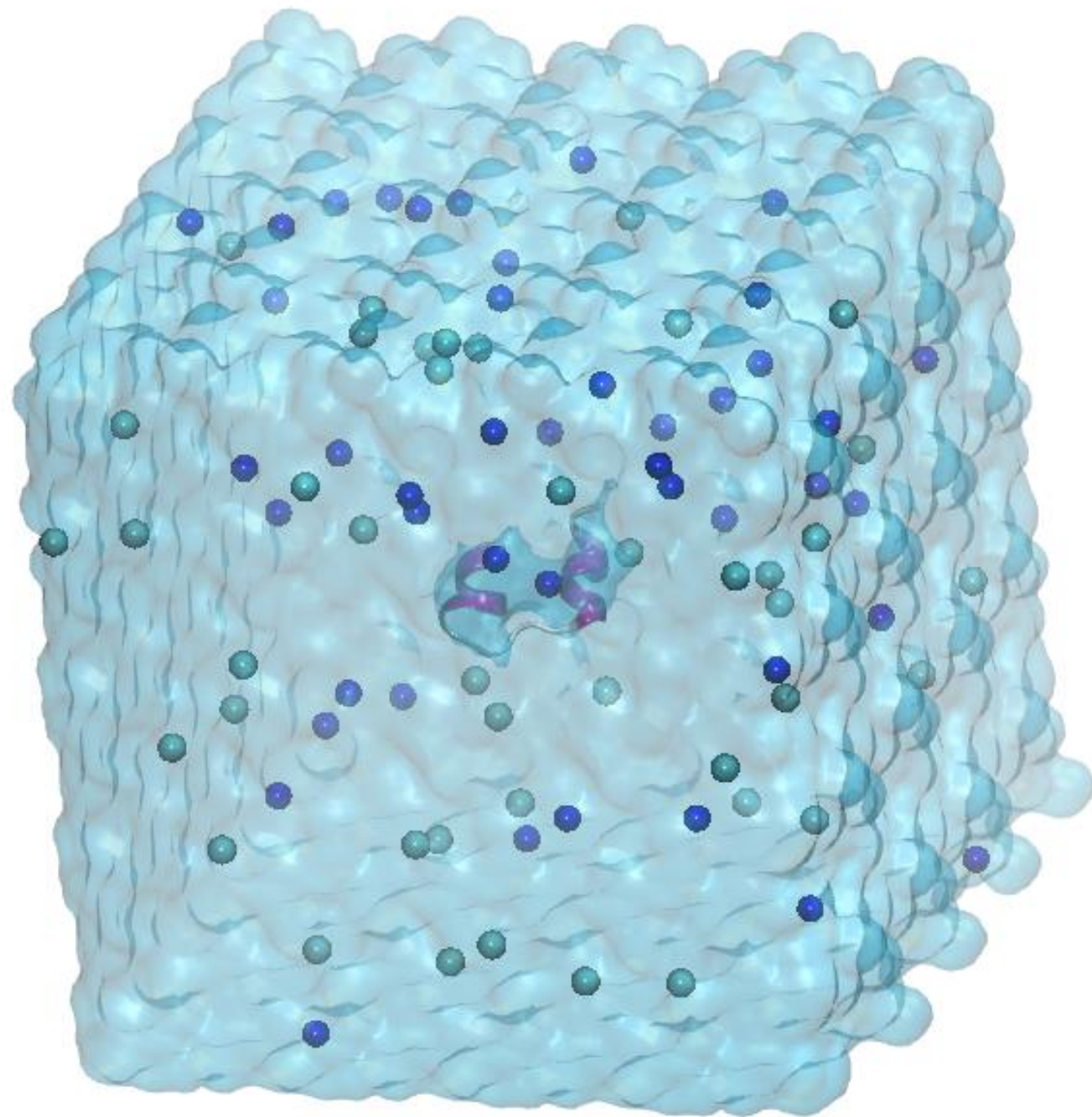
# Definindo a caixa de simulação (PBC)

- 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.









# Definindo a caixa de simulação (PBC)

```
gmx editconf -f insulina.gro -o box.gro -c -d 2.5 -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 solv.gro -p topol.top
```

```
# -cp = coordinates protein, coordenadas do nosso soluto  
(geralmente, proteína).
```

```
# -cs = coordinates solvent, coordenadas da molecula que será usada  
como solvente.
```

```
# -p = processing, para processar o arquivo de topologia do  
sistema.
```

