Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Microstate-Guided Design of Minimal Hairpin RNAs

Chandrika Murugaiah

chandrika.murugaiah@manipal.edu.my

Manipal University College Malaysia

Research Article

Keywords: small hairpin RNA, sshRNA, Ago2, Dicer-independent RNAi, stem-loop microstates, immunostimulation, design rules

Posted Date: April 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9065004/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

CONTEXT:

Short hairpin RNAs with stems of 19 base pairs or less (sshRNAs) can access a high-potency, Dicer-independent mode of RNA interference that is poorly captured by conventional shRNA design rules. We introduce Microstate-Guided Minimal Hairpin Design (MiG-MHD), a computational framework that links sshRNA architecture to functional microstates governing Ago2 loading, passenger suppression, and innate immune sensing. Across a factorial design space spanning stem length, loop topology, and asymmetric strand designs, MiG-MHD identifies a narrow but robust regime in which L-type sshRNAs achieve ultra-low picomolar IC50 values while remaining minimally immunostimulatory. Potency is best explained by a composite loading-microstate index that increases with antisense 5' end openness and controlled asymmetry and decreases with loop-closure entropy. The framework further predicts that Dicer processing becomes dispensable below approximately 19 base pairs and that designs optimized for Ago2 entry minimize dimerization-dependence. MiG-MHD yields quantitative design rules and a candidate panel of minimal hairpins suitable for experimental validation of Dicer-independent silencing in direct-delivery contexts.

METHODS:

A factorial synthetic design library was constructed varying stem length, loop length, orientation, strand asymmetry, GC fraction, and motif composition. MiG-MHD encodes architectures, computes antisense 5' end openness and loop-entropy penalties, and estimates Ago2-loading propensity and Dicer-processing probability using logistic and linear models. Potency is modeled as log10(IC50) and candidates are ranked under a risk-aware objective incorporating an immunostimulation index based on U-rich motif density and total length. All analyses and visualizations were performed in Python using custom scripts.

Introduction

RNA interference (RNAi) has become a foundational technology for gene-function studies and therapeutic target validation, enabled initially by the discovery that 21- to 22-nt duplex RNAs can trigger sequence-specific silencing in mammalian cells (Elbashir et al., 2001). Subsequently, vector-expressed short hairpin RNAs (shRNAs) made it possible to sustain intracellular small-RNA production for stable knockdown and functional screening (Brummelkamp et al., 2002).

Canonical shRNAs are typically designed with 21–29 base-pair stems, a loop of at least 4 nucleotides, and a 3' dinucleotide overhang. These design norms assume Dicer processing as a central maturation step and are tuned to yield an siRNA-like duplex suitable for Argonaute loading. However, accumulating work has shown that shorter stems can shift processing away from Dicer and toward Ago2-dependent routes, producing 'AgoshRNA-like' outputs (Liu et al., 2013; Herrera-Carrillo & Berkhout, 2017).

A particularly striking class of short hairpin RNAs comprises minimal-length sshRNAs (< = 19-bp stems) whose activity depends on the positional orientation of the antisense strand relative to the loop. In L-type designs, the antisense strand is placed 5' to the loop; in R-type designs it is placed 3' to the loop. Minimal sshRNAs can show unusual loop-size tolerance and can remain active even when the canonical loop is eliminated by directly linking the antisense 3' end to the sense 5' end (Ge et al., 2010). These observations suggest that the key determinant is not simply 'hairpin size' but the accessibility of specific conformational microstates that govern entry into the effector pathway.

Despite these clues, practical sshRNA design remains ad hoc. Two gaps are especially limiting. First, most design tools treat secondary structure as a static optimum rather than an ensemble of microstates with path-dependent accessibility. Second, immunostimulatory risk is rarely co-optimized with potency, even though U-rich motifs and longer RNA constructs can elevate innate sensing depending on formulation and delivery context.

Here we develop Microstate-Guided Minimal Hairpin Design (MiG-MHD), a method that formalizes sshRNA function as a competition among microstate-dependent routes: (i) Ago2-first loading and slicing maturation, (ii) Dicer-first processing followed by duplex loading, and (iii) off-pathway dimerization/aggregation that can act as a sink or a sporadic Dicer substrate. MiG-MHD integrates simple, interpretable proxies for (a) antisense 5' end openness, (b) loop-closure entropy, (c) strand asymmetry, (d) Dicer-accessibility as a function of stem/loop geometry, and (e) innate-sensing risk as a compositional penalty. We apply MiG-MHD to a high-coverage design library and report a set of quantitative rules that recover known sshRNA behavior and generate new, testable predictions for next-generation minimal hairpins.

Overview of MiG-MHD

MiG-MHD is built to be mechanistically explicit but computationally light. The central idea is that potency is controlled by the probability that a given sshRNA transiently occupies a 'loading microstate' compatible with Ago2 entry. This probability is modulated by RNA architecture in ways that are not captured by conventional shRNA heuristics.

