



The transformational role of GPU computing and deep learning in drug discovery

Mohit Pandey^{1,5}, Michael Fernandez^{1,5}, Francesco Gentile¹, Olexandr Isayev²,
Alexander Tropsha³, Abraham C. Stern⁴✉ and Artem Cherkasov¹✉

Deep learning has disrupted nearly every field of research, including those of direct importance to drug discovery, such as medicinal chemistry and pharmacology. This revolution has largely been attributed to the unprecedented advances in highly parallelizable graphics processing units (GPUs) and the development of GPU-enabled algorithms. In this Review, we present a comprehensive overview of historical trends and recent advances in GPU algorithms and discuss their immediate impact on the discovery of new drugs and drug targets. We also cover the state-of-the-art of deep learning architectures that have found practical applications in both early drug discovery and consequent hit-to-lead optimization stages, including the acceleration of molecular docking, the evaluation of off-target effects and the prediction of pharmacological properties. We conclude by discussing the impacts of GPU acceleration and deep learning models on the global democratization of the field of drug discovery that may lead to efficient exploration of the ever-expanding chemical universe to accelerate the discovery of novel medicines.

Originally developed to accelerate three-dimensional graphics, the benefits of GPUs for powerful parallel computing were quickly praised by the scientific community. The earliest attempts to use GPUs for scientific purposes employed the programmable shader language to run calculations. In 2007, NVIDIA released Compute Unified Device Architecture (CUDA) as an extension of the C programming language, together with compilers and debuggers, opening the floodgates for porting computationally intensive workloads into GPU accelerators. Further advances came from the release of common maths libraries such as fast Fourier transforms and basic linear algebra subroutines, which were foundational to scientific computing. In the same year, the first computational chemistry programs were ported to GPUs, enabling efficient parallelization of molecular mechanics and quantum Monte Carlo¹ calculations.

In September 2014, NVIDIA released cuDNN, a GPU-accelerated library of primitives for deep neural networks (DNNs) implementing standard routines such as forward and backward convolution, pooling, normalization and activation layers. The architectural support for training and testing subprocesses enabled by GPUs seemed to be particularly effective for standard deep learning (DL) procedures. As a result, an entire ecosystem of GPU-accelerated DL² platforms has emerged. While NVIDIA's CUDA is a more established GPU programming framework, AMD's ROCm³ represents a universal platform for GPU-accelerated computing. ROCm introduced new numerical formats to support common open-source machine learning libraries such as TensorFlow and PyTorch; it also provides the means for porting NVIDIA CUDA code into AMD hardware⁴. It is important to note that AMD not only is catching up to the ROCm platform in the GPU computing race, but also recently introduced the new flagship GPU architecture AMD Instinct MI200 Series⁵ to compete with the latest NVIDIA Ampere A100 GPU architecture⁶.

The fields of bioinformatics, cheminformatics and chemogenomics in particular, including computer-aided drug discovery

(CADD), have taken advantage of DL methods running on GPUs. Most challenges in CADD have routinely faced combinatorics and optimization problems, and machine learning has been effective at providing solutions for them⁷. Thus, major progress has been made in DL for CADD applications such as virtual screening, de novo drug design, absorption, distribution, metabolism, excretion and toxicity (ADMET) properties prediction and so on (Fig. 1).

Herein, we discuss the effects of GPU-supported parallelization and DL model development and application on the timescale and accuracy of simulations of proteins and protein–ligand complexes. We also provide examples of DL algorithms used for structure determination in cryo-electron microscopy (cryo-EM) and 3D structure prediction of proteins.

GPU computing and DL for molecular simulations

GPU acceleration comes from massive data parallelism, which arises from similar independent operations performed on many elements of the data. In graphics, an example of a common data parallel operation is the use of a rotation matrix across coordinates describing the positions of objects as a view is rotated. In a molecular simulation, data parallelism can be applied to independent calculation of atomic potential energies. Similarly, DL model training involves forward and backward passes that are commonly expressed as matrix transformations that are readily parallelizable (Fig. 2).

Accelerating molecular dynamics simulations on GPUs. The development of GPU-centred molecular dynamics codes in the past decade led hundred-fold reductions in the computational costs of simulations compared with central processing unit (CPU)-based algorithms⁸. Consequently, most molecular dynamics engines (such as AMBER (assisted model building with energy refinement)⁹, GROMACS (Groningen machine for chemical simulations)¹⁰ and NAMD (nanoscale molecular dynamics)¹¹) now provide GPU-accelerated implementations. GPUs not only are well suited

¹Vancouver Prostate Centre, Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, Canada. ²Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA. ³Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁴Nvidia Corporation, Santa Clara, CA, USA. ⁵These authors contributed equally: Mohit Pandey, Michael Fernandez. ✉e-mail: astern@nvidia.com; acherkasov@prostatecentre.com