# Definindo a caixa de simulação (PBC)

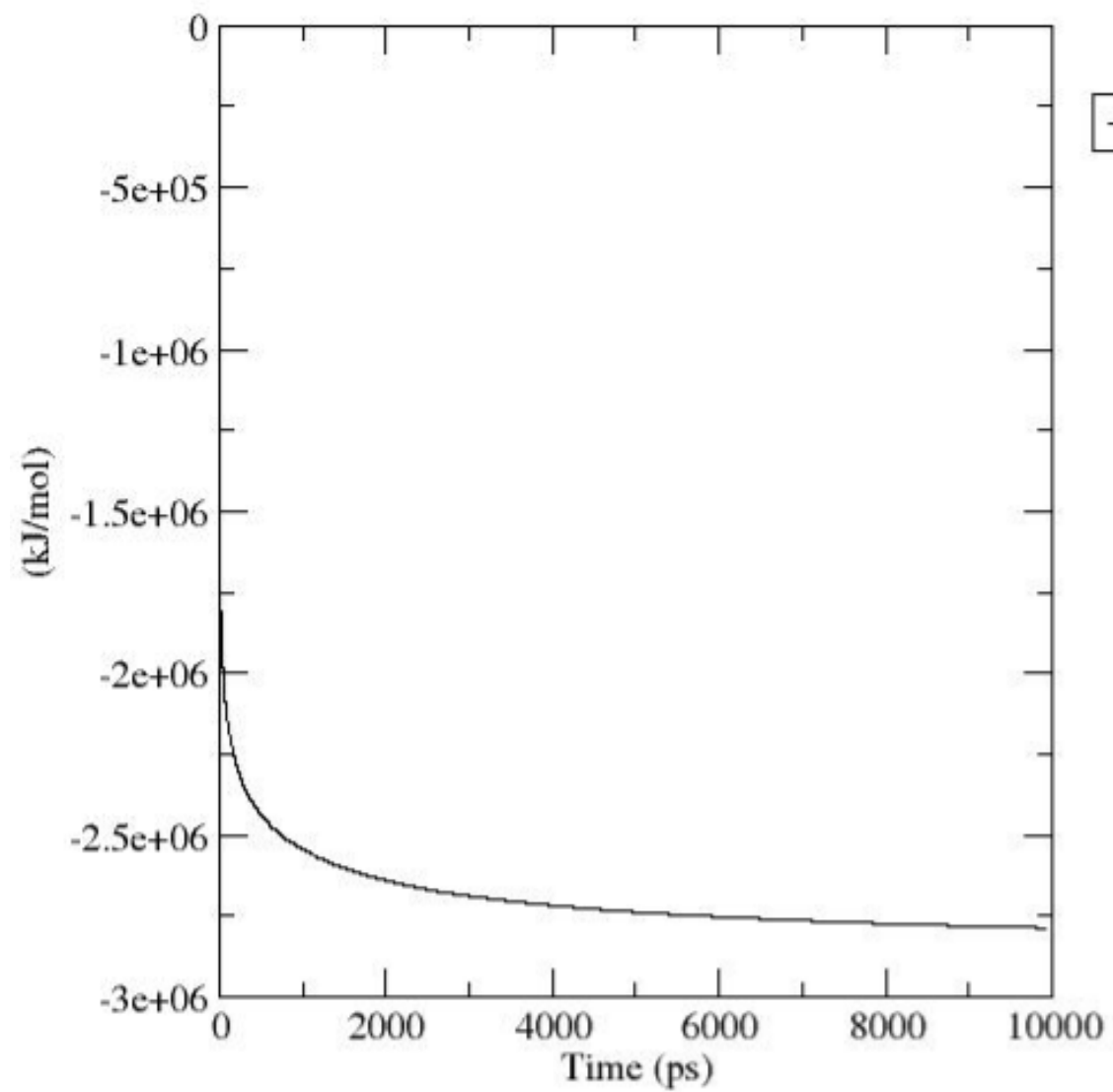
```
gmx grompp -v -f inputs/ions.mdp -o ions.tpr -c solv.gro -p topol.top  
  
# -c = coordenates, arquivo com as coordenadas do sistema.
```

```
gmx genion -s ions.tpr -o solv_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.

## GROMACS Energies



# Minimização do sistema

