Supplementary Information (SI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2026

# Supporting Information

# Design of permeability-optimized target-binding macrocycles via direct preference optimization

Heqi Sun[1], Hong Tan[1], Yanyi Chu[2], Jiayi Li[1], Ruixuan Wang[1], Dong-Qing Wei[1,3,4,*]

[1] State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200040, P. R. China.
[2] Key Laboratory of RNA Innovation, Science and Engineering, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China.
[3] Hebi Branch, Henan Academy of Sciences, Qishui Guang East, Qibin District, Hebi Henan, 458030, P. R. China.
[4] Qihe Laboratory, Qishui Guang East, Qibin District, Hebi, Henan, 458030, P. R. China.

**Corresponding author:**
[*]Dong-Qing Wei, email: dqwei@sjtu.edu.cn

# S1 Supplementary figures

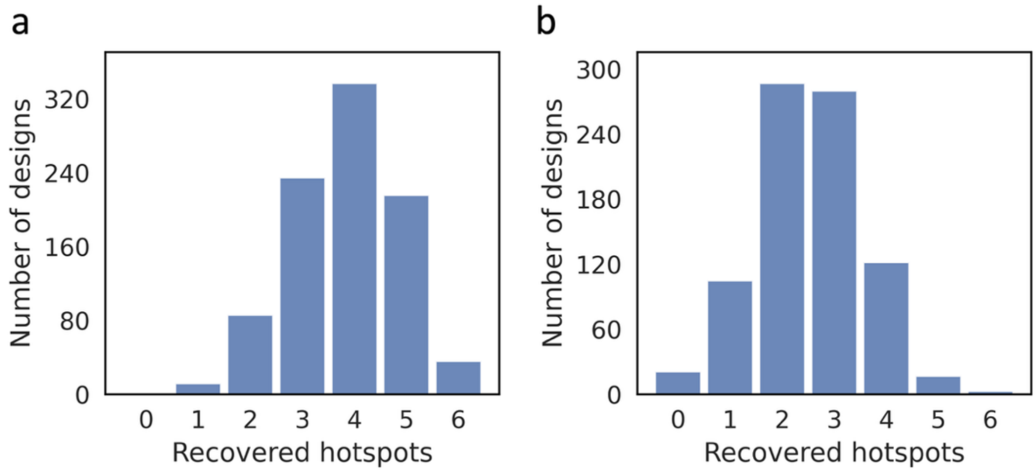


**Fig. S1.** Hot-spot engagement of CycDiff-DPO designs for the two PPI targets in the case studies. Histograms show the number of designed cyclic peptides versus the number of hotspot residues recovered through peptide–target contacts per design for (a) Keap1–Nrf2 and (b) SPSB2–iNOS.

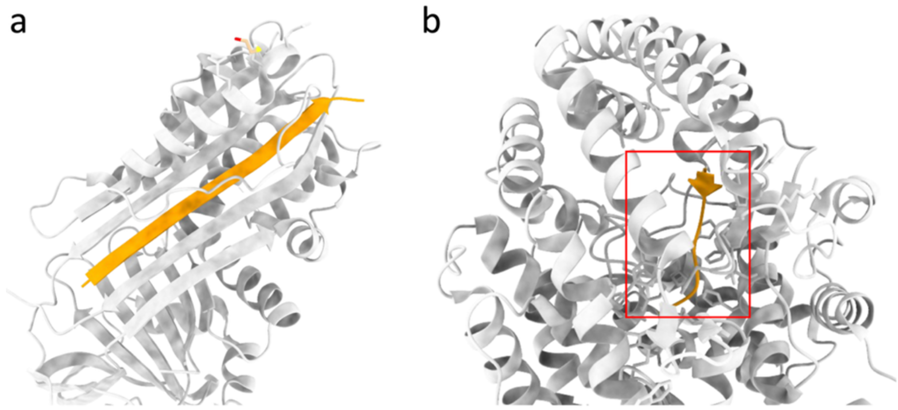


**Fig. S2.** Representative LNR complexes incompatible with cyclic peptide scaffolds. (a) PDB 1JRR: the ligand adopts an extended β-sheet conformation that is unlikely to be accommodated by a cyclic peptide of similar length. (b) PDB 4APH: the ligand is buried within the protein core, where cyclization would likely cause steric clashes and poor shape complementarity.

