De novo design of neutralizing protein binders against CCL26 and POSTN, two convergent targets in eosinophilic esophagitis

Authors: [Author list] **Affiliations:** [Affiliations] **Correspondence:** [Corresponding author]

Abstract

Eosinophilic esophagitis (EoE) is a chronic, food-antigen–driven type-2 inflammatory disease of the esophagus. Approved and late-stage therapies converge on a small set of upstream nodes — the IL-4/IL-13 receptor (dupilumab), IL-5/IL-5Rα (mepolizumab, benralizumab), and TSLP — leaving the downstream effector chemokine axis and the tissue-remodeling axis pharmacologically unexploited. Here we combine a nine-cohort EoE transcriptomic meta-analysis with literature and clinical-pipeline mining to nominate two targets that are strongly and reproducibly dysregulated yet have no direct therapeutic: **CCL26** (eotaxin-3), the eosinophil-selective chemokine that is the single most discriminating EoE transcript, and **POSTN** (periostin), a matricellular remodeling protein that remains elevated in histologic remission alongside persistent desmoglein-1 (DSG1) loss. For each target we carried out structure-based de novo binder design — RFdiffusion backbone generation against a defined functional epitope, SolubleMPNN sequence design, and Boltz-2 co-folding validation with an on-target epitope-overlap analysis. From 80 backbones and 1,920 designed sequences we validated 60 binder–target complexes; 59/60 passed interface-confidence thresholds (ipTM > 0.5, complex pLDDT > 0.7). The two leads engage their intended functional surfaces — the CCR3/glycosaminoglycan (GAG) face of CCL26 (ipTM 0.942, 4/5 epitope hotspots, 1,220 Å² buried) and the FAS1 integrin-interaction patch of POSTN (ipTM 0.901, complex pLDDT 0.911, 3/4 hotspots, 842 Å² buried). All designs are computational and require experimental validation; we present them as a mechanistically-grounded starting point for two differentiated EoE therapeutic strategies — anti-chemokine ligand trapping and anti-remodeling neutralization.

Keywords: eosinophilic esophagitis, CCL26/eotaxin-3, periostin/POSTN, de novo protein design, RFdiffusion, ProteinMPNN, Boltz-2, ligand trap

1. Introduction

Eosinophilic esophagitis (EoE) is a chronic immune-mediated disease characterized by eosinophil-predominant inflammation of the esophageal epithelium, symptomatic esophageal dysfunction, and — if untreated — progressive fibrostenotic remodeling. It is driven by a type-2 (Th2) immune response to food antigens, in which epithelial alarmins and Th2 cytokines (IL-4, IL-13, IL-5, TSLP) orchestrate eosinophil recruitment, mast-cell expansion, and barrier disruption.

The therapeutic pipeline reflects this cytokine-centric model. Dupilumab (anti-IL-4Rα, blocking both IL-4 and IL-13 signaling) is approved for EoE; anti-IL-5/IL-5Rα agents (mepolizumab, reslizumab, benralizumab), anti-TSLP (tezepelumab), and anti-IL-13 (cendakimab) are approved in related indications or in late-stage EoE trials. These agents share a common feature: they intercept the disease **upstream**, at the cytokine or eosinophil level. Two biologically central axes remain without a direct therapeutic:

1. **The effector chemokine axis.** CCL26/eotaxin-3 is the terminal, eosinophil-selective signal that converts upstream Th2 activation into physical eosinophil recruitment. It is the single most discriminating transcript separating EoE from gastroesophageal reflux disease (PMID 17900656), yet no approved or late-stage agent targets the eotaxin/CCR3 axis directly.

2. **The remodeling / barrier-persistence axis.** Periostin (POSTN) is a matricellular protein that amplifies subepithelial fibrosis and remodeling. Critically, POSTN accumulation and DSG1 loss persist in histologic remission (PMID 39343172), marking a residual disease process that current eosinophil-directed therapies do not resolve.

We reasoned that these two targets — one at the point of eosinophil recruitment, one at the point of tissue remodeling — represent complementary, unexploited intervention points, and that both are amenable to a neutralizing-biologic (ligand-trap) modality because both are compact, secreted proteins with biochemically-defined functional surfaces. We therefore undertook de novo design of protein binders against a defined functional epitope on each, using a contemporary structure-based pipeline (RFdiffusion → SolubleMPNN → Boltz-2), and report two computationally-validated leads.

---

2. Results

2.1 A nine-cohort meta-analysis nominates CCL26 and POSTN as reproducibly dysregulated, undrugged targets

We assembled a random-effects meta-analysis of nine public EoE case–control transcriptomic cohorts (GSE250595, GSE148381, GSE197702, GSE303169, GSE58640, GSE234973, GSE278888, GSE246323, GSE228083; 25,654 genes). Both candidate targets are strongly and consistently up-regulated: **CCL26** at pooled +4.56 $\log_2$ fold-change (adjusted $p = 3\times10^{-4}$, up in 9/9 cohorts) and **POSTN** at +5.55 (adjusted $p = 3\times10^{-16}$, up in 8/8 cohorts) (**Fig. 2A**, **Fig. S1**, **Table S1**). Both sit among the most strongly up-regulated genes in the disease, alongside effector-cell markers (CPA3, ALOX15) and opposite the concordantly down-regulated barrier genes (DSG1, SPINK7, FLG).

