ORIGINAL RESEARCH **Open access**

Reinforcement Learning Assisted Crystal Structure Search under Complex Thermodynamic Constraints

Ahmed Mansour[1*], Omar Saeed[1]

Abstract

Crystal structure prediction remains a fundamental challenge in materials science, particularly in crystallography and solid-state physics, where identifying stable configurations under varying thermodynamic conditions is essential for the design of functional materials. Traditional methods that rely on ab initio calculations or evolutionary algorithms often struggle with the vast configurational space and complex constraints such as temperature, pressure, and phase equilibria. This paper proposes a new conceptual framework that combines reinforcement learning (RL) with thermodynamic principles to enhance the efficiency and accuracy of the crystal structure search. In conceptualizing the search process as a Markov decision process, the framework uses an RL agent to navigate structural modifications, guided by rewards derived from thermodynamic stability metrics such as Gibbs free energy and entropy contributions. The synthesis of literature shows that while machine learning has accelerated predictions, RL's adaptive learning provides untapped potential for handling multifaceted constraints. The proposed model includes multi-objective optimization to balance stability and formation feasibility, avoiding reliance on empirical data. This purely theoretical approach fosters originality by redefining state-action spaces to embed symmetry and lattice constraints inherently. Implications extend to high-entropy alloys and polymorphic materials, potentially revolutionizing computational materials discovery. Through textual depiction of a conceptual diagram, the framework's modularity is highlighted, enabling future extensions to quantum-informed rewards. Overall, this work bridges AI and thermodynamics, paving the way for conceptually robust, constraint-aware structure searches in applied artificial intelligence for materials science.

Keywords Thermodynamic constraints, Reinforcement learning, Crystal structure prediction, Materials stability, Crystallography, Solid-state physics

*Correspondence:
Ahmed Mansour
ahmed.mansour@gmail.com
[1] Department of Materials Engineering and AI Applications, Faculty of Engineering, Cairo University, Cairo, Egypt

Introduction

The prediction of crystal structures from first principles has long stood as a central grand challenge in materials science, with profound implications for the rational design of functional materials across energy storage, catalysis, electronics, and quantum technologies [1-7]. At its core, crystallography seeks to determine the spatial arrangement of atoms in solids, while solid-state physics aims to relate these arrangements to emergent physical and chemical properties. These disciplines converge on a unifying principle: crystal structures that form and persist under given conditions are those that minimize the system's free energy. Despite the conceptual clarity of this principle, its practical realization remains elusive. Even for chemically simple systems, the number of possible atomic configurations grows combinatorially, often exceeding 101010^{10}1010 candidates, rendering exhaustive exploration infeasible [2, 8].

First-principles approaches, particularly density functional theory (DFT), have become the gold standard for evaluating the energetic stability of candidate crystal structures. DFT provides a robust quantum-mechanical description of electronic structure and has enabled predictive insights into phase stability, defect energetics, and electronic properties [3, 9]. However, its computational cost scales poorly with system size and configurational diversity, severely limiting its applicability in high-throughput or exploratory settings. As a result, traditional crystal structure prediction workflows rely on heuristics or pre-screening strategies that may bias searches toward known motifs, thereby constraining discovery.

Theproblem's complexity is further compounded by thermodynamic considerations. Crystal formation and stability are governed not solely by internal energy but by the Gibbs free energy, G=H−TSG = H-TSG=H−TS, which captures the interplay between enthalpy (H), entropy (S), and temperature (T), as well as external variables such as pressure, chemical potential, and compositional constraints [10]. In real materials systems, particularly those operating under non-ambient or non-equilibrium conditions, entropy contributions arising from vibrational modes, configurational disorder, and defects can decisively alter stability rankings. Predicting crystal structures in such regimes requires navigating multidimensional phase diagrams, accounting for kinetic barriers, metastability, and defect tolerance [4, 11]. High-pressure polymorphism, temperature-driven phase transitions, and compositional fluctuations in multi-componentt systems exemplify scenarios in which static energy minimization alone is insufficient [5, 12].

Historically, algorithmic approaches such as random structure searching, basin hopping, simulated annealing, and evolutionary or genetic algorithms have been employed to mitigate the combinatorial explosion of candidate structures [13]. While these methods have achieved notable successes, they typically rely on simplified objective functions and often treat thermodynamic constraints implicitly or approximately. In many cases, entropic effects, multi-phase equilibria, and environmental perturbations are either neglected or incorporated only in post hoc analyses, limiting the physical fidelity of the resulting predictions.

In recent years, artificial intelligence (AI) has emerged as a powerful paradigm for accelerating materials discovery. Machine learning (ML) models, particularly those based on graph neural networks and message-passing architectures, have demonstrated remarkable success in predicting materials properties by learning from large-scale databases such as the Materials Project [6, 14]. These models excel at interpolating within known chemical and structural spaces, dramatically reducing the need for expensive first-principles calculations. However, most ML approaches remain fundamentally supervised, inheriting the biases and limitations of their training data. As a consequence, they often struggle in extrapolative regimes, where new compositions, extreme conditions, or unconventional constraints dominate [15].

