RBX1 De Novo Binder Design

Campaign Report

Sergio E. Mares

March 23, 2026

Contents

1 Executive Summary

This report documents a large-scale de novo binder design campaign targeting the human RBX1 RING-H2 domain. Key outcomes:

- **20,724** unique binder designs were generated and scored with Boltz-2.
- **60** designs were validated with AlphaFold3 (AF3); **14** achieved AF3 ipTM ≥ 0.70 ("top hits").
- $\sim$**1,225** designs were validated with OpenFold3 (OF3).
- **81** designs were submitted to the ADAPTYV Boltz-2 competition and independently re-scored; **33** survived (competition ipSAE ≥ 0.6).
- **Key finding:** Boltz-2 pipeline scores (ipSAE, ipTM) do *not* predict AF3 validation ($r \approx 0.08$ for ipTM, $r = -0.044$ for ipSAE) or competition outcomes ($r \approx 0.31$ for ipSAE).
- **Best overall design:** `w2_design_1209` — AF3 ipTM = 0.77, competition Pb ipSAE = 0.81, OF3 ipTM = 0.75.

2 Target: RBX1

RBX1 (RING-Box protein 1, PDB: 5H0P) is a 108-amino-acid RING-H2 domain that coordinates three zinc ions and serves as the catalytic subunit of Cullin–RING E3 ubiquitin ligase complexes. The target presents two distinct binding interfaces:

- **E2 face** ($\sim$35 interface residues): the surface that recruits E2 ubiquitin-conjugating enzymes. This face was the primary design target, targeted with enhanced hotspot sets derived from DMS data.
- **Cullin face** ($\sim$31 interface residues): the surface that contacts the Cullin scaffold protein.

A 165-sequence multiple sequence alignment (MSA) was used for evolutionary conservation analysis. Zinc-coordinating cysteines and histidines are absolutely conserved and dominate ESM-2 deep mutational scanning (DMS) sensitivity profiles.

3 Design Pipeline

Two complementary pipelines were employed:

1. **RFdiffusion** → **ProteinMPNN** → **Boltz-2**: The primary pipeline. RFdiffusion generates backbone scaffolds conditioned on target hotspot residues; ProteinMPNN designs sequences; Boltz-2 scores complexes with MSA input.
2. **BoltzGen**: An alternative generative pipeline that directly samples binder sequences and structures.

Hotspot configurations included:

- **E2 Enhanced** (+DMS-derived contacts): the majority of E2-face designs.
- **E2 Standard**: canonical E2-face hotspots.

- **Cullin**: Cullin-face hotspots.

Table 1: Campaign breakdown by target face.

Campaign	Designs	Mean ipSAE	AF3-validated
E2 Face	12,112	0.227	45
Cullin Face	6,791	0.205	6
BoltzGen	1,820	0.446	9
Total	**20,723**	—	**60**

4 Scoring and Validation

Designs were evaluated at four levels of increasing cost and reliability:

1. **Boltz-2 (primary scoring):** All 20,724 designs scored using Boltz-2 with target MSA. Metrics: ipTM, pTM, pLDDT, ipSAE.

2. **OpenFold3 (secondary validation):** $\sim$1,225 top designs validated. OF3 provides an independent structure prediction check.

3. **AlphaFold3 (ground truth):** 60 designs validated via the AF3 web server. AF3 ipTM ≥ 0.70 is considered a "top hit."

4. **ADAPTYV Competition (Pb):** 81 designs submitted and independently re-scored by competition organizers using a fine-tuned Boltz-2 setup. Survival threshold: ipSAE ≥ 0.6.

5 Cross-Validation Analysis

A central finding of this campaign is that **Boltz-2 pipeline metrics do not predict AF3 or competition outcomes**. This section presents the key correlation analyses.

5.1 ipSAE vs. AF3 ipTM

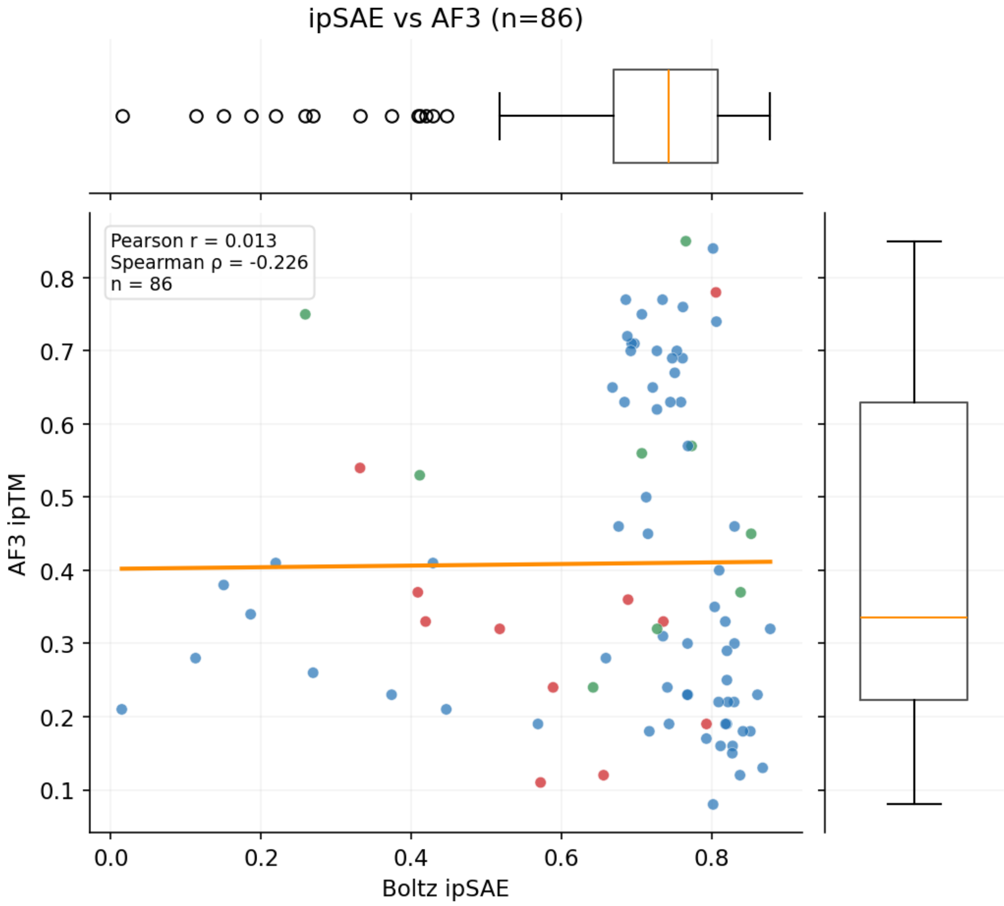

Figure 1: Boltz-2 ipSAE vs. AlphaFold3 ipTM for 60 validated designs. Pearson $r = -0.044$. The competition ranking metric does not predict real binding as assessed by AF3.

5.2 Boltz-2 ipTM vs. AF3 ipTM

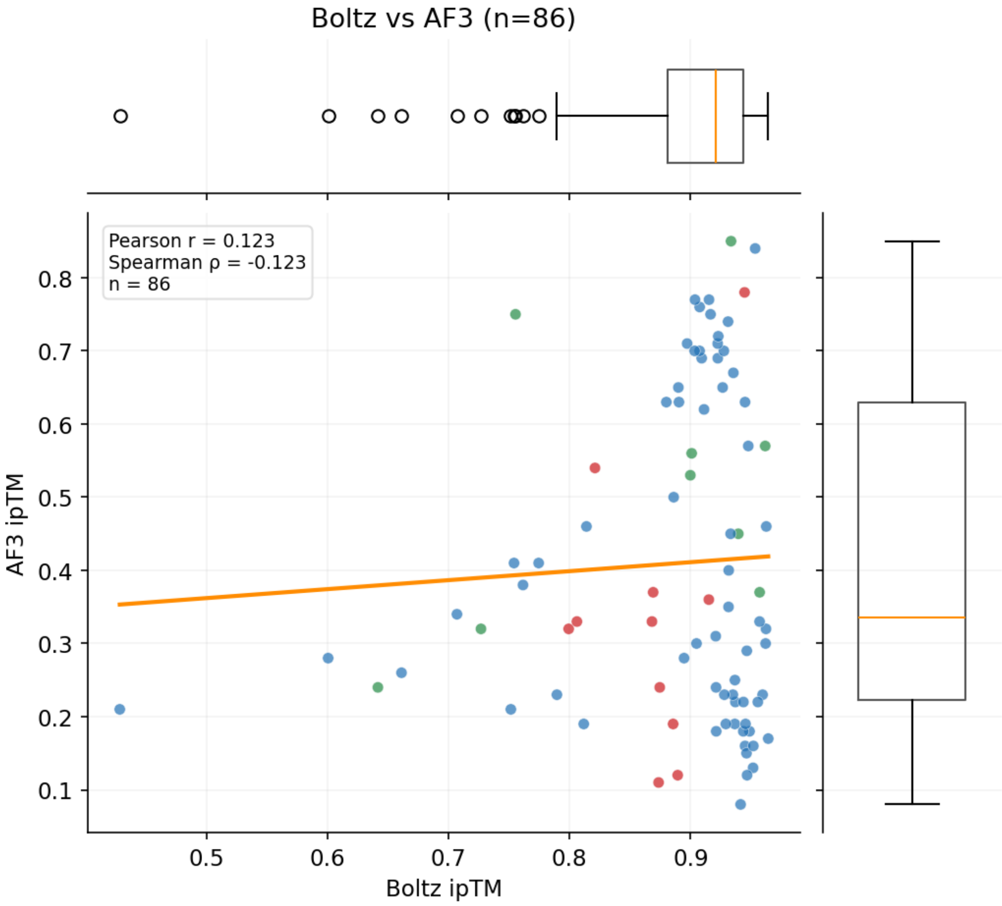

Figure 2: Boltz-2 ipTM vs. AlphaFold3 ipTM ($n = 60$). Pearson $r = 0.08$, Spearman $\rho = -0.12$. No predictive relationship.

