De Novo Binder Design Against RBX1 Using an Autonomous AI Agent

Lucia Urcelay Ganzabal, ML Research Scientist in the Centre for Genomic Regulation (CRG), Ferruz Lab

Background

I had been aming to find some time to work on a submission for this competition but I found myself with nothing 15 hours before the deadline. Because of this, I considered the lack of time as an opportunity to evaluate the capability of autonomous AI agents to independently execute a computational protein design campaign from literature review through to a validated submission FASTA, without manual intervention in the design decisions. I wanted to put to test these kinds of model, and see how far they can get with minimum interaction from the user. I chose the agent Biomni, developed by Phylo, as I had seen tweets talking about it in the last few weeks, that being the only reason. I decided i would only use a single prompt for the whole process. Prompt:

> *I am going to participate in a binder design competition for the target RBX1. These are links with information about the competition and rules: https://proteinbase.com/competitions/gem-adaptyv-rbx1 https://www.gembio.ai/gemadaptyv2026*
>
> *Perform a literature review on this target and make a plan for designing top performing binders for this target using the tools you have available. The output should be a submission ready fasta file that complies with competition rules.*

Approach

The agent began by retrieving and interpreting the primary literature on RBX1 structure and function, identifying the NMR structure 2LGV as the most suitable design template (the crystal structure 4F52 contains a bound E2 enzyme that would occlude the target surface). The agent extracted MODEL 1 from 2LGV (residues 9–108, retaining three structural zinc ions) and mapped hotspot residues onto the structure based on published mutagenesis data.

Backbone generation was performed using RFdiffusion (Watson et al., 2023) in binder-design mode, targeting the RBX1 RING domain (chain A, residues 9–108) with binder lengths of 60–100 amino acids. Three campaigns were run in parallel, each specifying a distinct set of hotspot residues to focus diffusion on different regions of the functional interface:

- Campaign 1: Secondary E2 surface: residues A60, A63, A66, A81, A82, A83 (200 backbones)
- Campaign 2: Cullin/E2 β-sheet edge: residues A40, A42, A44, A46, A48, A50 (200 backbones)
- Campaign 3: Core E2/Glomulin surface: residues A44, A55, A57, A87, A91 (200 backbones)

Sequence design on the 600 generated backbones was performed using ProteinMPNN (Dauparas et al., 2022) in soluble-model mode, designing only the binder chain (chain A) while holding the RBX1 target (chain B) fixed. Eight sequences were sampled per backbone at two temperatures (T = 0.1 and T = 0.2), yielding 9,600 candidate sequences in total.

Sequences were filtered and ranked through an automated pipeline also managed by the agent. Novelty filtering removed sequences with greater than 60% single-amino-acid composition (a threshold calibrated during the run after the agent identified that the default 40% threshold incorrectly flagged legitimate E/K-rich helical binders), any sequence with a 30-residue window identity above 75% to RBX1, and any sequence with a normalised edit distance below 25% relative to RBX1. Diversity selection used greedy clustering at a 50% pairwise identity threshold, with sequences ranked by ProteinMPNN score (lower = higher predicted sequence-structure compatibility). The top 100 diverse sequences were selected for submission.

In Silico Results

Of the 9,600 sequences generated, 9,557 (99.6%) passed all novelty and validity filters. The 43 rejected sequences were genuinely degenerate (>60% single amino acid, predominantly poly-alanine or poly-arginine). Greedy diversity clustering at 50% identity yielded 2,627 distinct sequence clusters, from which the top 100 by ProteinMPNN score were selected.

The 100 submitted sequences have the following characteristics:

- ProteinMPNN score range: 0.9335 – 1.0750 (median 1.038; scores below ~1.5 are generally considered high-quality)
- Length range: 63 – 100 amino acids (mean 86.9 AA)
- Minimum normalised edit distance to RBX1: 0.787 (competition minimum: 0.25)
- Maximum 30-residue window identity to RBX1: 0.300 (competition maximum: 0.75)
- Campaign representation: Campaign 3 core E2/GLMN surface (37 sequences), Campaign 1 secondary E2 surface (42 sequences), Campaign 2 Cullin/E2 edge (21 sequences)

The designed sequences are predominantly helical in character, consistent with the alpha-helical backbones generated by RFdiffusion for this target geometry. The top-ranked sequence (score 0.9335, 75 AA) targets the core E2/Glomulin hotspots and is representative of the coiled-coil-like amphipathic helices that dominate the submission. All 100 sequences passed competition rule validation with wide margins on all criteria.

Notes on Autonomous Agent Performance

Several challenges arose during execution that required the agent to diagnose and recover autonomously. These included HPC command format errors (incorrect Hydra configuration keys for RFdiffusion, non-existent ProteinMPNN flags), GPU concurrency limits requiring dynamic job scheduling, S3 download failures for large trajectory files requiring direct API queries to selectively retrieve only final design PDBs, and a mid-session HPC token expiry that prevented retrieval of a fourth campaign's results. In each case the agent seemed to identify the root cause from error logs and adapted its approach without human intervention.

One scientifically meaningful correction was made autonomously: the agent identified that its initial low-complexity filter (40% single-amino-acid threshold) was incorrectly discarding ~2,200 legitimate E/K-rich helical binders, a known characteristic of ProteinMPNN designs for helical targets, and recalibrated the threshold to 60% before finalising the submission.

On a personal note, I'm making the submission right before the deadline hour so I have not been able to verify the output of the model carefully, if I get lucky enough to have some of the designs tested we will be able to actually assess the capabilities of this tool. I also want to mention that as impressive as it seems, I felt totally detached from the project and from the science of it. I can't say that I have learned about the target and that I have been excited to launch experiments and wait nervously for the results. I suppose this may be one of the trade-offs of using such models. That said, I think that if applied thoughtfully, they can be valuable tools. We'll see what comes of this!

References

1. Duda DM et al. (2012). Structure of a Glomulin-RBX1-CUL1 complex: inhibition of a RING E3 ligase through masking of its E2-binding surface. Mol Cell 47(3):371–382.
2. Watson JL et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature 620:1089–1100.
3. Dauparas J et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. Science 378(6615):49–56.
4. Li Y et al. (2023). RBX1 overexpression in cancer and its therapeutic implications. (Representative citation for RBX1 oncology rationale.)