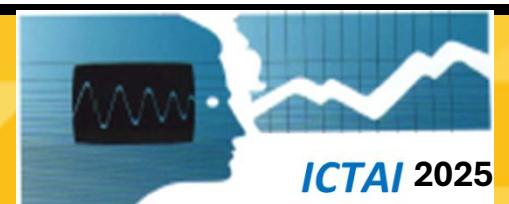




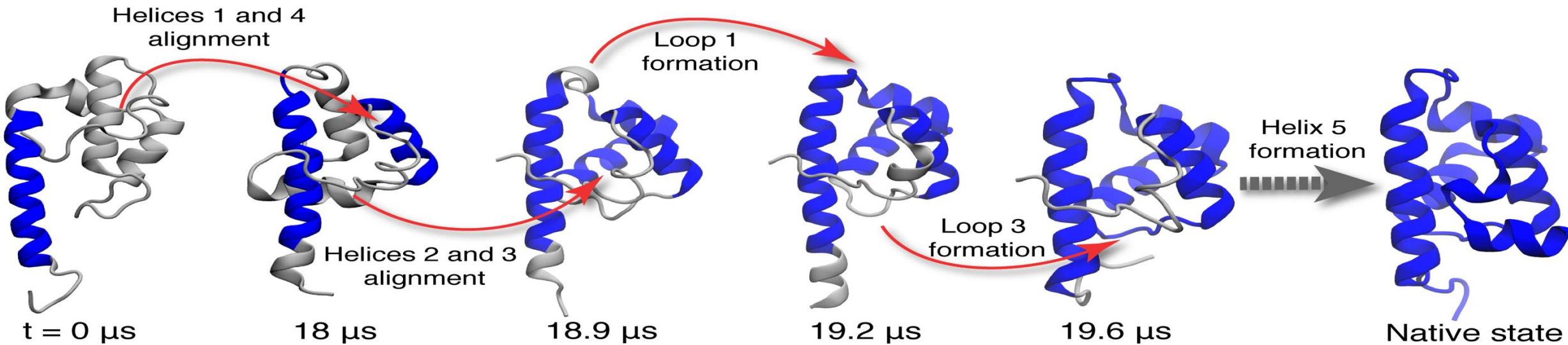
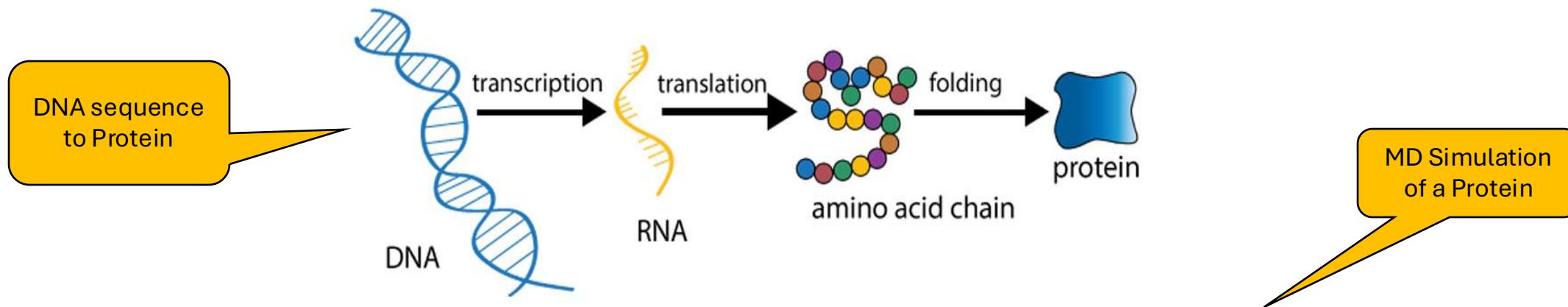
AccelMD: A Self-adaptive AI-enabled Framework for Accelerating Molecular Dynamics Simulations

Authors: Kazi Fahim Ahmad Nasif, Bobin Deng, Lingtao Chen, Yixin Xie, Shaolei Teng, Jiang Li, Liang Zhao, Syed Md Shamsul Alam, Nobel Dhar, Kun Suo and Dan Chia-Tien Lo

Presented by
Kazi Fahim Ahmad Nasif
PhD Candidate
Kennesaw State University
United States
knasif@students.Kennesaw.edu



Molecular Dynamics Simulation (MD)



Why MD Matters (Motivation)

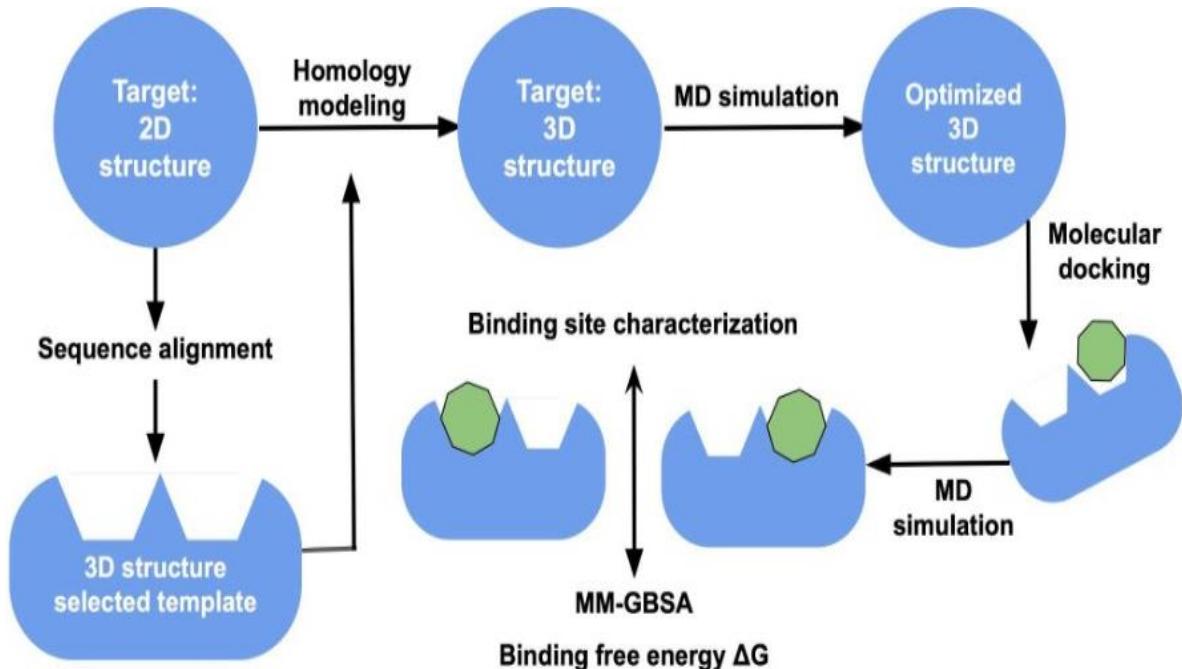


Fig: Drug Discovery Process

Computational Bottleneck

20 nanosecond simulation of 13,137 atoms (1653 residues) takes 5 hours on high-performance hardware.