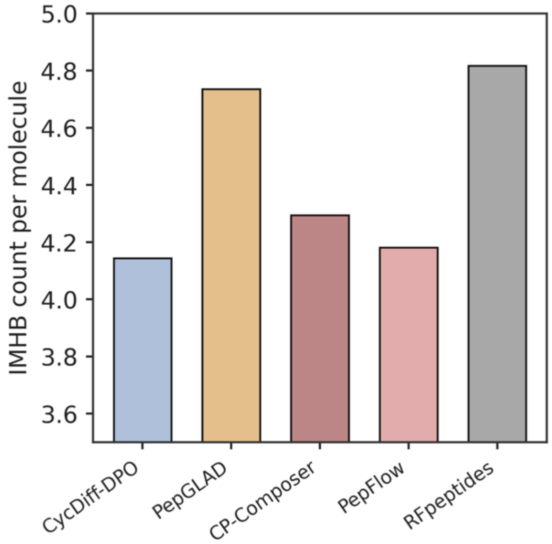


**Fig. S3.** Average intramolecular hydrogen-bond (IMHB) count per generated cyclic peptide across five generative models.

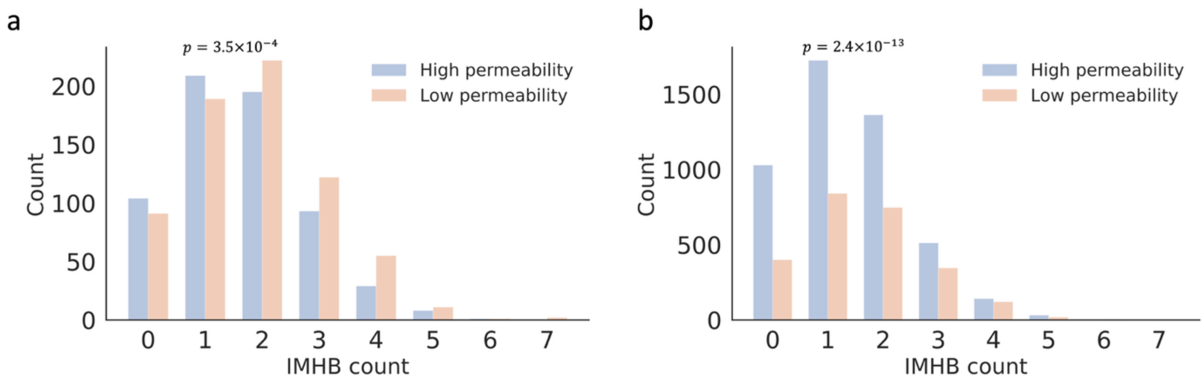


**Fig. S4.** Distribution of intramolecular hydrogen-bond (IMHB) counts for cyclic peptides in the CycPeptMPDB database, stratified by experimental permeability. Peptides with $\log P_{exp} >$ -6 cm/s are classified as high permeability; all others are classified as low permeability. P values were computed using a two-sided Mann–Whitney U test, with $P < 0.05$ considered statistically significant. (a) Compounds with Caco-2 permeability measurements. (b) Compounds with PAMPA permeability measurements.

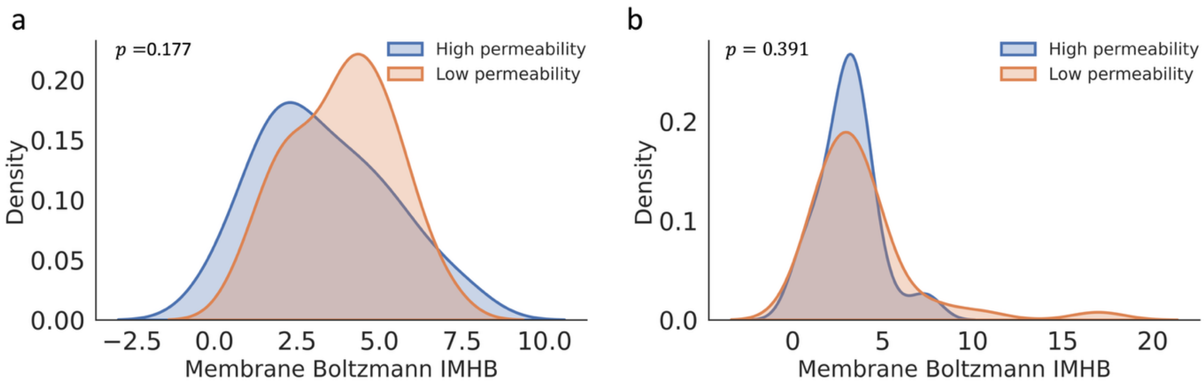


**Fig. S5.** Boltzmann-weighted IMHB counts for cyclic peptides in a membrane-mimicking environment (ALPB chloroform). Peptides with $\log P_{exp} > -6$ cm/s are classified as high permeability; all others are classified as low permeability. P values were computed using a two-sided Mann–Whitney U test, with $p < 0.05$ considered statistically significant. (a) Compounds with Caco-2 permeability measurements. (b) Compounds with PAMPA permeability measurements.