5.3 AF3 vs. OF3

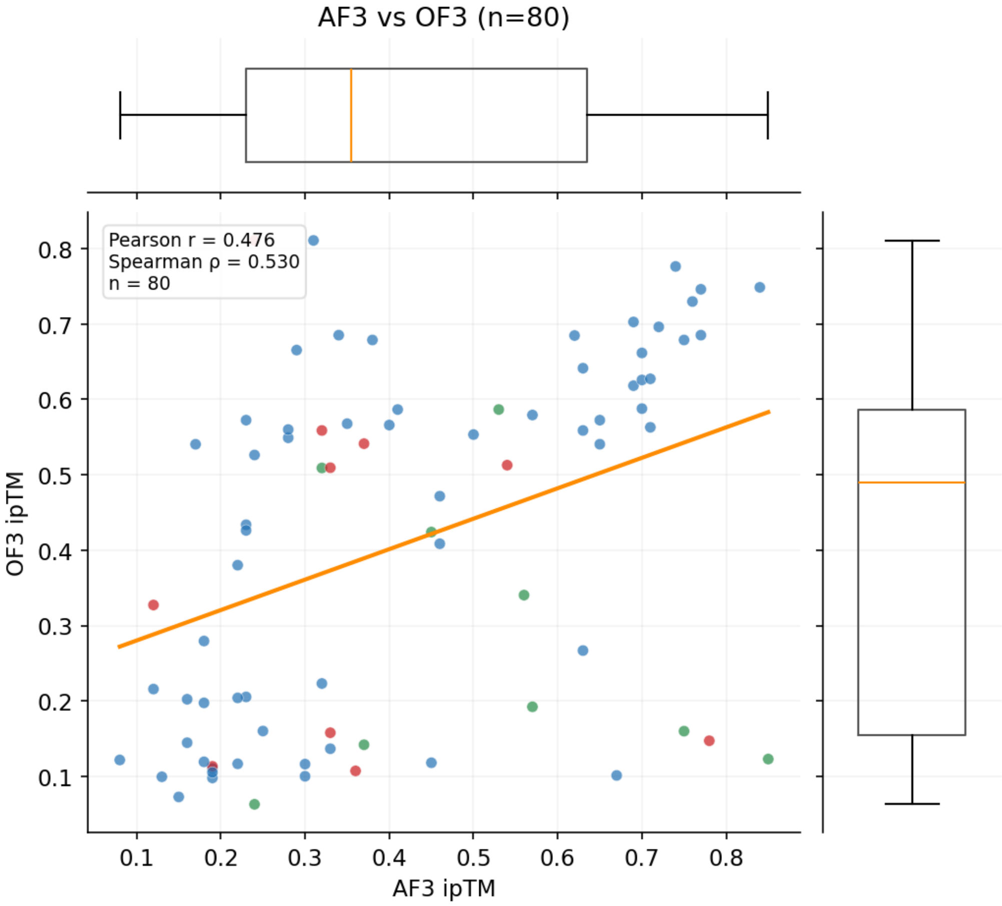

Figure 3: AlphaFold3 ipTM vs. OpenFold3 ipTM for 49 designs with both scores. Pearson $r = 0.57$, Spearman $\rho = 0.49$. Moderate positive correlation.

5.4 Boltz-2 vs. OF3

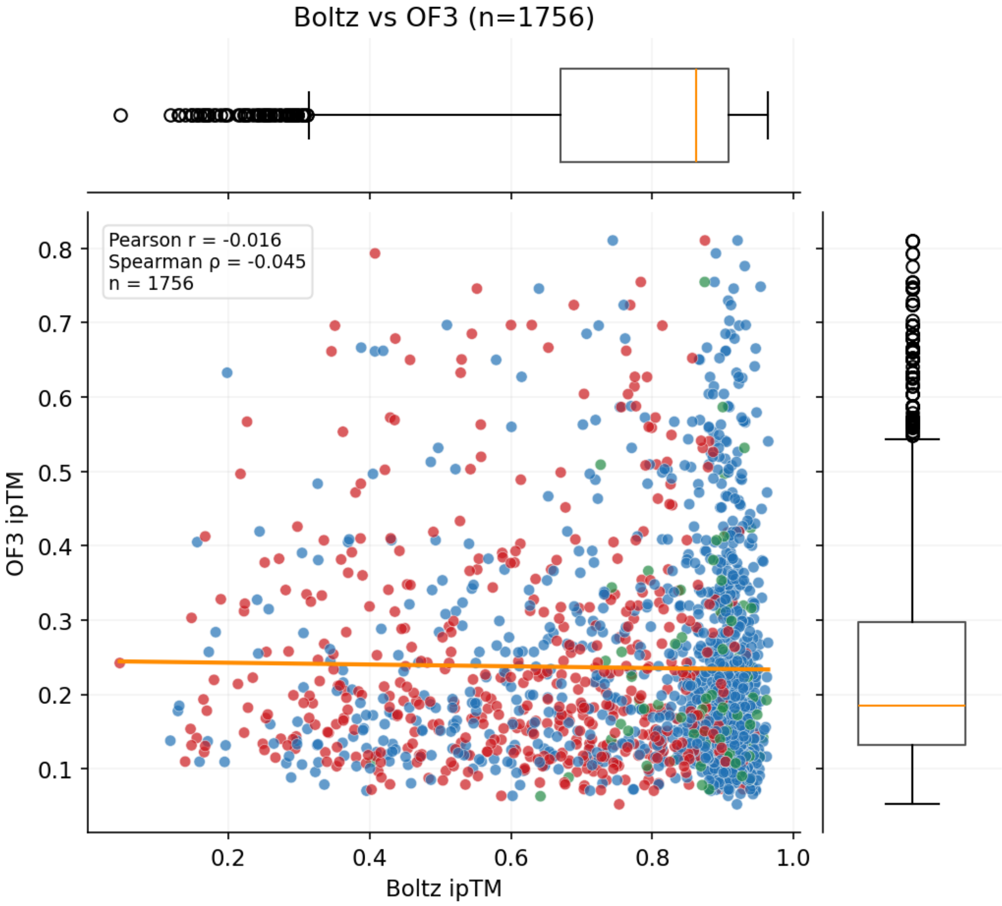

Figure 4: Boltz-2 ipTM vs. OpenFold3 ipTM ($n = 1{,}225$). Pearson $r = -0.01$, Spearman $\rho = -0.04$. No correlation.

6 Top Designs

6.1 Top 18 by AlphaFold3 ipTM

Table 2 lists all designs with AF3 ipTM ≥ 0.50, ranked by AF3 ipTM. The best design by AF3 is `wbg_design_1817` (AF3 = 0.85), a BoltzGen design.

Table 2: Top designs ranked by AlphaFold3 ipTM (≥ 0.50).

