MiST: Understanding the Role of Mid-Stage Scientific Training in Developing Chemical Reasoning Models

Anonymous Author(s)
Affiliation
Address
email

Abstract

Large Language Models (LLMs) acquire emergent reasoning capabilities when fine-tuned in an online setting with simple rule-based rewards. Recent studies, however, indicate that success in this regard is conditioned on the latent solvability of tasks in the base LLM: RL can only amplify answers to which the base model already assigns non-negligible probabilities. This work investigates the emergence of chemical reasoning capabilities and what these prerequisites mean for chemistry. We identify two necessary conditions for RL-based chemical reasoning: 1) Symbolic competence, and 2) Latent domain knowledge. We propose MiST: a set of mid-stage training techniques to satisfy these, including data-mixing with SMILES-aware preprocessing and continued pre-training on a rich data mixture of 2.9B tokens. These steps raise the latent-solvability score on IUPAC to SMILES translation by 2x and enable RL to lift top-1 accuracy on reaction prediction from 4.1% to 25.2% on challenging chemical tasks, while producing faithful reasoning traces. Our results define clear prerequisites for chemical reasoning training and highlight the broader role of mid-stage pre-training in unlocking reasoning capabilities.

1 Introduction

Reasoning tasks in chemistry are fundamental yet notoriously challenging, requiring models to integrate multiple layers of chemical knowledge and logical deduction (Coley et al., 2019; Alampara et al., 2024). While traditional chemoinformatics approaches rely primarily on supervised architectures optimized for specific tasks, they lack generalization and human-like reasoning capacities, instead often performing as highly specialized pattern recognition systems (Schwaller et al., 2019; Mirza et al., 2024a). Recently, reinforcement learning (RL) driven frameworks (Guo et al., 2025b) have shown promising advances in generating sophisticated emergent reasoning capabilities without explicit step-level supervision, achieving remarkable results across general-purpose domains like math and coding. Nevertheless, independent follow-ups have shown that such capabilities do not simply appear, but emerge instead as amplified patterns already existing in the base model's output distribution —even if with low likelihoods (Guo et al., 2023; Flam-Shepherd & Aspuru-Guzik, 2023). Consequently, whether RL succeeds on a new domain depends crucially on the latent solvability of the tasks for that specific base model.

Chemistry presents a severe stress test for this premise. Unlike arithmetic or programming, chemical problems combine highly specialized symbol systems (Weininger, 1988) (e.g. SMILES, IUPAC) with domain-specific physical constraints (valence, stereochemistry). Off-the-shelf LLMs typically fail to generate syntactically valid SMILES, let alone perform any tasks involving SMILES manipulation and generation (Bran et al., 2025). Empirically, we find that direct application of RL methods to such

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

models fails: the reward signal vanishes because the correct answer never appears in the candidate set, except for the simpler examples.

These observations raise a fundamental question: What pre-training and prerequisites must an LLM satisfy so that RL can reliably unlock chemical reasoning? In this paper, we answer that question by 1) proposing quantitative diagnostics that measure a model's latent solvability for chemical tasks, 2) systematically creating and ablating two proposed domain-specific prerequisites, and 3) showing that RL and other reasoning post-training techniques succeed if those diagnostics cross certain thresholds.

We propose symbolic competence and latent chemical knowledge as two necessary prerequisites for reasoning in chemistry. The former requires that models must be able to read and generate syntactically valid chemical strings, like SMILES, IUPAC names, or CIF files. The second requirement means that answers must exist in the long-tail of the model's prior distribution, so that these can be exploited by the RL training. We demonstrate, through a diagnostic benchmark for latent solvability, that improving on these requisites boosts model post-RL performance by up to 20%, yielding highly capable chemical reasoning models.

In addition, we propose a representative set of tasks in chemistry that are suitable for reasoning, i.e., tasks that expert humans can typically solve through some reasoning process, see Section 1. We perform a range of ablations and generalization tests on RL performance, and show that removing any single prerequisite collapses RL gains, confirming their necessity. We release 1) a diagnostic benchmark for latent solvability, and 2) our pre-training corpus. Our findings meaningfully inform how reasoning-oriented RL methods can generalize to complex scientific domains and provide a foundational roadmap toward flexible, robust chemical reasoning AI systems.

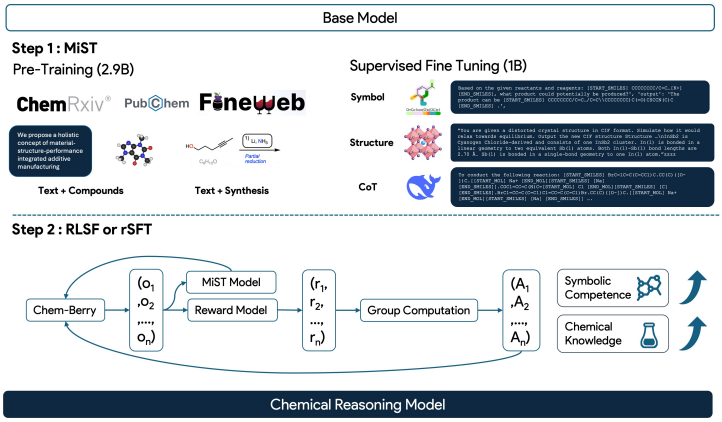

Figure 1: Multi-stage pipeline for training a chemical-reasoning language model. Step1 (*MiST*, 3.9 B tokens) Continued Pretraining exposes a general-purpose base model to a chemistry-centric corpus that interleaves plain text with compound & synthesis information. A subsequent 1 B-token supervised fine-tuning phase teaches three formats: (i) symbol-level molecular or material understanding, (ii) structure-aware question & answers, and (iii) chemical chain-of-thought (CoT). In Step2 the MiST backbone is further specialized with either RLSF (reinforcement learning from scientific feedback) or rSFT (reasoning-style supervised fine-tuning). A pool of candidate answers $(o_1, \ldots, o_n)$ generated by the MiST model is scored by a task-specific reward model $(r_1, \ldots, r_n)$; a group-computation module aggregates these signals to update the policy, iteratively refining the model into a *Chemical Reasoning Model*.

