AutoRNA: Advancing RNA Motif Structural Analysis with a Fully Automated AI Agent

Ramez Mosad[1], **Burhan Khan**[1], **Vivek Shenoy**[1], **Rahul Jasthi**[2], **Brent M. Znosko**[2], **Jie Hou**[1]

[1]Department of Computer Science, [2]Department of Chemistry,
Saint Louis University, St. Louis, MO

Abstract

RNA underlies gene expression and protein synthesis, yet experimentally resolving its structure is slow, costly, and difficult. Protein-folding research leapt forward with AlphaFold-like models, but RNA prediction still lags because of greater molecular complexity and limited structural data. A deeper grasp of RNA motifs—the recurring elements that dictate folding and function—is critical for next-generation 3D RNA structure prediction, with broad impact on personalized medicine, drug design, and biotechnology. Currently, motif analysis requires intensive programming and computation, placing it out of reach for many students and labs. To lower that barrier, we are building an AI agent that uses large-language-model (LLM) reasoning to automate and streamline diverse RNA motif analysis workflows.

Project History & Status

The team has worked on implementing 3 sets of functions that work in tandem with each other inside our AI agent pipeline (autoRNA)

After running through all of these functions, it should output the reported number of downloaded pdb files along with any other necessary information in regards to the file.

Key Features

1. **Autonomous Multi-Agent Architecture** uses intelligent agents that collaborate independently to complete complex tasks
2. **Minimal Human Intervention** is based around a User Proxy that only approves plans. The rest is fully handled by agents—our system is designed to be mostly self-driving.
3. **Easily Extensible** support allows for new scientific tools to be added, simply by defining a function and updating a config—no need to retrain or re-engineer the pipeline!

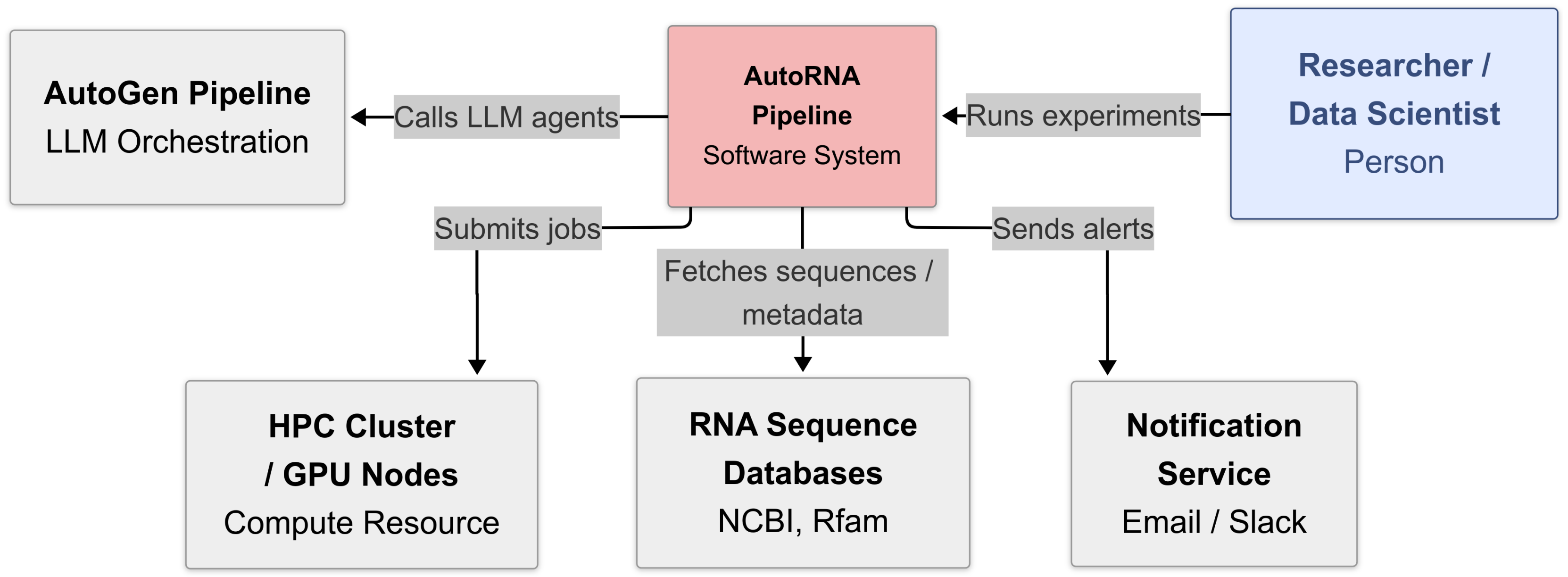
Methodology Infrastructure

- **Frameworks:** AutoGen (Microsoft), OpenAI API
- **Programming Language:** Python
- **Scientific Libraries:** Biopython, Pandas, NumPy
- **File Formats:** PDB, JSON, CSV
- **Environments:** Google Colab, Jupyter Notebook
- **Version Control:** Git, GitHub