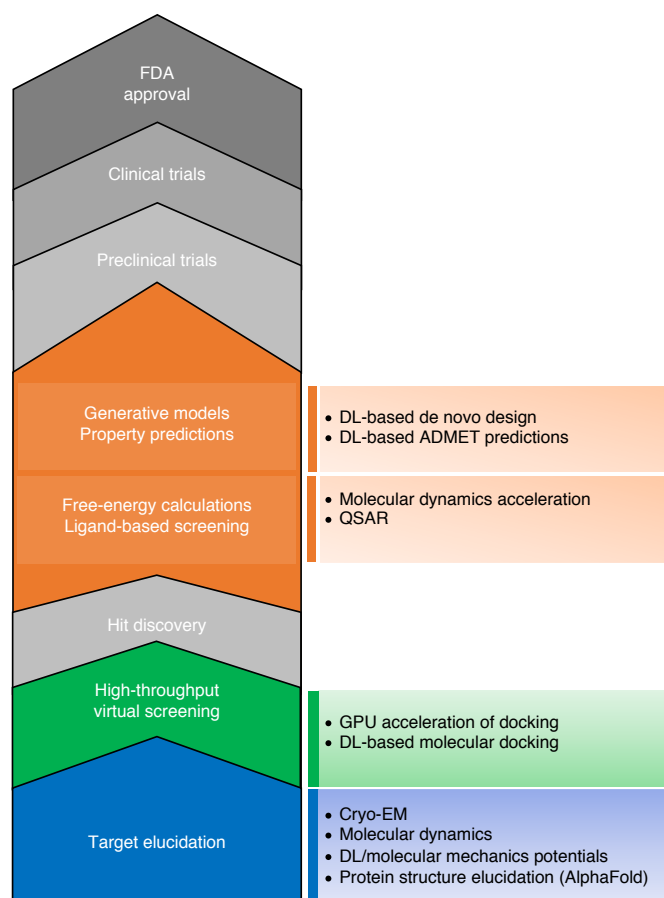


Fig. 1 | CADD workflow. GPU accelerators find applications in each step of the drug discovery and development process (shaded in colour). FDA, US Food and Drug Administration.

to accelerating molecular dynamics simulations but also scale well with system size using spatial domain decomposition¹². As a result, molecular dynamics simulations extend to a broader range of biomolecular phenomena, approaching the viral and cell level and coming closer to experimental timescales. Recent methodological and algorithmic advances enabled molecular dynamics simulations of molecular assemblies of up to 2×10^9 atoms (Fig. 3)¹³, with overall simulation times of microseconds or even milliseconds.

Free-energy simulations represent another area that continues to benefit from progress in GPU development. Methods such as relative binding free-energy calculations, thermodynamic integration and free-energy perturbation¹⁴ now allow reliable binding affinities for a large number of protein–ligand complexes to be computed. In this regard, the recent development of neural network-based force fields such as ANI (accurate neural network engine for molecular energies)¹⁵ and AIMNet (atoms-in-molecules net)¹⁶ provides industry-standard accuracy of free-energy simulations. The benchmarks with inhibitors for tyrosine-protein kinase 2 from the Schrödinger *Journal of the American Chemical Society* benchmark set¹⁷ showed that the simulations with ANI machine learning potential reduced the absolute binding free-energy errors by 50%. Frameworks such as ANI provide a systematic approach for generating atomistic potentials and drastically reduce the human effort required to fit a force field, thus automating force field development¹⁸. More recently, other DL frameworks have been proposed to further push the boundaries of molecular simulations in drug discovery¹⁹. Exemplifying these approaches, the reweighted auto-encoder variational Bayes for enhanced sampling²⁰ method was

employed successfully to simulate ligand–protein dissociation. It processed notably faster than conventional molecular dynamics, yet generated accurate estimates of binding free energies²¹ and loop conformation sampling²². Similarly, Drew Bennett et al.²³ used DNNs to predict water-to-cyclohexane transfer energies of small molecules derived from molecular dynamics simulations. The use of hybrid DL and molecular mechanics potentials²⁴ for ligand–protein simulations has also been proposed, supported by the development of open-source frameworks^{25,26}. These methods employ quantum mechanics-based DL potentials for the ligand and molecular mechanics for the surrounding environment, and have shown superior performances in reproducing binding poses²⁷ compared with conventional potentials.

Quantum mechanics and GPUs. The availability of CUDA²⁸ and OpenCL²⁹ application programming interfaces (APIs) has been key to the success of GPU applications, although programming GPUs to run chemistry codes efficiently is not trivial. To achieve high efficiency, computational threads that are grouped into blocks need to be executed simultaneously. TeraChem was the first quantum chemistry code to be written specifically for GPUs³⁰. The mixed-precision arithmetic allowed very efficient computation of Coulomb and exchange matrices³¹. The latest algorithmic developments in TeraChem allowed entire proteins to be simulated with density functional theory (DFT)³². Hybrid quantum mechanics–molecular mechanics simulations of the nonadiabatic dynamics of Bacteriorhodopsin provided insight into the light-activation machinery and a molecular-level understanding of the conversion of light energy into work³³. DFT calculations are now routine for studying protein–ligand interactions. For instance, the best calculations resulted in mean absolute errors of ~ 2 kcal mol^{−1} for protein–ligand interaction energies³³. DFT calculations of serine protease factor X and tyrosine-protein kinase 2 showed that the obtained geometries are close to the co-crystallized protein–ligand structures³⁴.

Future exascale supercomputers will provide high levels of parallelism in heterogeneous CPU and GPU environments. This scaling requires the development of new hybrid algorithms and, essentially, a complete rewrite of the scientific codes. These new developments are now being implemented as a part of the NWChemEx package³⁵. NWChemEx will offer the possibility of performing quantum mechanics and molecular mechanics simulations for systems that are several orders of magnitude larger than those that are tractable by canonical formulations of theoretical methods³⁵.

GPU acceleration of protein structure determination. High-throughput and automation of cryo-EM have become increasingly important as the state-of-the-art experimental technique used for protein structure determination for use in structure-based drug design³⁶. DL-based approaches, such as DEEMap³⁷ and DeepPicker³⁸, have been developed to accelerate processing of cryo-EM images. The DEEMap method directly extracts structure dynamics associated with hidden atomic fluctuations by combining DL and molecular dynamics simulations that learn the relationships between local density data. DeepPicker employs convolutional neural networks (CNNs) and cross-molecule training to capture common features of particles from previously analysed micrographs, which facilitates automatic particle picking in single-particle analysis. This tool serves to illustrate that DL integration can successfully address current gaps towards fully automated cryo-EM pipelines, paving the way for a new multidisciplinary approach to protein science^{37,38}.