Cross-referencing the meta-analysis against literature attention (591 EoE abstracts, 2015–2026) and the interventional trial landscape (157 EoE trials) shows that CCL26 and POSTN occupy a distinctive position: high omics effect size and mechanistic literature support, but **zero direct interventional trials** — in contrast to IL-13, IL-5, and TSLP, which dominate both literature attention and the trial pipeline (**Fig. 2B**, **Fig. S2**, **Table S3**). On an omics-derived druggability shortlist that filters for effect size, direction concordance, and antibody tractability, POSTN ranks 3rd of 20 and CCL26 11th of 20 (**Fig. S4**, **Table S2**).

Single-cell differential expression localizes each target to a distinct epithelial compartment: CCL26 to suprabasal epithelium and POSTN to basal epithelium (**Fig. 2C**, **Fig. S3**), consistent with CCL26 as an epithelially-secreted chemokine and POSTN as a basally-produced remodeling signal.

2.2 Mechanistic rationale and epitope definition

CCL26 (eotaxin-3). CCL26 signals exclusively through CCR3 and is an eosinophil-selective chemoattractant induced by IL-4 (≈100× more potent than IL-13) and IL-13, synergizing with TNF-α and suppressed by glucocorticoids (PMID 12061839). Its chemotactic function requires two molecular interactions, either of which is a valid neutralization target: (i) CCR3 engagement via the chemokine N-terminus and N-loop, and (ii) glycosaminoglycan (GAG) binding via basic residues in the α-helical region, which immobilizes the chemokine into the haptotactic gradient eosinophils follow (PMID 35742962, PMID 26701132). We defined the design epitope as the composite CCR3/GAG basic face spanning the N-loop (residues 16–17) and the C-terminal α-helix basic cluster (R54/K55/K56) (**Fig. S6**). The neutralization logic is an **anti-ligand mini-binder / trap** that sequesters secreted CCL26 and occludes this functional surface — preferred over an anti-CCR3 approach because CCR3 is a broadly-expressed seven-transmembrane GPCR, whereas mature CCL26 is a compact (~71-residue) secreted protein with a clean extracellular target surface.

POSTN (periostin). The primary epithelial barrier lesion in EoE is IL-13–driven loss of DSG1; periostin is the top induced gene when DSG1 is lost, and DSG1 loss potentiates inflammation partly *through* POSTN (PMID 24220297). POSTN acts as an integrin ligand that drives myofibroblast differentiation and matrix remodeling (PMID 28918442, PMID 23019192), and luminal eotaxin-3 and periostin both correlate with remodeling severity (PMID 35405206). We defined the design epitope as an exposed patch on FAS1 domain IV corresponding to the integrin-interaction surface (**Fig. S6**). Importantly, a POSTN binder is best framed as an **anti-remodeling / loop-breaking** agent: it would interrupt the IL-13→DSG1↓→POSTN↑ feed-forward circuit and thereby *indirectly* support barrier recovery, rather than directly re-assembling desmosomes. This mechanism is differentiated from every agent in the current EoE pipeline and directly addresses the disease process that persists in histologic remission (PMID 39343172, PMID 35888006).

2.3 A structure-based de novo design pipeline

For each target we prepared a clean single-chain structure of the functional unit — CCL26 from the NMR structure 1G2S (mature 71-residue eotaxin-3) and POSTN from 5WT7 (the isolated FAS1 domain IV, 140 residues) — and defined a spatially-compact hotspot patch on the functional epitope (**Fig. S6**). RFdiffusion generated 40 binder backbones per target (binder length 65–90 aa) targeting these hotspots using the complex/PPI checkpoint. Backbone quality control confirmed that all 40 CCL26 backbones and 34/40 POSTN backbones engaged ≥3 epitope hotspot residues; the flatter, larger FAS1 face produced more variable docking geometry (**Fig. 3B**, **Fig. S5**). SolubleMPNN then designed 1,920 sequences (8 sequences × 3 sampling temperatures × 80 backbones), holding the target fixed and designing only the binder chain (**Fig. 3C**, **Fig. S7**). A sequence-complexity filter (entropy ≥ 2.6, maximum single-residue run ≤ 5, top-amino-acid fraction ≤ 0.40) removed low-diversity designs, retaining 1,074/1,920 (**Fig. S8**), from which 30 designs per target were selected by a composite of MPNN score and sequence entropy for structural validation (**Fig. 3A**).

