AI and HPC in Molecular Simulations and Drug Discovery

Bojana Koteska

Faculty of Computer Science and Engineering, Skopje

EuroCC4SEE Workshop in Belgrade, May 2025

Traditional Drug Discovery

- Drug discovery traditionally spans 10-15 years with an estimated cost of $2.6 billion per drug
- The success rate is low (~10%) due to high failure rate during clinical trials
- Traditional methods rely on trial and error, extensive lab work, and repetitive screening

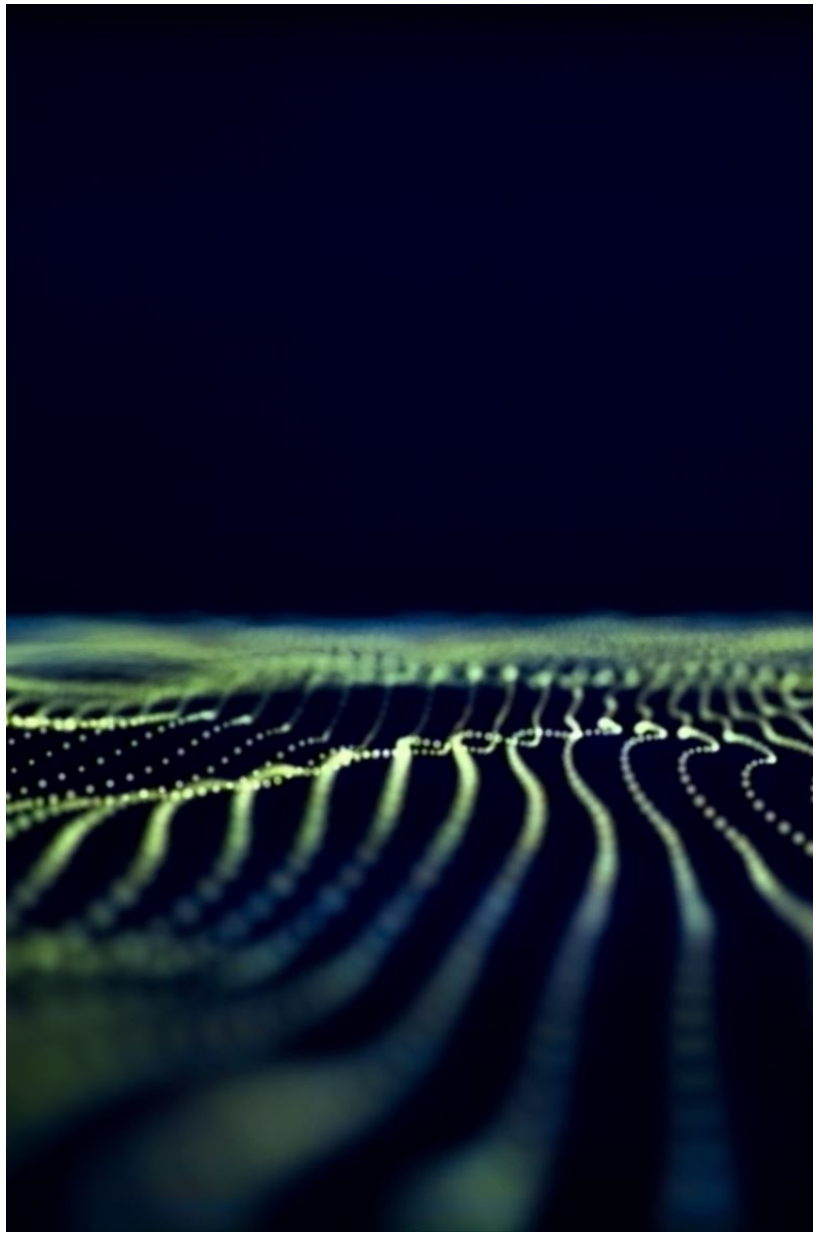
Key Challenges in Drug Discovery

R&D timelines are extremely long and expensive

Biological systems are complex and difficult to model

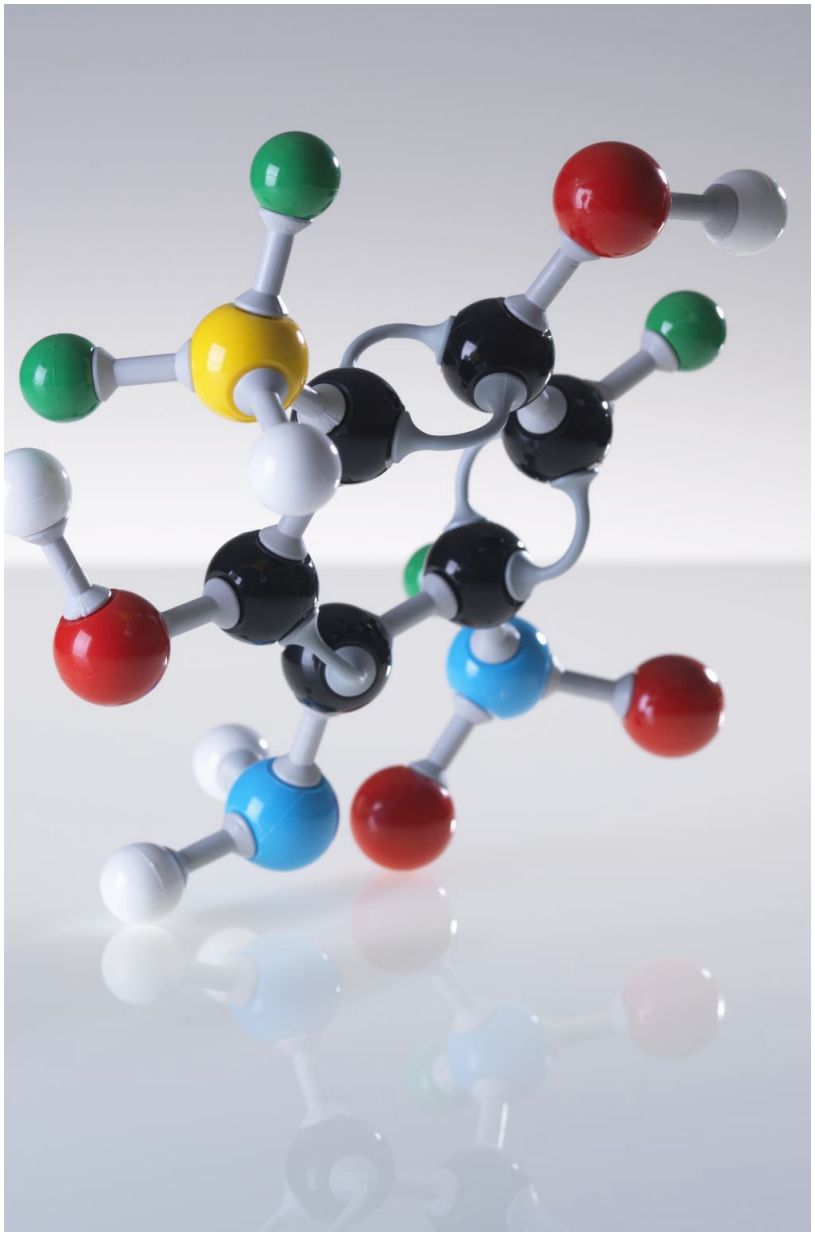
Many drugs fail in late-stage trials due to unforeseen toxicity or inefficacy

There is an urgent need for more predictive, scalable, and data-driven approaches



