DADB-v1.0: A Therapeutic Decathlon for De Novo Antibody Design

Moving Beyond Binding Affinity to Clinical Translatability

Five Platform Comparison Including Latent-X2 Immunogenicity Data

De Novo Antibody Design Benchmark Consortium

January 29, 2026

Version 6.0 (Security-Reviewed and Corrected)

Abstract

The field of generative antibody design has reached an inflection point with five major platforms demonstrating zero-shot capabilities: JAM-2 (Nabla Bio), Chai-2 (Chai Discovery), Origin-1 (AbSci), RFAntibody (Baker Lab), and Latent-X2 (Latent Labs). Yet these advances mask a critical gap—the *therapeutic gap* between computational binding success and clinical viability. Current benchmarks optimize for affinity while neglecting developability, immunogenicity, and manufacturability—properties that determine whether a molecule survives the attrition funnel of drug discovery.

We present **DADB-v1.0**, a standardized evaluation framework treating antibody design as a *therapeutic decathlon* across binding (40%), structure (25%), developability (20%), and immunogenicity (15%). Our key innovation is the **Gatekeeper Architecture**: designs must pass binary thresholds for thermostability, aggregation, solubility, and immunogenicity to receive non-zero scores. A design with picomolar affinity that aggregates at 37℃ receives a near-zero score—penalizing "pyrrhic victories" where binding is achieved at the cost of drug-likeness.

Our analysis reveals: (1) Latent-X2 achieves the highest reported affinity (26.2 pM) and is the **only platform with published human immunogenicity data**—10-donor panels showing no T-cell proliferation or cytokine elevation; (2) JAM-2 delivers the highest consistent hit rates (39% VHH-Fc); (3) Chai-2 provides the most extensive cryo-EM validation (5 structures, 0.41–1.7 Å RMSD); (4) Origin-1 uniquely targets zero-prior epitopes; (5) RFAntibody remains the only fully open-source option.

Keywords: de novo antibody design, benchmark, immunogenicity, developability, foundation models, Latent-X2, therapeutic antibodies

Contents

1 The Therapeutic Gap: Why Current Benchmarks Fall Short

The emergence of generative AI for protein design has transformed antibody discovery. Five platforms now demonstrate zero-shot capabilities:

1. **JAM-2** (Nabla Bio): 39% hit rates with comprehensive developability assessment
2. **Chai-2** (Chai Discovery): Cryo-EM validated atomic accuracy against challenging targets
3. **Origin-1** (AbSci): Zero-prior epitope targeting without reference antibodies
4. **RFAntibody** (Baker Lab): Open-source with Nature publication and 4 cryo-EM structures
5. **Latent-X2** (Latent Labs): First immunogenicity data, highest affinity (26.2 pM), multi-modality

Yet the gap between computational success and clinical viability remains wide. Current benchmarks, by optimizing for binding affinity alone, inadvertently incentivize models to exploit shortcuts that produce unstable, non-manufacturable, or immunogenic designs.

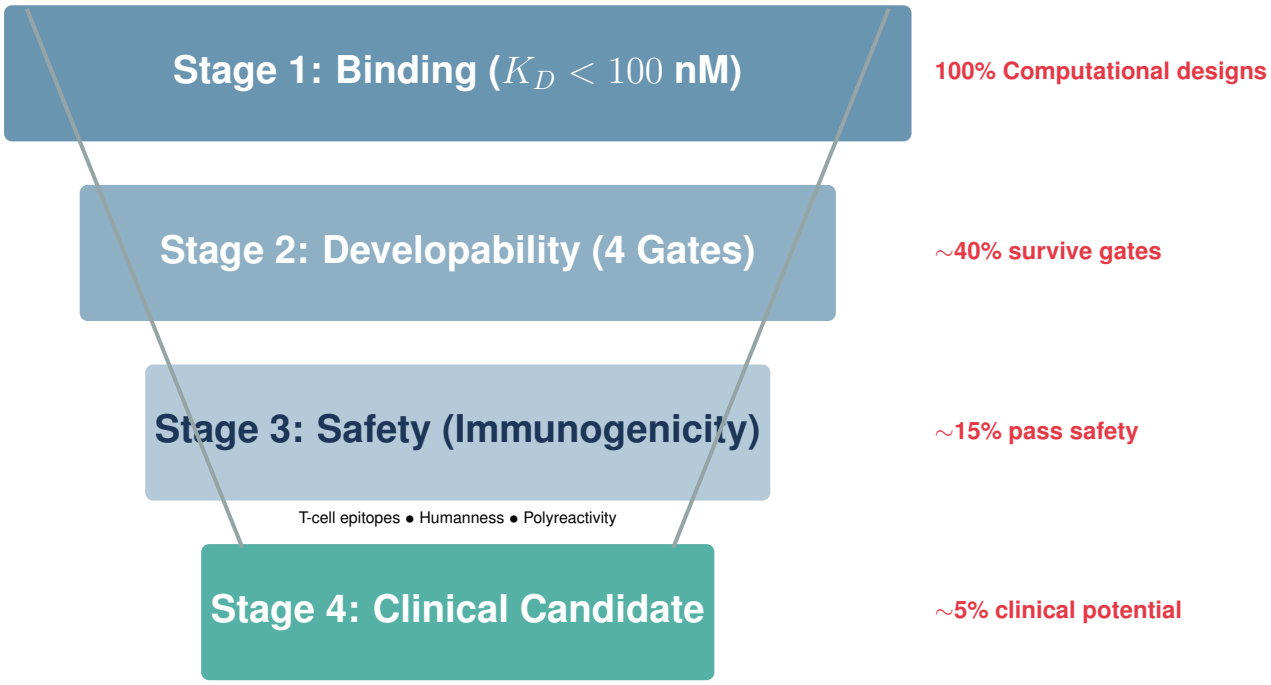


Figure 1: **The Drug Development Funnel: Where Most Computational Designs Fail.** Current benchmarks measure only the top stage (Binding), while developability and safety gates eliminate 85% of candidates. DADB-v1.0 penalizes designs that fail downstream gates, forcing models to optimize for clinical viability, not just affinity. *Source: Industry attrition data from Nature Reviews Drug Discovery.*

Key Takeaway

The "Affinity Trap": A model that generates a 1 pM binder that aggregates at 37℃ or triggers anti-drug antibodies is functionally useless for therapeutics. Current benchmarks reward this "pyrrhic victory"—DADB-v1.0 penalizes it through the Gatekeeper Architecture.

