Fine-tuning thermostats for coarse-grained simulations

Antonella Alba[1], Lucianna Silva dos Santos[1] , Andrés Camilo Ballesteros[1,2] , Martín Soñora[1] , Sergio Pantano[1,2]

[1]Institut Pasteur de Montevideo, Mataojo 2020, 11400, Montevideo, Uruguay,
[2]Área Bioinformática,Departamento DETEMA, Facultad de Química, Universidad de la República, Montevideo, Uruguay.
Corresponding author: spantano@pasteur.edu.uy

Melting peptides with SIRAH

What is this work about and which methods have been used?

When performing MD simulations, several aspects must be considered, forcefield of preference and solvent model in first place, but also when designing a protocol, among another relevant choices, the one of which thermostat to use, given the relevance of temperature control, specially if we want to reproduce melting temperatures.

Key aspects to consider regarding this work:

- Simulations were performed using SIRAH 2.0 as forcefield and AMBER20 as MD engine
- This assessment involves Langevin (varying its collision frequency value to fine-tune it) and Canonical Sampling Velocity Rescaling (CSVR) thermostats
- Molecular case studies chosen for this work were pure water, Crambin (PDB: 1CRN), Chignolin (PDB ID: 1UAO), SYR_4E_4, p31-43 (PDB ID: 6QAX)
- Simulations were performed at constant and increasing temperature

Performing coarse-grained molecular dynamics simulations (CG-MD) over protein systems modifying the thermostat choice to study its effects on protein dynamics at constant and increasing temperatures.

SIRAH
Southamerican Initiative for a Rapid and Accurate Hamiltonian

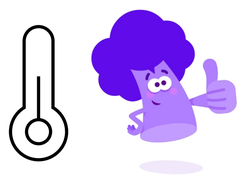
AMBER

CG- MD
SIRAH[1] FF and AMBER[2] as MD engine

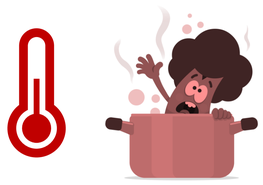
Langevin

CSVR

Langevin[3] and CSVR[4]
Are two stochastic thermostats implemented in AMBER



CONSTANT TEMPERATURE SIMULATIONS
For Crambin and p31-43, using Langevin and CSVR thermostats.



INCREASING TEMPERATURE SIMULATIONS
For WT4, SYR_4E_4 and Chignolin, using Langevin and CSVR thermostats.

ANALYSIS

VMD

SIRAH Tools[6]

Clustering analysis
CPPTRAJ[5] (AmberTools23)

Results

The molecular model of choice to perform an initial test was Crambin.
A cluster analysis was performed over the generated trajectories to compare both thermostats regarding their effect on conformational sampling (**Fig. 1**).

