💬 **PERSPECTIVE**

🔓 OPEN ACCESS

Generative AI for computational chemistry: A roadmap to predicting emergent phenomena

Pratyush Tiwary[a,b,1], Lukas Herron[b,c,2], Richard John[d,2] 🆔, Suemin Lee[b,c,2], Disha Sanwal[a,2], and Ruiyu Wang[a,2] 🆔

The recent surge in generative AI has introduced exciting possibilities for computational chemistry. Generative AI methods have made significant progress in sampling molecular structures across chemical species, developing force fields, and speeding up simulations. This Perspective offers a structured overview, beginning with the fundamental theoretical concepts in both generative AI and computational chemistry. It then covers widely used generative AI methods, including autoencoders, generative adversarial networks, reinforcement learning, flow models, and language models, and highlights their selected applications in diverse areas including force field development, and protein/RNA structure prediction. A key focus is on the challenges these methods face before they become truly predictive, particularly in predicting emergent chemical phenomena. We believe that the ultimate goal of a simulation method or theory is to predict phenomena not seen before and that generative AI should be subject to these same standards before it is deemed useful for chemistry. We suggest that to overcome these challenges, future AI models need to integrate core chemical principles, especially from statistical mechanics.

generative AI | computational chemistry | molecular modeling

The last few years have seen a surge of excitement and explosion of generative AI methods across scientific fields, with computational chemistry being no exception. Pioneering efforts include sampling structures and thermal distributions of complex molecular systems, developing transferable force fields, and performing accelerated simulations (1–4). With numerous tools emerging, a clear Perspective is now essential to highlight progress and critically examine pitfalls. While the scope of generative AI's impact in chemistry is broad, this Perspective will focus exclusively on molecular simulation–driven computational chemistry. Molecular simulations offer an efficient platform for validating and iterating on new generative AI techniques, continually improving them using force field-based approximations of reality. Additionally, using generative AI on molecular simulation data can explore chemical and physical spaces that are otherwise difficult to access.

This Perspective is structured as follows. We begin with Section 1, summarizing the essential theoretical concepts and terminology of both generative AI and computational chemistry. Next, in Section 2, we provide an overview of widely used generative AI methods (Fig. 1), including autoencoders (AE) and their derivatives, generative adversarial networks (GANs), reinforcement learning (RL), flow-based methods, and recurrent neural networks (RNNs) and language models. The Perspective then highlights selected applications in Section 3 out of very many, focusing on ab initio quantum chemistry, coarse-grained force fields, protein structure prediction, and RNA structure prediction. Following this, Section 4 explores common themes and characteristics of generative AI tools that are particularly desirable for chemistry applications. Of particular note here is predicting emergent phenomena, which lies at the heart of chemistry and all science. Emergent phenomena occur when new properties arise even in systems with simple underlying interactions if they are large enough and/or studied long enough, as Phil Anderson discussed in his classic essay "More Is Different" (5). Current AI approaches struggle with capturing these emergent behaviors. Recent literature on some of the most powerful generative AI frameworks such as large language models (LLMs) (6) and diffusion models (7) has highlighted and quantified the limitations of current AI tools in capturing any emergent behavior, often showing that these tools primarily excel at impressive but essentially memorization and interpolation. We conclude with brief Section 5, providing an honest evaluation of the progress so far and challenges that need to be addressed before generative AI becomes a reliable member of the molecular simulator's toolbox for predicting emergent phenomena.

1. The Theoretical Minimum

In this section, we will summarize the theoretical concepts and terms key to this Perspective. We do this for computational chemistry and for generative AI. We recommend ref. 8 for a deeper understanding of computational chemistry

Author affiliations: [a]Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742; [b]University of Maryland Institute for Health Computing, Bethesda, MD 20852; [c]Biophysics Program and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742; and [d]Department of Physics and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742

[1]To whom correspondence may be addressed. Email: ptiwary@umd.edu.

[2]L.H., R.J., S.L., D.S., and R.W. contributed equally to this work.
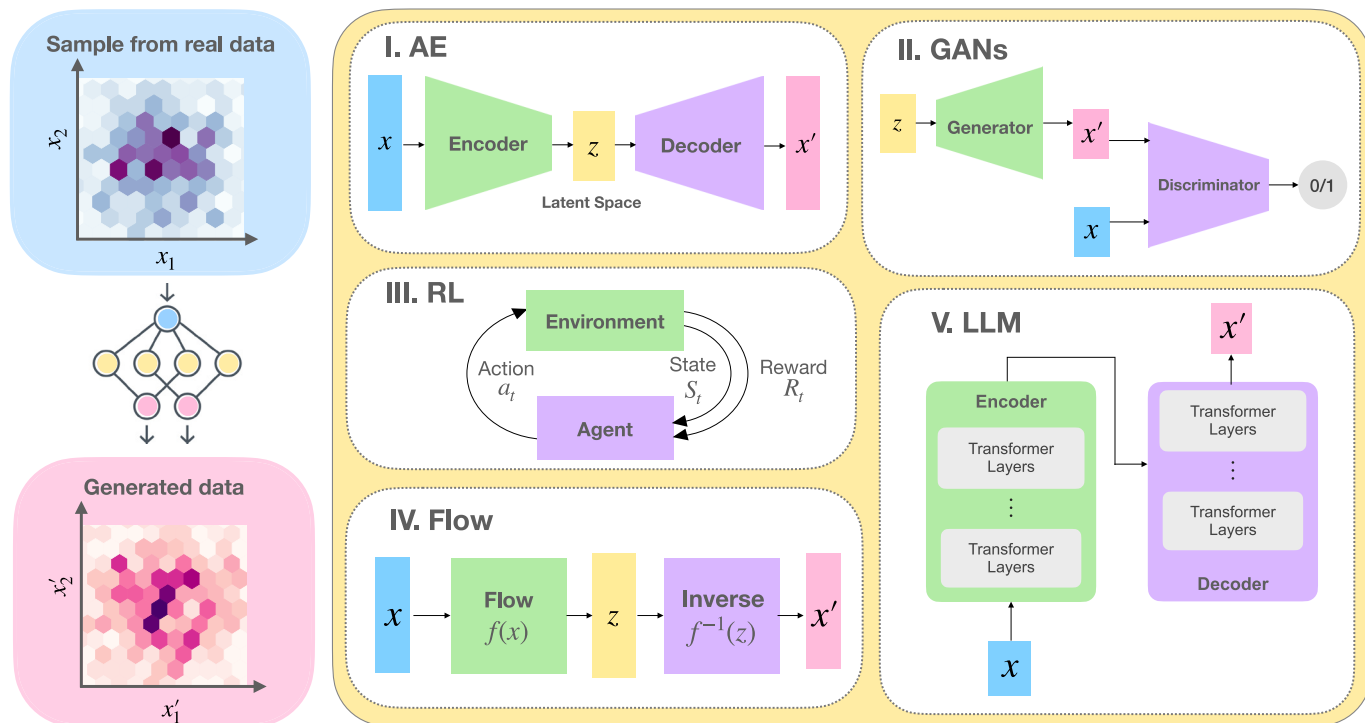
Fig. 1. Overall framework of generative AI and the methods discussed in this Perspective. The central task in generative AI is to generate new data that are similar to the training data or to model the underlying distribution of the data when the probability distribution is not explicitly available. This challenge holds particular importance in chemistry, where data can be structured (e.g., molecular graphs) or unstructured, time series, or static. The methods discussed in this Perspective include (*I*) AE: an architecture where the encoder compresses the input data x into a latent space z, and the decoder reconstructs the data from z to produce an output x' which should closely match the input. (*II*) GANs: a framework comprising a generator that produces synthetic data x' from latent variables z and a discriminator that distinguishes between real data x and generated data x'. The generator and discriminator are trained together in an adversarial process, with the generator improving its ability to create realistic data as the discriminator refines its ability to detect fakes. (*III*) RL: a learning paradigm where an agent, typically a decision-making entity, interacts with an environment over time t. The agent takes actions a_t based on the current state S_t of the environment, and in return, it receives rewards R_t. Through this process, the agent learns to maximize cumulative rewards by refining its strategy or policy over successive iterations. (*IV*) Flow models: these models learn to transform complex probability distributions of the data x into simpler, tractable prior distributions z using invertible functions $f(x)$. Given data from a complex true distribution, these models enable the mapping to a simpler latent space, from which new data can be generated by inverting the transformation. (*V*) LLMs: a typical LLM consists of an encoder and a decoder, both composed of multiple transformer layers. These layers use self-attention mechanisms to understand and focus on the most relevant parts of the input sequence, facilitating the generation of coherent and contextually appropriate outputs in a variety of natural language processing tasks.