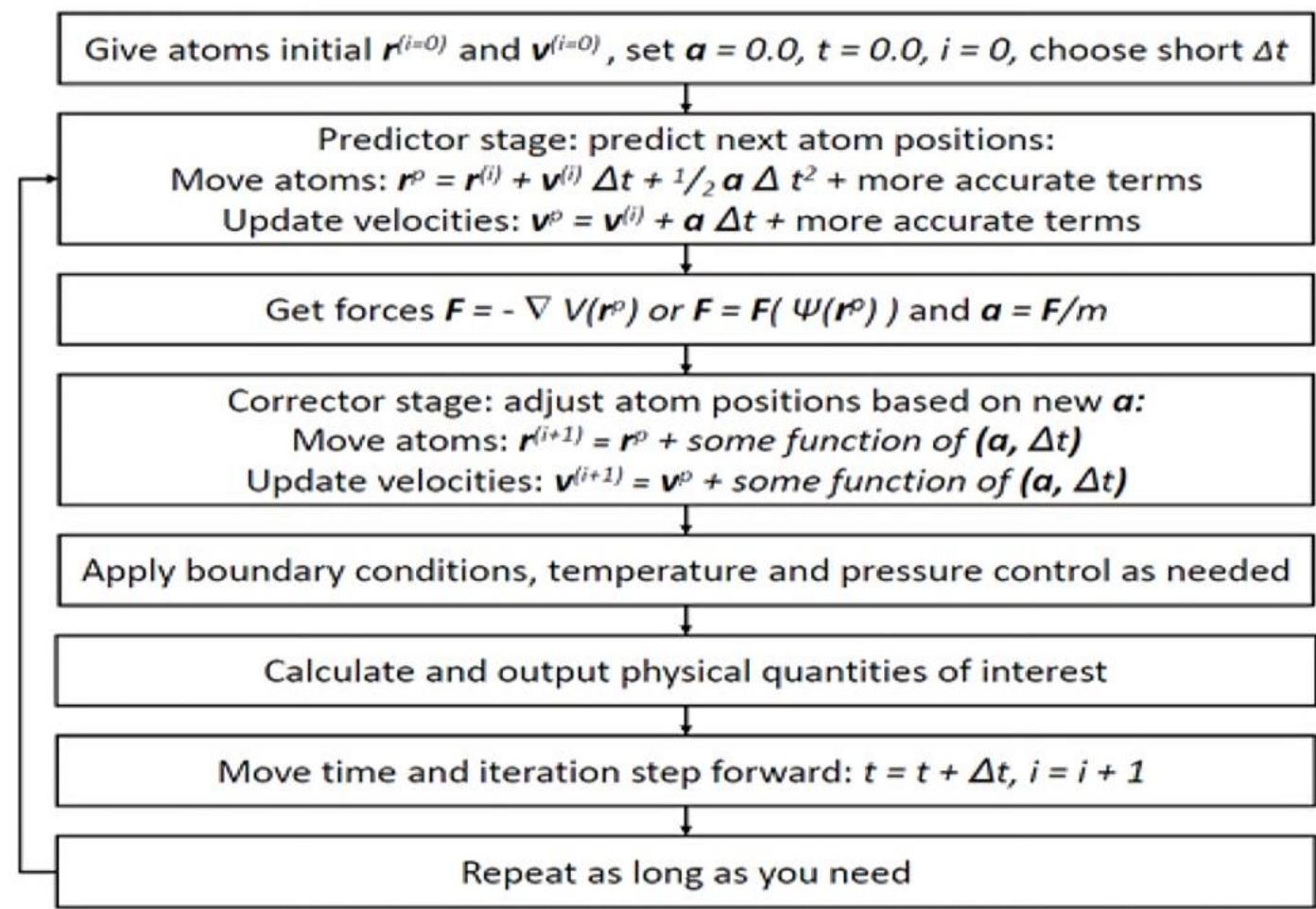
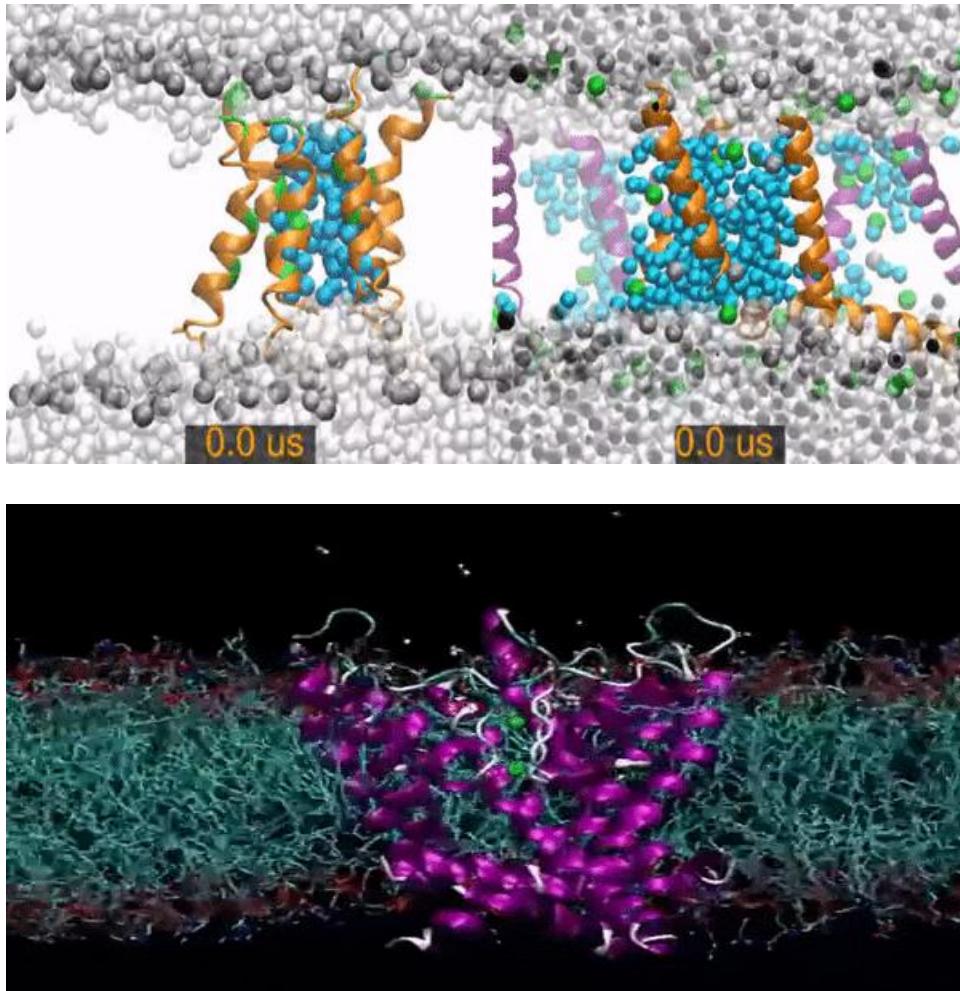
Sequential Nature

Each timestep depends on the previous one, limiting parallel computing acceleration and creating a fundamental barrier.

