Academic Grant Program

Dynamic Q-Former: Bridging Molecular Graphs with LLMs for Drug Discovery

Project Title

Project Name: Dynamic Q-Former: Bridging Molecular Graphs with LLMs for Drug Discovery

Principal Investigator Information

Name: Pingzhao Hu
Title: Canada Research Chair in Computational Approaches to Health Research, Associate Professor
University Name: University of Western Ontario

Project Collaborators

Research Team Member Name: Qiuhao Zeng
Role: PhD Student, Department of Computer Science
University Name: University of Western Ontario

Research Team Member Name: Zihao Jing
Role: PhD Student, Department of Computer Science
University Name: University of Western Ontario

Abstract

Generative AI has advanced molecular science through domain-specific foundation models, yet molecular AI remains siloed across modalities, limiting large language models in molecular reasoning and mechanism elucidation. Existing multimodal approaches suffer from two key challenges: loss of structural fidelity due to shallow, symmetric fusion with fixed-length queries, and heavy fine-tuning requirements that jointly train connectors with large LLMs, resulting in inefficient and fragile optimization.

We propose the Dynamic Query-Former (DQ-Former), the first alignment bridge that connects frozen molecular encoders with frozen LLMs. DQ-Former introduces dynamic queries in correspondence with entropy-guided informative sub-graphs, preserving substructures that are essential for chemical reasoning. This design provides structure-aware alignment while eliminating the need to fine-tune the backbone. The project will deliver (1) an entropy-guided DQ-Former that ensures structural fidelity through substructure-aware queries, (2) a frozen-backbone multimodal bridge for efficient and stable training, and (3) state-of-the-art improvements in retrieval accuracy and reasoning fidelity on molecular benchmarks.

Project Keywords

Multimodal Fusion, Generative Drug Discovery, Entropy-Guided Alignment, Molecule Modeling.

NVIDIA Platforms

A100-40GB GPUs with CUDA and NCCL are for efficient Torch and Transformers training framework. Containers from NVIDIA NGC will support reproducibility. Our team has prior experience with CUDA-enabled PyTorch on HPC clusters, enabling smooth adoption of NVIDIA cloud resources.

Dataset and Model

- Data: SmolInstruct (pretraining), Mol-Instructions (fine-tuning), MolecularQA, TDC benchmarks.
- Type: SMILES (1D), molecular graphs (2D), conformers (3D), biomedical text.
- Source: Open-source datasets published on Huggingface, no NDA required.
- Models: Frozen graph encoders (UniMol, MuMo), DQ-Former fusion module (~100M parameters), and frozen LLMs (Mistral-7B, LLaMA-2, and LLaMA-3 backbones).

Introduction

Background and Challenges. Domain-specific foundation models have reshaped biomedicine, such as Galactica for scientific text-data (Taylor et al., 2022a), and Uni-Mol for molecular modeling (Zhou et al., 2023). However, existing models remain modality-siloed, limiting generative LLMs in molecule reasoning and mechanism elucidation. Existing multimodal systems such as Zeng's molecule-text bridge (Zeng et al., 2022), 3D-MolT5 (Zhao et al., 2024), and Mol-LLM (Lee et al., 2025a) attempt to align molecular modalities with language but face two challenges: (1) **loss of structure**, as symmetric, shallow fusion with fixed-length queries compresses stereochemistry and substructures, causing representation collapse and

low reasoning fidelity (Fang et al., 2023); and (2) **heavy fine-tuning**, since most bridges jointly train the connector and fine-tune the LLM to achieve alignment, leading to compute-inefficient.

Contribution and Impact. We propose the Dynamic Query-Former (DQ-Former), the first alignment bridge that connects frozen molecular encoders with frozen LLMs in chemical field. Unlike prior methods that require fine-tuning and lake of structural detail, DQ-Former introduces dynamic queries correspondence with entropy-based informative sub-graph, ensuring that functional groups are explicitly preserved. This structure-aware alignment allows LLMs to access chemically faithful representations while keeping both backbones frozen for efficiency and stability. This project is expected to deliver:

- An entropy-guided DQ-Former that preserves structural fidelity through substructure-aware queries.
- A frozen-backbone multimodal bridge, eliminating the need to fine-tune graph encoders or LLMs.
- State-of-the-art improvements on both reasoning and traditional molecular and chemical benchmarks.