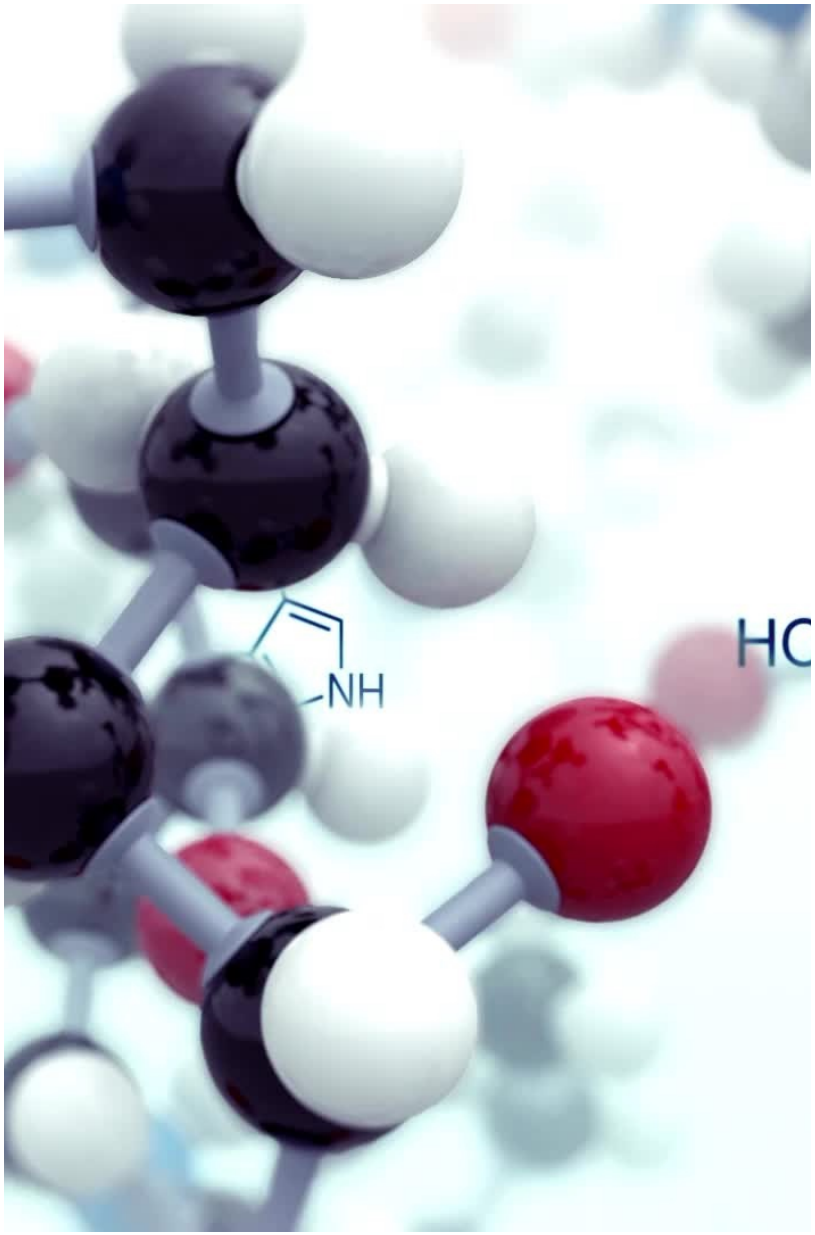
Why We Need Innovation?

- Biomedical data is growing exponentially from genomics, proteomics, and clinical records
- Drug-resistant pathogens and rare diseases demand more tailored therapies
- Precision medicine requires integrating vast, heterogeneous datasets
- Innovations like AI and HPC offer faster, cheaper, and more reliable alternatives



What Are Molecular Simulations?

- Molecular simulations model the behavior of atoms and molecules to predict structure, dynamics, and interactions
- Applications include drug binding, protein folding, and solvent effects
- Simulations bridge theoretical chemistry and experimental biology

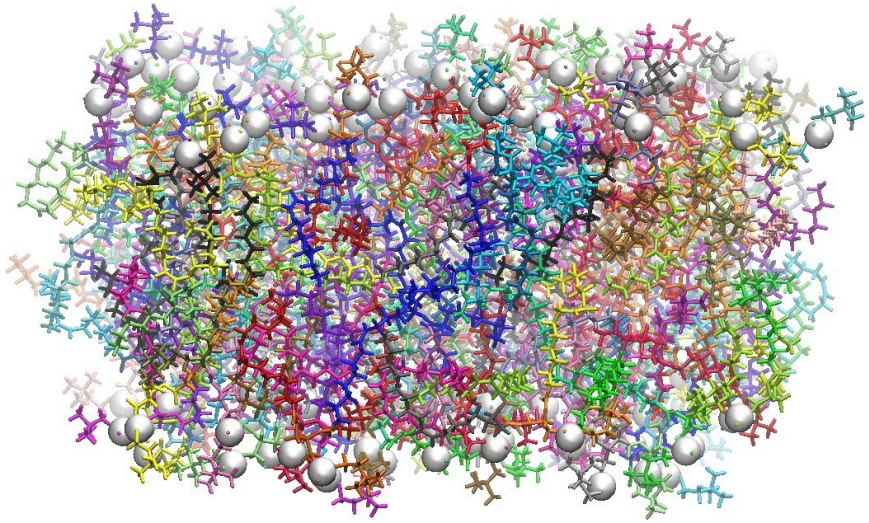


Molecular Dynamics (MD)

- MD simulations compute the time-dependent behavior of molecular systems using Newton's laws of motion
- Simulations can reveal how a drug binds to a receptor or how a protein changes shape
- Timescales: from picoseconds to microseconds, depending on system size