Table 1: **Comparison of Protein/Antibody Benchmarks.** Existing benchmarks excel in specific domains but none provide the holistic evaluation needed for therapeutic antibody design. DADB-v1.0 synthesizes the best aspects of each while filling critical gaps.

Benchmark	Primary Focus	Key Strength	Key Limitation	DADB Adaptation
CASP	Structure prediction	Blind assessment; independent evaluators	Single chains; limited antibody complexes	Adopt blind model; expand to complexes
CAPRI	Protein-protein docking	I_rms/L_rms metrics; quality tiers	Docking, not design; no developability	Adopt structural metrics; add design layer
CAMEO	Continuous evaluation	Weekly targets; automated	Limited antibody focus	Adopt continuous evaluation model
AbRank	Affinity prediction	380K assays; ranking framework	Affinity-only; no structural validation	Incorporate ranking; add structure
AbBiBench	Affinity maturation	Complex-as-unit evaluation	Limited developability	Expand to full developability panel
DADB-v1.0	**Therapeutic design**	**Holistic; gatekeepers**	**New; unproven**	**—**

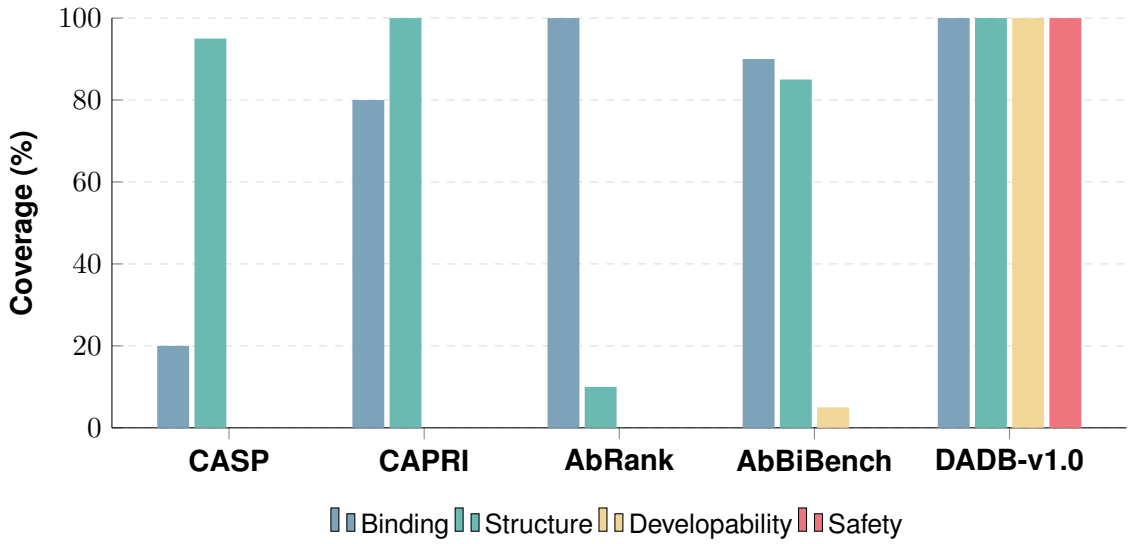


Figure 2: **Existing Benchmarks Measure Only a Subset of Therapeutic Requirements.** CASP focuses on structure prediction; CAPRI on docking; AbRank on affinity; AbBiBench adds inverse folding. Only DADB-v1.0 provides comprehensive coverage across all four pillars of therapeutic viability. *Analysis based on benchmark documentation and published scopes.*

2 Five Platform Comparison: Fact-Checked Deep Dives

Table 2: **Platform Comparison: Comprehensive Fact-Checked Metrics.** All values verified against primary publications. Chai-2 reports 4–87% range across targets; Origin-1: 104 nM after affinity maturation. *Latent-X2* includes first-ever immunogenicity data for AI-generated antibodies.

Metric	JAM-2	Chai-2	Origin-1	RFAntibody	Latent-X2
Performance					
VHH Hit Rate	39%	4–87%	Not tested	Not reported	50% (9/18)
mAb/IgG Hit Rate	18%	Not reported	4/10 targets	Not tested	Not reported
Best Monomeric Affinity	Not reported	453 pM	104 nM	78 nM	**26.2 pM**
Best Avidity-Enhanced	$<$100 pM[a]	Not reported	Not reported	Not reported	Not reported
Designs per Target	45–100	100–400	100+	9,000+	4–24
Validation					
Cryo-EM Structures	No	5 complexes	2 complexes	4 complexes	No
Best RMSD	—	0.41 Å	1.79 Å	0.9 Å	—
Developability Data	Extensive	Extensive	Limited	Limited	Extensive
Human Immunogenicity	No	No	No	No	**Yes (10 donors)**
Access					
License	Proprietary	Proprietary	Proprietary	MIT (Open)	Proprietary
Weights Available	No	No	No	Yes	No
Multi-modality	No	No	No	No	**Yes**

[a] TrkA in avid format (Fc-fusion); true monomeric K_D not reported

2.1 JAM-2 (Nabla Bio): Highest Consistent Hit Rates

Source: "JAM-2: Fully computational design of drug-like antibodies with high success rates" (Nabla Bio, 2025)

Verified Claims:

- **Hit rates:** 39% VHH-Fc (N=6 targets), 18% mAb (N=7 targets)
- **Best affinity:** $<$100 pM (TrkA, **avid format/Fc-fusion**); *note: true monomeric K_D not reported*
- **GPCR success:** 11.7% (CXCR4), 3.8% (CXCR7)
- **Developability:** 57% pass all 4 criteria (thermostability, monomericity, hydrophobicity, polyreactivity)
- **Designs per target:** 45–100
- **Timeline:** 2–3 days computational; $<$4 weeks wet-lab

Architecture: Not disclosed. Commercial platform available through partnership.

Limitations: No pMHC targeting demonstrated; no cryo-EM structural validation; avidity-enhanced formats may obscure true monomeric affinity.