In addition to accelerated experimental characterization of protein structures by cryo-EM, the recent ground-breaking success of DeepMind with the AlphaFold-2 method in the Critical Assessment of Protein Structure Prediction (CASP) challenge hints at the future impacts of DL algorithms in protein structural characterization and

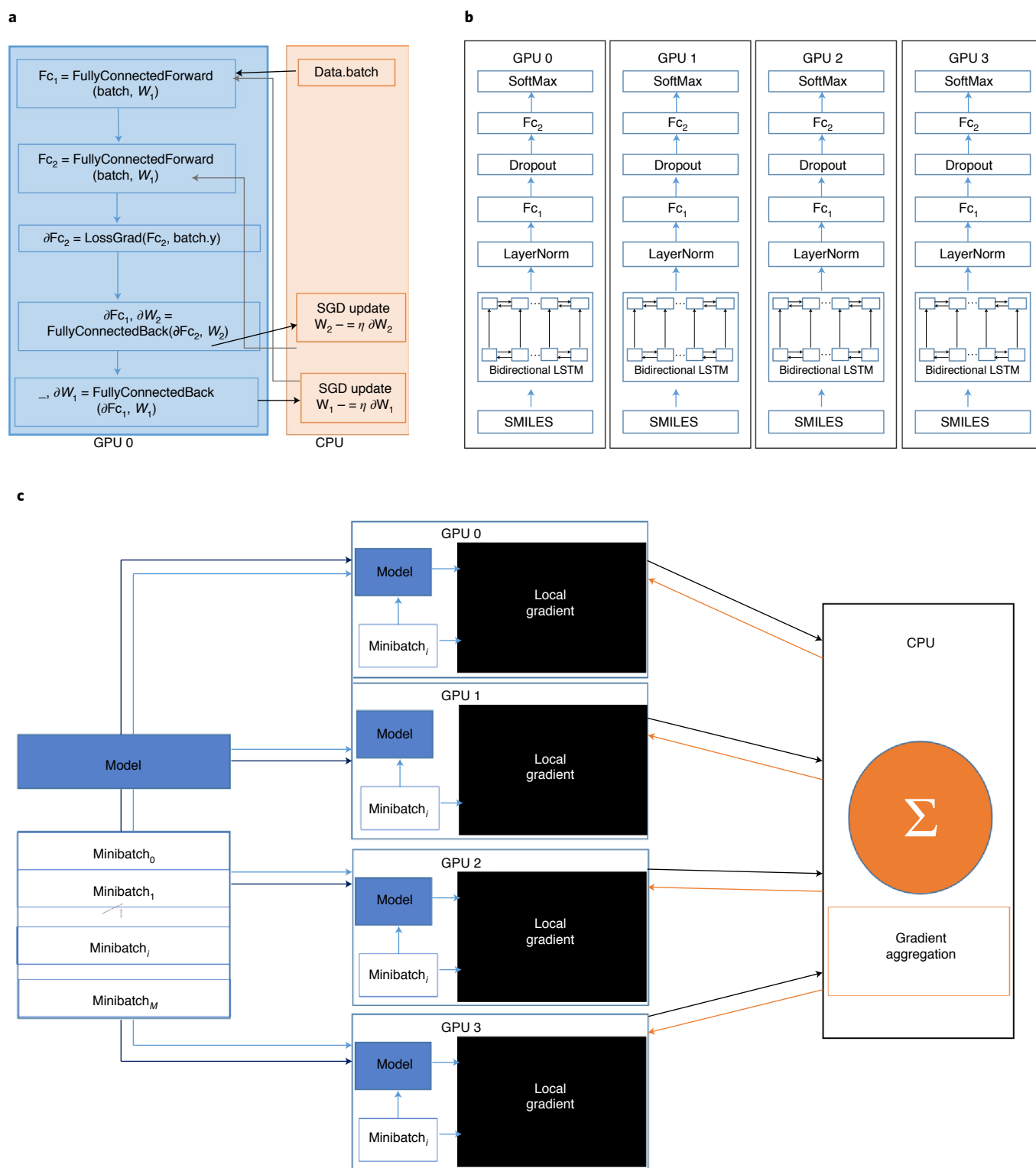


Fig. 2 | Parallelization of DL architectures in single- and multi-GPU environments. Neural network arithmetic operations are based on matrix multiplications that are parallelized by GPUs using block multiplication and aggregation¹³¹. **a**, Distribution of computational graph over one GPU for a two-layered multilayer perceptron (MLP). W , trainable parameters; SGD, stochastic gradient descent algorithm; η , learning rate of the stochastic gradient descent algorithm. **b**, Data parallelization. Each GPU stores a network copy. Data parallelization is the most commonly adopted GPU paradigm for accelerating DL¹³². A copy of the network resides in each GPU, and each GPU gets its own dedicated minibatch of data to train on. The computed gradients and losses are then transferred to a shared device (typically the CPU) for aggregation before being rebroadcast to GPUs for parameter updates. LayerNorm, Dropout, Fc, SoftMax and Bidirectional LSTM (long short-term memory) are modules of an arbitrary neural network topology used for demonstration. **c**, Forward and backpropagation for a gradient minibatch descent algorithm. M , total mini-batches for the data.

