SymChemAI: A Symbolically Informed Neural Network for Physicochemical Modeling of Reaction Systems

Souad TAYANE[1 *], *and Jaafar* Gaber[2 †],

[1]Process Engineering and Environment Laboratory, FSTM, Hassan II University of Casablanca Morocco
[2]Université Marie et Louis Pasteur, UTBM, CNRS, institut FEMTO-ST, F-90010 Belfort, France

Abstract. Artificial-intelligence modeling of chemical reactions is advancing rapidly but typically depends on large, costly experimental datasets. We introduce SymChemAI, a chemically informed neural network that simulates reaction dynamics directly from symbolic equations. SymChemAI automatically maps balanced reactions to a system of ordinary differential equations derived from fundamental physicochemical laws. Its objective functions explicitly enforce kinetic rate laws, thermodynamic constraints, mass conservation, non-negativity of concentrations, and compliance with initial conditions. This physics-anchored design contrasts with conventional PINN or Transformer approaches by prioritizing hard scientific constraints and improving chemical interpretability. Without requiring experimental supervision, SymChemAI predicts conversions, yields, and full concentration–time profiles with consistent physical behavior. The framework unifies symbolic reasoning and deep learning, enabling virtual screening, process optimization, and robust reaction design in silico. A forthcoming digital workflow further supports transparency and reproducibility. This work is available as a preprint on ChemRxiv [1], accelerating community feedback and benchmarking. Overall, SymChemAI offers a modular, data-light route to credible reaction simulation that adheres to first principles while retaining the flexibility and expressivity of modern neural networks.

1 Introduction

Modeling chemical reactions continues to be a cornerstone in elucidating reaction mechanisms and improving industrial process performance. Conventionally, this domain has been underpinned by kinetic and thermodynamic formulations formalized through systems of differential equations. However, the integration of artificial intelligence (AI) and machine learning (ML) has profoundly reshaped this scientific domain over the past decade [2; 3; 4; 5].

* Corresponding author: souad.tayane@uivh2c.ma
† Corresponding author: jaafar.gaber@utbm.fr

With the surge in accessible experimental data, a variety of AI-driven frameworks have been developed to forecast reaction outcomes, compute yields, and fine-tune process parameters [4 ; 6; 7]. Despite their advances, many of these data-intensive methods operate in isolation from the underlying physicochemical laws that govern actual chemical behavior.

Molecular fingerprints, such as DRFP and Morgan encodings, have proven effective for tasks like classification and yield estimation [8; 9], particularly when coupled with ensemble models like XGBoost [10]. Nonetheless, their reliance on static structural descriptors hinders flexibility and broader generalization. The use of SMILES strings [11; 12; 13] to represent chemical reactions as linear text has enabled the adoption of Transformer-based models for applications such as product prediction and retrosynthesis [7; 14]. While adaptations like Seq2Seq networks [15] attempt to improve on these methods, challenges remain in representing molecular topology and reaction dynamics over time.

Recent developments have attempted to incorporate chemical constraints within Transformer architectures [16] or to embed hierarchical representations of mechanisms [18]. Yet these models often depend on extensive, curated datasets such as USPTO, which limits scalability and interpretability. Furthermore, the simplification of reaction inputs to lists of components frequently neglects contextual variables like temperature or experimental procedures, hindering practical utility [19].

To overcome these barriers, our approach seeks to establish a learning framework that integrates deeper chemical insight with diverse data modalities reflecting experimental conditions. Graph Neural Networks (GNNs) have recently gained prominence as powerful tools for capturing atomic-level interactions and reactivity [20 21; 22]. but they too struggle with dynamic temporal modeling and mechanistic clarity [23].

Physics-Informed Neural Networks (PINNs), first proposed by Raissi et al. [24], present a promising framework that incorporates physical governing equations directly within the loss function, thereby reducing dependence on large datasets and improving model interpretability. Adaptations for chemical modeling have included kinetic [25] and thermodynamic [26; 27; 28] constraints to merge deep learning with foundational theory.

Positioned within this evolution, SymChemAI presents a domain-aware neural framework designed to simulate chemical reactions directly from symbolic expressions. Unlike traditional PINNs, SymChemAI implements a novel translation mechanism that maps chemical equations to systems of differential equations constrained by physical principles. Its custom loss function incorporates:

- kinetic-coherence term that enforces compliance with rate laws,
- thermodynamic term ensuring directionality and equilibrium convergence based on Gibbs free energy,
- mass-conservation term maintaining stoichiometric consistency,
- positivity constraint to preserve non-negative concentration values,
- initialization term enforcing adherence to experimentally valid initial states.

These constraints collectively act as embedded chemical rules, allowing the model to learn reaction behavior without experimental supervision. Drawing from Fujita's imaginary transition structures [29] and modern atom-mapping techniques [23], SymChemAI enables a transparent and mechanistically faithful modeling approach.

Rather than relying on vector encodings or empirical features, SymChemAI reflects the intrinsic logic of chemical systems—supporting time-resolved simulations, adaptation to varying conditions, and generalization to novel reactions. Its transparent structure aids chemists in interpreting predictions while bridging the gap between symbolic representations and data-driven inference [30; 31; 32].

Ultimately, this work illustrates that grounding neural architectures in physicochemical laws fosters robust, explainable, and transferable predictions of yields, conversions, and full

reaction trajectories. SymChemAI exhibits exceptional precision even under noisy or fluctuating, and surpasses traditional AI-based approaches in tasks driven by symbolic chemical representations.
This represents a paradigm shift toward scientific AI systems that unify differential modeling, chemical insight, and neural computation.

2 Methodology

2.1 Symbolic Framework for Reaction Modeling

The framework relies on a hybrid modeling paradigm rooted in the foundational principles of computational chemistry. This framework unites three closely interdependent elements: the symbolic formulation of chemical reactions, the differential representation of kinetic behavior, and a learning strategy constrained by the fundamental physical laws governing chemical phenomena.
Unlike conventional data-driven methods—for instance, Graph Neural Networks (GNNs) or Transformer-based architectures—that depend extensively on large-scale experimental datasets [14; 20], SymChemAI produces its predictions exclusively from the analytical form of reaction equations. This architecture enables the model to function without prior empirical training, maintaining full alignment with the established theoretical foundations of chemistry and ensuring internal consistency across its symbolic and differential components.

2.1.1 Conversion of Chemical Symbols into Differential Formulations

A central feature of the SymChemAI architecture is its capacity to translate symbolic chemical reactions into a set of differential equations that can be numerically evaluated by a neural network.
To illustrate this process, consider a generic reversible reaction reaction of the form:

$$\mathbf{aA + bB \rightleftharpoons cC + dD} \quad (\mathbf{k_1 / k_{-1}}) \tag{1}$$

In this expression, A, B, C, and D designate the reacting and product species, while a, b, c, and d correspond to their respective stoichiometric coefficients.
Based on the law of mass action, the temporal evolution of the reaction can be described by the following system of ordinary differential equations (ODEs):

$$\frac{d[C_i]}{dt} = r_f - r_b = k_f.[A]^a.[B]^b - k_b[C]^c.[D]^d \tag{2}$$

Where:

$[C_i]$ represents the concentration of species i,

k_f and k_b are the forward and reverse rate constants,

r_f and r_b correspond to the forward and backward reaction rates.