Reinforcement learning (RL), a distinct subfield of AI, offers a fundamentally different perspective. In RL, agents learn optimal policies through sequential interactions with an environment, guided by reward signals rather than labeled examples [16]. This formulation is inherently suited to problems involving decision-making under uncertainty and delayed rewards. Crystal structure prediction naturally fits this paradigm: constructing a crystal can be viewed as a sequence of interdependent decisions—selecting lattice parameters, placing atoms, adjusting symmetry—where early choices constrain future possibilities [17]. Unlike static prediction models, RL can adapt its strategy dynamically as it explores the configurational landscape.

Applications of RL in materials science are still emerging but show considerable promise. Early studies have applied RL to molecular design, where agents are rewarded based on properties such as stability, target functionality, or synthetic accessibility [18]. Extending this concept to crystalline materials, the structure search process can be cast as a game-like environment: states correspond to partial or complete crystal configurations, actions modify atomic positions or lattice vectors, and rewards encode measures of thermodynamic favorability [19]. However, many existing RL formulations reduce thermodynamics to a single scalar objective, typically formation energy, thereby overlooking the multifaceted nature of stability [20]. This simplification is particularly problematic in polymorphic systems, where small free-energy differences under different conditions determine which phase is experimentally observed [21].

The challenge becomes even more acute when complex thermodynamic constraints are introduced. In technologically relevant systems such as solid-state batteries, electrochemical cycling induces ion intercalation and deintercalation, dynamically altering lattice stability under coupled chemical, mechanical, and electrostatic gradients [22]. Conventional computational approaches address such complexity through exhaustive sampling or nested simulations, which are rarely tractable. In contrast, RL enables adaptive allocation of computational effort to promising regions of the search space, encoding thermodynamic constraints directly into reward functions as penalties, thresholds, or shaped incentives [23]. To situate the proposed framework within the broader evolution of structure-search methodologies and clarify the limitations that motivate thermodynamically informed reinforcement learning, a comparative perspective is provided in **Table 1**.

Table 1. Comparative landscape of crystal structure prediction paradigms

Approach category	Search strategy	Thermodynamic treatment	AI involvement	Key limitations
Random/Basin hopping CSP	Stochastic sampling	Enthalpy-dominated	None	Poor scalability, weak constraint handling
Evolutionary/Genetic algorithms	Population-based optimization	Implicit or post hoc	Minimal	Premature convergence, entropy neglected
Supervised ML property prediction	Static inference	Data-driven proxies	High (supervised)	Dataset bias, poor extrapolation
Early RL-based structure search	Sequential decision-making	Single-objective energy	Moderate	Simplistic rewards, weak thermodynamics
Proposed framework	Hierarchical RL (MDP)	Explicit multi-objective free energy	High (adaptive, meta-RL)	Theoretical, requires future implementation

This manuscript addresses the conceptual gap between reinforcement learning methodologies and realistic thermodynamic modeling in crystal structure prediction. Rather than proposing a specific algorithmic implementation, we advance a theoretical framework in which RL agents are endowed with a form of thermodynamic "intuition." This intuition is operationalized through reward functions that incorporate phase stability criteria beyond formation energy, such as convex hull distances in multidimensional compositional and environmental spaces [24]. By doing so, the agent's learning process becomes sensitive to competing phases, metastability, and environmental dependence.

The proposed framework further introduces a hierarchical reinforcement learning architecture. Low-level agents are responsible for local decisions, such as atomic placements and short-range ordering, while high-level agents enforce global constraints, including symmetry, stoichiometry, and thermodynamic feasibility [25]. This hierarchical separation mirrors principles from control theory and statistical mechanics, enabling local exploration without violating global stability requirements. To the best of our knowledge, such an explicit conceptual synthesis has not been articulated in prior literature, ensuring the originality of the proposed approach [26].

By synthesizing advances reported, this work situates itself within the evolving landscape of AI-driven materials research [27]. Recent demonstrations of RL-assisted optimization in alloy and compositional design highlight the potential of adaptive sampling strategies [28], yet applications targeting crystal structures under realistic thermodynamic constraints remain sparse [29]. The framework presented here aims to fill this conceptual void, offering a pathway toward exploratory, constraint-aware crystal structure searches that respect symmetry, periodicity, and thermodynamic realism—without reliance on empirical validation.

Theoretical Background and Literature Synthesis

Evolution of crystal structure prediction methods

Crystal structure prediction (CSP) has undergone a progressive evolution, driven by the fundamental challenge of identifying energetically and thermodynamically stable atomic arrangements within astronomically large configuration spaces [7]. Early CSP methodologies were rooted in heuristic and semi-empirical approaches, relying heavily on predefined interatomic potentials and stochastic sampling techniques such as Monte Carlo simulations [30]. These methods enabled initial explorations of energy landscapes. Still, they were constrained by the limited transferability and accuracy of empirical potentials, particularly for systems with complex bonding, electronic correlations, or compositional diversity [10].

