

# CompactMD: AI-Driven Compression of Molecular Trajectories

Rubén Adiel Luna Medina,<sup>†</sup> Thanh Duy Cao,<sup>†</sup> Srishti Karanth,<sup>†</sup> Sebastien Lafond,<sup>†</sup> Hergys Rexha,<sup>†</sup> and Marc Baaden<sup>‡</sup>

<sup>†</sup>Faculty of Science and Engineering, EDISS, Åbo Akademi University, Finland

<sup>‡</sup>Institut de Biologie Physico-Chimique, CNRS, Laboratoire de Biochimie Théorique, France

## 01 Introduction

**Molecular Dynamics (MD)** is a computational technique that simulates the physical motion of atoms and molecules over time to analyze structural, dynamical, and interaction properties of molecular systems.



Biomolecular study

Use cases



Drug discovery

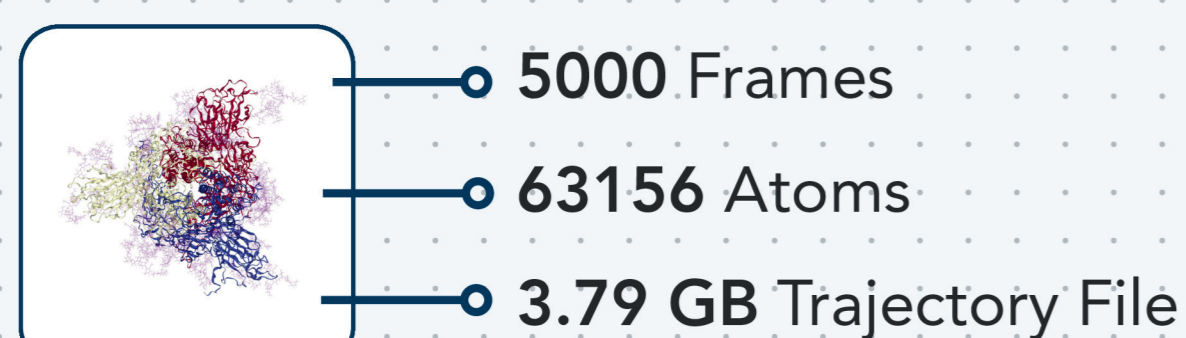
Motivation :

- **CompactMD** explores an AI-driven approach that learns the structural and temporal properties of MD trajectories to efficiently compress large molecular dynamics trajectories to make storage, sharing, and analysis easier.



## 02 Benchmarking

1 Traditional (XTC, LCP) and AI-based (**compressTraj**) compression techniques were benchmarked with:



SARS-CoV-2 spike glycoprotein

Lower the time, better

Algorithm	Compression	Decompression
XTC	29.54 ± 0.13	29.61 ± 1.18
LCP	34.46 ± 0.41	11.46 ± 0.55
compressTraj	57.36 ± 2.58	6.00 ± 0.01
compressTraj + training	31205.6 ± 233.60	6.00 ± 0.01

Table 1 : Compression and Decompression Time (s) of Trajectory Files

2 Performance was evaluated using **compression/decompression time** (Table 1) for efficiency, **compression ratio** (Fig. 1) for size reduction, and **RMSD** (Fig. 2) for structural preservation.