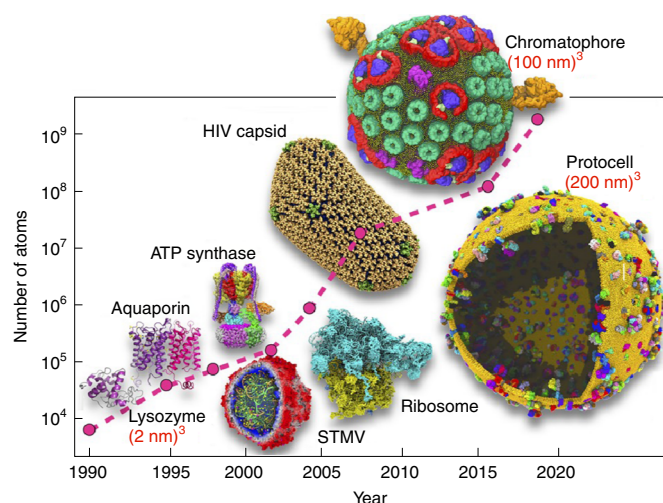


Fig. 3 | Timeline of the complexity of biological systems that could be simulated with molecular dynamics. Continuous development effort over the years towards simulating with NAMD realistic biological objects of increasing complexity from a small, solvated protein, on the thousand-atom size scale, in the early 1990s, to a full protocell, on the billion-atom size scale, now. ATP, adenosine triphosphate; HIV, human immunodeficiency virus; STMV, satellite tobacco mosaic virus. Figure reproduced with permission from ref.¹³, AIP Publishing.

the expansion of the druggable proteome³⁹. AlphaFold-2 can regularly predict protein geometry with atomic accuracy without being previously exposed to similar structures. The recently updated neural network-based model demonstrated an accuracy competitive with experiments in most cases, and greatly outperformed other methods at the 14th CASP competition. The DL model behind AlphaFold-2 incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments to crack one of the oldest problems in biology. AlphaFold-2 was employed to predict the structures of nearly every known human protein and other organisms important to medical research, a total of 350,000 proteins, which represents an impressive achievement for biomedical research³⁹.

The emergence of DL in CADD

Advances in DL, particularly in computer vision and language processing, revived the recent interest of CADD researchers in neural networks. Merck is credited with popularizing DL for CADD through the Kaggle competition on Molecular Activity Challenge in 2012 (ref.⁴⁰). The winning solution by Dahl et al.⁴¹ leveraged a multitask learning approach to train a DNN. Thereafter, many researchers embraced such models for drug discovery problems. These include the evaluation of the predictors of the pharmacokinetic behaviour of therapeutics and their adverse effects⁴², the prediction of small molecule–protein binding⁴³, the determination of chemotherapeutic responses of carcinogenic cells⁴⁴, the quantitative estimation of drug sensitivity⁴⁵ and quantitative structure–activity relationship (QSAR) modelling⁴⁶, among others.

The emergence of GPU-enabled DL architectures, along with the proliferation of chemical genomics data, has led to meaningful CADD-enabled discoveries of clinical drug candidates. Furthermore, artificial intelligence (AI)-driven companies (such as BenevolentAI, Insilico Medicine and Exscientia, among others) are reporting successes in augmented drug discovery. For example, Exscientia developed a drug candidate, DSP-1181, to be used against obsessive-compulsive disorder that entered phase 1 clinical trials less than 12 months from its conception using AI approaches⁴⁷. Insilico Medicine just began a clinical trial with its first

AI-developed drug candidate to treat idiopathic pulmonary fibrosis and BenevolentAI identified baricitinib⁴⁸ as a potential treatment for COVID-19 (ref.⁴⁹). These recent success cases indicate that further promotion and application of AI-driven approaches supported by GPU computing could greatly accelerate the discovery of novel and improved medicines.

DL architectures for CADD. From discriminative neural networks that find applications in virtual screening of existing or synthetically feasible chemical libraries to the recent success of DL generative models that has inspired their use in de novo drug design, Fig. 4 depicts the general scheme of commonly used state-of-the-art DL architectures. Table 1 enumerates their adoption in CADD.

MLPs. Multilayer perceptrons (MLPs) are fully connected networks with input, hidden and output layer(s) and nonlinear activation functions (sigmoid, tanh, ReLU (rectified linear unit) and so on) that are the basis of DNNs⁵⁰. Their large learning capacity and relatively small numbers of parameters made MLPs the earliest successful application of artificial neural networks in drug discovery for QSAR studies⁵¹. Modern GPU machines render MLPs inexpensive models that are suitable for the large cheminformatics datasets that are having a renewed impact on CADD⁵².

CNNs. Arguably the most utilized DNNs, CNNs are guided by hierarchical principles and utilize small receptive fields to process local subsections of the input. CNNs have been the go-to architecture for image and video processing, while they also enable success in biomedical text classification⁵³. A typical CNN operates on a 3D volume (height, width, channel), generates translation-invariant feature maps based on learnable kernels and pools these maps to produce scale- and rotation-invariant outputs.

The parallelizable nature of convolution operation makes CNNs suitable for implementation on GPUs. The Toxic Color⁵⁴ method was first developed with the Tox21 benchmark data using simple 2D drawings of chemicals, demonstrating that GPU-enabled CNN predictions, without employing any chemical descriptors, were comparable to state-of-the-art machine learning methods. Goh et al.⁵⁵ subsequently introduced Chemception, a CNN trained on molecular drawings to predict chemical properties such as toxicity, activity and solvation, which showed comparable performance to MLPs trained with extended-connectivity fingerprints. Their model was further improved by encoding atom- and bond-specific chemical information into the CNN⁵⁵.