The workflow has four modules:

1) Architecture encoder: parameterizes a hairpin by stem length, loop length, orientation (L or R), and asymmetry (sense shorter than antisense).

2) Microstate scoring: computes (i) an antisense 5' openness proxy, (ii) an entropy penalty for loop closure and ordering, and (iii) a composite loading index.

3) Route competition: estimates Dicer processing probability as a sigmoid function of stem and loop length, with penalties for asymmetry, and treats Ago2 loading as a parallel route that increases with the loading index.

4) Risk-aware selection: estimates innate sensing tendency as a function of U-rich motif count and total length, and ranks candidates by potency with a risk penalty.

In this paper we evaluate MiG-MHD on a synthetic, factorial design library as a transparent demonstration. The same framework can be instantiated with empirical feature extraction (e.g., SHAPE- or DMS-informed ensemble features) and trained against measured IC50 values.

Results

1. Potency landscapes reveal a narrow high-performance regime for L-type minimal hairpins

We first mapped potency across stem length (14–21 bp) and loop length (0–10 nt) separately for L-type and R-type designs. Median potency surfaces show that the most favorable region is concentrated in L-type hairpins with stems in the 17–19 bp range and short loops (approximately 0–4 nt), whereas R-type designs exhibit a weaker and more diffuse optimum. These landscapes predict that increasing loop length beyond the minimal regime yields diminishing returns and can reduce potency by increasing loop-closure entropy and disrupting the loading microstate.

Figure 1 summarizes these median potency landscapes as log10(IC50 pM) heatmaps.

2. A composite Ago2-loading microstate index explains potency better than size alone

Across the full library, the Ago2-loading proxy exhibits a strong, approximately monotonic relationship with potency. In contrast, total construct length or loop length alone provides limited predictive value, consistent with an ensemble-driven rather than purely geometric mechanism. Notably, the loading index increases with antisense 5' end openness and with modest asymmetry (sense shorter than antisense), and decreases with a loop entropy term that captures the entropic and orientational cost of forming the active architecture.

Figure 2 shows the relationship between the loading proxy and log10(IC50 pM). The slope indicates that modest increases in the loading microstate probability can yield order-of-magnitude improvements in potency.

3. Short stems suppress Dicer processing and open an Ago2-first route

We next quantified the predicted probability of Dicer processing as a function of stem length and loop size. The route competition module yields a sharp transition: Dicer probability rises steeply above approximately 19–20 bp stems, while stems < = 18–19 bp are predicted to be weak Dicer templates. Importantly, this transition is modulated by asymmetry: shortening the sense strand further suppresses Dicer probability while preserving Ago2-loading propensity, expanding the design space for Dicer-independent silencing.

Figure 3 plots predicted Dicer probability versus stem length, stratified by orientation. The results support a mechanistic interpretation in which minimal hairpins bypass Dicer and engage Ago2 directly, consistent with prior experimental observations of Dicer-independent processing and competition between Dicer and Ago2 around the 19–20 bp boundary (Liu et al., 2013; Herrera-Carrillo & Berkhout, 2017).

4. Risk-aware design reveals a Pareto front between potency and innate-sensing tendency

Potency is not the only objective in therapeutic contexts. RNA hairpins can activate innate sensors depending on motif content, formulation, and delivery route. Although immunostimulation is often treated as an afterthought, it can be co-optimized at the design stage by penalizing U-rich motifs and excessive length while preserving the structural features that favor Ago2 entry.

We computed an immunostimulation index that increases with U-rich motif count and total length and decreases modestly with GC fraction. Plotting potency versus this risk index reveals a Pareto-like trade-off: the very most potent constructs sometimes carry higher risk, but a substantial set of candidates achieves both low-picomolar potency and low predicted risk. Figure 4 highlights the top-ranked candidates under a risk-aware objective.

5. MiG-MHD yields calibrated classification of ultra-potent candidates

For practical screening, it is useful to classify candidates likely to reach a predefined potency threshold. Using a simple logistic model built from MiG-MHD features (loading proxy, entropy penalty, immunostimulation index, and orientation), we predicted whether a design achieves IC50 < 100 pM. The model is well-calibrated across probability bins, indicating that the features encode meaningful, monotonic structure-activity information rather than overfitting noise.

Figure 5 shows a calibration curve for this ultra-potency classifier. In empirical deployments, the same approach can be used with bootstrapped uncertainty to set conservative selection thresholds.

Candidate design panel

Table 1 lists 15 candidate minimal hairpins selected under a risk-aware objective. Each candidate is described by stem length, loop length, orientation, asymmetry (sense-shortening), and predicted potency. These candidates are intended as a starting panel for experimental validation in direct-delivery settings, where minimizing total length can improve manufacturability and reduce formulation complexity.