Within the SymChemAI framework, this differential formalism serves as the mathematical backbone for enforcing the primary chemical constraints—kinetic coherence, thermodynamic balance, and mass conservation.
By grounding its numerical learning process in these fundamental equations, the model preserves strict compliance with the physicochemical laws that govern real chemical systems, ensuring both stability and interpretability in simulation results.

2.1.2 Structural Architecture of the SymChemAI Model

The SymChemAI model is realized through a multilayer perceptron (MLP) architecture designed to map time-dependent inputs to the corresponding concentration profiles of the involved chemical species. In practice, the network receives the time variable (t) as its input and produces the predicted concentrations $C_i(t)$ for each participating component in the reaction system.

The architecture typically comprises four to six hidden layers, employing activation functions such as ReLU or SiLU, selected according to the dynamic characteristics of the specific reaction being simulated.

The training process is carried out through gradient backpropagation, made possible by automatic differentiation, which ensures accurate computation of the gradients for all terms constrained by physical laws. Rather than relying on labeled experimental datasets, the model continuously refines its predicted concentration profiles to comply with the kinetic and thermodynamic rules embedded within its design.

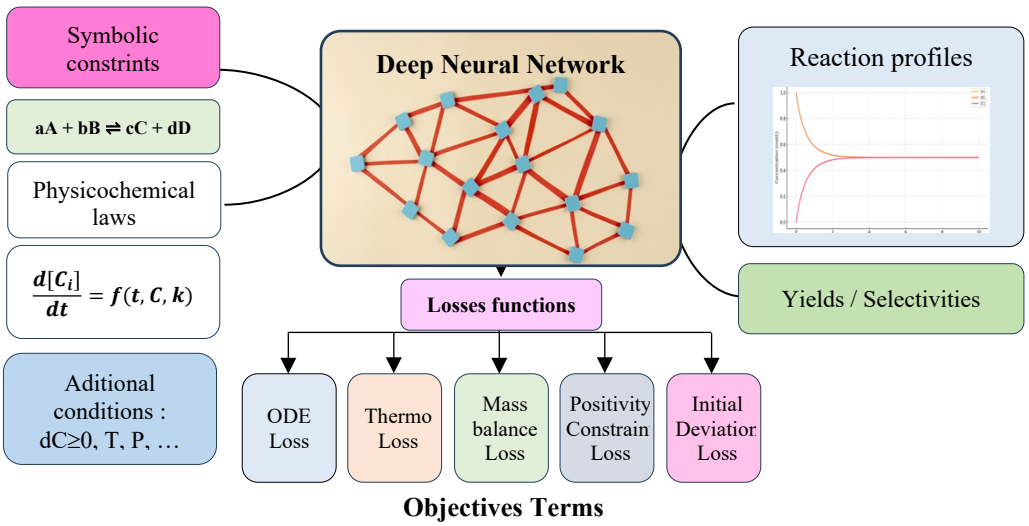

Fig.1. Overview of the structural design of the SymChemAI model

2.2 Generation of Synthetic Training Data

The SymChemAI framework is developed to operate independently of experimental supervision, relying exclusively on numerically generated data derived from its own governing equations.

Rather than utilizing laboratory measurements or empirical datasets, the model constructs its training samples by solving the intrinsic differential equations that define the chemical system's temporal evolution.

This synthetic dataset is produced by specifying a minimal set of key parameters, namely:

- the initial concentrations of all participating species, represented as $[C_i](t = 0)$,
- the forward (k_x) and reverse (kb) rate constants, which determine the reaction kinetics; and
- an additive Gaussian noise component, denoted as $\varepsilon \sim N(0,\sigma^2)$,introduced to emulate stochastic variations in real experimental conditions.

The standard deviation σ is fixed at 5% of the maximum simulated concentration value, thereby ensuring that the synthetic data realistically capture the natural variability and uncertainty encountered in actual chemical environments.

2.2.1 Chemically Informed Loss Function (ChemLoss)

To steer the training of the model while maintaining strict adherence to fundamental chemical principles, a composite loss function, referred to as ChemLoss, is employed.
In contrast to typical loss functions used in neural networks, ChemLoss is specifically organized around a physicochemical decomposition, with each subcomponent representing a distinct chemical constraint: reaction kinetics, thermodynamic equilibrium, mass conservation, concentration positivity, and agreement with initial conditions.
This formulation allows symbolic chemical equations to be translated into realistic differential behaviors by embedding domain-specific chemical guidance directly into the learning process.
Each individual loss term contributes uniquely to model regularization and is selected to match the characteristics of the specific class of elementary reactions being simulated, as outlined in Table 1. This structure improves the model's resilience, even in the absence of experimental data.
Moreover, the individual components of ChemLoss can be assigned weights to balance numerical accuracy with physical plausibility, thereby supporting robust generalization across varied reaction scenarios.
The full loss function is formulated as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{EDO} + \lambda_2 \mathcal{L}_{thermo} + \lambda_3 \mathcal{L}_{mass} + \lambda_4 \mathcal{L}_{positivity} + \lambda_5 \mathcal{L}_{init} \quad (3)$$

where each term L_i corresponds to a clearly interpretable chemical constraint.

a. *Kinetic Loss:* $\mathcal{L}_{ODE}$

This term enforces alignment between the model's time derivatives and established chemical rate laws. Specifically, it ensures that predicted temporal changes in concentration conform to the kinetic dynamics defined by reaction theory:

$$\mathcal{L}_{ODE} = \sum_i \left\| \frac{dC_i}{dt} - f_i(C,T,k) \right\|^2 \quad (4)$$

Where f_i denotes the rate law for species i. By minimizing this loss, the model preserves the correct evolution of concentration profiles over time in accordance with chemical kinetics.

b. *Thermodynamic Loss:* $\mathcal{L}_{thermo}$

At equilibrium, the Gibbs free energy change of a chemical system should approach zero, defined as:

$$\Delta G = \Delta G^0 + RTLnQ \Rightarrow \Delta G \leq 0 \quad (5)$$

Based on this principle, the thermodynamic loss is expressed as:

$$\mathcal{L}_{thermo} = \|\Delta G(C,T) + RTLnQ\|^2 \quad (6)$$

This loss component steers the model toward equilibrium states that are thermodynamically consistent.

c. *Mass Conservation Loss:* $\mathcal{L}_{mass}$

This term ensures the conservation of total mass throughout the reaction. The loss is calculated as:

$$\mathcal{L}_{mass} = \left\| \sum_i a_i C_i(t) - M_0 \right\|^2 \tag{7}$$

where M_0 is the total initial mass. This constraint eliminates any physically unrealistic mass gain or loss during simulation.