2 Related Work

Post-training methods for reasoning The standard recipe for aligning LLMs augments supervised fine-tuning (SFT) with reinforcement learning from human or synthetic feedback (RLHF/RLAIF) (Lee et al., 2024). While RLHF reliably improves helpfulness and stylistic alignment, it is often insufficient for multi-step reasoning. Subsequent work therefore introduced chain-of-thought distillation (Wei et al., 2022; Li et al., 2023), step-aware reward models (Weng et al., 2025), and tree search with

self-consistency (Xie et al., 2024b). A recent, influential result by Guo et al. (2025b) showed that even *rule-based* rewards can unlock strong mathematical and coding skills, provided that the base model already allocates non-negligible probability mass to correct answers. Independent analyses confirmed that RL mainly acts as an *amplifier*: it can only surface solutions that lie somewhere in the base distribution (Yue et al., 2025). For weaker bases, SFT on traces generated by a larger model often outperforms RL (Guo et al., 2025b). Our work adopts this "RL as amplifier" view and asks what pre-training conditions make chemical problems *latently solvable* so that RL can succeed.

Chemical language modeling Language models have been adapted and used for a range of chemical tasks (Caldas Ramos et al., 2025). These models typically operate on linearized molecular strings, such as SMILES (Weininger, 1988), SELFIES (Krenn et al., 2020), or IUPAC names. Masked pre-training approaches (ChemBERTa (Chithrananda et al., 2020), MolBERT (Fabian et al., 2020)) learn molecular fingerprints that are useful for QSAR, whereas the Molecular Transformer family targets forward and retrosynthesis prediction (Schwaller et al., 2019, 2020). More recently, LLMs have been adapted and applied for tackling chemical tasks (Frey et al., 2023; Zhang et al., 2024; Jablonka et al., 2024; Xie et al., 2023b); extending general LLMs for molecule generation, property prediction, and Q&A. Other works have adapted LLMs for use as chemistry agents, integrating robotic labs and other tools (Bran et al., 2023; Boiko et al., 2023), hypothesis generation (Yang et al., 2025), and more recently, workflows have been designed for molecular design and synthesis planning (Wang et al., 2024a; Bran et al., 2025).

Mid-stage domain adaptation Continuing pre-training on an in-domain corpus—often called *domain-adaptive pre-training* (DAPT) or *continued pre-training* (CPT)—has become the dominant recipe for turning a general LLM into a domain specialist. Early successes such as BioMegatron for biomedicine (Shin et al., 2020), Legal-BERT (Chalkidis et al., 2020), and Code-Llama for programming (Rozière et al., 2023) demonstrated sizable gains with only a few billion extra tokens. A recent wave of work scales the idea to scientific domains: (1) AdaptLLM (Cheng et al., 2023) shows that a 7B parameters model, after just 10–15B of financial tokens, rivals BloombergGPT (50B parameters) on in-domain QA; (2) Tag-LLM (Shen et al., 2024) and Efficient-CPT (Xie et al., 2024a) report similar jumps while using parameter-efficient adapters; (3) SciLitLLM (Li et al., 2024b) uses a 12.7B token corpus of textbooks and full-text papers and beats much larger baselines on scientific-literature understanding; (4) domain-specific studies in materials science (Lu et al., 2025), radiation oncology (Holmes et al., 2023), Japanese finance (Hirano & Imajo, 2024), and cybersecurity (Bayer et al., 2024) confirm that CPT injects *latent* domain knowledge that survives further instruction tuning. However, two caveats emerge: CPT can erode zero-shot prompting ability if done naively (Cheng et al., 2023), and very small models ($<$ 2B parameters) often fail to develop new capabilities even after extensive CPT (Lu et al., 2025; Hsieh, 2025). Crucially, none of these works evaluate whether the adapted model becomes *latently solvable* for multi-step reasoning tasks that reinforcement learning could later amplify.

The chemical domain remains comparatively under-explored. ChemBERTa-2 (Maziarka et al., 2023) continues a BERT-style encoder on $\sim$1B SMILES tokens and improves fingerprint-style QSAR, while ChemLLM (Brand et al., 2023), DARWIN-Chem (Xie et al., 2023a), and SciDFM (Sun et al., 2024) incorporate reaction patents or literature but still operate in a single-shot, pattern-recognition regime.

LLM capability diagnostics Benchmark accuracy and perplexity offer only coarse snapshots of a model; they ignore the richer signal contained in the full conditional probability distribution. Holistic evaluation suites such as HELM (Liang et al., 2022) and LiveBench (White et al., 2024) log likelihoods but still aggregate them into single numbers. Probability-based *intrinsic* probes provide finer insight. BLiMP minimal pairs measure grammatical preference gaps (Meister & Cotterell, 2021), an idea later reused to analyze in-context learning brittleness (Zhao et al., 2024) and out-of-domain (OOD) intent detection (Wang et al., 2024b). For factual QA, calibration studies show that token-level probabilities reveal when models "know what they know" (Jiang et al., 2021; Kadavath et al., 2022). Distributional uncertainty metrics now underpin OOD detection (Liu et al., 2024a), self-correction pipelines (Liu et al., 2024b), and medical-reasoning assessment (Li et al., 2024a). Pezeshkpour (2023) and Wang et al. (2024c) formalize diagnostics as distribution-matching problems using KL divergence or Wasserstein distance, while (Ye et al., 2024) links dispersion measures to downstream robustness.

3 Preliminaries

We formalize the notions used throughout the paper and introduce the metrics that constitute our diagnostic suite.

3.1 Prerequisite 1: Symbolic Competence

To assess the symbolic competence of models, we compute the likelihood of generating a given sequence, in our case, a set of SMILES strings. We use a dataset of 10,000 molecules obtained from PubChem (Kim et al., 2025), and use the following definitions to compute a symbolic competence score.

Token log-likelihood extraction Given a model p_θ and a SMILES string $s = (t_1, ..., t_L)$.

At position i we compute the log-likelihood $r_{i,p_\theta}(s)$ of ground-truth token s_i within p_θ's next-token distribution $r_{i,p_\theta}(s) := p_\theta(t_i|t_1...t_{i-1})$. The mean of the whole string is taken as:

$$r_{p_\theta}(s) = \frac{1}{L}\sum_{i=1}^{L} r_{i,p_\theta}(s) \tag{1}$$

Symbolic competence score We define the symbolic competence score (SCS) on the assumption that a symbolically competent model should assign better likelihoods to chemically correct strings than to corrupted or invalid ones. We therefore measure the separation in the distributions of mean ranks between valid (canonical) SMILES and corrupted ones:

$$SCS := \frac{\bar{r}(corrupt(m)) - \bar{r}(canon(m))}{\sigma_{pool}}, \tag{2}$$

$$\sigma_{\text{pool}} = \sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}} \tag{3}$$

where σ_1 and σ_2 are each set's standard deviations, and σ_{pool} is the pooled standard deviation of the two sets. SCS is the Cohen's d effect size, where higher values indicate a cleaner separation and therefore stronger symbolic competence. A score of 0 means the model cannot distinguish canonical from corrupted strings, while SCS ≈ 2 corresponds to > 95 % separation. $corrupt$ is a SMILES corruption operator that randomly deletes grammar characters with a probability of 0.2, effectively yielding invalid but similar SMILES. For the material science (MatSci) task, instead of corrupting and calculating the SCS on SMILES, the calculations are performed on compositions, which specify their elements and space group in the format: A B A B <sg*X*>, where A and B are elements, and *X* represents the space group number.