Methods

Dynamic Q-Former Architecture. We propose a DQ-Former as a lightweight alignment bridge between molecular encoders and frozen LLMs. In contrast to the widely used Q-Former (Li et al., 2023), which relies on a fixed set of query tokens and risks collapsing modality-specific information, the DQ-Former introduces a hybrid query bank: fixed anchor tokens ensure stable cross-modal alignment, while dynamic fragment-conditioned tokens capture substructure-level fidelity (Figure 1).

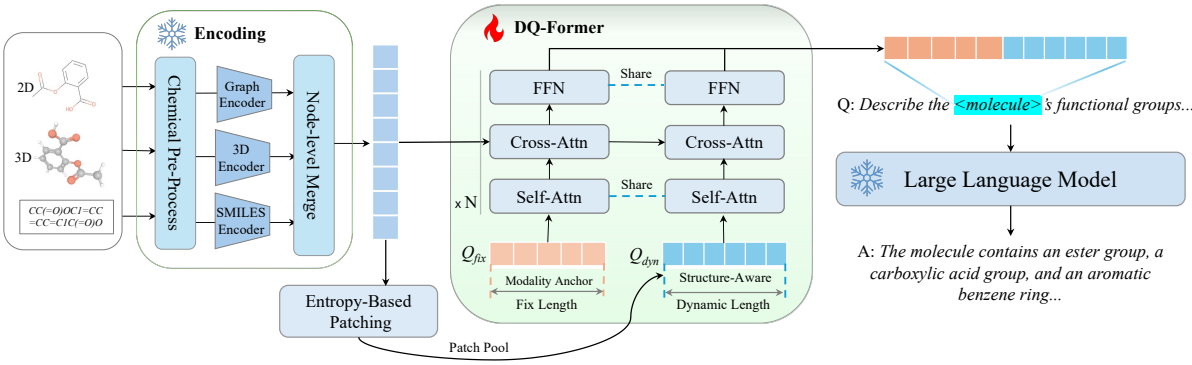

Figure 1: Architecture of the Dynamic Q-Former with frozen molecular encoders and LLM.

In Figure 2, we plot atom-level entropy along a SMILES string; taller peaks mean the next-atom predictor is most uncertain. At each pronounced local maximum, we terminate the current segment and start a new substructure at the next atom. This yields variable-length patches (dashed separators in Figure 2) that concentrate around functional groups such as carboxyls and amines, without hand-crafted rules. We then map patch indices back to graph nodes. These entropy-guided fragments serve as the pooling units that DQ-Former uses for graph-LLM alignment.

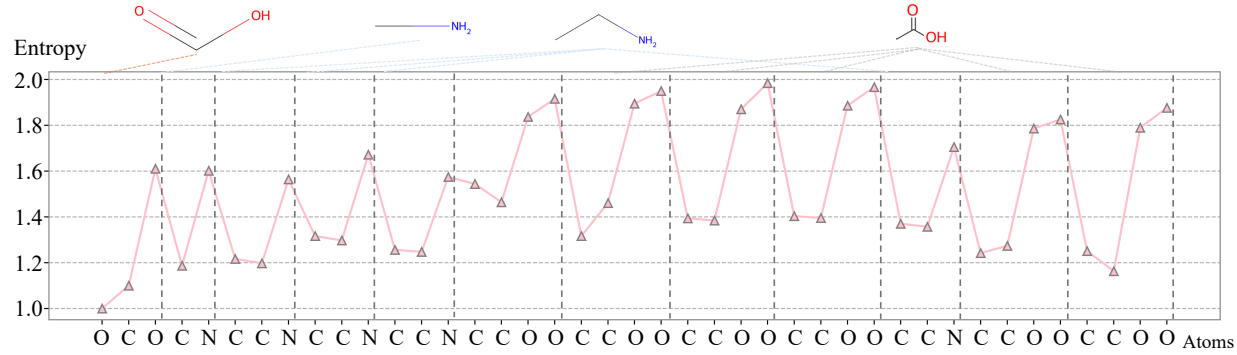

Figure 2: Atom-wise entropy illustration of the molecule polyaminopolycarboxylate.

