

Integration of LigandDesigner systems: ligand design for laboratory applications and HPC environments

Author: Slawomir A. Wisniewski, SCI4BIZ Research & Development Dept.

e-mail: sa.wisniewski@sci4biz.edu.pl

Abstract

Developing computer-assisted ligand design tools is key to accelerating drug discovery and studying molecular interactions. The article presents two complementary systems: **LigandDesigner - HomeLab**, optimized for laboratories with limited computational resources, and **LigandDesigner - HPC**, designed for high-performance supercomputing environments. Both systems use advanced machine learning models (LSTM, GAN) and integrate molecular dynamics, pharmacophore generation, and chemical validation techniques. The article presents the system architecture, predicted performance test results, and potential applications in life sciences.

Keywords: ligand design, receptor signaling, molecular dynamics simulations, targeted therapy, generative models, breast cancer, high-performance computing

Github repositories:

<https://github.com/SlawomirWisniewski73/LigandDesigner>

INTRODUCTION

The discovery and development of new drugs require significant time and financial resources. In recent years, computational technologies have become a key element in accelerating these processes, particularly in ligand design. Ligand design develops molecules capable of interacting with binding sites in proteins that can act as therapeutic targets. A key challenge is predicting potential ligands' chemical, pharmacological, and physical properties.

The development of machine learning (ML) techniques and the increased availability of computing infrastructure have opened new opportunities in this field. Generative models, such as GAN and LSTM, allow the design of new chemical structures from existing data, and molecular dynamics simulations allow a detailed analysis of their interactions with receptors (Gómez-Bombarelli et al., 2018; Kadurin et al., 2017).17). Molecular dynamics simulations play a key role in the analysis of ligand-receptor interactions, and their integration with machine learning opens new opportunities in research (Hollingsworth & Dror, 2018; Noé et al., 2019). Recent advances in deep learning, including graph neural networks and reinforcement learning, have shown significant promise in predicting molecular interactions and accelerating MD simulations (Zhang et al., 2018).

Previous work indicates that activation of the EDAR/XEDAR pathway by administration of appropriate ligands may be an effective therapeutic strategy in breast cancer (Wisniewski, 2024). To realize this concept, developing a tool to design specific ligands is necessary. This paper presents LigandDesigner, which uses generative models and molecular simulations to support ligand design. The diversity of working environments - from simple laboratories to supercomputers - means no universal solution exists. In response to this need, two complementary systems have been developed:

- **LigandDesigner - HomeLab** - a system for smaller teams and labs to work in a limited computing environment.
- **LigandDesigner - HPC** - a tool dedicated to large projects requiring high throughput and scalability in HPC environments.

This paper presents these systems as an example of the synergy between simplicity, accessibility, and advanced computing architecture.

SYSTEMS ARCHITECTURE

LigandDesigner - HomeLab

LigandDesigner - HomeLab is a tool optimized for environments with limited hardware resources:

- **Components:** Analysis of binding sites using pharmacophores. Ligand generation using conformational optimization and predictive models. Validation of ligands in the chemical and synthetic range.
- **Technologies:** Python, RDKit, scikit-learn. Optional GPU support for performance-sensitive computing.
- **Applications:** Local ligand design. Studies of the mechanisms of protein binding to ligands.

LigandDesigner - HPC

LigandDesigner - HPC uses a supercomputing infrastructure to perform large-scale computing:

- **Components:** Generative models (LSTM and GAN) for designing new chemical compounds. The GAN and LSTM models in LigandDesigner - HPC build upon the foundation laid by previous works, demonstrating the ability of AI to generate realistic and chemically valid ligand structures (Behler & Parrinello, 2007; Gómez-Bombarelli et al., 2018). Parallel molecular dynamics simulation (MPI). Distributed workflow management with support for ML, molecular docking, and result analysis.
- **Technologies:** PyTorch, OpenMPI, CUDA. Optimization of communication between GPU nodes using NCCL.
- **Applications:** High-throughput research in the search for ligands. Collaboration with cloud platforms and HPC clusters.

Workflow and Applications (Fig. 1 – A, B)

Both systems are based on a similar workflow, which includes:

1. **Binding site analysis:** LigandDesigner - HomeLab uses pharmacophores to identify key features of the binding site. LigandDesigner - HPC additionally integrates real-time molecular simulations.
2. **Ligand generation:** GAN models create new 3D structures using chemical constraints. LSTM models allow for the generation of pharmacophore-matched chemical sequences. The use of AI-enhanced MD simulations is inspired by approaches such as deep potential molecule dynamics and Boltzmann generators, which balance computational efficiency with quantum-level accuracy (Zhang et al., 2018; Noé et al., 2019).
3. **Optimization and validation:** Conformation optimization is the key for both systems. Integrating pharmacophore modeling and chemical property prediction is based on established tools such as RDKit and machine learning models for ADMET predictions (Landrum, 2013; Chan et al., 2019). Chemical validation and ML predictions (e.g., logP, toxicity) will be more detailed in LigandDesigner - HomeLab, while LigandDesigner - HPC will focus on high-throughput binding assays.

