

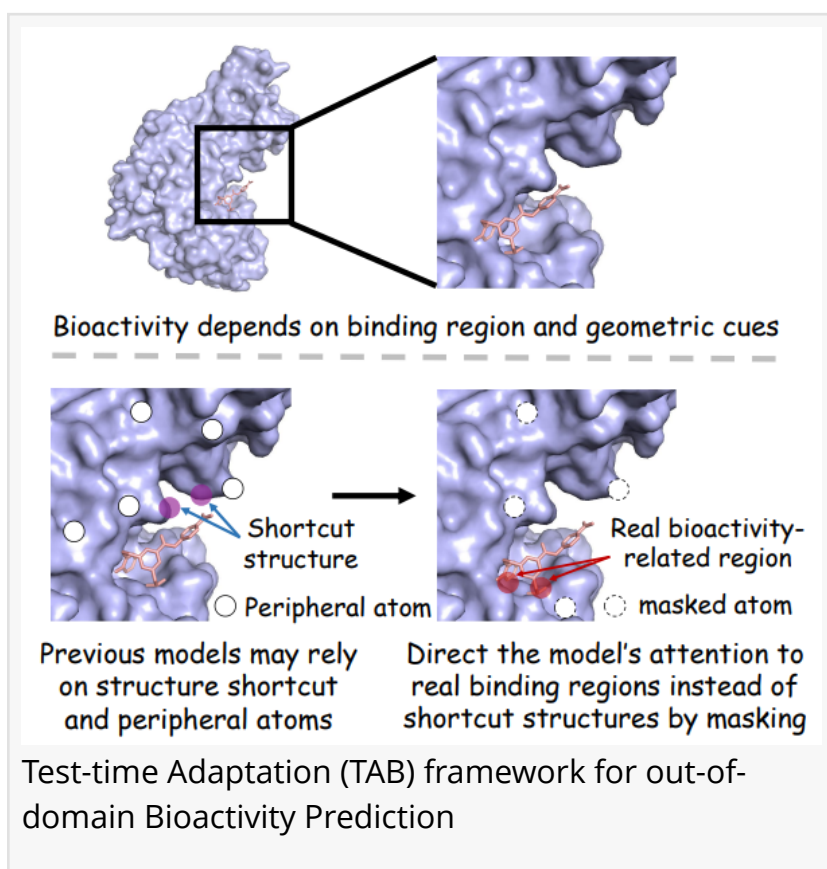
AI Breakthrough Helps Drug Discovery Models to Adapt Beyond Training Data

At ICLR 2026, a Singapore research team proposed a framework to allow AI models to make out-of-domain bioactivity prediction

SINGAPORE, July 21, 2026

/EINPresswire.com/ -- The

pharmaceutical industry faces a persistent data paradox. While artificial intelligence holds immense promise for discovering novel therapeutics, these systems consistently stumble when encountering out-of-domain (OOD) scenarios, such as previously unseen proteins or novel molecular scaffolds. Currently, predictive AI is usually constrained by the limited training data used in the model. Anything outside the training data, AI models' predictions will not be accurate.



A team of researchers from Singapore has presented a new methodology to address this potential bottleneck. Presented at the International Conference on Learning Representations (ICLR) in April 2026, this proposed Test-time Adaptation (TAB) framework for out-of-domain Bioactivity prediction allows AI models to adapt to completely new domains without accessing the original full training data.

[This paper](#) is titled "Test-Time Adaptation without Source Data for Out-of-Domain Bioactivity Prediction." Yiming Yang, Zhiyuan Zhou, and Yueming Yin co-authored the research with Associate Professors Hoi-Yeung Li and Adams Wai-Kin Kong.

TAB addresses two specific failure modes in current bioactivity prediction.

First, a drug molecule's activity is not determined by the molecule in isolation. It depends entirely on how the molecule fits into a specific pocket on the target protein. The spatial arrangement of atoms at that interface is what matters. Geometric features or the three-dimensional positioning of atoms relative to one other, are central to whether binding occurs. TAB is therefore designed to direct the model's attention toward the actual binding region and the physical geometry of the interaction. It stops the model from wandering across irrelevant parts of the molecule.

Second, AI models are vulnerable to a shortcut learning problem. Certain chemical substructures, such as a particular ring shape or a recurring protein surface pattern, frequently appear in active drug compounds. But they have no causal role in binding. The model spots these coincidental patterns in training data and leans on them as predictive signals. When the test data lacks those same coincidences, predictions collapse. TAB counters this by randomly masking portions of the atomic structure to obscure peripheral noise and focus entirely on the binding regions that dictate a drug's bioactivity, suppresses reliance on false privileged substructures, and promotes the learning of invariant, bioactivity-aware representations. Monte Carlo dropout evaluates prediction confidence during the process.

The TAB framework runs each prediction multiple times with slight randomness, measuring how consistent the answers are. Consistent predictions carry more weight; erratic ones are down-weighted. This prevents unreliable guesses from steering the adaptation process.

Next, a two-track learning system sharpens the model's understanding without supervision. One track teaches the model that modified versions of the same drug-protein pair should produce similar descriptions, while versions of different pairs should look distinct with a memory queue storing thousands of past examples as reference. The other track uses a momentum encoder: a slowly updating copy of the main model that serves as a stable anchor, preventing the system from drifting as it adapts.

Through extensive experiments on the Drug-Target Interaction Graph Neural Network or DTIGN (Yin et al., 2024), SIU 0.6 (Huang et al., 2025), and DrugOOD (Ji et al., 2023) benchmarks, the research demonstrates that the TAB framework achieves performance enhancement under scaffold, protein, and assay based OOD settings respectively, delivering improvements in both predictive accuracy and ranking consistency. On the eight subsets of DTIGN, TAB improves Pearson's R by 8.2% (a measure of predictive accuracy) and Kendall's Tau τ (a measure of ranking consistency) by 5.8% on average over the best baseline, underscoring its effectiveness as a source data-absent solution for OOD bioactivity prediction.

Effectively addressing OOD scenarios, the TAB framework allows researchers to deploy predictive models across companies and labs without sharing proprietary data, bypassing a long-standing bottleneck to AI-driven drug discovery. The TAB framework for out-of-domain Bioactivity Prediction represents a further innovation milestone, with a provisional patent application already filed and assigned to Nanyang Biologics (NYB). TAB is also part of [Vecura](#), NYB's agentic AI platform for molecular discovery and life science research.

About NYB

NYB is an [AI-native](#) biotechnology company developing discovery platforms, including Vecura and the Vecurate library, to accelerate innovation across pharmaceuticals, consumer health, and cosmeceuticals.

Jaxon Tan

Momentum AI Communications on behalf of NYB

[email us here](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/927325966>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.