## S2 Supplementary tables

**Table S1.** Per-fold performance of the in-house Caco-2 permeability predictor under five-fold cross-validation.

| **Fold** | $R^2$ | **MSE** | **Spearman's ρ** |
|---|---|---|---|
| 1 | 0.78 | 0.14 | 0.87 |
| 2 | 0.79 | 0.12 | 0.89 |
| 3 | 0.72 | 0.20 | 0.83 |
| 4 | 0.72 | 0.17 | 0.85 |
| 5 | 0.72 | 0.17 | 0.84 |

**Table S2.** Caco-2 permeability predictions for cyclic peptides from five generative models using two predictors.

| **Predictor** | **CycDiff-DPO** | **CP-Composer** | **RFpeptides** | **PepFlow** | **PepGLAD** |
|---|---|---|---|---|---|
| CPMP | -6.10±0.53 | -6.32±0.53 | -6.60±0.36 | -6.50±0.28 | -6.55±0.52 |
| PharmPapp | -5.22±0.03 | -5.35±0.05 | -5.32±0.05 | -5.49±0.05 | -5.37±0.04 |

Values represent mean ± SD of predicted $\log P_{exp}$ (cm/s).

**Table S3.** PAMPA permeability predictions for cyclic peptides from five generative models using three predictors.

| **Predictor** | **CycDiff-DPO** | **CP-Composer** | **RFpeptides** | **PepFlow** | **PepGLAD** |
|---|---|---|---|---|---|
| DMPNN | -7.85±0.98 | -8.66±1.12 | -9.08±1.10 | -8.55±0.77 | -8.66±0.88 |
| attentiveFP | -8.94±1.69 | -9.14±1.46 | -9.48±1.43 | -9.08±1.20 | -9.12±1.34 |
| CPMP | -6.89±0.68 | -7.13±0.78 | -7.88±0.54 | -7.87±0.58 | -7.04±0.88 |

Values represent mean ± SD of predicted $\log P_{exp}$ (cm/s).

**Table S4.** Correlation between membrane-mimicking IMHB and experimental permeability

| **Dataset** | **Spearman r** | **Spearman p** | **High-perm group mean IMHB** | **Low-perm group mean IMHB** | **Mann-Whitney p** |
|---|---|---|---|---|---|
| Caco-2 (n = 70) | -0.24 | 0.04 (*) | 3.32 | 3.85 | 0.18 (n.s.) |
| PAMPA (n = 70) | -0.16 | 0.19 (n.s.) | 3.09 | 3.79 | 0.39 (n.s.) |

* p < 0.05, ** p < 0.01, *** p < 0.001, n.s. = not significant.

**Table S5.** Permeability prediction of top-ranked designs and reference peptides for the Keap1–Nrf2 and SPSB2–iNOS case studies.

| Case study | Peptide | CPMP (Caco-2) | PharmPapp (Caco-2) | DMPNN (PAMPA) | AttentiveFP (PAMPA) | CPMP (PAMPA) |
|---|---|---|---|---|---|---|
| Keap1-Nrf2 | Design | -5.80 | -5.08 | -5.93 | -6.70 | -6.73 |
| | Ref. | -6.61 | -5.10 | -11.00 | -7.95 | -8.25 |
| SPSB2-iNOS | Design | -5.70 | -5.03 | -6.82 | -8.92 | -7.17 |
| | Ref. | -6.24 | -5.06 | -8.10 | -12.54 | -8.01 |

**Table S6.** Physicochemical properties of top-ranked designs and reference peptides for the Keap1–Nrf2 and SPSB2–iNOS case studies.

| Case study | Peptide | Sequence | MW | HBD | TPSA | logP |
|---|---|---|---|---|---|---|
| Keap1-Nrf2 | Design | c[LLLVVVK] | 765.05 | 8 | 229.72 | 1.63 |
| | Ref. | c[DhA-GDPETGE] | 628.59 | 9 | 297.94 | -4.37 |
| SPSB2-iNOS | Design | c[KVLDIHLL] | 932.18 | 11 | 324.80 | 0.29 |
| | Ref. | c[RGDINNNV] | 882.93 | 15 | 461.27 | -7.42 |

**Table S7.** Predicted Caco-2 permeability under different training strategies and DPO inverse temperature (β)

| DPO setting | CPMP | PharmPapp |
|---|---|---|
| SFT | -6.25±0.53 | -5.18±0.05 |
| Without DPO | -6.32±0.53 | -5.15±0.05 |
| $\beta = 0.1$ | -6.22±0.52 | -5.16±0.05 |
| $\beta = 0.5$ | -6.10±0.53 | -5.22±0.03 |
| $\beta = 5$ | -6.21±0.52 | -5.17±0.03 |
| $\beta = 50$ | -6.20±0.53 | -5.22±0.04 |

Values represent mean ± SD of predicted $\log P_{exp}$ (cm/s).