2.4 Boltz-2 fold-back validation and lead selection

We validated the 60 selected binder–target pairs by Boltz-2 co-folding (three recycling steps, three diffusion samples per prediction) and assessed both interface confidence and on-target engagement (**Fig. 4**). All 60 complexes folded successfully. **59/60 passed** the interface-confidence thresholds (ipTM > 0.5 and complex pLDDT > 0.7): 30/30 CCL26 designs (maximum ipTM 0.942, median 0.852) and 29/30 POSTN designs (maximum ipTM 0.904, median 0.780) (**Fig. 4A**, **Fig. S9**, **Table S4**).

Because interface confidence alone can be optimistic about geometry that is confident but off-target, we scored each design by an on-target composite, ipTM × (0.5 + 0.5 × epitope-coverage) × min(pLDDT/0.7, 1), which rewards designs that are both confident and docked onto the intended epitope hotspots (**Fig. 4C**, **Fig. S11**). Epitope engagement differed between targets: 13/30 CCL26 designs contacted ≥3 of the 5 epitope hotspots, versus 5/30 POSTN designs contacting ≥3 of 4, consistent with the flatter POSTN integrin face being harder to engage focally (**Fig. S10**).

2.5 Two validated lead binders engage their functional epitopes

The top-ranked design per target was selected as the lead (**Fig. 1**, **Fig. 5**).

CCL26 lead (design `ccl26_design_39_T0.3`, 67-aa binder): interface ipTM 0.942, complex pLDDT 0.779, pTM 0.956; buries 1,220 Å² across a 25-residue target and 27-residue binder interface; contacts 4/5 epitope hotspots (residues 16, 54, 55, 56), engaging both the N-loop and the C-terminal α-helix basic cluster — i.e. the composite CCR3/GAG face targeted by design (**Fig. 5A**, **Fig. S12**).

POSTN lead (design `postn_design_39_T0.1`, 81-aa binder): interface ipTM 0.901, complex pLDDT 0.911, pTM 0.941; buries 842 Å² across a 22-residue target and 20-residue binder interface; contacts 3/4 epitope hotspots (residues 83, 85, 108) on the FAS1 integrin-interaction patch (**Fig. 5B**, **Fig. S12**).

Residue-level interface distance maps confirm that the closest binder–target contacts localize to the epitope hotspots rather than off-target surfaces for both leads (**Fig. 5C**, **Fig. S12**). Interactive 3D rotation videos of both lead complexes are provided as supplementary media.

---

3. Discussion

We report two computationally-validated de novo protein binders directed at complementary, currently-undrugged nodes of EoE pathogenesis. The design rationale is grounded in a reproducible nine-cohort omics signal cross-referenced against the mechanistic literature and the clinical pipeline, which distinguishes this work from binder-design exercises that begin from a target chosen for tractability alone.

Mechanistic differentiation. CCL26 and POSTN sit downstream of the cytokine nodes that current therapies target. A CCL26 trap would block the terminal eosinophil-recruitment step regardless of which upstream cytokine (IL-4, IL-13, TSLP-driven Th2) initiated it, making it complementary to — not redundant with — approved anti-IL-4Rα and anti-IL-5 agents. A POSTN binder addresses the remodeling and barrier-persistence axis that eosinophil-directed therapy leaves behind; because DSG1 loss and periostin accumulation persist in histologic remission, this axis is a genuine residual unmet need.

The barrier question. A central motivation was whether targeting POSTN could restore esophageal barrier function. The evidence supports a precise, limited claim. POSTN is not the primary junctional disruptor — that is DSG1 loss — but it is the top gene induced by DSG1 loss and a driver of the feed-forward remodeling loop. A POSTN binder would therefore be expected to support barrier recovery **indirectly**, by interrupting the IL-13→DSG1↓→POSTN↑ circuit, rather than by directly re-assembling desmosomes. The junctional machinery is restorable in principle when the driving insult is removed (PMID 34458999), which makes loop interruption a plausible route to barrier benefit, but this should be framed as a downstream consequence, not a direct binding effect.

Limitations. These designs are entirely in silico. Interface confidence metrics (ipTM, pLDDT) and Boltz-2 co-folding are predictive, not experimental, and de novo binder success rates at the bench remain modest even for high-confidence designs. Three specific caveats apply. First, the lead binder sequences are strongly helical and acidic, a known compositional bias of SolubleMPNN on these backbone types; the acidic character is electrostatically complementary to the basic CCL26 GAG face but warrants explicit expression, solubility, and specificity testing. Second, the POSTN target was the isolated FAS1 domain IV, not the full-length multi-domain protein, so binder behaviour in the context of the intact protein and its integrin partners is untested. Third, POSTN has physiological roles in tissue repair, so a therapeutic binder would need to be confirmed neutralizing (not agonist) and sufficiently selective.

Next steps. The immediate priorities are experimental: recombinant expression of both leads and their targets, biophysical affinity measurement (SPR/BLI), and functional neutralization assays — CCR3/GAG-dependent eosinophil chemotaxis for the CCL26 trap and integrin-signaling / myofibroblast-differentiation readouts for the POSTN binder. Computationally, orthogonal co-folding (Chai-1, AlphaFold3-class models), explicit affinity prediction, and sequence-optimization to reduce the acidic/helical bias would strengthen the candidate set before synthesis.