3.2 Prerequisite 2: Latent Chemical Knowledge

As has recently been shown, the role of RL in training reasoning LLMs seems to be that of an amplifier, i.e., correct answers already exist in the base model's prior distribution with non-negligible probability.

With this in mind, we aim to assess the latent chemical knowledge of a given base model. As a proxy to this, we adopt the same strategy as that we use with the symbolic data, by measuring the Chemical-Competence Score (CCS), defined as the difference in the distributions of mean ranks between factually correct chemical statements and wrong ones. Given a list of chemical statements, such as the SMolInstruct Molecule Description subset (Yu et al., 2024b), we generate corrupted data by randomly swapping one sentence from each original statement with that from another randomly chosen statement in the pool.

3.3 Post-training methods

Large-scale pre-training furnishes the *prerequisites* discussed in Sections 3.1–3.2. We now describe the two post-training methods that we use throughout this work to surface and amplify these capabilities.

Supervised fine-tuning on reasoning traces Recent research (Guo et al., 2025a) has revealed that small base models can be trained with SFT on reasoning traces, resulting in small reasoning models that mimic the behavior demonstrated in the SFT training data, even if such data does not directly target the specific downstream task the models are evaluated on. The reason is that SFT transfers the response style and not only the task-specific capabilities, thus serving as an *amplifier* of latent knowledge. Following this, some reasoning traces were distilled from DeepSeek-R1 and used to perform SFT on our pretrained models. We generated $\sim$ 600,000 solutions for two canonical tasks: IUPAC $\rightarrow$ SMILES and SMILES $\rightarrow$ IUPAC, based on PubChem compounds.

Reinforcement learning with verifiable rewards Following recent works (Wang et al., 2025), we adopt Reinforcement Learning with Verifiable Rewards (RLVR) as a post-training method for our models. In this context, models are trained online with rule-based rewards that depend entirely on the final outcome. The goal of this type of training, as exemplified in previous works (Wang et al., 2025) is to encourage the model to achieve good results on the training tasks, while developing intermediate strategies to achieve this, that might involve reasoning.

We designed and used different types of reward functions for our GRPO experiments: (1) formatting rewards to ensure separation between the model reasoning and answer, (2) accuracy rewards to verify the correctness of the model answer, (3) helper rewards to penalize the model if the completions are ill-formed (such as very short completions, repetitive behaviors etc.). For the accuracy rewards, we employed different approaches to compare the answer and the solution, such as exact matches, Tanimoto similarity between SMILES, or Levenshtein distance.

Downstream reasoning tasks To train and evaluate the reasoning capabilities of our models, we implemented a suite of challenging tasks relevant to chemistry. The tasks have been selected with the following criteria in mind: (1) Difficulty: the task must be challenging enough to be unsolvable by base models alone, (2) Reasoning-suitable: tasks must be suitable for reasoning, i.e. solving an instance of the task would require more System-2 thinking from human experts than System-1 (see 5), and (3) Dataset availability: Datasets must be readily available such that, upon adaptation, an input-outcome dataset can be built that is representative of the task. The final list of tasks is listed in Table 4, and implementation details are provided in the Appendix 1.

4 MiST: Mid-stage Scientific Training

The purpose of this mid-stage training is to enhance the model's ability to generate valid SMILES, accurately follow chemistry-focused instructions, and strengthen its general chemical knowledge. We do this by continuing pretraining (next token prediction objective) on chemical and SMILES-related data, and then by performing SFT to better follow instructions and increase the thinking context window.

4.1 Datasets

The FineWeb chemistry dataset was filtered from FineWeb-Edu (Penedo et al. (2024)) using a custom non-ML classifier built using word frequency. The entire FineWeb-Edu dataset was fetched, and about 10,000 texts were manually labeled as chemistry and 50,000 as non-chemistry (based on the text source). These texts were lemmatized before building word frequency vectors for the two classes. The frequencies of the lemma k in chemistry texts and non-chemistry texts are denoted f_k^c and f_k^n, respectively. The text chemistry score (TCS) is computed using the following formula:

$$\text{TCS(text)} := \frac{1}{N_{\text{lemmas}}} \sum_{k \in \text{lemmas in text}} w_k, \quad w_k = \begin{cases} \frac{f_k^c}{f_k^n}, & \text{if } \frac{f_k^c}{f_k^n} > 1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

This labeling strategy was applied to the entire FineWeb-Edu corpus, and the texts with $TCS > 4$ were retained, yielding a pretraining set of 1.4 billion tokens of high-quality chemistry-labeled texts.

The first three million compounds from the PubChem database (Kim et al. (2025)) were dumped and filtered using the following pipeline: the compounds with ambiguous SMILES (different RDKit canonical SMILES from the IUPAC, InChI (Heller et al., 2015), or PubChem SMILES) were discarded, and the duplicates (for SMILES and IUPAC) were filtered out. Four SMILES variants (non-canonical SMILES) were generated from the canonical SMILES for each valid compound. Based on this strategy, the first million compounds from PubChem were filtered to around 600,000 compounds. The same approach was applied to the rest of the compounds, and the dataset was split in the following manner: the first million compounds (CID from 1 to 1,000,000) were used for pretraining, the second million compounds (CID from 1,000,001 to 2,000,000) were used for GRPO training, and the third million compounds (CID from 2,000,001 to 3,000,000) were used as the test split. Multiple derived datasets were also generated for the different chemical tasks used with GRPO training.

To construct the pretraining data, we used the data mixture as described in Table 1. All the data underwent the same preprocessing pipeline to interleave SMILES with text whenever a molecule name appeared (e.g. IUPAC, common name, short form, etc), this type of interleaved data was also used in (Taylor et al., 2022). We additionally generated a synthetic dataset using RDkit (RDKit, online) extracted properties of molecules (like QED, TPSA, etc) and filled it in a template. Furthermore, we include a "replay" dataset aiming to preserve the model's natural language abilities while furthering it's learning about chemical knowledge. We chose the Qwen2.5-3B base model to perform the pretraining for 3 epochs.