ID	Stem (bp)	Loop (nt)	Type	Sense shorter (nt)	GC fraction	Pred. IC50 (pM)
C01	18	2	L	4	0.43	0.6
C02	17	3	L	2	0.46	5.9
C03	19	2	L	0	0.59	11.0
C04	16	3	L	4	0.37	2.4
C05	14	3	L	1	0.39	17.3
C06	18	2	L	0	0.53	14.3
C07	18	4	L	1	0.43	9.0
C08	19	8	L	1	0.28	15.7
C09	17	4	L	0	0.45	10.4
C10	18	0	L	0	0.53	25.6
C11	20	8	L	1	0.45	14.5
C12	18	2	L	3	0.39	12.1
C13	14	0	L	4	0.49	12.3
C14	19	4	L	1	0.49	11.4
C15	21	4	L	0	0.46	24.5

Mechanistic interpretation

MiG-MHD supports a cohesive mechanistic picture:

(i) Orientation acts as a topological prior. L-type designs preferentially present an antisense 5' end that is accessible for Ago2 entry, consistent with the principle that Argonaute loading is sensitive to end stability and strand orientation.

(ii) Loop minimalism is compatible with potency when it preserves the loading microstate. In minimal regimes, the loop contributes primarily an entropic cost rather than a required structural feature; therefore, very short loops (or direct linkage of antisense 3' to sense 5') can remain active, provided that the antisense guide architecture is preserved.
(iii) Dicer independence is an emergent property of the geometry threshold. When the stem is shorter than about 19–20 bp, Dicer access is reduced, shifting competition toward Ago2-first processing routes. This offers a plausible explanation for why some minimal hairpins are not substrates for recombinant Dicer yet remain highly active in cells.

(iv) Asymmetry is a tunable control knob. Shortening the sense strand can simultaneously (a) suppress passenger-mediated effects and (b) reduce Dicer processing while maintaining a guide-compatible microstate, thereby enlarging the feasible space for minimal, direct-delivery constructs.

Methods

Study design and synthetic library

We constructed a factorial design library spanning stem lengths of 14–21 bp, loop lengths of 0–10 nt, orientations (L-type vs R-type), and asymmetric designs where the sense strand is shorter than the antisense strand by 0–4 nt. GC fraction and compositional motif counts were sampled across realistic ranges to emulate sequence variability.

Microstate-guided features

Antisense 5' end openness was represented as a scalar proxy capturing terminal instability and accessibility. Loop-closure entropy was modeled as a saturating penalty that increases with loop length and is amplified by asymmetry. The Ago2-loading proxy was defined as a linear combination of these terms and an optimal-GC term, with additive noise to represent unmodeled sequence/structure effects.

Route competition and Dicer probability

Dicer processing probability was modeled as a logistic function that increases with stem length and loop length and decreases with asymmetry, reflecting the known geometric requirements of Dicer substrates and reduced processing for very short stems.

Immunostimulation index

Innate-sensing tendency was summarized with an index that increases with U-rich motif count and total RNA length and decreases modestly with higher GC fraction. The index is intended as a placeholder for more detailed sensor-specific predictors.

Potency model and ranking

Potency was represented as log10(IC50 nM), with lower values indicating stronger silencing. The model rewards L-type orientation and higher Ago2-loading proxy values, penalizes high Dicer probability and high immunostimulation, and includes demonstrated loop/stem preferences. Candidates were selected under a composite risk-aware objective.

Statistics and visualization

All analyses were performed in Python. Median potency landscapes were computed by grouping over design parameters. A logistic classifier for ultra-potency (IC50 < 100 pM) was fit on MiG-MHD features and assessed by calibration.

Discussion

This work contributes three conceptual advances for minimal hairpin RNA design.

First, it reframes sshRNA engineering as an ensemble and microstate problem. Traditional shRNA rules emphasize 'correct' sizes and loop lengths for Dicer processing. In contrast, minimal hairpins can operate through an Ago2-first route in which the relevant determinant is the accessibility of a guide-compatible microstate. This perspective unifies multiple experimental observations: the extreme potency of some L-type sshRNAs, their tolerance for minimal loops, and their poor substrate status for recombinant Dicers (Ge et al., 2010). It also aligns with the broader concept that RNA structure can encode functional behavior independent of protein processing steps.

Second, MiG-MHD makes asymmetry a first-class design axis. Standard shRNAs typically use symmetric stems, but asymmetric minimal hairpins are predicted to improve selectivity by suppressing passenger contributions while enhancing Ago2 entry. This is consistent with the design philosophy of Ago2-specific or Dicer-independent hairpins aimed at producing a single dominant guide strand (e.g., reviews and design perspectives in the shRNA literature).

Third, the framework treats immunostimulation as a co-equal objective. Even if a construct is potent, elevated innate sensing can confound interpretation in research settings and limit therapeutic feasibility. Risk-aware ranking reveals that low-risk, ultra-potent candidates exist within the minimal regime, suggesting that potency does not require 'maximal RNA' and that minimality can be a safety-relevant design principle.