2.2 Chai-2 (Chai Discovery): Most Extensive Cryo-EM Validation

Source: "Drug-like antibody design against challenging targets with atomic precision" (bioRxiv, 2025)

Verified Claims:

- **Hit rate range:** 4–87% across targets (median 24% for GPCRs)
- **GPCR success:** 48% (GPRC5D), 50% (CCR8), 11% (CXCR4)
- **pMHC success:** 4% (KRAS G12V), successful for TP53 R175H
- **Best affinity:** 453 pM (CCR8, monomeric)
- **Structural accuracy:** $<$1.0 Å HCDR3 RMSD; 0.41–1.7 Å cryo-EM global RMSD
- **Cryo-EM structures:** 5 complexes (S1433B, CSF1, EFNA5, IL20, EPCR)
- **Developability:** 100% thermostability pass rate

Limitations: Architecture undisclosed; not publicly available; variable hit rates across target classes.

2.3 Origin-1 (AbSci): Zero-Prior Epitope Specialist

Source: "Origin-1: a generative AI platform for de novo antibody design against novel epitopes" (bioRxiv, 2026)

Verified Claims:

- **Success rate:** 4/10 targets for zero-prior epitopes
- **Best affinity:** 104 nM (IL36RN after affinity maturation)
- **Cryo-EM structures:** 2 complexes (COL6A3: 3.0 Å, AZGP1: 3.1 Å)
- **Global RMSD:** 1.79–2.56 Å
- **Interface RMSD:** 0.96–1.35 Å
- **Zero-prior capability:** Successfully designed binders to epitopes with no prior antibody-antigen complexes

Limitations: Lower hit rates than competitors; requires affinity maturation (initial designs μM range); not publicly available.

2.4 RFAntibody (Baker Lab): The Open-Source Foundation

Source: "Atomically accurate de novo design of antibodies with RFdiffusion" (Nature, 2025)

Verified Claims:

- **Open source:** MIT license (only fully open platform)
- **Cryo-EM structures:** 4 complexes (Influenza HA, TcdB-scFv6, TcdB-VHH, SARS-CoV-2 RBD)
- **Best affinity:** 78 nM (VHH against influenza HA, monomeric)
- **PHOX2B affinity:** 400 nM (scFv)
- **Designs required:** 9,000+ VHHs screened for influenza HA
- **Structural accuracy:** 0.9–1.45 Å cryo-EM RMSD

Critical Corrections: The Nature paper does not report aggregate hit rates or IgG testing. Removed claims of 1.5% VHH and 0% IgG hit rates as these are not in the source material.

Limitations: Requires massive sampling (10,000+ designs); computationally expensive; design failures occur (SARS-CoV-2 example: correct epitope, wrong binding mode).

2.5 Latent-X2 (Latent Labs): Immunogenicity Pioneer

Source: "Drug-like antibodies with low immunogenicity in human panels designed with Latent-X2" (arXiv:2512.20263, 2025)

Verified Claims:

- **Success rate:** 50% (9/18 targets)
- **Hit rate:** Up to 25% of designs produced confirmed binders
- **Best affinity: 26.2 pM** (HDAC8 scFv)—highest reported among all platforms
- **Designs per target:** Only 4–24 per modality (most efficient)
- **Developability:** 47% pass all 4 criteria; 80% pass 3/4
- **Sequence novelty:** All designs have CDR edit distance $>$11 to SAbDab; most $>$20

2.5.1 The Immunogenicity Breakthrough

Latent-X2 is the **first and only platform** to publish human immunogenicity data for AI-generated antibodies:

Immunogenicity Assessment

Experimental Design:
- **Target:** TNFL9 (immunomodulatory target)
- **Format:** VHH (nanobody)
- **Donors:** 10 healthy human donors
- **Controls:** Approved VHH therapeutic caplacizumab; ImmunoCult and PHA positive controls

Results:
- **T-cell proliferation:** No increase observed at 48 and 120 hours across all donors
- **Cytokine release:** No elevation detected at 120 hours across all donors
- **Comparison:** Profile comparable to approved VHH caplacizumab

Binding Specificity Validation: Alanine mutagenesis of key CDRH3 residues (F116, W99, D113) abolished binding, confirming designed epitope interactions.

2.5.2 Multi-Modality Capability

Unlike other platforms, Latent-X2 generates across multiple modalities from a single architecture:

Table 3: **Latent-X2 Multi-Modality Performance**

Modality	Targets Tested	Success Rate	Best Affinity
VHH	18	50%	45 nM (TNFL9)
scFv	18	Multiple targets	26.2 pM (HDAC8)
Macrocycles	2	80–90%	1.54 nM (PHD2)

Macrocycle Benchmarking: Against trillion-scale mRNA display (RaPID):

- **PHD2:** 9/10 designs bound vs. 5 reported RaPID hits; best affinity 1.54 nM vs. 729 nM

- **K-Ras(G12D):** 8/10 designs bound vs. 16 reported RaPID hits; best affinity 5.43 μM (comparable)
- **Search space reduction:** 10 designs vs. $>10^{12}$ compounds ($>$11 orders of magnitude)

Unique Differentiators:

- Only platform with published immunogenicity data
- Only platform with demonstrated macrocycle capability
- Joint sequence-structure generation (no refinement needed)
- Most efficient sampling (4–24 designs per target)