concepts. For generative AI, there is no one book that can stay up-to-date with the dizzying pace of development, though ref. 9 covers several key underlying concepts.

1.1. Computational Chemistry.

1. *Potential energy surface (PES)*: The PES is a multidimensional surface representing the energy of a molecular system as a function of its atomic positions. The minima on the PES correspond to stable molecular structures, while the pathways connecting these minima represent possible reaction mechanisms ignoring entropic effects relevant at temperatures above 0 K.

2. *Force fields*: Force fields are mathematical models used to describe the PES. They consist of a set of parameters and equations that define the interactions between atoms, including bond stretching, angle bending, torsional angles, and nonbonded interactions (e.g., van der Waals forces, electrostatic interactions). The choice of force field significantly influences the accuracy of molecular simulations, as it dictates how well the model can replicate real physical behaviors. Popular force fields include AMBER, CHARMM, OPLS, and in recent years, also machine learning force fields (MLFFs) (10).

3. *Thermodynamic ensemble*: A thermodynamic ensemble is a statistical representation of a system in which all possible microstates are considered according to specific environmental constraints like temperature, pressure, and volume. The choice of ensemble is crucial for accurately modeling real-world conditions and predicting system behavior from molecular simulation or generative AI.

4. *Collective variables (CVs) and reaction coordinate*: CVs simplify molecular analysis by reducing dimensionality and capturing essential degrees of freedom. They are key in enhanced sampling techniques for exploring rare events. Choosing the right CVs, which approximate the reaction coordinate (RC), is vital for capturing system dynamics. The RC tracks a system's progress along a reaction pathway, often identifying the transition state. Often the committor is considered the ideal RC (11), as it quantifies the probability of a system evolving toward a specific product state, accounting for both energetic and entropic effects.

5. *Free energy surfaces (FES)*: FES extend the concept of PESs. They quantify the probability of observing the system as a function of one or more CVs, by integrating out all other degrees of freedom. In practice, FES are almost

invariably directly sampled at finite temperatures, rather than simply integrating out degrees of freedom from the PES at 0 K. Depending on how closely the chosen CVs approximate the true RC, the FES can unfortunately be mechanistically quite misleading it could be masking out true barriers.

6. *Molecular simulations*: Molecular dynamics (MD) and Monte Carlo (MC) are essential for simulating molecular systems. MD solves Newton's equations to model atomic trajectories, while MC uses random sampling to explore configurational space. Ab initio MD combines quantum mechanical calculations like density functional theory (DFT) with MD for greater accuracy but at higher computational cost.

1.2. Generative AI.

1. *Latent Variables*: Latent variables are hidden factors that capture underlying structures in data. In models like AE and pre-deep-learning methods such as principal component analyses, these variables represent the reduced-dimensional space that captures essential features of the data, facilitating generation and reconstruction processes. Generally latent variables are often entangled, meaning they mix multiple underlying factors. Disentangling them improves interpretability and control, enabling precise manipulation of control parameters.

2. *Prior*: The prior is a distribution over the latent variables that encodes initial beliefs about their values before observing data. The use of priors can simplify model fitting and help achieve more intuitive models.

3. *Loss Function*: The loss function quantifies the difference between generated samples and the true data distribution, with different metrics used depending on the method. RMS error measures prediction accuracy, cross-entropy evaluates how similar predicted and actual distributions are, Kullback–Leibler (KL) divergence assesses how much one distribution diverges from a reference, and Wasserstein distance, while more computationally expensive, provides stability and clearer interpretation by assessing the cost of transforming one distribution into another.

4. *Training, Testing, and Validation*: Training fits the model to data, validation tunes it on unseen data to avoid overfitting, and testing evaluates its generalization on a separate dataset. Ideally, the test data should be entirely unseen, but in practice, overlap often occurs, especially in chemistry. Thus, careful data curation is crucial for reliable generative AI in computational chemistry (12).

5. *Regularization and Mode Collapse*: Regularization methods, such as dropout, weight decay, and early stopping, are used to prevent overfitting by penalizing overly complex models and ensuring they generalize well to new data. Especially in generative models, care must be taken to avoid mode collapse, where the model generates limited, repetitive outputs, missing the diversity in the data.

6. *Embedding*: Embeddings are dense, low-dimensional representations that map discrete data, like words or items, into continuous vector spaces where similar items are closer together. While latent space emphasizes compressing data and extracting essential features, embedding space prioritizes capturing relationships and semantic meaning.

7. *Attention*: Attention mechanisms allow models to focus on different parts of the input data when generating outputs. This is critical in transformer models and helps improve performance in tasks like natural language processing by enabling the model to weigh different input components.

2. Generative AI Methods for Computational Chemistry

Here, we provide an overview of popular generative AI methods relevant to computational chemistry. We describe the central ideas and highlight what makes these methods appealing, their limitations, and new research directions toward improving them.

2.1. AE and Derived Methods. AE have become increasingly visible in the field of computational chemistry due to their powerful ability to learn and represent complex high-dimensional molecular data as points in a low-dimensional latent space. Generally speaking, an AE is a type of neural network that compresses input data into a lower-dimensional latent space and reconstructs it as accurately as possible. By sampling from this latent space, new high-dimensional molecular data can be generated, potentially discovering novel molecular configurations. Points close in the latent space correspond to similar molecules, allowing the AE to explore chemical diversity (by sampling near the edges of the latent space) and classify molecular similarity (on the basis of proximity within the latent space). This strategy has been applied to classify and explore chemical space, improving similarity searches and clustering of compounds from experiments or calculations like DFT. AEs are also used in RC discovery and enhanced MD, enabling visualization and exploration of complex, high-dimensional landscapes by reducing dimensionality while retaining key features (4).

AEs are popular in computational chemistry for their easy-to-visualize latent variables, but they are prone to misuse. A central issue is the temptation of assuming Euclidean geometry in the latent space, which can lead to uncontrolled mapping of distances between the latent and high-dimensional spaces. Additionally, latent variables may be correlated or lack physical relevance. Understanding these problems requires a deeper look into AE construction, where traditionally an encoder maps input data to the latent space, and a decoder maps it back, with the goal of minimizing reconstruction loss. With this generic recipe, one can construct different types of AE-inspired methods by varying the following:

1. *Prior for the latent variable*: AEs can minimize training loss significantly with expressive encoders and decoders, risking overfitting. To prevent this, a prior distribution is imposed on the latent variable, adding a regularization loss that keeps the latent variable close to the prior while maintaining low reconstruction loss. The choice of prior, such as a Gaussian distribution in variational AE (VAEs) or a mixture of Gaussians (13) is an active research area.

