

HotBind-RBX1: Hotspot-Guided De Novo Peptide Binder Design against RBX-1 Using BindCraft and Multi-Stage Docking Validation

Rik Ganguly^{1,*}, Siddhant Poudyal^{2,*}, Gautam Ahuja^{1,3,*}

*Equal Contribution. Email: (rik.ganguly; gautam.ahuja; siddhant.poudyal)@ashoka.edu.in

¹Mphasis AI & Applied Teach Lab at Ashoka, Ashoka University, Haryana, India; ²Ashoka University; ³Department of Computer Science, Ashoka University, Haryana, India.

Author Designation: Rik Ganguly: Junior Engineer; Siddhant Poudyal: Project Research Scientist-I; Gautam Ahuja: Multimodal Biomedical Data Scientist & ML Engineer

1 Target Characterisation and Hotspot Identification

1.1 Binding pocket identification via CASTpFold

The three-dimensional structure of RBX-1 was subjected to surface cavity analysis using CASTpFold (Computed Atlas of Surface Topography of Proteins), which employs an alpha-shape algorithm to identify and characterise solvent-accessible pockets based on geometric and volumetric criteria. A well-defined surface cavity (Pocket ID 1) was identified, presenting a solvent-accessible area of 24.823 Å² and a volume of 9.626 Å³, and comprising 26 residues: ALA31, GLN32, TRP35, ASP36, ASP40, ALA43, ILE44, ARG46, ASN47, MET50, CYS56, GLN57, ALA58, ASN59, THR64, GLU67, VAL70, ALA71, TRP72, CYS75, HIS80, PHE81, TRP87, LEU88, LYS105, and TYR106.

Structural mapping and multiple sequence alignment (MSA) revealed a conserved aromatic cluster TRP35, TRP72, TRP87, and PHE81 lining the pocket interior, indicative of a hydrophobic stacking environment suitable for substrate accommodation. The acidic polar centre formed by ASP36 and ASP40 was identified as a conserved motif consistent with electrostatic stabilisation or catalytic acid-base chemistry. The physicochemical composition of the pocket is diverse, spanning aromatic (TRP, PHE), acidic (ASP), polar (GLN, ASN, THR), basic (ARG, HIS, LYS), hydrophobic (ALA, ILE, VAL, MET, LEU), and sulphur-containing (CYS, MET) residues.

1.2 Evolutionary conservation analysis

Multiple sequence alignment of RBX-1 orthologs across representative eukaryotic species was performed to assess residue-level conservation within the identified pocket. Positions TRP35, ASP36, ASP40, ARG46, ASN47, TRP72, HIS80, PHE81, TRP87, and CYS75 exhibited high conservation scores, indicating strong evolutionary constraint and functional indispensability. In contrast, peripheral cavity residues including ALA31, VAL70, MET50, LEU88, LYS105, and TYR106 showed variable conservation, consistent with structural scaffolding roles rather than direct functional activity.

1.3 Energetic frustration analysis via Frustratometer

To delineate functionally significant regions within the pocket from a biophysical energetics perspective, local configurational and mutational frustration were mapped using the Frustratometer server. The Frustratometer quantifies the degree to which local residue interactions deviate from their lowest-energy configuration by comparing observed interaction energies against a statistical ensemble of decoy configurations, yielding a frustration index for each residue pair.

Pairwise configurational frustration mapping revealed clusters of highly frustrated interactions concentrated in residue spans 35-50 and 70-90, spatially coinciding with the conserved aromatic-acidic core identified by CASTpFold. Local frustration density analysis within a 5 Å sphere confirmed that residues TRP35, ASP36, ASP40, ARG46, ASN47, TRP72, HIS80, and TRP87 collectively exhibit elevated frustration density, a hallmark of catalytic or binding-competent sites that require conformational flexibility for functional activity. Minimally frustrated contacts were found to dominate the peripheral structural regions, consistent with their role in maintaining global fold stability. Mutational frustration analysis further corroborated that the aromatic-acidic core exhibits a high density of highly frustrated mutational contacts, indicating strong selective pressure against substitution at these positions.