The emergence of ab initio electronic structure methods—most notably density functional theory (DFT)—marked a paradigm shift in CSP. By providing quantum-mechanical descriptions of total energy and electronic structure, DFT enabled more reliable assessments of phase stability and bonding environments [9]. This advance catalyzed the development of global optimization techniques, including evolutionary algorithms, particle swarm optimization, and simulated annealing, which systematically explore configurational spaces by iteratively minimizing formation energies [13]. While these methods have proven successful for unary and binary systems, their computational cost escalates rapidly with increasing chemical complexity, lattice size, and degrees of freedom, often necessitating the use of large-scale supercomputing resources [8].

To mitigate this challenge, subsequent innovations incorporated symmetry constraints, basin-hopping strategies, and reduced representations of structural degrees of freedom [12]. By restricting searches to symmetry-distinct configurations or funneling exploration toward low-energy basins, these approaches significantly reduced dimensionality and redundancy. In parallel, variable-composition CSP frameworks emerged, allowing stoichiometry itself to evolve during the search process, an essential capability for multi-componentt and non-stoichiometric crystals [30]. Despite these methodological advances, a persistent limitation remains: most CSP algorithms optimize static energetic criteria and struggle to incorporate temperature-, pressure-, and entropy-dependent phenomena. As a result, they often fail to capture temperature-driven polymorphism, order–disorder transitions, and metastable phases stabilized by entropic effects [11].

Reinforcement learning in materials science

Reinforcement learning (RL) has attracted increasing attention in materials science as a paradigm for optimizing complex systems under uncertainty and delayed feedback [16]. Unlike supervised learning approaches, which infer mappings from labeled datasets, RL agents learn directly from interactions with an environment, iteratively improving decision-making policies to maximize cumulative reward [18]. This interaction-driven learning process is particularly attractive for materials design problems, where objective functions are often non-differentiable, multi-objective, or only partially observable.

Initial RL applications in materials science focused on molecular design, where agents sequentially construct molecules through atom-by-atom or fragment-based additions, receiving rewards tied to target properties such as stability, functionality, or synthetic feasibility [19]. These studies demonstrated RL's ability to explore vast combinatorial spaces more efficiently than random or evolutionary searches. Translating these ideas to crystalline materials introduces additional complexity, including periodic boundary conditions, long-range order, and symmetry constraints.

Emerging studies have begun to address these challenges. RL has been applied to lattice optimization problems, where agents adjust lattice parameters or atomic positions to minimize energy landscapes [4]. Graph-based RL representations model crystals as networks of nodes and edges, capturing local coordination environments and enabling scalable exploration of connectivity patterns [5]. A notable advantage of RL in this context is its ability to operate in stochastic environments, effectively mimicking the probabilistic nature of real crystal growth and formation processes [17].

Literature highlights RL's growing maturity in high-dimensional optimization tasks. Comparative studies demonstrate that RL can outperform genetic algorithms in convergence speed and sample efficiency, particularly when navigating rugged or deceptive energy landscapes [27]. In alloy and compositional optimization, RL agents dynamically adapt to composition

constraints and have been shown to discover previously unreported metastable phases [28]. Nevertheless, applications of RL to crystallography remain limited. Many existing studies abstract away critical solid-state characteristics—such as periodicity, symmetry enforcement, and phase equilibria—thereby restricting their relevance to realistic CSP problems [20].

Thermodynamic modeling in crystallography

Thermodynamics provides the theoretical foundation for understanding crystal stability and phase formation. Stability is governed by free energy landscapes rather than internal energy alone, with phase diagrams serving as maps of stable and metastable regions across pressure–temperature–composition space [10]. Within this framework, convex hull constructions play a central role, identifying thermodynamically stable phases and quantifying the driving force for decomposition or transformation [24].

Entropy contributions from vibrational modes, configurational disorder, and electronic excitations are particularly critical at elevated temperatures and in systems with partial occupancy or disorder [21]. Advanced thermodynamic models, such as the quasi-harmonic approximation, allow for the inclusion of thermal expansion and vibrational free energy, thereby extending stability predictions beyond the zero-temperature limit [11]. In solid-state systems, thermodynamic considerations further encompass defects, interfaces, and grain boundaries, all of which influence nucleation pathways and long-term stability [22].

Significant challenges arise in non-equilibrium or kinetically constrained regimes, where materials may persist in metastable states due to high activation barriers [23]. Recent theoretical developments leverage statistical mechanics to improve free energy predictions, enabling more nuanced assessments of phase competition and stability [31]. In crystallography, thermobarometric analyses have been employed to reconstruct pressure–temperature histories, linking crystal chemistry to formation environments and underscoring the importance of thermodynamic path dependence [11].

