**BoltzNotebook: An Interactive Google Colaboratory Platform for Protein-Ligand Structure Prediction and Affinity Analysis**

Atharva Tilewale1, Dhaval Patel1

1Department of Industrial Biotechnology, Gujarat Biotechnology University, Near Gujarat International Finance Tec (GIFT)-City, Gandhinagar, 382355, Gujarat, India

**Abstract**

The accurate prediction of protein-ligand structures and binding affinities is fundamental to modern drug discovery and structural biology. Deep learning models like Boltz2 have emerged as powerful tools for these tasks, yet their reliance on command-line interfaces and complex computational environments presents a significant barrier to many researchers. To address this, we present BoltzNotebook, an open-source Google Colaboratory notebook that provides an accessible, end-to-end platform for running the Boltz2 pipeline. BoltzNotebook features a feature-rich graphical user interface (GUI) that simplifies the input of protein sequences and ligand identifiers (CCD or SMILES), an automated workflow that handles environment setup and execution, and an advanced post-prediction module for automated analysis. This module generates interactive 3D visualisations of the predicted structure, per-residue confidence (pLDDT) and inter-domain error (PAE) plots, and a comprehensive binding affinity dashboard with metrics such as predicted IC50​ and ΔG. By encapsulating the complexity of the Boltz2 framework within a user-friendly, web-based environment, BoltzNotebook democratizes access to state-of-the-art biomolecular modelling for a broader scientific audience.

**1. Introduction**

The advent of deep learning has revolutionised the field of protein structure prediction. Models such as AlphaFold2 and RoseTTAFold have achieved unprecedented accuracy in predicting the three-dimensional structures of monomeric and multimeric proteins from their amino acid sequences (Jumper et al., 2021; Baek et al., 2021). Building on these successes, new architectures have been developed to tackle the even more complex challenge of modelling protein-ligand interactions, a critical step in computational drug discovery.

Boltz2 is a recent deep learning framework that leverages diffusion models to co-predict protein and ligand structures, as well as their binding affinity (Passaro et al., 2025). Its ability to model both covalent and non-covalent interactions and provide quantitative affinity estimates makes it a powerful tool for virtual screening and lead optimisation. However, like many advanced computational biology tools, its usage requires proficiency with the command line, management of complex software dependencies (e.g., CUDA), and access to sufficient computational resources. These requirements can limit its adoption by bench scientists, educators, and researchers without specialised bioinformatics support.

To bridge this gap between cutting-edge tools and their user base, platforms like ColabFold have successfully demonstrated the value of packaging complex pipelines into accessible Google Colaboratory notebooks (Mirdita et al., 2022). Following this paradigm, we developed BoltzNotebook, an interactive and comprehensive Colaboratory environment designed to make the full power of Boltz2 accessible to all researchers, regardless of their computational background.

**2. Implementation and Workflow**

BoltzNotebook is implemented as a self-contained Jupyter Notebook designed to run on the Google Colaboratory platform. It automates the workflow from setup to analysis in a series of logical, user-executable cells.

**2.1 Pipeline Architecture**

The notebook follows a sequential, five-stage workflow (Figure 1):

1. **Environment Setup:** The notebook first installs all necessary dependencies, including PyTorch, Biopython, and the Boltz2 software with CUDA acceleration. This step ensures a consistent and functional environment for every user.
2. **Parameter Generation:** A custom GUI allows the user to define all input parameters for the prediction job.
3. **Model Execution:** The notebook constructs and executes the boltz predict command, handling the search for Multiple Sequence Alignments (MSAs) and running the core diffusion-based structure prediction model.
4. **Structure Visualisation:** Upon completion, the predicted 3D structure is immediately displayed in an interactive py3Dmol viewer.
5. **Confidence and Affinity Analysis:** A dedicated analysis module automatically parses the output files to generate a dashboard of confidence metrics (pLDDT, PAE) and binding affinity predictions.

*Figure 1: The end-to-end automated workflow of the BoltzNotebook platform, from environment setup to final analysis.*

**2.2 Interactive Input and Configuration**

A key innovation of BoltzNotebook is its graphical user interface (GUI), built with embedded HTML and JavaScript. This interface eliminates the need for users to manually write the YAML configuration files required by Boltz2. The GUI allows users to:

* Define multiple protein chains with their corresponding amino acid sequences.
* Add ligands by specifying either their Chemical Component Dictionary (CCD) identifier or their SMILES string.
* Enable or disable the prediction of binding affinity for a specified ligand.
* Configure advanced run parameters, such as the number of recycling steps, sampling steps, and MSA depth.