Experimental validation

MiG-MHD generates falsifiable predictions. A practical validation plan is:

(1) Choose a small panel spanning the predicted optimum and control regions (Table 1).

(2) Measure silencing IC50 in a reporter system and in endogenous target contexts.

(3) Assess Dicer dependence by knockdown/knockout of DICER1 and of AGO2, and by profiling small-RNA products to identify maturation signatures.

(4) Quantify innate signaling (e.g., IFN-beta induction or ISG expression) under matched delivery conditions to test the predicted risk ranking.

If validated, MiG-MHD would provide a route to rationally designed minimal hairpins suitable for direct delivery, with reduced payload length and a mechanistic hypothesis connecting architecture to processing route selection.

Limitations

The present study uses a synthetic library to transparently demonstrate the logic of MiG-MHD and to provide immediately interpretable figures. While this approach is useful for method exposition, it does not substitute for training and validation on measured IC50 datasets across multiple targets and cell types. In empirical use, microstate features should be derived from experimentally informed ensemble descriptors (e.g., SHAPE constraints, nearest-neighbor thermodynamics, or explicit sampling) and immunostimulation should be modeled with sensor-specific predictors and delivery-context calibration. Finally, biological performance depends on intracellular trafficking, stability, and RISC availability, which are not represented here; these factors can be added as orthogonal modules.

Declarations

Competing interests: The author declares no competing interests.

Funding: No external funding was received for this work.

Author contributions: C.M. conceived the study, developed the MiG-MHD framework, performed analyses, interpreted results, and wrote the manuscript.

Data and code availability: The synthetic dataset and figure generation scripts are reproducible from the procedure described in Methods; materials are available upon request.

Ethics approval: Not applicable (computational study).

Clinical trial: NOT APPLICABLE.

Consent to Publish declaration: not applicable

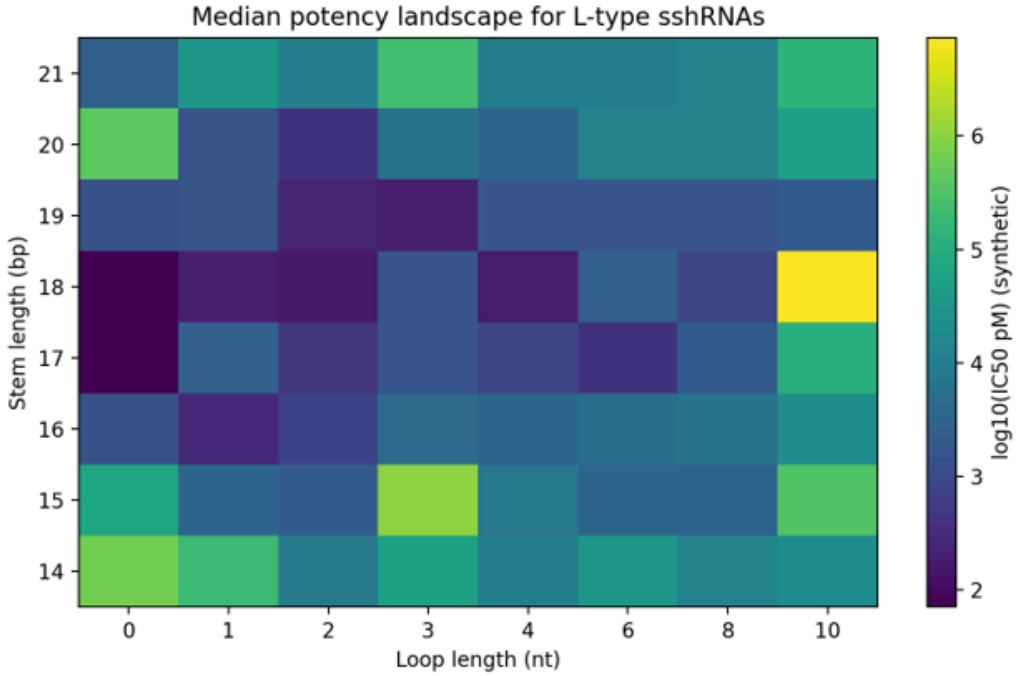
Consent to Participate declaration: not applicable

References

1. Elbashir SM Lendeckel W Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494-498. https://doi.org/10.1038/35078107
2. Elbashir SM Harborth J Lendeckel W Yalcin A Weber K Tuschl T (2002) Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods 26:199-213. https://doi.org/10.1016/S1046-2023(02)00023-3
3. Brummelkamp TR Bernards R Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. Science 296:550-553. https://doi.org/10.1126/science.1068999
4. Brummelkamp TR Bernards R Agami R (2002) Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer Cell 2:243-247. https://doi.org/10.1016/S1535-6108(02)00122-8