Scientific Impact

Long-timescale phenomena like protein folding remain inaccessible, slowing drug discovery and biological research.

MD pipeline recap (where time goes)



Related work (classical & parallel)

- **Classical MD Simulation**
 - Solves Newton's equations of motion via iterative force-field calculations (e.g., AMBER, GROMACS, NAMD, LAMMPS, OpenMM).
 - extremely high computational cost.
 - Larger biomolecular systems → months of runtime even on HPC clusters.
- **Parallel Computing & HPC Optimizations**
 - GPU acceleration, and MPI reduce
 - MD is sequentially dependent (each timestep → next timestep)
 - Amdahl's Law → theoretical ceiling

Motivating new AI-driven acceleration strategies

Related work (AI for MD)

Neural Network Potentials (NNPs):

- Examples: **SchNet**, **PhysNet**, **ANI**, **NequIP**, **MACE**.
- Learn **energy and force fields** to replace classical potentials.
- **Limitations:** Fails to predict long-range interactions.
- Computationally expensive and limited robustness.
- Limited adaptability to small proteins.

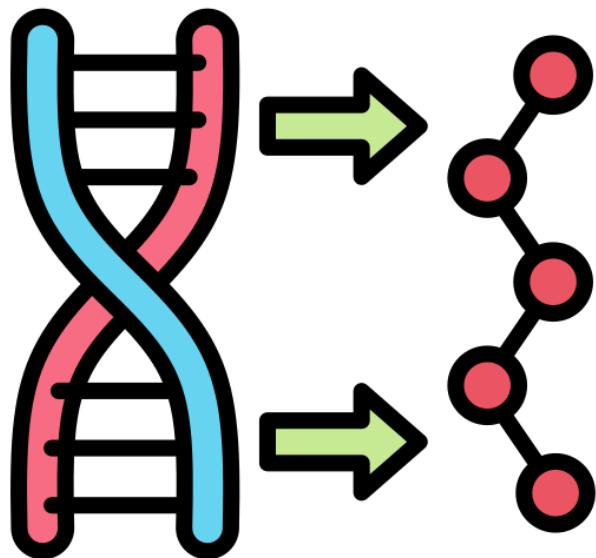
Microsoft BioEmu-1

- **Goal:** Generate *equilibrium ensembles* of proteins using large generative diffusion models.
- **Limitation:** Captures ensemble diversity but not explicit time evolution or kinetic transitions.
- BioEmu-1 focuses on *static equilibrium*.

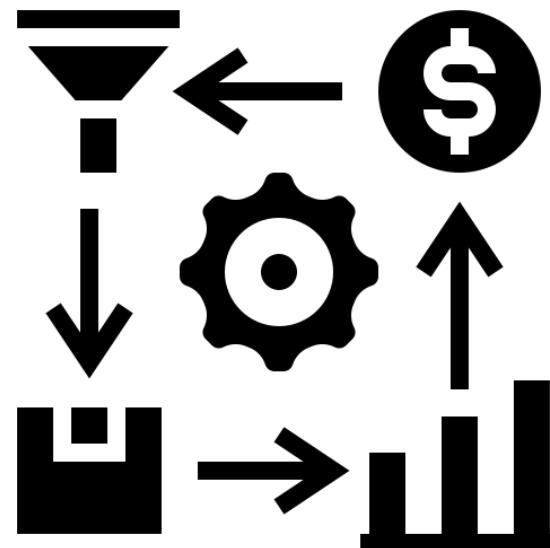
MIT DeepJump (arXiv 2025)

- **Model:** Predict conformational transitions over variable “jump sizes.”
- Generates long-timescale trajectories and folding pathways using generative sampling.
- **Limitations:**
 - Accuracy degrades for large jump sizes (> 10 ns).
 - Poor generalization to small/disordered structures.