Common Software Tools

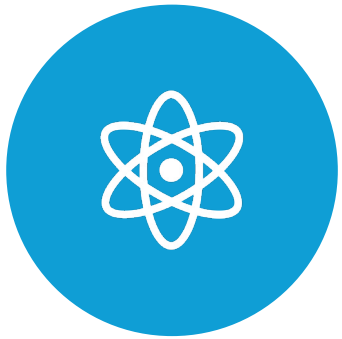
- GROMACS, NAMD: classical MD
- AutoDock: molecular docking
- AMBER, CHARMM: force-field based simulations



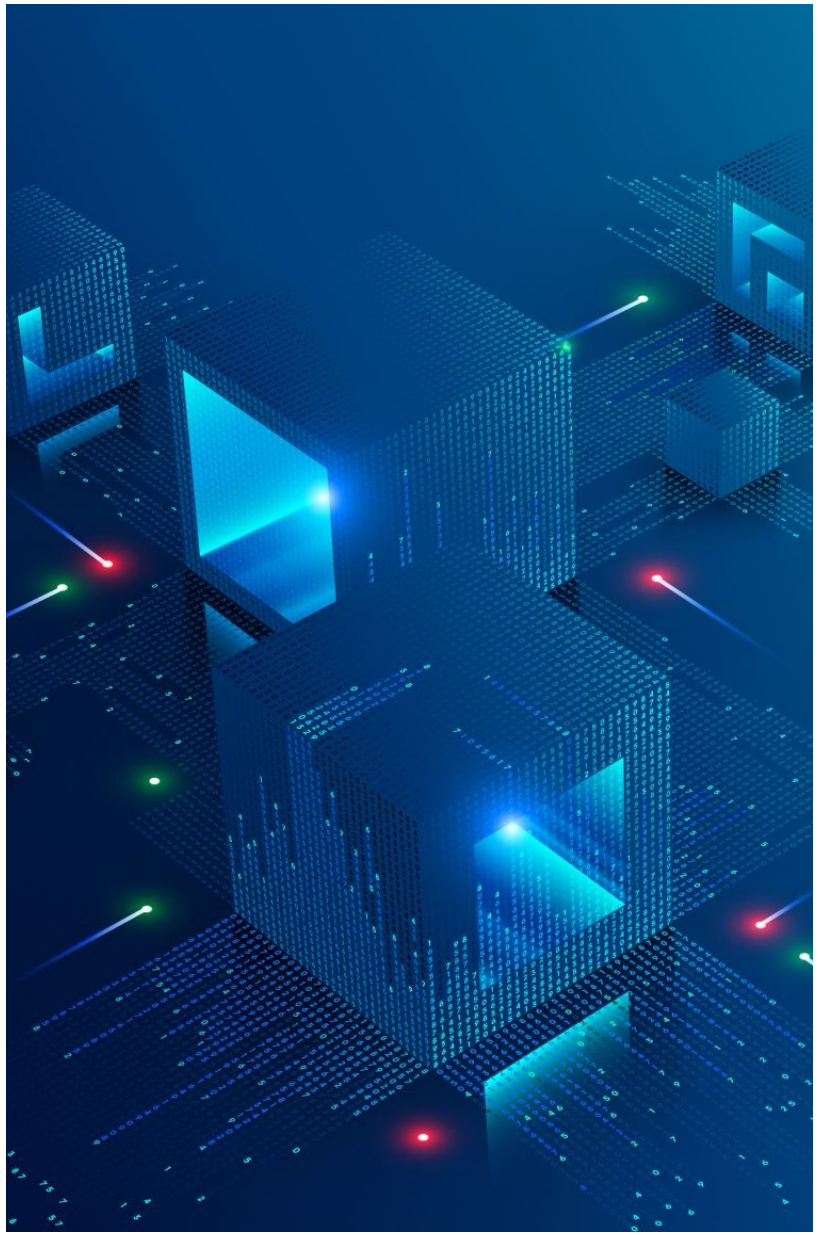
Limitations of Classical Simulations

REQUIRE SIGNIFICANT COMPUTATIONAL RESOURCES

CANNOT EASILY SIMULATE RARE EVENTS OR LONG TIMESCALES



SIMPLIFIED MODELS MIGHT MISS IMPORTANT QUANTUM EFFECTS OR HOW MOLECULES CHANGE SHAPE WHEN INTERACTING



What is HPC?

- HPC uses parallel computing architectures to solve large-scale problems quickly
- Clusters of CPUs and GPUs process millions of calculations simultaneously
- Enables researchers to simulate complex biomolecular systems

Why Use HPC in Simulations?