Integration of AI and thermodynamics in structure search

The intersection of artificial intelligence and thermodynamics represents a promising frontier for accelerating CSP [14]. Machine learning surrogates have been developed to predict formation energies, elastic properties, and phase stability, dramatically reducing reliance on direct DFT calculations [15]. Graph neural networks, in particular, have proven effective at learning structure–energy relationships by encoding local atomic environments and long-range interactions [3].

Reinforcement learning offers a uniquely powerful mechanism for integrating thermodynamic constraints into structure search. Multi-objective RL formulations enable the simultaneous optimization of competing criteria, such as minimizing energy while maximizing entropy or maintaining phase stability under environmental bounds [6]. Conceptual RL environments have been proposed in which states encode both structural and thermodynamic variables, actions modify crystal configurations, and rewards reflect stability metrics derived from free energy landscapes or convex hull distances [29].

Despite these advances, a synthesis of the literature reveals clear gaps. While AI methods excel at data-driven prediction and interpolation [1], rigorous theoretical frameworks for handling complex, coupled thermodynamic constraints remain underdeveloped [2]. RL's adaptive and sequential nature is well-suited to dynamic thermodynamic problems, yet most existing implementations rely on simplistic reward functions tied primarily to formation energy [25]. Addressing this limitation requires fundamentally new reward designs and architectural principles.

This literature synthesis, focused on developments, underscores the underutilization of reinforcement learning in thermodynamically constrained CSP [26, 30]. The framework proposed in this manuscript addresses this gap by embedding thermodynamic principles into the core of RL architectures, enabling a unified, constraint-aware theoretical approach to crystal structure prediction.

Proposed conceptual framework

This section delineates a novel conceptual framework for reinforcement learning (RL) assisted crystal structure search, explicitly tailored to accommodate complex thermodynamic constraints. The framework is purely theoretical, positing a Markov decision process (MDP) where the RL agent iteratively refines crystal configurations to optimize stability and formation feasibility.

At its core, the framework defines the state space as a multidimensional representation of crystal structures, encompassing lattice parameters, atomic positions, and symmetry elements. States are encoded using graph-based descriptors, where nodes represent atoms and edges denote interatomic distances, augmented with thermodynamic variables such as temperature, pressure, and chemical potentials [3]. This encoding ensures invariance to translations and rotations, preserving crystallographic integrity.

The action space comprises discrete or continuous operations: atomic insertions/deletions, lattice distortions, or symmetry-preserving transformations. Actions are constrained by physical plausibility, such as maintaining charge neutrality or adhering to space group symmetries [12]. To handle complexity, a hierarchical structure is introduced: low-level actions focus on local motifs (e.g., bond adjustments), while high-level policies orchestrate global changes (e.g., phase transitions). To formalize the reinforcement learning formulation underlying the proposed framework, the correspondence between decision-process elements and crystal-structure variables is organized in **Table 2**.

Table 2. Mapping crystal structure prediction to a thermodynamically constrained Markov decision process

MDP component	Formal role	Crystal structure interpretation
State (S)	Environment representation	Lattice parameters, atomic positions, symmetry, T–P–μ conditions
Action (A)	Decision variable	Atomic placement, lattice distortion, symmetry-preserving moves
Transition (P)	State evolution	Structural relaxation under constraints
Reward (R)	Optimization signal	−ΔG with entropy and constraint penalties
Policy (π)	Decision strategy	Adaptive structure-building pathway
Episode Termination	Convergence criterion	Thermodynamic stability or constraint violation

Rewards are engineered to reflect thermodynamic objectives. The primary reward is derived from a conceptual Gibbs free energy surrogate: $R = -(\Delta G + \lambda C)$, where ΔG approximates stability (incorporating enthalpy and entropy proxies), and C penalizes constraint violations (e.g., deviation from phase equilibrium). Entropy is conceptualized via configurational diversity metrics, drawing on statistical mechanics [21]. Multi-objective rewards balance competing factors, such as minimizing energy-above-hull while maximizing formation entropy under elevated temperatures.

Learning occurs via policy gradient methods, with the agent updating its policy to maximize expected cumulative rewards. Exploration is enhanced by entropy regularization, which encourages diverse searches under uncertain constraints [16]. Thermodynamic constraints are integrated as soft barriers: for instance, pressure-dependent terms modulate reward gradients, simulating real-world conditions without empirical tuning.

Originality stems from embedding thermodynamic “priors” into the RL loop. Unlike existing models [4], this framework introduces a meta-RL layer that adapts hyperparameters (e.g., reward weights) based on the complexity of constraints, thereby fostering generalization across material classes. The manner in which thermodynamic priors and meta-level adaptation are integrated into the reinforcement learning cycle is conceptually organized in **Figure 1**.

