From Data to Physically Meaningful Equation in Minutes: How an Agentic Reasoning-and-Coding Large Language Model Solved a Competitive Adsorption Puzzle.

\*\*Abstract\*\*

Scientific modeling often requires navigating a trade-off between physical interpretability and empirical accuracy—a task that can take weeks of iteration, especially in systems with partial observability or structural complexity. In this work, we document how a state-of-the-art agentic reasoning-and-coding large language model (LLM), known as o3, autonomously solved a modeling challenge in surface chemistry that puzzled us for months: quantifying the competitive adsorption of carboxylic acids on metal–organic layers (MOLs). When provided with experimental data and a concise problem formulation, o3 rapidly formulated a physically grounded adsorption model, implemented the code to fit it, revised its assumptions when the initial fit failed, and ultimately derived a two-parameter equation that matched experimental data across more than a dozen molecular systems. The resulting model—simple, mechanistically transparent, and quantitatively robust—incorporates both classical Langmuir competition and structural constraints such as site accessibility and residual formate. Beyond solving this specific problem, our study demonstrates a shift in scientific practice: from manually-driven trial-and-error to AI-catalyzed insight generation. We argue that o3’s success represents a new modality of research—where the language model not only aids computation but actively participates in scientific reasoning.

\*\*Introduction\*\*

Scientific modeling is often an exercise in disciplined imagination—crafting equations that explain observations while remaining grounded in physical laws. Yet, even for experienced researchers, the path from data to model can be slow and uncertain, especially when a system’s true degrees of freedom are not directly measurable. In this paper, we describe an instance where a general-purpose large language model (LLM) not only assisted but cracked a scientific modeling problem that had resisted traditional approaches.

Our system of interest involves competitive coordination exchange on the surface of metal–organic layers (MOLs), a class of two-dimensional crystalline materials whose functional performance depends on how surface-bound molecules interact with metal clusters. Despite clear experimental trends—stronger acids displacing weaker ones, saturation of modification at sub-theoretical levels—attempts to model this behavior quantitatively had either lacked predictive power or relied on ad hoc fitting. The deeper problem was that any physically meaningful model had to account for a partially accessible surface, buried coordination sites, and pH-sensitive equilibria—features that complicate classical Langmuir-type adsorption models.

This impasse became the starting point for an AI collaboration. We engaged a frontier reasoning-and-coding LLM, o3, and posed a clear but open-ended challenge: derive a simple, physically interpretable model that explains the ratio of adsorbed modifier to formate on MOLs as a function of their concentration ratio in solution. Without being pre-programmed with chemical heuristics, o3 quickly proposed a classical competitive Langmuir model, identified its failure to capture saturation behavior, and then introduced structural corrections accounting for site accessibility and bound formate. Over several reasoning–coding–testing iterations and further discussion with the researcher, it converged on a compact equation with three interpretable parameters: an equilibrium constant (*K*), the fraction of exchangeable sites (*f*). This model not only matched dozens of experimental curves with high fidelity, but also provided mechanistic insight into how acid strength and MOL structure jointly govern surface functionalization.

In this paper, we detail this process. We describe how o3 reasoned through the scientific problem, present the model it derived, and evaluate its performance against experimental data. Just as importantly, we reflect on what this episode reveals about the future of scientific modeling: one in which agentic AI can serve not merely as a calculator, but as a genuine collaborator—one capable of proposing, coding, and validating hypotheses in minutes.

Results

Metal–organic layers (MOLs) are a class of crystalline materials with an ultra-thin, sheet-like structure. They are built by linking metal clusters and organic molecules into a two-dimensional network. What makes MOLs significant is their exposed surface: unlike traditional three-dimensional materials, MOLs have accessible reactive sites that allow for direct surface modification.

In MOL synthesis, the metal clusters—typically Hf₆ oxo-clusters—are capped with small molecules like formate. These formates stabilize the structure and prevent the layers from stacking vertically, thus preserving the 2D morphology. After synthesis, these surface-bound formates can be selectively exchanged with other carboxylic acids (called modifiers), enabling programmable chemical functionalization of the MOL surface.

This exchange process forms the core of our modeling problem: different modifiers replace formate to different extents, depending on their chemical strength, concentration, and steric bulk. But the relationship between these factors and the final composition on the surface had never been captured in a general, physically interpretable equation. We turned to o3 for help in constructing such a model: to describe how much of the original formate could be replaced by a given modifier acid under different solution conditions.

Upon receiving detailed background information—including two full research summaries, fitting trials, conceptual attempts, and weekly experimental summaries in PowerPoint—alongside raw adsorption data from multiple systems, o3 recognized that the problem resembled competitive adsorption and proposed a classical Langmuir competitive model as a starting point. However, when this model was coded and applied to the data, o3 itself noted discrepancies—it could not account for the observed saturation levels.