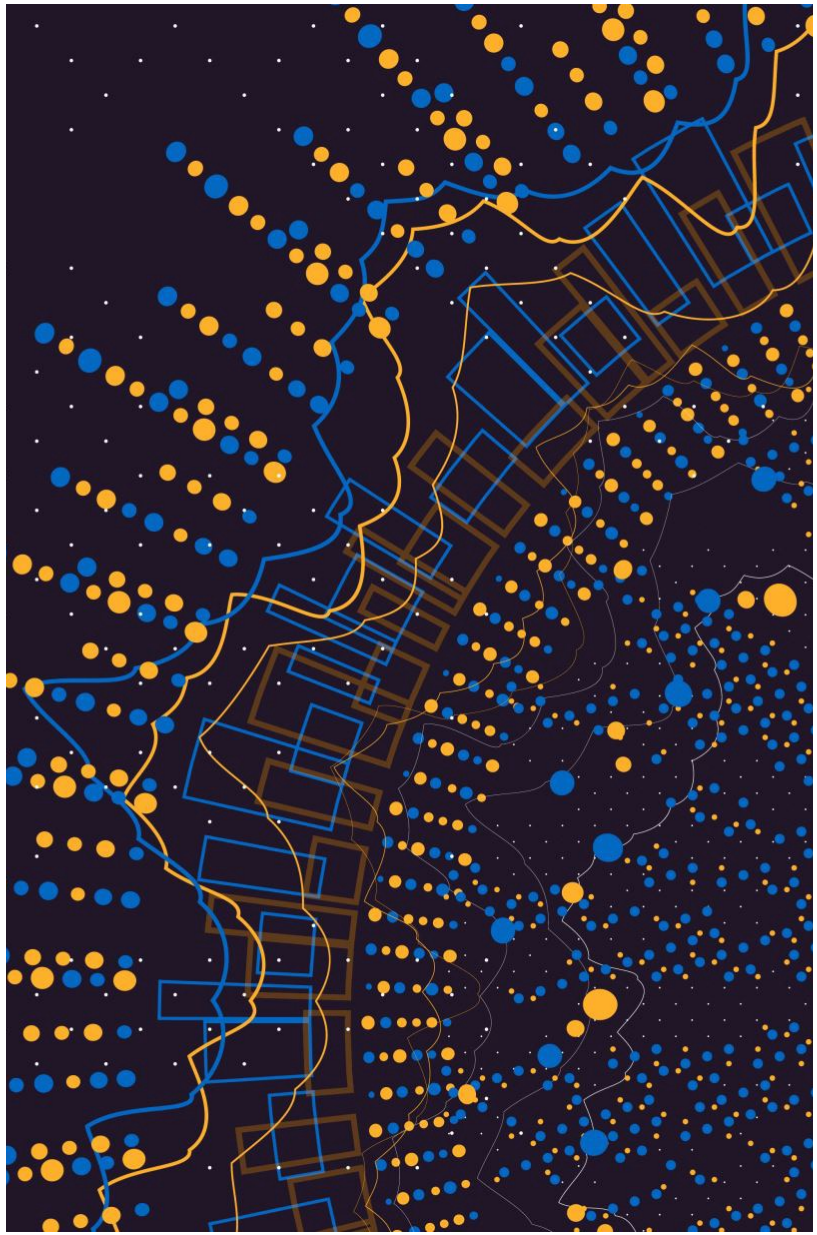
- Allows modeling of large biological assemblies (e.g., virus capsids, cell membranes)
- Enables longer simulations to observe slow dynamics like protein folding
- Reduces time-to-results from weeks to hours

Infrastructure Requirements

- HPC systems include compute nodes, fast interconnects, storage arrays, and job schedulers
- Software must be optimized for parallel execution
- Facilities include national supercomputing centers and cloud-based HPC

Real Applications in Simulations

- SARS-CoV-2 spike protein simulations supported vaccine design
- Antibody-antigen binding simulations assist in developing immunotherapies
- HPC accelerated virtual screening of billions of compounds during the COVID-19 pandemic



What is AI & ML?