---

4. Methods

Omics meta-analysis and target selection. Nine public EoE case–control transcriptomic cohorts were combined by random-effects meta-analysis (25,654 genes), yielding pooled $\log_2$ fold-changes, standard errors, heterogeneity (I^2), and Benjamini–Hochberg-adjusted *p*-values. Targets were cross-referenced against a literature corpus (591 EoE abstracts, 2015–2026, PubMed) and the interventional trial landscape (157 EoE trials, ClinicalTrials.gov), and against an omics-derived druggability shortlist filtered for effect size, direction concordance, single-cell specificity, and antibody tractability. Single-cell differential expression was drawn from an EoE single-cell dataset (cell types: endothelial, basal/suprabasal epithelial, mast, myeloid, T cell).

Target structure preparation. CCL26 was taken from PDB 1G2S (solution NMR, mature 71-residue eotaxin-3); POSTN from PDB 5WT7 (NMR, FAS1 domain IV, 140 residues). Structures were cleaned to the first model, chain A, standard residues (Biopython). Per-residue solvent-accessible surface area (Shrake–Rupley) confirmed all candidate epitope residues were surface-exposed. Design hotspots were chosen for spatial compactness on the functional epitope: CCL26 residues 16, 17, 54, 55, 56 (N-loop + C-helix GAG/CCR3 basic face); POSTN residues 83, 85, 108, 116 (exposed integrin-interaction patch on FAS1-IV).

Backbone generation (RFdiffusion). RFdiffusion (complex/PPI checkpoint, `Complex_base_ckpt.pt`) generated 40 binder backbones per target with the target held fixed and hotspot residues specified via `ppi.hotspot_res`; binder length 65–90 residues, denoiser noise scales set to 0. Backbones were QC'd for binder length, radius of gyration, and number of binder Cα atoms within 10 Å of hotspot Cα atoms.

Sequence design (SolubleMPNN). ProteinMPNN in SolubleMPNN mode (`v_48_020` weights) designed 8 sequences per backbone at each of three sampling temperatures (0.1, 0.2, 0.3), designing only the binder chain (chain B) with the target chain (A) fixed — 1,920 sequences total. Sequences were filtered for complexity (Shannon entropy ≥ 2.6 bits, maximum single-residue run ≤ 5, top-amino-acid fraction ≤ 0.40); 30 designs per target were selected by a composite of MPNN score and entropy for validation.

Structure validation (Boltz-2). Each selected binder–target pair was co-folded with Boltz-2 (`--use_msa_server`, 3 recycling steps, 3 diffusion samples), reporting interface ipTM, complex pLDDT, pTM, and an aggregate confidence score. Pass criteria were ipTM > 0.5 and complex pLDDT > 0.7. On-target engagement was scored as ipTM × (0.5 + 0.5 × epitope-coverage) × min(pLDDT/0.7, 1), where epitope-coverage is the fraction of design hotspots with any target atom within 5 Å of a binder atom. The top-ranked design per target by on-target score was selected as the lead. Interface residues (any atom < 5 Å across chains) and buried surface area ((SASA_target + SASA_binder − SASA_complex)/2, Shrake–Rupley) were computed for each lead.