2. *Additional loss terms to enforce physics*: Physically meaningful latent representations can be obtained by using disentanglement-based loss terms, like in the β-VAE approach (14), or through dynamics-based priors (15). In chemistry, meaningful latents can also be obtained by adding loss terms that maximize specific physical attributes, as shown in ref. 16, where the two dimensions of the latent space correspond to entropic and enthalpic degrees of freedom.

3. *Generalizing output task*: A traditional autoencoder can be extended to predict other quantities, not just reconstruct inputs. For example, using the information bottleneck framework (17), it can predict which metastable state a molecular system will visit after a time delay. This approach (18, 19) closely mirrors desirable attributes of the committor function.

2.2. GANs. GANs (20) have gained particular attention for their ability to produce high-quality realistic images, audio, video, and chemical molecules. GANs possess two unique features: the discriminator and the generator (Fig. 1), which compete against each other in a zero-sum game. Here, the generator synthesizes new data while the discriminator tries to distinguish between the synthetic data and real data from a training set. Through continuous feedback, the generator progressively improves its ability to generate realistic data until the discriminator is fully deceived by the generator's newly synthesized data. This central idea underlying GANs has been improved through various variants. Conditional GANs (cGANs) (21), for instance, generate data conditioned on specific attributes, which is especially useful in chemistry for creating molecules or materials with desired properties. Wasserstein GANs (WGANs) (22) enhance training stability and mitigate issues like mode collapse by optimizing the Wasserstein distance between synthetic and real data.

GANs have successfully demonstrated great potential for chemistry also through integration with other neural networks. For molecular discovery, You et al. (23) have shown that combining graph neural networks with GANs, as in the Graph Convolutional Policy Network (GCPN), can effectively generate novel molecular structures by optimizing desired properties with RL (Section 2.3). In addition, the idpGAN (24) model utilizes GANs by incorporating them with transformer architectures (Section 2.5) to generate sequence-dependent protein conformational ensembles, which can capture protein dynamics and interactions. Ref. 25 provided an early and novel application of GANs to molecular coarse-grained simulations. Sidky et al. (26) proposed latent space simulators that integrate the VAMPnet model (27) with WGANs. This allows generating long synthetic MD trajectories that can accurately reproduce atomistic structures and kinetics observed in training trajectories, at a much lower computational cost.

Despite these forays in chemistry, GANs have several limitations that restrict their applicability. These limitations include training instability, mode collapse, and a heavy dependence on large datasets. Training instability arises due to the delicate balance between the generator and discriminator, which leads to nonconvergence and contributes to the issue of mode collapse. Reliance on training data makes it challenging to generate data that are out-of-distribution (OOD) (28, 29). As a result, GANs are gradually going out of fashion for chemical applications, and the field is shifting toward new state-of-the-art methods, such as diffusion models and RL-based approaches, which offer solutions to some of the inherent limitations of GANs (30).

2.3. RL. In RL a proverbial agent learns to make optimal decisions by interacting with an environment to maximize cumulative rewards. RL is usually modeled as a Markov decision process, comprising states, actions, an environment, and rewards; the agent receives rewards based on its actions and the current state of the environment, with the reward signal guiding subsequent actions and state transitions. The agent employs trial and error to explore various strategies which helps it improve its actions over time. One of the most widely used applications of RL in the pharmaceutical industry is in context of computationally driven chemistry through the method REINVENT (31) and many others that have followed (32), which use RL often combined with other DL approaches, to generate optimized molecules consistent with user-defined properties. In computational chemistry, it has been used for the learning of transition states (33), diffusive dynamics (34), and sampling protein conformational dynamics (35). In spite of its promise, RL for chemistry continues to be plagued with a few issues fundamental to molecular systems. These include:

1. *Curse of dimensionality*: Molecular systems exist in high-dimensional spaces, which RL algorithms can struggle to explore and learn efficiently. Often this is dealt with an adaptive use of RL where RL is trained on some existing dataset and the trained surrogate model is used to perform further exploration and data generation (35).
2. *Data scarcity problem*: Important molecular events, like chemical reactions or conformational changes, occur infrequently, and are hard to capture, resulting in incomplete datasets. Consequently, RL models trained on limited data may bias toward common states and miss rare but crucial phenomena.
3. *Novelty problem*: Molecular systems often have multimodal data corresponding to different metastable states. RL models often fail to generate a diverse set of outputs, leading to the mode collapse problem (Section 1.2). For instance, in ref. 32, the most potent design had a high structural similarity to an existing drug molecule, a problem somewhat common in the use of RL for discovery of chemical matter.

We conclude by highlighting how concepts from chemistry, particularly statistical mechanics, are enhancing RL. In Maximum Entropy RL (MaxEnt RL) (36), the goal is to maximize both expected reward and policy entropy, promoting stochasticity in actions. This approach improves RL algorithms' adaptability to real-world complexities. Recent work on Maximum Diffusion RL (37), based on the Principle of Maximum Caliber (38), has shown efficiency gains over traditional MaxEnt RL. The GFlowNet framework (39) treats state-action trajectories as network flows, ensuring robust sampling with detailed balance and importance sampling, outperforming Markov Chain MC methods in certain cases. The intersection of statistical mechanics and RL continues to be a promising area for developing innovative RL methods.

2.4. Flow Based Methods. Having discussed recent RL variants inspired by statistical physics in the last subsection, we now turn to models rooted in these principles from their inception (40). Flow models aim to sample from an inaccessible target probability distribution using limited available samples—an empirical dataset $\mathcal{D}$. Physics-inspired methods like simulated tempering (41) and annealed importance sampling (42) transform samples from the simple prior into samples from the target by constructing a bridge between the distributions. However, constructing the bridge relies on modifying the energy functions of the target and prior distributions. Generative flow models use neural networks to bridge distributions without requiring access to the energy function.

Flow-based models transform a simple prior distribution $q(\mathbf{z})$ into a more complex target distribution $p(\mathbf{x})$ through a series of learnable mappings (Fig. 1). Normalizing flows is a popular framework for sampling from $p(\mathbf{x})$. A normalizing flow model parameterizes an invertible transformation $\mathcal{M}$. The transformation acts as a change of variables that deforms the prior into the target distribution so that sampling from the prior and applying $\mathcal{M}$ produces a sample from the target distribution, i.e., $\mathbf{x} = \mathcal{M}(\mathbf{z})$. The change in probability associated with the change of coordinates is given by the identity

$$\log p(\mathbf{x}) = \log q(\mathbf{z}) - \log \left| \det J_{\mathcal{M}}(\mathbf{z}) \right|. \qquad [1]$$

The identity in Eq. **1** reweights samples between the prior and target distributions via the Jacobian determinant of $\mathcal{M}$. The optimal change of variables $\mathcal{M}$ maximizes the likelihood for samples in $\mathcal{D}$ and can be obtained by training a neural network $\mathcal{M}_\theta$.

However, computing the determinant is a (usually) prohibitively expensive operation that scales as $\mathcal{O}(d^3)$ in the general case, where d is the number of components of $\mathbf{x}$. Normalizing flows address this by employing architectures that result in structured Jacobians [for example, alternatingly upper- and lower-triangular (43)] for which the determinant computation is faster. However, the computationally tractable determinant results in reduced expressivity, and much recent effort has been devoted to addressing this tradeoff (44, 45).