```
gmx grompp -v -f inputs/minim.mdp -c solv_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.
```

# minim.mdp

[Code](#)[Blame](#)

11 lines (11 loc) · 807 Bytes

[Raw](#)

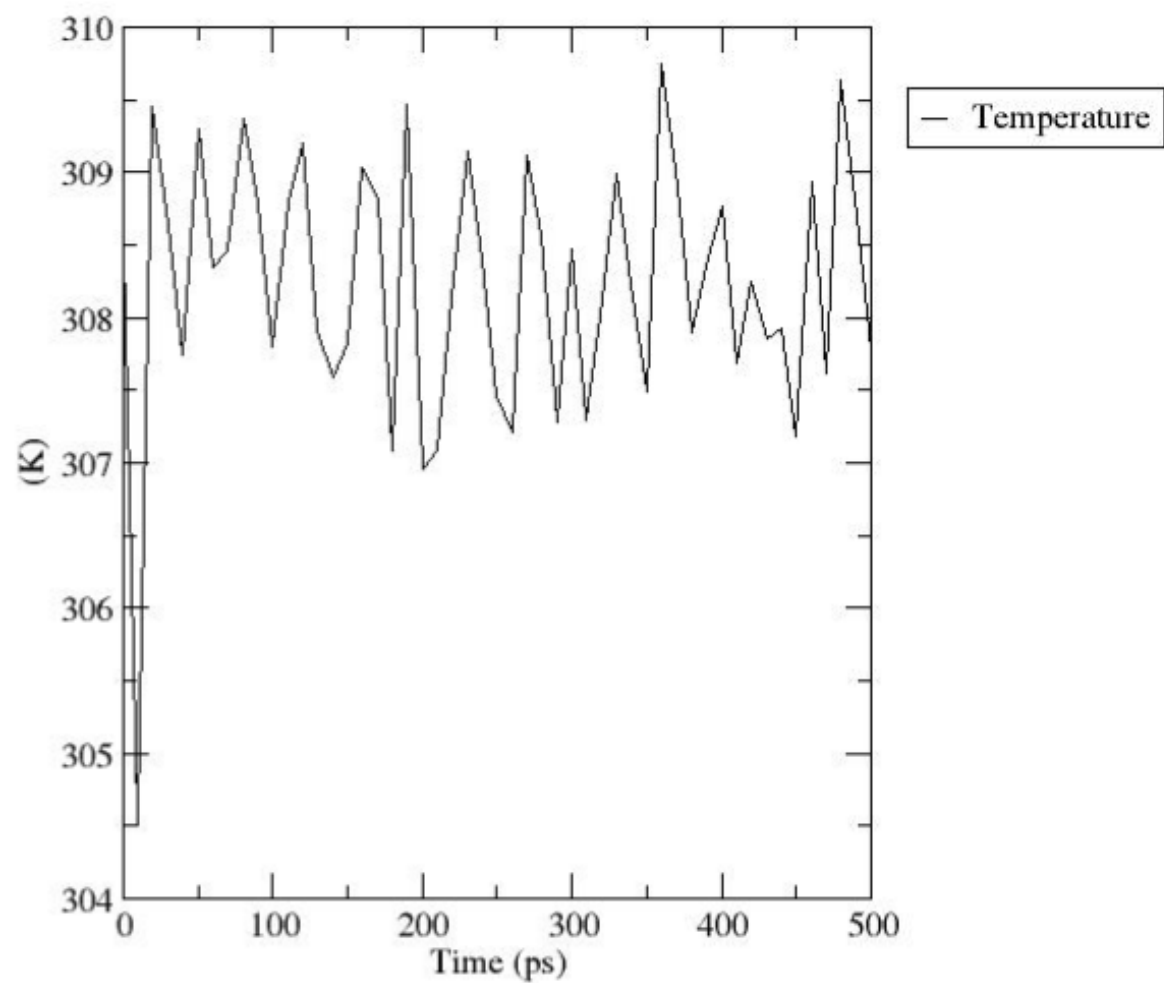
```
1  integrator = steep      ; Algorithm (steep = steepest descent minimization)
2  emtol      = 1000.0     ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
3  emstep     = 0.01      ; Minimization step size (nm)
4  nsteps     = 50000     ; Maximum number of (minimization) steps to perform
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
```

# Equilíbrio canônico: NVT e NPT

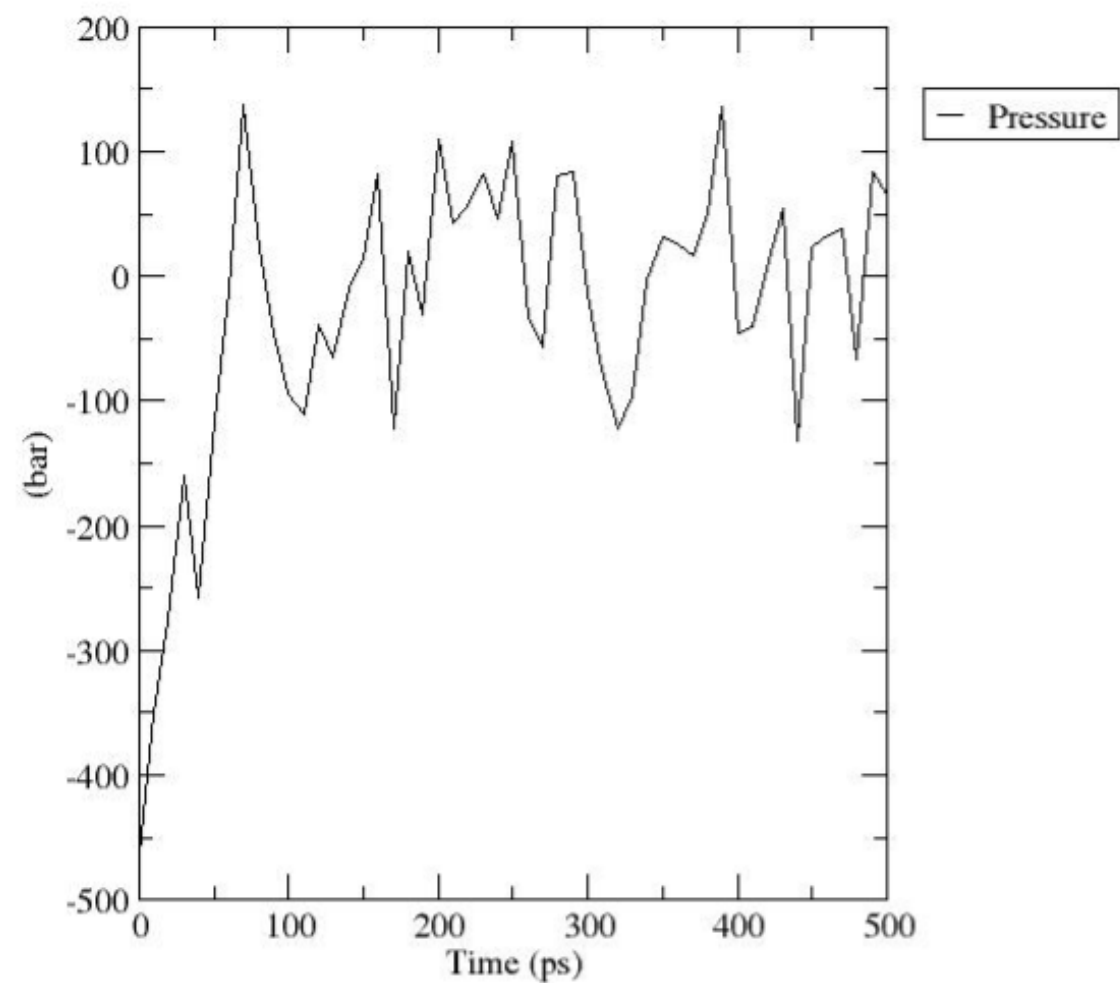
- Equilíbrio canônico NVT (6): número de partículas, volume e temperatura são mantidos constantes.
  - Berendsen, Nosé-Hoover, Andersen e V-rescale\*.
- Equilíbrio canônico NPT (6): número de partículas, pressão e temperatura são mantidos constantes.
  - Berendsen, Parrinello-Rahman e C-rescale\*.
-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



### GROMACS Energies



### GROMACS Energies



# 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
```