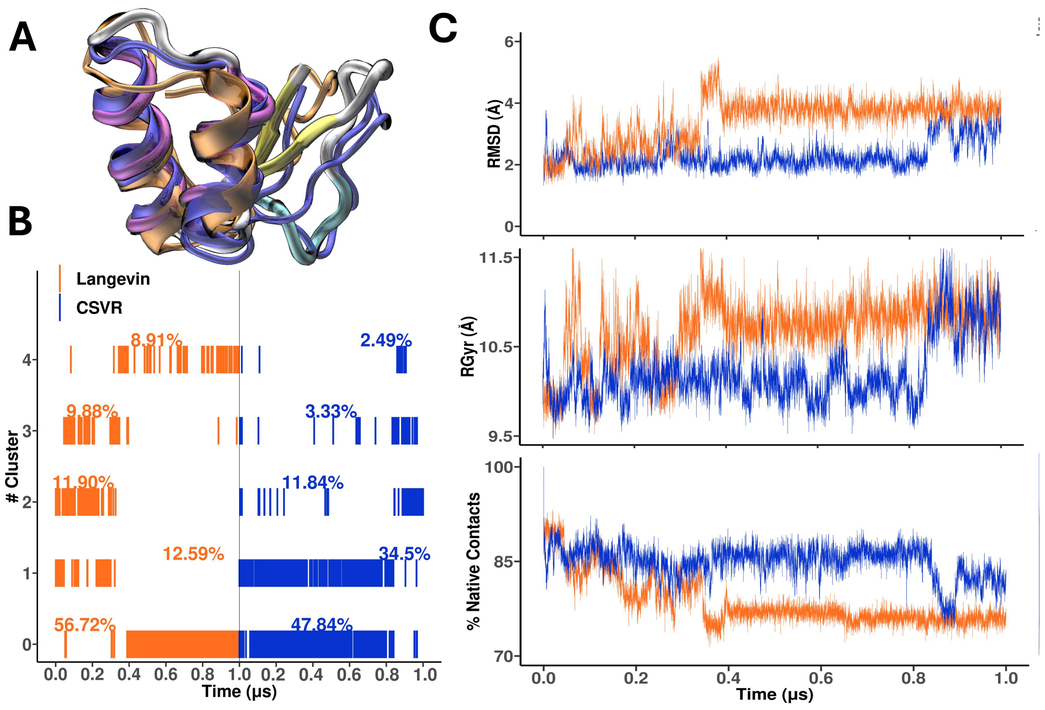


Fig. 1 – A) Three superposed 3D structures of Crambin. Experimental structure colored by secondary structure, and the representative frame of the most populated cluster for the dynamics performed employing Langevin (in orange) and CSVR (in blue). **B)** Cluster analysis performed using k-means algorithm comparing structures obtained from the Langevin (orange) and CSVR (blue) simulation based on their RMSD. **C)** RMSD, Radius of gyration and Native Contacts from the Langevin (orange) and CSVR (blue) simulation.

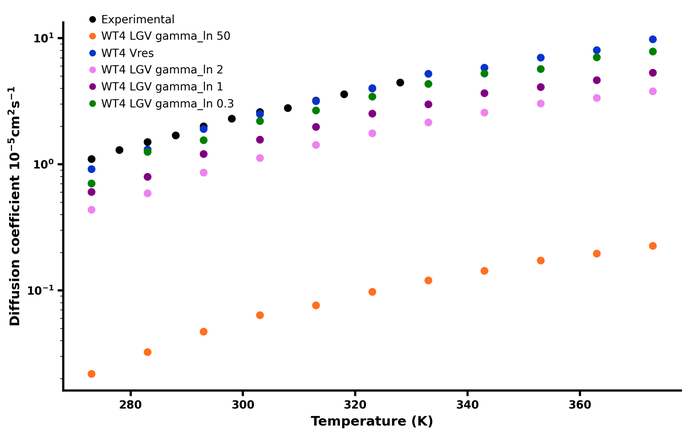


Fig. 2 – Water diffusion coefficient obtained from MD simulations of WT4 at increasing temperature for CSVR (denoted as Vres) and Langevin. For Langevin, several gamma_ln values were tested.

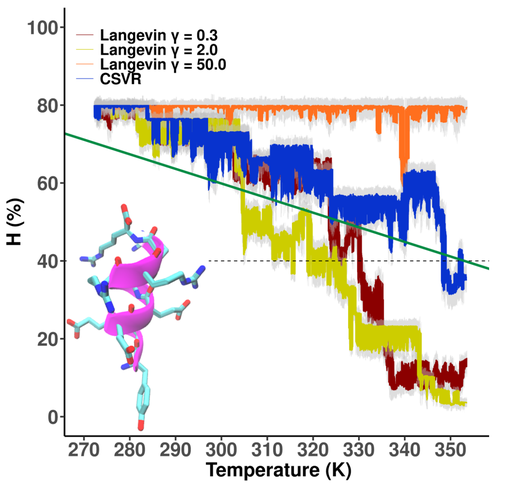


Fig. 3 – Linear increase of temperature on SYR_4E_4. In green, experimental data. Dotted line indicates cutoff for helical structure.

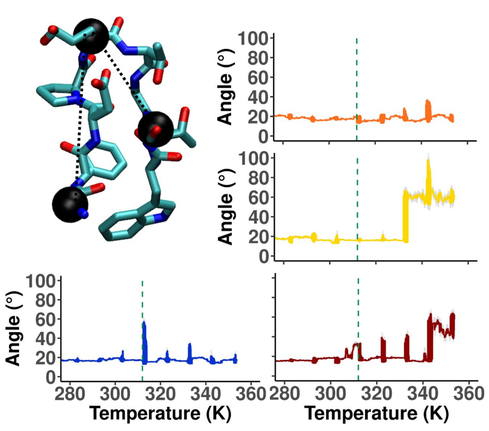


Fig. 4 – Temperature increase simulated over Chignolin. Tm = 312 K. The caption shows the angle selected as unfolding descriptor