Diffusion models (46) are generative algorithms that do not require access to the Jacobian determinant during training. Diffusion models learn to invert an Ornstein–Uhlenbeck (OU) diffusion process. The diffusion process generates a probability flow $p(\mathbf{x}, t)$ that transports the target distribution $p(\mathbf{x}, t = 0)$ to the prior $q(\mathbf{z}) \equiv p(\mathbf{x}, t = 1)$ (47). Diffusion models are made possible by a theorem stating that the drifts of the forward- and reverse-time diffusion processes differ only by the score $\nabla_{\mathbf{x}} \log p(\mathbf{x}, t)$ (48). Similar to normalizing flows, a neural network is trained to estimate the score from realizations of the forward process. Once parameterized, the score network can be used to simulate the time-reversed diffusion process, thereby generating samples from the target distribution.

The ideas behind diffusion models are deeply rooted in statistical physics. The OU diffusion process is naturally studied using techniques from nonequilibrium thermodynamics (40), and parameterizing the gradient of a potential (the score) rather than the potential itself is similar in spirit to force-matching in coarse-grained models (49). These early applications in coarse-graining of force-matching with a loss function were an early form of machine learning in statistical physics, but not utilizing neural nets to encode the coarse-grained interactions. Looking forward, tools from statistical physics will be central to understanding how diffusion models operate. For example, using techniques from spin-glass theory, it was found that the dynamics of the diffusion processes can be divided into distinct regimes distinguishing generalization and memorization of the training data (50). Recent exciting advancements such as flow-matching (51) and stochastic interpolants (52) further introduce ideas from optimal transport to model arbitrary diffusion processes.

2.5. RNNs and LLMs. RNNs, like long short-term memory (LSTM) networks, and transformer-based architectures (53, 54), have made significant strides in natural language processing, speech recognition, and computational chemistry, notably in protein structure prediction with AlphaFold2 (AF2). These models excel in handling sequential data, whether predicting the next value in a time series, generating sequences, or classifying entire sequences. LSTM networks use memory cells and gating mechanisms to maintain long-term dependencies, while transformers, such as those in LLMs, use self-attention mechanisms to capture complex relationships, processing sequences more efficiently and effectively. A typical LLM can contain an encoder and a decoder (Fig. 1) both composed of multiple transformer layers, each leveraging self-attention to understand the input context. The attention mechanism is key in aligning and focusing on the most relevant parts of the sequence during both encoding and decoding, enabling the model to generate coherent and contextually appropriate outputs.

These models are particularly valuable for chemistry, where many processes are non-Markovian and require long-term dependencies for accurate predictions, such as in reaction prediction and MD simulations. However, the success of LLMs in other domains does not easily transfer to chemistry due to their limitations in extrapolating beyond the training data and not accounting for hidden biases (55). This is critical in a field where novel molecules and reactions often lie outside previously explored chemical spaces. To address this, specialized approaches are needed, such as the one proposed in ref. 56, which integrates statistical physics into LSTM training using a path sampling approach based on the Principle of Maximum Caliber (38).

3. Selected Applications

3.1. Ab Initio Quantum Chemistry and Coarse-Grained Force Fields. In quantum chemistry, deep neural networks are used to achieve high quantum-level accuracy while reducing computational costs. For instance, AI algorithms have been developed to solve electronic Schrödinger equations for ground and excited states, reducing the computational complexity from $O(N^7)$ to $O(N^4)$ (57–59). As for MD simulations, MLFFs allow ab initio quantum-quality calculations to approach the speed of classical MD simulations (60). Another approach directly using generative AI is applying coarse-grained models for macromolecules, which effectively reduce the number of atoms in MD simulations (25, 61).

Ab initio MD is computationally expensive, but MLFFs can estimate energies and forces from atomic configurations without electron calculations (62), enabling quantum-quality simulations of hundreds of atoms over nanoseconds. MLFFs have been used to study liquid dielectric constants (63), phase behaviors of water (64), proton transfer for energy materials (65), and prebiotic chemical reactions (66). Beyond MLFFs, diffusion models can generate molecular structures by sampling the Boltzmann distribution, optimizing geometry without force and energy calculations (67), speeding up ground-state searches. Additionally, ML can model charge density from structures (68), which, while not always needed for MD, provides the missing charge density in MLFFs, expanding AI applications to complex quantum chemistry problems like predicting vibrational spectra (69). Developing MLFFs that can generalize beyond training data remains an area of concern and active research interest (66, 70). A related open challenge in MLFFs involves accounting for effects of long range forces such as electrostatics and treating very inhomogeneous systems (71, 72).

While MLFFs maintain quantum-level accuracy in systems typically studied with classical MD, CG models simplify simulations by representing larger systems with reduced atomic detail. An open challenge in CG models is backmapping to all-atom configurations. One approach uses an autoencoder architecture, where the encoder learns the CG model, and the decoder backmaps it to an ensemble's average structure (73). Denoising Diffusion Probabilistic Models can also generate atom-resolution structures (74). Additionally, learning the score term of a denoising diffusion process can approximate CG force fields and generate equilibrium Boltzmann distributions around local energy minima (75).

3.2. Protein Structure and Conformation Prediction.
Understanding a protein's structure, both in native and nonnative states, is crucial for determining its function, stability, and interactions. Traditionally, structure prediction relied on experimental techniques like X-ray diffraction (XRD), NMR, and cryoelectron microscopy (Cryo-EM) (76). The success of AI-driven approaches like AF2 and RoseTTAFold, which can now predict crystal-like protein structures, is largely due to the availability of such high-quality experimental data deposited in the Protein Data Bank (PDB) (77, 78). AF2 uses coevolutionary information from multiple sequence alignments (MSAs) to predict structures from amino acid sequences, while RoseTTAFold enhances accuracy by integrating evolutionary information with 3D coordinate refinement (78). Recently, AlphaFold3 (AF3) and RFDiffusion have further advanced predictions by incorporating diffusion-based models capable of handling complex structures (79, 80).

While AI methods have transformed biomolecular structure prediction, they remain limited by the quality of their training datasets and struggle to predict metastable nonnative structures or the effects of point mutations (81). To explore hidden conformations, approaches like MSA perturbation have been developed, including reduced-MSA (82), AF2Cluster (83), and SPEECH-AF (84), all providing structural diversity. However, proteins are dynamic systems with fluctuations that depend precisely on thermodynamic parameters such as temperature, pressure, and chemical potential (85, 86). To capture this conformational diversity, methods like AF2RAVE (87) and AF2 integrated with MSM (88) generate hypothetical structures through MSA perturbation and rank them according to their Boltzmann weight via MD simulations. AF2RAVE, for example, uses reduced-MSA for initial predictions, clusters them with the State Predictive Information Bottleneck approach (17), and ranks them using enhanced sampling MD simulations. This method has been effective in capturing conformational changes, such as DFG-in and -out conformations of tyrosine-protein kinases, improving accuracy in downstream docking (89, 90).

In some cases, MD simulations are used as training data for AI-driven methods. The Boltzmann generator (91) combines MD simulations with normalizing flows to generate Boltzmann-weighted molecular conformations, potentially applicable to proteins. AlphaFlow (92) integrates AlphaFold with flow matching (93) to produce diverse protein conformations, training on MD simulations from the ATLAS dataset (94) to enhance diversity and capture fast mode dynamics. However, it struggles with slow, emergent properties, limiting its ability to model long-term protein behavior. Extending training with longer MD simulations could address this but raises questions about the added value of generative AI. Two recent approaches, distributional graphformer (DiG) (95) and NeuralPLexer (96), combine generative AI with MD to predict conformational distributions. DiG uses a diffusion algorithm within a graphformer architecture to generate protein conformations from a primary sequence, with MD simulations as training data, while NeuralPLexer employs a diffusion-based model to predict state-specific protein–ligand complex structures, capturing dynamic conformational changes by integrating biophysical constraints.

