

# SlytheRINs: A tool for analyzing and comparing protein Residue Interaction Networks (RINs)

Laura Bradaschia Shimohara<sup>1</sup>, Matheus Assunção<sup>2</sup>, Marília V. A. de Almeida<sup>1</sup>, Ândrea Kelly Ribeiro dos Santos<sup>2</sup> and João Paulo M. S. Lima<sup>1</sup>.

<sup>1</sup>Bioinformatics Multidisciplinary Environment - BioME. Universidade Federal do Rio Grande do Norte (UFRN), Natal, RN, Brazil. <sup>2</sup>Universidade Federal do Pará (UFPA).

## Application Documentation

<b>SlytheRINs: A tool for analyzing and comparing protein Residue Interaction Networks (RINs).....</b>	<b>1</b>
<b>Application Documentation.....</b>	<b>1</b>
1. Introduction.....	1
2. Key Features.....	2
3. Requirements & Installation.....	3
4. How to Run the App.....	3
5. User Interface Overview.....	4
5.1. Sidebar.....	4
5.2. Main Dashboard Area.....	5
6. Analysis Sections (Tabs).....	5
6.1. Part 1: RIN Comparison & Network Complexity.....	5
6.2. Part 2: Chemical Interaction Analysis.....	7
6.3. Part 3: AlphaMissense Integration.....	8
7. Output/Downloads.....	8
8. Troubleshooting/Notes.....	9
9. Citation/Contact (Example).....	9

## 1. Introduction

**SlytheRINs** is an interactive Python web application built with Streamlit (<https://streamlit.io>), designed to analyze and compare protein Residue Interaction Networks (RINs). These graphs represent the amino acid residues as the nodes and their chemical interactions as the edges. They can be easily generated from a single or multiple protein structures, using computational modeling methods, normal-based mode, or molecular dynamics simulation. SlytheRINs provides an analysis dashboard to explore and compare network parameters, identify significant residue-level changes across different conformations, analyze chemical interaction types, and

integrate these findings with predicted mutation effects from AlphaMissense (in the case of human proteins).

Currently, SlytherINs uses [RING](#) ([Del Conte et al. 2024](#)) output files as input files. Due to its ease of use and adequate chemical parameter calculations, we strongly recommend using it to generate your Residue Interaction Networks. We plan to develop a custom RIN generator to integrate into future versions of SlytherINs, but RING is the best in class for this job.

The application was developed by the EvoMol-Lab team at the Bioinformatics Multidisciplinary Environment (BioME) at the Federal University of Rio Grande do Norte (UFRN) in Brazil.

## 2. Key Features

- **Comparative RIN Analysis:** Upload multiple RIN `.edges.txt` files to compare network statistics across different protein conformations or structures.
- **Residue-Level Metrics:**
  - Calculate and visualize variance for degree, betweenness centrality, and clustering coefficient per residue.
  - Mean and Standard Deviation (STD) plots for degree and triangle counts per residue.
  - Degree Standard Deviation (DSD) plot, a SlytherINs-specific metric (inherited from [CoRINs](#)) to highlight residues with stable vs. dynamic connectivity.
- **Statistical Significance Testing:**
  - Paired t-tests to identify residues with degrees significantly different from the network average.
  - Chi-squared tests to assess if the variation in a residue's degree across networks is statistically significant (compared to a Poisson process).
- **Global Network Metrics:**
  - Assortativity coefficient calculation for each network, with an overall t-test for significance.
  - Per-network complexity analysis, including:
    - Network density.
    - Average clustering coefficient.
    - Average shortest path length (for the most significant connected component).
    - Interactive degree distribution plot (with linear/log scaling options).
- **Chemical Interaction Analysis:**
  - Quantification and visualization of different types of chemical interactions

(e.g., HBOND, VDW, IONIC) per residue, if this information is present in the input files.

- **AlphaMissense Integration:**

- Direct download and processing of AlphaMissense substitution prediction data from the [AlphaFold EBI database](#) using a [UniProt](#) ID (for human proteins).
- Option to upload custom AlphaMissense data files (.tsv/.csv).
- Visualization of AlphaMissense classification counts and score distributions per position.