Dataset Source	**Tokens**	**Percentage**
ChemRxiv + S2ORC	1.2B	41.38%
FineWeb chemistry filtered data	1.4B	48.28%
PubChem synthetic data (600k compounds)	220M	7.59%
CommonCrawl Replay dataset	80M	2.75%
Total	**2.9B**	**100%**

Table 1: Dataset composition and token distribution used for the pretraining step.

For SFT, we utilized question-answering (QA) training examples derived from SmolInstruct (Yu et al., 2024a), specifically employing only the SMILES↔IUPAC and molecule captioning subsets. We also collect examples from MPtrj dataset(Deng et al., 2023). Additionally, we incorporated MMLU and chain-of-thought (CoT) reasoning traces from DeepSeek-R1, which were preprocessed to maintain coherence with our pretraining data. In this phase, we also expanded the model's context window from 4,096 to 8,192 tokens to accommodate longer reasoning sequences. The pretrained model underwent SFT for approximately 8 epochs, continuing until the previously observed loss spikes were fully mitigated. During fine-tuning, two distinct question types were used:

∗ Questions requiring explicit reasoning traces, with solutions prefixed by the tag "<think>".
∗ Questions directly presenting the final answers, prefixed by the tag "<answer>".

Dataset Source	**Notes / Sample Count**
DeepSeek reaction traces	∼7K samples
DeepSeek relaxation traces	∼2K samples
MPtrj dataset	∼20K samples
SmolInstruct dataset	I2S, S2I, Molecule captioning and generation tasks
MMLU	Train: ∼350 samples, Chemistry: ∼300 samples
CoT Chain	∼27K samples
Total Tokens	**1B**

Table 2: Instruction tuning dataset composition used for the SFT step.

The model, despite using about 3B tokens for continued pretraining and 1B tokens for SFT, performs better on some tasks in comparison with models like NatureLM (Xia et al., 2025), which has used hundreds of billions of tokens for pretraining and SFT. This was made possible by the high-quality interleaved text produced by our preprocessing pipeline.

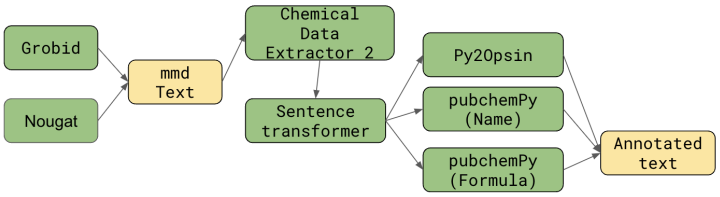

Figure 2: Overview of the preprocessing pipeline

An overview of our preprocessing pipeline is depicted as follows. Initially, we leveraged Nougat (Blecher et al., 2023) and GROBID (Meuschke et al., 2023) libraries for converting PDF documents into textual formats. Nougat demonstrated superior performance in accurately transforming complex structures such as tables, formulae, bibliographic references, and figure captions into LaTeX-formatted text. Conversely, GROBID excelled at extracting plain textual content from PDFs. The output of the authors were merged with explicit tags assigned to each structural element: tables were encapsulated with [START_TABLE] and [END_TABLE], formulas marked by [START_FORMULA] and [END_FORMULA], bibliographic references enclosed within [START_BIBREF] and [END_BIBREF], and figure descriptions bracketed by [START_FIGURE] and [END_FIGURE]. Subsequently, this structured text was processed through the Chemical Data Extractor 2 (Swain & Cole, 2016), identifying candidate molecule entities along with their positional context within the text. To ensure high precision in entity identification, candidates were further validated using a custom-trained sentence transformer model, designed specifically to discern genuine molecular entities from contextual information. Validated molecular entities were then translated from their IUPAC nomenclature to SMILES notation using py2opsin, a Python interface for OPSIN (Lowe et al., 2011). In cases where OPSIN failed to yield a definitive conversion, entities were cross-referenced against PubChem (Kim et al., 2025). Ultimately, during the pretraining phase alone, our model encountered approximately 800,000 unique chemical compounds along with their corresponding SMILES representations.

5 Post-training Experiments

This section quantifies how much of the potential unlocked by Mid-stage Scientific Training (MiST) can actually be surfaced with standard post-training recipes. We therefore keep the mid-training configuration fixed (Section 4) and vary the post-training stack:

1. BASE: original Qwen2.5-3B
2. +MIST: after MiST continued pre-training (checkpoint `v6-1`)
3. +MIST+SFT: MiST backbone after SFT on 120k DeepSeek-R1 traces (see Section 3.3).
4. +MIST+SFT+RL(TASK i): previous model further optimized with RLVR (see Section 3.3). (TASK i) specifies the single task the model is trained on with RLVR.

As an initial downstream test of our pipeline's performance, we use ChemBench (Mirza et al., 2024b) to evaluate the general chemistry knowledge of LLMs; the results are shown in Table 3.

The results in Table 3 demonstrate that MiST together with reasoning SFT, proposed in this work, has a largely positive effect on downstream tasks of general chemistry knowledge, across most chemistry sub-domains. The role of MiST is particularly important in Organic, Inorganic, and General Chemistry, with improvements of up to 6-7% over the Qwen+SFT baseline, and more than 11% over the base (instruction-tuned) model, suggesting benefits of both post-training stages on model's chemical performance. These results serve already as diagnostic measures of success of a given mid-training methodology, and serve as a basis to select models for the following stage of RLVR, with the goal of enhancing reasoning and problem-solving capabilities.

We then proceed to evaluate model's capacity of learning in an online setting from verifiable rewards through RLVR. As explained in Section 1, we implement a number of chemical tasks that are suitable

Table 3: ChemBench sub-domain Accuracy (%)

	Models		
Sub-domain	Qwen-2.5-3B Instruct	Qwen+SFT	MiST+SFT (ours)
Organic Chem	44.99	46.15	**50.12**
Inorganic Chem	46.70	51.08	**57.60**
Toxicity/Safety	21.33	**26.52**	26.37
Material Sci	35.84	42.50	**48.75**
General Chem	33.56	38.25	**44.30**
Chem Preference	45.40	50.00	**52.10**
Analytical Chem	25.00	34.20	**40.70**
Technical Chem	42.11	44.74	**50.00**
Physical Chem	20.60	35.10	**38.78**
Total	35.06	40.95	**45.41**

for reasoning, and for which verifiable rewards can be defined. Several models were trained on this setting and, in the following, we evaluate task performance as a function of the base model used. We employ two different inference techniques to generate the results reported here, namely using System-1 (direct answer) and System-2 (employing reasoning) thinking, as defined by (McGlynn, 2014). We do this by appending the tags "<answer>" or "<reasoning>" respectively upon generation, which induces models into either type of thinking. The results shown in Table 4 show the performances of our models across the multiple tasks defined in Section 1, along with the diagnostic metrics defined in Section 3.1.