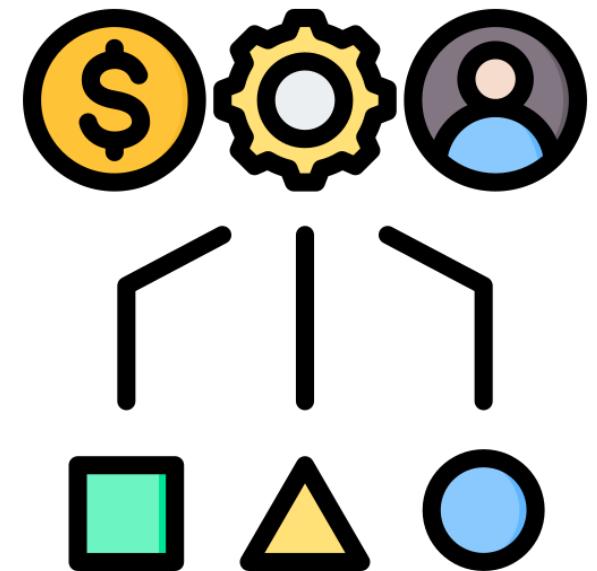
Research objective



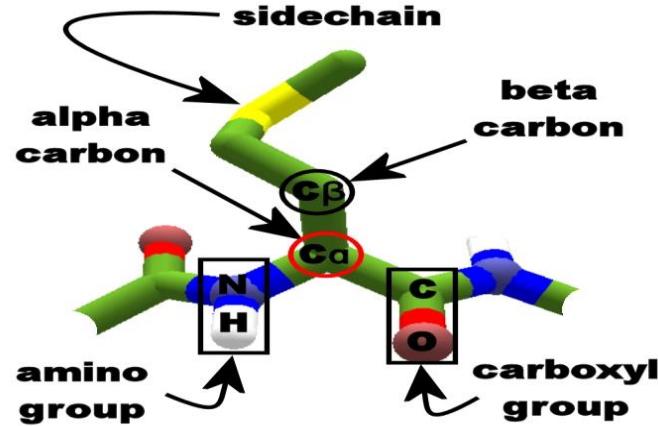
Accurate prediction



Faster Inference



Low Resource

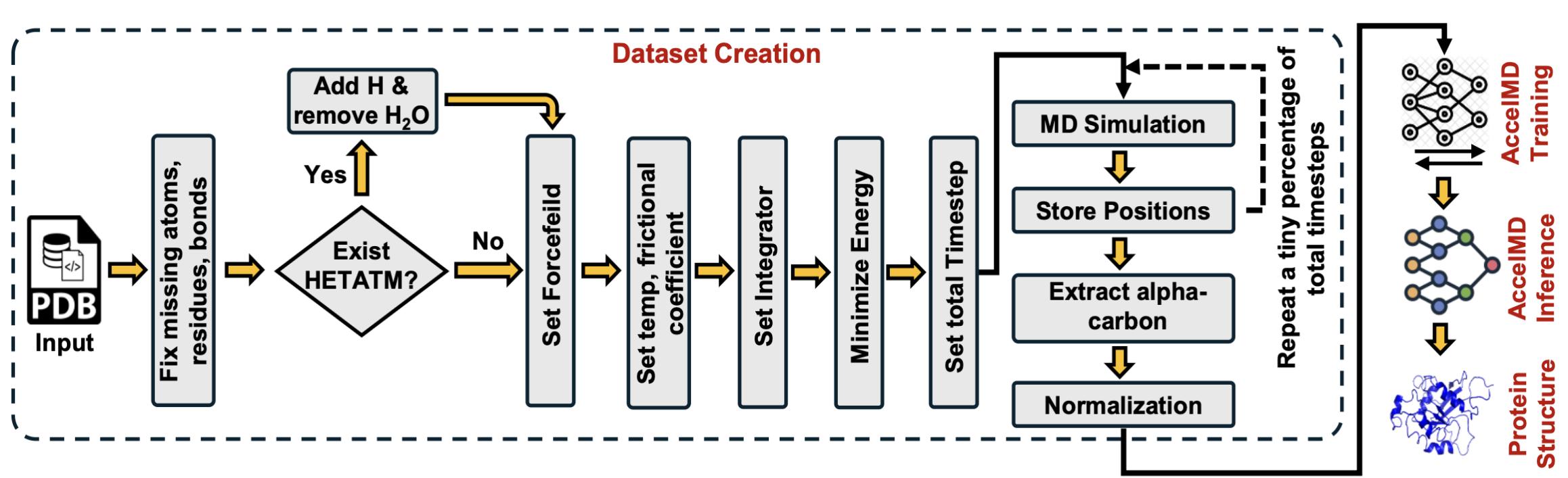


Structure of PDB file

	Amino Acid	Sequence/Residue Number	Coordinates			(etc.)
Element	Chain		X	Y	Z	
ATOM 1 N	MET A	1	19.353	41.547	-3.887	...
ATOM 2 CA	MET A	1	20.513	40.939	-4.592	...
ATOM 3 C	MET A	1	20.150	39.658	-5.355	...
ATOM 4 O	MET A	1	19.053	39.551	-5.903	...
ATOM 5 CB	MET A	1	21.642	40.678	-3.592	...
ATOM 6 CG	MET A	1	21.233	39.903	-2.360	...
ATOM 7 SD	MET A	1	22.533	39.928	-1.113	...
ATOM 8 CE	MET A	1	23.771	38.881	-1.885	...
ATOM 9 N	ASP A	2	21.068	38.694	-5.390	...
ATOM 10 CA	ASP A	2	20.856	37.440	-6.117	...
ATOM 11 C	ASP A	2	20.124	36.371	-5.299	...
ATOM 12 O	ASP A	2	20.680	35.818	-4.351	...

Element position within amino acid

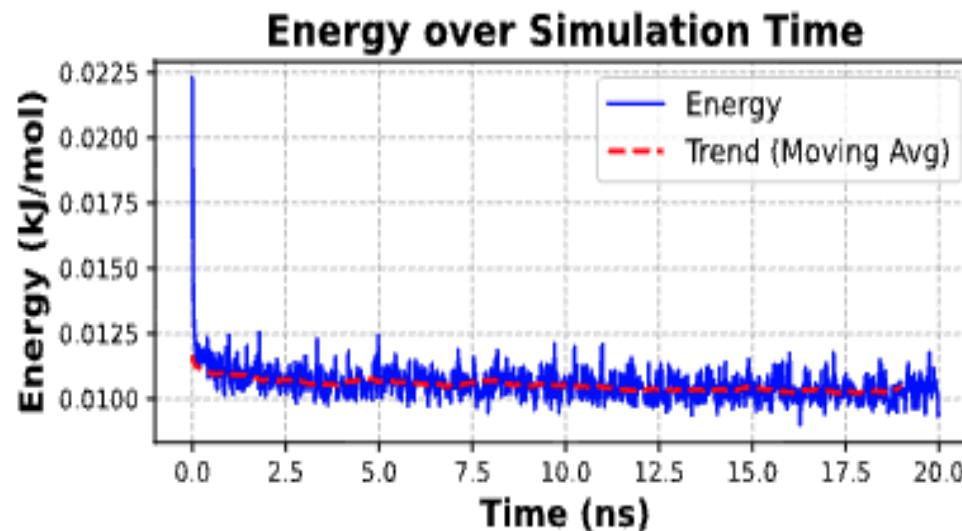
AccelMD framework



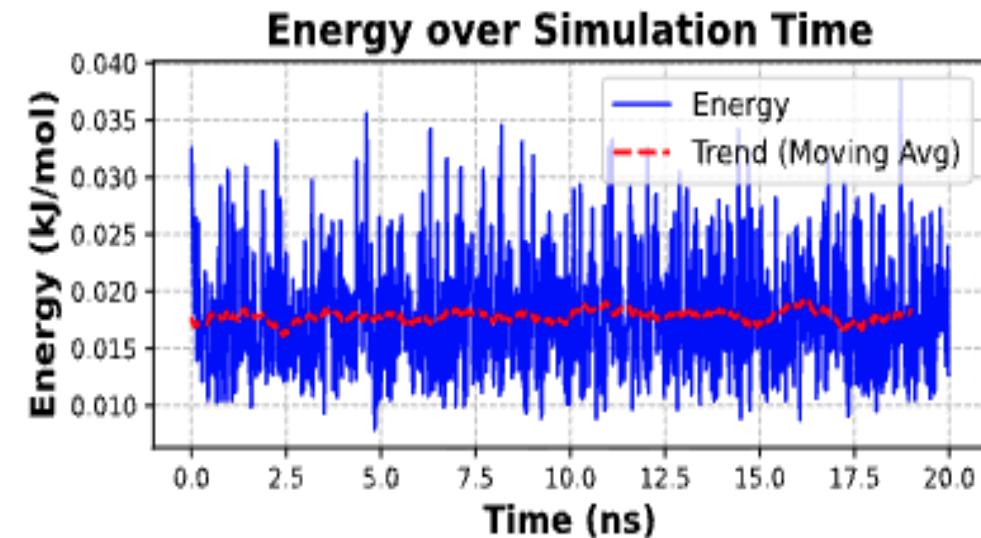
Dynamic dataset creation

- **PDB fixing:** Repair missing atoms, residues, and bonds using *PDBFixer*.
- **Energy minimization → Equilibration:**
Run a short simulation (e.g., **Langevin integrator @ 300 K, 2 fs timestep**)
- **Data collection:**
Collect **α -carbon (C_α) coordinates** every k steps over ≈ 20 ns
- **Dataset organization:**
Store data in a **3-dimensional NumPy array** with shape: (timesteps, residues, 3)
 - $residues$ = number of C_α atoms in the protein (e.g., 1653)
 - 3 = (x, y, z) coordinate values
 - **Example:** $(1000 \times 1653 \times 3)$
- * C_α atoms are responsible for protein's structural backbone