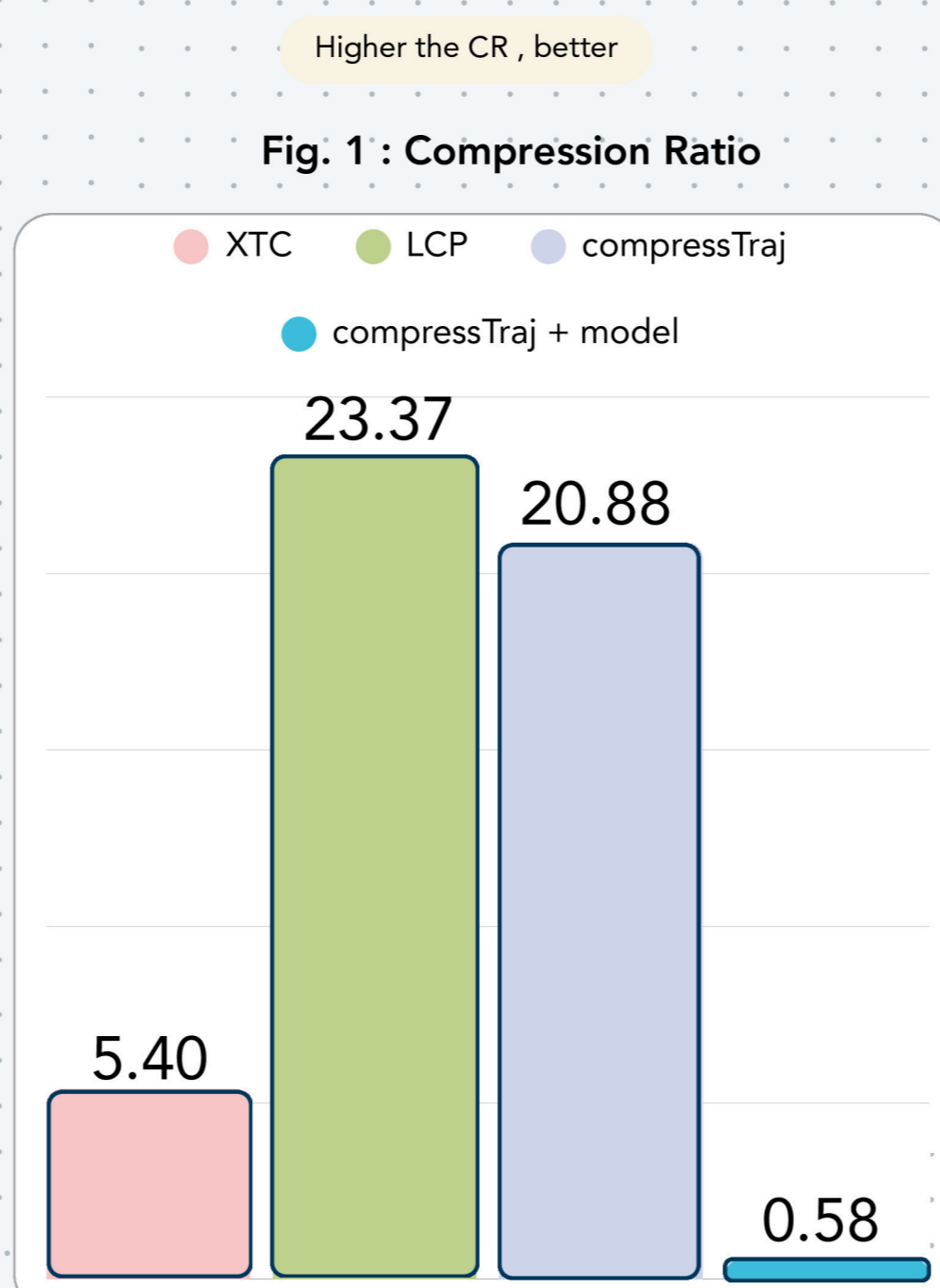


Fig. 1 : Compression Ratio