Table 4: Effect of MiST and each post-training stage on downstream reasoning tasks. SCS = symbolic-competence score, CCS = chemical-competence score; both are unitless effect-size measures ranging from 0 (no separation) to 2 (near-perfect separation); higher is better, see Section 3.1. I2S = IUPAC→SMILES translation, RxP = forward reaction prediction, RxN = reaction-naming, CMG = conditional material generation. For the three downstream tasks we report top-1 accuracy. The value outside the parentheses is obtained with a "direct answer" (system-1) prompt. Values inside parentheses are the accuracy when "reasoning" (system-2 chain-of-thought) prompting is enabled.

	Metrics		Reasoning tasks			
Model	SCS ↑	CCS ↑	I2S ↑	RxP ↑	RxN ↑	CMG ↑
Qwen-2.5 3B	0.95	0.352	0.03	0.6	10.33 (10.87)	58.6
+MiST	1.639	0.443	49.12	4.1	12.8 (11.30)	1.2
+SFT	1.906	0.771	68.2 (34.5)	8.20 (21.2)	11.9 (11.0)	34.8
OrgChem Tasks						
+RL(I2S)	1.825	0.759	**68.39** (67.5)	—	—	—
+RL(RxP)	1.880	0.782	—	**8.80 (25.2)**	—	—
+RL(RxN)	1.906	0.789	—	—	**22.87 (35.17)**	—
MatSci Tasks						
+RL(CMG)	0.893	0.777	—	—	—	**73.8**
Ablations						
no MiST + SFT	1.853	0.788	22.00	5.10	2.6(4.80)	—

The results reveal the large effect that the MiST proposed here has on symbolic competence, as demonstrated by the SCS column. Clearly pretrained models like Qwen2.5-3B lack the symbolic abilities needed to complete tasks requiring SMILES understanding and writing. However, this is overcome with MiST. Furthermore, the results show that RL generally improves the performance of LLMs on specific chemical tasks, and this effect is remarkably stronger on tasks requiring SMILES synthesis, like *Reaction Prediction* or *Iupac 2 SMILES*.

One important observation from these results is that activating reasoning on RL-trained LLMs generally yields better results; however in certain cases this trend reverses, as is the case of *Iupac 2 SMILES*. In this task, we measure better performance when reasoning is not activated, however the gap is smaller for the RL-trained models. We attribute this to the ability already being present *and*

amplified in the LLM after SFT, which during RL training hinders learning of different task-solving patterns as models already perform well at that stage. Further research should go into this direction.

6 Discussion

This paper set out to answer a concrete, practical question: What conditions must a general-purpose LLM satisfy so that light-weight, rule-based post-training methods (SFT + RLVR) can unlock reliable chemical reasoning? We conducted a series of experiments using the Qwen2.5-3B model as a base. Our results suggest that our proposed Mid-stage Scientific Training (MiST) is at least necessary for unlocking chemical reasoning capabilities in LLMs. We show that both symbolic competence and latent chemical knowledge increase under MiST, and that these gains translate downstream to better achieved performances on post-training methods like RLVR and SFT.

Furthermore, as already indicated in other recent reports, RL remains an amplifier of already existing behaviors and knowledge in base LLMs. However more importantly for the case of chemistry, and other scientific fields that make heavy use of domain-specific terminology and symbolic systems, symbolic competence remains the true bottleneck, at least for small-scale LLMs. As our results demonstrate on SMILES-heavy tasks, like *Reaction Prediction* or *IUPAC to SMILES*, base models barely perform on these tasks, with results nearing 0% accuracy. SFT only boosts the Reaction Prediction results to 5.10%, however MiST is necessary to boost accuracies to 25.2% when reasoning is activated, also shows the improvements in the material science knowledge that not specifically trained (in CMG task),indicating a strong role of MiST in enabling the solution of scientific tasks.

Our findings generalize beyond chemistry. Any scientific field that (1) relies on specialized symbol systems and (2) has access to verifiable rewards can likely benefit from the same two-stage recipe: (1) ensure symbol mastery via targeted MiST, (2) apply RLVR or other post-training techniques to amplify latent solutions. With this we show that, small, compute-efficient models can already reach useful competence if those prerequisites are met. MiST demonstrates that carefully crafted, mid-stage scientific training is a powerful lever for unlocking reasoning in specialized domains. Rather than chasing ever larger parameter counts, we advocate investing in domain-specific data pipelines and intrinsic diagnostics—ingredients that, as chemistry shows, can turn an otherwise myopic LLM into a competent scientific assistant.

7 Limitations

While MiST demonstrates that targeted mid-stage pre-training can unlock chemical reasoning in a 3B-parameter model, several caveats remain. First, we have only probed a single backbone size; larger or smaller architectures may exhibit different symbolic–competence thresholds, limiting direct extrapolation of our findings. Second, the RLVR rewards we use focus on syntactic agreement with ground truth (e.g. exact SMILES or high Tanimoto similarity) and thus do not discourage chemically implausible or unsafe outputs, leaving open the possibility of reward hacking. Third, our evaluation suite—reaction prediction, IUPAC to SMILES translation, and conditional material generation, cover a narrow slice of chemistry; tasks that hinge on stereochemistry, kinetics, spectroscopy, or three-dimensional conformations remain unexplored. Finally, our pre-training corpus is dominated by small-molecule, organic literature and patents, potentially biasing the model against inorganic, macromolecular, or bio-chemical domains. Addressing these scale, reward, coverage, and data-bias issues will be critical before MiST-style models can be relied upon as general scientific assistants.

References

Nawaf Alampara, Mara Schilling-Wilhelmi, Martiño Ríos-García, Indrajeet Mandal, Pranav Khetarpal, Hargun Singh Grover, NM Krishnan, and Kevin Maik Jablonka. Probing the limitations of multimodal language models for chemistry and materials research. *arXiv preprint arXiv:2411.16955*, 2024.