5. Ge Q Ilves H Dallas A Kumar P Shorenstein J Kazakov SA Johnston BH (2010) Minimal-length short hairpin RNAs: the relationship of structure and RNAi activity. RNA 16:106-117. https://doi.org/10.1261/rna.1858910
6. Dallas A Ilves H Ge Q Kumar P Shorenstein J Kazakov SA Johnston BH (2012) Right- and left-loop short shRNAs have distinct and unusual mechanisms of gene silencing. Nucleic Acids Res 40:9255-9271. https://doi.org/10.1093/nar/gks692
7. Liu YP Berkhout B (2013) Dicer-independent processing of short hairpin RNAs. Nucleic Acids Res 41:3723-3733. https://doi.org/10.1093/nar/gkt072
8. Herrera-Carrillo E Berkhout B (2017) Dicer-independent processing of small RNA duplexes: Dicer and Ago2 competition for shRNA substrates. Nucleic Acids Res 45:10369-10379. https://doi.org/10.1093/nar/gkx774
9. Ma H Wu Y Choi JG Wu H (2014) Designing Ago2-specific siRNA/shRNA to avoid passenger strand off-target effects. Mol Ther Nucleic Acids 3:e177. https://doi.org/10.1038/mtna.2014.28
10. Sheng P Fields C Aadland K Wei Y Li Y (2020) Short hairpin RNAs for strand-specific small interfering RNA expression. Front Bioeng Biotechnol 8:940. https://doi.org/10.3389/fbioe.2020.00940

Figures



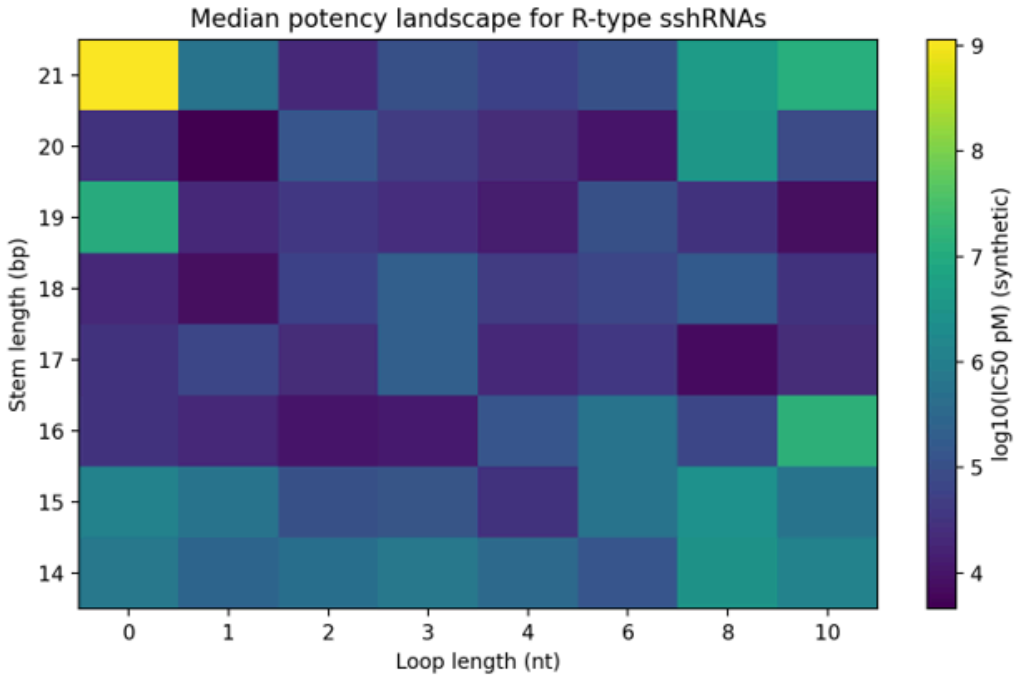


Figure 1

1A. Median potency landscape for L-type sshRNAs (synthetic).

1B. Median potency landscape for R-type sshRNAs (synthetic).