- Artificial Intelligence: systems that mimic human intelligence
- Machine Learning: algorithms that learn from data
- Deep Learning: uses neural networks to detect patterns and make predictions
- These technologies enable intelligent data analysis and decision-making in drug discovery

AI in Drug Discovery

- Predicts protein structure, drug-likeness, toxicity, and side effects
- Facilitates de novo drug design
- Automates screening of millions of molecules
- AI reduces reliance on costly and time-consuming lab experiments

How Molecules Become Machine Learning Inputs

- Molecules are stored in a text format called SMILES (e.g., "CCO" for ethanol)
- These are converted into numerical features like:
  - Molecular size
  - Atom types
  - Presence of rings or charges
- These features are used to train AI models to make predictions
- *molecule icon → feature vector (e.g., [1.4, 3.2, 0, 1, ...])*

What AI Can Predict in Drug Discovery

- Will this molecule bind to a biological target?
- Is it toxic or safe?
- Can it dissolve in water - solubility?
- Is it likely to be absorbed in the body - bioavailability?

Introduction to Generative AI

- **Generative AI** refers to models that can generate new data, such as images, text, molecules, etc., based on learned patterns from existing data
- **Key models**: Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Reinforcement Learning
- **Applications**: De novo drug design, molecular optimization, protein structure prediction

How Generative AI Works in Drug Discovery

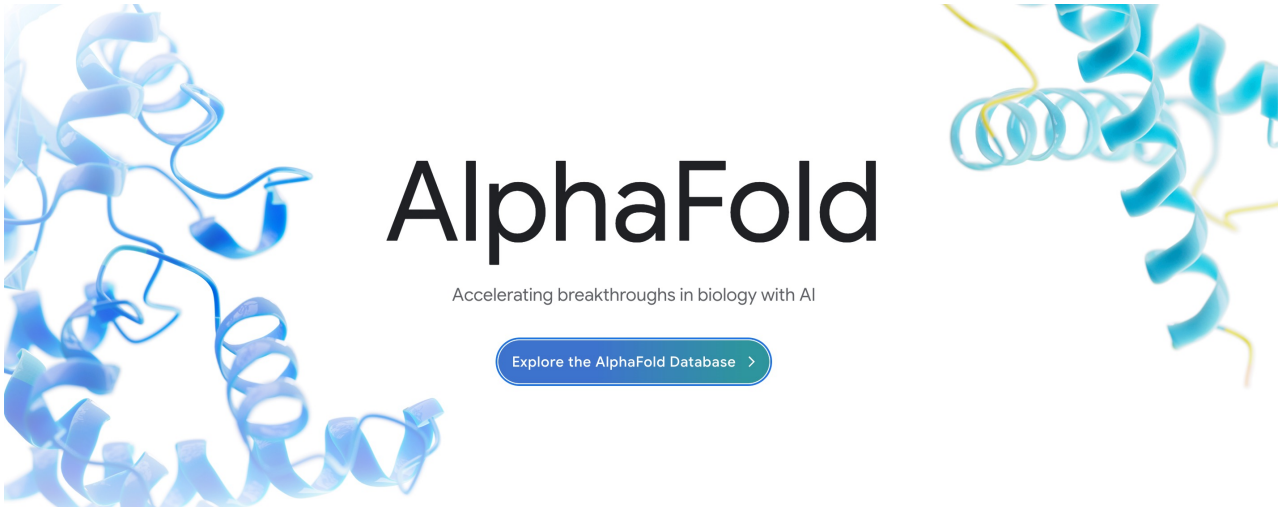
- Training on existing datasets: Generative AI models are trained on large datasets of known compounds, protein-ligand interactions, and chemical properties
- Creating new molecules: Using tools like GANs or VAEs, these models can generate novel chemical structures based on learned properties
- Optimization: After generating new molecules, models can further refine the drug candidates, adjusting their structures to enhance potency and reduce toxicity

What Generative AI Does in Drug Discovery

- Learns from thousands of real drug molecules
- Generates entirely new candidate molecules
- Filters them by properties like drug-likeness
- *drug dataset → AI model → new molecule suggestion*

Example: AlphaFold

- AlphaFold by DeepMind predicts protein structures from amino acid sequences
- Achieves accuracy comparable to X-ray crystallography
- Freely available database with over 200 million protein structures
- https://deepmind.google/technologies/alphafold/