Predicted performance results

LigandDesigner - HPC achieves scalability comparable to state-of-the-art distributed computing frameworks, leveraging optimized GPU communication protocols like NCCL (Gropp, 2014). The prediction demonstrates the potential of AI-driven MD simulations to

achieve quantum mechanical accuracy, as supported by works like SchNet and Deep Molecular Dynamics (Schütt et al., 2018).

LigandDesigner - HomeLab

Assuming tests on hardware with a CPU and a single GPU (NVIDIA RTX 3060):

- The generation time of a single ligand: ~2-3 min.
- GPU memory consumption: 2-3 GB.
- Pharmacological validation: <1 min/ligand.

LigandDesigner - HPC

Assuming testing on a cluster (Xeon Platinum 8268, NVIDIA V100, 8 GPU nodes):

- Generation of 10,000 ligands: ~4 hrs.
- Scaling efficiency in 8 nodes: ~92%.
- Molecular dynamics (50 ns): ~6 hrs.

DISCUSSION

Using artificial intelligence to predict ADMET properties and synthetic feasibility aligns with the vision of using data-driven models for drug discovery (Vamathevan et al., 2019). The projected performance test results of LigandDesigner - HomeLab and LigandDesigner - HPC assume their complementarity in diverse research environments. **LigandDesigner - HomeLab** will be helpful in labs with standard hardware, such as multi-core processors and single GPU cards. Thanks to its flexibility and ease of installation, it can be used in academic research, teaching, and the initial stages of drug design.

On the other hand, **LigandDesigner - HPC**, run on high-performance clusters, will provide the high throughput needed for projects requiring the generation and analysis of tens of thousands of ligands. The efficiency of scaling across multiple GPU nodes makes it ideal for scientific consortia and industrial projects where speed and accuracy are required. LigandDesigner - HPC complements tools such as AlphaFold and Deep Potential MD by focusing on ligand-receptor interactions and high-throughput ligand generation (Jumper et al., 2021).

Comparing the two systems, the key factor in choosing a tool is the availability of computing infrastructure and the project's specifics. In the case of LigandDesigner - HomeLab, the main limitation is computation time, which increases significantly with a more significant number of ligands. In contrast, LigandDesigner - HPC, while powerful, requires specialized hardware and knowledge of cluster configuration.

CONCLUSIONS

- LigandDesigner - HomeLab fills a gap in access to advanced ligand design tools for laboratories with limited resources.
- LigandDesigner - HPC supports large-scale research projects, enabling scalable computing on supercomputers.

Future directions for both systems include integrating cloud computing and developing AI modules to predict pharmacological properties. Special attention should be paid to aligning both systems with requirements related to regulatory issues, such as compliance with the International Council for Harmonization (ICH) guidelines and the ability to export results in formats supported by open-access platforms.

AUTHOR CONTRIBUTIONS STATEMENT

SAW conceived and designed the study, performed the experiments, analyzed the data, wrote the manuscript, and approved the final version.

CONFLICT OF INTEREST STATEMENT

The author declares that there is no conflict of interest. The research was carried out without commercial or financial relationships that could be construed as a potential conflict of interest.

Bibliography

1. Behler, J. & Parrinello, M. (2007). Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. 10.1103/PhysRevLett.98.146401. Physical Review Letters. vol. 98. no. 14
2. Chan, H. S., Shan, H., Dahoun, T., Vogel, H. & Yuan, S. (2019). Advancing Drug Discovery via Artificial Intelligence. 10.1016/j.tips.2019.06.004. Trends in Pharmacological Sciences. vol. 40. no. 8
3. Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P. & Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. 10.1021/acscentsci.7b00572. ACS Central Science. Vol. 4. no. 2
4. Gropp, W., (2014). Using MPI: Portable Parallel Programming with the Message Passing Interface., MIT Press
5. Hollingsworth, S. A. & Dror, R. O. (2018). Molecular Dynamics Simulation for All. 10.1016/j.neuron.2018.08.011. Neuron. Vol. 99. no. 6
6. Jumper, J., Evans, R., Pritzel, A., Er, ., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgl, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S. et al. (2021). Highly accurate protein structure prediction with AlphaFold. 10.1038/s41586-021-03819-2. Nature. vol. 596. no. 7873
7. Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A. & Zhavoronkov, A. (2017). druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. 10.1021/acs.molpharmaceut.7b00346. Molecular Pharmaceutics. vol. 14. no. 9
8. Landrum, G. (2013). RDKit: Open-Source Cheminformatics Software. 2024-12-17
9. Noé, F., Olsson, S., Köhler, J. & Wu, H. (2019). Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. 10.1126/science.aaw1147. Science. vol. 365. no. 6457
10. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A. & Müller, K. R. (2018). SchNet - A deep learning architecture for molecules and materials. 10.1063/1.5019779. the Journal of Chemical Physics. vol. 148. no. 24
11. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M. & Zhao, S. (2019). Applications of machine learning in drug discovery and development. 10.1038/s41573-019-0024-5. Nature Reviews Drug Discovery. vol. 18. no. 6
12. Wisniewski, S. (2024). The Development of Differentiation-Based Therapy for Breast Cancer Using EDAR and XEDAR Signaling Pathways: A Conceptual Framework. 10.32388/1CLGH0
13. Zhang, L., Han, J., Wang, H., Car, R. & E, W. (2018). Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. 10.1103/PhysRevLett.120.143001. Physical Review Letters. vol. 120. no. 14