- **Data Export & Reporting:**

- Download options for all generated data tables as .tsv files.
- Download options for all plots as high-resolution .png images.
- Generation of a comprehensive PDF report summarizing the analysis.
- Option to download all generated data and plots in a single .zip archive.

### 3. Requirements & Installation

To run SlytherINs locally, you will need Python 3.x and the following packages:

- streamlit
- pandas
- networkx
- plotly
- scipy
- requests (for AlphaMissense download from AlphaFold DB)
- fpdf2 (for PDF report generation)
- kaleido (for exporting Plotly figures to static images like PNG)

You can install these packages using pip:

```
pip install streamlit pandas networkx plotly scipy requests fpdf2 kaleido
```

DejaVu Font files (open-source) are also included in SlytherINs' repository fonts folder for full PDF report functionality. The script is configured to look for them there. If these fonts are not found, the PDF report will use a default font, and special characters (like "", "", "") might not render correctly, potentially leading to FPDFUnicodeEncodingException errors.

### 4. How to Run the App

For a small number of RIN files (or up to 200 mb of text files), you can use the tool hosted in Streamlit, at the following address:

[slytherins.streamlit.app](https://slytherins.streamlit.app)

If you need to execute several analyses or have large multiple conformations files, we recommend a local execution.

1. Download the repository files or execute the command:

```
git clone https://github.com/jpmslima/SlytheRINs.git
```

2. Enter in SlytheRINs folder and ensure all required packages and font files are installed/placed as described above.
3. Open your terminal or command prompt and install all dependencies:

```
pip install streamlit pandas networkx plotly scipy requests fpdf2 kaleido
```

4. Navigate to the directory where you saved the SlytheRINs Python script (e.g., SlytheRINs.py).
5. Run the command:

```
streamlit run SlytheRINs.py
```

6. The application should open automatically in your default web browser, using localhost.

## 5. User Interface Overview

The SlytheRINs application is organized with a sidebar for controls and a central dashboard area for displaying results.

### 5.1. Sidebar

- **Data Upload Section:**

- **1. RIN Data:**
  - **Upload RIN .edges.txt files:** Allows uploading multiple .edges.txt files generated by [RING](#). These files define the edges (interactions) in your residue networks.
  - **Example RIN files:** A checkbox to load pre-packaged example RIN files for quick testing and demonstration.
- **2. AlphaMissense Data:**
  - **Enter UniProt Accession:** Text input to provide a human protein's UniProt ID (e.g., P03891).
  - **Download & Process from AlphaFold DB:** Button to fetch AlphaMissense substitution data directly from the AlphaFold EBI database for the entered

UniProt ID.

- **Upload AlphaMissense .tsv/.csv file:** This allows uploading a custom file containing AlphaMissense predictions. The app can handle both the official AlphaFold DB CSV format and a simpler 4-column TSV/CSV format (often a result of grep).
- **Example AlphaMissense file:** A checkbox to load a pre-packaged example AlphaMissense data file.
- **Report Generation Section:**
  - **Prepare Data for Reports:** After completing your analyses, click this button to gather all generated metrics, plots, and data tables.
  - **Download PDF Report:** Becomes active after preparing data; downloads a comprehensive PDF summary of the analyses.
  - **Download All Data & Plots (.zip):** Becomes active after preparing data; downloads a ZIP archive containing all generated TSV data files and PNG plot images.

## 5.2. Main Dashboard Area

The main results are organized into three tabs:

- **Tab 1: RIN Comparison & Network Complexity**
- **Tab 2: Chemical Interactions**
- **Tab 3: AlphaMissense Integration**

## 6. Analysis Sections (Tabs)

### 6.1. Part 1: RIN Comparison & Network Complexity

This section compares multiple RINs and analyzes the complexity of individual networks.

- **Summary Metrics:**
  - **Number of Networks Loaded:** Total count of successfully processed RIN files.
  - **Total Unique Residues Identified:** Count all unique residues found across all loaded networks.
- **Variance Plots (Across All Networks):** These plots highlight residues whose network properties change significantly across input RINs.
  - **Degree Variance:** This shows the variance in the number of connections (degrees) for each residue. High variance suggests dynamic connectivity.
  - **Betweenness-centrality Variance:** Shows the variance in how often a residue lies on the shortest paths between other residues. High variance suggests a changing role in network communication.