Training. The DQ-Former is pre-trained with contrastive learning strategy, where fixed anchor tokens enforce modality alignment and dynamic queries capture substructure fidelity. Multimodal inputs (1D to 3D) are projected into a shared latent space through a combination of contrastive

alignment, modality-matching, and substructure reconstruction, ensuring both global cohesion and fine-grained chemical semantics. After pretraining, we perform end-to-end fine-tuning while keeping the molecular encoders and LLM frozen. Only the DQ-Former is updated, allowing the model to efficiently adapt to chemistry instructions. This process enables the LLM to acquire chemically grounded understanding of molecular structures without retraining large backbone models.

Software & Hardware help. We implement the framework in PyTorch with Transformers Trainer, leveraging A100-40GB GPUs with CUDA/NCCL for efficient multimodal LLM training at scale. This integration enables scalable experiments that would be infeasible on our own standard compute.

Expected Results

Resource Estimation and Outcomes. As shown in Figure 3, the project is planned as a four-month effort requiring ~8,000 GPU hours on NVIDIA A100-40GB. Resources will be concentrated in the middle phase for DQ-Former pre-training, fine-tuning, and benchmarking. Expected outcomes are as follows:

- Release of an entropy-guided Dynamic Q-Former model, related code, and constructed data.
- A multimodal framework enabling efficient alignment between frozen encoder and LLM.
- Demonstrated SOTA performance on molecular representation and reasoning benchmarks, with results submitted to ICML2026 conference.

GPU Support. NVIDIA GPUs are indispensable for this project, providing the compute scale needed for multimodal training, high-throughput inference, and instruction tuning with 7~8B LLMs. The training of millions of molecules demand thousands of GPU hours, which are not feasible on academic clusters. With NVIDIA's accelerated infrastructure, we can implement and test DQ-Former at scale.

Preliminary Results. DQ-Former outperforming fine-tuned baselines and surpassing the original Q-Former bridge (Table 1). With NVIDIA support, these promising results can be scaled into a robust framework for multimodal learning field.

Table 1 Results on Pampa from TDC benchmark.

Models	Base model	Acc/%
GPT-4o	GPT-4o	48.65
Mol-Instructions	Llama2-7B	49.63
Mol-LLaMA	Llama2-7B	<u>75.68</u>
3D-MoLM	Llama3-8B	46.93
LLaMo	Llama3-8B	49.25
UniMol + Q-Former	Llama3-8B	73.46
UniMol + DQ-Former	**Llama3-8B**	**82.56**

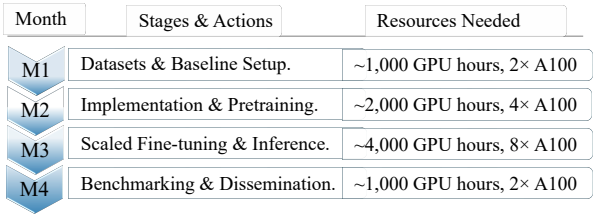

Figure 3: Project timeline and resource estimation

Project Support Details

This project requires cloud resources to complete pretraining, finetuning, and benchmarking. The total request is well within the CFP limits and justified by the scale of datasets, model size, and training process in Figure 3.

- Hours in the cloud needed: ~8,000 GPU hours (A100-40GB)
- Number of concurrent GPUs needed: up to 8
- Amount of cloud storage needed: ~5 TB (molecular datasets, checkpoints, intermediate results)
- Physical hardware needed: None

Cloud Readiness

Our development environment runs on Ubuntu Linux. Upon receiving cloud GPU access, we will:

- Import molecular datasets (e.g. MoleculeQA, Mol-Instructions, TDC) from Huggingface and TDC.
- Install dependencies via pip (e.g. PyTorch, Transformers, RDKit).
- Set up and run our codebase using Cursor editor.
- Regularly back up intermediate results to persistent storage.
- Install and integrate CUDA® and NVIDIA SDKs for training acceleration.
- Build and run containers for reproducibility and scalability.

Appendix A – CV(s)

Principal Investigator: Pingzhao Hu, Ph.D.