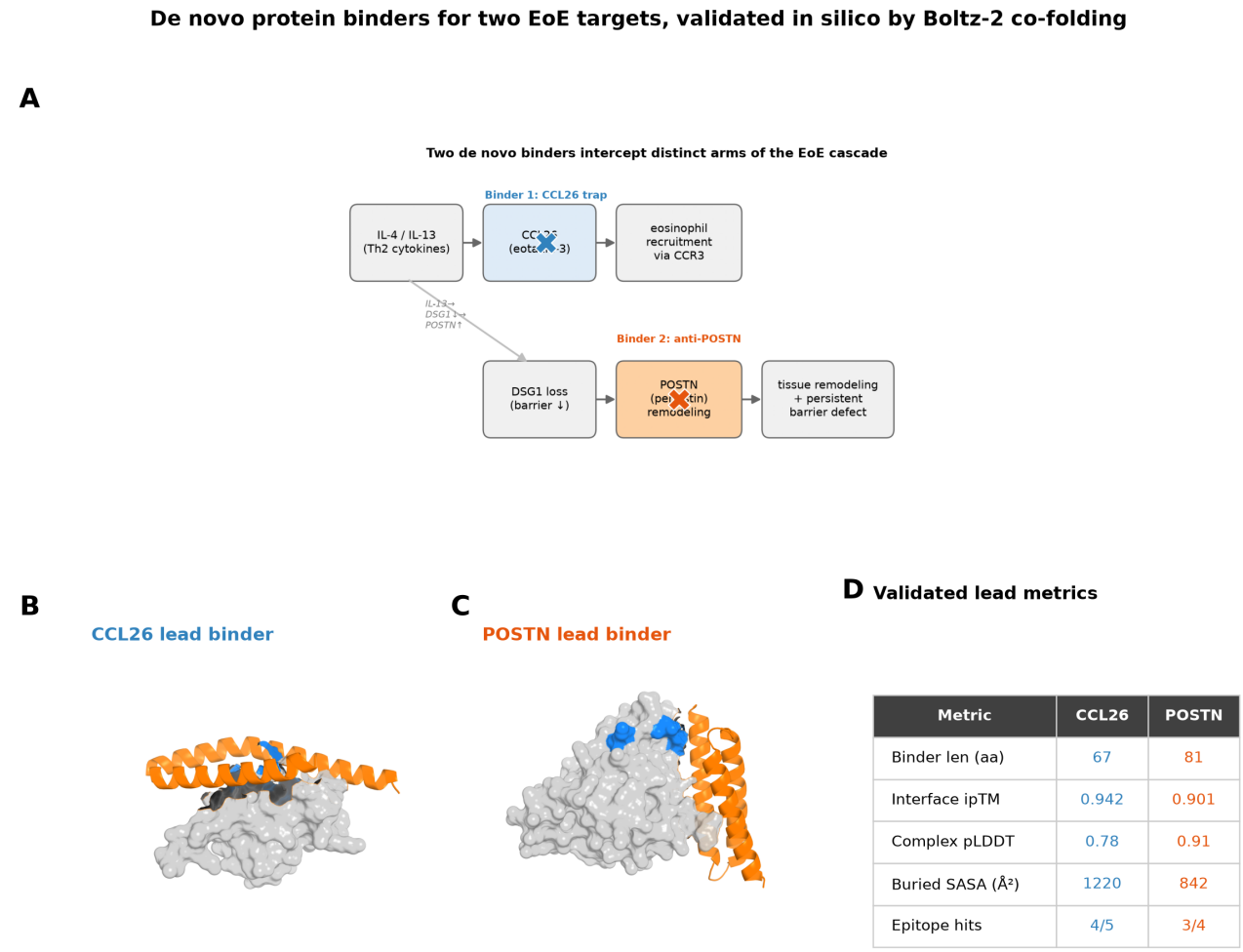
Compute and visualization. Backbone generation, sequence design, and co-folding were run on GPU (NVIDIA A100/A10G). Lead complexes were rendered and animated with PyMOL (target surface, epitope hotspots highlighted, binder cartoon) and encoded to video with ffmpeg. Figures were produced in matplotlib.

---

5. Data and code availability

All intermediate and final data products are available as project artifacts: the meta-analysis signature, druggability shortlist, literature×omics cross-reference, single-cell differential expression, backbone QC tables, the 1,920-sequence SolubleMPNN table, the 60-complex validation table, lead complex structures (PDB), and lead specification sheets (see **Supplementary Information** for the full index).

Figure legends (main)



Metric	CCL26	POSTN
Binder len (aa)	67	81
Interface ipTM	0.942	0.901
Complex pLDDT	0.78	0.91
Buried SASA (Å²)	1220	842
Epitope hits	4/5	3/4

Figure 1. De novo protein binders for two EoE targets, validated in silico by Boltz-2 co-folding. (A) The EoE cascade with the two intervention points: a CCL26 trap blocks the eosinophil-recruitment arm (IL-4/IL-13 → CCL26 → CCR3), and an anti-POSTN binder blocks the remodeling arm (DSG1 loss → POSTN → tissue remodeling + persistent barrier defect), interrupting the IL-13→DSG1↓→POSTN↑ feed-forward loop. (B, C) Rendered lead binder–target complexes: target as grey surface, epitope hotspots in blue, de novo binder as orange cartoon. (D) Validated lead metrics.

Target selection: convergent transcriptomic, literature, and single-cell evidence for CCL26 and POSTN