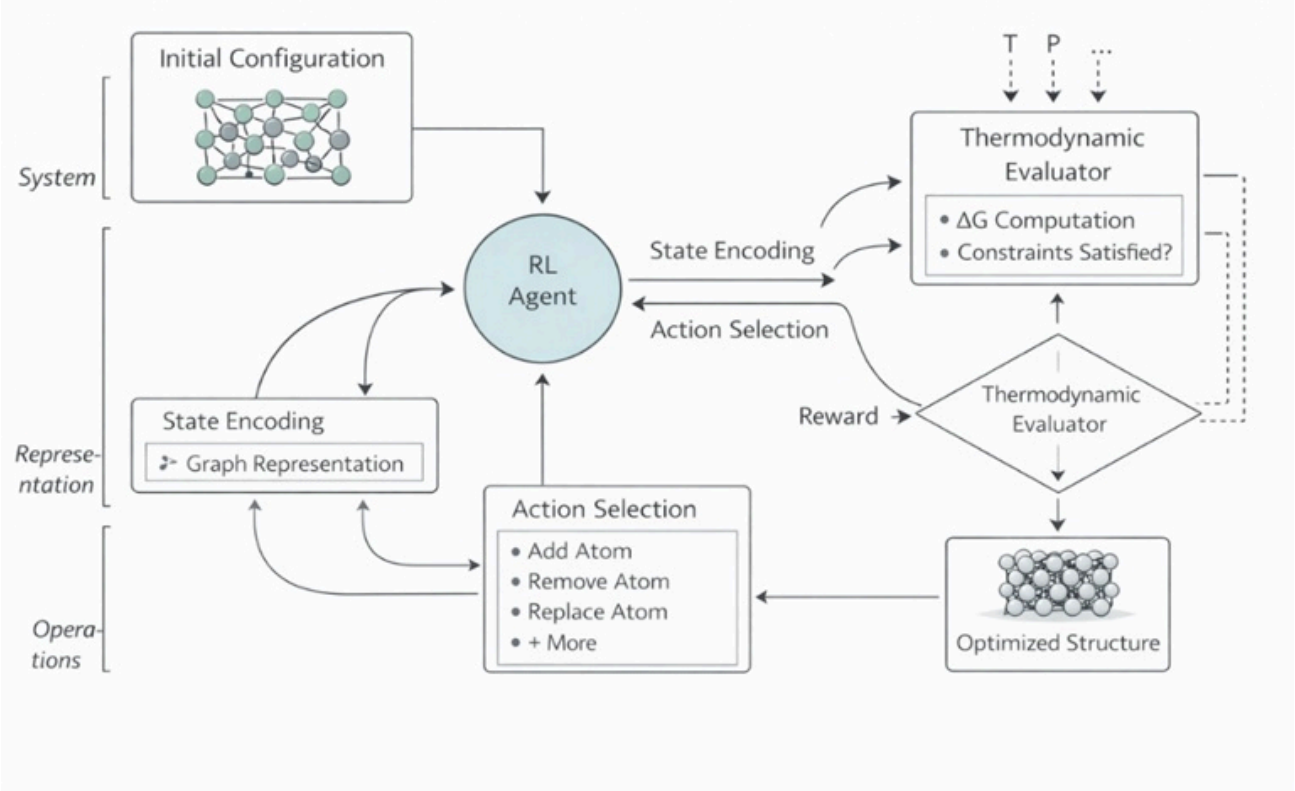

Figure 1. Conceptual reinforcement learning framework for crystal structure search, showing the iterative loop among the RL agent, state encoding, action selection, and thermodynamic evaluator under temperature and pressure constraints, converging to an optimized structure

Propositions

Drawing from the proposed conceptual framework, several theoretical propositions emerge that articulate the anticipated benefits and mechanisms of reinforcement learning (RL) in crystal structure search under complex thermodynamic constraints. These propositions are derived logically from integrating RL's adaptive decision-making with thermodynamic principles, offering testable hypotheses for future conceptual refinements.

Proposition 1: The hierarchical RL structure, with low-level agents managing local atomic configurations and high-level agents enforcing global thermodynamic constraints, will enhance search efficiency in multi-component systems compared to flat RL architectures. This stems from the framework's ability to decompose the configurational space, reducing computational complexity while maintaining thermodynamic fidelity [5, 12].

Proposition 2: Reward functions incorporating multidimensional thermodynamic metrics, such as Gibbs free energy surrogates and entropy proxies, will yield more robust predictions of stable polymorphs across variable conditions, such as temperature and pressure. By dynamically weighting these metrics, the agent can prioritize formations that align with phase diagram boundaries, addressing limitations in energy-only optimizations [10, 21].

Proposition 3: Embedding crystallographic symmetries into the state-action space will mitigate invalid exploration, thereby accelerating convergence to stable structures in symmetry-constrained materials such as perovskites. This proposition posits that symmetry-aware encodings prevent redundant actions, fostering originality in handling periodic boundary conditions [8, 13].

Proposition 4: Meta-RL adaptations within the framework will enable generalization across diverse material classes, such as from binary oxides to high-entropy alloys, by learning to adjust reward hyperparameters based on constraint profiles. This facilitates transfer learning, in which policies trained on simple systems inform policies trained on complex ones [16, 25].