# 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
```

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

<b>Barostato</b>	<b>Características</b>	<b>Vantagens</b>	<b>Limitações</b>
<b>Berendsen</b>	Ajusta pressão rapidamente durante o equilíbrio	Simples, ideal para pré-produção	Não reproduz corretamente as flutuações canônicas
<b>Parrinello-Rahman</b>	Permite flutuações de volume e forma da caixa (NPT)	Correto para simulações de produção	Pode ser instável sem bom equilíbrio inicial
<b>C-rescale*</b>	Versão estocástica rigorosa de controle de pressão. Mantém flutuações canônicas corretas no ensemble NPT	Produz 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

# 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-veret, estocástico etc.
- Sem restrições.

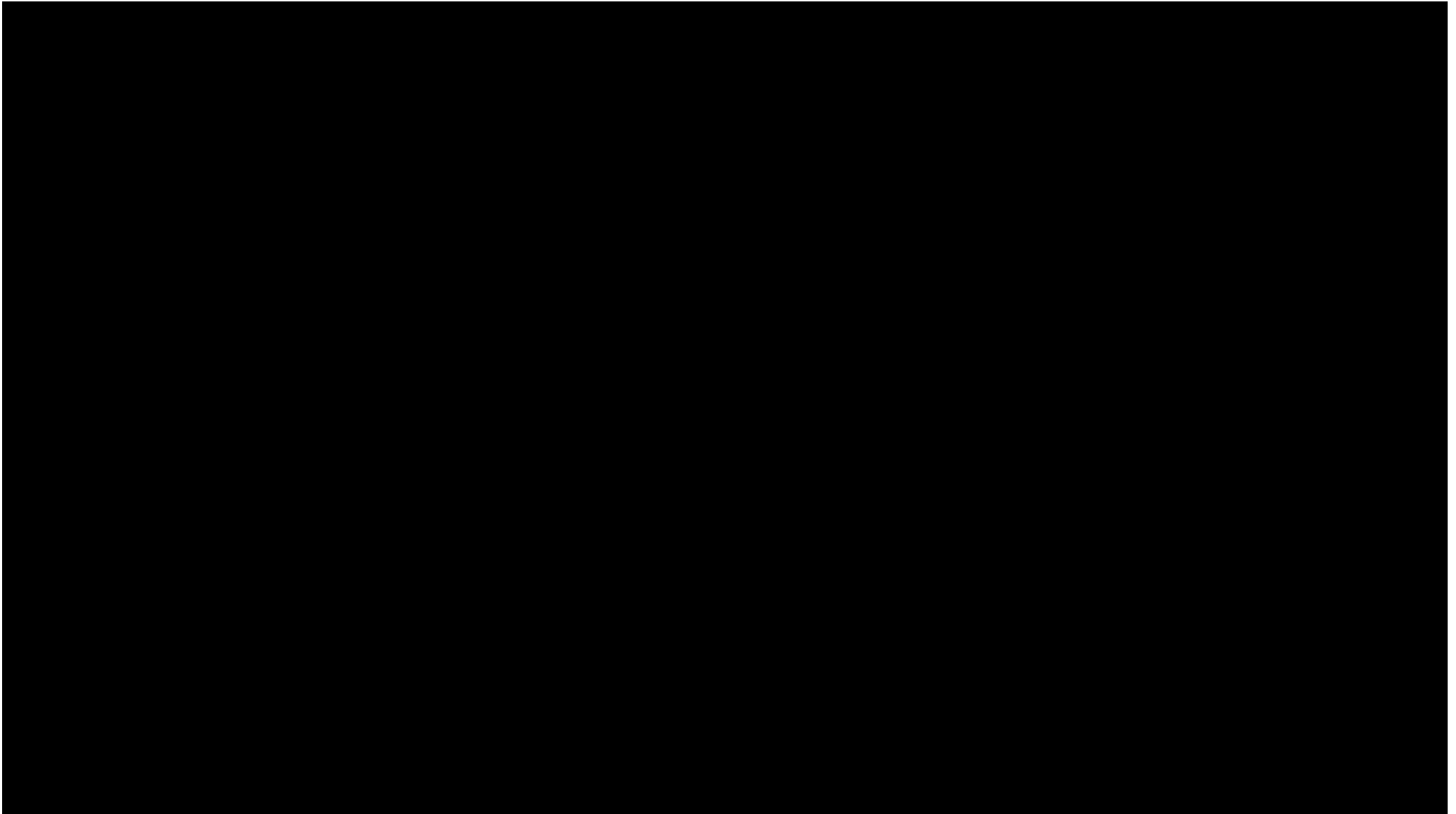
<b>Integrador</b>	<b>Características</b>	<b>Vantagens</b>	<b>Limitações / Quando evitar</b>	<b>Uso típico</b>
<b>md</b>	Leap-frog Verlet. Passo de tempo curto (1–2 fs). Conserva bem energia e momento.	Robusto, padrão, eficiente.	Velocidades não coincidem com posições.	Produção em proteínas, membranas,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





# Análise de resultados

- Analises 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 solvatação: 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)
```

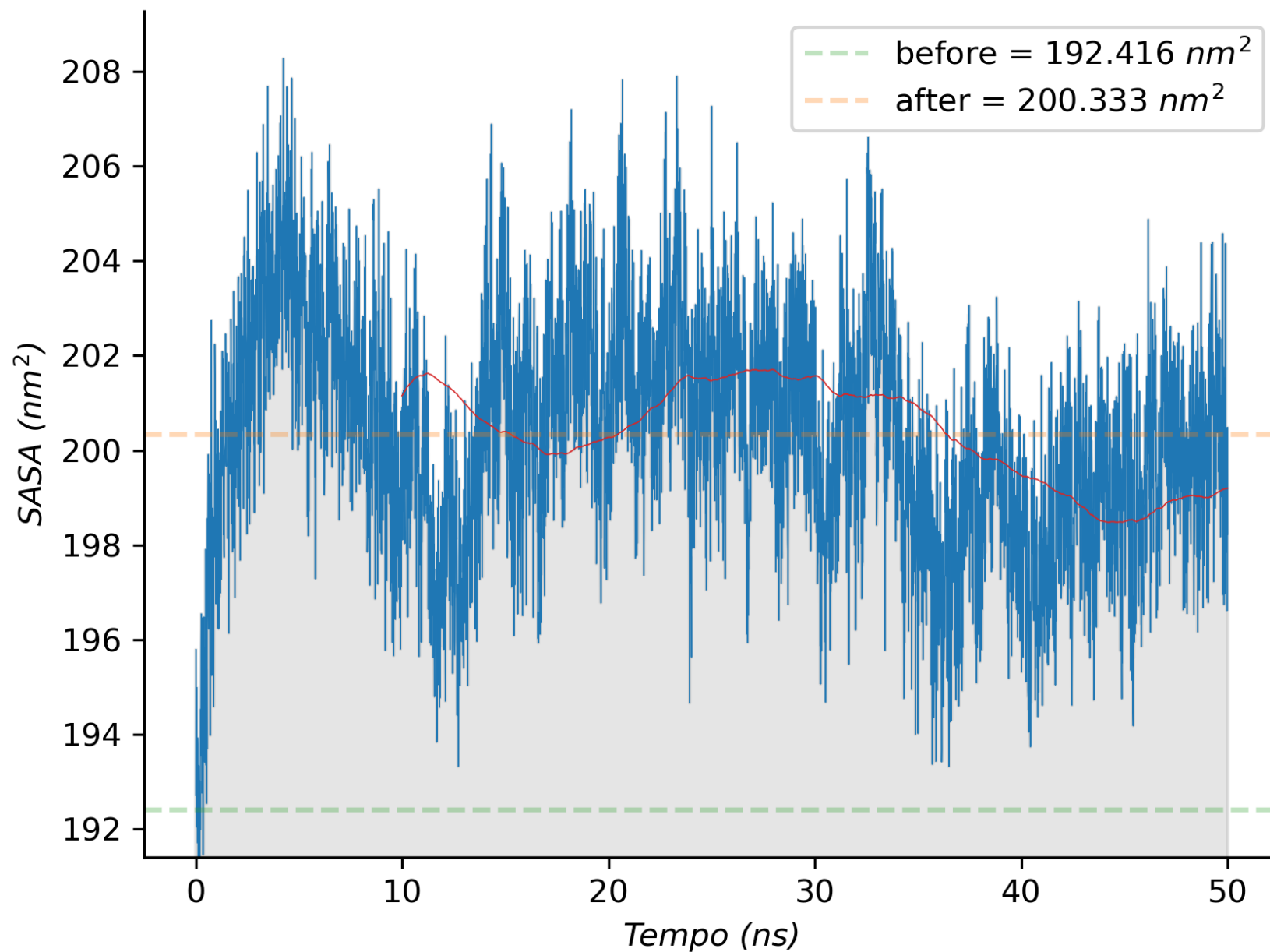
```
++++ PLEASE READ AND CITE THE FOLLOWING REFERENCE +++++
```