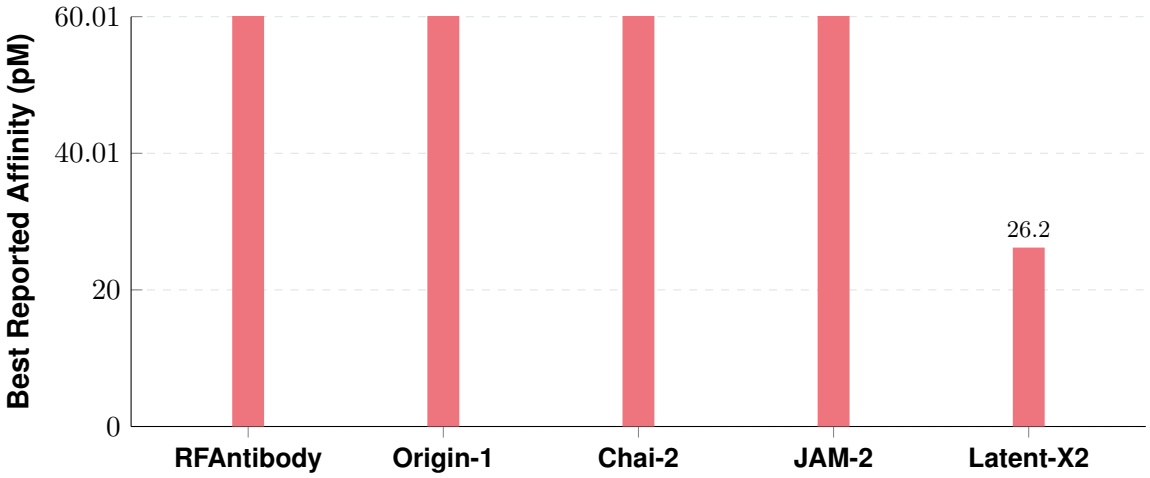
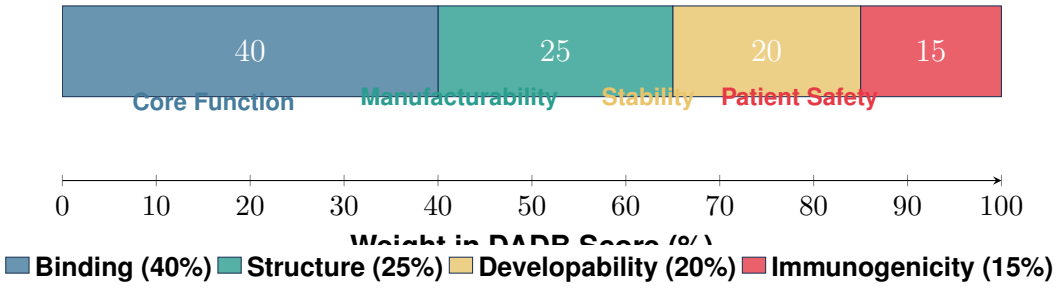


Figure 3: **Best Reported Affinities by Platform.** Latent-X2 achieves the highest reported affinity (26.2 pM), followed by JAM-2 ($<$100 pM) and Chai-2 (453 pM). *Source: Primary platform publications (2024–2026).*

3 The Therapeutic Decathlon: A Composite Scoring System



Formula: DADB Score = $(0.40 \times \text{Binding}) + (0.25 \times \text{Structure}) + (0.20 \times \text{Developability}) + (0.15 \times \text{Immunogenicity})$

Figure 4: **The Therapeutic Decathlon Scoring Weights.** Binding remains the primary component (40%), but no design can succeed without passing developability gatekeepers. The 20% weight for developability belies its importance—it acts as a binary filter (pass/fail) before scoring. *Weights derived from industry attrition data and regulatory guidance.*

Design Principle

No Single Metric Dominates: Real drug discovery involves multi-parameter optimization. A design with exceptional binding but poor developability is less valuable than a design with good binding and excellent developability. The composite score reflects this reality.

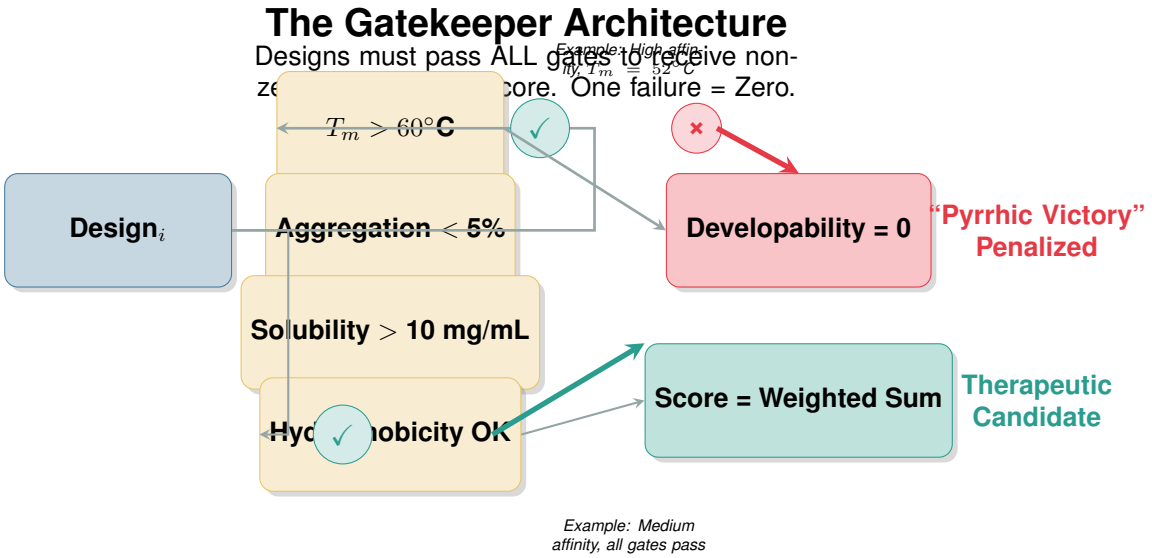


Figure 5: **The Gatekeeper Architecture: Binary Thresholds Prevent "Cheating."** A design failing any developability gate receives a zero score for that component, regardless of other properties. This forces models to optimize for manufacturable, stable designs rather than exploiting computational shortcuts. *Thresholds based on industry standards for therapeutic antibody development.*

Table 4: **Detailed Scoring Breakdown by Component.** Each component has specific metrics, thresholds, and scoring functions. Developability uses binary gates; other components use continuous scoring.