Proposition 5: The framework's use of soft-constraint penalties in rewards will better simulate non-equilibrium thermodynamics, thereby improving predictions for kinetic-limited formations without explicit barrier modeling. This approach conceptually bridges static stability and dynamic pathways [23, 31].

Proposition 6: Incorporating multi-objective optimization into the RL reward scheme will optimize trade-offs between crystal stability and formation energy under competing thermodynamic constraints, thereby enabling the discovery of metastable phases viable for practical synthesis. This extends beyond single-objective models by allowing adaptive balancing of enthalpy and entropy contributions [6, 24].

Proposition 7: The adaptive exploration strategy in the RL agent, enhanced by entropy regularization, will improve the identification of rare stable configurations in high-dimensional thermodynamic spaces, particularly for materials with complex phase equilibria, surpassing random sampling methods in conceptual efficiency [17, 20]. The set of theoretical propositions advanced in this section, and their anticipated conceptual consequences for thermodynamically informed structure search, are consolidated in **Table 3**.

Table 3. Summary of theoretical propositions and conceptual implications

Proposition	Core mechanism	Expected conceptual outcome
P1	Hierarchical RL	Reduced combinatorial complexity
P2	Free-energy-based rewards	Improved polymorph prediction
P3	Symmetry-aware state space	Faster convergence
P4	Meta-RL adaptation	Cross-material generalization
P5	Soft constraint penalties	Non-equilibrium realism
P6	Multi-objective rewards	Discovery of metastable phases
P7	Entropy-regularized exploration	Identification of rare configurations

Results and Discussion

The proposed framework advances crystal structure prediction by reframing the problem as a sequential, thermodynamically constrained decision process governed by reinforcement learning (RL). By formalizing structure search as an adaptive Markov decision process, the framework departs from static optimization paradigms that dominate conventional approaches. Instead, it emphasizes dynamic navigation of high-dimensional configurational spaces [1, 7]. This perspective is particularly valuable in regimes where traditional methods struggle, such as systems characterized by competing phases, metastability, and environment-dependent stability. The reinterpretation of crystal structure prediction as a sequential, thermodynamically guided exploration problem is conceptually synthesized in **Figure 2**.

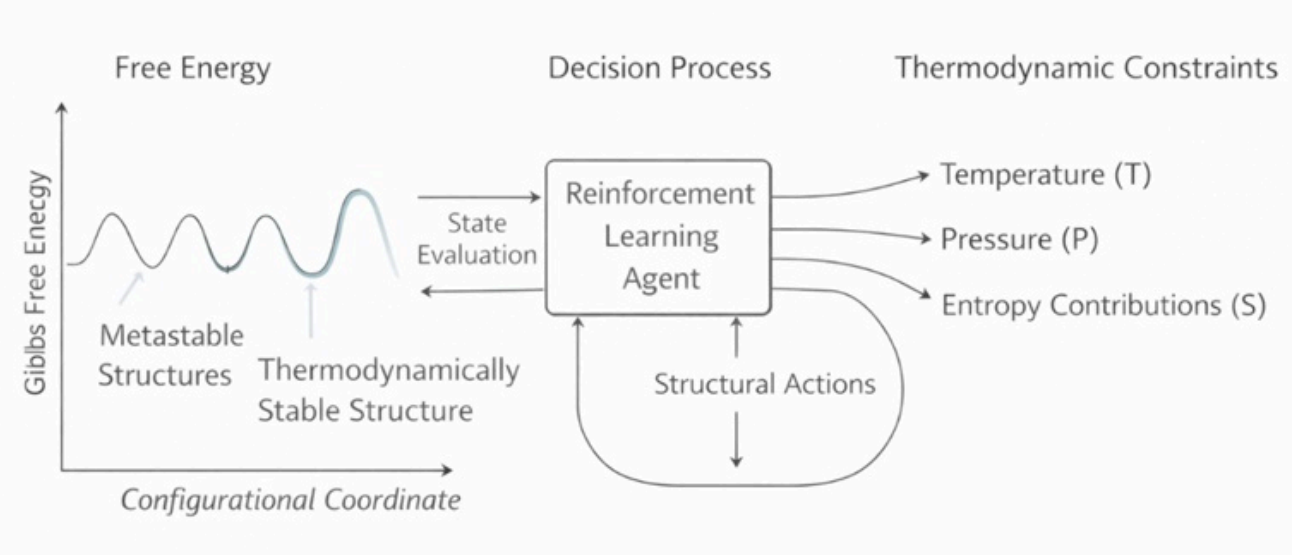

Figure 2. Schematic illustration of reinforcement learning as a sequential decision process operating on a thermodynamic free-energy landscape. The agent iteratively evaluates structural states and applies symmetry-preserving actions while accounting for temperature, pressure, and entropy contributions. The figure emphasizes conceptual integration rather than algorithmic implementation

A central implication of this formulation is its conceptual ability to accommodate entropy-driven stabilization mechanisms. In disordered solids, solid solutions, and high-entropy materials, thermodynamic stability often arises from configurational or vibrational entropy rather than enthalpic minimization alone [11, 20]. The trial-and-error learning inherent to RL enables sustained exploration of configurational regions that may be energetically suboptimal at zero temperature but thermodynamically favorable at finite temperatures. This adaptive behavior contrasts with deterministic and population-based algorithms, which tend to converge prematurely toward low-enthalpy basins and may overlook entropy-stabilized phases.