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

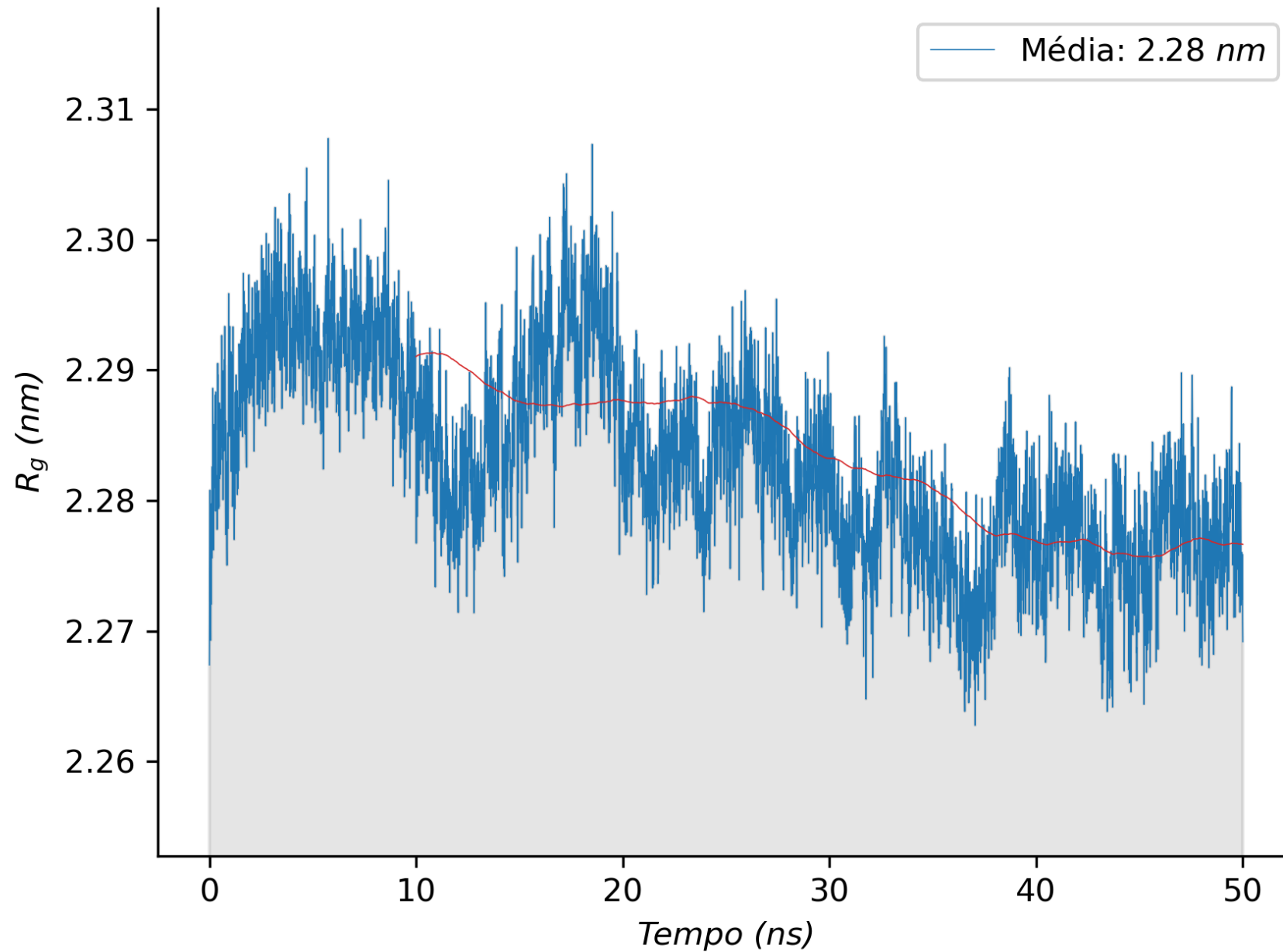
```
----- Thank You -----
```

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

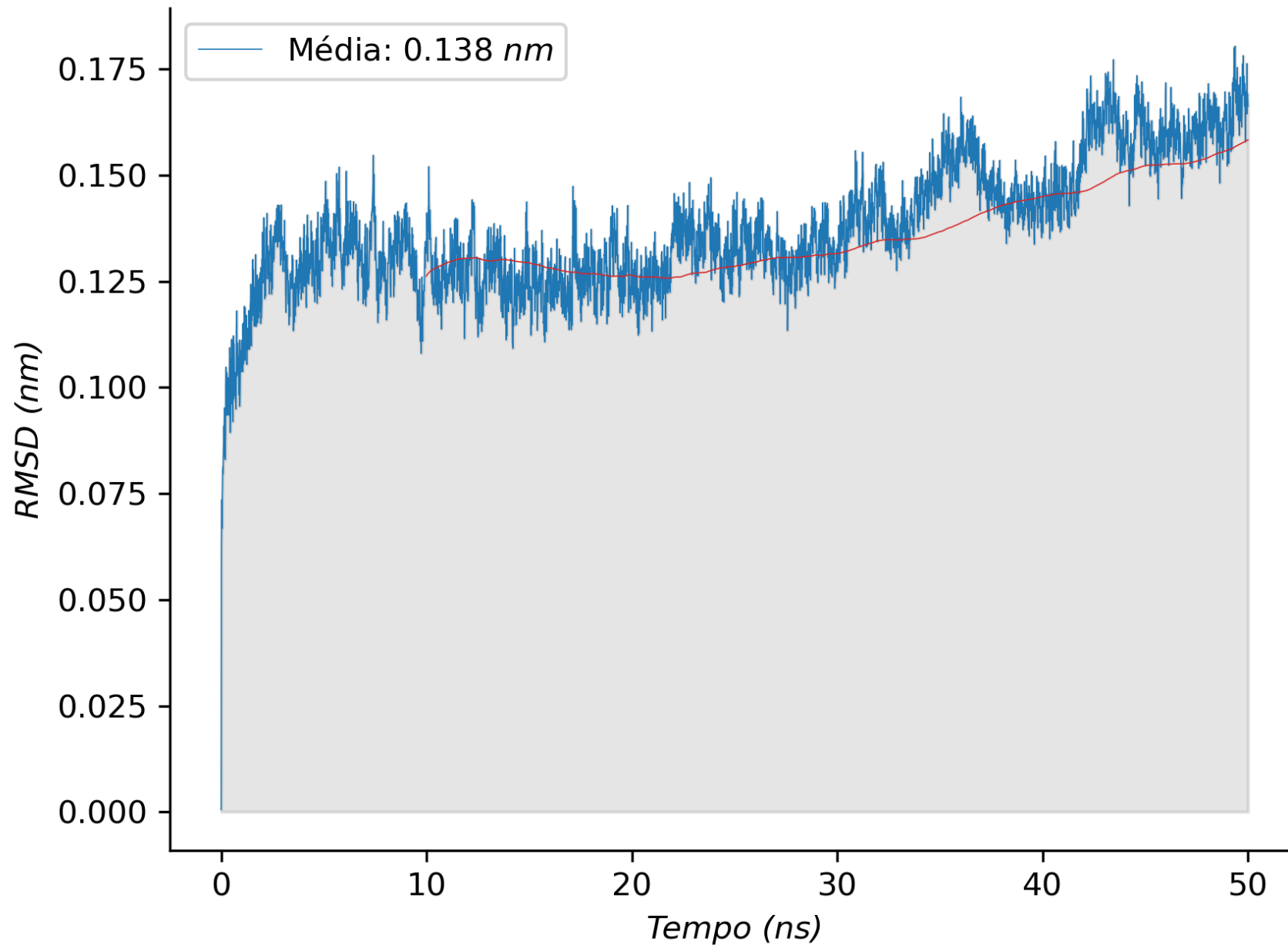