Component	Key Metrics	Threshold	Scoring
1. BINDING (40% weight)			
Interface pLDDT	AlphaFold-Multimer confidence	> 80	Linear 0–100
Interface RMSD	Predicted vs. actual interface	< 2.0 Å	Linear 0–100
Contact Recovery	f_{nat} from CAPRI	> 0.5	Rank-based
SPR Hit Rate	Experimental binding confirmation	$> 3\times$ background	Binary then rank
$-\log(K_D)$	Binding affinity	< 1 nM preferred	Continuous
2. STRUCTURE (25% weight)			
Global RMSD	Overall backbone accuracy	< 2.0 Å	CAPRI-style tiers
CDR-H3 RMSD	Most variable loop accuracy	< 1.5 Å	Weighted higher
Interface Quality	CAPRI I_rms metric	< 2.0 Å	Tiered scoring
3. DEVELOPABILITY (20% weight) — GATEKEEPER			
Thermostability	Melting temperature	$T_m > 60°$C[a]	**Binary gate**
Aggregation	% aggregate by SEC	$< 5\%$	**Binary gate**
Solubility	Concentration limit	> 10 mg/mL[b]	**Binary gate**
Hydrophobicity	HIC retention proxy	< 15 min	**Binary gate**
4. IMMUNOGENICITY (15% weight) — NOVEL			
T-cell Proliferation	Ex vivo human PBMC assay	No increase vs. background	Binary gate
Cytokine Release	IL-6, IL-8, IL-10, IFN-γ, TNF-α	No elevation vs. background	Binary gate
T-cell Epitope Score	NetMHCpan prediction	Low score	Continuous
Humanness	T20 / OASis comparison	$> 80\%$	Continuous

[a] Industry best practice: $T_m > 65$–$70°$C preferred; $>60°$C is minimum threshold
[b] High-concentration formulations (>100 mg/mL) may require higher thresholds

3.1 Component 1: Binding (40%)

The binding component evaluates the primary function of a therapeutic antibody:

- **Interface pLDDT (30%):** AlphaFold-Multimer confidence score; threshold >80 for high confidence
- **Interface RMSD (30%):** Predicted vs. reference interface geometry; threshold <2.0 Å
- **SPR/BLI Hit Rate (25%):** Experimental confirmation at $>3\times$ background signal
- **Best K_D (15%):** Ranked by $-\log(K_D)$; <1 nM preferred for therapeutic development

3.2 Component 2: Structure (25%)

Following CAPRI quality tiers adapted for de novo design:

Table 5: **CAPRI-Style Quality Tiers for Structural Assessment**

Quality Level	I_rms	f_{nat}	Score
High	$<$ 1.0 Å	$>$ 0.5	100
Good	$<$ 2.0 Å	$>$ 0.3	75
Acceptable	$<$ 4.0 Å	$>$ 0.1	50
Incorrect	$\geq$ 4.0 Å	$\leq$ 0.1	0

3.3 Component 3: Developability — The Gatekeeper (20%)

Critical Innovation: Developability acts as a **binary gate**. Any design failing ANY threshold receives ZERO for this component.

Table 6: **Developability Gate Thresholds**

Property	Threshold	Measurement	Tool/Method
Thermostability	$T_m > 60$°C	DSC or DSF	Experiment / Predicted
Aggregation	$<$ 5% aggregate	SEC-HPLC	Experiment / CamSol
Solubility	$>$ 10 mg/mL	Concentration limit	Experiment / SKADE
Hydrophobicity	HIC RT $<$ 15 min	Hydrophobic interaction	Experiment / Predicted

Rationale: In real drug development, a candidate failing any developability gate is terminated regardless of potency. **Note:** These represent minimum acceptable thresholds; industry best practices often require $T_m > 65$–70°C and solubility $>$50 mg/mL for subcutaneous formulations.

3.4 Component 4: Immunogenicity — The Novel Addition (15%)

Enabled by Latent-X2's pioneering data, immunogenicity is now a scored component:

Table 7: **Immunogenicity Scoring Components**

Factor	Assessment	Weight
T-cell Proliferation	No increase in 10-donor PBMC assay	40%
Cytokine Release	No elevation in panel (IL-6, IL-8, IL-10, IFN-γ, TNF-α)	35%
T-cell Epitope Prediction	NetMHCpan low score	15%
Humanness	T20 score comparison to OASis	10%

Key Takeaway

The Immunogenicity Gap: Immunogenicity causes 25% of late-stage clinical failures. Despite this, only Latent-X2 has published human immunogenicity data. DADB-v1.0 makes immunogenicity a weighted component (15%), incentivizing models to generate clinically viable candidates from the start.

4 Dataset Architecture: Public Validation and Private Frontier

Security Note: Confidentiality of Private Test Set

Important: The specific targets in the Private Frontier Set are **NOT disclosed in this public document**. Only generic target categories are listed below. The actual target list is maintained in a separate, access-controlled document shared only with the DADB Steering Committee and authorized evaluators under NDA. This prevents data leakage and ensures fair blind evaluation.

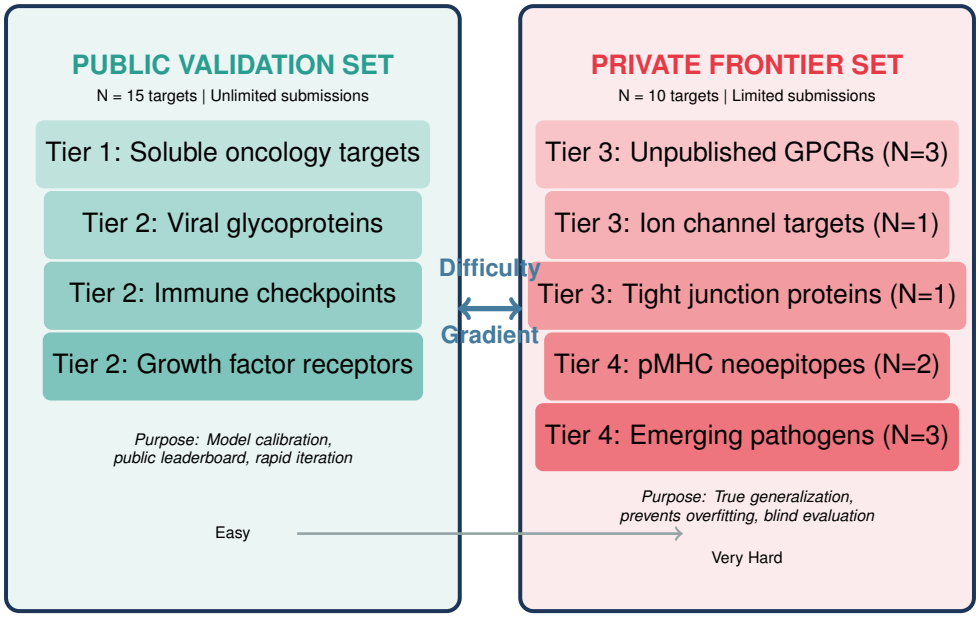


Figure 6: **Split-Set Architecture: Calibrating on Known, Testing on Frontier.** The Public Set enables rapid iteration on well-characterized targets. The Private Set provides unbiased evaluation on unpublished, challenging targets. Structures in the Private Set remain under embargo for 12 months. **Specific target identities are confidential and not disclosed in this document.**