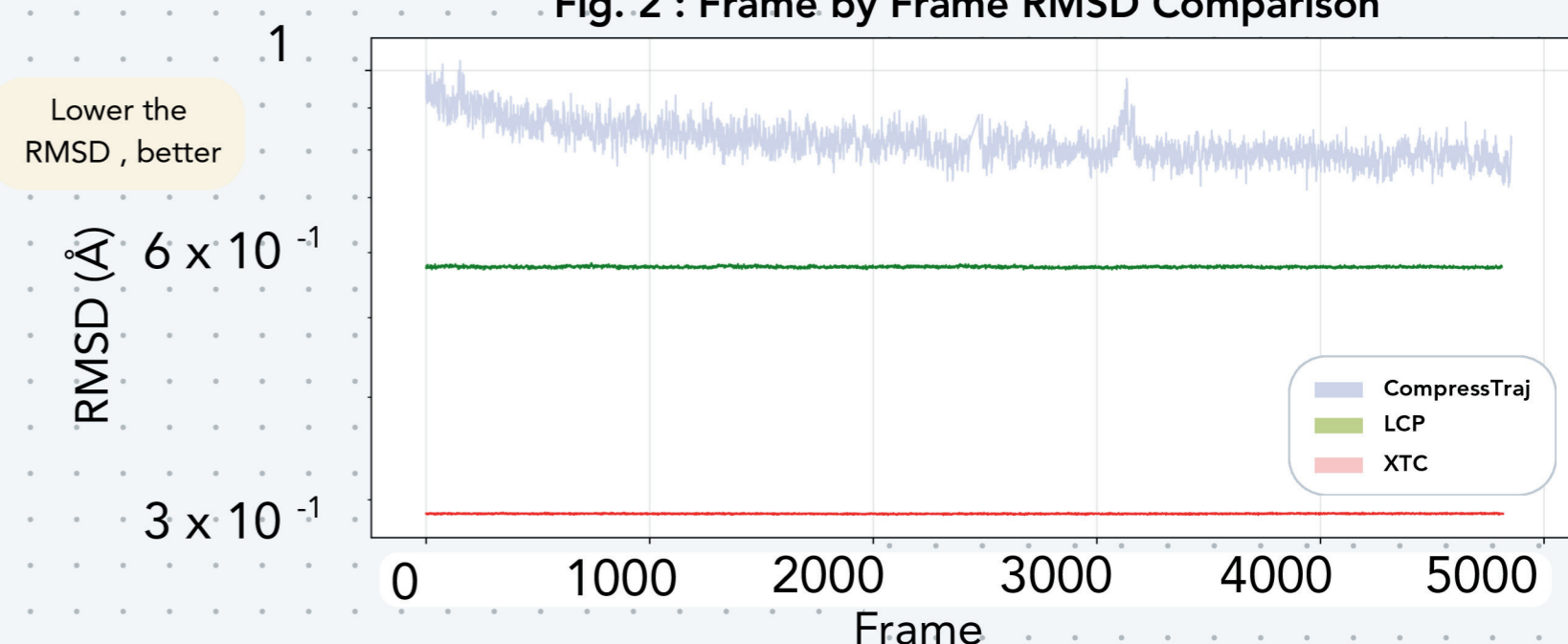