### SASA da proteína 1UKC/Acetato em fase aquosa



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

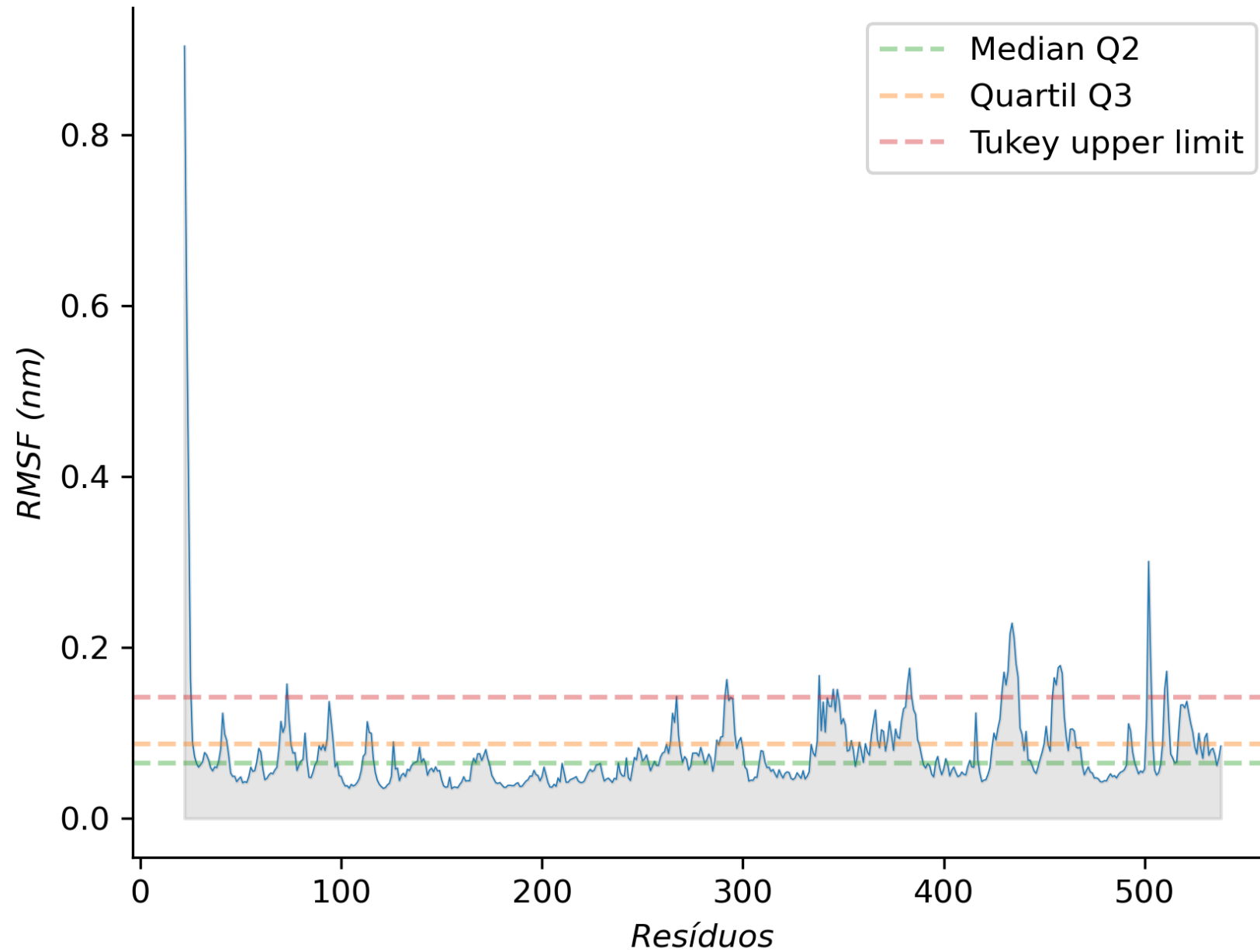


RMSD da proteína 1UKC/Acetato em fase aquosa (ref. backbone)