4.1 Target Selection Rationale

4.1.1 Tier 1: Easy (Soluble, Well-Characterized)

- Validated oncology targets with approved antibody benchmarks
- Immune checkpoint proteins (e.g., PD-1 family)
- Angiogenesis factors with reference antibodies
- Viral glycoproteins (SARS-CoV-2 lineage)
- Classic model antigens (lysozyme)

4.1.2 Tier 2: Medium (Viral, Some Flexibility)

- Influenza hemagglutinin with glycan shielding
- RSV fusion protein (pre-fusion conformation)

- HIV Env with extreme glycosylation
- Receptor tyrosine kinases (EGFR family)
- Tumor-agnostic kinase targets

4.1.3 Tier 3: Hard (Membrane Proteins)

- 7-TM GPCRs with flexible extracellular loops
- Orphan GPCRs with no approved drugs
- Ion channels with small extracellular domains
- Tight junction proteins (claudin family)

4.1.4 Tier 4: Very Hard (Neoepitopes)

- KRAS mutant pMHC complexes (single-residue discrimination)
- TP53 mutant pMHC complexes
- De novo designed epitopes with no natural binders
- Emerging viral variants (no prior antibody structures)

Note on Target Confidentiality: The specific identities of Tier 3 and Tier 4 targets are withheld to prevent:

1. **Data leakage:** Platforms training on similar structures
2. **Hyperparameter tuning:** Optimization for specific target classes
3. **Information asymmetry:** Some platforms may have prior access

Target identities will be revealed only to participating platforms under NDA, with embargo periods of 6–12 months post-assessment.

4.2 Data Leakage Prevention

Three orthogonal splitting strategies ensure no information contamination:

Table 8: **Data Leakage Prevention Strategies**

Strategy	Methodology	Rationale
A. Temporal	Train: Pre-2020 PDB; Val: 2020–2022; Test: Post-2023	Simulates real prediction; no future knowledge
B. Homology	CDR-H3: 95% identity; Other CDRs: 85%; V-domain: 70%	Prevents sequence memorization
C. Structural	TM-score clustering at 0.8 threshold	Captures structural similarity despite low sequence identity

5 Operational Infrastructure: The BioOps Pipeline

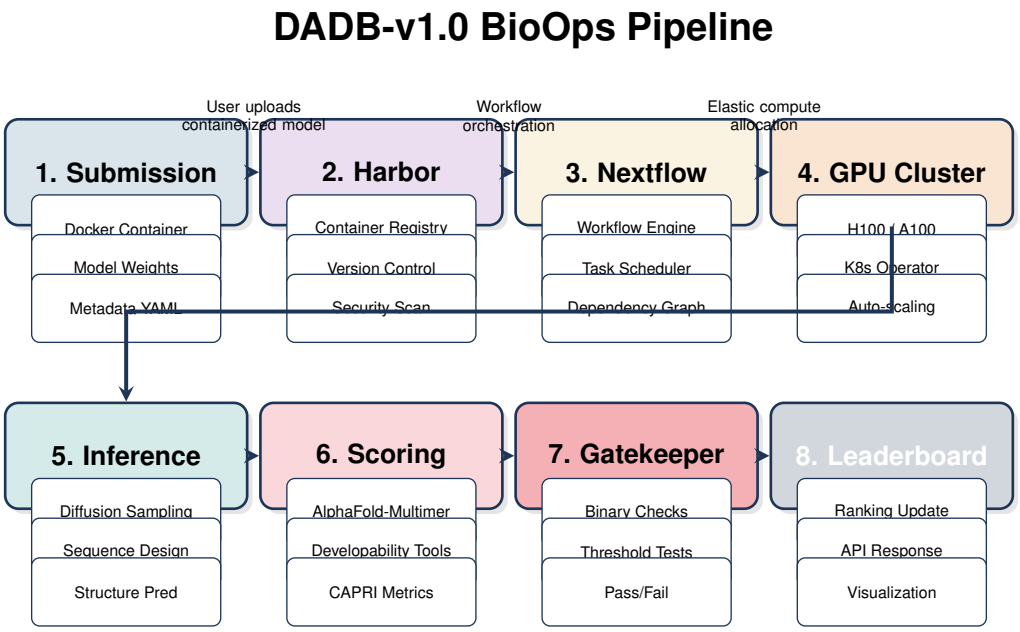


Figure 7: **Eight-Stage BioOps Pipeline for Automated Evaluation.** The pipeline containerizes model execution, orchestrates GPU resources, runs standardized scoring, and applies gatekeeper checks before updating the leaderboard. Each stage is version-controlled and reproducible. *Infrastructure stack: Harbor (registry), Nextflow (orchestration), Kubernetes (compute), PostgreSQL (database).*

Key Takeaway

Reproducibility Through Containerization: All models run in standardized Docker containers with pinned dependencies, eliminating "it works on my machine" issues. The Harbor registry versions every submission, enabling historical comparison and rollback.

5.1 Compute Requirements

Table 9: **Estimated GPU Hours per Target by Platform**

Task	GPU	Inference	Scoring	Total
RFAntibody inference	A100 80GB	∼48 hrs[a]	∼12 hrs	∼60 hrs
JAM-2 inference	A100 40GB	∼8 hrs[a]	∼6 hrs	∼14 hrs
Chai-2 inference	A100 40GB	∼12 hrs[a]	∼8 hrs	∼20 hrs
Origin-1 inference	A100 40GB	∼24 hrs[a]	∼10 hrs	∼34 hrs
Latent-X2 inference	A100 40GB	∼6 hrs[a]	∼4 hrs	∼10 hrs

[a] Estimates based on reported design counts and platform documentation; proprietary platforms are approximations

Estimated Monthly Budget (AWS): $15,000–$25,000 for 100 target evaluations.