As demonstrated in these examples, generative AI and its further incorporation with MD simulations have proven effective in accurately predicting protein structures and exploring protein conformations, bringing out the best of both AI and MD. We expect this will remain an area of active research with the continued development of more robust and accurate methods. That said, we urge caution in training generative AI on synthetic MD data. It is essential that these models respect physical laws, as deviations could lead to nonphysical predictions. Like with DALL-E and ChatGPT, deepfakes could proliferate, leading to unreliable outcomes and unreal chemistry.

3.3. RNA Structure Prediction.
RNAs are an emerging frontier in medicinal chemistry (97), yet despite the growing interest in them, experimentally solved RNA structures remain scarce (98). As a result, computational methods have become indispensable for modeling RNA tertiary structures. Among all methods, physics-based computational methods—like minimum free energy secondary structure prediction algorithms (99–102) and energy-based simulations (103–105)—are especially prevalent. These methods model energetic and entropic contributions to RNA structure formation and are a starting point for successful generative approaches. An intuitive approach to integrating

AI with physics-based methods is using a neural network to predict geometries that are later refined with an energy function (106–108). While these methods have the potential to generalize well across diverse RNA sequences, their accuracy currently lags behind methods that make use of large databases (109, 110) of RNA sequences. Future advancements in this space might be more improved energy (111, 112) or scoring functions (113, 114) and conformational sampling algorithms (115, 116).

Language models for RNA were developed following their widespread success in machine translation and text generation (117–122). These models parameterize embeddings that can serve as inputs for other networks fine-tuned for specific tasks like structure or function prediction. AF3's MSA module and Pairformer (79) learn embeddings from MSAs, while foundation models like RNA-FM (118) or ATOM-1 (121) learn from single sequences or chemical mapping data. The premier application of RNA language models to structure prediction is as a module within a larger multicomponent AI system.

Recent years have seen the emergence of large-scale pretrained generative AI models for RNA structure prediction (79, 123–126). These are neural networks composed of different "blocks" or "modules" that learn specific aspects of biomolecular structure. Together, the modules predict tertiary structures from sequence and template information. AF3 (79) and RoseTTAFold-2NA (124) (RF2NA) are state-of-the-art in RNA tertiary structure prediction, though the accuracy of predicted RNA structures does not yet rival that of proteins structures (127). Still, both methods aim to predict one structure per sequence, operating within a modeling paradigm that is increasingly challenged by advances in structural biology.

The field of structural biology is undergoing a paradigm shift, moving away from the historically prevalent native-centric view of biomolecular structure (128–131). A single structure may be insufficient to account for biomolecular function; the appropriate description is instead an ensemble—a set of representative Boltzmann-weighted conformers. Currently, the goal of most generative AI methods for RNA structure prediction is to predict one native structure per sequence (86). To fully embrace this paradigm shift, future generative methods should aim to predict Boltzmann-weighted structural ensembles.

To complicate matters further, the ensemble is not static: it responds to environmental factors like temperature (132), ions (133), or small molecules (134). While AF3 can model the environment explicitly to some extent, it is perhaps more feasible to model the environment implicitly. Thermodynamic maps (135) (TMs) follow a generative framework that infers the structural ensemble's dependence on environmental conditions. TMs are diffusion models where thermodynamic parameters (e.g., temperature) are implicitly represented within the Langevin dynamics of the diffusion process. The central idea is to map molecular conformations onto those of a simple, idealized system where the effect of the environment is straightforward to account for. Though still in its early stages, TMs have demonstrated the ability to predict the temperature dependence of RNA structural ensembles.

4. Desirables from Generative AI for Chemistry

We believe the ultimate predictive power of any tool—whether theory, molecular simulations, or generative AI—lies in starting from chemical identity and accurately predicting function while rigorously accounting for environmental conditions. However, function is not an inherent property of a given sequence or a chemical formula; rather, it is an emergent property that arises from the dynamic interactions and feedback loops across multiple scales. Achieving this requires navigating through increasing complexity—structure, thermodynamic ensemble, and environment (Fig. 2)—where modeling challenges intensify due to intricate fluctuations governed by equilibrium and nonequilibrium statistical mechanics. While current generative AI methods like AlphaFold2 predict the most stable structure given chemical identity, much more remains to be desired. This section outlines key attributes for advancing generative AI in this direction.

1. *Chemistry and AI, Not AI for Chemistry*: It is indubitable that AI's ability to handle large, complex datasets, and uncover hidden structures makes it especially valuable in computational chemistry. At the same time, integrating chemistry, particularly statistical mechanics, quantum mechanics, and thermodynamics, with AI creates a powerful synergy. These fields provide essential priors,
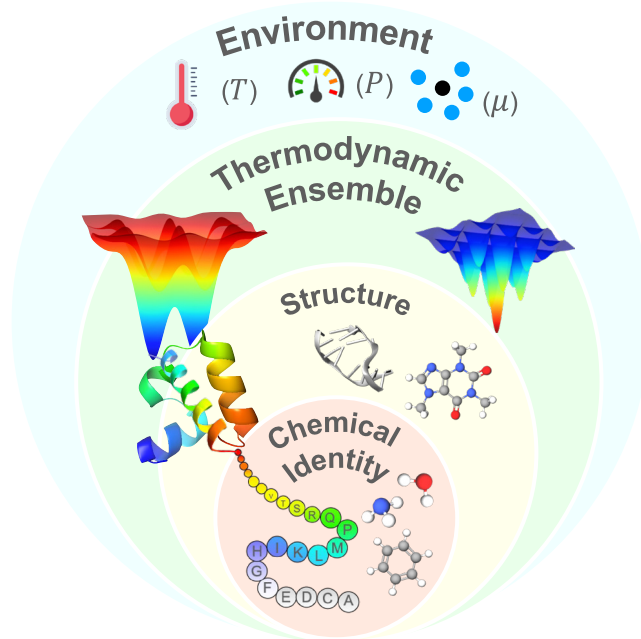
Fig. 2. Desirables for predicting emergent phenomena in chemistry with generative AI. The goal for generative AI, molecular simulations, and computational chemistry is to start from the chemical identity—whether sequence or composition—and accurately predict function while considering the relevant environmental conditions. To achieve this, the intermediate rungs of structure, thermodynamic ensemble, and environment must be accounted for, where the environment can be quantified through parameters such as the temperature T, pressure P, and chemical potential μ. As we move up each level, modeling becomes more challenging due to the nuances of fluctuations governed by laws of equilibrium and nonequilibrium statistical mechanics. This will require generative AI models deeply grounded in statistical mechanics, with precise priors to account for complex interactions, dynamics, and environmental effects.

frameworks for hypothesis testing, and tools for extrapolation that enhance generative AI's effectiveness in chemistry.

2. *Interpretability and Reliability Testing*: The interpretability and reliability of AI models are paramount in chemistry. Current internal confidence measures, such as AlphaFold's pLDDT scores, have shown limitations in providing reliable assessments of model predictions (81). A focus on developing more robust interpretability frameworks and reliability testing is necessary to ensure that AI models not only generate accurate predictions but also provide meaningful insights into their confidence levels and potential errors.

3. *Out-of-Distribution Generalization and Efficacy in Data-Sparse Regimes*: AI excels at smart interpolation within well-covered data regions, but the real challenge is generalizing to out-of-distribution data to reduce hallucinations and spurious predictions. In chemistry, where data are often sparse, generative AI must be effective with limited data, transferring learned knowledge across diverse chemical types to ensure models remain robust, versatile, and reliable in scientific discovery.