Markus Bayer, Philip D Kuehn, Ramin Shanehsaz, and Christian A Reuter. CySecBERT: A domain-adapted language model for the cybersecurity domain. *ACM Transactions on Privacy and Security*, 2024.

Lukas Blecher, Guillem Cucurull, Thomas Scialom, and Robert Stojnic. Nougat: Neural optical understanding for academic documents, 2023. URL https://arxiv.org/abs/2308.13418.

Daniil A Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, 2023.

Andrés M Bran, Sam Cox, Oliver Schilter, Carlo Baldassari, Andrew D. White, and Philippe Schwaller. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 6:525 – 535, 2023. URL https://api.semanticscholar.org/CorpusID:258059792.

Andres M Bran, Theo A Neukomm, Daniel P Armstrong, Zlatko Jončev, and Philippe Schwaller. Chemical reasoning in llms unlocks steerable synthesis planning and reaction mechanism elucidation, 2025. URL https://arxiv.org/abs/2503.08537.

Ulf N. Brand, Zhengxiao Du, Ali Taheri, and Philippe Schwaller. ChemLLM: A large language model for chemistry. arXiv preprint arXiv:2310.01890, 2023.

Mayk Caldas Ramos, Christopher J. Collison, and Andrew D. White. A review of large language models and autonomous agents in chemistry. *Chemical Science*, 16(6):2514–2572, 2025. doi: 10.1039/D4SC03921A. URL https://pubs.rsc.org/en/content/articlelanding/2025/sc/d4sc03921a. Publisher: Royal Society of Chemistry.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. LEGAL-BERT: The Muppets straight out of law school. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

Daixuan Cheng, Shaohan Huang, and Furu Wei. Adapting large language models via reading comprehension. *arXiv preprint arXiv:2309.09117*, 2023.

Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta: Large-scale self-supervised pretraining for molecular property prediction. *Preprint at https://arxiv.org/abs/2010.09885*, 2020.

Connor W Coley, Wengong Jin, Luke Rogers, Timothy F Jamison, Tommi S Jaakkola, William H Green, Regina Barzilay, and Klavs F Jensen. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.*, 10:370–377, 2019.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

Benedek Fabian, Thomas Edlich, H'el'ena Gaspar, Marwin H. S. Segler, Joshua Meyers, Marco Fiscato, and Mohamed Ahmed. Molecular representation learning with language models and domain-relevant auxiliary tasks. *ArXiv*, abs/2011.13230, 2020. URL https://api.semanticscholar.org/CorpusID:227209142.

Daniel Flam-Shepherd and Alán Aspuru-Guzik. Language models can generate molecules, materials, and protein binding sites directly in three dimensions as xyz, cif, and pdb files, 2023.

Nathan C Frey, Ryan Soklaski, Simon Axelrod, Siddharth Samsi, Rafael G´omez-Bombarelli, Connor W. Coley, and Vijay Gadepally. Neural scaling of deep chemical models. *Nature Machine Intelligence*, 5:1297 – 1305, 2023. URL https://api.semanticscholar.org/CorpusID:262152780.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jiawei Wang, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Shengfeng Ye, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wanjia Zhao, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanhong Xu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025a. URL `https://arxiv.org/abs/2501.12948`.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025b.

Taicheng Guo, Bozhao Nan, Zhenwen Liang, Zhichun Guo, Nitesh Chawla, Olaf Wiest, Xiangliang Zhang, et al. What can large language models do in chemistry? a comprehensive benchmark on eight tasks. *Advances in Neural Information Processing Systems*, 36:59662–59688, 2023.

Stephen R Heller, Alan McNaught, Igor Pletnev, Stephen Stein, and Dmitrii Tchekhovskoi. Inchi, the iupac international chemical identifier. *J. Cheminf.*, 7(1):1–34, 2015.

Masanori Hirano and Kentaro Imajo. Construction of domain-specified japanese large language model for finance through continual pre-training. In *16th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2024.

Jason Holmes, Zhengliang Liu, Lian Zhang, Yuzhen Ding, Terence T. Sio, Lisa A. McGee, Jonathan B. Ashman, et al. Evaluating large language models on a highly-specialized topic, radiation oncology physics. *Frontiers in Oncology*, 2023.

Sheng-Kai Hsieh. Continual pre-training is (not) what you need in domain adaption. *arXiv preprint arXiv:2501.01234*, 2025.

Kevin Maik Jablonka, Philippe Schwaller, Andres Ortega-Guerrero, and Berend Smit. Leveraging large language models for predictive chemistry. *Nat. Mac. Intell.*, 6:161–169, 2024. URL `https://api.semanticscholar.org/CorpusID:267538205`.

Zhengbao Jiang, Jun Araki, Haibo Ding, and Graham Neubig. How can we know when language models know? on the calibration of language models for question answering. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 4661–4673, 2021.

Saurav Kadavath, Tom Conerly, Amanda Askell, Yuntao Bai, Deep Ganguli, Danny Hernandez, Nicholas Schiefer, et al. Language models (mostly) know what they know. *arXiv preprint arXiv:2207.05221*, 2022.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. PubChem 2025 update. *Nucleic Acids Research*, 53(D1):D1516–D1525, January 2025. ISSN 1362-4962. doi: 10.1093/nar/gkae1059. URL https://doi.org/10.1093/nar/gkae1059.

Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. Self-Referencing Embedded Strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn.: Sci. Technol.*, 1:045024, 2020.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, and Sushant Prakash. RLAIF vs. RLHF: Scaling Reinforcement Learning from Human Feedback with AI Feedback, September 2024. URL http://arxiv.org/abs/2309.00267. arXiv:2309.00267 [cs].

Liunian Harold Li, Jack Hessel, Youngjae Yu, Xiang Ren, Kai-Wei Chang, and Yejin Choi. Symbolic chain-of-thought distillation: Small models can also "think" step-by-step. In *Annual Meeting of the Association for Computational Linguistics*, 2023. URL https://api.semanticscholar.org/CorpusID:259251773.

Shuyue Stella Li, Vidhisha Balachandran, Shangbin Feng, Jonathan S. Ilgen, Emma Pierson, Pang Wei Koh, and Yulia Tsvetkov. Mediq: Question-asking llms and a benchmark for reliable interactive clinical reasoning. In *Neural Information Processing Systems*, 2024a. URL https://api.semanticscholar.org/CorpusID:270219405.