- **Clustering Coefficient Variance:** Shows the variance in how well-connected a residue's immediate neighbors are. High variance indicates changes in local network density.
- **Mean & STD Plots (Across All Networks):**
  - **Triangle Counts:** Displays each residue's average (mean) involvement in triangles (3 mutually connected residues) and its variability (standard deviation) across conformations.
  - **Degree Values:** Displays each residue's average (mean) degree and variability (standard deviation) across conformations.
- **Degree Distribution per Residue (Box Plot - Across All Networks):**
  - For each residue, this box plot shows the distribution (median, quartiles, outliers) of its degree values across all analyzed networks.
- **Degree Standard Deviation (DSD) Plot:**
  - A SlytherINs-specific metric displaying each residue's Degree Standard Deviation (DSD). It highlights how variable a residue's number of interactions is. Low DSD indicates stable connections; high DSD suggests dynamic behavior. Shaded regions indicate thresholds for low DSD ( $\pm 0.5/\pm 1.0$ ).
- **Assortativity & Significance (Per Network):**
  - Plots the degree assortativity coefficient for each network (Network Index on x-axis).
  - Provides an overall t-test result comparing the mean assortativity across all networks to zero.
- **Significance Plots (Across All Networks):**
  - **Degree Significance (Paired T-test vs. Others):** For each residue, this tests if its degree (across networks) significantly differs from the average degree of all *other* residues in those same networks. Low p-values (e.g., below the red line at 0.05) suggest statistically significant differences.
  - **Degree Variation Significance (vs. Poisson):** For each residue, this tests if the variation of its degree values (across networks) is significantly different from what would be expected by a random (Poisson) process. Low p-values suggest the variation is not merely random.
- **Network Complexity Analysis (Per Selected Network):**
  - Allows selection of a single network from the uploaded set via a dropdown menu.
  - **Metrics Displayed:**
    - Number of Nodes
    - Number of Edges
    - Network Density
    - Average Clustering Coefficient

- Average Shortest Path Length (calculated for the most significant connected component if the graph is disconnected).
- **Degree Distribution Plot:** A bar chart showing the distribution of node degrees (how many nodes have k connections). Checkboxes allow toggling log scales for the X and Y axes to help identify potential power-law distributions characteristic of scale-free networks.

## 6.2. Part 2: Chemical Interaction Analysis

This section is active if the uploaded .edges.txt files contain an "Interaction" column (as typically generated by [RING](#)).

- It quantifies and visualizes the mean count (and standard deviation) of different types of chemical interactions (e.g., HBOND, VDW, IONIC, PIPISTACK, etc.) for each residue across all networks.
- Each interaction type gets its plot, displayed in a panel layout (typically two plots per row).
- The interactions calculated by RING are:

Symbol	Interaction
HBOND	Hydrogen bond
SSBOND	Disulfide bridge (covalent bond)
IONIC	Ionic bond
VDW	van der Waals interaction
PICATION	$\pi$ -cation
PIPISTACK	$\pi$ - $\pi$ stacking
PIHBOND	$\pi$ -hydrogen
METAL_ION	Metal ion coordination
HALOGEN	Halogen bond
IAC	Inter-Atomic Contact

**SlytherINs uses RING's attributes; it does not modify them.** If you want to alter the interaction parameters, please execute RING with these modifications, using their [webpage](#) or standalone package [HERE](#).

### 6.3. Part 3: AlphaMissense Integration

This section allows integration and visualization of AlphaMissense pathogenicity predictions.