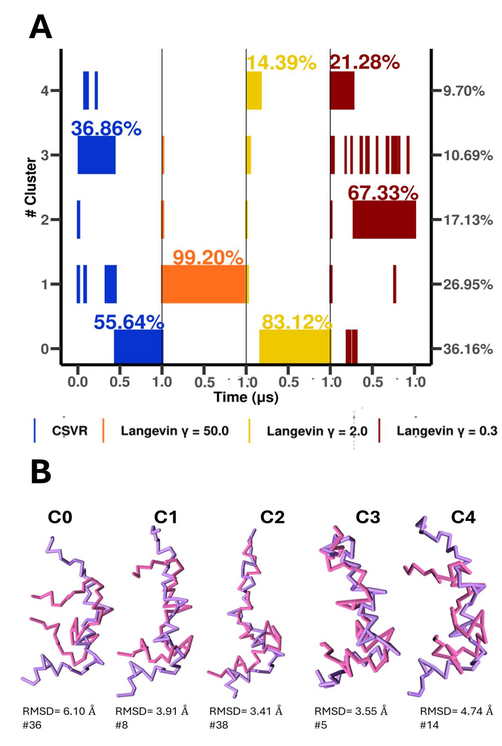


Fig. 5 – A) Cluster analysis performed using k-means algorithm comparing structures obtained from the Langevin and CSVR simulation based on their RMSD. **B)** 3D structures of the most representative frame for each cluster aligned against the NMR conformer with lower RMSD.

A parameter whose value must be stated when performing simulations with Langevin thermostat is γ, the collision frequency value which is denoted by the keyword gamma_ln on AMBER inputs. It has a damping effect on the velocities. The analysis performed for **Fig. 1** uses gamma_ln equal to $50s^{-1}$. We performed pure water simulations with WT4[7] as solvent model comparing the effect that gamma_ln value has on water diffusion **(Fig. 2)** and compare them with experimental data to fine-tune it. From that assessment it emerged that when gamma_ln equals $0.3s^{-1}$ the water diffusion coefficient is according to the experiment **(Fig. 2)**. So, from that point on, the analysis involved three different values for gamma_ln, $50s^{-1}$, $2.0s^{-1}$ (the value suggested in AMBER manual) and $0.3s^{-1}$.

Then we simulated peptides at increasing temperature **(Fig. 3** and **4)**, comparing our results with experimental data obtained from literature, for two systems, SYE_4R_4[8], **(Fig. 3)** and Chignolin[9] **(Fig. 4)**.
The next step was to assess the thermostat choice effect on IDPs conformational sampling, and the model case study of choice was p31-43 **(Fig. 5)**, a gliadin derived peptide previously studied by our group[10]. NMR studies have identified that p31-43 adopts preferentially a loose L-shaped conformation, so we choose to start our simulation from an unrealistic alpha helix to ensure the capacity of the force field to overcome rotational barriers and sample unstructured conformations.

References

[1] P. D. Dans, A. Zeida, M. R. Machado, S. Pantano, *J. Chem. Theory Comput*. **2010**, *6*, 1711–1725.
[2] Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D. et al., *J. Chem. Inf. Model.* 2023, 63, 6183–6191
[3] T. Schneider, E. Stoll, *Ferroelectrics* **1980**, *24*, 67–74.
[4] G. Bussi, D. Donadio, M. Parrinello, *The Journal of Chemical Physics* **2007**, *126*, DOI 10.1063/1.2408420.
[5] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput*. **2013**, *9*, 3084–3095.
[6] M. R. Machado, S. Pantano, *Bioinformatics* **2016**, *32*, 1568–1570.
[7] L. Darré, M. R. Machado, P. D. Dans, F. E. Herrera, S. Pantano, *J. Chem. Theory Comput.* **2010**, *6*, 3793–3807.
[8] R. F. Sommese, S. Sivaramakrishnan, R. L. Baldwin, J. A. Spudich, *Protein Science* **2010**, *19*, 2001–2005.
[9] S. Honda, K. Yamasaki, Y. Sawada, H. Morii, *Structure* **2004**, *12*, 1507–1518.
[10] M. F. G. Castro, E. Miculán, M. G. Herrera, C. Ruera, F. Perez, E. D. Prieto, et al., Front Immunol **2019**, *10*, DOI 10.3389/fimmu.2019.00031.

Conclusion

- Both CSVR and Langevin can sample the same conformational space, but Langevin provokes a damped dynamical behavior that strengthens as the gamma_ln value increases.
- Thermostat choice can modulate the kinetics of the protein.
- SIRAH FF can display unfolding and is able to sample unstructured conformations.
- For the protocol and systems studied in simulations at increasing temperature, CSVR behaves according to the experimental data.

Acknowledgement

UNIVERSIDAD DE LA REPÚBLICA URUGUAY
FACULTAD DE QUÍMICA
Institut Pasteur de Montevideo
SIRAH Southamerican Initiative for a Rapid and Accurate Hamiltonian
ANII AGENCIA NACIONAL DE INVESTIGACIÓN E INNOVACIÓN
PEDECIBA MEC-UDELAR
FOCEM