4. *Rethinking Data - More data are not always better data*: In chemistry, the typical scaling laws seen in LLMs—where more data improve test loss—may not hold true. More data do not always enhance performance and can sometimes obscure meaningful insights. For instance, in an MD trajectory trapped in metastable states, adding more data might amplify noise rather than provide useful information about rare transitions. Capturing such rare events in the training dataset remains an open problem, thereby emphasizing the continued importance of enhanced sampling methods (4). These concerns underscore the need to rethink data handling in AI for chemistry, ensuring that additional data are genuinely useful and do not obscure key physical insights. Moreover, traditional AI methods of splitting data into training, testing, and validation sets can be problematic due to data leakage and challenges in quantifying overlap.

5. *Emergent Phenomena and correct coupling to environmental variables*: The power and thrill of MD and computational chemistry lie in predicting new chemistry and physics that emerge naturally when simulating a large system for a long time-phenomena that could not be guessed from a simple force field or Hamiltonian. Emergent behavior, sensitive to control parameters and environmental variables, often arises in such simulations, closely aligning with theoretical predictions. However, generative AI often struggles to produce novel physics or chemistry beyond its training data, with emergent phenomena sometimes being artifacts. Most AI models also fail to account for environmental conditions, limiting their predictive power in new scenarios. Hybrid approaches like AF2RAVE (87) and TMs (135) show promise in this context by integrating AI with physical principles.

5. Critical Assessment and Outlook

Generative AI has made impressive strides in computational chemistry, particularly in force field development, structure prediction, and accelerated molecular simulations, showing its potential to tackle complex chemical challenges. However, significant obstacles remain before AI can fully integrate into the molecular simulation toolbox. The ultimate goal of any simulation method or theory is to reliably predict chemical function directly from chemical identity—a dream yet unrealized. We argue that the same aspirations should be applied to generative AI for physical sciences. By integrating chemistry, particularly statistical mechanics, into AI models, considering the roles of structural and dynamical ensembles with precise fluctuations, and accounting for environmental influences, this goal may be achievable. One central issue is that AI models are "data hungry," requiring large datasets to make accurate predictions. In contrast, a core goal of theoretical chemistry has been to derive emergent phenomena from general underlying theories (e.g., the Schrödinger equation, Boltzmann statistical mechanics), which strive for predictive power even with minimal data. Theoretical frameworks, such as Marcus theory for electron transfer, demonstrate how such predictions can stem from simple principles. While chemistry has much to gain from generative AI, it also has much to teach it. Grounding AI in chemistry's principles can create more accurate, adaptable, and interpretable models. This integration could transform AI into a powerful tool for predicting novel emergent phenomena, driving discoveries, and deepening our understanding of chemical processes.

1.   D. M. Anstine, O. Isayev, Generative models as an emerging paradigm in the chemical sciences. *J. Am. Chem. Soc.* **145**, 8736–8750 (2023).
2.   Y. Du *et al.*, Machine learning-aided generative molecular design. *Nat. Mach. Intell.* **6**, 589–604 (2024).
3.   G. M. Rotskoff, Sampling thermodynamic ensembles of molecular systems with generative neural networks: Will integrating physics-based models close the generalization gap? *Curr. Opin. Solid State Mater. Sci.* **30**, 101158 (2024).
4.   S. Mehdi, Z. Smith, L. Herron, P. Zou, P. Tiwary, Enhanced sampling with machine learning. *Annu. Rev. Phys. Chem.* **75**, 347–370 (2024).
5.   P. W. Anderson, More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* **177**, 393–396 (1972).
6.   R. Schaeffer, B. Miranda, S. Koyejo, Are emergent abilities of large language models a mirage? *Adv. Neural Inf. Process. Syst.* **36**, 1–17 (2024).
7.   G. Biroli, M. Mézard, Generative diffusion in very large dimensions. *J. Stat. Mech: Theory Exp.* **2023**, 093402 (2023).
8.   D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Elsevier, 2023).
9.   A. D. White, Deep learning for molecules and materials. *Living J. Comput. Mol. Sci.* **3**, 1499 (2021).
10. P. S. Nerenberg, T. Head-Gordon, New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **49**, 129–138 (2018).
11. R. B. Best, G. Hummer, Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6732–6737 (2005).
12. E. Heid, C. J. McGill, F. H. Vermeire, W. H. Green, Characterizing uncertainty in machine learning for chemistry. *J. Chem. Inf. Model.* **63**, 4012–4029 (2023).
13. J. Tomczak, M. Welling, "VAE with a VampPrior" in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2018), pp. 1214–1223.