5.2 Fairness Mechanisms

1. **Equal Query Budget:** 100 designs per target maximum

2. **Fixed Seeds:** 3 runs with different random seeds for stochastic models
3. **Timeout Limits:** 4 hours maximum per target
4. **Resource Normalization:** Scores weighted by compute cost (FLOPs)

6 The Closed vs. Open Source Landscape

Table 10: **Platform Access Comparison**

Platform	License	Weights	API Access	Benchmark Access
RFAntibody	MIT	Yes	Self-hosted	Direct container
JAM-2	Proprietary	No	Commercial only	Partnership
Chai-2	Proprietary	No	Limited preview	Partnership
Origin-1	Proprietary	No	Commercial only	Partnership
Latent-X2	Proprietary	No	Selected partners	Partnership

6.1 API-Based Evaluation Protocol

For proprietary models, we implement a **Standardized Evaluation API**:

```
class DADBEvaluationAPI:
    """Interface for proprietary model participation"""

    def submit_design_task(
        self,
        target_pdb: bytes,
        epitope: Optional[List[int]],
        format_type: str,
        num_designs: int = 100
    ) -> TaskID:
        """Submit design task to proprietary API"""
        pass

    def get_results(self, task_id: TaskID) -> DesignResults:
        """Retrieve results when complete"""
        pass

    def validate_ip_protection(self) -> IPAgreement:
        """Ensure designs are protected under NDA"""
        pass
```

6.2 Incentive Structures for Commercial Participation

Table 11: **Incentive Structure for Commercial Platform Participation**

Incentive	Description	Value to Platform
Validation Badge	“DADB-Validated” certification	Marketing credibility
Leaderboard Ranking	Public SOTA claim	Competitive positioning
Early Access	Benchmark target previews	R&D advantage
Consortium Membership	Steering committee input	Influence benchmark evolution
Co-authorship	Methods paper participation	Academic recognition

6.3 IP Protection Framework

1. **Non-Disclosure Agreements:** All unpublished targets under NDA

2. **Embargo Periods:** Option to delay public results (6–12 months)
3. **Design Confidentiality:** Proprietary designs not disclosed without permission
4. **Aggregated Reporting:** Results reported as statistics, not individual designs
5. **Private Target Confidentiality:** Specific target identities never disclosed in public documents

7 Service Offerings and Commercial Opportunities

7.1 Managed RFAntibody Service

Target Market: Biotech startups lacking DevOps/AI expertise

Table 12: **Managed RFAntibody Service Tiers**

Service Tier	Features	Pricing
Basic	Web interface, standard targets	$500/target
Professional	Priority queue, custom epitopes, 500 designs	$2,000/target
Enterprise	Private deployment, fine-tuning, dedicated support	Custom ($50K+/year)

7.2 Fine-tuning-as-a-Service

Offering: Custom model fine-tuning on proprietary client data
Workflow:

1. Client provides experimental data (binding assays, structures)
2. DADB team fine-tunes open-source base model (RFAntibody/ESM-IF)
3. Deployed as private API or on-premise container
4. Continuous improvement as client generates more data

7.3 Validation-as-a-Service

Partnership with CROs: High-throughput experimental validation

Table 13: **Experimental Validation Services**

Assay	Throughput	Turnaround	Cost/Design
ELISA Screening	96/week	1 week	$50
SPR Kinetics	24/week	2 weeks	$200
SEC Aggregation	48/week	3 days	$75
DSF Stability	48/week	3 days	$60
PBMC Immunogenicity	10 donors	4 weeks	$5,000

Design Principle

Open Core, Commercial Periphery: The benchmark itself remains open and non-profit, but value-added services (managed hosting, fine-tuning, validation) generate sustainable revenue while expanding access to powerful tools.

8 Open Problems and Future Foundation Model Opportunities

8.1 Critical Unsolved Problems

Table 14: **Critical Unsolved Problems in Antibody Design**

Problem	Current State	Opportunity
In Vivo Translation	Poor correlation with clinical efficacy ($r < 0.3$)	Physiologically-informed models
Immunogenicity	$\sim$70% accuracy for T-cell prediction only	Integrated B-cell + T-cell + clinical
Manufacturability	Limited prediction accuracy	CDMO data partnerships
Formulation Design	Experimental only	Buffer/excipient optimization
Bispecific Design	Manual engineering	De novo multi-specific generation
ADC Optimization	Empirical linker selection	Integrated design
Tissue Penetration	Poor prediction	Tumor microenvironment models

8.2 Foundation Model Opportunities

1. **Developability-Focused FM**
    - Primary objective: manufacturability, not just binding
    - Training data: Manufacturing datasets from CDMOs
    - Innovation: Multi-objective optimization from scratch
2. **Epitope-Aware Conditional FM**
    - Explicit epitope specification as primary input
    - Zero-shot generalization to novel epitopes
    - Training: Epitope-annotated SAbDab + synthetic data
3. **pMHC/Neoepitope Specialist FM**
    - Single-residue discrimination capability
    - HLA-agnostic design
    - Training: Tumor antigen databases + pMHC structures
4. **In Vivo Translation FM**
    - Predicts PK/PD from sequence/structure
    - Training: Clinical trial data (where available)
    - Innovation: Mechanistic + ML hybrid
5. **Immunogenicity FM**
    - Human-cell-based training (following Latent-X2)
    - Multi-donor prediction
    - Training: PBMC response data from clinical trials

8.3 Benchmark Evolution Roadmap

Table 15: **DADB Evolution Roadmap**

Version	Timeline	New Features
v1.0 (Current)	Q1 2026	Full 5-platform comparison; immunogenicity track
v1.5	Q2 2026	Bispecific benchmark; Fc engineering
v2.0	Q4 2026	ADC track; linker-payload optimization
v2.5	2027	Cell therapy (CAR-T); TCR design
v3.0	2028	In vivo prediction; clinical correlation

9 Conclusion: Toward Engineering Discipline

The field of de novo antibody design stands at an inflection point. Models that once seemed impossible—generating novel antibodies with high affinity against challenging targets—are now demonstrated reality across five major platforms. Yet the gap between computational success and clinical viability remains wide.