Energy minimization



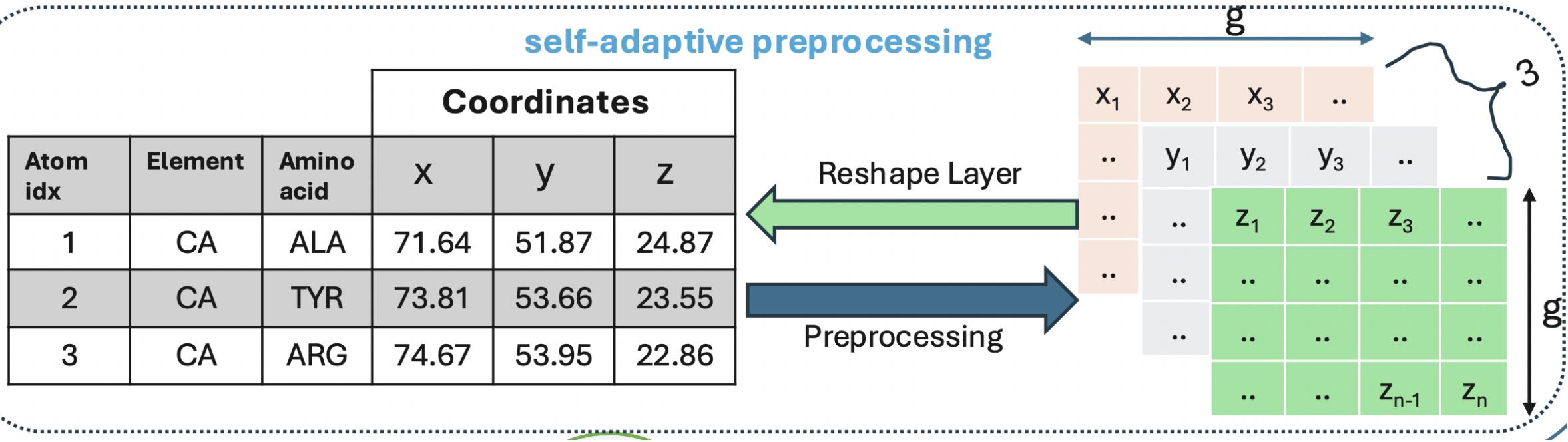
(a) Protein ID '8sk8'



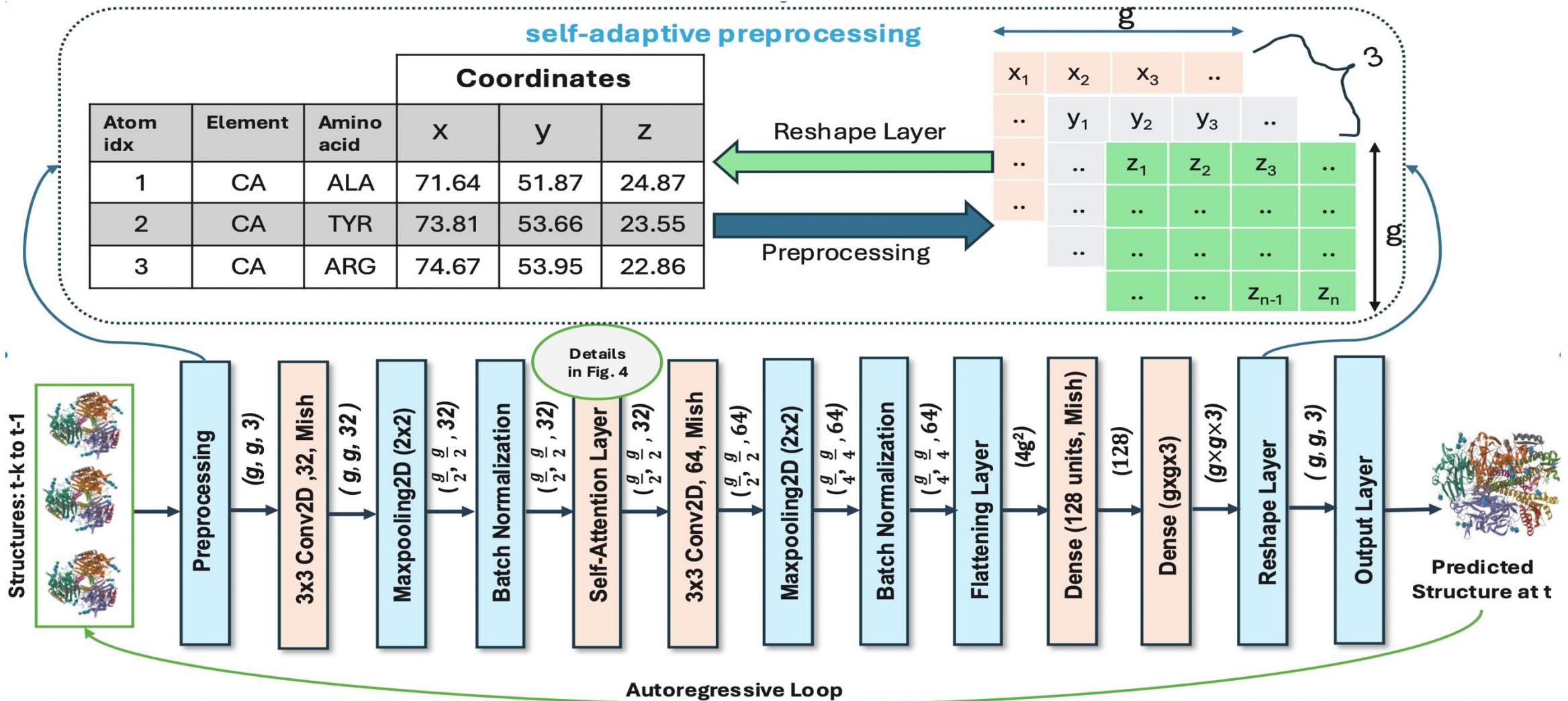
(b) Protein ID '5ghd'