14. I. Higgins *et al.*, beta-VAE: Learning basic visual concepts with a constrained variational framework. *ICLR (Poster)* **3** (2017).
15. D. Wang, Y. Wang, L. Evans, P. Tiwary, From latent dynamics to meaningful representations. *J. Chem. Theory Comput.* **20**, 3503–3513 (2024).
16. E. R. Beyerle, P. Tiwary, Thermodynamically optimized machine-learned reaction coordinates for hydrophobic ligand dissociation. *J. Phys. Chem. B* **128**, 755–767 (2024).
17. D. Wang, P. Tiwary, State predictive information bottleneck. *J. Chem. Phys.* **154**, 134111 (2021).
18. E. R. Beyerle, S. Mehdi, P. Tiwary, Quantifying energetic and entropic pathways in molecular systems. *J. Phys. Chem. B* **126**, 3950–3960 (2022).
19. R. Zhao, Z. Zou, J. D. Weeks, P. Tiwary, Quantifying the relevance of long-range forces for crystal nucleation in water. *J. Chem. Theory Comput.* **19**, 9093–9101 (2023).
20. I. Goodfellow *et al.*, "Generative adversarial nets" in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Weinberger, Eds. (Curran Associates, Inc., 2014), vol. 27.
21. M. Mirza, S. Osindero, Conditional generative adversarial nets. arXiv [Preprint] (2014). https://doi.org/10.48550/arXiv.1411.1784 (Accessed 6 November 2024).
22. M. Arjovsky, S. Chintala, L. Bottou, "Wasserstein generative adversarial networks" in *Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research*, D. Precup, Y. W. Teh, Eds. (PMLR, 2017), vol. 70, pp. 214–223.
23. J. You, B. Liu, Z. Ying, V. Pande, J. Leskovec, "Graph convolutional policy network for goal-directed molecular graph generation" in *Advances in Neural Information Processing Systems*, S. Bengio, *et al.*, Eds. (Curran Associates, Inc., 2018), vol. 31.
24. G. Janson, G. Valdes-Garcia, L. Heo, M. Feig, Direct generation of protein conformational ensembles via machine learning. *Nat. Commun.* **14**, 774 (2023).
25. A. E. Durumeric, G. A. Voth, Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *J. Chem. Phys.* **151**, 124110 (2019).
26. H. Sidky, W. Chen, A. L. Ferguson, Molecular latent space simulators. *Chem. Sci.* **11**, 9459–9467 (2020).
27. A. Mardt, L. Pasquali, H. Wu, F. Noé, VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **9**, 5 (2018).
28. J. Gui, Z. Sun, Y. Wen, D. Tao, J. Ye, A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* **35**, 3313–3332 (2023).
29. H. Thanh-Tung, T. Tran, *Catastrophic forgetting and mode collapse in GANs* (IEEE, 2020), pp. 1–10.
30. P. Dhariwal, A. Q. Nichol, "Diffusion models beat GANs on image synthesis" in *Advances in Neural Information Processing Systems*, A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan, Eds. (Curran Associates, Inc., 2021).
31. T. Blaschke *et al.*, Reinvent 2.0: An AI tool for de novo drug design. *J. Chem. Inf. Model.* **60**, 5918–5922 (2020).
32. A. Zhavoronkov *et al.*, Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).
33. J. Zhang *et al.*, Deep reinforcement learning of transition states. *Phys. Chem. Chem. Phys.* **23**, 6888–6895 (2021).
34. A. Das, D. C. Rose, J. P. Garrahan, D. T. Limmer, Reinforcement learning of rare diffusive dynamics. *J. Chem. Phys.* **155**, 134105 (2021).
35. D. E. Kleiman, D. Shukla, Multiagent reinforcement learning-based adaptive sampling for conformational dynamics of proteins. *J. Chem. Theory Comput.* **18**, 5422–5434 (2022).
36. B. D. Ziebart *et al.*, *Maximum Entropy Inverse Reinforcement Learning* (AAAI, Chicago, IL, 2008), vol. 8, pp. 1433–1438.
37. T. A. Berrueta, A. Pinosky, T. D. Murphey, Maximum diffusion reinforcement learning. *Nat. Mach. Intell.* **6**, 504–514 (2024).
38. K. Ghosh, P. D. Dixit, L. Agozzino, K. A. Dill, The maximum caliber variational principle for nonequilibria. *Annu. Rev. Phys. Chem.* **71**, 213–238 (2020).
39. Y. Bengio *et al.*, Gflownet foundations. *J. Mach. Learn. Res.* **24**, 10006–10060 (2023).
40. J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics" in *International Conference on Machine Learning* (PMLR, 2015), pp. 2256–2265.
41. E. Marinari, G. Parisi, Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett.* **19**, 451–458 (1992).
42. R. M. Neal, Annealed importance sampling. *Stat. Comput.* **11**, 125–139 (2001).
43. L. Dinh, J. Sohl-Dickstein, S. Bengio, Density estimation using real NVP. arXiv [Preprint] (2016). https://doi.org/10.48550/arXiv.1605.08803 (Accessed 2 September 2024).
44. C. Durkan, A. Bekasov, I. Murray, G. Papamakarios, Neural spline flows. *Adv. Neural. Inf. Process. Syst.* **32**, 1–12 (2019).
45. H. Wu, J. Köhler, F. Noé, Stochastic normalizing flows. *Adv. Neural. Inf. Process. Syst.* **33**, 5933–5944 (2020).
46. J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models. *Adv. Neural. Inf. Process. Syst.* **33**, 6840–6851 (2020).
47. Y. Song *et al.*, Score-based generative modeling through stochastic differential equations. arXiv [Preprint] (2020). https://arxiv.org/abs/2011.13456.
48. B. D. Anderson, Reverse-time diffusion equation models. *Stochastic Process. Appl.* **12**, 313–326 (1982).
49. J. Jin, A. J. Pak, A. E. P. Durumeric, T. D. Loose, G. A. Voth, Bottom-up coarse-graining: Principles and perspectives. *J. Chem. Theory Comput.* **18**, 5759–5791 (2022).
50. G. Biroli, T. Bonnaire, V. de Bortoli, M. Mezard, Dynamical regimes of diffusion models. *Nat. Commun.* **15**, 9957 (2024).
51. Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, M. Le, Flow matching for generative modeling. arXiv [Preprint] (2022). https://arxiv.org/abs/2210.02747 (Accessed 2 September 2024).
52. M. S. Albergo, N. M. Boffi, E. Vanden-Eijnden, Stochastic interpolants: A unifying framework for flows and diffusions. arXiv [Preprint] (2023). https://arxiv.org/abs/2303.08797.
53. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
54. A. Vaswani, Attention is all you need. arXiv [Preprint] (2017). https://arxiv.org/abs/1706.03762.
55. E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big? (2021), pp. 610–623.
56. S. T. Tsai, E. Fields, Y. Xu, E. J. Kuo, P. Tiwary, Path sampling of recurrent neural networks by incorporating known physics. *Nat. Commun.* **13**, 7231 (2022).
57. J. Hermann, Z. Schätzle, F. Noé, Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **12**, 891–897 (2020).
58. D. Pfau, S. Axelrod, H. Sutterud, I. von Glehn, J. S. Spencer, Accurate computation of quantum excited states with neural networks. *Science* **385**, eadn0137 (2024).
59. L. E. Ratcliff *et al.*, Challenges in large scale quantum mechanical calculations. *WIREs Comput. Mol. Sci.* **7**, e1290 (2017).
60. P. Tiwary, Modeling prebiotic chemistries with quantum accuracy at classical costs. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2408742121 (2024).
61. M. Majewski *et al.*, Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **14**, 5739 (2023).
62. F. Noé, A. Tkatchenko, K. R. Müller, C. Clementi, Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71**, 361–390 (2020).
63. C. Zhang, S. Yue, A. Z. Panagiotopoulos, M. L. Klein, X. Wu, Why dissolving salt in water decreases its dielectric permittivity. *Phys. Rev. Lett.* **131**, 076801 (2023).
64. T. E. Gartner, P. M. Piaggi, R. Car, A. Z. Panagiotopoulos, P. G. Debenedetti, Liquid-liquid transition in water from first principles. *Phys. Rev. Lett.* **129**, 255702 (2022).
65. V. Quaranta, M. Hellström, J. Behler, Proton-transfer mechanisms at the water–ZnO interface: The role of presolvation. *J. Phys. Chem. Lett.* **8**, 1476–1483 (2017).
66. Z. Benayad, R. David, G. Stirnemann, Prebiotic chemical reactivity in solution with quantum accuracy and microsecond sampling using neural network potentials. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2322040121 (2024).
67. D. Rothchild *et al.*, Investigating the behavior of diffusion models for accelerating electronic structure calculations. *Chem. Sci.* **15**, 13506–13522 (2024).
68. P. Pope, D. Jacobs, "Towards combinatorial generalization for catalysts: A Kohn–Sham charge-density approach" in *Advances in Neural Information Processing Systems*, A. Oh, *et al.*, Eds. (Curran Associates, Inc., 2023), vol. 36, pp. 60585–60598.
69. L. Zhang *et al.*, Deep neural network for the dielectric response of insulators. *Phys. Rev. B* **102**, 041121 (2020).
70. Y. Zhai, A. Caruso, S. L. Bore, Z. Luo, F. Paesani, A "short blanket" dilemma for a state-of-the-art neural network potential for water: Reproducing experimental properties or the physics of the underlying many-body interactions? *J. Chem. Phys.* **158**, 084111 (2023).
71. A. Gao, R. C. Remsing, Self-consistent determination of long-range electrostatics in neural network potentials. *Nat. Commun.* **13**, 1572 (2022).
72. D. M. Anstine, O. Isayev, Machine learning interatomic potentials and long-range physics. *J. Phys. Chem. A* **127**, 2417–2431 (2023).
73. W. Wang, R. Gómez-Bombarelli, Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **5**, 125 (2019).
74. M. S. Jones, K. Shmilovich, A. L. Ferguson, Diamondback: Diffusion-denoising autoregressive model for non-deterministic backmapping of cα protein traces. *J. Chem. Theory Comput.* **19**, 7908–7923 (2023).
75. M. Arts *et al.*, Two for one: Diffusion models and force fields for coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **19**, 6151–6159 (2023).
76. B. C. Goh *et al.*, Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annu. Rev. Biophys.* **45**, 253–278 (2016).
77. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
78. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
79. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature* **630**, 493–500 (2024).
80. J. L. Watson *et al.*, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
81. G. R. Buel, K. J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* **29**, 1–2 (2022).
82. D. del Alamo, D. Sala, H. S. Mchaourab, J. Meiler, Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **11**, e75751 (2022).
83. H. K. Wayment-Steele *et al.*, Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
84. R. A. Stein, H. S. Mchaourab, Speach AF: Sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLoS Comput. Biol.* **18**, 1–16 (2022).
85. K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
86. G. R. Bowman, AlphaFold and protein folding: Not dead yet! The frontier is conformational ensembles *Annu. Rev. Biomed. Data Sci.* **7**, 51–57 (2024).
87. B. P. Vani, A. Aranganathan, D. Wang, P. Tiwary, AlphaFold2-rave: From sequence to Boltzmann ranking. *J. Chem. Theory Comput.* **19**, 4351–4354 (2023).
88. A. Meller, S. Bhakat, S. Solieva, G. R. Bowman, Accelerating cryptic pocket discovery using AlphaFold. *J. Chem. Theory Comput.* **19**, 4355–4363 (2023).
89. B. P. Vani, A. Aranganathan, P. Tiwary, Exploring kinase Asp-Phe-Gly (DFG) loop conformational stability with AlphaFold2-rave. *J. Chem. Inf. Model.* **64**, 2789–2797 (2024).
90. X. Gu, A. Aranganathan, P. Tiwary, Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-rave. *eLife* **13**, RP99702 (2024).
91. F. Noé, S. Olsson, J. Köhler, H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019).