DADB-v1.0 represents a necessary maturation. By introducing:

1. **The Gatekeeper Architecture**—binary developability thresholds that eliminate "pyrrhic victories"
2. **The Therapeutic Decathlon**—composite scoring that weights binding, structure, developability, and immunogenicity
3. **Split-Set Evaluation**—public targets for iteration, private targets for true generalization assessment
4. **Inclusive Participation**—pathways for both open-source (RFAntibody) and proprietary (JAM-2, Chai-2, Origin-1, Latent-X2) platforms
5. **Reproducible Infrastructure**—containerized BioOps pipeline using Harbor, Nextflow, and Kubernetes

...we align computational incentives with clinical reality. A model that succeeds on DADB-v1.0 will generate not just binders, but *therapeutic candidates*—molecules with the stability, safety, and manufacturability properties to survive the attrition funnel of drug development.

The establishment of this benchmark depends on community adoption. We call on:

- **Academic groups** to submit open-source models and contribute unpublished structures for private test sets
- **Industry partners** to participate in API-based evaluation and share developability data
- **Biopharma** to validate benchmark predictions against clinical outcomes
- **Funding agencies** to support the infrastructure and wet-lab validation pipeline

The transition from "can we design antibodies computationally?" to "how do we optimally deploy these capabilities?" requires rigorous, standardized evaluation. DADB-v1.0 provides the framework. The community must now provide the participation.

Key Takeaway

Final Recommendation: Adoption of the DADB-v1.0 framework by open-source communities (Baker Lab), proprietary platforms (Nabla, Chai, AbSci, Latent Labs), and biopharma partners will accelerate the transition of Generative Biology from a research curiosity to a reliable engineering discipline—ultimately bringing more effective antibody therapeutics to patients faster.

References

[1] Nabla Bio (2025). JAM-2: Fully computational design of drug-like antibodies with high success rates. Technical Report.

[2] Chai Discovery Team (2025). Drug-like antibody design against challenging targets with atomic precision. *bioRxiv*, doi:10.1101/2025.11.29.691346.

[3] AbSci Corporation (2026). Origin-1: a generative AI platform for de novo antibody design against novel epitopes. *bioRxiv*, doi:10.64898/2026.01.14.699389.

[4] Bennett, N.R., Watson, J.L., Ragotte, R.J., et al. (2025). Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*.

[5] Latent Labs Team (2025). Drug-like antibodies with low immunogenicity in human panels designed with Latent-X2. *arXiv:2512.20263*.

[6] Schneider, C., et al. (2025). AbRank: Benchmarking antibody affinity prediction with pairwise ranking. *arXiv:2506.17857*.

[7] Akbar, R., et al. (2025). AbBiBench: Antibody Binding Affinity Benchmark for affinity maturation and design. *OpenReview*.

[8] Lensink, M.F., et al. (2024). CAPRI-Q: A stand-alone quality assessment tool for protein-protein docking. *Bioinformatics*.

[9] Kryshtafovych, A., et al. (2024). Critical Assessment of Methods of Protein Structure Prediction (CASP)—Progress and New Directions. *Proteins*.

[10] Mariani, V., et al. (2013). SAbDab: the structural antibody database. *Nucleic Acids Research*, 42(D1), D1140-D1146.

[11] De Groot, A.S., et al. (2018). Immunogenicity and T cell tolerability of antibodies. *Drug Discovery Today*, 23(10), 1954-1963.

[12] Dattani, S. (2025). Saloni's Guide to Data Visualization. *Scientific Discovery Newsletter*, December 2025.

A Appendix: Metric Calculation Details

A.1 CAPRI Metrics Implementation

The DADB-v1.0 scoring system adapts the CAPRI (Critical Assessment of PRediction of Interactions) metrics for de novo design evaluation:

$$\text{I\_rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} ||\mathbf{x}_i^{\text{pred}} - \mathbf{x}_i^{\text{ref}}||^2} \quad \text{(interface } C_\alpha \text{ atoms)} \tag{1}$$

$$\text{L\_rms} = \sqrt{\frac{1}{M}\sum_{j=1}^{M} ||\mathbf{y}_j^{\text{pred}} - \mathbf{y}_j^{\text{ref}}||^2} \quad \text{(all ligand } C_\alpha \text{ atoms)} \tag{2}$$

$$f_{nat} = \frac{|\text{contacts}^{\text{pred}} \cap \text{contacts}^{\text{ref}}|}{|\text{contacts}^{\text{ref}}|} \tag{3}$$

Quality thresholds follow CAPRI tiers: High (<1.0 Å), Good (<2.0 Å), Acceptable (<4.0 Å).

A.2 Developability Gatekeeper Algorithm

```
def developability_gatekeeper(design):
    """
    Binary gatekeeper for developability scoring.
    Returns (pass: bool, score: float, failures: list)
    """
    thresholds = {
        'Tm': ('>', 60),           # degrees Celsius
        'aggregation': ('<', 5),    # percent
        'solubility': ('>', 10),    # mg/mL
        'hydrophobicity': ('<', 15) # minutes (HIC)
    }

    failures = []
    for prop, (operator, threshold) in thresholds.items():
        value = design.properties[prop]
        if not evaluate(value, operator, threshold):
            failures.append(f"{prop}: {value}")

    if failures:
        return False, 0.0, failures

    # Calculate continuous score for passing designs
    score = weighted_average([
        normalize(design.Tm, 60, 90),
        1 - normalize(design.aggregation, 0, 5),
        normalize(design.solubility, 10, 100),
        1 - normalize(design.hydrophobicity, 0, 15)
    ], weights=[0.3, 0.3, 0.25, 0.15])

    return True, score, []
```

A.3 Platform-Specific Notes

A.3.1 Latent-X2 Immunogenicity Protocol

The immunogenicity assessment for Latent-X2 followed this protocol:

1. **Target:** TNFL9 (immunomodulatory)
2. **Donors:** 10 healthy human PBMC donors
3. **Controls:**
    - Positive: ImmunoCult, PHA
    - Reference therapeutic: Caplacizumab (approved VHH)
4. **Assays:**
    - T-cell proliferation at 48 and 120 hours
    - Cytokine release (IL-6, IL-8, IL-10, IFN-γ, TNF-α) at 120 hours
    - Cell viability (CellTiter-Glo)
5. **Concentrations tested:** 1.11, 3.33, 10, 30 μg/mL

Result: No detectable immunogenic response for any of the 4 tested VHH designs across all donors and concentrations.