Within minutes, o3 hypothesized that not all adsorption sites were accessible—some remained buried due to MOL stacking. It modified the Langmuir equation accordingly, introducing two structural parameters: *A*, the number of accessible sites per SBU, and *B*, the number of residual formates that could not be displaced. These concepts were not given but inferred from the shape of the experimental curves. The resulting expression,

Ron=AKRsolA+B+BKRsolR\_\text{on} = \frac{A K R\_\text{sol}}{A + B + B K R\_\text{sol}}

linked the surface ratio Ron=qMod/qFAR\_\text{on} = q\_\text{Mod}/q\_\text{FA} to the solution ratio Rsol=CMod/CFAR\_\text{sol} = C\_\text{Mod}/C\_\text{FA}, capturing both competitive binding and saturation behavior.

When applied across multiple acids—ranging from trifluoroacetic acid (TFA) to hexanoic acid (HAC)—the model exhibited excellent agreement with experiment. Fitted *K* values correlated strongly with acid strength (pKa), confirming that thermodynamic favorability governs exchange. For strong acids like TFA, *K* exceeded 40; for weak acids like HAC, *K* fell below 1. The saturation level A/BA/B predicted the upper limit of modification, typically around 3–5 for double-layer MOLs. Intriguingly, polyacids like citric acid induced structural changes in the MOL, effectively increasing *A* to near the theoretical maximum of 6, while *B* approached zero.

To further simplify the framework, o3 engaged in an iterative dialogue with the researcher to refine the model's complexity. Through this collaboration, the model was reduced to its essential physical form using just two parameters: the equilibrium constant *K* and the fraction of accessible sites *f*. The final expression adopted by the o3–researcher team was:

Ron=fKRsol1−f+fKRsolR\_\text{on} = \frac{f K R\_\text{sol}}{1 - f + f K R\_\text{sol}}

Here, *K* represents the thermodynamic preference for the modifier over formate, and *f* denotes the fraction of sites that are accessible for exchange. This equation maintained the equilibrium structure of Langmuir-type adsorption while abstracting the saturation effect into a single structural parameter.

At first glance, the derivation of this equation may appear deceptively simple. Yet, this particular form had eluded human researchers across multiple rounds of modeling attempts. One reason is the cognitive entanglement between physical constraints (such as accessibility of sites) and equilibrium behavior—features that are often treated independently or implicitly. Moreover, human modelers faced practical barriers in testing hypotheses: each proposed equation required manual coding, debugging, and fitting—a process that discouraged rapid iteration. As a result, most previous efforts either added unnecessary complexity or settled for empirical fits with limited interpretability. In contrast, o3 combined symbolic reasoning with real-time coding and data fitting, enabling it to test and revise models on the fly. This seamless integration of hypothesis-driven and data-driven reasoning allowed o3 to uncover a minimal yet mechanistically grounded solution that had escaped notice through traditional workflows.

To further interpret the physical meaning of the two parameters, we asked o3 to examine their relationship with molecular descriptors. When fitted across different acids, K inversely correlated with pKₐ, indicating that it reflects the thermodynamic driving force of ligand exchange: stronger acids (with lower pKₐ) more readily displace formate due to their greater binding affinity. Meanwhile, f decreased with increasing molecular size—measured by the number of non-hydrogen atoms in each molecule—a simple proxy for steric bulk. This revealed that f encodes a structural limitation: bulkier acids are more likely to encounter steric hindrance when approaching or coordinating with surface sites, especially in partially buried or crowded environments. These relationships were not built into the model but emerged naturally from fitting, reinforcing the notion that K and f correspond to physically grounded aspects of reactivity and accessibility.

Discussion

The two-parameter model derived by o3 was not the product of regression alone, but the result of a structured reasoning process grounded in physical interpretation. When asked to explain its approach, o3 pointed to three key features in the data: the steepness of the initial slope, the saturation behavior, and the persistent presence of formate. These were abstracted into two core parameters: *K*, the equilibrium preference of the modifier over formate, and *f*, the fraction of accessible sites. The final model retained the competitive structure of Langmuir adsorption while accounting for geometric constraints.

What sets o3 apart is its ability to reason symbolically and test hypotheses immediately. This contrasts with traditional workflows, where each new model requires manual coding and fitting, slowing the pace of exploration. o3’s integration of abstraction and validation enabled the discovery of a model that is both minimal and mechanistically grounded.

The physical meaning of *K* and *f* was validated by their correlation with molecular descriptors: *K* with acid strength (pKₐ), and *f* with molecular size. These relationships, not built into the model but revealed through fitting, reinforce its robustness.

This work illustrates a new mode of collaboration in science: where AI models like o3 contribute not only as tools, but as reasoning partners that accelerate the path from data to understanding.

Conclusion

This study presents a collaborative human–AI process that led to a physically meaningful and experimentally validated model for competitive adsorption on MOL surfaces. The role of the o3 agent went far beyond coding assistance: it interpreted trends, formulated hypotheses, tested models in real time, and guided abstraction—all while maintaining transparency in its reasoning. The final model, built on just two parameters, captures the essence of the adsorption process while remaining accessible and predictive.

The speed and quality of this discovery underscore a new research paradigm. Rather than relying solely on manual insight or black-box fitting, the integration of agentic AI offers a pathway for building mechanistic models with both speed and interpretability. As AI continues to advance, we envision these tools becoming standard collaborators in scientific inquiry—bridging the gap between complex data and conceptual understanding.