92. B. Jing, B. Berger, T. Jaakkola, Alphafold meets flow matching for generating protein ensembles. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2402.04845. Accessed 2 September 2024.
93. Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, M. Le, "Flow matching for generative modeling" in *The Eleventh International Conference on Learning Representations* (2023).
94. Y. Vander Meersche, G. Cretin, A. Gheeraert, J. C. Gelly, T. Galochkina, ATLAS: Protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Res.* **52**, D384–D392 (2023).
95. S. Zheng *et al.*, Predicting equilibrium distributions for molecular systems with deep learning. *Nat. Mach. Intell.* **6**, 558–567 (2024).
96. Z. Qiao, W. Nie, A. Vahdat, T. F. Miller, A. Anandkumar, State-specific protein–ligand complex structure prediction with a multiscale deep generative model. *Nat. Mach. Intell.* **6**, 195–208 (2024).
97. J. P. Falese, A. Donlic, A. E. Hargrove, Targeting RNA with small molecules: From fundamental principles towards the clinic. *Chem. Soc. Rev.* **50**, 2224–2243 (2021).
98. M. Szikszai *et al.*, RNA3DB: A structurally-dissimilar dataset split for training and benchmarking deep learning models for RNA structure prediction. *J. Mol. Biol.* **436**, 168552 (2024).
99. R. Nussinov, A. B. Jacobson, Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6309–6313 (1980).
100. M. Zuker, On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48–52 (1989).
101. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
102. X. Xu, P. Zhao, S. J. Chen, Vfold: A web server for RNA structure and folding thermodynamics prediction. *PLoS ONE* **9**, e107504 (2014).
103. M. Parisien, F. Major, The MC-fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55 (2008).
104. P. Sripakdeevong, W. Kladwang, R. Das, An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20573–20578 (2011).
105. A. M. Watkins, R. Rangan, R. Das, FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. *Structure* **28**, 963–976 (2020).
106. R. Pearce, G. S. Omenn, Y. Zhang, De novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.05.15.491755 (Accessed 27 November 2024).
107. W. Wang *et al.*, trRosettaRNA: Automated prediction of RNA 3D structure with transformer network. *Nat. Commun.* **14**, 7266 (2023).
108. Y. Li *et al.*, Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nat. Commun.* **14**, 5745 (2023).
109. S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, S. R. Eddy, Rfam: An RNA family database. *Nucleic Acids Res.* **31**, 439–441 (2003).
110. The RNAcentral Consortium, RNAcentral: A hub of information for non-coding RNA sequences. *Nucleic Acids Res.* **47**, D221–D229 (2019).
111. J. Stasiewicz, S. Mukherjee, C. Nithin, J. M. Bujnicki, QRNAS: Software tool for refinement of nucleic acid structures. *BMC Struct. Biol.* **19**, 1–11 (2019).
112. P. Xiong, R. Wu, J. Zhan, Y. Zhou, Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement. *Nat. Commun.* **12**, 2777 (2021).
113. J. Wang, Y. Zhao, C. Zhu, Y. Xiao, 3dRNAscore: A distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res.* **43**, e63 (2015).
114. R. J. Townshend *et al.*, Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
115. J. Li, S. J. Chen, RNAJP: Enhanced RNA 3D structure predictions with non-canonical interactions and global topology sampling. *Nucleic Acids Res.* **51**, 3341–3356 (2023).
116. K. Chen, Y. Zhou, S. Wang, P. Xiong, RNA tertiary structure modeling with BRiQ potential in CASP15. *Proteins: Struct., Funct., Bioinf.* **91**, 1771–1778 (2023).
117. M. Akiyama, Y. Sakakibara, Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning. *NAR Genomics Bioinf.* **4**, lqac012 (2022).
118. J. Chen *et al.*, Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. arXiv [Preprint] (2022). https://arxiv.org/abs/2204.00300 (Accessed 27 November 2024).
119. X. Wang *et al.*, Uni-RNA: Universal pre-trained models revolutionize RNA research. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.07.11.548588 (Accessed 27 November 2024).
120. K. Chen *et al.*, Self-supervised learning on millions of pre-mRNA sequences improves sequence-based RNA splicing prediction. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.01.31.526427 (Accessed 27 November 2024).
121. N. Boyd *et al.*, ATOM-1: A foundation model for RNA structure and function built on chemical mapping data. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.12.13.571579 (Accessed 27 November 2024).
122. Y. Chu *et al.*, A 5'UTR language model for decoding untranslated regions of mRNA and function predictions. *Nat. Mach. Intell.* **6**, 449–460 (2024).
123. T. Shen *et al.*, E2Efold-3D: End-to-end deep learning method for accurate de novo RNA 3D structure prediction. arXiv [Preprint] (2022). https://arxiv.org/abs/2207.01586 (Accessed 27 November 2024).
124. M. Baek, R. McHugh, I. Anishchenko, D. Baker, F. DiMaio, Accurate prediction of nucleic acid and protein–nucleic acid complexes using RoseTTAFoldNA. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.09.09.507333 (Accessed 27 November 2024).
125. Y. Kagaya *et al.*, NuFold: A novel tertiary RNA structure prediction method using deep learning with flexible nucleobase center representation. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.09.20.558715 (Accessed 27 November 2024).
126. R. Krishna *et al.*, Generalized biomolecular modeling and design with RoseTTAFold all-atom. *Science* **384**, eadl2528 (2024).
127. R. Das *et al.*, Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins: Struct., Funct., Bioinf.* **91**, 1747–1770 (2023).
128. K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
129. S. L. Bonilla, J. S. Kieft, The promise of cryo-EM to explore RNA structural dynamics. *J. Mol. Biol.* **434**, 167802 (2022).
130. M. L. Ken *et al.*, RNA conformational propensities determine cellular activity. *Nature* **617**, 835–841 (2023).
131. S. L. Bonilla, A. N. Jones, D. Incarnato, Structural and biophysical dissection of RNA conformational ensembles. *Curr. Opin. Struct. Biol.* **88**, 102908 (2024).
132. J. Kortmann, F. Narberhaus, Bacterial RNA thermometers: Molecular zippers and switches. *Nat. Rev. Microbiol.* **10**, 255–265 (2012).
133. P. Brion, E. Westhof, Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137 (1997).
134. N. I. Orlovsky, H. M. Al-Hashimi, T. G. Oas, Exposing hidden high-affinity RNA conformational states. *J. Am. Chem. Soc.* **142**, 907–921 (2019).
135. L. Herron, K. Mondal, J. S. Schneekloth, P. Tiwary, Inferring phase transitions and critical exponents from limited observations with thermodynamic maps. arXiv [Preprint] (2023). https://arxiv.org/abs/2308.14885 (Accessed 27 November 2024).