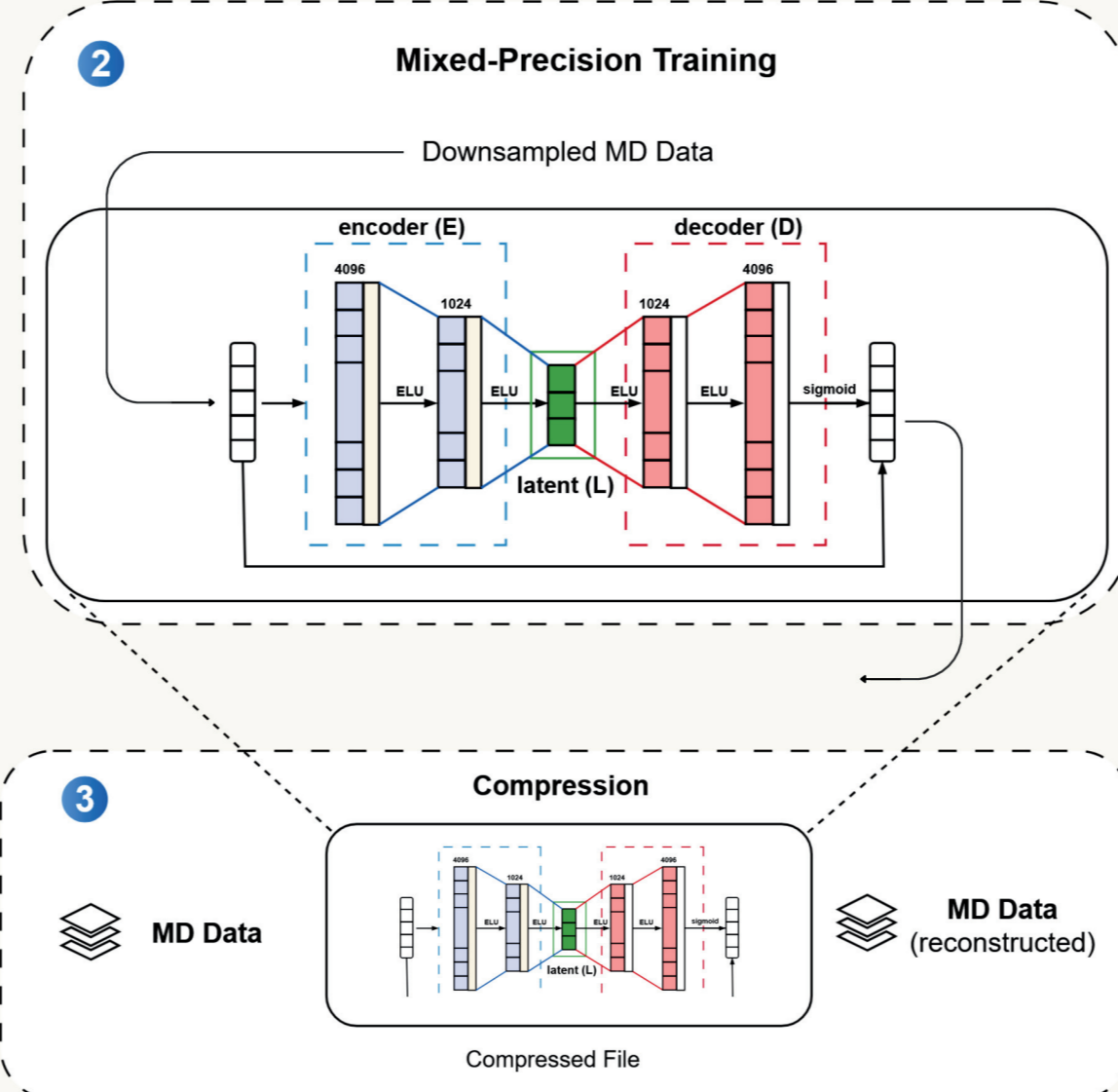
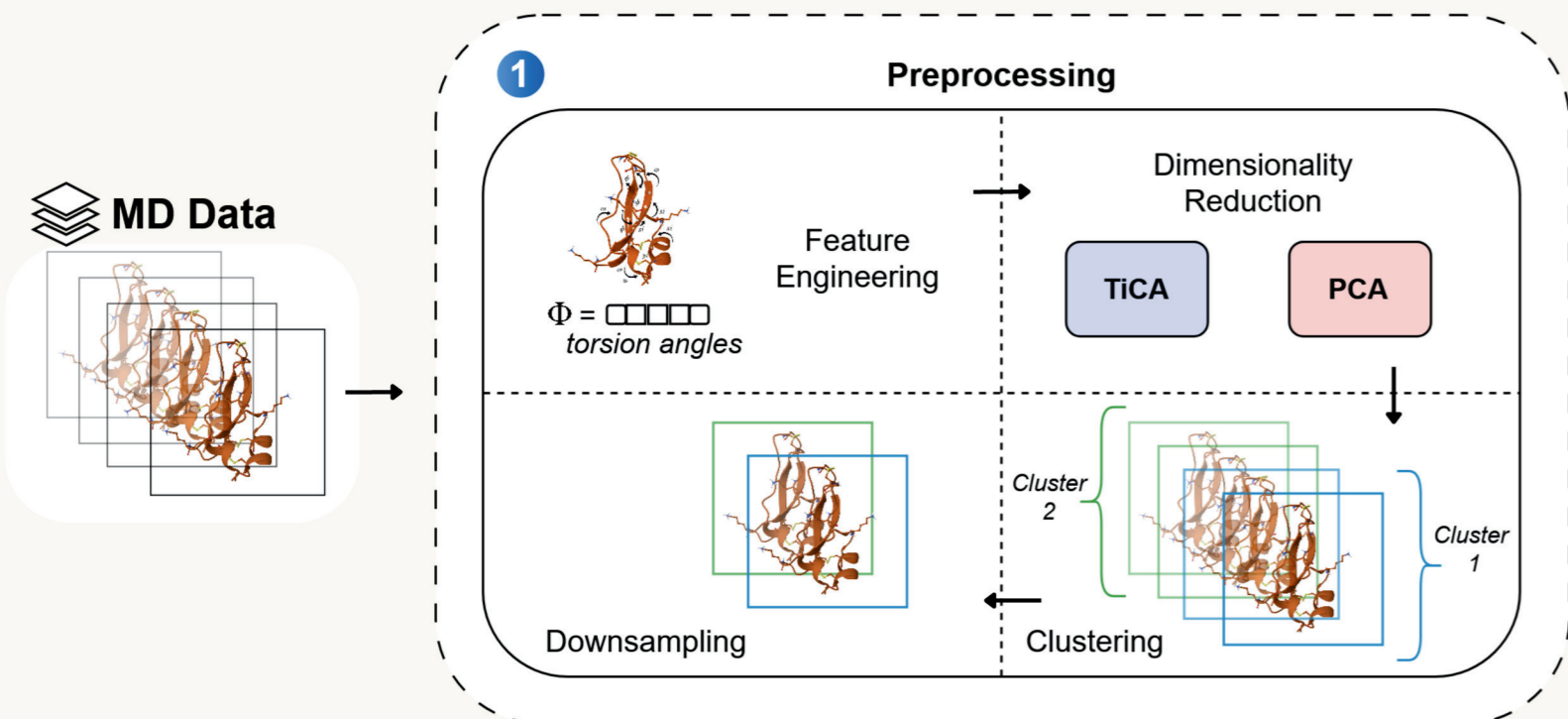