Rk	ID	Camp.	Len	AF3	AF3 pTM	OF3	Boltz	Pb	Flag
1	wbg_design_1817	BoltzGen	64	0.85	0.73	0.12	0.93	0.13	top hit
2	w2_design_0283	E2 Face	49	0.84	0.67	0.75	0.95	0.71	top hit
3	w1_design_0357	Cullin	60	0.78	0.67	0.15	0.94	0.69	promote
4	w2_design_1209	E2 Face	65	0.77	0.69	0.75	0.92	0.81	top hit
5	w3_design_3061	E2 Face	57	0.77	0.67	0.69	0.90	0.70	top hit
6	w2_design_5499	E2 Face	63	0.76	0.66	0.73	0.91	0.18	top hit
7	wbg_design_0198	BoltzGen	59	0.75	0.66	0.16	0.76	—	top hit
8	w3_design_1459	E2 Face	65	0.74	0.66	0.78	0.93	0.76	top hit
9	w2_design_0217	E2 Face	41	0.72	0.62	0.70	0.92	0.72	top hit
10	w3_design_3538	E2 Face	43	0.70	0.62	0.63	0.93	0.76	top hit

Rk	ID	Camp.	Len	AF3	AF3 pTM	OF3	Boltz	Pb	Flag
11	w2_design_1208	E2 Face	65	0.70	0.64	0.59	0.91	—	top hit
12	w2_design_0743	E2 Face	57	0.69	0.62	0.62	0.92	—	promote
13	w2_design_4025	E2 Face	64	0.69	0.63	0.70	0.91	—	promote
14	w3_design_2356	E2 Face	45	0.67	0.61	0.10	0.94	0.05	promote
15	w2_design_0218	E2 Face	41	0.65	0.60	0.57	0.93	0.75	promote
16	w1_design_0389	E2 Face	108	0.63	0.67	0.27	0.89	0.69	promote
17	w2_design_6139	E2 Face	41	0.63	0.59	0.64	0.94	0.11	promote
18	w2_design_8092	E2 Face	50	0.62	0.59	0.68	0.91	—	promote

6.2 Top 20 by Competition Pb ipSAE

Table 3 lists the 20 highest-scoring designs in the ADAPTYV competition re-scoring.

Table 3: Top 20 designs ranked by competition Pb ipSAE.

Rk	ID	Camp.	Len	Pb	AF3	OF3	Boltz	ipSAE	Flag
1	w3_design_3063	E2 Face	57	0.83	0.31	0.81	0.92	0.73	discard
2	w2_design_1209	E2 Face	65	0.81	0.77	0.75	0.92	0.73	top hit
3	wcm_design_0709	Cullin	41	0.79	0.33	0.51	0.81	0.42	discard
4	w3_design_1459	E2 Face	65	0.76	0.74	0.78	0.93	0.81	top hit
5	w3_design_3538	E2 Face	43	0.76	0.70	0.63	0.93	0.75	top hit
6	w2_design_0218	E2 Face	41	0.75	0.65	0.57	0.93	0.67	promote
7	wbg_design_1458	BoltzGen	64	0.74	0.53	0.59	0.90	0.41	promote
8	w2_design_3061	E2 Face	53	0.73	0.34	0.69	0.71	0.19	discard
9	w2_design_5942	E2 Face	65	0.72	0.65	0.54	0.89	0.72	promote
10	w2_design_0217	E2 Face	41	0.72	0.72	0.70	0.92	0.69	top hit
11	w2_design_0283	E2 Face	49	0.71	0.84	0.75	0.95	0.80	top hit
12	w2_design_0316	E2 Face	41	0.71	—	0.70	0.93	0.69	—
13	w2_design_6225	E2 Face	48	0.70	0.50	0.55	0.89	0.71	promote
14	w3_design_3061	E2 Face	57	0.70	0.77	0.69	0.90	0.69	top hit
15	w2_design_5293	E2 Face	57	0.70	0.63	0.56	0.88	0.68	promote
16	w1_design_0357	Cullin	60	0.69	0.78	0.15	0.94	0.81	promote
17	w1_design_0389	E2 Face	108	0.69	0.63	0.27	0.89	0.76	promote
18	w2_design_5764	E2 Face	65	0.69	—	0.54	0.90	0.69	—
19	wcm_design_3063	Cullin	43	0.68	0.24	0.81	0.87	0.59	discard
20	w2_design_1458	E2 Face	50	0.68	0.41	0.59	0.75	0.43	review

6.3 Full Sequences of Top 5 AF3 Designs

1. **wbg_design_1817** (AF3 = 0.85, 64 AA, BoltzGen):
`SAAAAAGEKAAAAAEAGASTDELAKLWVEAAKLAAASSDPAAANAAFIGIAKAIVIRMKRLLGL`

2. **w2_design_0283** (AF3 = 0.84, 49 AA, E2 Face):
`LDELMKKLKEELEELREKLREAVKAGDEEKAEELRKRLFDCIIELSKLK`

3. **w1_design_0357** (AF3 = 0.78, 60 AA, Cullin Face):
`SRMEENRRARAERAEALRAAGDLAALEALLAAEAAAQAALAAAAERERAERERRRAEEEA`

4. **w2_design_1209** (AF3 = 0.77, 65 AA, E2 Face):
`LSEEELEELERERRELEELREELERKIEEAESEEEKKELEKKLEEVEAHIRCLEARELILEAAEP`

5. **w2_design_5499** (AF3 = 0.76, 63 AA, E2 Face):
`PKDRKTLEELKKELEERRERAEELLKSPSPSTQSLGKAQLLVCEEAEEKIKKLEEELRKEEEA`

7 Pipeline Performance

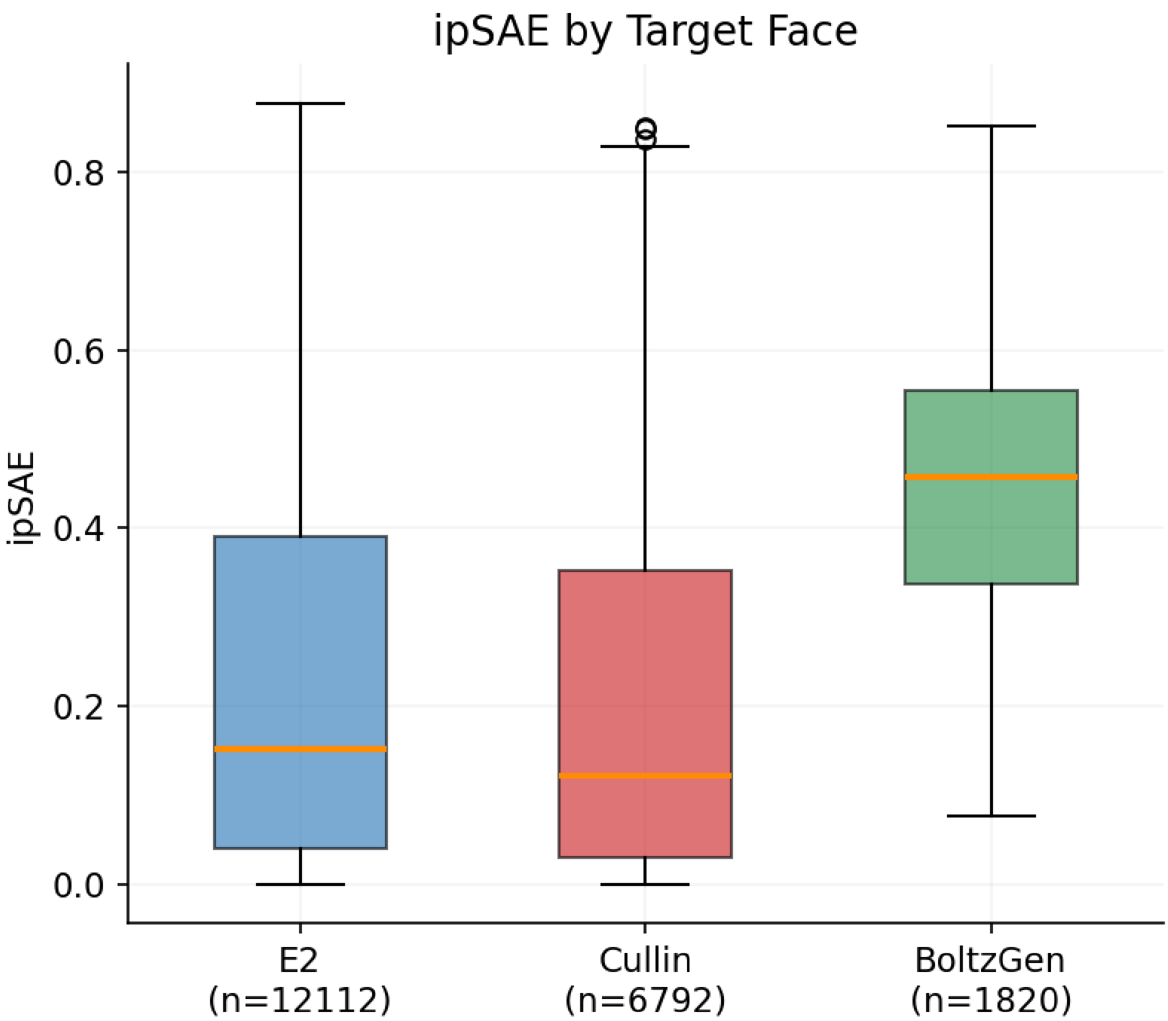

Figure 5: ipSAE distribution by target face. BoltzGen designs show inflated ipSAE (mean 0.446) compared to RFdiffusion campaigns (E2: 0.227, Cullin: 0.205).