RNNs. Historically, computational chemists have relied extensively on topological fingerprints such as extended-connectivity fingerprints⁵⁶ or other descriptors for molecular characterization⁵⁷. One popular linear Goh representation is SMILES (simplified molecular input line entry system)⁵⁸. String representations of fixed length are useful because they can be treated as sequences and efficiently modelled within temporal networks such as recurrent neural networks (RNNs). RNNs may be viewed as an extension of Markov chains with memory that are capable of learning long-range dependencies through its internal states, and hence modelling autoregression in molecular sequences.

The capacity of DL algorithms to learn latent internal representations for the input molecules without the need for hand-crafted descriptors allows syntactically and semantically meaningful representations specific to the dataset and problem at hand. SMILES2vec⁵⁹ was trained to learn continuous embeddings from SMILES representations to make predictions for several datasets and tasks (toxicity, activity, solvation and solubility). The lower dimensionality of these vectors speeds training and reduces memory requirements—both of which are critical aspects of training neural networks. Inspired by the success of popular word-embedding algorithm word2vec, Jaeger et al.⁶⁰

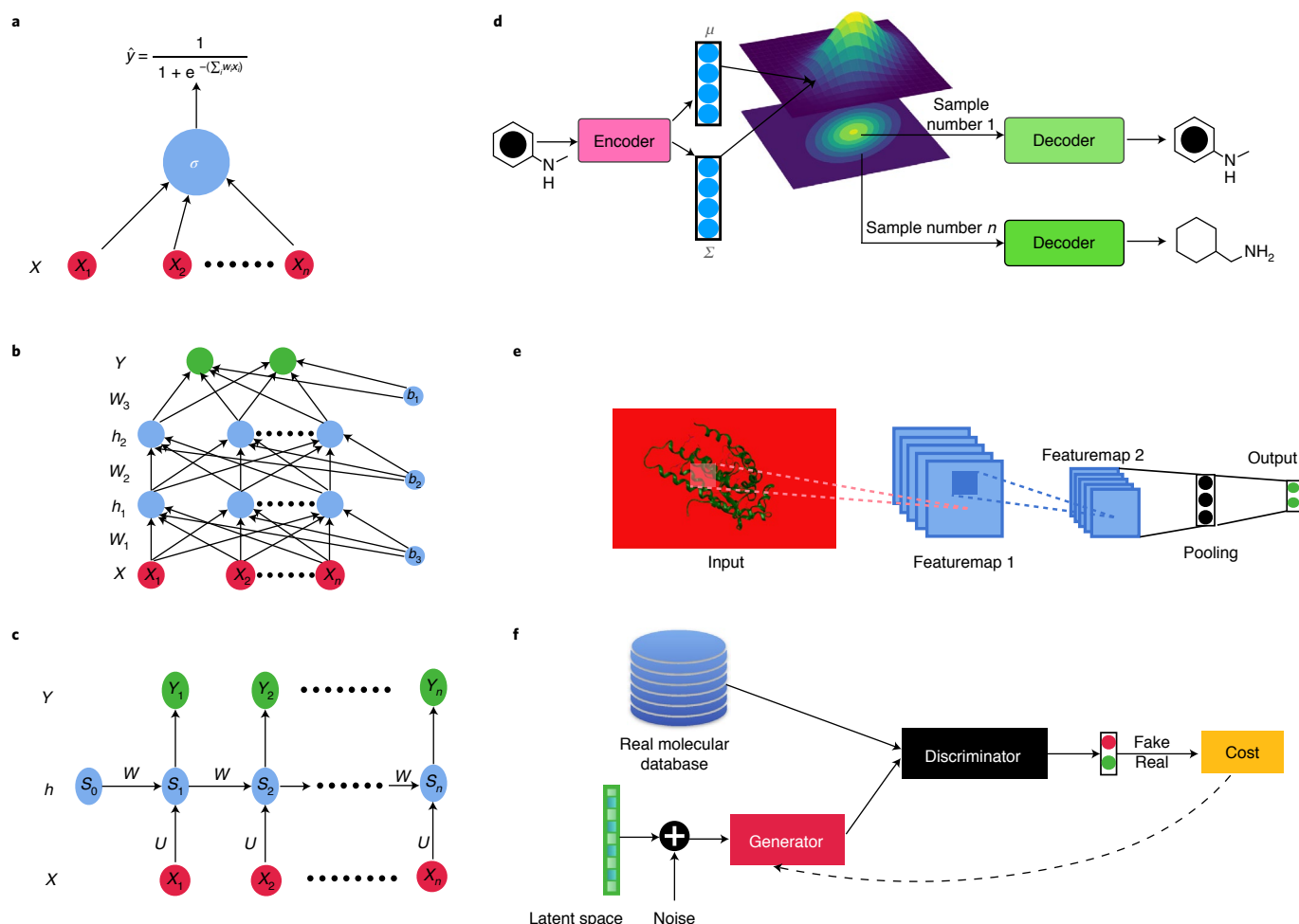


Fig. 4 | Architectures of several popular neural networks. **a**, Sigmoid neuron as a building block for neural networks. A sigmoid neuron is a perceptron with sigmoid nonlinearity. **b**, A fully connected feed-forward neural network (MLP) consists of an input layer, hidden layer(s) and output layer with non-linear activations such as sigmoid. X and Y represent input and output, respectively, from the models. h , hidden layer; b , bias term. **c**, A simplified unfolded representation of an RNN. U and W are trainable model parameters; S_i is the latent state at the 'ith' timestep of an RNN input. **d**, VAE. A probabilistic encoder maps the input into a latent space under a Gaussian assumption. μ and Σ are the parameter vectors of learned multivariate Gaussian distribution. Samples are drawn from this latent space and decoder attempts to reconstruct original input from these samples. **e**, CNN. Kernels are convolved over input image and subsequently over feature maps to progressively generate higher-order feature maps. Pooling further reduces the dimensionality of the feature maps. **f**, GAN. The discriminator and generator are two arbitrary neural networks that compete in a zero-sum game to synthetically generate new samples. These large-capacity DL models cannot be reasonably trained without using a hardware accelerator such as a GPU. It is implied (unless otherwise stated) that such models are deployed on GPUs.