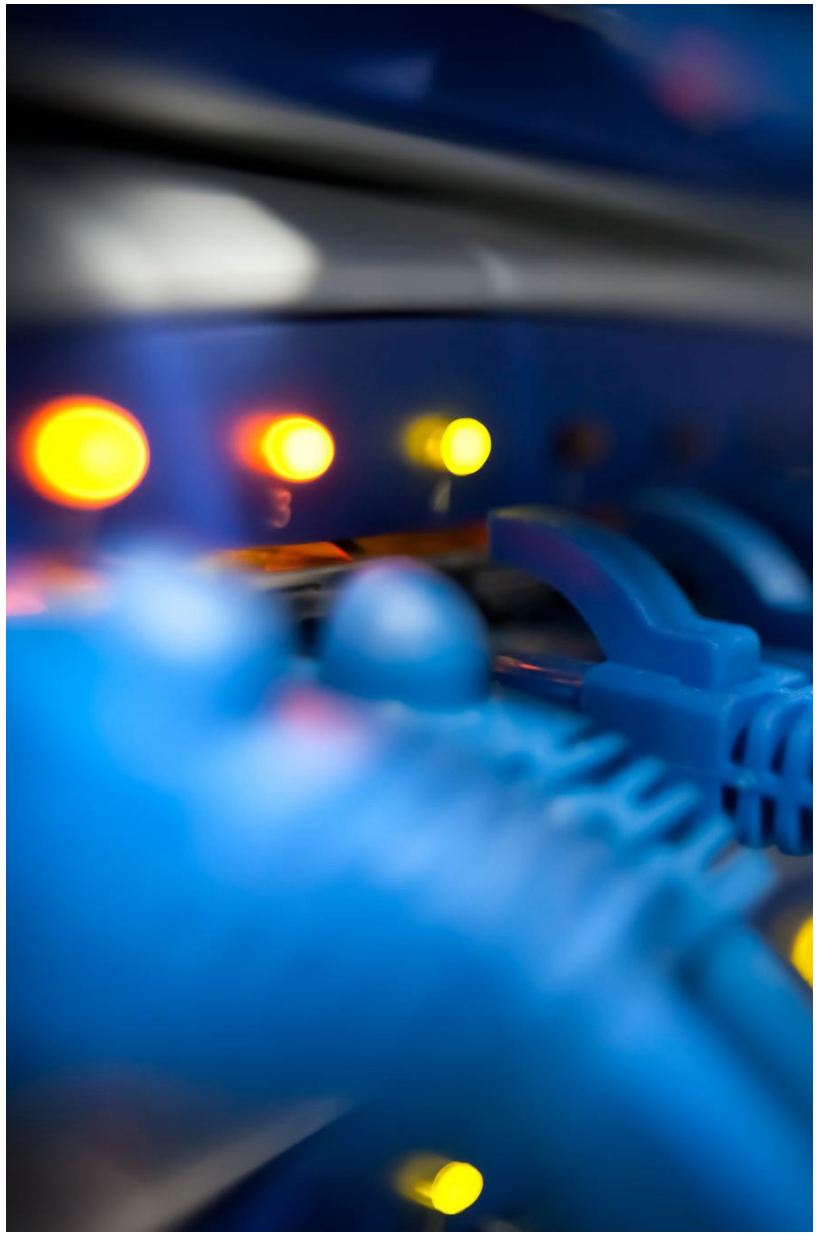


From Data to Drug - Where AI Fits

1. Start with known data (molecular structures, test results)
2. Train an AI model
3. Generate or screen new molecules
4. Lab scientists test top hits in real experiments

What You Can Do

- Try open tools:
  - https://deepchem.io
  - https://rdkit.org
- Run beginner models in Google Colab
- Ask your Computer Science friends for help building models :)
- Look into collaborative drug discovery projects