Energy Minimization ensures the stability of the system

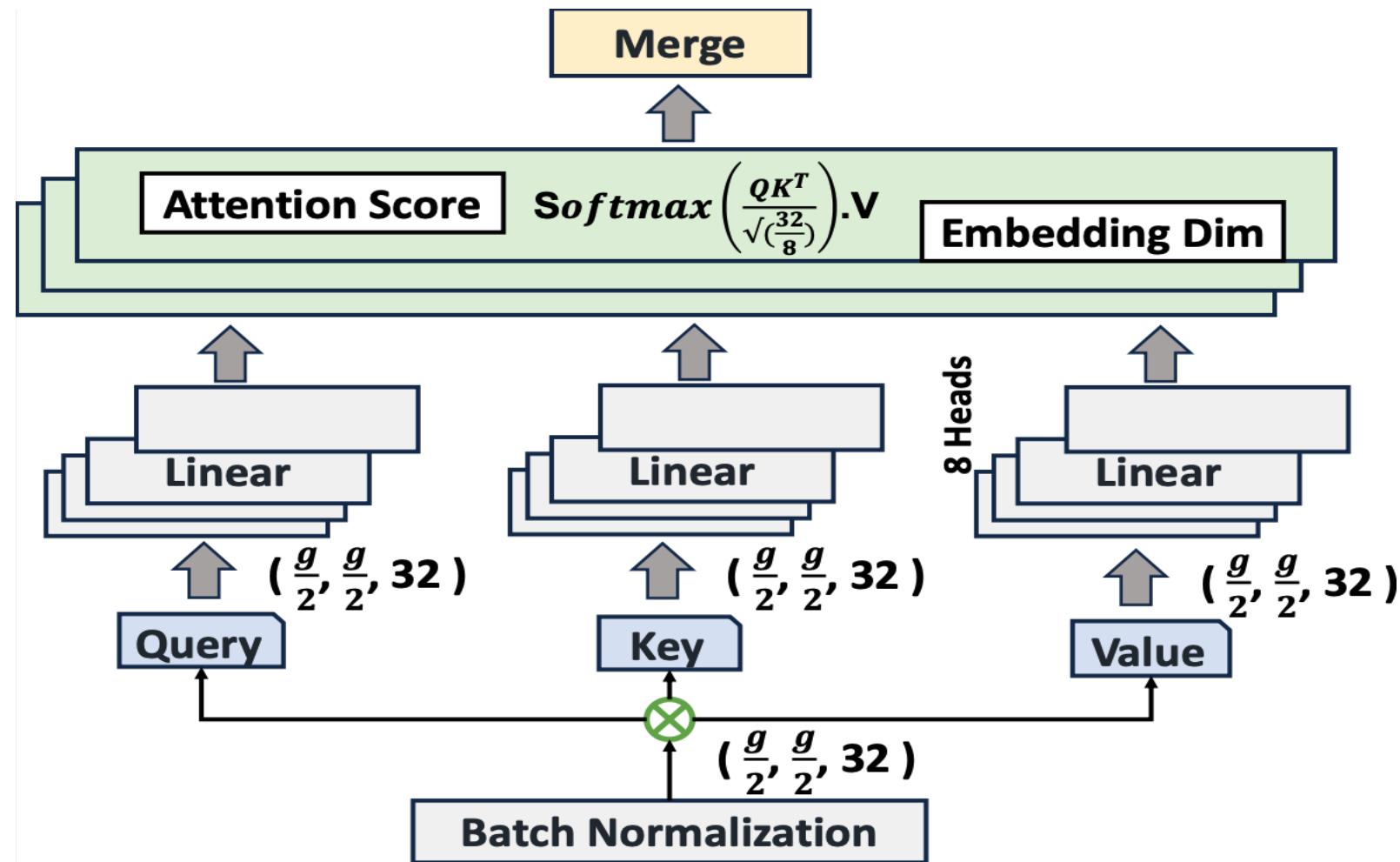
Self-adaptive preprocessing



Model architecture



Model architecture (cont.)



Experimental Setup

- Dataset
 - Protein Data Bank (PDB)
 - 20 proteins spanning 97–2976 residues
- Baselines
 - ResNet50
 - AlexNet
 - LeNet
 - GoogleNet
 - DenseNet

System Configuration:

- 256 AMD EPYC 7742 64-Core Processor CPU
- eight NVIDIA A100-SXM4-80 GB GPU
- 2 TB RAMs

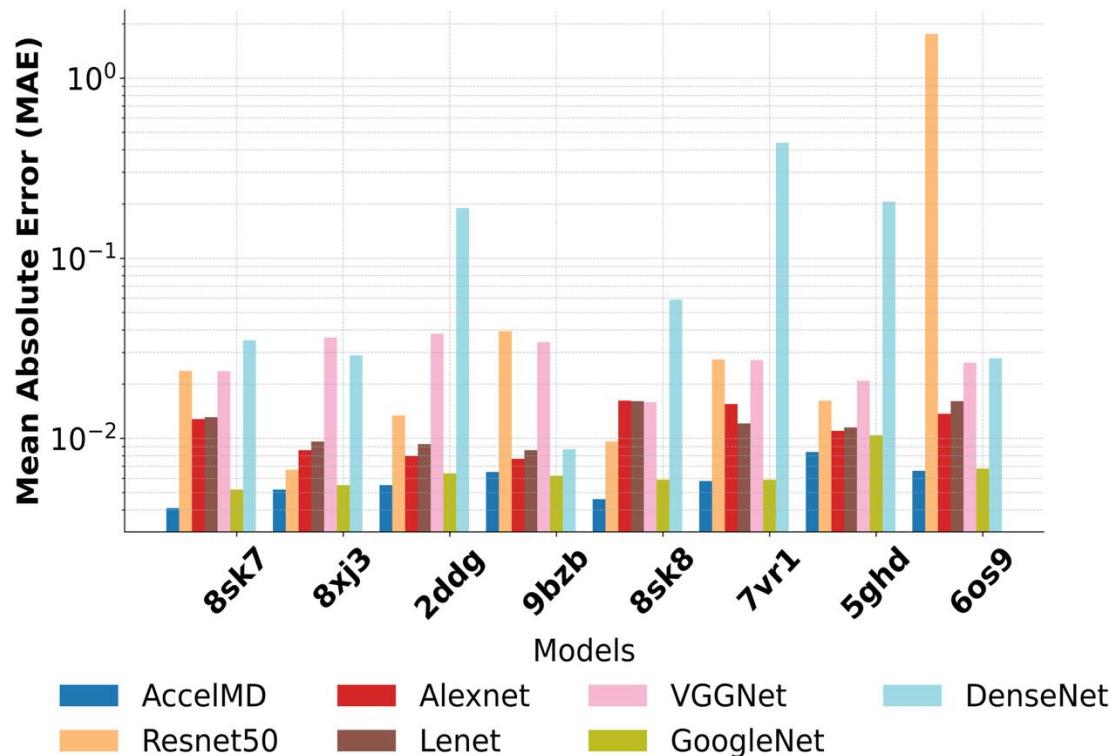
Evaluation Metrics for Structural Accuracy

- **MAE (\AA) :** (1 Angstrom = 10^{-10} m)
- **TM-score (Template Modeling score):**
 - Quantifies the structural similarity between two protein conformations.
Ranges from **0 to 1**, where 1 indicates a perfect match.