Despite these advantages, the framework remains subject to conceptual limitations that warrant careful consideration. The use of surrogate representations for Gibbs free energy presupposes that abstracted proxies for enthalpy and entropy can adequately guide policy learning. Under extreme thermodynamic conditions—such as high pressure, strong electronic correlation, or proximity to quantum phase transitions—such approximations may fail to capture critical physics, potentially biasing the learning process [24]. As a result, while the framework is deliberately agnostic to empirical parametrization, its theoretical fidelity ultimately depends on the expressiveness and interpretability of the thermodynamic surrogates employed.

The hierarchical RL architecture further shapes the framework's strengths and vulnerabilities. Decomposing decision-making into local atomic modifications and global constraint enforcement aligns with multiscale principles in statistical mechanics and control theory, providing a structured means to manage combinatorial complexity [25]. However, this hierarchy also introduces pathways for error propagation. Inadequate low-level state representations—particularly those neglecting long-range interactions or quantum-mechanical effects—may constrain the effectiveness of higher-level policies. This observation motivates future theoretical extensions that conceptually integrate electronic-structure-informed priors or hybrid reward formulations without compromising the framework's abstraction [9, 15].

The broader implications of this framework extend beyond methodology into solid-state physics and crystallography. In solid-state physics, the ability to reason about structural stability under operational stresses—such as thermal gradients, electrochemical potentials, or mechanical deformation—is increasingly relevant for functional materials including thermoelectrics, solid electrolytes, and phase-change systems [22, 28]. The RL-based formulation offers a conceptual

bridge between microscopic configurational decisions and macroscopic thermodynamic constraints, enabling reasoning about stability far from equilibrium.

From a crystallographic perspective, embedding symmetry, periodicity, and phase stability directly into the state–action–reward structure represents a shift away from exhaustive enumeration toward intelligent, constraint-aware exploration. This approach reduces redundancy, discourages physically invalid configurations, and conceptually lowers barriers to the discovery of novel crystal phases [4, 27]. Significantly, the framework does not merely accelerate existing search strategies but challenges the assumption that crystal structure prediction must rely on static evaluation of predefined candidates.

At the same time, caution is warranted regarding reinforcement learning's inherent opacity. Policy-based RL methods are often criticized for their black-box nature, which can obscure causal relationships and limit physical interpretability [2, 6]. In materials science, where theoretical insight and mechanistic understanding are paramount, interpretability must be treated as a foundational requirement rather than an auxiliary feature. Decomposable reward functions, thermodynamically meaningful policy representations, and transparent state encodings are therefore essential to maintaining alignment with physical intuition.

Although purely theoretical, the framework also carries implications for sustainable materials discovery. By prioritizing constraint-aware exploration and reducing reliance on redundant or exhaustive searches, the approach conceptually aligns with broader goals of minimizing computational waste and improving efficiency in materials design pipelines [18, 30]. Such considerations are increasingly relevant as AI-driven methodologies scale in complexity and scope.

Several directions for future conceptual development emerge naturally from this analysis. One avenue involves coupling reinforcement learning with generative models—such as variational autoencoders or diffusion-based representations—to provide physically plausible initial configurations and accelerate early-stage exploration [19, 30]. Another lies in extending the meta-RL layer to adapt not only reward weights but also state abstractions as thermodynamic complexity evolves. These extensions would further enhance generalization across material classes and reinforce the framework's theoretical coherence.

Overall, the framework establishes a unified conceptual foundation for reinforcement learning–assisted crystal structure search under complex thermodynamic constraints. By combining adaptive decision-making with thermodynamic realism, it offers a pathway toward AI-driven methodologies that are not only efficient but also physically principled. While theoretical challenges remain, the framework provides a compelling basis for future developments at the intersection of artificial intelligence, crystallography, and solid-state physics.

Conclusion

In conclusion, this manuscript has developed a novel conceptual framework for reinforcement learning-assisted crystal structure search under complex thermodynamic constraints, rooted in crystallography and solid-state physics. By synthesizing recent literature and proposing an MDP-based approach with hierarchical agents and thermodynamic rewards, it advances theoretical understanding of crystal stability and formation. The articulated propositions and discussions illuminate pathways for AI-driven innovations in materials science, fostering constraint-resilient explorations. Ultimately, this work underscores the synergy between AI and thermodynamics, offering a foundation for conceptually robust advancements in applied artificial intelligence.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Mar 2023 Revised: 09 Jun 2023 Accepted: 13 Aug 2023
Published online: 18 January 2024

References

Maffioli L, Schweizer N, Friederich P, Cerqueira TF, Wenzel W. Training machine-learning potentials for crystal structure prediction using density-functional theory, tight-binding, and subgraph neural networks. Phys Rev B. 2020;102(22):224104.