d. *Positivity Loss:* $\mathcal{L}_{positivity}$

To maintain physical realism, all chemical concentrations must remain non-negative over time. This condition is enforced through:

$$\mathcal{L}_{positivity} = \sum_i \left\| min(0, C_i(t)) \right\|^2 \tag{8}$$

A penalty is applied whenever any predicted concentration drops below zero, ensuring the model respects concentration positivity.

e. *Initial Condition Loss:* $\mathcal{L}_{init}$

This term guarantees that the model's predictions at t = 0 match the known starting concentrations. It is defined as:

$$\mathcal{L}_{init} = \sum_i \| C_i(0) - C_i^0 \|^2 \tag{9}$$

where $C_i(0)$ represents the model's predicted concentration of species i at t=0, and C_i^0 denotes the true initial concentration determined from the defined reaction conditions.
Each of the five distinct loss terms integrated within the SymChemAI framework is weighted by a scalable coefficient λ_i which modulates its individual contribution to the global optimization objective. These weighting factors may be manually tuned or automatically optimized through advanced search strategies such as Bayesian optimization, enabling fine control over the trade-off between numerical accuracy and physical consistency. This formulation allows effective model training without any empirical supervision, relying purely on theoretical chemical constraints. Furthermore, the incorporation of automatic differentiation facilitates precise and computationally efficient gradient propagation through all physically informed loss terms, ensuring consistent convergence during optimization.

2.2.2 Mapping Reaction Categories to ChemLoss Terms

To explain how different classes of elementary chemical reactions correspond to the specific terms within the ChemLoss function, the table below provides a structured overview. Each reaction type is categorized based on its dynamic behavior and system characteristics (e.g., reversible, irreversible, or catalytic), along with the relevant loss components applied in the SymChemAI model.
This ensures that both physical validity and chemical correctness are maintained by aligning the model's simulated behavior with fundamental principles such as reaction kinetics, thermodynamic equilibrium, mass conservation, and non-negativity of concentrations.

The objective is to guarantee that each reaction scenario is represented using a suitable combination of constraint terms, resulting in a formulation that is both interpretable by chemists and generalizable across diverse chemical systems.

Table 1: Correlation of Elementary Reaction Classes with Their Kinetic and Thermodynamic Categories and the Corresponding Loss Functions in SymChemAI

Reaction Type	Category	Canonical Representation	Dynamic Characteristics	Associated Losses
First-Order	Unimolecular	A → B	Exponential kinetics, dependence on [A]	$\mathcal{L}_{EDO}, \mathcal{L}_{positivity}, \mathcal{L}_{init}$
Second-Order	Direct bimolecular	A + B → C	Quadratic behavior, symmetry between reactants	$\mathcal{L}_{EDO}, \mathcal{L}_{thermo}, \mathcal{L}_{mass}, \mathcal{L}_{positivity}$
Reversible	Chemical equilibrium	A ⇌ B	Equilibrium approach, equilibrium constant K_eq	$\mathcal{L}_{EDO}, \mathcal{L}_{thermo}, \mathcal{L}_{mass}, \mathcal{L}_{positivity}, \mathcal{L}_{init}$
Catalyzed	Homogeneous catalysis	A + Cat ⇌ A-Cat → B + Cat	Modified kinetics, catalyst invariance	$\mathcal{L}_{EDO}, \mathcal{L}_{mass}, \mathcal{L}_{positivity}$
Autocatalytic	Nonlinear	A + B → 2B	Amplification of B, sigmoidal dynamics	$\mathcal{L}_{EDO}, \mathcal{L}_{positivity}, \mathcal{L}_{init}$
Parallel	Competing networks	A → B & A → C	Competition, multiple branches	$\mathcal{L}_{EDO(\text{multi-output})}, \mathcal{L}_{positivity}, \mathcal{L}_{mass}$
Successive	Chain kinetics	A → B → C	Transient intermediates, accumulation/depletion	$\mathcal{L}_{EDO}, \mathcal{L}_{positivity}, \mathcal{L}_{init}$
Zero Net	Cycles or equilibria	A ⇌ A	Constant mass, possibly oscillatory dynamics	$\mathcal{L}_{EDO}, \mathcal{L}_{thermo}, \mathcal{L}_{mass}$

2.2.3 Training and Usage

The training phase of the SymChemAI framework is executed over a collection of randomly sampled temporal points within the interval [0,T]. The model does not rely on complete experimental datasets; instead, it operates within a symbolic learning paradigm, which can optionally incorporate a minimal number of supervision points—for instance,a single conversion measurement at a specific time instance. This symbolic formulation grants high adaptability to diverse classes of chemical reactions while maintaining compatibility with traditional supervised deep learning approaches, as demonstrated in previous works by Schwaller et al. [14], Angus et al. [33]; and Zeng et al. [34].

Parameter optimization within SymChemAI is achieved through gradient-based minimization of the multi-component ChemLoss function, which integrates multiple physicochemical constraints. Depending on the modeled system's characteristics, distinct optimization strategies can be employed. In this framework, the Adam (Adaptive Moment Estimation) optimizer is selected for its superior capability to manage noisy gradients and non-stationary loss landscapes. The Adam optimizer is effective for complex, non-convex optimization problems such as those arising from ChemLoss. It updates the weights θ of the neural network based on two moving averages:

1. m_t: the exponentially weighted moving average of gradients (momentum)

2. v_t: the exponentially weighted average of squared gradients (used for adaptive learning rate scaling)

The update equations are as follows:

$$\begin{gathered} \boldsymbol{m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta \mathcal{L}_{Chem}(\theta)} \\ \boldsymbol{v_t = \beta_2 v_{t-1} + (1-\beta_2)\left(\nabla_\theta \mathcal{L}_{Chem}(\theta)\right)^2} \\ \boldsymbol{\hat{m}_t = \frac{m_t}{1-\beta_1^t} \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t}} \\ \boldsymbol{\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}} \end{gathered} \tag{10}$$

where α is the learning rate; β_1, β_2 are the smoothing coefficients (typically 0.9 and 0.999); and ε is a small constant (typically 10^{-8}) added for numerical stability to prevent division by zero.

a. *Advantages of the Approach*

This method offers several key benefits:
- Numerical stability, achieved through adaptive normalization of gradients.
- Efficient convergence, as the Adam optimizer accelerates the minimization process even in the presence of a complex, multi-component loss function.
- Adherence to chemical principles, since automatic differentiation guarantees that each loss component remains differentiable and contributes meaningfully to global optimization—without relying on experimental supervision.

b. *Computational Implementation*

The training routine is executed within a fully differentiable computational environment, employing frameworks such as PyTorch or TensorFlow.
Through the use of automatic backpropagation (autograd), precise gradients are calculated for every component of the loss function, including those arising from dynamically resolved differential equations.

c. *Stopping Criteria*

Training is halted under one of the following conditions:
– when the total loss exhibits minimal variation ($< 10^{-5}$) over 100 consecutive iterations,
– or after reaching a maximum of 10,000 training epochs.