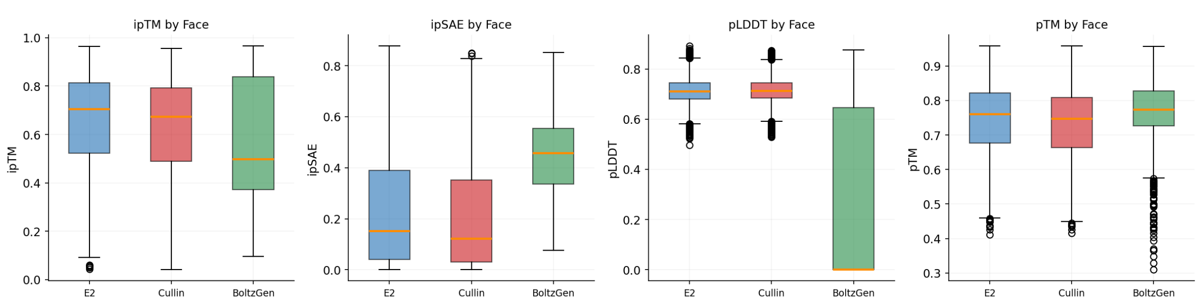

Figure 6: Multi-metric comparison across target faces and pipeline types.

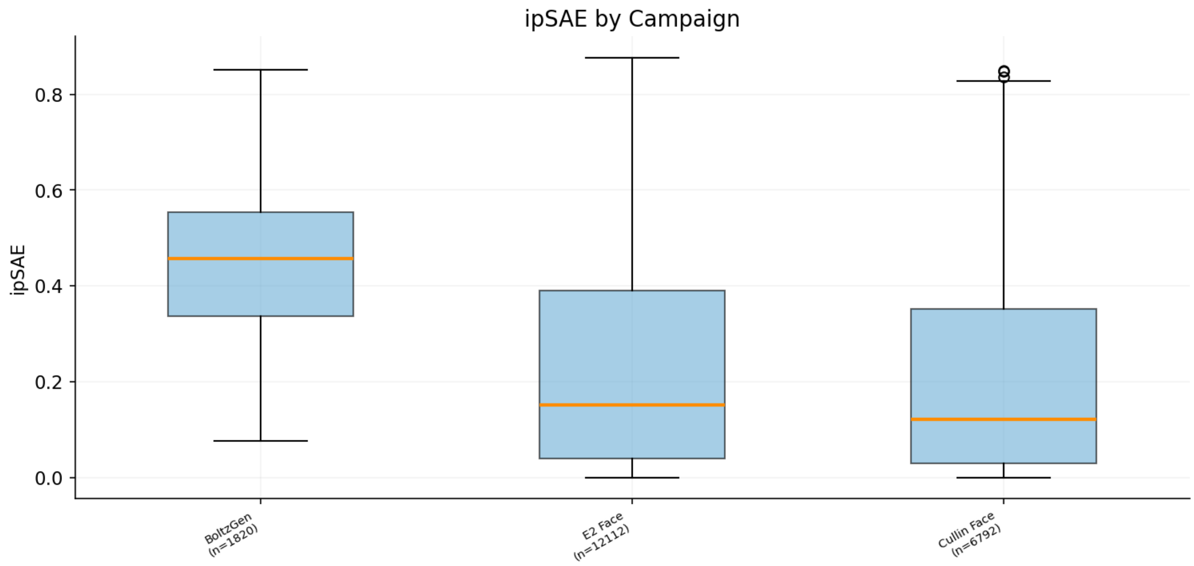

Figure 7: Side-by-side comparison of Boltz-2, OpenFold3, and AlphaFold3 scores for top designs.

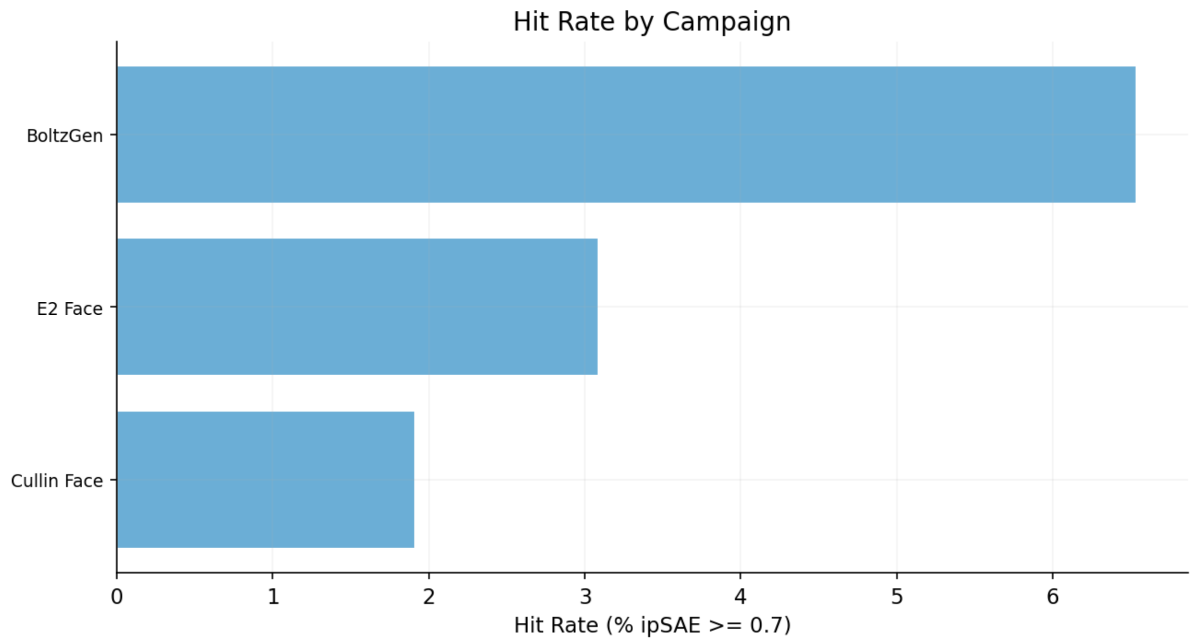

Figure 8: Hit rates across scoring methods and campaigns.

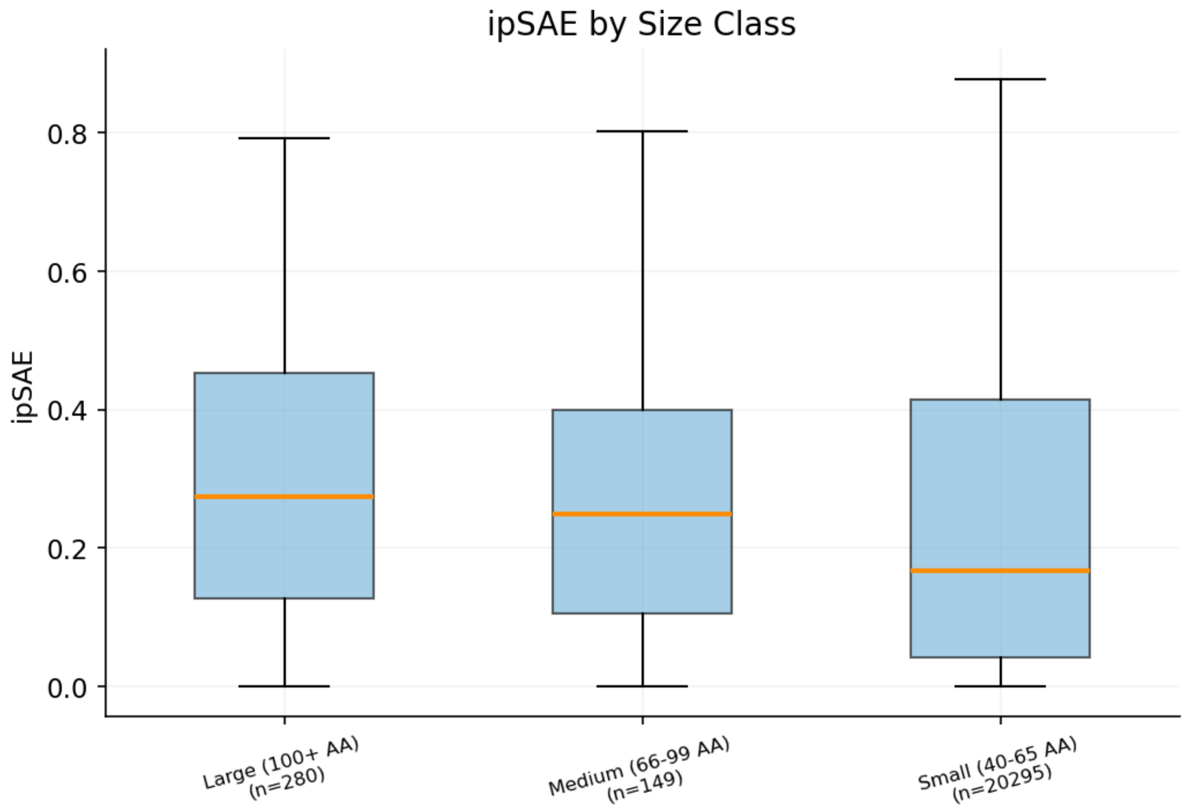

Figure 9: ipSAE distribution by binder size class.