developed mol2vec. Based on unsupervised pretraining of word2vec on ZINC and ChEMBL datasets, the learned representations achieved state-of-the-art performance and were better suited to regression tasks than Morgan fingerprints.

VAEs. Variational autoencoders (VAEs)⁶¹ are deep generative models that are revolutionizing cheminformatics owing to their capacity to probabilistically learn latent space from observed data that can later be sampled to generate new molecules with fine-tuned functional properties. VAEs support direct sampling, and hence generation, of molecules from a learned distribution over the latent space without the need for expensive Monte Carlo sampling. Blaschke et al.⁶² generated new molecules targeting dopamine receptor 2 using a VAE model. These molecules were further validated using a support vector machine model trained for activity prediction. Sattarov et al.⁶³ explored Seq2Seq VAEs to selectively design compounds with desired properties. A generative topographic mapping was used to sample from the latent representation learned by the

VAE. Other studies investigated VAEs in conjunction with molecular graphs to generate new molecules⁶⁴.

GANs. Recently, generative adversarial networks (GANs) have established themselves as powerful and diverse deep generative models. GANs are based on an adversarial game between a generator and a discriminator module. The objective of the discriminator network is to differentiate between real and fake datapoints generated by the generator network. A concurrently trained generator network attempts to create novel datapoints such that the discriminator is manipulated into believing the generated results to be real. Following the empirical success of GANs, several improvements and modifications were proposed⁶⁵. These methods were promptly utilized by researchers in drug discovery to artificially synthesize data across subproblems⁶⁶. Méndez-Lucio et al.⁶⁷ investigated a GAN-based generative modelling approach at the intersection of systems biology and molecular drug design. Their attempt to bring biology and chemistry together was demonstrated in the generation of active-like molecules given

Table 1 | State-of-the-art DL categories and their applications in drug discovery

Learning category	Architecture/algorithm	Application
Generative models	Graph neural network (GNN)	Molecular representation ¹¹¹
		ADMET prediction ¹¹¹
		Energy prediction ¹¹²
		Drug–target interaction ⁷⁸
		Molecular property prediction ¹¹³
	Generative adversarial networks (GANs)	De novo drug design ¹¹⁴
	Variational autoencoder (VAE)	De novo drug design ¹¹⁵
	Transformer	Drug–target interaction ⁷⁰
		Molecular representation ⁷¹
		De novo drug design ⁷²
Representation learning	SMILES2vec	Continuous molecular embedding ⁵⁹
	ProtVec	Continuous molecular embedding ¹¹⁶
Transfer learning	DenseNet (CNN)	virtual screening ¹¹⁷
	Multitask learning	Toxicity prediction ¹¹⁸
		Multiple pharmacokinetic parameter prediction ¹¹⁹
		Binding mode, affinity and activity prediction ¹²⁰
		ADMET prediction ¹¹¹
Federated learning	Horizontal federated learning	QSAR analysis ¹²¹
		Solubility prediction ¹²²
Deep reinforcement learning	Stack-RNN + {reward-based RL}	De novo drug design ¹²³
	Long short-term memory (LSTM) + {policy gradient, regularized policy gradient, advantage actor-critic network, proximal policy optimization}	De novo drug design ¹²⁴
	LSTM + {actor-critic network}	De novo drug design ¹²⁵
	RNN + {policy-based RL}	De novo drug design ⁸²
Interpretable artificial intelligence	Uncertainty estimation	Activity prediction ¹²⁶
		Reaction prediction ¹²⁷
		Molecular property prediction ¹²⁸
	GNN	ADMET ¹²⁹
		Chemical reactivity prediction ¹³⁰

{ } indicates the policy iteration algorithm used for reinforcement learning (RL).

the gene expression signature of the target. To this end, they used a combination of conditional GANs and a Wasserstein GAN with a gradient penalty. GANs have also been explored in conjunction with genetic algorithms to combat mode collapse and hence incrementally explore a larger chemical space⁶⁸.

Transformer networks. Inspired by tremendous success of the use of transformer networks⁶⁹ in natural language processing, DL researchers in drug discovery were motivated to explore its power for training long-term dependencies for sequences. Using self-attention, Shin et al.⁷⁰ performed end-to-end neural regressions to predict affinity scores between drug molecules and target proteins. In doing so, they learned molecular representations for the drug molecules by aggregating molecular token embedding with position embedding, as well as learning new representations for proteins using a CNN. In the same vein, Huang et al.⁷¹ introduced MolTrans to predict drug–target interactions. Grechishnikova formulated target-specific molecular generation as a translation task

between amino acid chains and their SMILES representations using a transformer encoder and decoder⁷².