- **Data Input:**
  - **Direct Download:** Enter a UniProt ID for a human protein and click "Download & Process from AlphaFold DB". The app fetches the aa-substitutions.csv file.
  - **File Upload:** Upload a .tsv or .csv file. The app attempts to parse:
    1. The official AlphaFold DB CSV format (columns: protein\_variant, am\_pathogenicity, am\_class).
    2. A 4-column format (ProteinID, Mutation, AM\_Score, AM\_Classification), often from grep output.
  - **Example File:** Use a pre-packaged example.
- **Data Processing:**
  - Extracts residue position from the mutation string (e.g., A123G -> 123).
  - Converts AlphaFold DB classification terms (LPath, LBen, Amb) to pathogenic, benign, and ambiguous for consistency with the internal color map.
- **Plots:**
  - **AlphaMissense Classification Counts:** A stacked bar chart showing the count of 'pathogenic', 'benign', and 'ambiguous' classifications for mutations at each residue position.
  - **AlphaMissense Score Distribution:** A box plot showing the distribution of AlphaMissense pathogenicity scores (0 to 1) for all possible amino acid changes at each residue position. Shaded regions indicate thresholds for likely benign, ambiguous, and likely pathogenic.

## 7. Output/Downloads

For most generated plots and data tables, SlytherINs provides download options:

- **Data (TSV):** Download the underlying numerical data for a plot or analysis as a Tab-Separated Values file.
- **Plot (PNG):** Download the currently displayed plot as a high-resolution PNG image. (This requires the Kaleido package.)

### Report Generation (from Sidebar):

1. **Prepare Data for Reports:** Click this button after running your desired analyses. It collects all generated metrics, plot figures, and data table references into st.session\_state.
2. **Download PDF Report:** Once data is prepared, this button becomes active. It generates a multi-page PDF document containing:



- A title page and generation timestamp.
  - Chapter titles for each analysis part.
  - Key summary metrics.
  - Plot captions/descriptions followed by the plot images.
  - (Note: TSV data tables are *not* directly embedded in the PDF to keep it concise, but are available in the ZIP download).
3. **Download All Data & Plots (.zip):** After preparing data, this button generates a ZIP archive containing:
- A Report\_Summary.txt file with key metrics and a list of included files.
  - A data/ subfolder with all generated TSV data files.
  - A plots/ subfolder with all generated PNG plot images.

## 8. Troubleshooting/Notes

- **PNG Export Failure:** If PNG downloads fail, ensure the kaleido package is correctly installed in your Python environment (pip install kaleido).
- **PDF Generation Issues / Unicode Errors:**
  - Ensure fpdf2 is installed (pip install fpdf2).
  - For correct rendering of special characters (e.g., "", "", "") in the PDF, download the **DejaVu Sans font family** (DejaVuSans.ttf, DejaVuSans-Bold.ttf, DejaVuSans-Oblique.ttf, DejaVuSans-BoldOblique.ttf). Place these .ttf files in a subdirectory named fonts within the same directory as your SlytheRINs script. If these are not found, the app will issue a warning and fall back to a default font, which may not support all characters.
- **AlphaFold DB Download:** The direct download feature for AlphaMissense data using a UniProt ID is specific to the AlphaFold EBI database and primarily works for human proteins for which this data is available.
- **File Formats:**
  - **RIN Data:** Expects .edges.txt files, typically tab-separated, with at least 'NodeId1' and 'NodeId2' columns. An Interaction column is also required for chemical interaction analysis (Part 2). Node IDs are expected in the format Chain:Position:ResID:AtomName (e.g., A:123:ALA:CA).
  - **AlphaMissense Upload:** Can handle the official CSV from AlphaFold DB or a 4-column (ProteinID, Mutation, Score, Classification) TSV/CSV, often without headers if it's a grep output.
- **Performance:** Analyzing a large number of RIN files or large individual networks can be computationally intensive and may take time. Generating PDF and ZIP reports with many plots can also be resource-intensive.

## 9. Citation/Contact (Example)

If you use SlytherINs in your research, please consider citing it.

Please visit the [EvoMol-Lab GitHub](#) page or contact the developers for issues, suggestions, or contributions.

## Disclaimer

The developer team used generative AI tools for the following tasks:

- Code revision and optimization.  
Writing code for the following tasks:
  - Generate PDF reports.
  - Generate download links for `.tsv` and `.png` files.
  - Generate report summaries and `.zip` file download for all the analysis plots and files.
- Elaborate documentation topic structure.
- Review the english language.