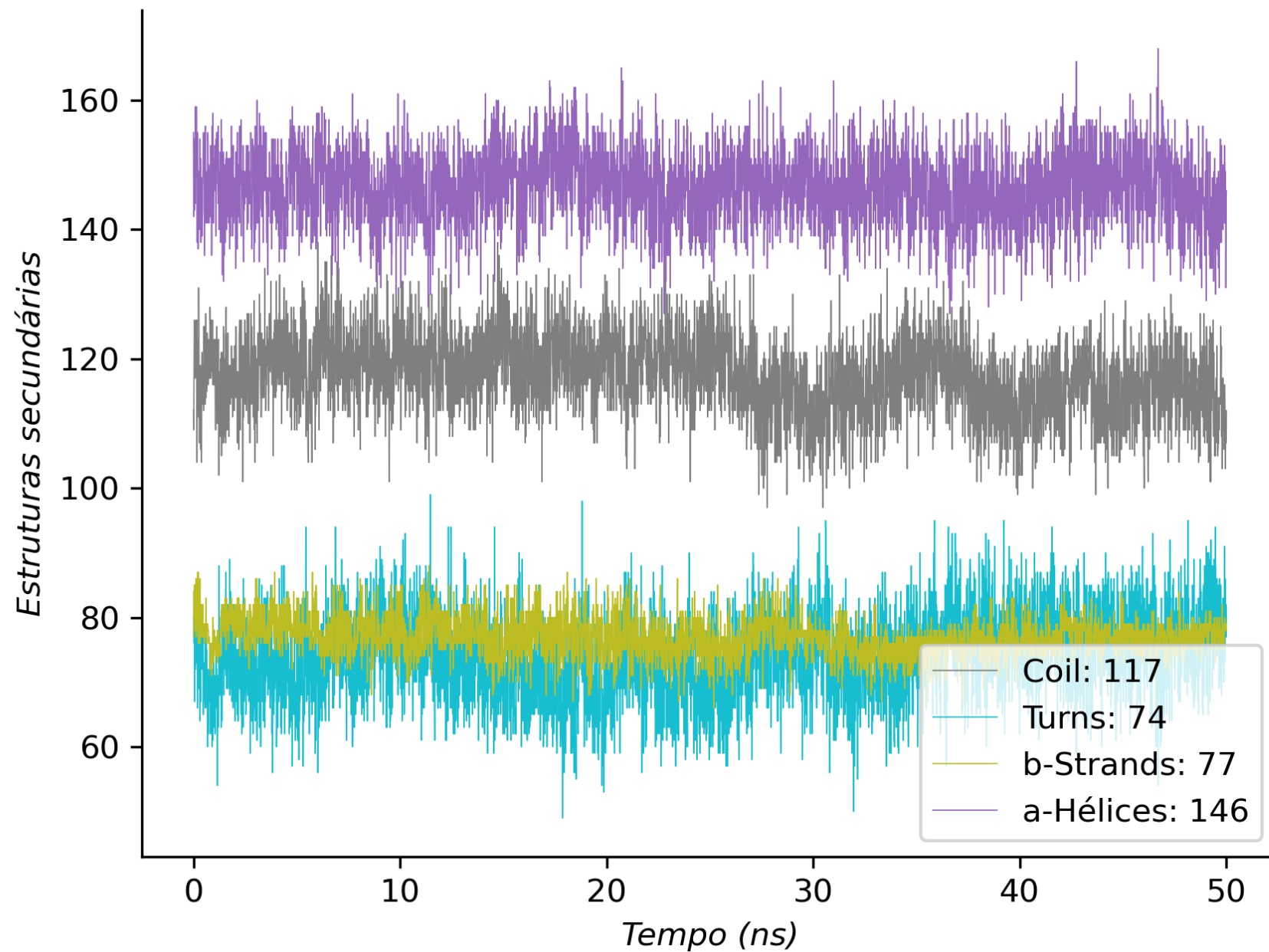




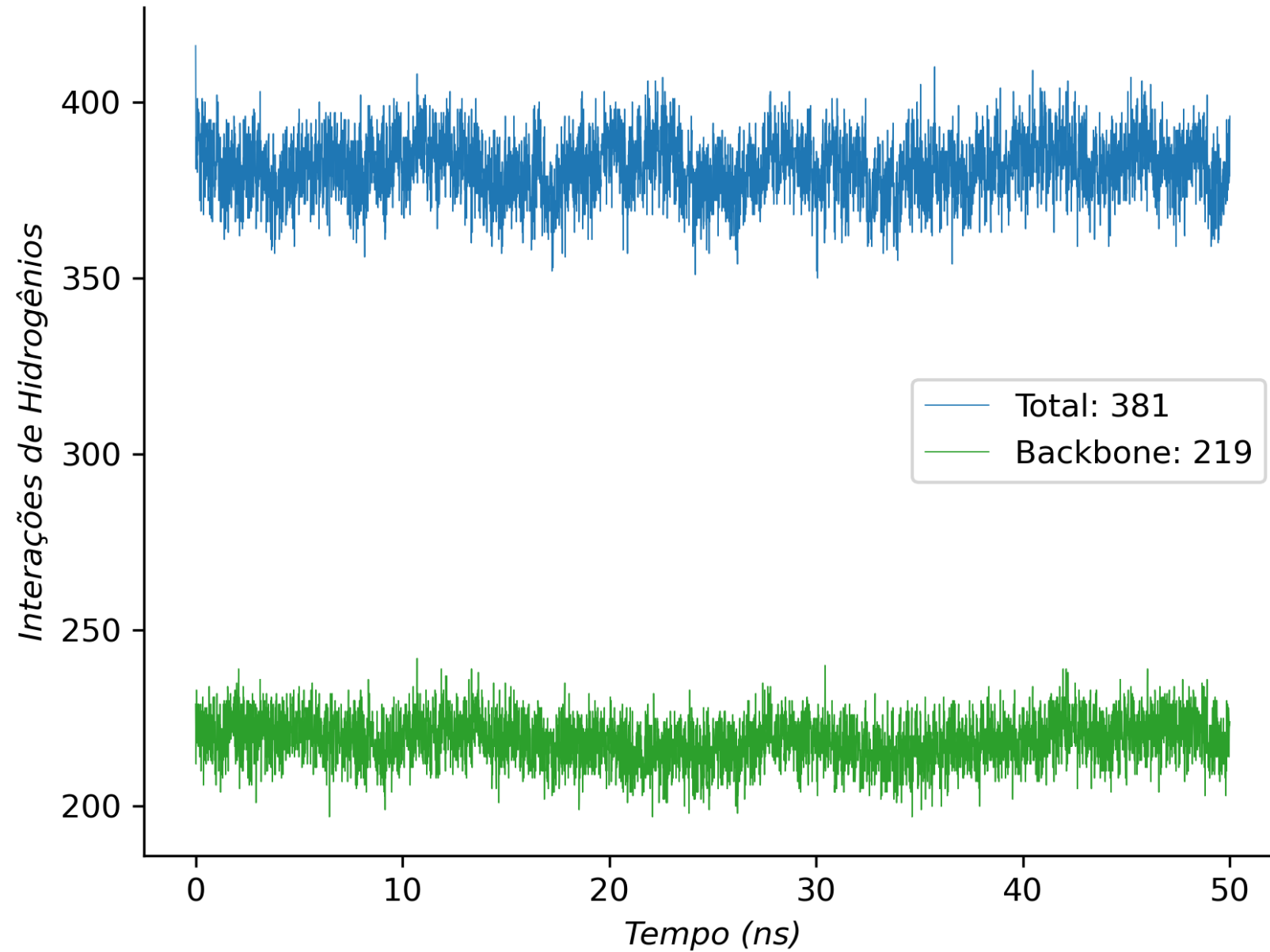
# RMSF da proteína 1UKC/Acetato em fase aquosa (ref. Ca)



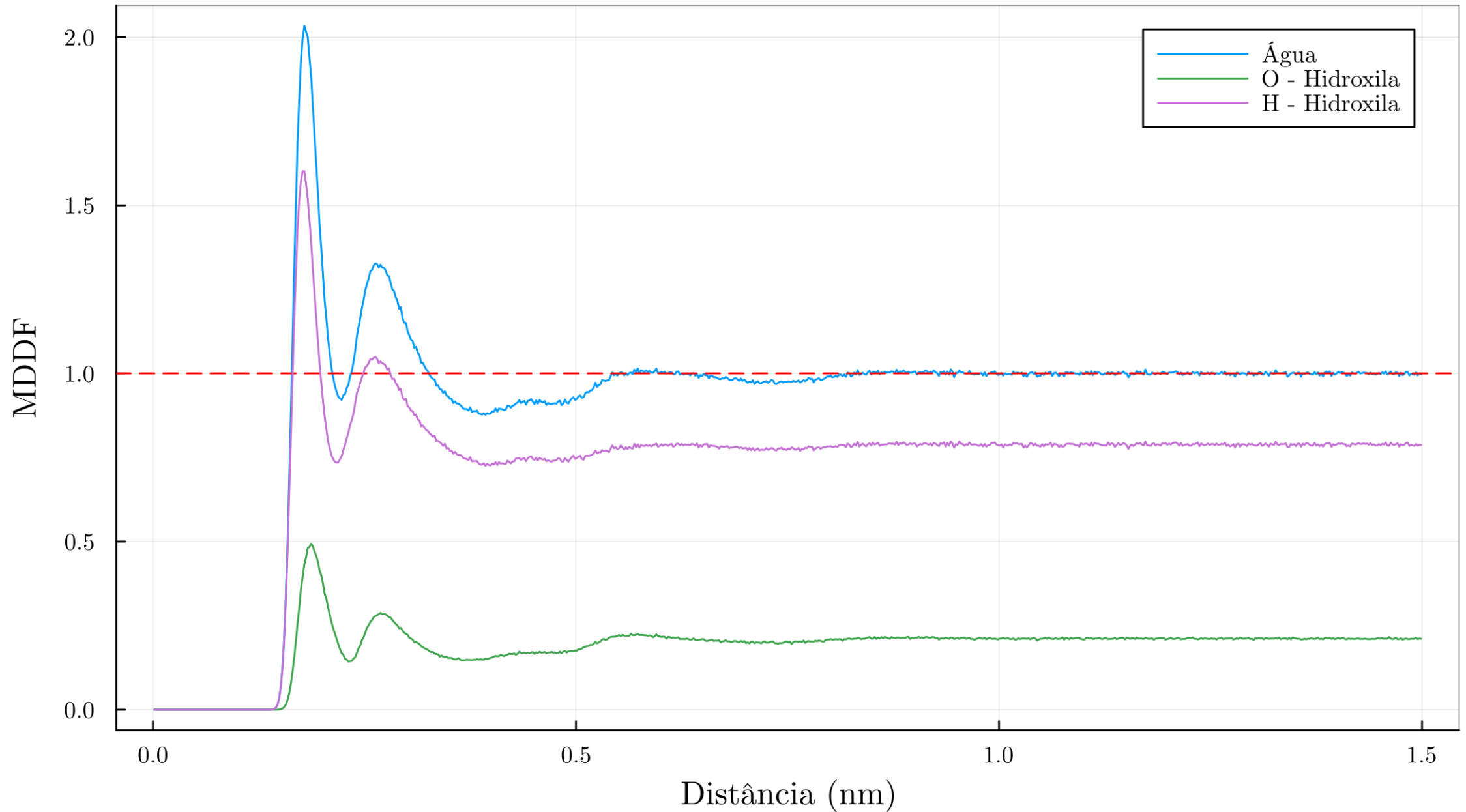
## DSSP da proteína 1UKC/Acetato em fase aquosa



# 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.

# Praticando...



[github.com/patrickallanfaustino](https://github.com/patrickallanfaustino)

[patrick.faustino@unesp.br](mailto: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! 🍌

**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.