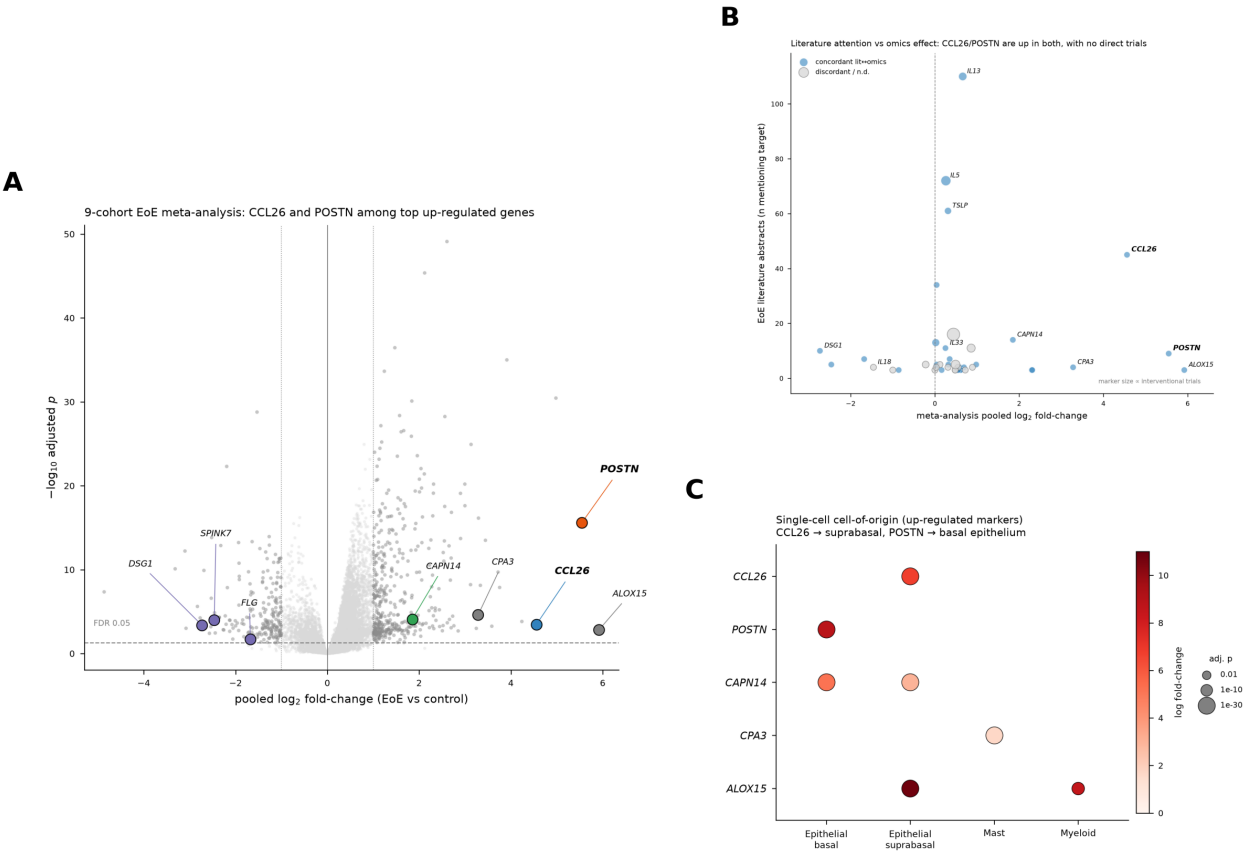


Figure 2. Target selection: convergent transcriptomic, literature, and single-cell evidence. (A) Nine-cohort meta-analysis volcano; CCL26 and POSTN (bold) are among the most strongly up-regulated genes, opposite concordantly down-regulated barrier genes (DSG1, SPINK7, FLG). (B) Literature attention vs. omics effect size, marker size proportional to interventional-trial count; CCL26/POSTN combine high omics effect with zero direct trials. (C) Single-cell cell-of-origin: CCL26 in suprabasal epithelium, POSTN in basal epithelium.