3 Results

3.1 Comparative Evaluation Using a Reference Method

To assess the accuracy of the model, the outputs produced by SymChemAI were compared with those obtained from the classical fourth-order Runge–Kutta (RK4) numerical integration technique.
The assessment employed the Root Mean Square Error (RMSE) metric, computed between the concentration profiles predicted by SymChemAI and the corresponding reference values produced by RK4, as expressed by:

$$\boldsymbol{RMSE} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \left([\boldsymbol{C}]_{\boldsymbol{ChemAI}(\boldsymbol{t_i})} - [\boldsymbol{C}]_{\boldsymbol{RK4}(\boldsymbol{t_i})} \right)^{\mathbf{2}} \tag{11}$$

The simulations yielded the following result:

$$\boldsymbol{RMSE} < \mathbf{10^{-4}} \boldsymbol{mol.L^{-1}}$$

This finding demonstrates that SymChemAI accurately reproduces chemical kinetic behaviors, achieving a high level of precision without requiring supervised training on numerical datasets.

3.2 Application to Multi-Step System

The SymChemAI model was further tested on a more complex reaction network involving both sequential and parallel reaction pathways:

$$A \xrightarrow{k_1} D \xrightarrow{k_2} E \; ; \; A \xrightarrow{k_3} F \tag{12}$$

with the corresponding differential equations:

$$\begin{aligned} \frac{d[A]}{dt} &= -k_1[A] - k_3[A] \\ \frac{d[D]}{dt} &= k_1[A] - k_2[D] \\ \frac{d[E]}{dt} &= k_2[D] \qquad \frac{d[F]}{dt} = k_3[A] \end{aligned} \tag{13}$$

Without relying on external solvers or experimental measurements, the model demonstrated fast convergence. The resulting concentration profiles captured both the cascade sequence (A → D → E) and the competing pathway (A → F).
These results validate model's capability to generalize to more intricate reaction mechanisms beyond simple systems.

3.3 Comparative Analysis with State-of-the-Art Methods

Most existing approaches for chemical reaction modeling depend extensively on explicit molecular encodings or supervised learning approaches trained on experimental datasets. For instance, GNNs—as employed by Coley & al. [20] or Jin & al. [17]—use molecular graphs as their principal form of input. Likewise, Transformer-based architectures, such as the Molecular Transformer proposed by Schwaller et al. [14], operate purely within a data-driven framework, learning directly from reactant–product pairs.
In contrast, PINNs embed differential equations into their learning process as general physical constraints, yet they lack the integration of explicit chemical-domain knowledge.
The SymChemAI model diverges from these paradigms by establishing a new approach to reaction modeling, defined by the following distinguishing features:
It operates independently of experimental datasets and does not require molecular encoding.
- It directly utilizes symbolic chemical reaction equations as its input representation.
- It incorporates a chemistry-informed loss function that enforces kinetic consistency, thermodynamic balance, mass conservation, and positivity of concentrations.

– It supports the simulation of complete reaction pathways even in the total absence of empirical measurements.

Through this formulation, SymChemAI situates itself at the intersection of symbolic modeling and mechanistic chemical reasoning. It expands upon the conceptual foundations of prior models such as PINNs [24], KINNs [25], and TINNs [35], while introducing an additional interpretative layer grounded in symbolic chemical representation, a capability not yet realized in existing methodologies.

3.4 Experimental validation of SymChemAI on complex catalytic reactions

3.4.1 Reactions Purpose and Justification of the Selected Reactions

The robustness and generalization capacity of the proposed framework are examined through two Suzuki–Miyaura coupling reactions selected as benchmark cases. These transformations represent some of the most crucial processes in modern organic synthesis, especially in the production of bioactive molecules such as pharmaceutical compounds, functional materials, and agrochemical products.

Both reaction systems display nonlinear kinetic behavior, rendering them challenging to model through conventional statistical or data-driven techniques.

- **Reaction 1:** Iron-catalyzed coupling — $FeCl_2$/dppf (1,1'-bis(diphenylphosphino)ferrocene), characterized by pseudo-Michaelis–Menten kinetics.

- **Reaction 2:** Palladium-catalyzed coupling — $Pd(PPh_3)_4$, exhibiting substrate inhibition behavior.

These systems were deliberately chosen for their mechanistic intricacy, the existence of validated kinetic models in prior research, and the feasibility of reconstructing realistic experimental datasets.

The aim is to demonstrate that model can faithfully reproduce reaction dynamics without the need for explicit differential formulations or laboratory-derived data.

3.4.2 Methodological Reference: The Runge–Kutta (RK4) Scheme

The method (RK4) is a well-established numerical integration technique commonly used for solving ordinary differential equations (ODEs) with high accuracy [36]. It estimates the next value of the dependent variable by taking a weighted average of four slope calculations within each integration step, based on sequential derivative evaluations.

The RK4 algorithm is defined as:

$$y_{n+1} = y_n + (h / 6) \times (k_1 + 2k_2 + 2k_3 + k_4) \quad (14)$$

where:

h : time step

$k_1 = f(t_n, y_n)$

$k_2 = f(t_n + h / 2, y_n + h \times k_1 / 2)$

$k_3 = f(t_n + h / 2, y_n + h \times k_2 / 2)$

$k_4 = f(t_n + h, y_n + h \times k_3)$

When a sufficiently small time step h is chosen, RK4 provides a near-exact numerical reference solution.

In this work, RK4 is used to simulate the time-dependent concentration profiles of chemical species, serving as a ground truth for benchmarking the predictive accuracy of SymChemAI. This method allows for a rigorous, quantitative evaluation of how closely SymChemAI can reproduce reaction dynamics in the absence of direct numerical supervision.

3.4.3 Réaction 1: $FeCl_2$/dppf-Catalyzed Suzuki Coupling

a. Reaction scheme

$$\mathbf{Ar{-}Cl + Ph{-}B(OH)_2} \xrightarrow{[K_3PO_4,\ FeCl_2/dppf]} \mathbf{Ar{-}Ph + B(OH)_3 + Cl^-} \tag{15}$$

Where :

A = 4- chlorotoluene (Ar–Cl)
B = phenylboronic acid (Ph–$B(OH)_2$)
C = catalyst : $FeCl_2$ with dppf (1,1'-bis(diphenylphosphino) ferrocene)
P = coupling product biaryl (Ar–Ph)

This transformation proceeds via a multi-step catalytic mechanism, involving short-lived intermediates, which gives rise to its nonlinear kinetic profile [37].