WORK ADDRESS：Department of Biochemistry, Department of Computer Science
Schulich School of Medicine & Dentistry, Western University
Phone: (204) 898-0360 (Cell) Email: phu49@uwo.ca
Website: https://phulab.org/

Employment

Associate Professor, Department of Biochemistry, Western University	2022 – Present
Associate Professor, Department of Computer Science, Western University	2022 – Present
Associate Professor, Department of Biochem & Medical Genetics, University of Manitoba	2020 – 2022
Adjunct Professor, Department of Computer Science, University of Manitoba	2017 – 2022
Assistant Professor**,** (Status-Only position) Division of Biostatistics, University of Toronto	2013 – Present
Assistant Professor, Department of Biochem & Medical Genetics, University of Manitoba	2014 – 2020
Manager, Statistical Analysis Facility of The Centre for Applied Genomics, SickKids, Toronto	2006 – 2014
Research Biostatistician, Program in Genetics and Genomic Biology, SickKids, Toronto	2004 – 2005
Research Biostatistician, The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital	2003 - 2003

Education

2005– 2012	PhD, Computer Science, York University, Toronto, ON, Canada
2001– 2002	MSc, Computer Science, Dalhousie University, Halifax, NS, Canada

Selected Honors and Career Awards

2022 Tier 2 Canada Research Chair in Computational Approaches to Health Research, Western Univ.
2020 Terry G. Falconer Memorial Rh Institute Foundation Emerging Researcher Award, U. Manitoba
2018 The Interstellar Initiative Award**,** New York Academy of Sciences and AMRD, Japan

Selected Research Funding between 2020 and 2025

2025-2030	CIHR, 2025 Spring Project Grants	$1,067,176 CAD
2023–2028	CIHR, 2023 Spring Project Grants	$830,026 CAD
2023–2024	CIHR, 2022 Fall Project Grants (Priority Grant)	$100,000 CAD
2022–2027	CIHR, Tier 2 Canada Research Chair Award	$600,000 CAD
2021–2022	Mitacs, Accelerate Program	$90,000 CAD
2021–2026	Natural Science and Engineering Research Council of Canada	$210,000 CAD
2020– 2022	CancerCare Manitoba Foundation, Research Operating Funding	$150,000 CAD
2020– 2023	Manitoba Medical Service Foundation Career Research Award	$85,000 CAD

Selected Peer-reviewed Journal articles (I am the corresponding author for the articles) in 2025

1. Y Sun, Y Lu, YY Li, Z Jing, P Hu (2025) MolGraph-xLSTM: A graph-based dual-level xLSTM framework with multi-head mixture-of-experts for enhanced molecular representation and interpretability. *Communications Chemistry,* Accepted.
2. Y Lu, YY Li, Y Sun, P Hu (2025) FusionLCLM: Enhanced Molecular Property Prediction via Knowledge Fusion of Large Chemical Language Models. *Journal of Cheminformatics*.
3. G Zhou, S Janarthanan, Y Lu, P Hu (2025). CL-MFAP: A contrastive learning-based multimodal foundation model for antibiotic property prediction. The Thirteenth *International Conference on Learning Representations* (ICLR 2025). Singapore
4. H Hadipour, YY Li, Y Sun, A Deng, L Lac, R Davis, ST Cardona**, P Hu** (2025). GraphBAN: An Inductive Graph-Based Approach for Enhanced Prediction of Compound-Protein Interactions. *Nature Communications*. 16:2541.

Project Collaborator 1: ZENG QIUHAO

Email: qzeng53@uwo.ca — LinkedIn Profile — Google Scholar

Professional Summary

Fourth-year PhD candidate at the University of Western Ontario, Machine Learning Group, supervised by Prof. Charles Ling (Fellow, Canadian Academy of Engineering) and Prof. Boyu Wang. Research focuses on: Efficient transformer architectures utilizing Triton kernels and Transfer learning in dynamic temporal domains.

Previous experience: Research Associate in the Brain-Computer Interface Group at Nanyang Technological University under IEEE Fellow Prof. Cuntai Guan, focusing on EEG signal processing and classification.