8 AlphaFold3 Insights

Comprehensive analysis of what predicts AF3 validation success across 60 validated designs (14 top hits with AF3 ipTM ≥ 0.70).

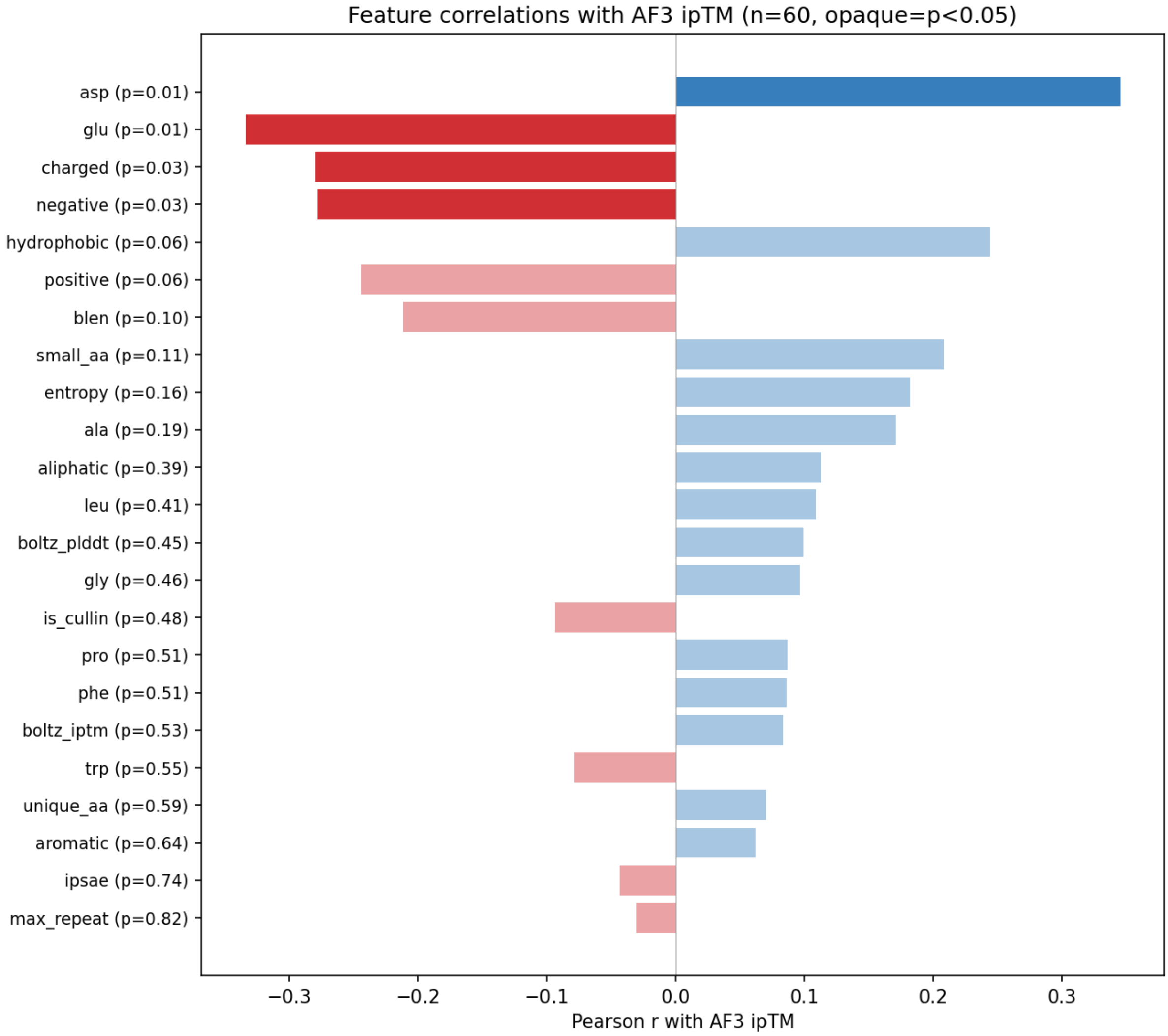

Figure 10: LASSO regression coefficients predicting AF3 ipTM from sequence and structural features. Asp content is the strongest positive predictor; Glu content is negative.

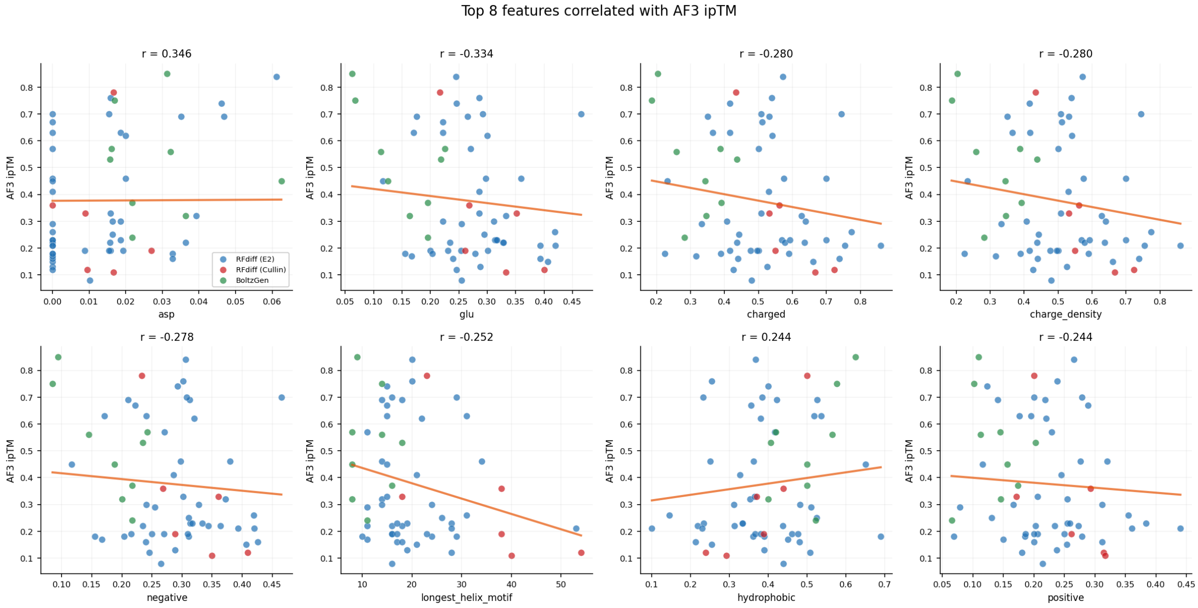

Figure 11: Random forest feature importance for AF3 ipTM prediction.

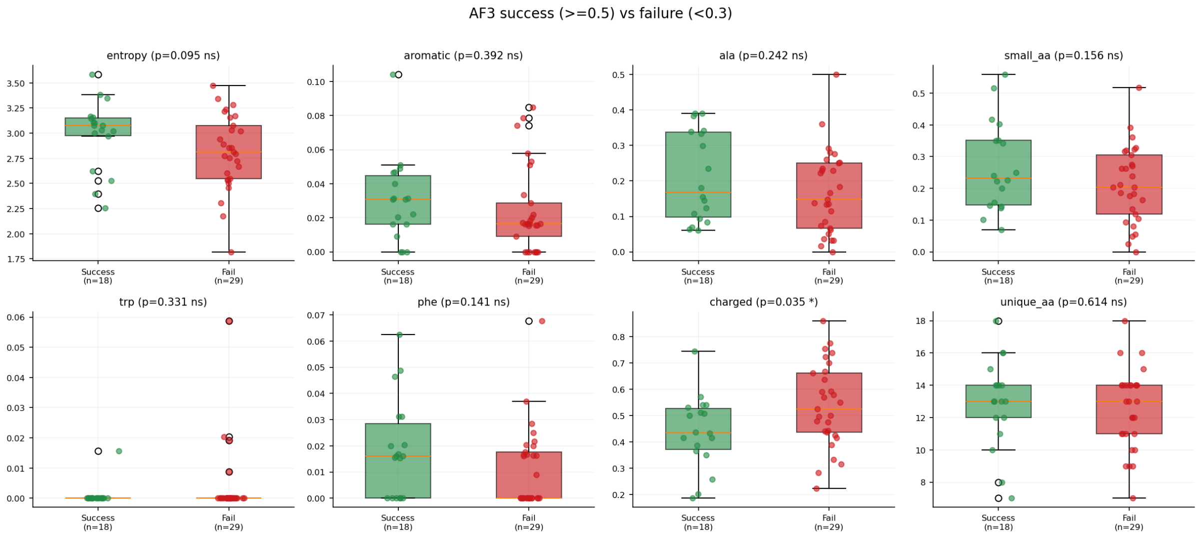

Figure 12: Feature comparison between AF3-successful (≥ 0.70) and AF3-failed (< 0.50) designs.