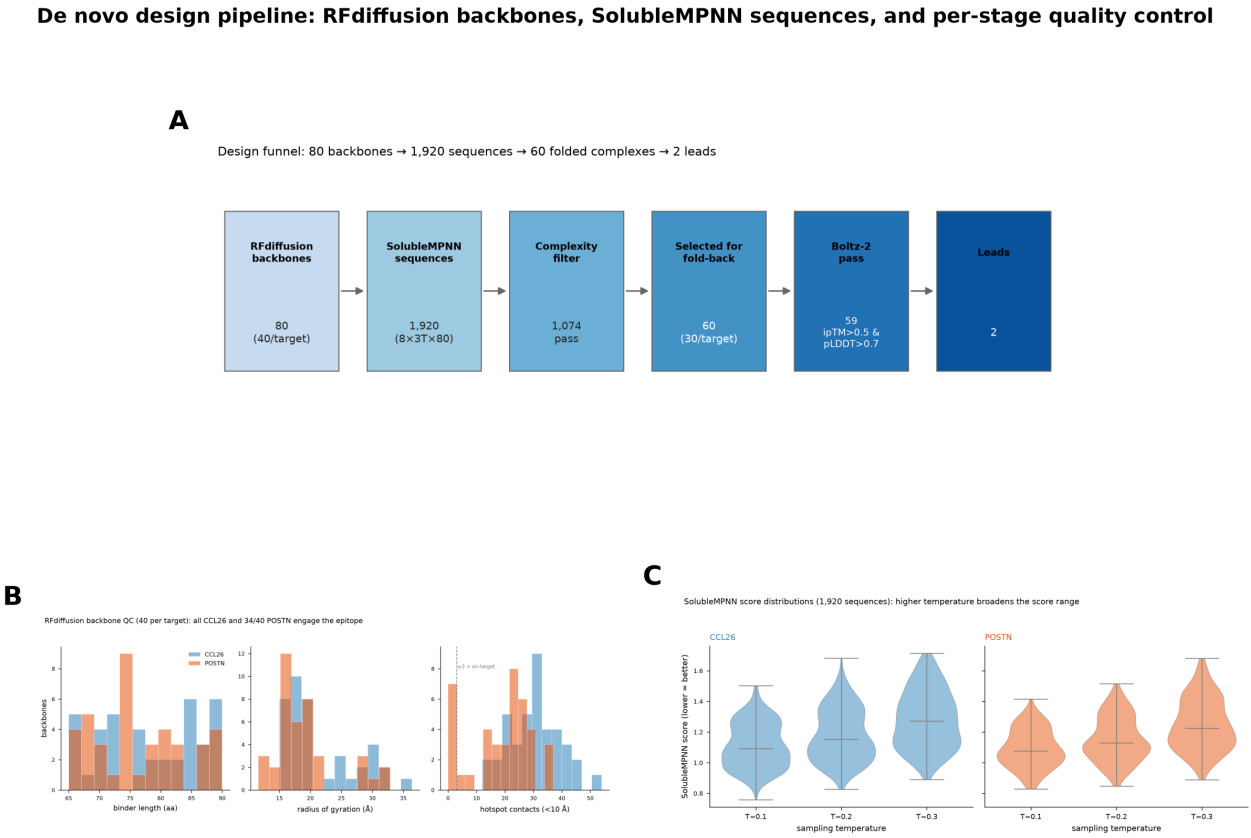


Figure 3. De novo design pipeline. (A) Design funnel from 80 RFdiffusion backbones through 1,920 SolubleMPNN sequences and complexity filtering to 60 folded complexes and 2 leads. (B) Backbone QC distributions (binder length, radius of gyration, hotspot contacts) for 40 backbones per target. (C) SolubleMPNN score distributions by sampling temperature.

Fold-back validation: interface confidence, epitope engagement, and on-target ranking select two leads