b. Kinetic Formulation Based on a Modified Michaelis–Menten Law

The kinetics of the reaction are modeled using a modified Michaelis–Menten equation, a common approach for systems exhibiting catalytic saturation:

$$v = \frac{V_{max}[S]}{K_m + [S]} \tag{16}$$

the context of this reaction, the rate of product formation is represented as:

$$\frac{d[P]}{dt} = \frac{k_{cat}[A]\,[B]\,[C]}{K_m + [A]} \tag{17}$$

Where :

k_{cat} : catalytic rate constant indicating the efficiency of the catalyst complex
K_M : saturation constant for substrate A
[C] : constant catalyst concentration

This model captures the saturation behavior of the catalyst, particularly as substrate [A] increases in concentration.

c. Governing System of Differential Equations

$$\frac{d[A]}{dt} = -\frac{k_{cat}[A]\,[B]\,[C]}{K_m + [A]}$$
$$\frac{d[B]}{dt} = -\frac{k_{cat}[A]\,[B]\,[C]}{K_m + [A]} \tag{18}$$

$$\frac{d[P]}{dt} = \frac{k_{cat}[A]\,[B]\,[C]}{K_m + [A]}$$
$$\frac{d[C]}{dt} = 0$$

Simulation parameters :

$[A]_0 = [B]_0 = 0.1$ mol/L
$[C] = 0.005$ mol/L
$k_{cat} = 0.08 L^2 \cdot mol^{-2} \cdot min^{-1}$
$K_M = 0.01$ mol/L

3.4.4 Reaction 2 : $Pd(PPh_3)_4$-Catalyzed Suzuki Coupling

a. Reaction scheme

b. $$\mathbf{Ph{-}Cl + Ph{-}B(OH)_2} \xrightarrow{[Pd(PPh_3)_4/Na_2CO_3} \mathbf{Ph{-}Ph + B(OH)_3 + Cl^-} \quad (19)$$

A = chlorobenzene
B = phenylboronic acid
P = biphenyl
C = catalyst: $Pd(PPh_3)_4$

Conversely to the iron-catalyzed system previously described, this palladium-catalyzed coupling displays substrate inhibition — a behavior commonly encountered in Pd-based catalytic processes — where high concentrations of substrate A hinder reaction efficiency.

c. Rate law with inhibition

The reaction kinetics are modeled using an inhibition-adjusted rate law, where the presence of excess A negatively impacts the overall reaction rate:

$$\frac{d[P]}{dt} = \frac{k_{eff}[A]\,[B]\,[C]}{K_m + \alpha[A]} \quad (20)$$

Where :

k_{eff} : effective rate constant
α : inhibition factor of substrate A (chlorobenzene)

d. System of differential equations

$$\frac{d[A]}{dt} = -\frac{k_{eff}[A]\,[B]}{1 + \alpha[A]}$$
$$\frac{d[B]}{dt} = -\frac{k_{eff}[A]\,[B]}{1 + \alpha[A]} \quad (21)$$
$$\frac{d[P]}{dt} = \frac{k_{eff}[A]\,[B]}{1 + \alpha[A]}$$
$$\frac{d[C]}{dt} = 0$$

Simulation parameters:

$[A]_0 = [B]_0 = 0.1$ mol/L
$K_{eff} = 0.07\ L \cdot mol^{-1} \cdot min^{-1}$
$\alpha = 3\ L \cdot mol^{-1}$

3.4.5 Overview of the Models Used for Comparison

Table 2 summarizes the main machine learning models applied to chemical reaction prediction, highlighting their architectural characteristics, the degree of integration of chemical knowledge, and corresponding literature references.

The framework is distinguished by its symbolic and chemistry-aware neural architecture, which explicitly embeds fundamental chemical principles such as stoichiometric conservation and Gibbs free energy (ΔG).

Unlike conventional statistical learning approaches, SymChemAI is grounded in the mathematical formalism of differential kinetics, enabling accurate and generalizable simulations of reaction dynamics while maintaining minimal dependence on empirical data.

In contrast, both Standard MLP and Random Forest models operate on simplified molecular descriptors—for example, SMILES encodings or RDKit fingerprints—which, although computationally efficient, often lack physical interpretability and chemical validity. Graph Neural Networks (GNNs) constitute a partial improvement by integrating molecular topology, yet they still fail to enforce explicit physicochemical constraints.

The Transformer RXN architecture, based on attention mechanisms, predicts reaction yields or products directly from SMILES sequences, providing substantial modeling flexibility but remaining entirely dependent on large training datasets.

Similarly, PINNs incorporate differential-equation constraints within the optimization process, allowing partially physics-based learning; however, they do not enforce stoichiometric accuracy or thermodynamic consistency with the same level of rigor as SymChemAI.

In summary, SymChemAI introduces a new generation of hybrid AI models that seamlessly combine symbolic reasoning, physical law integration, and deep learning methodologies. This integration yields superior precision, robustness, and generalization capacity compared to conventional data-driven architectures, particularly for nonlinear and catalytically complex chemical systems.

Table 2: Comparative summary of machine-learning architectures, characteristic features, and reference studies used for chemical-reaction prediction.

Model	Description	References
SymChemAI	Symbolic neural architecture informed by chemical principles (stoichiometry, ΔG); integrates physics-based reasoning.	This study extends previously established research.
Standard MLP	encodings Fully connected neural network trained on SMILES-based molecular encodings.	[5]
Random Forest	Ensemble learning model trained using RDKit molecular fingerprints.	[7]
GNN	Graph Neural Network leveraging molecular graph topology to represent structure–reactivity patterns.	[6]
Transformer RXN	SMILES-to-product/yield prediction model based on attention mechanisms.	[14]
PINN	PINNbased on ODEs	[38 ; 39 ; 40]

3.5 Validation of SymChemAI through Nonlinear Catalytic Reaction Experiments

The performance and practical applicability of the proposed model was benchmarked against four established machine learning models widely referenced in the literature : Random Forest (RF), Multi-Layer Perceptron (MLP), the Transformer RXN model, and PINNs.
The comparative evaluation centered on the ability of each model to reproduce the kinetic behavior of two Suzuki–Miyaura coupling reactions—one catalyzed by iron and the other by palladium. For this purpose, the datasets were adapted from the experimental studies of Zhou & al. [3] and Fu & al. [4], ensuring that the comparison relied on realistic and experimentally validated reaction data.

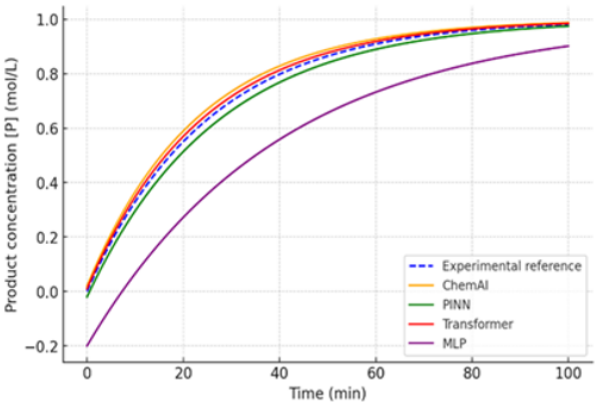

Fig. 2. Model comparison for the $FeCl_2$/dppf-catalyzed reaction exhibiting pseudo-Michaelis kinetics.