Sihang Li, Jian Huang, Jiaxi Zhuang, Yaorui Shi, Xiaochen Cai, Mingjun Xu, Xiang Wang, Linfeng Zhang, and Guolin Ke. SciLitLLM: How to adapt LLMs for scientific literature understanding. *arXiv preprint arXiv:2408.04567*, 2024b.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, et al. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110*, 2022.

Bo Liu, Liming Zhan, Zexin Lu, Yujie Feng, Lei Xue, and Xiao-Ming Wu. How good are llms at out-of-distribution detection?, 2024a. URL https://arxiv.org/abs/2308.10261.

Dancheng Liu, Amir Nassereldine, Ziming Yang, Chenhui Xu, Yuting Hu, Jiajie Li, Utkarsh Kumar, Changjae Lee, and Jinjun Xiong. Large language models have intrinsic self-correction ability. *ArXiv*, abs/2406.15673, 2024b. URL https://api.semanticscholar.org/CorpusID:270703467.

Daniel M. Lowe, Peter T. Corbett, Peter Murray-Rust, and Robert C. Glen. Chemical name to structure: Opsin, an open source solution. *J. Chem. Inf. Model.*, 51(3):739–753, 2011. doi: 10.1021/ci100384d.

Wei Lu, Rachel K. Luu, and Markus J. Buehler. Fine-tuning large language models for domain adaptation: Exploration of training strategies, scaling, model merging and synergistic capabilities. *npj Computational Materials*, 2025.

Łukasz Maziarka, Krzysztof Rataj, Tomasz Danel, Piotr Warchoł, and Stanisław Jastrzębski. ChemBERTa-2: Large-scale self-supervised pre-training for molecules. *arXiv preprint arXiv:2309.12948*, 2023.

N. F. McGlynn. Thinking fast and slow. *Australian veterinary journal*, 92 12:N21, 2014. URL https://api.semanticscholar.org/CorpusID:36031679.

Corina Meister and Ryan Cotterell. Language model evaluation beyond perplexity. *arXiv preprint arXiv:2106.04638*, 2021.

Norman Meuschke, Apurva Jagdale, Timo Spinde, Jelena Mitrović, and Bela Gipp. *A Benchmark of PDF Information Extraction Tools Using a Multi-task and Multi-domain Evaluation Framework for Academic Documents*, pp. 383–405. Springer Nature Switzerland, 2023. ISBN 9783031280320. doi: 10.1007/978-3-031-28032-0_31. URL http://dx.doi.org/10.1007/978-3-031-28032-0_31.

Adrian Mirza, Nawaf Alampara, Sreekanth Kunchapu, Martiño Ríos-García, Benedict Emoekabu, Aswanth Krishnan, Tanya Gupta, Mara Schilling-Wilhelmi, Macjonathan Okereke, Anagha Aneesh, et al. Are large language models superhuman chemists? *arXiv preprint arXiv:2404.01475*, 2024a.

Adrian Mirza, Nawaf Alampara, Sreekanth Kunchapu, Martiño Ríos-García, Benedict Emoekabu, Aswanth Krishnan, Tanya Gupta, Mara Schilling-Wilhelmi, Macjonathan Okereke, Anagha Aneesh, Amir Mohammad Elahi, Mehrdad Asgari, Juliane Eberhardt, Hani M. Elbeheiry, María Victoria Gil, Maximilian Greiner, Caroline T. Holick, Christina Glaubitz, Tim Hoffmann, Abdelrahman Ibrahim, Lea C. Klepsch, Yannik Köster, Fabian Alexander Kreth, Jakob Meyer, Santiago Miret, Jan Matthias Peschel, Michael Ringleb, Nicole Roesner, Johanna Schreiber, Ulrich S. Schubert, Leanne M. Stafast, Dinga Wonanke, Michael Pieler, Philippe Schwaller, and Kevin Maik Jablonka. Are large language models superhuman chemists?, November 2024b. URL `http://arxiv.org/abs/2404.01475`. arXiv:2404.01475 [cs].

Guilherme Penedo, Hynek Kydlíček, Loubna Ben allal, Anton Lozhkov, Margaret Mitchell, Colin Raffel, Leandro Von Werra, and Thomas Wolf. The FineWeb Datasets: Decanting the Web for the Finest Text Data at Scale, October 2024. URL `http://arxiv.org/abs/2406.17557`. arXiv:2406.17557.

Pouya Pezeshkpour. Measuring and modifying factual knowledge in large language models. In *Proceedings of the 22nd International Conference on Machine Learning and Applications*, pp. 992–999, 2023.

RDKit, online. RDKit: Open-source cheminformatics. `http://www.rdkit.org`, 2023.

Baptiste Rozière, Guillaume Lample, Gautier Izacard, Jean Simón, Alexis Palmer, Shuyin Ruan, Myle Ott Nguyen, Nathan Scales, et al. Code Llama: Open foundation models for code. Technical report, Meta AI, 2023.

Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A Hunter, Costas Bekas, and Alpha A Lee. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS Cent. Sci.*, 5(9):1572–1583, 2019.

Philippe Schwaller, Riccardo Petraglia, Valerio Zullo, Vishnu H Nair, Rico Andreas Haeuselmann, Riccardo Pisoni, Costas Bekas, Anna Iuliano, and Teodoro Laino. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.*, 11: 3316–3325, 2020.

Junhong Shen, Neil Tenenholtz, James Hall, David Alvarez-Melis, and Nicoló Fusi. Tag-LLM: Repurposing general-purpose LLMs for specialized domains. *arXiv preprint arXiv:2402.07927*, 2024.

Jeongwoo Shin, Chunting Wang, Zhixuan Yu, Manling Ho, Jeremy R. Smith, Chris Pugh, Hannaneh Hajishirzi, Mari Ostendorf, Ali Farhadi, and Wen-tau Yih. Biomegatron: Larger biomedical domain language model. *arXiv preprint arXiv:2010.09889*, 2020.

Liangtai Sun, Danyu Luo, Da Ma, Zihan Zhao, Zhe-Wei Shen, Su Zhu, Lu Chen, Xin Chen, and Kai Yu. SciDFM: A large language model with mixture-of-experts for science. *arXiv preprint arXiv:2409.01234*, 2024.

Matthew C. Swain and Jacqueline M. Cole. Chemdataextractor: A toolkit for automated extraction of chemical information from the scientific literature. *Journal of Chemical Information and Modeling*, 56(10):1894–1904, 2016. doi: 10.1021/acs.jcim.6b00207. URL `https://doi.org/10.1021/acs.jcim.6b00207`. PMID: 27669338.

Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. Galactica: A large language model for science, 2022. URL `https://arxiv.org/abs/2211.09085`.