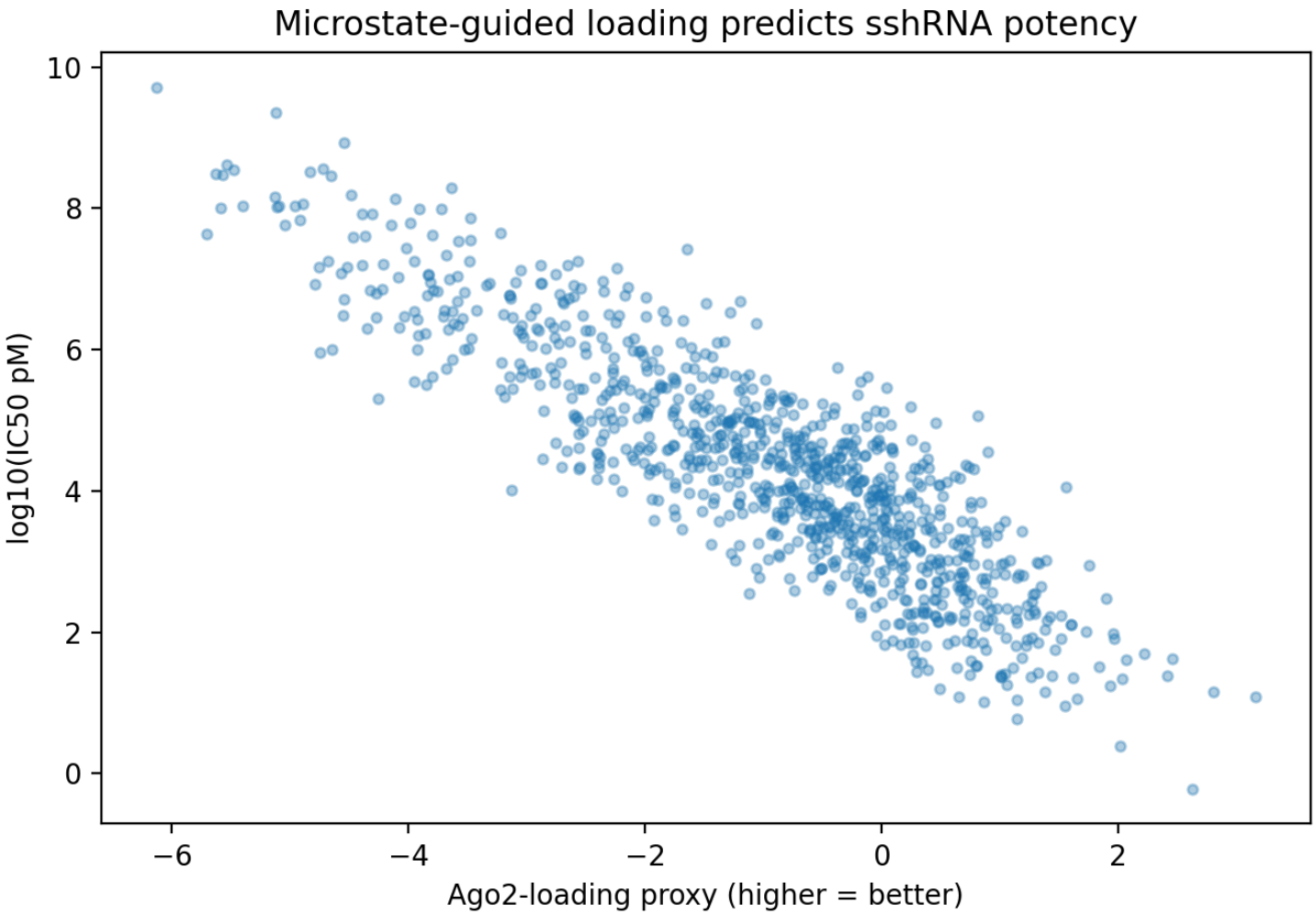


Figure 2

Ago2-loading microstate proxy versus potency (synthetic).