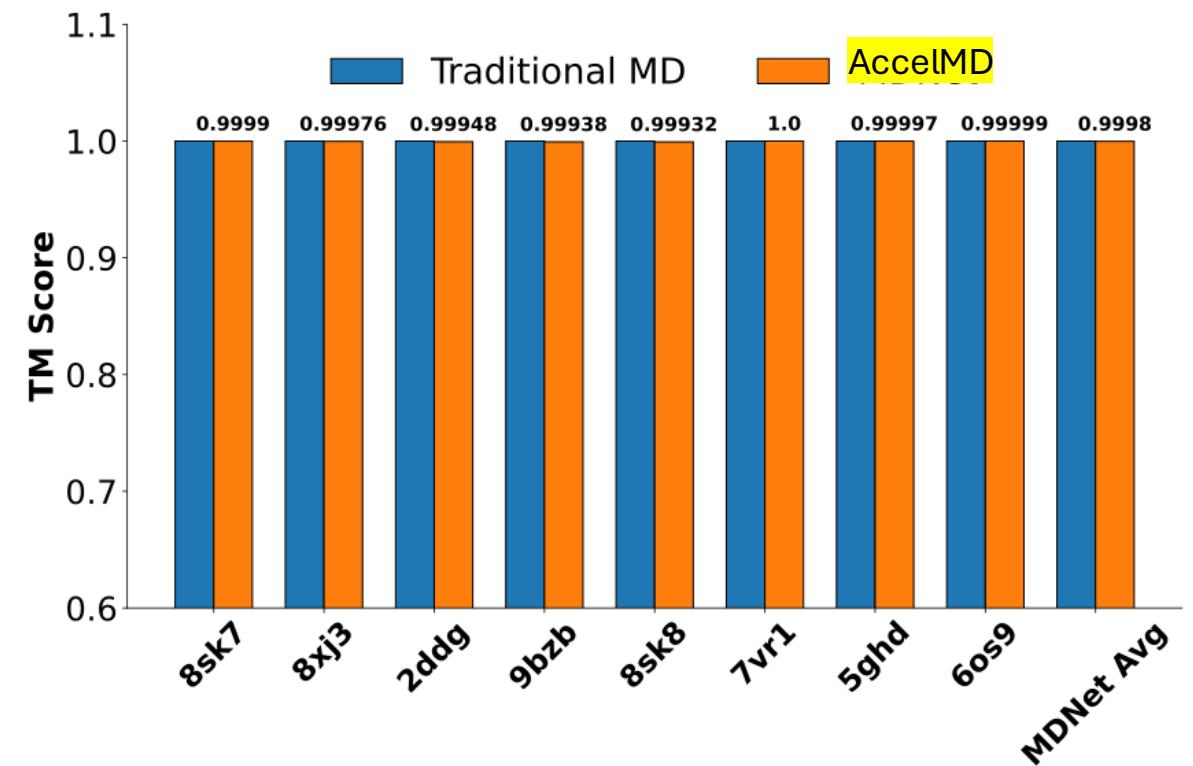
$$\text{TM-score} = \max \left(\frac{1}{L_{target}} \sum_{i=1}^{L_{common}} \frac{1}{1 + (\frac{d_i}{d_0})^2} \right)$$

- L_{target} : length of target protein
- L_{common} : number of aligned residues
- d_i : distance between the i -th pair of aligned residues
- $d_0 = 1.24 \sqrt[3]{L_{target} - 15} - 1.8$: scale factor

Results

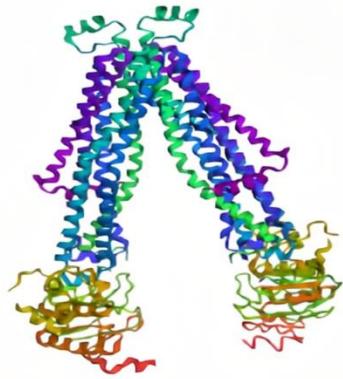


MAE ~ **0.0058 Å** (best overall)



TM-score ~ **0.9998** (near-MD)

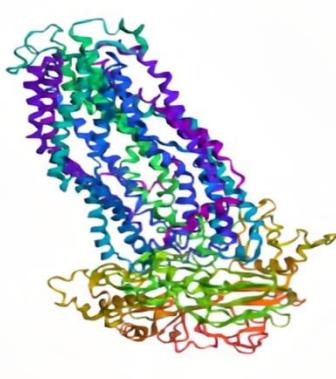
Visual Comparison



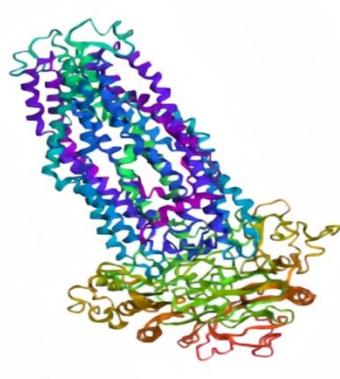
Original



Fixed Protein



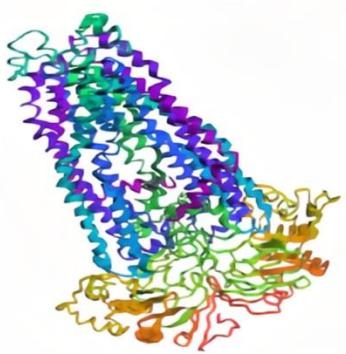
after 20ns (Simulated)



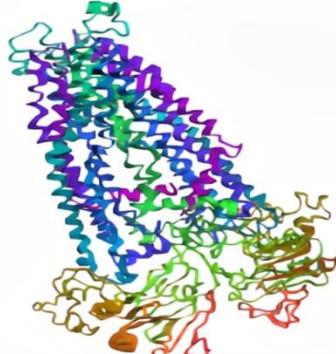
after 20ns (Predicted)



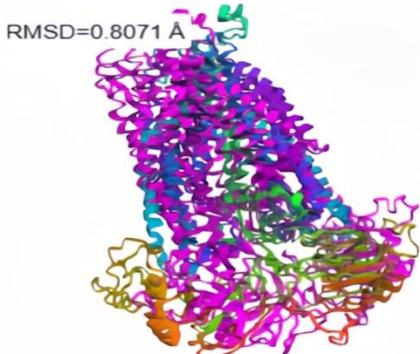
simulated vs predicted (20ns)



after 30ns (Simulated)