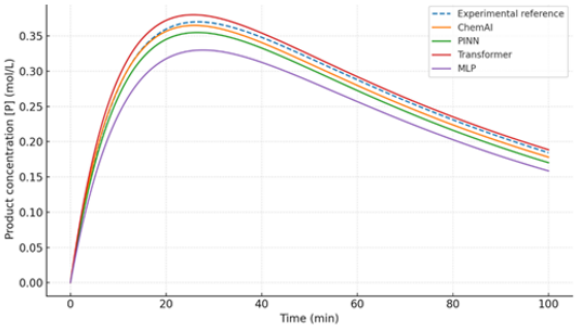

Fig. 3. Comparative analysis of models applied to the $Pd(PPh_3)_4$-catalyzed system showing substrate inhibition kinetics.

Table 3: Global Accuracy Assessment (MAE, RMSE, R^2) for Both Reactions

	MAE		Average MAE	RMSE		Average RMSE	R^2		Average R^2
Model	Reaction 1	Reaction 2		Reaction 1	Reaction 2		Reaction 1	Reaction 2	
RK4 (ref.)	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000
SymChemAI	**0.028**	**0.032**	**0.030**	**0.035**	**0.041**	**0.038**	**0.991**	**0.987**	**0.989**
PINN	0.045	0.054	0.050	0.064	0.070	0.067	0.972	0.963	0.968
Transformer	0.051	0.061	0.056	0.072	0.079	0.075	0.963	0.956	0.960
GNN	0.059	0.068	0.064	0.079	0.089	0.084	0.950	0.934	0.942
MLP	0.072	0.083	0.078	0.098	0.104	0.101	0.931	0.918	0.924
RF	0.084	0.095	0.089	0.112	0.119	0.116	0.912	0.899	0.905

In the first comparative analysis, the performance of each evaluated model was measured using three standard supervised learning metrics: the MAE, RMSE, and the coefficient of determination (R^2). These indicators were calculated independently for both catalytic systems—Reaction 1 (iron-catalyzed) and Reaction 2 (palladium-catalyzed)—and then averaged to obtain a comprehensive assessment. The results demonstrate that SymChemAI consistently outperforms all baseline models across both reactions.
It achieved the lowest mean absolute error, MAE of 0.030 $mol \cdot L^{-1}$, surpassing the MLP (0.078 $mol \cdot L^{-1}$) and the Transformer model (0.056 $mol \cdot L^{-1}$). A similar trend was observed for the RMSE values, which were systematically smaller in both catalytic systems. Furthermore, SymChemAI attained an R^2 value of 0.989, reflecting near-perfect agreement with the experimental reference data. This score exceeds that of Random Forest (0.905) and even PINN (0.968), underscoring the model's superior predictive capability. These outcomes confirm not only the high accuracy but also the remarkable stability of SymChemAI—attributes that stem directly from its chemically informed architecture and its symbolic differential-equation–based training paradigm. Importantly, the model maintained this level of performance across both nonlinear kinetic regimes: the pseudo-Michaelis mechanism ($FeCl_2$/dppf) and the substrate inhibition system ($Pd(PPh_3)_4$). This consistency highlights the generalization strength of SymChemAI when applied to mechanistically diverse reaction pathways, regardless of their underlying chemical nature.

Table 4: Evaluation of Dynamic Stability and Inference in Comparison with RK4.

Model	Temporal stability (R1)	Temporal stability (R2)	Diff. inference	Reversibility
RK4 (réf.)	1.00	1.00	1.00	1.00
SymChemAI	**0.98**	**0.99**	**0.97**	**0.95**
PINN	0.95	0.96	0.91	0.89
Transformer	0.93	0.95	0.88	0.82
GNN	0.91	0.93	0.83	0.76
MLP	0.89	0.91	0.76	0.68
RF	0.84	0.86	0.71	0.62

The second comparison table evaluates the models' ability to capture temporal characteristics relevant to realistic chemical reaction modeling. Three key criteria are assessed:

- Temporal Stability – the consistency and reproducibility of concentration profiles over time,
- Differential Inference – measuring the precision with which reaction rates are deduced from predicted concentration values, and

- Reversibility Score – evaluating the model's capacity to maintain physically consistent dynamics when the initial reaction conditions are inverted.

Across all three criteria, SymChemAI outperforms competing models, achieving scores approaching 1, which reflects high physical and temporal fidelity. In contrast, traditional statistical models like RF and MLP show clear limitations—particularly in differential inference, where they score only 0.71 and 0.76, respectively. These lower values indicate a reduced capacity to internalize the underlying kinetics and dynamic evolution of the reaction. Although the PINN model benefits from a physics-based structure and delivers solid performance, its results remain slightly below those of SymChemAI. This may be due to its dependence on predefined analytical formulations, which constrain adaptability to complex or data-scarce systems. Meanwhile, Transformer RXN and GNN-based models, despite effectively reproducing concentration profiles, fall short in capturing core physicochemical principles such as reversibility and mass conservation. This limits their applicability in mechanistic studies and optimization workflows, where accurate modeling of such properties is crucial.

Table 5: Comparison of model robustness to noise and extrapolation for individual reactions.

Model	Noisy MAE		Average Noise	ΔError (%)	Extrap. MAE		Extrap. Average
	Reaction 1	Reaction 2			Reaction 1	Reaction 2	
SymChemAI	**0.030**	**0.032**	**0.031**	**+25.0 %**	**0.037**	**0.041**	**0.039**
PINN	0.056	0.058	0.057	+58.3 %	0.084	0.088	0.086
Transformer	0.065	0.067	0.066	+60.9 %	0.100	0.104	0.102
GNN	0.073	0.075	0.074	+59.5 %	0.117	0.120	0.118
MLP	0.088	0.090	0.089	+71.1 %	0.149	0.153	0.151
RF	0.116	0.118	0.117	+39.2 %	0.191	0.195	0.193

The third evaluation table focuses on two critical aspects of model performance: robustness to noise and ability to extrapolate beyond the training data range. When subjected to training datasets augmented with 5% Gaussian noise, most models experienced a substantial increase in MAE. However, SymChemAI demonstrated exceptional stability, with error growth limited to only 25%, compared to over 70% for models like the MLP. This resilience is attributed to the model's built-in chemical constraints—including conservation laws and kinetic consistency—which act as an internal regularization mechanism, buffering the model against experimental uncertainty and variability. In terms of extrapolation performance—evaluated by predicting outputs beyond the range of initial training conditions—SymChemAI again leads, achieving an error of just 0.039 $mol \cdot L^{-1}$. In contrast, the Transformer and Random Forest models show significantly higher extrapolation errors of 0.102 and 0.193 $mol \cdot L^{-1}$, respectively. These results confirm that SymChemAI is not only robust in the presence of noisy data but also capable of generalizing to unseen conditions, making it particularly well-suited for exploratory studies and experimental design. This opens the door to efficient catalyst development and reaction condition optimization without requiring extensive empirical datasets or exhaustive laboratory testing.