Haorui Wang, Marta Skreta, Cher-Tian Ser, Wenhao Gao, Lingkai Kong, Felix Strieth-Kalthoff, Chenru Duan, Yuchen Zhuang, Yue Yu, Yanqiao Zhu, et al. Efficient evolutionary search over chemical space with large language models. *arXiv preprint arXiv:2406.16976*, 2024a.

Pei Wang, Keqing He, Yejie Wang, Xiaoshuai Song, Yutao Mou, Jingang Wang, Yunsen Xian, Xunliang Cai, and Weiran Xu. Beyond the known: Investigating llms performance on out-of-domain intent detection. In *International Conference on Language Resources and Evaluation*, 2024b. URL https://api.semanticscholar.org/CorpusID:268032564.

Weixuan Wang, Barry Haddow, Alexandra Birch, and Wei Peng. Assessing the reliability of large language model knowledge. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 1234–1249, 2024c.

Yiping Wang, Qing Yang, Zhiyuan Zeng, Liliang Ren, Lucas Liu, Baolin Peng, Hao Cheng, Xuehai He, Kuan Wang, Jianfeng Gao, Weizhu Chen, Shuohang Wang, Simon Shaolei Du, and Yelong Shen. Reinforcement Learning for Reasoning in Large Language Models with One Training Example, April 2025. URL http://arxiv.org/abs/2504.20571. arXiv:2504.20571 [cs].

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

David Weininger. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, 28:31–36, 1988.

Wanjiang Weng, Xiaofeng Tan, Hongsong Wang, and Pan Zhou. Realign: Bilingual text-to-motion generation via step-aware reward-guided alignment. 2025. URL https://api.semanticscholar.org/CorpusID:278394818.

Colin White, Samuel Dooley, Manley Roberts, Arka Pal, Ben Feuer, Siddhartha Jain, Ravid Shwartz-Ziv, Neel Jain, Khalid Saifullah, Siddartha Naidu, et al. Livebench: A challenging, contamination-free benchmark for large language models. *arXiv preprint arXiv:2403.12345*, 2024.

Yingce Xia, Peiran Jin, Shufang Xie, Liang He, Chuan Cao, Renqian Luo, Guoqing Liu, Yue Wang, Zequn Liu, Yuan-Jyue Chen, Zekun Guo, Yeqi Bai, Pan Deng, Yaosen Min, Ziheng Lu, Hongxia Hao, Han Yang, Jielan Li, Chang Liu, Jia Zhang, Jianwei Zhu, Ran Bi, Kehan Wu, Wei Zhang, Kaiyuan Gao, Qizhi Pei, Qian Wang, Xixian Liu, Yanting Li, Houtian Zhu, Yeqing Lu, Mingqian Ma, Zun Wang, Tian Xie, Krzysztof Maziarz, Marwin Segler, Zhao Yang, Zilong Chen, Yu Shi, Shuxin Zheng, Lijun Wu, Chen Hu, Peggy Dai, Tie-Yan Liu, Haiguang Liu, and Tao Qin. Nature language model: Deciphering the language of nature for scientific discovery, 2025. URL https://arxiv.org/abs/2502.07527.

Tong Xie, Yuwei Wan, Wei Huang, Zhenyu Yin, Yixuan Liu, Shaozhou Wang, Qingyuan Linghu, Chunyu Kit, Clara Grazian, Wenjie Zhang, and Bram Hoex. Darwin series: Domain specific large language models for natural science. *arXiv preprint arXiv:2308.09913*, 2023a.

Tong Xie, Yuwei Wan, Wei Huang, Zhenyu Yin, Yixuan Liu, Shaozhou Wang, Qingyuan Linghu, Chunyu Kit, Clara Grazian, Wenjie Zhang, Imran Razzak, and Bram Hoex. Darwin series: Domain specific large language models for natural science. *ArXiv*, abs/2308.13565, 2023b. URL https://api.semanticscholar.org/CorpusID:274142505.

Yong Xie, Karan Aggarwal, and Aitzaz Ahmad. Efficient continual pre-training for building domain specific large language models. In *Findings of the Association for Computational Linguistics (ACL)*, 2024a.

Yuxi Xie, Anirudh Goyal, Wenyue Zheng, Min-Yen Kan, Timothy P. Lillicrap, Kenji Kawaguchi, and Michael Shieh. Monte carlo tree search boosts reasoning via iterative preference learning. *ArXiv*, abs/2405.00451, 2024b. URL https://api.semanticscholar.org/CorpusID:269484186.

Zonglin Yang, Wanhao Liu, Ben Gao, Tong Xie, Yuqiang Li, Wanli Ouyang, Soujanya Poria, Erik Cambria, and Dongzhan Zhou. Moose-chem: Large language models for rediscovering unseen chemistry scientific hypotheses, 2025. URL https://arxiv.org/abs/2410.07076.

Fanghua Ye, Mingming Yang, Jianhui Pang, Longyue Wang, Derek F. Wong, Emine Yilmaz, Shuming Shi, and Zhaopeng Tu. Benchmarking large language models via uncertainty quantification. *arXiv preprint arXiv:2401.12321*, 2024.

Botao Yu, Frazier N. Baker, Ziqi Chen, Xia Ning, and Huan Sun. Llasmol: Advancing large language models for chemistry with a large-scale, comprehensive, high-quality instruction tuning dataset, 2024a. URL https://arxiv.org/abs/2402.09391.

Botao Yu, Frazier N Baker, Ziqi Chen, Xia Ning, and Huan Sun. Llasmol: Advancing large language models for chemistry with a large-scale, comprehensive, high-quality instruction tuning dataset. *arXiv preprint arXiv:2402.09391*, 2024b.

Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Yang Yue, Shiji Song, and Gao Huang. Does reinforcement learning really incentivize reasoning capacity in llms beyond the base model?, 2025. URL https://arxiv.org/abs/2504.13837.

Di Zhang, Wei Liu, Qian Tan, Jingdan Chen, Hang Yan, Yuliang Yan, Jiatong Li, Weiran Huang, Xiangyu Yue, Wanli Ouyang, Dongzhan Zhou, Shufei Zhang, Mao Su, Han-Sen Zhong, and Yuqiang Li. Chemllm: A chemical large language model, 2024. URL https://arxiv.org/abs/2402.06852.

Siyan Zhao, Tung Nguyen, and Aditya Grover. Probing the decision boundaries of in-context learning in large language models. *arXiv preprint arXiv:2406.01234*, 2024.