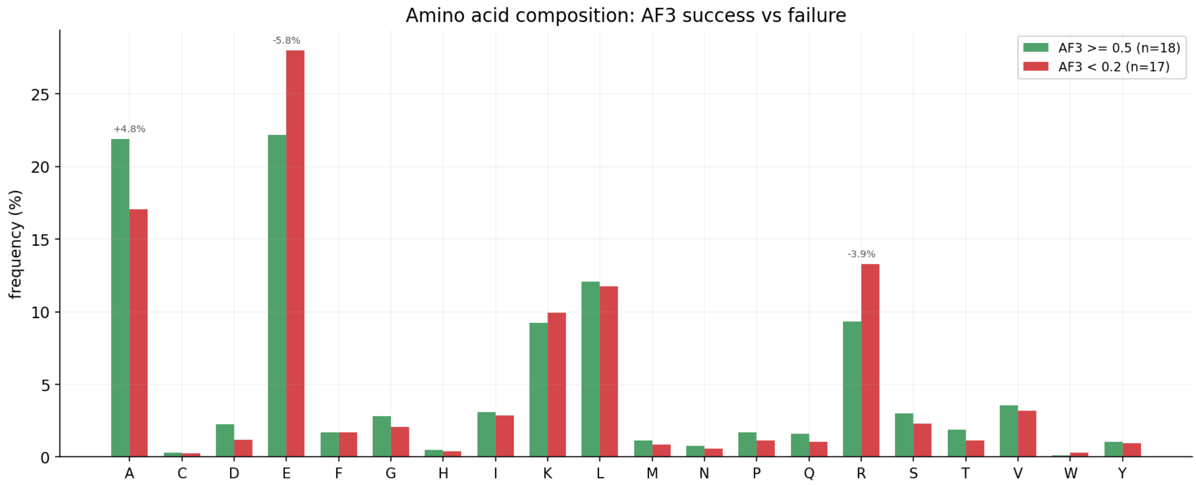

Figure 13: Amino acid composition and sequence-level analysis of AF3-validated designs.

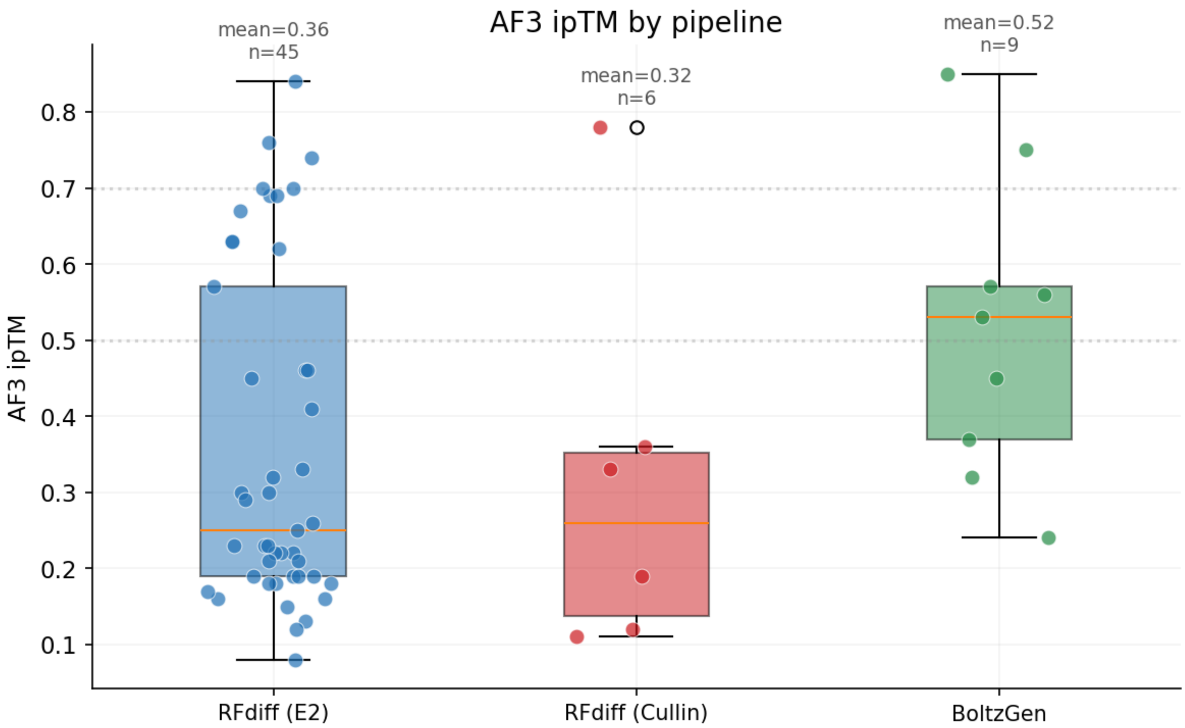

Figure 14: BoltzGen vs. RFdiffusion+MPNN designs in AF3 validation. BoltzGen designs validate at a higher rate despite lower sequence complexity.

Key findings from AF3 analysis:

- Aspartate (D) content is the strongest positive predictor of AF3 success.
- Glutamate (E) content is negatively associated — despite E and D being chemically similar.
- BoltzGen designs validate at a higher rate than RFdiffusion+MPNN designs.
- Shannon entropy and aromatic amino acid content are positively associated with AF3 success.
- Boltz-2 ipTM and ipSAE have near-zero predictive power for AF3 outcomes.

9 Post-Submission Analysis

After submitting 81 designs to the ADAPTYV competition, the organizers independently re-scored all designs with their own Boltz-2 setup. Of 81 submitted, 33 survived (ipSAE ≥ 0.6) and 31 failed (ipSAE < 0.3).

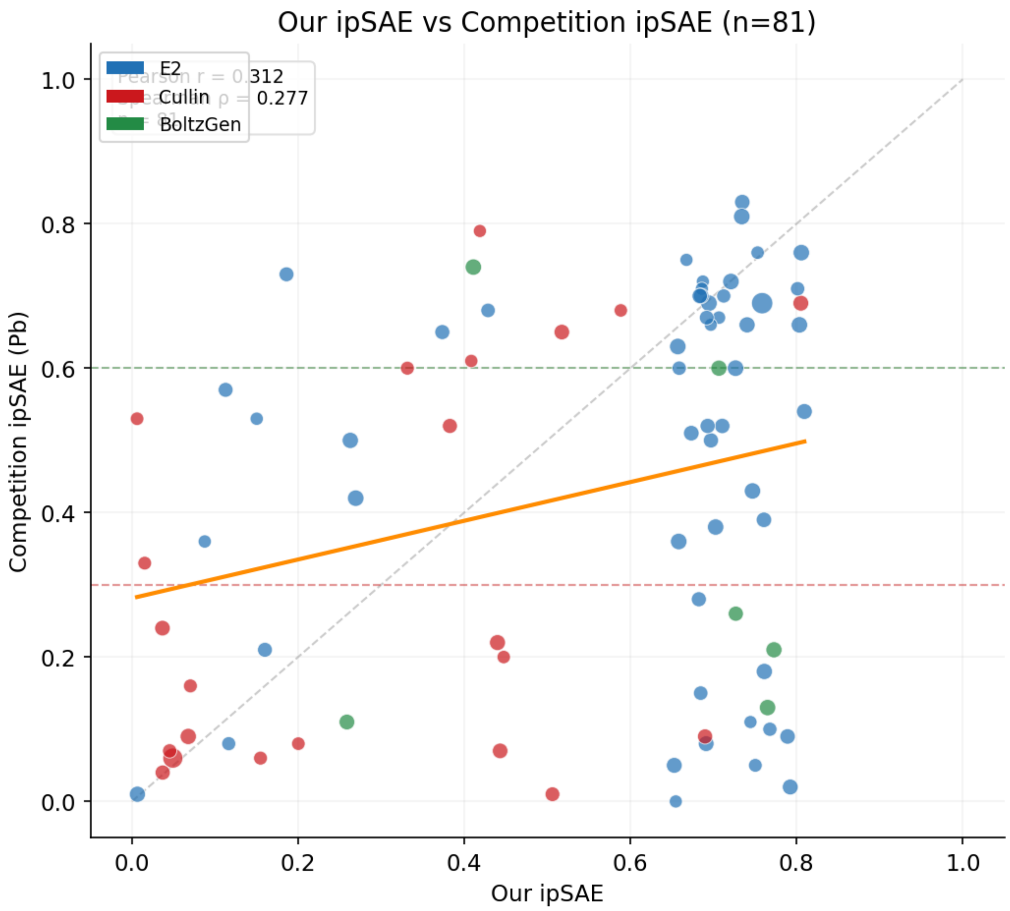

Figure 15: Our ipSAE vs. competition ipSAE. Notable reversals: 7 designs we scored below 0.5 survived in competition (≥ 0.6), while 9 designs we scored above 0.7 failed (< 0.3).