Collectively, the integration of geometric pocket identification, evolutionary conservation analysis, and dual-mode energetic frustration profiling delineated a functionally significant active-site cluster centred on the aromatic-acidic core (TRP35-ASP36-ASP40-ARG46-ASN47-TRP72-HIS80-TRP87). These eight residues were designated as primary hotspot residues for all downstream binder design and docking stages.

2. Peptide Binder Design, Validation, and Refinement

2.1 De novo peptide binder design via BindCraft

De novo peptide binders targeting the eight designated hotspot residues (TRP35, ASP36, ASP40, ARG46, ASN47, TRP72, HIS80, TRP87) were designed using BindCraft, a deep learning-based framework for the computational design of high-affinity binders against user-defined interface hotspots.

Two independent design runs were performed with differing parametric configurations. In Run 0, chain A of the RBX-1 structure was used as the target, with hotspot residues specified at positions A35, A36, A40, A46, A47, A72, A80, and A87. Peptide lengths were sampled in the range of 20-40 residues using the Peptide design protocol, HardTarget prediction protocol, AlphaFold2 interface protocol, and Default template protocol, with the Peptide filter option applied. This run yielded 25 candidate peptide sequences, which were subsequently folded into three-dimensional structures using ESMFold for downstream docking. Run 1 retained the same chain and hotspot specification but introduced a shorter length range of 6-24 residues, substituted the template protocol with Masked, and adopted the `Peptide_Relaxed` filter option which is a more permissive sampling strategy aimed at capturing compact binder geometries. This run produced 3 pre-folded PDB complexes directly.

All designed structures, the 25 ESMFold-generated models from Run 0 and the 3 complexes from Run 1 were subjected to protein-peptide docking using two complementary pipelines: HADDOCK 2.4 via the web server interface, and HADDOCK 3 executed locally via the command-line terminal. This dual-pipeline approach enabled cross-validation of binding modes and scoring across software versions, providing greater confidence in the ranked outputs.

2.2 Binding affinity estimation via PRODIGY and candidate prioritisation

Final HADDOCK-refined complexes were submitted to PRODIGY for binding affinity prediction. PRODIGY employs a linear regression model trained on experimentally determined protein-protein binding affinities, using interfacial contact counts decomposed by physicochemical class: charged-charged, charged-polar, charged-apolar, polar-polar, polar-apolar, and apolar-apolar, as predictive features. Candidates with predicted K_d in the nanomolar range (1-100 nM) were prioritised as high-confidence leads.

Final candidate selection integrated evidence from all pipeline stages. A candidate was classified as a high-priority lead if it satisfied all of the following criteria: (i) passed all BindCraft filters ($i_{pTM} > 0.70$, $i_{pAE} < 0.30$, $dG/dSASA < -2.5$, binder RMSD < 1.5 Å, zero relaxed clashes); (ii) demonstrated a convergent binding pose between BindCraft; (iii) ranked within the top HADDOCK cluster with a negative total score and low intra-cluster RMSD; and (iv) exhibited a PRODIGY-predicted K_d below 1 μ M.

2.3 Novelty validation against UniRef50

To confirm sequence novelty, all 28 binder sequences were searched against UniRef50 (~60M sequences) using two independent tools. MMseqs2 (v18.8cc5c, local alignment mode, --min-seq-id 0.75, E-value < 0.001) returned zero hits across all sequences. This was cross-validated using DIAMOND (sensitive mode, BLOSUM62, --min-score 10), which returned three weak local alignments (mpnn1, mpnn2, mpnn4; 44–50% identity) - all well below the 75% competition threshold and with E-values of 6×10^2 – 1×10^3 , orders of magnitude above the significance cutoff of 0.001, indicating noise-level alignments rather than true homology. All 28 sequences therefore satisfy the de novo novelty criterion.

3. Future Scope:

Affinity maturation would have provided an opportunity to systematically modulate binding strength, enabling the design of moderately affine binders that may be more likely to elicit cytotoxic effects. However, the primary objective of this study was the de novo generation of peptide binders while preserving the core structural framework with minimal modifications. Within this constraint, there remains scope to further explore a range of optimized binding variants that balance affinity and functional activity.

4. Material:

Drive Link: https://drive.google.com/drive/folders/12ExnXcK0EGHACz4FtH7FzW_7Z5z4vbwQ