Education

- **PhD in Computer Science**, University of Western Ontario. Jan 2022 – Present
- **M.Sc in Electrical Engineering**, National University of Singapore. Aug 2017 – Jun 2018
- **Bachelor in Engineering Mechanics**, Harbin Institute of Technology Sep 2013 – Jul 2017

Work Experience

- **Associate Researcher**, Huawei Noah's Ark Lab NLP Team, Montreal. Jul 2024 – Present
  - Researching efficient attention architecture for next-generation Large Language Models.
- **Software Engineer**, LITEON Singapore. Jul 2018 – Mar 2019
  - Developed machine vision algorithms for camera manufacturing testing and validation.
- **Research Associate**, Nanyang Technological University. Mar 2019 – Jun 2021
  - Developed rehabilitation games based on EEG brain-computer interfaces.

Publications

- ZETA: Leveraging *Z*-order Curves for Efficient Top-*k* Attention

***Qiuhao Zeng**, Jerry Huang, Peng Lu, Gezheng Xu, Boxing Chen, Charles Ling, Boyu Wang*
International Conference on Learning Representations (ICLR), 2025.

- Calibrated Language Models and How to Find Them with Label Smoothing

*Jerry Huang, Peng Lu, **Qiuhao Zeng***
International Conference on Machine Learning (ICML), 2025.

- Towards Understanding Evolving Patterns in Sequential Data

***Qiuhao Zeng**, Long-Kai Huang, Qi Chen, Charles Ling, Boyu Wang*
Conference on Neural Information Processing Systems (NeurIPS), 2024 (**Spotlight, Top 2.1%**)

- On the Benefits of Attribute-Driven Graph Domain Adaptation

*Ruiyi Fang, Bingheng Li, Zhao Kang, **Qiuhao Zeng**, Nima Hosseini Dashtbayaz, Ruizhi Pu, Boyu Wang,* Charles Ling
International Conference on Learning Representations (ICLR), 2025.

- Homophily Enhanced Graph Domain Adaptation

*Ruiyi Fang, Bingheng Li, Jingyu Zhao, Ruizhi Pu, **Qiuhao Zeng**, Gezheng Xu, Charles Ling, Boyu Wang* International Conference on Machine Learning (ICML), 2025.

- **Latent Trajectory Learning for Limited Timestamps under Distribution Shift over Time**

***Qiuhao Zeng**, Changjian Shui, Long-Kai Huang, Peng Liu, Xi Chen, Charles Ling, Boyu Wang International Conference on Learning Representations (ICLR), 2024* (**Oral, Top 1.2%**).

- Generalizing across Temporal Domains with Koopman Operators

***Qiuhao Zeng**, Wei Wang, Fan Zhou, Gezheng Xu, Ruizhi Pu, Changjian Shui, Christian Gagn´e, Shichun* Yang, Charles Ling, Boyu Wang
AAAI Conference on Artificial Intelligence (AAAI), 2024.

- Foresee What You Will Learn: Data Augmentation for Domain Generalization in Non-Stationary Environments

***Qiuhao Zeng**, Wei Wang, Fan Zhou, Charles Ling, Boyu Wang*
AAAI Conference on Artificial Intelligence (AAAI), 2023.

Project Collaborator 2: Zihao Jing
zjing29@uwo.ca

EDUCATION

Western University (UWO) — London, CA
PhD of Computer Science · 2024.09 – Present
Beihang University — Beijing, China
Bachelor of Engineering in Software Engineering · 2019.09 – 2024.06

PUBLICATIONS

- Huang, J., Hu, Z., **Jing, Z.**, Gao, M., & Wu, Y. (2024). *Piccolo2: General text embedding with multi-task hybrid loss training.* arXiv:2405.06932. DOI: 10.48550/arXiv.2405.06932.
- Y Sun, Y Lu, YY Li, Z Jing, P Hu (2025) MolGraph-xLSTM: A graph-based dual-level xLSTM framework with multi-head mixture-of-experts for enhanced molecular representation and interpretability. Communications Chemistry, Accepted.
- **Jing, Z.** (2024). *Graph Pruning Based Spatial and Temporal Graph Convolutional Network with Transfer Learning for Traffic Prediction.* arXiv:2409.16532.
- **Jing, Z.** (2023). *The Application and Implementation of Blockchain Techniques in the Medical Industry.* In *Proceedings of the 4th International Conference on Economic Management and Model Engineering (ICEMME 2022)* (pp. 555–561). SciTePress.