GNNs. A recent innovation in the use of DL on non-Euclidean data such as graphs, point clouds and manifolds promoted graph neural networks (GNNs)⁷¹. The central form taken by the majority of GNN variants is neural message parsing in which messages from each node in the graph are exchanged and updated iteratively using neural networks, thereby generating robust representations. PyTorch Geometric⁷³ provides CUDA kernels for message parsing APIs by leveraging sparse GPU acceleration. Deep Graph Library-LifeSci⁷⁴ unifies several seminal works to introduce a platform-agnostic API for the easy integration of GNNs in life sciences with a particular focus on drug discovery. The mathematical representation for graphs succinctly captures the graphical structure of molecules, meaning that GNNs are potentially of great use in CADD.

Duvenaud et al.⁷⁵ showed that learned graph representations for drugs outperform circular fingerprints on several benchmark

datasets. Inspired by gated GNNs, PotentialNet⁷⁶ showed improved performance at ligand-based multitasks (electronic property, solubility and toxicity prediction). Several other studies demonstrated improved predictive performance when geometric features such as atomic distances were also considered⁷⁷. Torng et al.⁷⁸ used graph autoencoders to learn protein representations from their amino acid residues, along with graph representations of protein pockets. These vectors were then concatenated with graph representations for drug molecules and fed into an MLP to predict drug–protein associations. Gao et al.⁷⁹ learned protein and drug embeddings using RNNs and GNNs on protein sequences and atomic graphs of drugs, respectively. One popular approach to the repurposing of drugs involves the completion of knowledge graphs; these large knowledge graphs are built from the known similarities between diseases, drugs and indications⁸⁰. Gaudelot et al. presented an extensive review of GNNs for CADD applications⁸¹.

Reinforcement learning. Reinforcement learning is a branch of AI that simulates decision-making through the optimization of reward- and penalty-based policies. With the penetration of DL, deep reinforcement learning has found applications in CADD, particularly in de novo drug design, by enabling molecules to have desired chemical properties^{82,83}. Deep reinforcement learning trained on GNNs was further shown to improve the validity of the molecular structures generated⁸⁴. Enforcing chemically meaningful actions simultaneously with optimizing rewards around chemical properties generates useful leads while imparting chemistry domain knowledge to otherwise largely black-box DL solutions⁸⁵.

Scaling up virtual screening with GPUs and DL. Structure-based virtual screening and ligand-based virtual screening aim to rank chemical compounds on the basis of their computed binding affinity to a target, and to extrapolate structural similarities between small molecules to functional equivalence, respectively. With the exponential growth of purchasable ligand libraries, already comprising tens of billions of synthesizable molecules⁸⁶, there is increasing interest in expanding the scale at which conventional virtual screening operates with the parallelization of docking calculations or DL-based acceleration.

A number of structure-based virtual screening methods have been developed recently to efficiently screen billion-entry chemical libraries. VirtualFlow⁸⁷ represents the first example of such platforms, allowing a billion molecules to be screened on large CPU clusters (~10,000 cores) in a couple of weeks while displaying a linear scaling behaviour. Differently from VirtualFlow and other CPU-based methods⁸⁸, GPU acceleration of docking algorithms using OpenCL and CUDA libraries has partially addressed the high-throughput bottleneck by dividing the whole protein surface into arbitrary independent regions (or spots)⁸⁹ or by combining both multicore CPU architectures and GPU accelerators in heterogeneous computing systems⁹⁰. A recent example of such strategies is Autodock-GPU, which allows a billion molecules to be screened in a day on large GPU clusters such as the Summit supercomputer (~27,000 GPUs) by parallelizing the pose search process⁹¹. These approaches that leverage GPU computing on high-performance computing will therefore probably become instrumental in identifying novel lead compounds from large, diverse chemical libraries, or accelerating other structure-based methods such as inverse docking⁹². Still, the costs of computing remain high and can be prohibitive for drug discovery organizations that cannot access elite supercomputing clusters.

On the other hand, alternative structure-based virtual screening platforms have recently emerged, leveraging DL predictions and molecular docking to boost the selection of active compounds from large libraries with limited computational resources. The common strategy among these methods is the implementation of DL

emulators of classical computational screening scores that rely on an order-of-magnitude higher inference speed than conventional docking. Predictive DL models are built using a variety of chemical structure representations, from molecular fingerprints to more sophisticated embeddings, to filter out large portions of a chemical library. One of the earliest developed methods, Deep Docking⁹³, relies on a fully connected MLP model that is trained with chemical fingerprints and scores of a small portion of a library, then used to predict the docking score classes of the remaining molecules, allowing low-ranked entries to be removed without docking them. Deep Docking was initially deployed by Ton et al.⁹⁴ to screen 1.3 billion molecules from ZINC15 using Glide against SARS-CoV-2 main protease. More recently, it was also applied sequentially on different docking programs to screen 40 billion commercially available molecules against SARS-CoV-2 main protease by Gentile et al., leading to the identification of novel experimentally confirmed inhibitor scaffolds⁹⁵. Other similar methods have been proposed that rely on DL models that predict docking outcomes, such as MolPAL (molecular pool-based active learning)⁹⁶ and AutoQSAR/DeepChem⁹⁷. Hofmarcher et al.⁹⁸ also performed ligand-based virtual screening on the ZINC database with over 1 billion compounds to rank potential SARS-CoV-2 inhibitors using an RNN. Compared with brute-force methods, these DL-based approaches may play an important role in making the chemical space accessible to academic research groups and small/medium industry alike.