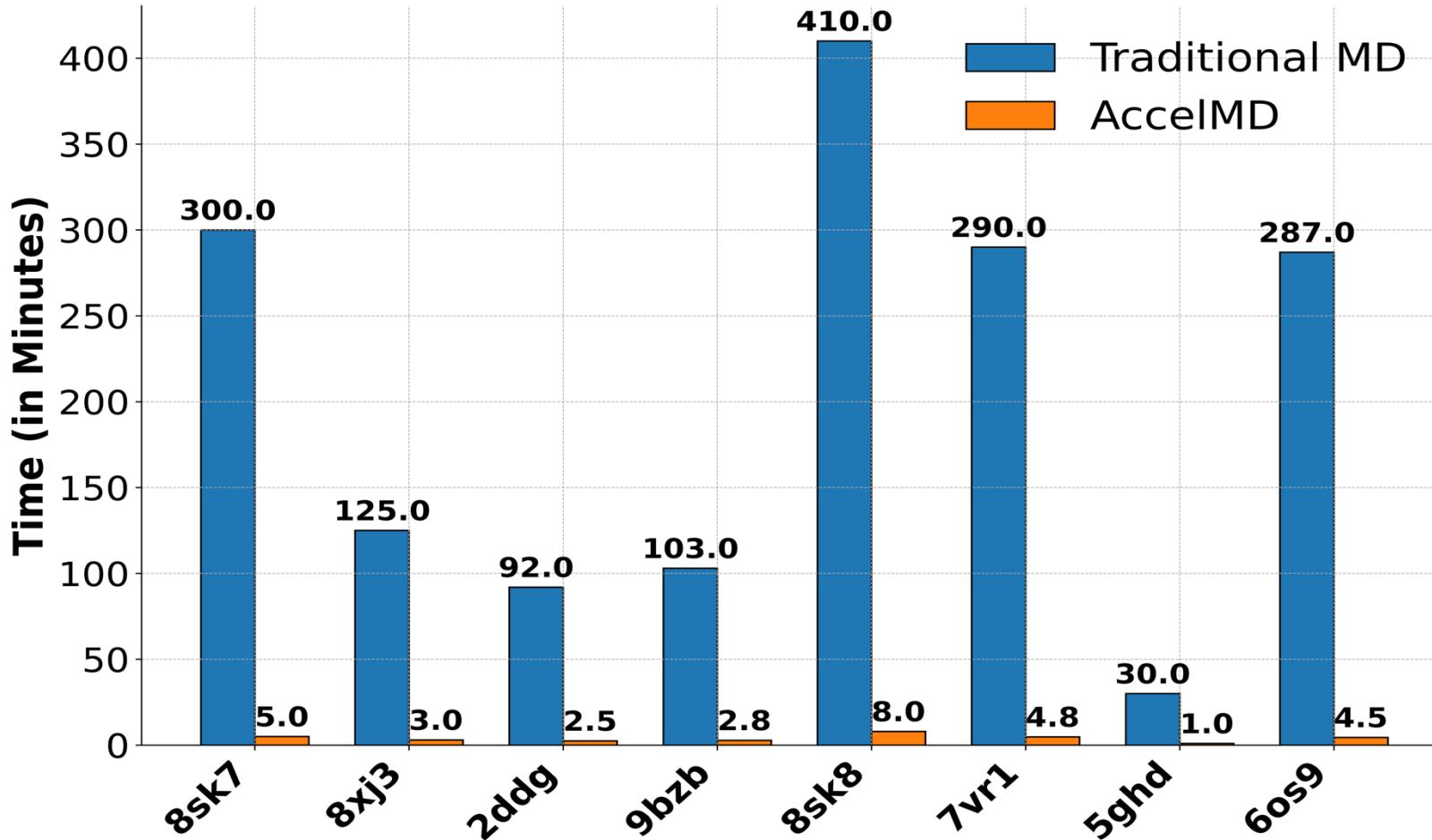


after 30ns (Predicted)



simulated vs predicted (30ns)

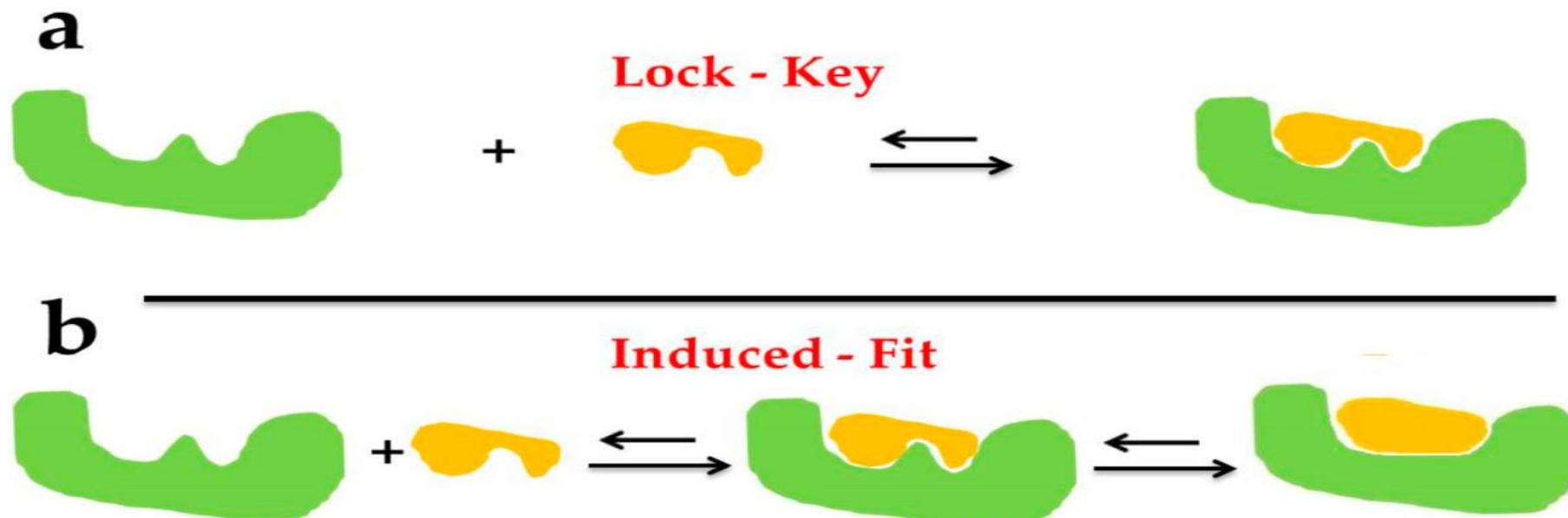
Speed & Robustness



- ~47.6× average speedup (max ~63.8×)
- Handles all sizes
- structural overlays (RMSD < ~1 Å examples)

Limitations & Future work

- Waiting time for initial simulations (5 hours)
- Integrating learned energies/forces
- Enhance protein-protein docking accuracy (Induced-Fit)



Contributors



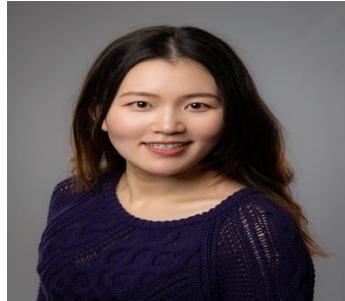
Kazi Fahim Ahmad Nasif
PhD. Candidate
Kennesaw State University
knasif@students.Kennesaw.edu



Bobin Deng
Assistant Professor
Kennesaw State University
bdeng2@Kennesaw.edu



Lingtao Chen
PhD. Candidate
Kennesaw State University
lchen25@students.Kennesaw.edu



Yixin (Chloe) Xie
Assistant Professor
Kennesaw State University
yxie11@Kennesaw.edu



Shaolei Teng
Associate Professor
Howard University
shaolei.teng@Howard.edu



Jiang Li
Associate Professor
Howard University
jli@Howard.edu



Liang Zhao
Assistant Professor
Kennesaw State University
lzhao10@Kennesaw.edu



Syed Md Shamsul Alam
PhD. Student
Northern Arizona
University
ss5687@nau.edu



Nobel Dhar
PhD. Candidate
Kennesaw State University
ndhar@students.Kennesaw.edu



Kun Suo
Associate Professor
Kennesaw State University
ksuo@Kennesaw.edu



Dan Lo
Professor
Kennesaw State University
dlo2@Kennesaw.edu