SAINT LOUIS UNIVERSITY
SCHOOL OF SCIENCE AND ENGINEERING

System Architecture

Context of the Software



AutoGen Pipeline
LLM Orchestration

AutoRNA Pipeline
Software System

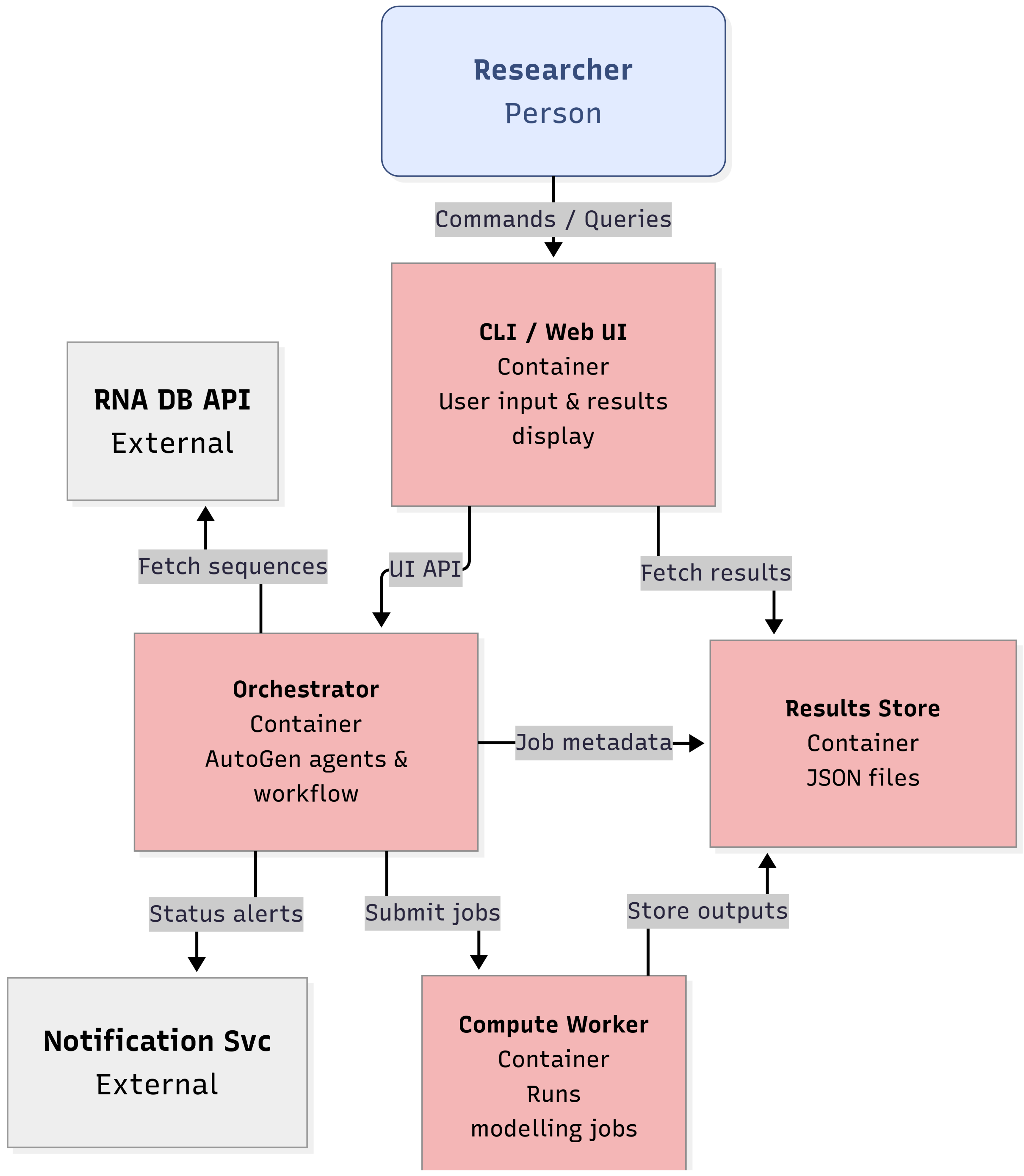
Researcher / Data Scientist
Person

HPC Cluster / GPU Nodes
Compute Resource

RNA Sequence Databases
NCBI, Rfam

Notification Service
Email / Slack

Containers of the Software



Researcher
Person

CLI / Web UI
Container
User input & results display

RNA DB API
External

Orchestrator
Container
AutoGen agents & workflow

Results Store
Container
JSON files

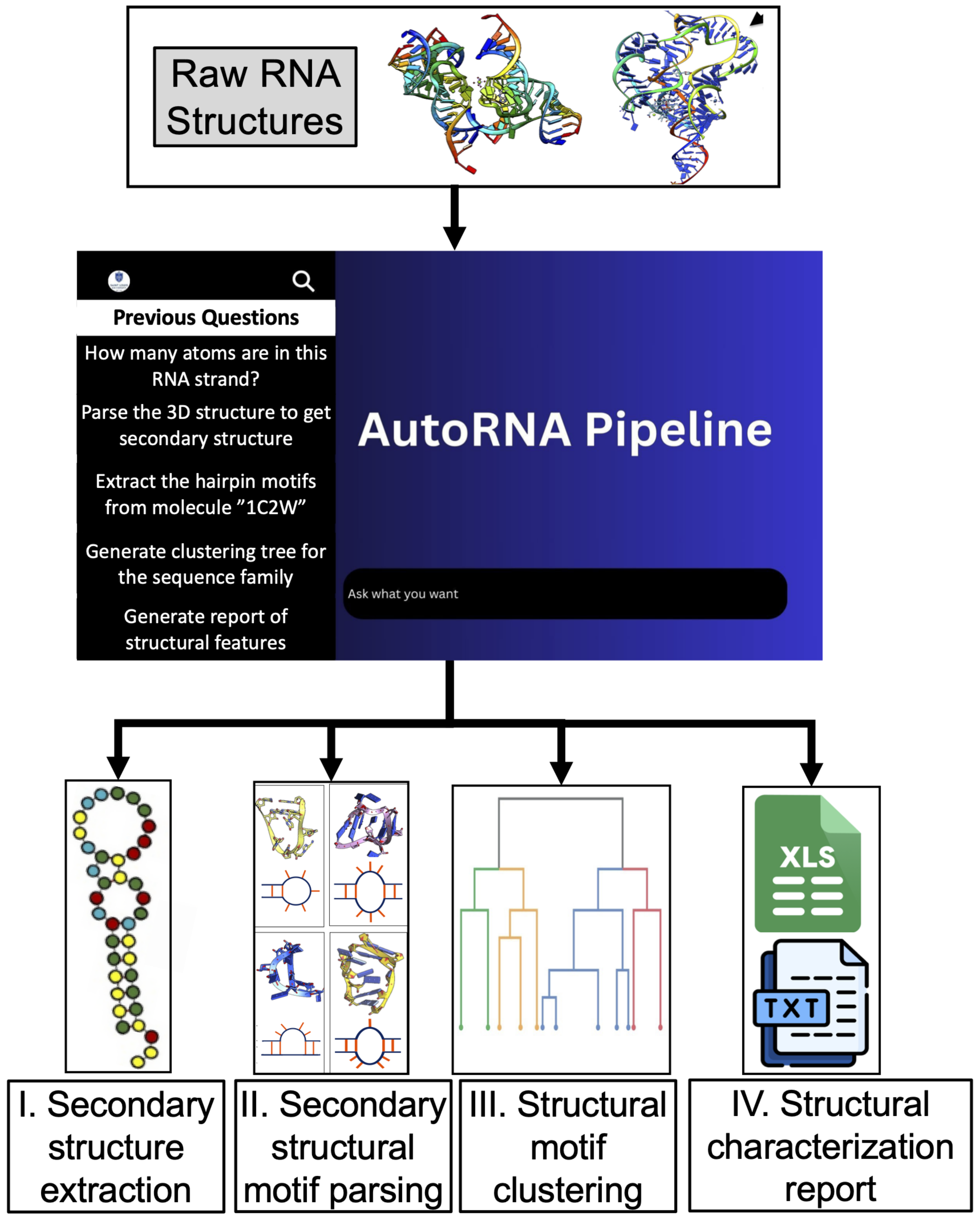
Notification Svc
External

Compute Worker
Container
Runs modelling jobs

Znosko Lab CoSSMos Database

GitHub: https://github.com/JieHou-SLU/autoRNAMotifs

Highlighted Example



Previous Questions

How many atoms are in this RNA strand?

Parse the 3D structure to get secondary structure

Extract the hairpin motifs from molecule ”1C2W”

Generate clustering tree for the sequence family

Generate report of structural features

Future Work

Currently, our group has successfully added functions for the “Download and Clip the .pdb files” and “Reorder Residues” steps of the manual RNA analysis workflow pipeline. Using this as a precedent, future work will include implementing the remaining steps of the current workflow as functions in the autoRNA agent. Furthermore, optional functions beyond the steps of the pipeline will be added such as one that can summarize the characterization output report or recommend values for the RMSD cutoff. Finally, a GUI for the autoRNA agent will be developed and published for student use.

Acknowledgement

This work was supported by the SLU Foundational Interdisciplinary Research Experience (FIRE) program and conducted under the guidance of Drs. Jie Hou and Brent M. Znosko.