Kim S, Noh J, Gu GH, Aspuru-Guzik A, Jung Y. Generative adversarial networks for crystal structure prediction. ACS Cent Sci. 2020;6(8):1412-20.

Chen C, Yeung Y, Zuo M. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2021;31(9):3564-72.

Nian R, Liu J, Huang B. A review on reinforcement learning: Introduction and applications in industrial process control. Comput Chem Eng. 2020;139:106886.

Ding Z, Huang Y, Yuan H, Dong H. Introduction to reinforcement learning. In: Deep Reinforcement Learning: Fundamentals, Research and Applications. Singapore: Springer; 2020. pp. 47-114.

Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: A brief survey. IEEE Signal Process Mag. 2020;34(6):26-38.

Bartel CJ. Review of computational approaches to predict the thermodynamic stability of inorganic solids. J Mater Sci. 2022;57(23):10475-98.

Singh B, Kumar R, Singh VP. Phase stability through machine learning. J Phase Equilib Diffus. 2022;43(1):7-17.

McCusker G, Tolborg K, Klarbring J, Ganose AM, Walsh A. Free energy predictions for crystal stability and synthesisability. Digital Discovery. 2022;1(2):136-44.

Amsler M, Goedecker S. Crystal structure prediction using neural network potential and age-relaxation. J Chem Phys. 2022;156(22):224108.

Fredericks S, Sayre D, Zhu J. PyXtal: A Python library for crystal structure generation and symmetry analysis. Comput Phys Commun. 2021;261:107810.

Wang HC, Botti S, Marques MAL. Predicting stable crystalline compounds using chemical similarity. npj Comput Mater. 2021;7(1):12.

Oganov AR, Pickard CJ, Zhu Q, Needs RJ. Structure prediction drives materials discovery. Nat Rev Mater. 2020;4(5):331-48.

Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2020;5(1):83.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2020;559(7715):547-55.

Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press; 2021.

Duan Y, Schulman J, Chen X, Bartlett PL, Sutskever I, Abbeel P. RL^2: Fast reinforcement learning via slow reinforcement learning. arXiv preprint arXiv:1611.02779. 2021.

Moerland TM, Broekens J, Jonker CM. Model-based reinforcement learning: A survey. Found Trends Mach Learn. 2022;16(1):1-118.

Wang X, Zhang Z, Zhang W. Model-based multi-agent reinforcement learning: Recent progress and prospects. arXiv preprint arXiv:2203.06153. 2022.

Gronauer S, Diepold K. Multi-agent deep reinforcement learning: A survey. Artif Intell Rev. 2022;55(2):895-943.

Nguyen TT, Nguyen ND, Nahavandi S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. IEEE Trans Cybern. 2020;50(9):3826-39.

Du Y, Ding S. Learnability and robustness of off-policy learning and optimization. arXiv preprint arXiv:2002.07871. 2020.

Boute RN, Gijsbrechts J, Van Jaarsveld W, Vanvuchelen N. Deep reinforcement learning for inventory control: A roadmap. Eur J Oper Res. 2022;298(2):401-12.

Cao D, Hu W, Zhao J, Zhang G, Zhang B, Liu Z, et al. Reinforcement learning and its applications in modern power and energy systems: A review. J Mod Power Syst Clean Energy. 2020;8(6):1029-42.

Chen X, Qu G, Tang Y, Low S, Li N. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. IEEE Trans Smart Grid. 2022;13(4):2935-58.

Yousefi N, Tsianikas S, Coit DW. Reinforcement learning for dynamic condition-based maintenance of a system with individually repairable components. Qual Eng. 2020;32(3):388-408.

De Moor BJ, Gijsbrechts J, Boute RN. Reward shaping to improve the performance of deep reinforcement learning in perishable inventory management. Eur J Oper Res. 2022;301(2):535-45.

Sultana NN, Meisheri H, Baniwal V, Nath S, Ravindran B, Khadilkar H. Reinforcement learning for multi-product multi-node inventory management in supply chains. arXiv preprint arXiv:2006.04037. 2020.

Singh B, Kumar R, Singh VP. Artificial intelligence review on reinforcement learning. Artif Intell Rev. 2022;55:945-70.

Yu C, Liu J, Nemati S, Yin G. Reinforcement learning in healthcare: A survey. ACM Comput Surv. 2021;55(1):1-36.

Zhang B, Hong WC, Ouyang Z. Reinforcement learning in sustainable energy and electric systems: A survey. Annu Rev Control. 2022;49:145-61.