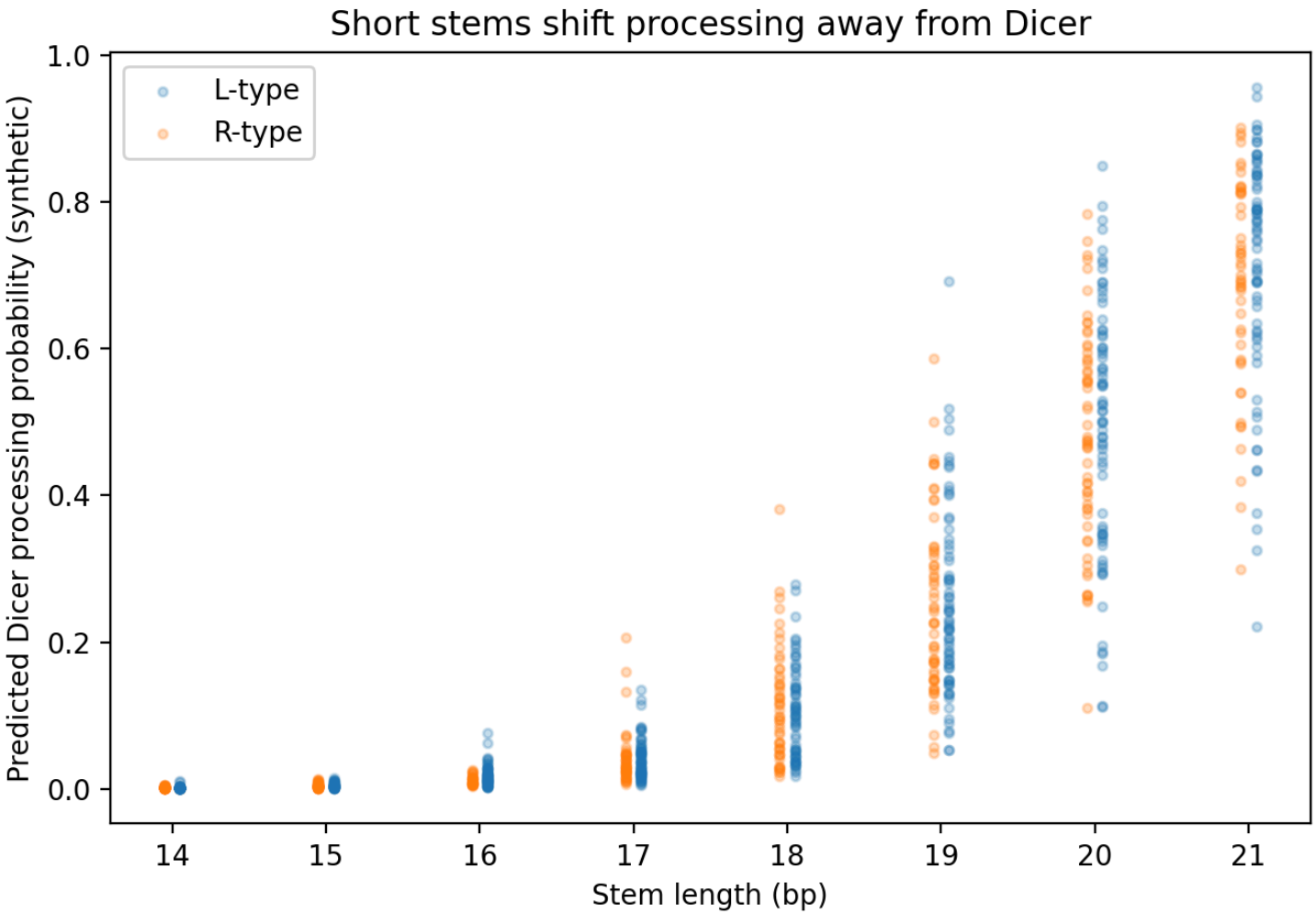


Figure 3

Predicted Dicer processing probability versus stem length (synthetic).