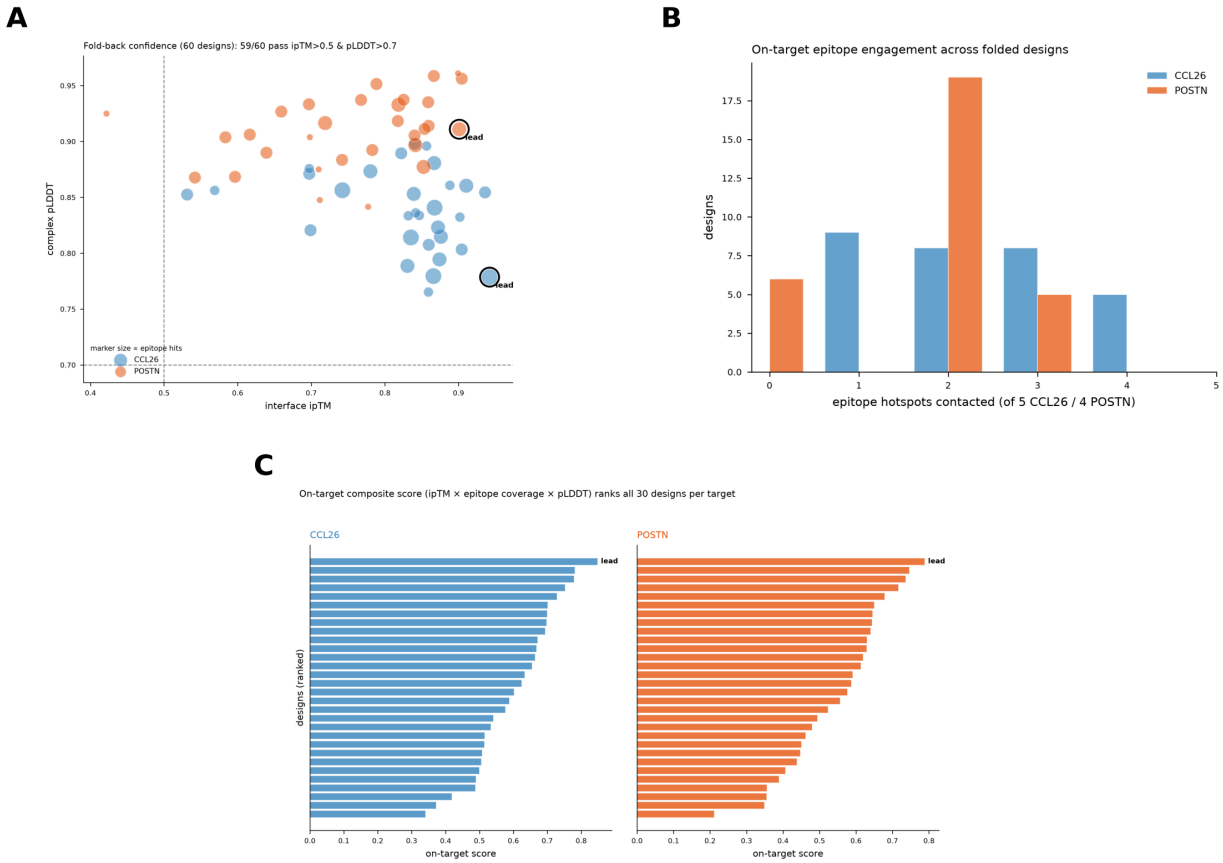


Figure 4. Fold-back validation. (A) Interface ipTM vs. complex pLDDT for all 60 designs; 59/60 pass thresholds (dashed lines), leads circled, marker size proportional to epitope hotspot hits. (B) Distribution of epitope hotspots contacted per design. (C) On-target composite score ranking all 30 designs per target; leads marked.

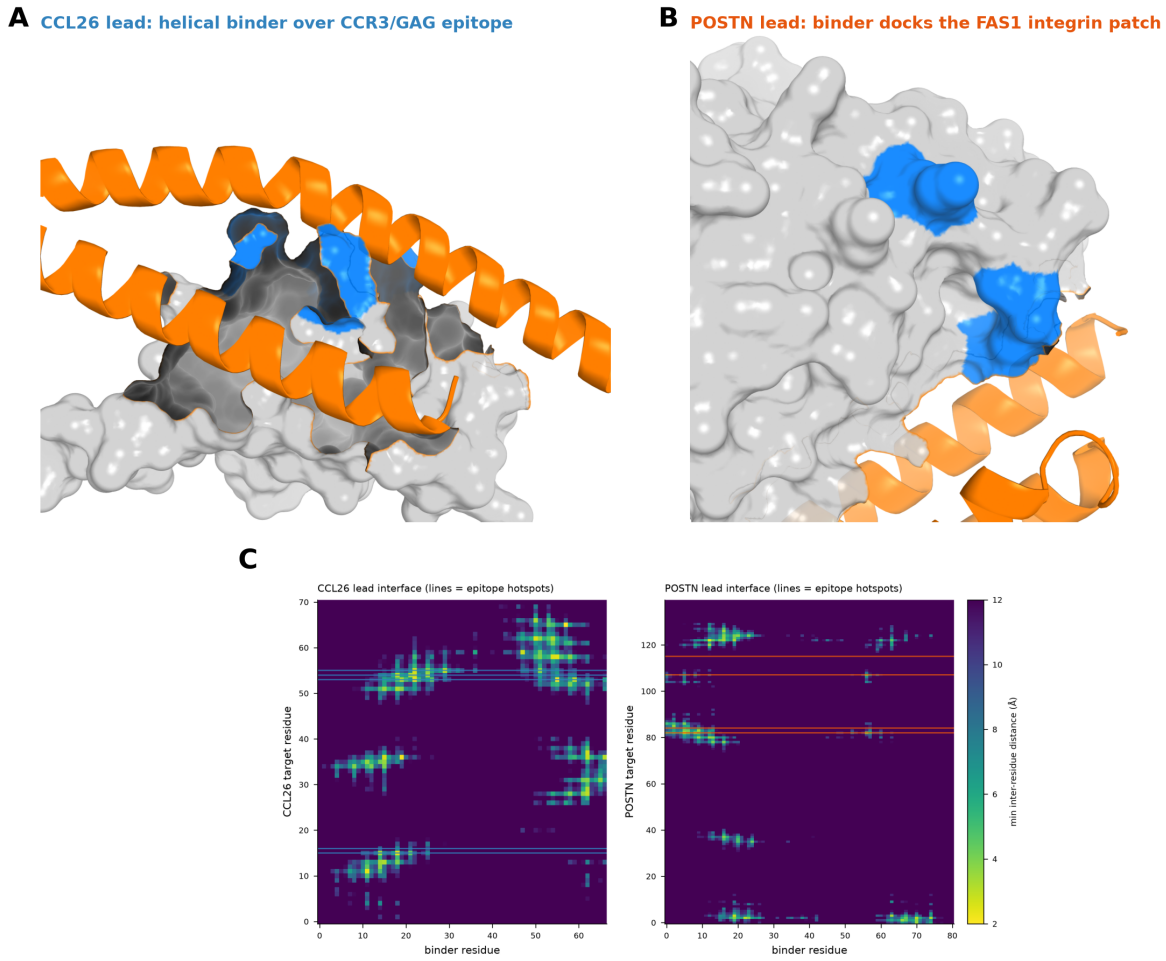


Figure 5. Lead binders engage the intended functional epitopes. (A) CCL26 lead close-up: helical binder over the CCR3/GAG basic face (epitope hotspots blue). (B) POSTN lead close-up: binder docked onto the FAS1 integrin patch. (C) Residue-level interface distance maps; close-contact bands (yellow) align with epitope hotspot rows (colored lines).

Manuscript prepared as a hackathon deliverable (Built with Claude: Life Sciences). All designs are computational and require experimental validation.