This interactive system includes real-time input validation to prevent common errors and guides the user toward creating a valid prediction job.

**2.3 Prediction Backend**

The core structure prediction is performed by the unmodified Boltz2 framework. The notebook dynamically constructs the appropriate command-line arguments based on the user's GUI input. This ensures that the predictions are fully consistent with the original Boltz2 implementation, leveraging its advanced diffusion models and recycling steps to generate accurate 3D coordinates for both the protein and any specified ligands.

**2.4 Automated Analysis and Visualisation Module**

A major contribution of BoltzNotebook is its automated post-prediction analysis module. After a successful run, the notebook automatically processes the raw output files to generate a rich, interpretable report. This report includes:

* **pLDDT Plot:** A plot of the per-residue local distance difference test (pLDDT) score, a confidence metric ranging from 0 to 100 for each amino acid residue. The notebook generates a separate plot for each protein chain.
* **PAE Heatmap:** A heatmap of the Predicted Aligned Error (PAE), which provides confidence estimates for the relative positions and orientations of different protein domains.
* **Affinity Dashboard:** When affinity prediction is enabled, the notebook generates a specialised dashboard summarising the results. This includes the binary probability of binding and a quantitative prediction of the binding affinity, which is converted to more familiar pharmacological metrics like the predicted inhibitory concentration (IC50​) and the Gibbs free energy of binding (ΔG).

**3. Results: Example Use-Case**

To demonstrate the utility of BoltzNotebook, we performed a prediction of the human Cyclin-dependent kinase 2 (CDK2) in complex with the inhibitor Dinaciclib. The FASTA sequence of CDK2 (UniProt ID: P24941) and the SMILES string for Dinaciclib were provided as input through the GUI. The affinity prediction for the ligand was enabled.

The notebook successfully completed the run in under 30 minutes on a standard Colab T4 GPU. The output (Figure 2) included:

* The predicted 3D structure of the CDK2-Dinaciclib complex, with the ligand correctly positioned in the ATP-binding pocket.
* Confidence plots showing a high mean pLDDT score of 92.5, indicating a high-quality model. The PAE plot showed low error within the N- and C-terminal domains.
* The affinity analysis module reported a binding probability of 91.2% and a predicted log10​(IC50​) of -2.5 (equivalent to ~3 nM), which is in strong agreement with experimentally determined values.

*Figure 2: Results for the CDK2-Dinaciclib complex as generated by BoltzNotebook. (A) Interactive 3D view of the predicted structure. (B) Per-residue confidence (pLDDT) plot for the CDK2 chain. (C) Automated binding affinity dashboard.*

**4. Discussion**

BoltzNotebook provides a robust and user-friendly solution for performing advanced protein-ligand structure prediction. Its primary contribution lies in making the powerful Boltz2 engine accessible to a broad audience, including those in structural biology, medicinal chemistry, and education, who may lack the computational expertise to use command-line tools. The integration of an end-to-end workflow with automated analysis and visualisation significantly streamlines the process of generating and interpreting molecular models.

We acknowledge the limitations inherent in a cloud-based platform. The computational resources provided by Google Colab may be insufficient for predicting exceptionally large protein complexes, and run times are subject to platform constraints. Furthermore, the accuracy of the results is fundamentally determined by the underlying Boltz2 model. However, for a vast range of common research targets, BoltzNotebook provides a highly effective and efficient modelling solution.

**5. Conclusion**

BoltzNotebook successfully encapsulates a complex bioinformatics pipeline into a simple, interactive, and powerful tool. By removing technical barriers and integrating publication-quality analysis, it serves as an important resource that empowers researchers to leverage state-of-the-art deep learning models for protein-ligand modelling, thereby accelerating hypothesis generation and research in structural biology and drug discovery.

**Data and Code Availability**

BoltzNotebook is open-source and freely available on GitHub at: <https://github.com/AtharvaTilewale/Boltz-Notebook>.

**References**

1. Baek, M., DiMaio, F., Anishchenko, I., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*.
2. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*.
3. Mirdita, M., Schütze, K., Moriwaki, Y., et al. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*.
4. Passaro, S., Corso, G., Wohlwend, J., et al. (2025). Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*.