

LLM Agent Based Protein Function Prediction

Supplementary Material

July 31, 2025

1 Protein-centric agent creation and reasoning process

1.1 Metadata retrieval

Before creating an agent, we retrieve information from UniProtKB and PubMed. For example: for the protein with ID 11H.STRNX we retrieve the following information:

Uniprot metadata: Protein: Strychnine-11-hydroxylase — Gene: 11H — Organism: *Strychnos nux-vomica* — Function: Monooxygenase involved in the biosynthesis of curare monoterpene indole alkaloids (MIAs), natural products such as strychnine, a neurotoxic compound used as a pesticide to control rodents, and its pharmacologically active derivatives, including brucine, used to regulate blood pressure (PubMed:35794473). Curare alkaloids act as animal glycine receptor antagonists (PubMed:35794473). Catalyzes the conversion of beta-colubrine to 11-deMe brucine (PubMed:35794473) — Catalytic Activity: beta-colubrine + reduced [NADPH-hemoprotein reductase] + O₂ = 11-demethylbrucine + oxidized [NADPH-hemoprotein reductase] + H₂O + H(+) — Pathways: Alkaloid biosynthesis — Subcellular Location: Membrane — Keywords: Heme; Iron; Membrane; Metal-binding; Monooxygenase; Oxidoreductase; Transmembrane; Transmembrane helix — Additional Info: SIMILARITY: Belongs to the cytochrome P450 family

PubMed abstracts:

- ‘Strychnine is a natural product that, through isolation, structural elucidation and synthetic efforts, shaped the field of organic chemistry. Currently, strychnine is used as a pesticide to control rodents because of its potent neurotoxicity. The polycyclic architecture of strychnine has inspired chemists to develop new synthetic transformations and strategies to access this molecular scaffold, yet it is still unknown how plants create this complex structure. Here we report the biosynthetic pathway of strychnine, along with the related molecules brucine and diaboline. Moreover, we successfully recapitulate strychnine, brucine and diaboline biosynthesis in *Nicotiana benthamiana* from an upstream intermediate, thus demonstrating that this complex, pharmacologically active class of compounds can now be harnessed through metabolic engineering approaches.’

1.2 Agent creation

Using the retrieved information from UniProtKB ([**uniprot_info**]) and PubMed ([**abstracts**]), we create and agent and its role and context:



User:

You are a GO annotation curator that refines GO term predictions for UniProtKB protein entry 11H_STRNX Here is the general information about this proteins functions: [uniprot_info] Here are the article abstracts that are related to this protein: [abstracts].

You operate by revising external information of a protein sequence such as the interpro annotations or diamond score similarity. You operate in this way: you are given a term and you need to check (1) if the term is in the interpro annotations or if the definition is related to the definition of interpro annotations, (2) the diamond score for the term. You will be asked to increase or decrease the score of the term based on the information you have access to.

1.3 Reasoning process

To start the reasoning process, we select initial GO term predictions with scores ≥ 0.1 , and retrieve the following information:

- Initial MLP prediction score
- Diamond Score
- Definition and labels
- Taxonomical constraints

We denote this information as [go_terms_info]. Then, we instruct the agent to analyze the GO terms and suggest refinements of predictions:



User:

You have now this information about the GO terms you discussed before [go_terms_info].

For each relevant GO term suggested:

- Analyze annotation frequency: terms with low frequency should might be underrepresented and might be plausible. Consider a term underrepresented if its frequency is below 200
- Analyze supporting evidence (InterPro / Diamond / Abstracts) for each plausible term.
- If there is conflicting evidence, provide your resolution
- Provide Current score vs. recommended score. We want to minimize the amount of changes, so only update by incrementing or decrementing the score by 0.2 maximum.
- Confidence level (high/medium/low)

Output your report with all the points above in a structured format. Annotation frequency is based on the training data and is an important factor in your analysis.



Protein Agent:

Okay, I will analyze the provided GO terms and suggest score adjustments based on the available information.

- GO:0043227 - membrane-bounded organelle
 - Analysis: This term is a general cellular component. The protein is located in the membrane, so this term is plausible.
 - Supporting Evidence:
 - InterPro: Not found in the provided information.
 - Diamond: 0.442
 - Abstracts: The protein is located in the membrane.
 - Resolution: The diamond score is relatively high, and the protein is membrane-bound.
 - Recommendation: Increase score by 0.2.
 - Current Score: 0.340
 - Recommended Score: 0.540
 - Confidence: High
- GO:0005886 - plasma membrane
 - Analysis: The definition excludes the plasma membrane.
 - Annotation Frequency: N/A
 - Supporting Evidence:
 - InterPro: Not found in the provided information.
 - Diamond: 0.278
 - Abstracts: N/A
 - Resolution: The definition of the term conflicts with the protein function.
 - Recommendation: Decrease score by 0.2.
 - Current Score: 0.227
 - Recommended Score: 0.027
 - Confidence: High
- GO:0005783 - endoplasmic reticulum
 - Analysis: This term is a specific type of membrane-bound organelle. It could be plausible if the protein is specifically located in the ER.
 - Annotation Frequency: N/A
 - Supporting Evidence:
 - InterPro: Not found in the provided information.
 - Diamond: 0.236
 - Abstracts: N/A
 - Resolution: There is no specific evidence to support this term.
 - Recommendation: No change.
 - Current Score: 0.206
 - Recommended Score: 0.206
 - Confidence: Medium



User:

Apply your analysis to update GO term scores. Perform the update and also provide a rationale for each change. If no changes are needed, return 'No changes needed'.



Protein Agent:

Changes applied.

1.4 Runtime analysis

To assess the time in seconds needed to update the predictions, we averaged the inference time for both LLM agents (Gemini-Flash-2.0 and GPT-4.1 nano) over the first 10 proteins from the test set. We found that Gemini-Flash-2.0 is faster than GPT-4.1 nano across all three subontologies.

Time in seconds needed to update the predictions.

Method	MFO	BPO	CCO
Gemini-Flash-2.0	18.98	57.05	31.12
GPT-4.1 nano	31.14	70.91	44.97