RESEARCH & PROJECTs

Current: Multimodal LLM for Biomolecular & Chemistry — 2024.09 – Present
Western University — Computer Science

- Proposed MuMo, a structure-aware multimodal fusion framework for molecular representation learning, achieving state-of-the-art performance on 21 TDC and MoleculeNet benchmarks, with up to 27% improvement on LD50.
- Submit Paper: Structure-Aware Fusion with Progressive Injection for Multimodal Molecular Representation Learning, Nips 2025.

AI-Engine Intelligent Transportation — Senior Researcher Intern — 2022.07 – 2023.07
Beijing Intelligent Transportation Research Center

- Designed and implemented a spatio-temporal graph convolutional prediction network, achieving a prediction error (MAE) of 3.5 or less within 30 minutes, utilizing graph pruning and transfer learning with the paper accepted by the ICIAAI conference.
- Designed and developed a real-time camera vehicle monitoring and road condition indicator system based on the Target Detection algorithm.
- Launched product serving the Shanghai section of the Shanghai G60 Expressway and related toll stations.

WORK EXPERIENCE

SenseTime Co., Ltd. — LLM Researcher Intern — 2023.09 – 2024.06

- Fine-tuned 100B-parameter large language model for vertical livestream marketing; led model optimization and iteration; successfully deployed in production for Sina Weibo.
- Developed Chinese NLP embedding models; independently built training and evaluation pipelines; led model iterations; achieved top-1 performance on the MTEB Chinese benchmark.

Appendix B – Citations

Li, J., Li, D., Savarese, S., & Hoi, S. (2023). Bootstrapping language-image pre-training with frozen image encoders and large language models. *ICML 2023*, Proceedings of Machine Learning Research (PMLR), 202:19730–19742.

Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z., Zhang, L., & Ke, G. (2023). A universal 3D molecular representation learning framework. *ICLR 2023* (poster).

Zeng, Z., Yao, Y., Liu, Z., & Sun, M. (2022). A deep-learning system bridging molecule structure and biomedical text with comprehension comparable to human professionals. *Nature Communications, 13*, Article 862.

Zhao, Z., Ma, D., Chen, L., Sun, L., Li, Z., Xia, Y., Chen, B., Xu, H., Zhu, Z., Zhu, S., Fan, S., Shen, G., Yu, K., & Chen, X. (2025). Developing ChemDFM as a large language foundation model for chemistry. *Cell Reports Physical Science, 6*(4), 102523.

Fang, Y., Liang, X., Zhang, N., Liu, K., Huang, R., Chen, Z., Fan, X., & Chen, H. (2024). A large-scale biomolecular instruction dataset for large language models. *ICLR 2024* (poster).

Lee, C., Song, Y., Jeong, Y., Ko, H., Hormazabal, R., Han, S., Bae, K., Lim, S., & Kim, S. (2025). Generalist molecular LLM with improved graph utilization. *arXiv:2502.02810*.

Taylor, R., Borisov, V., Le Bras, R., et al. (2022). A large language model for science. *arXiv:2211.09085*.

Baselines (The order is following Table 1):

Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, Almeida D, Altenschmidt J, Altman S, Anadkat S, Avila R. Gpt-4 technical report. arXiv preprint arXiv:2303.08774. 2023 Mar 15.

Fang Y, Liang X, Zhang N, Liu K, Huang R, Chen Z, et al. Mol-Instructions: A Large-Scale Biomolecular Instruction Dataset for Large Language Models. International Conference on Learning Representations. 2023.

Kim D, Lee W, Hwang SJ. Mol-LLaMA: Towards General Understanding of Molecules in Large Molecular Language Model. arXiv.org. 2025.

Li S, Liu Z, Luo Y, Wang X, He X, Kawaguchi K, et al. Towards 3D Molecule-Text Interpretation in Language Models. International Conference on Learning Representations. 2024.

Park J, Bae M, Ko D, Kim HJ. LLaMo: Large Language Model-based Molecular Graph Assistant. Neural Information Processing Systems. 2024.