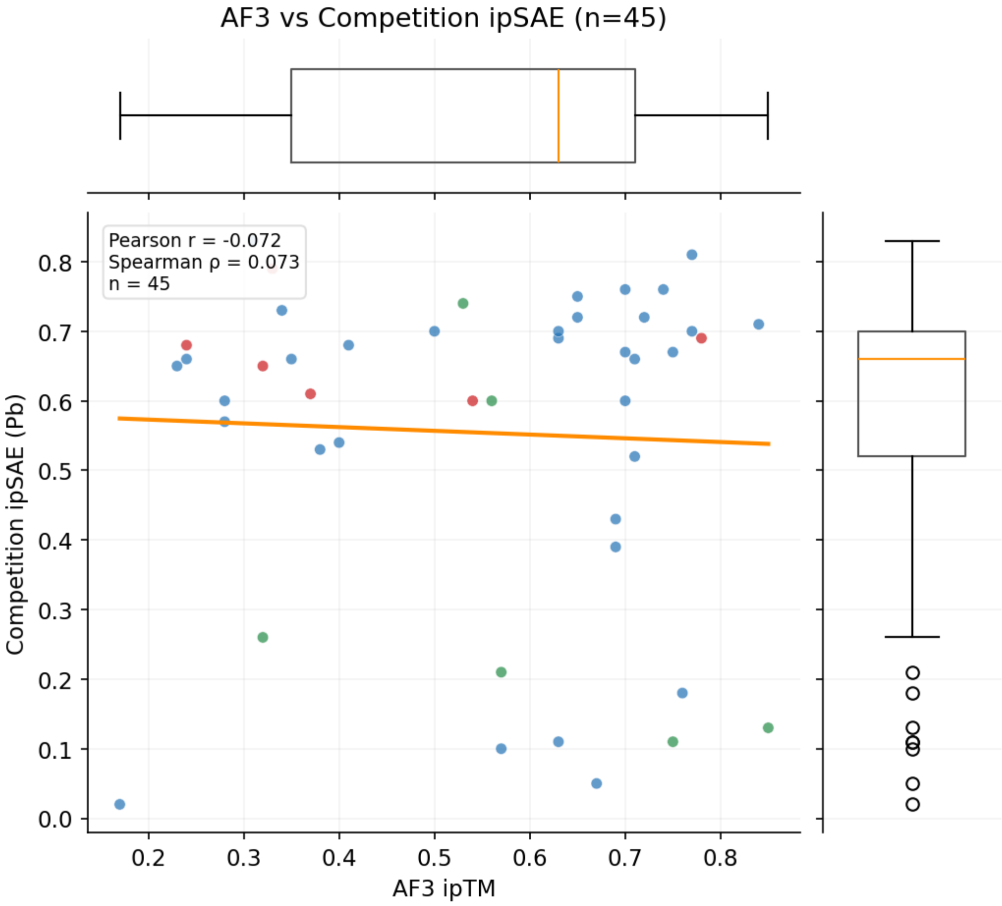

Figure 16: AF3 ipTM vs. competition Pb ipSAE for designs with both scores.

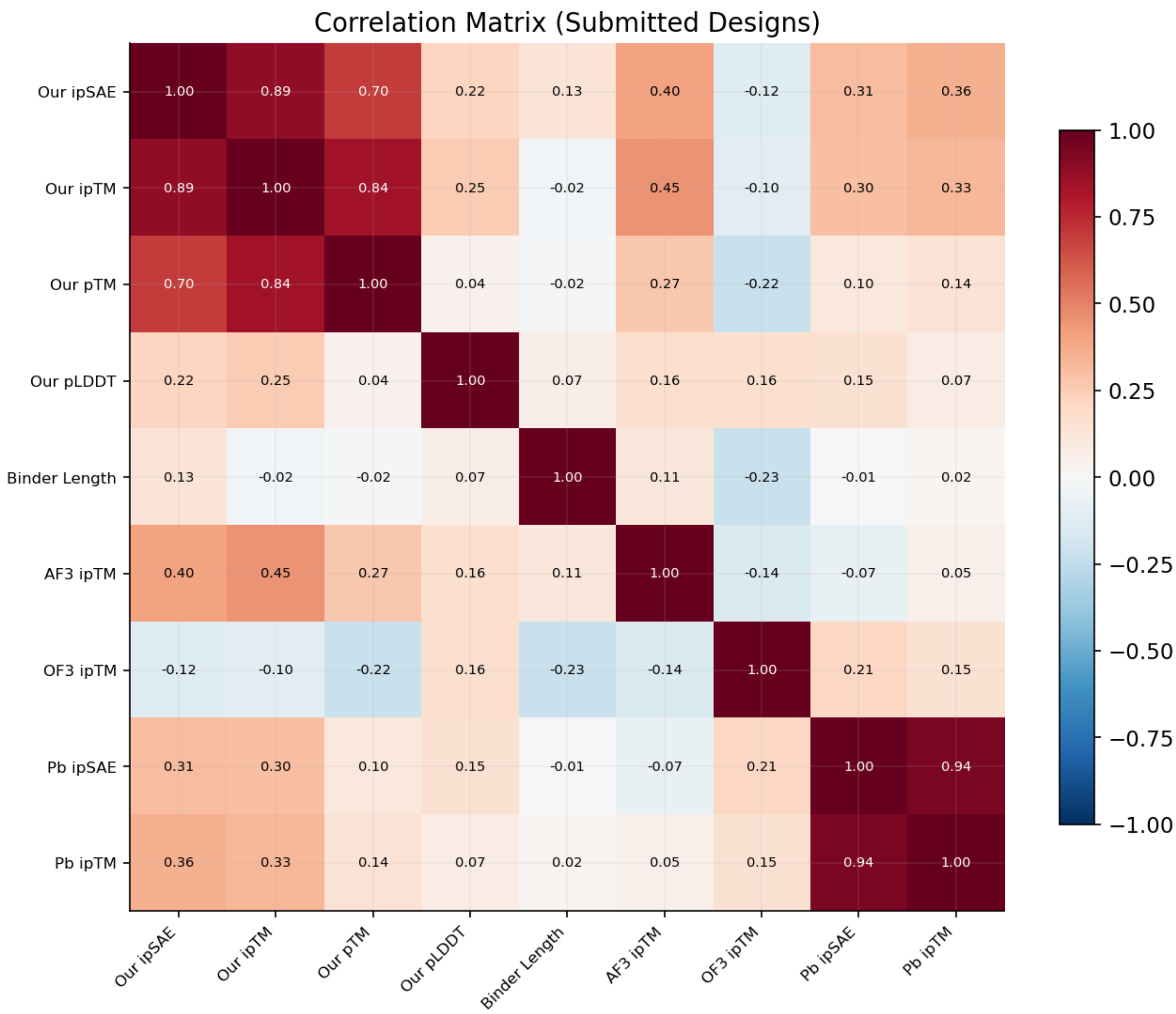

Figure 17: Full correlation matrix of our features vs. competition metrics. Competition metrics cluster separately from our pipeline metrics. Charge features show moderate associations with competition outcomes.

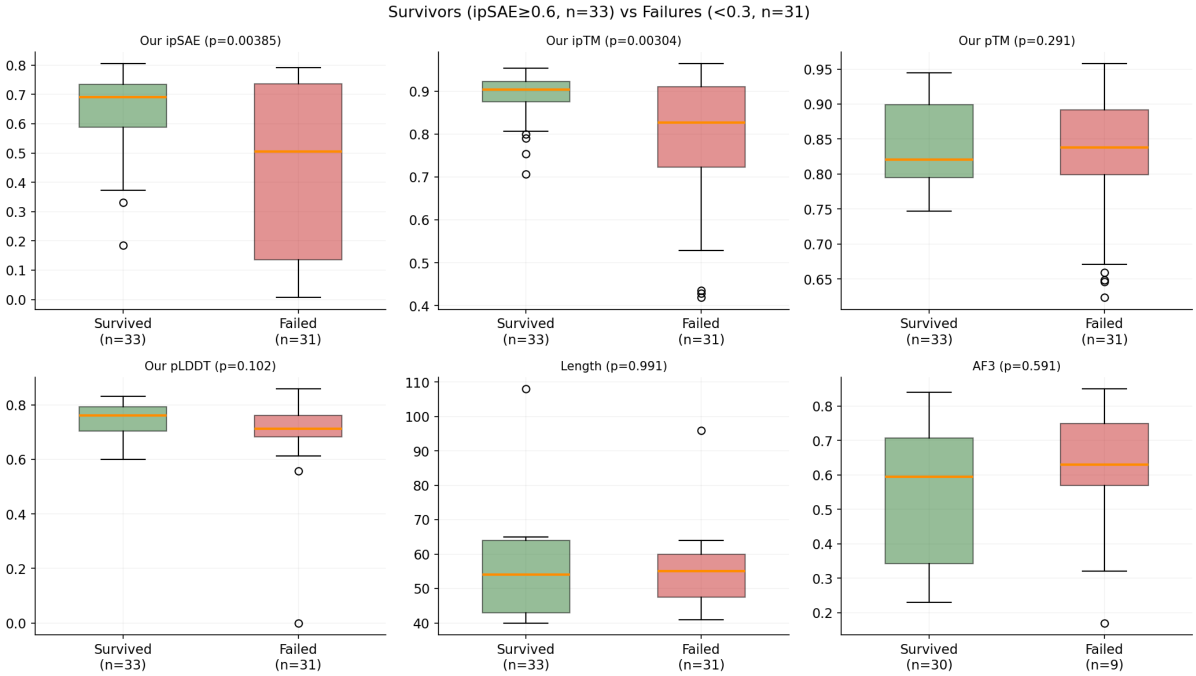

Figure 18: Feature comparison between competition survivors (≥ 0.6) and failures (< 0.3).