Fig. 2 : Frame by Frame RMSD Comparison

- 3 After aligning the RMSD magnitude across methods, **compressTraj** demonstrates **fast compression** and achieves a **comparable compression ratio**.
- 4 However, the **model training time** remains relatively **high** and the **trained model size** is still **substantial**, highlighting opportunities for further optimization.

## 03 Pipeline



- Based on the promising results of **compressTraj**, **CompactMD** extends its architecture as follows:
- **Preprocessing**: representative frame selection.
- **Mixed-precision Training**: BF16 encoder–decoder on downsampled MD.
- **Compression**: compress using above trained model.

## 04 Preliminary Results

32K seconds → 18K seconds

~43% total time reduction

20.88 → 40

~50% compression ratio increase (compressTraj)

0.58 → 1.16

~50% compression ratio increase (compressTraj + model)

Accuracy was preserved

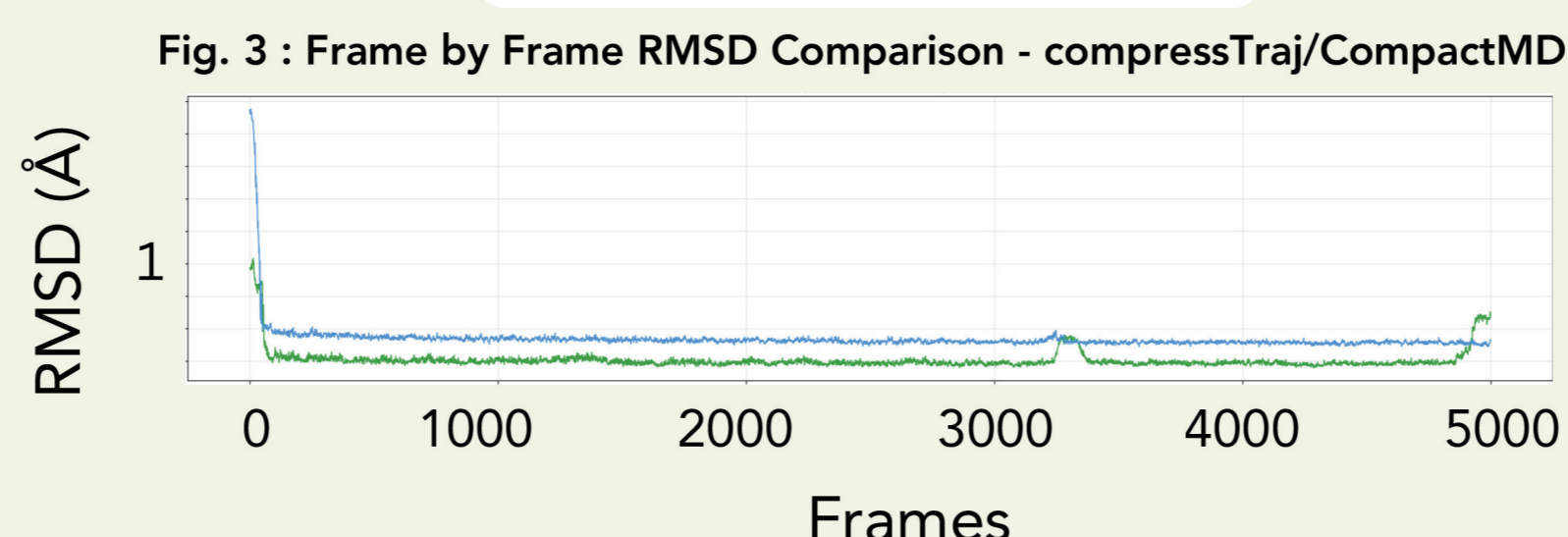


Fig. 3 : Frame by Frame RMSD Comparison - compressTraj/CompactMD

## 05 Conclusion

**CompactMD** demonstrates a promising approach that AI-driven compression can match and potentially exceed traditional methods with further optimization.

Ongoing work

- Exploring quantization methodologies.
- Scaling the dataset scope to larger and longer trajectories.
- Experimenting with different molecular systems.