**Table S8.** Predicted PAMPA permeability under different training strategies and DPO inverse temperature (β).

| DPO setting | DMPNN | attentiveFP | CPMP |
|---|---|---|---|
| SFT | -8.50±1.10 | -9.00±1.45 | -7.00±0.77 |
| Without DPO | -8.66±1.11 | -9.14±1.46 | -7.13±0.78 |
| $\beta = 0.1$ | -8.25±1.05 | -8.90±1.45 | -6.95±0.72 |
| $\beta = 0.5$ | -7.85±0.98 | -8.94±1.69 | -6.89±0.68 |
| $\beta = 5$ | -7.95±1.01 | -9.00±1.55 | -6.95±0.70 |
| $\beta = 50$ | -8.15±1.02 | -9.10±1.56 | -7.05±0.72 |

Values represent mean ± SD of predicted $\log P_{exp}$ (cm/s).

**Table S9.** List of the 35 molecular descriptors used as inputs to the XGBoost model for predicting Caco-2 permeability.

| Feature | Description |
|---|---|
| MolWt | Molecular Weight |
| MolLogP | Lipophilicity (LogP) |
| TPSA | Topological Polar Surface Area |
| NumHDonors | Number of H-Bond Donors |
| NumHAcceptors | Number of H-Bond Acceptors |
| NumRotatableBonds | Number of Rotatable Bonds |
| NumHeteroatoms | Number of Heteroatoms |
| NumAromaticRings | Number of Aromatic Rings |
| NumSaturatedRings | Number of Saturated Rings |
| NumAliphaticRings | Number of Aliphatic Rings |
| RingCount | Total Ring Count |
| FractionCSP3 | Fraction of SP3 Carbons |
| NumValenceElectrons | Number of Valence Electrons |
| NumRadicalElectrons | Number of Radical Electrons |
| HeavyAtomCount | Heavy Atom Count |
| HeavyAtomMolWt | Heavy Atom Molecular Weight |
| MaxPartialCharge | Maximum Partial Charge |
| MinPartialCharge | Minimum Partial Charge |
| MaxAbsPartialCharge | Maximum Absolute Partial Charge |
| MinAbsPartialCharge | Minimum Absolute Partial Charge |
| LabuteASA | Labute Approximate Surface Area |
| BalabanJ | Balaban J Index |
| BertzCT | Bertz CT Index |
| Chi0 | Chi0 Topological Index |
| Chi1 | Chi1 Topological Index |
| Chi0n | Chi0n (normalized) Topological Index |
| Chi1n | Chi1n (normalized) Topological Index |
| Kappa1 | Kappa1 Shape Index |
| Kappa2 | Kappa2 Shape Index |
| Kappa3 | Kappa3 Shape Index |
| NOCount | Nitro Group Count |
| NHOHCount | Hydroxylamine Count |
| MaxEStateIndex | Maximum E-State Index |
| MinEStateIndex | Minimum E-State Index |
| NumAmideBonds | Number of Amide Bonds |

**Table S10.** Hyperparameter configurations of the 10 XGBoost base models in the Caco-2 permeability predictor ensemble.

| Model | Number of trees | Maximum depth | Learning rate | Sub-sampling | Regularization |
|---|---|---|---|---|---|
| base | 210 | 7 | 0.08 | 0.74 | L1=0.01, L2=1.0 |
| more_trees | 250 | 7 | 0.07 | 0.76 | L1=0.01, L2=1.0 |
| deeper | 180 | 9 | 0.08 | 0.72 | L1=0.01, L2=1.0 |
| shallow | 280 | 5 | 0.06 | 0.78 | L1=0.01, L2=1.0 |
| fast_lr | 150 | 7 | 0.10 | 0.74 | L1=0.01, L2=1.0 |

| | | | | | |
|---|---|---|---|---|---|
| reg_strong | 250 | 6 | 0.07 | 0.75 | L1=0.1, L2=1.5 |
| reg_l1 | 230 | 7 | 0.075 | 0.74 | L1=0.2, L2=1.0 |
| reg_l2 | 230 | 7 | 0.075 | 0.74 | L1=0.05, L2=2.0 |
| reg_balanced | 240 | 6 | 0.07 | 0.76 | L1=0.08, L2=1.3 |
| conservative | 300 | 5 | 0.05 | 0.78 | L1=0.05, L2=1.2 |