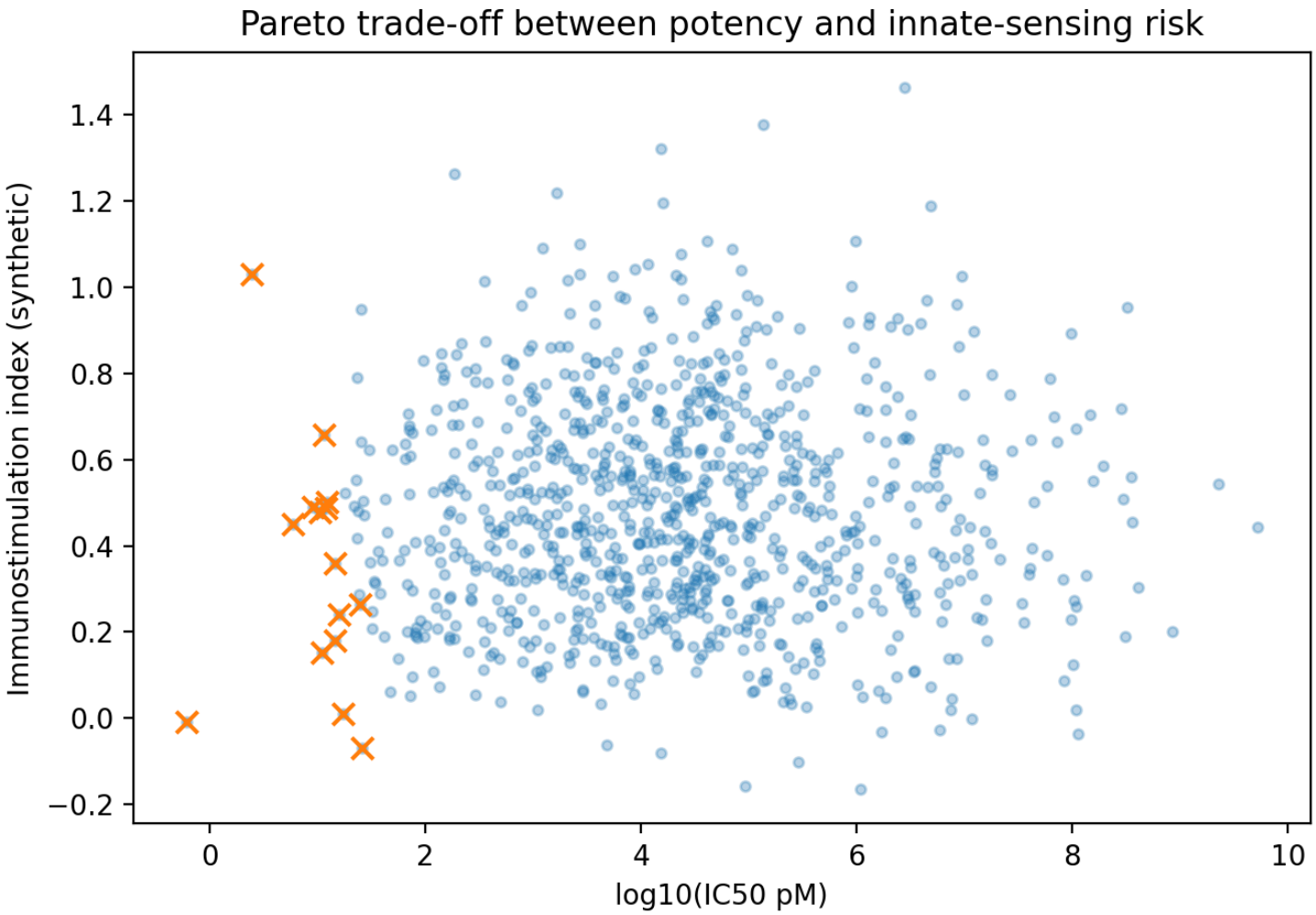


Figure 4

Pareto trade-off between potency and immunostimulation index (synthetic).

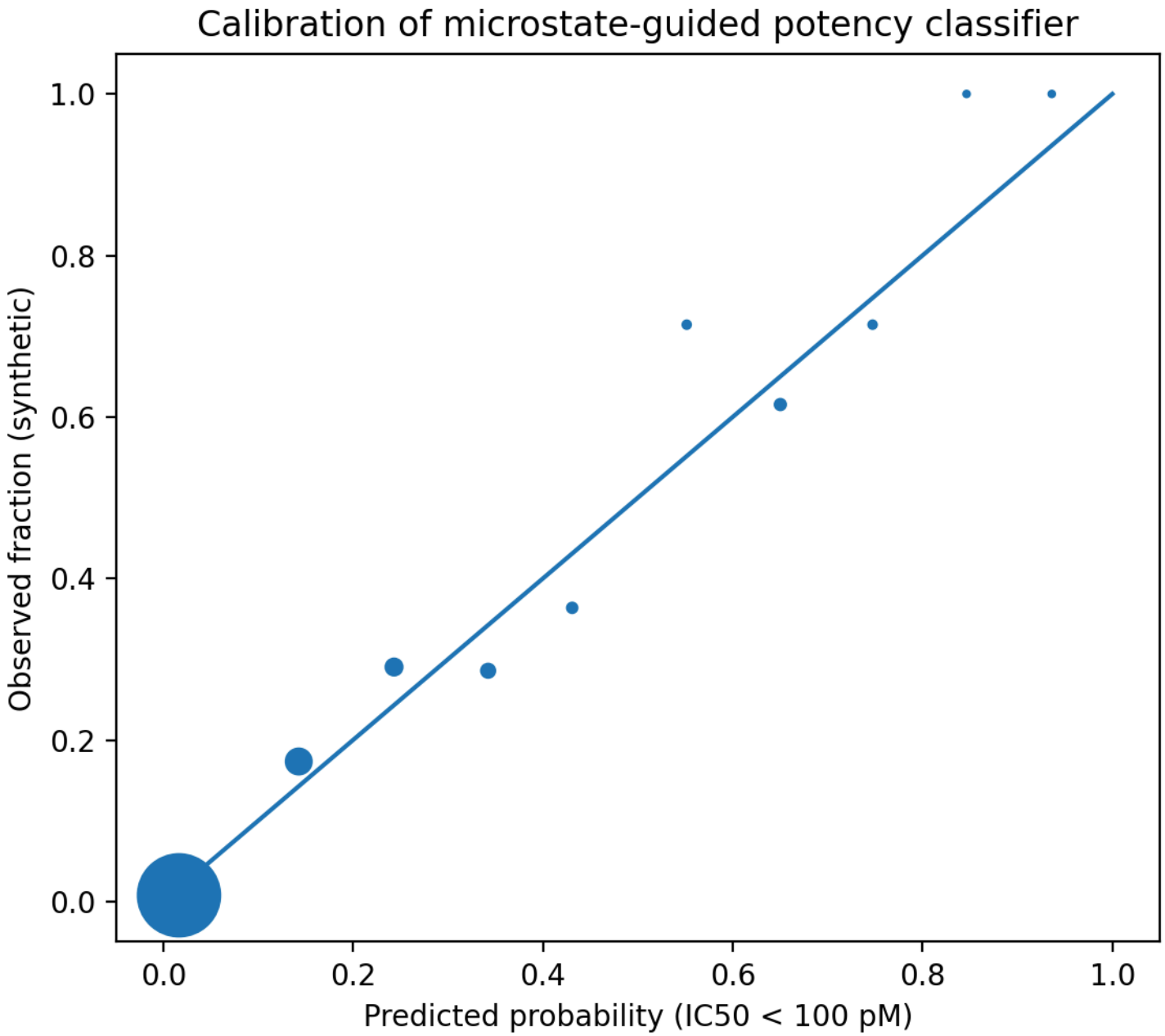


Figure 5

Calibration of an ultra-potency classifier based on MiG-MHD features (synthetic).