AI Needs HPC

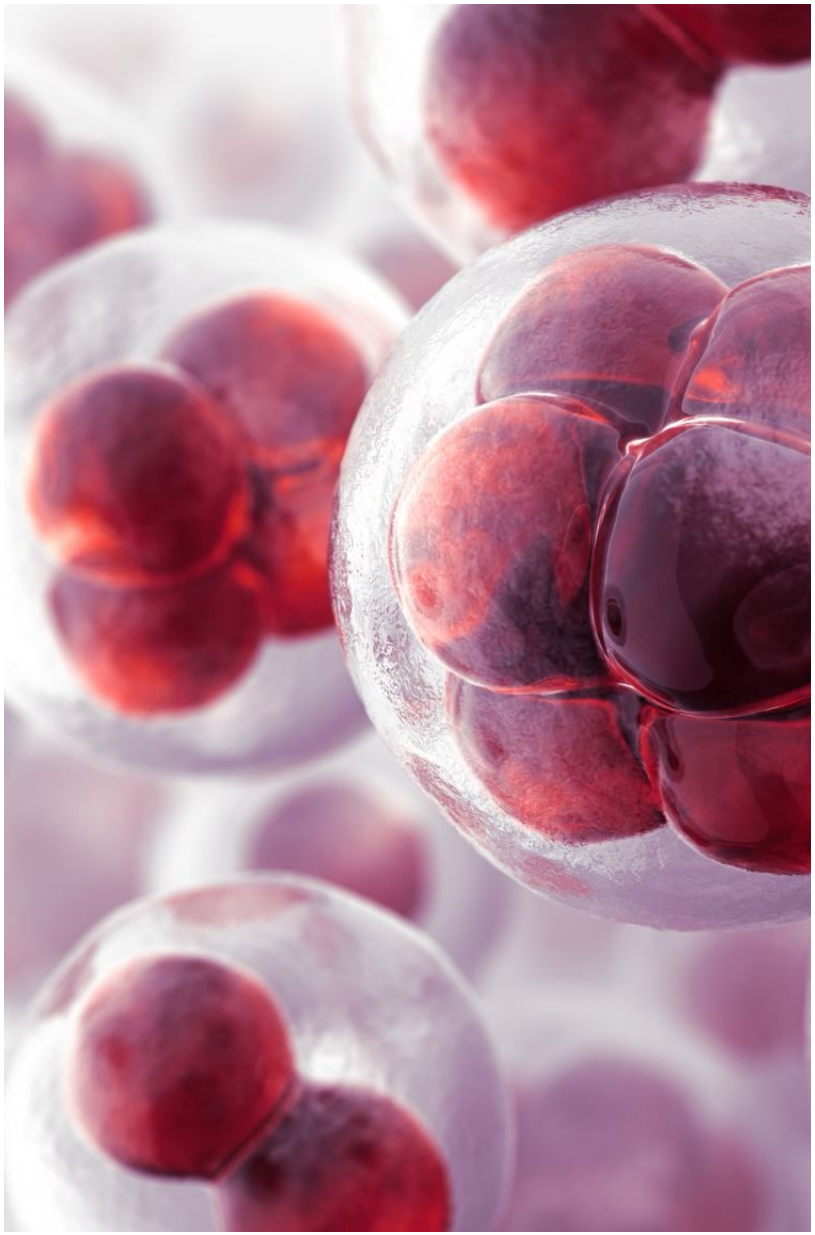
- Training deep neural networks requires massive computational power
- GPUs and TPUs handle parallel processing of large datasets
- HPC supports distributed training and hyperparameter optimization

Pipeline Example

- Simulation generates molecular interaction data
- AI models predict binding affinities or toxicity
- HPC runs large-scale simulations and model training in tandem
- Final candidates are validated with lab experiments

Case Study: Drug Repurposing for COVID-19

- AI and HPC helped repurpose existing drugs like remdesivir and ivermectin for COVID-19
- Simulations screened millions of compounds quickly
- AI models predicted the interactions of these drugs with the viral spike protein



Case Study: AI in Cancer Drug Development

- AI algorithms analyzed genetic mutations in cancer cells to suggest potential drug targets
- HPC simulations modeled the interactions of these drugs with mutated proteins
- Results: faster identification of promising cancer therapies

Benefits of AI-HPC Synergy

- Faster drug development timelines (months vs. years)
- Reduction in overall costs for early-stage drug discovery
- Improved accuracy in predicting drug efficacy and safety

Getting Started with HPC in Drug Discovery

- **Install GROMACS** or **OpenMM** on local machines or HPC clusters
  - GROMACS Installation Guide
  - OpenMM Tutorials
- **Free Resources**:
  - PRACE Training Portal
  - CSC HPC Tutorials
- **Platforms to Try**:
  - EuroHPC JU clusters
  - Google Cloud Research Credits

Getting Started with AI in Drug Discovery

Languages & Frameworks:

- Python, PyTorch, TensorFlow

Toolkits:

- RDKit (cheminformatics)
- DeepChem (ML + molecules)
- scikit-learn (traditional ML)

Datasets to Explore:

- ChEMBL
- ZINC15
- PubChem

The Future of AI in Drug Discovery

- AI models will become more sophisticated and accurate
- HPC will become more accessible with cloud computing
- Precision medicine will rely heavily on AI to match the right drug to the right patient

Conclusion

- AI and HPC are revolutionizing the speed and precision of drug discovery
- Their combination holds the potential to dramatically reduce the cost and time needed for bringing life-saving drugs to market
- The future of drug discovery is data-driven, with AI and HPC at the core

Thanks!

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101191697. The JU receives support from the Digital Europe Programme and Germany, Türkiye, Republic of North Macedonia, Montenegro, Serbia, Bosnia and Herzegovina.