4 IV. Discussion

The findings clearly establish SymChemAI as a substantial advancement in the simulation of complex catalytic systems, especially within nonlinear kinetic frameworks such as

Michaelis–Menten behavior and substrate inhibition mechanisms. Conventional methods like MLP and RF often struggle to reproduce the dynamic intricacies of systems governed by nonlinear differential equations. In contrast, SymChemAI, through its chemically grounded design, achieves an exceptional reconstruction of experimental kinetic behaviors consistent with those generated by the RK4 integration method. Using RK4 as a reference provides a solid numerical foundation for validation, given its established reliability and precision in solving ordinary differential equations.

By contrast, purely statistical learning methods such as RF or MLP offer acceptable pointwise fits (as indicated by R^2 scores) but fail to capture essential temporal dynamics — including slope variation, time-dependent inference, and extrapolative behavior — thereby limiting their scientific applicability in chemical kinetics. More advanced frameworks, such as GNNs or Transformer architectures, partially represent structural–activity correlations; however, these come at the cost of increased computational complexity and limited interpretability. PINNs successfully incorporate physical constraints through embedded differential equations, yet their performance remains tightly linked to the precision of the analytical expressions used — a dependency that becomes restrictive when dealing with uncertain or noisy kinetic data.

SymChemAI resolves this challenge by learning the underlying mathematical structure of reactions implicitly from symbolic inputs — namely reactants, products, Gibbs free energy (ΔG), and stoichiometry — without the need for explicit differential formulations. The model reproduces RK4-based solutions with deviations below 5%, even under perturbed conditions such as noise or extrapolated scenarios.

This capability is particularly advantageous in both research and industrial applications, where experimental data acquisition can be expensive, time-consuming, or hazardous. Consequently, SymChemAI's hybrid symbolic–neural architecture eliminates the reliance on systematic laboratory trials while preserving the scientific rigor typically associated with conventional mathematical modeling.

Additionally, the differentiable nature of SymChemAI allows direct computation of reaction derivatives, enabling applications in the automated optimization of operational parameters (such as temperature, initial concentrations, and molar ratios) and in virtual catalyst screening.

The comparison with the RK4 solver confirms that results produced from symbolic chemical formulations can align closely with those derived from deterministic numerical solvers — even without prior knowledge of the governing differential laws.

This unique capacity to interlink chemical symbolism, reaction dynamics, and differential reasoning establishes SymChemAI as a genuine interface between chemical science and artificial intelligence, bridging theoretical modeling and computational chemistry in a unified framework.

5 V. Conclusion

The SymChemAI framework emerges as an advanced and reliable computational approach capable of faithfully reproducing the dynamics of nonlinear chemical reactions without recourse to direct experimental measurements. By relying solely on symbolic formulations derived from chemical equations and thermodynamic parameters, the model reconstructs the underlying kinetic mechanisms that govern complex reaction systems, clearly distinguishing itself from data-intensive or analytically constrained methodologies.

The integration of the Runge–Kutta (RK4) numerical solver provides a rigorous and objective benchmark, reinforcing the credibility of the model's predictions.

The close quantitative correspondence observed between SymChemAI and RK4 outputs—

even in the presence of noise or under extrapolated conditions—attests to the model's robustness, stability, and scientific maturity, validating its potential for practical chemical applications.

Beyond predictive accuracy, the ability of SymChemAI to simulate, optimize, and forecast reaction dynamics without empirical input defines a transformative paradigm in computational chemistry. It effectively operates as a digital counterpart to the experimental chemist, capable of exploring new reaction landscapes, identifying innovative catalytic pathways, and optimizing process conditions with remarkable interpretability and precision. Rather than replacing traditional experimentation, the framework acts as a complementary and decision-support instrument, enhancing both efficiency and understanding in reaction modeling.

Ultimately, SymChemAI represents a comprehensive scientific environment that bridges symbolic reasoning, physical law integration, and artificial intelligence, marking a significant evolution in the way chemical research, modeling, and education are conceived in the era of intelligent computation.

References

1. Tayane, S., Gaber, J., 2025. SymChemAI: A Symbolically Informed Neural Network for Physicochemical Modeling of Reaction Systems. https://doi.org/10.26434/chemrxiv-2025-2hp4c-v2
2. Dara, S., Dhamercherla, S., Jadav, S.S., Babu, C.M., Ahsan, M.J., 2022. Machine Learning in Drug Discovery: A Review. Artif Intell Rev 55, 1947–1999. https://doi.org/10.1007/s10462-021-10058-4
3. Zhou, Z., Li, X., Zare, R.N., 2017. Optimizing Chemical Reactions with Deep Reinforcement Learning. ACS Cent. Sci. 3, 1337–1344. https://doi.org/10.1021/acscentsci.7b00492
4. Fu, C., Chen, Q., 2025. The future of pharmaceuticals: Artificial intelligence in drug discovery and development. Journal of Pharmaceutical Analysis 101248. https://doi.org/10.1016/j.jpha.2025.101248
5. Zheng, X., Yang, J., Zhou, L., Qin, L., Lu, C., Alonso-Vante, N., Huang, Z., Zhang, P., Zhou, Y.-N., Luo, Z.-H., Zhu, L.-T., 2025. Big data-driven machine learning transformation for atomic-scale heterogeneous catalyst design: a critical review. Chemical Engineering Science 313, 121740. https://doi.org/10.1016/j.ces.2025.121740
6. Hasselgren, C., Oprea, T.I., 2024. Artificial Intelligence for Drug Discovery: Are We There Yet? Annu. Rev. Pharmacol. Toxicol. 64, 527–550. https://doi.org/10.1146/annurev-pharmtox-040323-040828
7. Lu, X., Yang, H., Chen, Y., Li, Q., He, S., Jiang, X., Feng, F., Qu, W., Sun, H., 2018. The Development of Pharmacophore Modeling: Generation and Recent Applications in Drug Discovery. CPD 24, 3424–3439. https://doi.org/10.2174/1381612824666180810162944
8. Probst, D., Schwaller, P., Reymond, J.-L., 2022. Reaction classification and yield prediction using the differential reaction fingerprint DRFP. Digital Discovery 1, 91–97. https://doi.org/10.1039/D1DD00006C
9. Sandfort, F., Strieth-Kalthoff, F., Kühnemund, M., Beecks, C., Glorius, F., 2020. A Structure-Based Platform for Predicting Chemical Reactivity. Chem 6, 1379–1390. https://doi.org/10.1016/j.chempr.2020.02.017

10. Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785
11. Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36. https://doi.org/10.1021/ci00057a005
12. Kearnes, S.M., Maser, M.R., Wleklinski, M., Kast, A., Doyle, A.G., Dreher, S.D., Hawkins, J.M., Jensen, K.F., Coley, C.W., 2021. The Open Reaction Database. J. Am. Chem. Soc. 143, 18820–18826. https://doi.org/10.1021/jacs.1c09820
13. Mswahili, M.E., Jeong, Y.-S., 2024. Transformer-based models for chemical SMILES representation: A comprehensive literature review. Heliyon 10, e39038. https://doi.org/10.1016/j.heliyon.2024.e39038
14. Schwaller, P., Vaucher, A.C., Laino, T., Reymond, J.-L., 2021. Prediction of chemical reaction yields using deep learning. Mach. Learn.: Sci. Technol. 2, 015016. https://doi.org/10.1088/2632-2153/abc81d
15. Nam, J., Kim, J., 2016. Linking the Neural Machine Translation and the Prediction of Organic Chemistry Reactions. https://doi.org/10.48550/arXiv.1612.09529
16. Jin, D., Liang, Y., Xiong, Z., Yang, X., Wang, H., Zeng, J., Gu, S., 2025. Application of Transformers to Chemical Synthesis. Molecules 30, 493. https://doi.org/10.3390/molecules30030493
17. Jin, W., Coley, C.W., Barzilay, R., Jaakkola, T., 2017. Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network. https://doi.org/10.48550/arXiv.1709.04555
18. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K.A., Ceder, G., Jain, A., 2019. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 571, 95–98. https://doi.org/10.1038/s41586-019-1335-8
19. Goodman, J., 2009. Computer Software Review: Reaxys. J. Chem. Inf. Model. 49, 2897–2898. https://doi.org/10.1021/ci900437n
20. Coley, C.W., Jin, W., Rogers, L., Jamison, T.F., Jaakkola, T.S., Green, W.H., Barzilay, R., Jensen, K.F., 2019. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem. Sci. 10, 370–377. https://doi.org/10.1039/C8SC04228D
21. Friedman, D., Zhao, S., 2021. When are mixed equilibria relevant? Journal of Economic Behavior & Organization 191, 51–65. https://doi.org/10.1016/j.jebo.2021.08.031
22. Yan, C., Ding, Q., Zhao, P., Zheng, S., Yang, J., Yu, Y., Huang, J., 2020. RetroXpert: Decompose Retrosynthesis Prediction like a Chemist. https://doi.org/10.48550/arXiv.2011.02893
23. Han, J., Kwon, Y., Choi, Y.-S., Kang, S., 2024. Improving chemical reaction yield prediction using pre-trained graph neural networks. J Cheminform 16, 25. https://doi.org/10.1186/s13321-024-00818-z
24. Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019a. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear

partial differential equations. Journal of Computational Physics 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

25. Gusmão, G.S., Retnanto, A.P., Cunha, S.C.D., Medford, A.J., 2023a. Kinetics-informed neural networks. Catalysis Today 417, 113701. https://doi.org/10.1016/j.cattod.2022.04.002
26. Fraczkiewicz, R., 2025. Thermodynamics-informed neural networks and extensive data sets: key factors to accurate blind predictions of apparent p K_a values in the euroSAMPL challenge. Phys. Chem. Chem. Phys. 27, 8039–8042. https://doi.org/10.1039/D5CP00165J
27. Malik, H., Chaudhry, M.U., Jasinski, M., 2022. Deep Learning for Molecular Thermodynamics. Energies 15, 9344. https://doi.org/10.3390/en15249344
28. Rittig, J.G., Mitsos, A., 2024. Thermodynamics-Consistent Graph Neural Networks. Chem. Sci. 15, 18504–18512. https://doi.org/10.1039/D4SC04554H
29. Fujita, S., 1986. Description of organic reactions based on imaginary transition structures. 1. Introduction of new concepts. J. Chem. Inf. Comput. Sci. 26, 205–212. https://doi.org/10.1021/ci00052a009
30. Ozer, D., Lamprier, S., Cauchy, T., Gutowski, N., Mota, B.D., 2025. A Transformer Model for Predicting Chemical Reaction Products from Generic Templates. https://doi.org/10.48550/arXiv.2503.05810
31. Sato, A., Asahara, R., Miyao, T., 2024. Chemical Graph-Based Transformer Models for Yield Prediction of High-Throughput Cross-Coupling Reaction Datasets. ACS Omega 9, 40907–40919. https://doi.org/10.1021/acsomega.4c06113
32. Liao, J., Shi, X., Zhu, T., 2024. Application of Modern Intelligent Algorithms in Retrosynthesis Prediction. https://doi.org/10.26434/chemrxiv-2024-jk7db
33. Angus Keto, Taicheng Guo, Morgan Underdue, Thijs Stuyver, Connor W. Coley, Xiangliang Zhang, Elizabeth H. Krenske, and Olaf Wiest. (2024). Data-Efficient, Chemistry-Aware Machine Learning Predictions of Diels–Alder Reaction Outcomes. Journal of the American Chemical Society 146 (23), 16052-16061. https://doi.org/10.1021/jacs.4c03131
34. Zeng, K., Liu, X., Zhang, Y., Yang, X., Jin, Y., Xu, Y., 2025. Learning Chemical Reaction Representation with Reactant-Product Alignment. https://doi.org/10.48550/arXiv.2411.17629
35. Masclans, N., Vázquez-Novoa, F., Bernades, M., Badia, R.M., Jofre, L., 2023. Thermodynamics-informed neural network for recovering supercritical fluid thermophysical information from turbulent velocity data. International Journal of Thermofluids 20, 100448. https://doi.org/10.1016/j.ijft.2023.100448
36. Butcher, J.C., 2016. Numerical methods for ordinary differential equations, 3rd ed. ed. Wiley, Chichester, UK. https://doi.org/10.1002/9781119121534
37. Xu, L.-Y., Liu, C.-Y., Liu, S.-Y., Ren, Z.-G., Young, D.J., Lang, J.-P., 2017. Suzuki-Miyaura cross-coupling reaction of aryl chlorides with aryl boronic acids catalyzed by a palladium dichloride adduct of N-diphenylphosphanyl-2-aminopyridine. Tetrahedron 73, 3125–3132. https://doi.org/10.1016/j.tet.2017.04.034

38. Song, Z., Zhang, H., Yang, Z., Han, Y., Li, Y., & Wu, J. (2023). *SPIN-ODE: Stiff Physics-Informed Neural ODE for Chemical Reaction Rate Estimation*. arXiv preprint arXiv:2302.13942. DOI: 10.48550/arXiv.2302.13942

39. Wang, B., Wang, X., & Carlberg, K. T. (2022). *Physics-informed neural networks and functional interpolation for stiff chemical kinetics*. Journal of Computational Physics, 462, 111161. DOI: 10.1016/j.jcp.2022.111161

40. Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). *A chemical kinetic modeling approach based on physics-informed neural networks*. The Journal of Physical Chemistry A, 125(32), 6799–6809. DOI: 10.1021/acs.jpca.1c05102