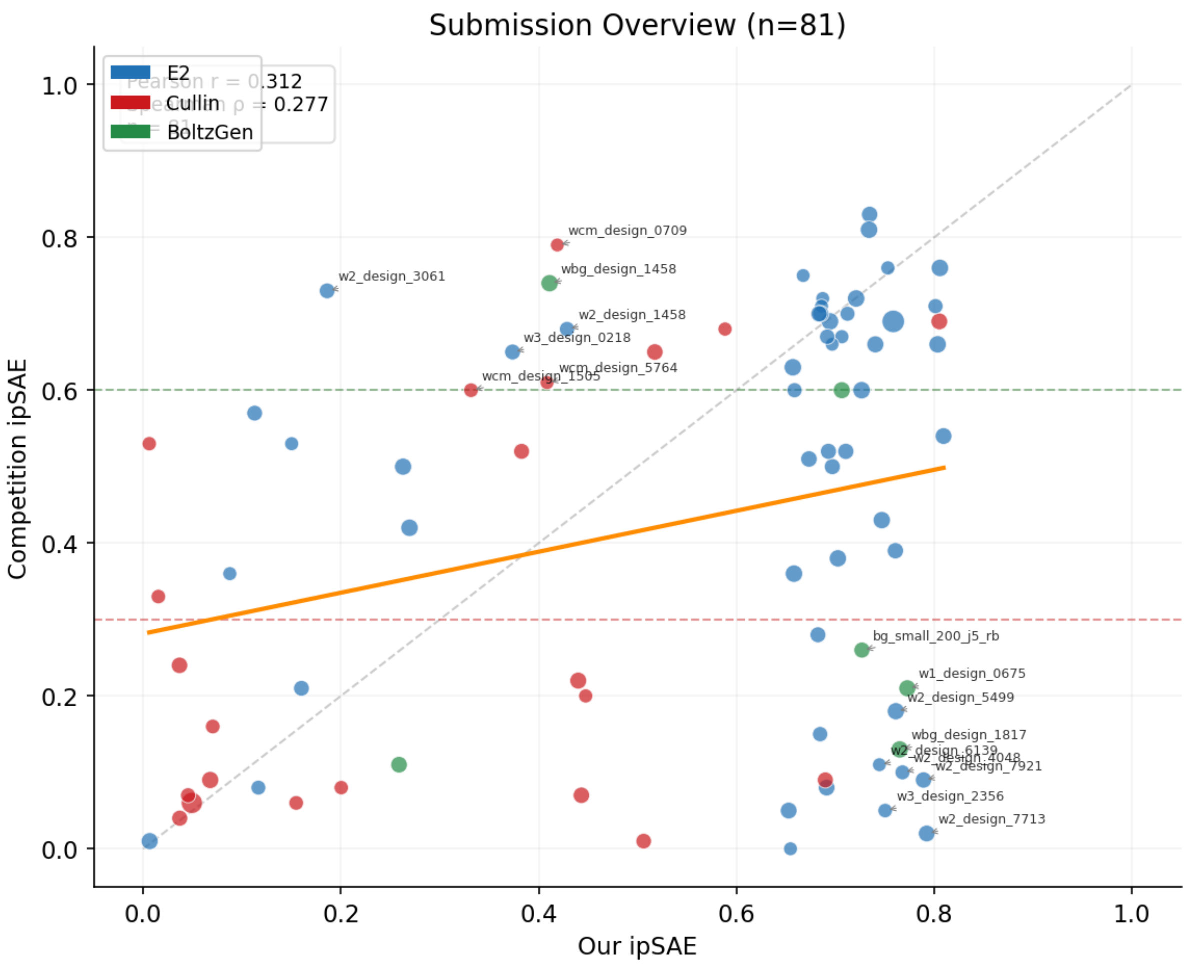

Figure 19: Overview of submission outcomes across all 81 designs.

9.1 Predictive Modeling of Competition Scores

Three regression models were trained to predict competition ipSAE from 64 sequence/structural features. All showed **negative cross-validated** R^2, confirming that competition scores are inherently unpredictable from our pipeline features:

Table 4: Cross-validated R^2 for competition ipSAE prediction.

Model	**CV** R^2
Ridge Regression	-2.22
ElasticNet	-0.69
Random Forest	-0.33

The top discriminating features between survivors and failures (Mann–Whitney U test):

Table 5: Top features discriminating competition survivors vs. failures.

Feature	p**-value**	**Importance (RF)**
Net charge	3.5×10^{-5}	0.105
Isoelectric point	3.4×10^{-4}	0.074
Our ipTM	3.0×10^{-3}	—
Charge density	3.2×10^{-3}	0.066
Our ipSAE	3.9×10^{-3}	0.041
Serine content	—	0.083

Net charge is the single strongest predictor of competition survival. The competition likely uses a fine-tuned or differently-configured Boltz-2 that is sensitive to electrostatic properties not captured by our scoring setup.

10 Evolutionary Analysis

ESM-2 deep mutational scanning (DMS) was performed on 2,160 single-point mutations across the 108-residue RBX1 target.

Key findings:

- Conservation (from 165-sequence MSA) correlates with ESM-2 DMS sensitivity (Spearman $r = 0.58$).
- Zinc-coordinating cysteines and histidines dominate both conservation and DMS sensitivity profiles.
- DMS-guided hotspot selection (E2 Enhanced) produced the highest AF3 hit rate among RFdiffusion campaigns.

11 Conclusions and Future Directions

11.1 Key Conclusions

1. **Boltz-2 scoring is unreliable as a filter.** ipSAE and ipTM show near-zero correlation with AF3 ground truth ($r \leq 0.08$). E2-face designs show systematically inflated Boltz-2 scores.

2. **AlphaFold3 is the reliable ground truth.** AF3 ipTM is the most trustworthy metric; OF3 shows moderate agreement ($r = 0.57$).
3. **Sequence diversity predicts AF3 success.** Shannon entropy, aromatic content, and Asp fraction are positive predictors; low-complexity poly-Ala/Glu sequences fail.
4. **Competition scoring is unpredictable.** The ADAPTYV Boltz-2 re-scoring produces results that are only weakly correlated with our scores ($r = 0.31$) and show dramatic reversals. Net charge is the strongest (but still weak) predictor.
5. **BoltzGen is a viable alternative.** Despite generating simpler sequences, BoltzGen designs validate at a higher rate in AF3.

11.2 Recommendations for Future Campaigns

- Use AF3 (or OF3 as a proxy) for all final design selection — do not rely on Boltz-2 scores alone.
- Optimize for sequence diversity: penalize low-entropy, poly-charged sequences.
- Incorporate net charge and isoelectric point as design filters if targeting Boltz-2-based competitions.
- Expand BoltzGen sampling, which shows promise for generating AF3-validated binders.
- Explore ensemble scoring (Boltz-2 + OF3 + AF3) for more robust design ranking.

A Design Naming Convention

Design identifiers follow the pattern `<wave>_design_<number>`:

- `w1`: Wave 1 — Cullin-face RFdiffusion designs.
- `w2`: Wave 2 — E2-face RFdiffusion designs (primary campaign).
- `w3`: Wave 3 — E2-face RFdiffusion designs (enhanced hotspots).
- `wcm`: Cullin-face mass generation wave.
- `wbg`: BoltzGen-generated designs.
- `bg_small_*`: Early BoltzGen prototypes.

B Top 100 Design Reference

The full top 100 candidate sequences (ranked by combined OF3/AF3 metrics) are available for download:

- FASTA: `data/rbx1_top100.fasta`
- CSV (all metrics): `data/rbx1_top100.csv`
- Interactive dashboard: https://sermare.github.io/protein-design.html