GPU-enabled DL promotes open science and the democratization of drug discovery

The integration of DL in CADD as presented here has contributed greatly to the global democratization of drug discovery and open science efforts. The open-source DL packages DeepChem⁹⁹, ATOM¹⁰⁰, Deep Docking⁹³, MolPAL⁹⁶, OpenChem¹⁰¹, GraphInvent¹⁰² and MOSES¹⁰³, among others, have simplified the integration of DL strategies into drug discovery pipelines using popular machine learning libraries including (but not limited to) scikit-learn, Tensorflow and Pytorch. The growing demand for large datasets for DL models is naturally encouraging data-sharing practices and calls for broader open data policies. Furthermore, GPU acceleration in cloud-native computing and micro-service-oriented architectures could make CADD methods free and widely available, contributing to standardizing computational modules and tools, as well as architectures, platforms and user interfaces. DL solutions can take advantage of public cloud services such as Amazon Web Services, Google Cloud Platform and Microsoft Azure to boost drug discovery by reducing the cost.

As exciting as these new DL-enabled modelling opportunities are, CADD scientists need to be cautious about the expected impact of DL technologies. Realistic expectations need to be derived from the lessons learned and best practices developed during more than 20 years of data-driven molecular modelling¹⁰⁴. For example, the quality, quantity and diversity of data can hamper not only the accuracy but also the overall generality of CADD models. Thus, data cleaning and curation will continue to play a major role that can alone determine the success or failure of such DL applications¹⁰⁴. On the other hand, the use of dynamic datasets derived from guided experiments or high-level computer simulations can facilitate the utilization of active learning strategies. Interactive training and validation can substantially improve model quality, as implemented by the AutoQSAR tool¹⁰⁵. Beyond predictive models, DL solutions are particularly useful when combining generative models and RL-based decision-making approaches. An optimization of reward- and penalty-based rules could enable unprecedented ‘à la carte’ design of chemical structures with desired chemical and functional properties^{82,83}. This method of simultaneously enforcing chemically and biologically meaningful actions into de novo drug design represents a drastic departure from the more traditional black-box DL solutions.

Open science efforts are benefiting from recent end-to-end DL models that can be implemented at all stages of drug discovery using GPUs¹⁰⁶. One such recently developed platform is IMPECABLE¹⁰⁷, which integrates multiple CADD methods. Al Saadi et al.¹⁰⁷ combined the strength of molecular dynamics in predicting binding free energies with the strength of docking in pose prediction. Their solution automates not just virtual screening, but also lead refinement and optimization.

NVIDIA Clara Discovery is a collection of GPU-accelerated frameworks, tools and applications for computational drug discovery spanning molecular simulation, virtual screening, quantum chemistry, genomics, microscopy and natural language processing¹⁰⁸. These platforms are intended to be open and cross-compatible, and are expected to accelerate the integration of different data sources across the biopharmaceutical spectrum from research papers, patient records, symptoms and biomedical images to genes, proteins and drug candidates.

Many major hardware producers now use their computing expertise to enter the realm of supercomputing by employing multiple GPU clusters to train large-capacity DL models for reaction prediction, molecular optimization and de novo molecular generation. The adoption of DL emulation of pharmaceutical endpoints⁹³ by CADD platforms can make drug discovery on libraries containing tens of billions of compounds affordable, even for small companies and academic labs without access to elite computational facilities.

Owing to the legal complexities, sharing of proprietary data between institutions continues to act as a bottleneck in streamlined drug discovery research. Federated learning allows participating institutions to perform localized training on their respective unshared data. Trained local models are then aggregated in a central server for broader accessibility. Federated learning thus supports democratization by alleviating data-exchange challenges to some degree, although effective model aggregation remains an active area of research.

Conclusions and outlook

Modern drug discovery has benefited from the recent explosion of DL models and GPU parallel computing. Driven by hardware advances, DL has demonstrated excellence in drug discovery problems ranging from virtual screening and QSAR analysis to generative drug design. De novo drug design in particular has been one of the major beneficiaries of advancements in GPU computation as it leverages large capacity and highly parameterized models such as VAE and GANs that cannot be reasonably deployed without using hardware accelerators such as GPUs. The ever-improving price-to-performance ratio of GPU hardware, reliance of DL on GPU and wide adoption of DL in CADD in recent years are all evident from the fact that over 50% of all 'AI in chemistry' documents in CAS Content Collection have been published in the past 4 years (ref. ¹⁰⁹). Furthermore, hybrid AI methods have been adopted that combine conventional molecular simulations with DL for fast and accurate screening of ultra-large chemical libraries approaching hundreds of billions of molecules. We expect that the growing availability of increasingly powerful GPU architectures, together with the development of advanced DL strategies and GPU-accelerated algorithms, will help to make drug discovery affordable and accessible to the broader scientific community worldwide.

Another key driver of DL algorithms is the availability of 'big data'. With the growing ease of genetic sequencing and high-throughput screening, large volumes of pristine data are now readily available to researchers in data-driven computational chemistry. However, the high-quality labelled data that are essential for supervised learning methods are still expensive to curate. Methods that build on learning from auxiliary datasets, knowledge transfer using transfer learning and label-conservative methods such as zero-shot learning have thus become a central piece of DL for drug discovery. The reliability and generalizability of any DL method developed for drug

discovery critically depends on the quality of the sourced data. Thus, data cleaning and curation play a major role that can solely define the success or failure of such DL applications¹¹⁰ and, consequently, in-depth exploration of the putative benefits of centralized, processed and well-labelled data repositories remains an open field of research.

Overall, researchers in drug discovery and machine learning have efficiently collaborated to identify CADD subproblems and corresponding DL tools. We believe that the next few years will see these applications be fine-tuned and mature, and this collaboration will further evolve to other underexplored areas of the life sciences. As such, federated learning and collaborative machine learning are gaining traction, and we believe they will be the forebears of the democratized drug discovery revolution.

Received: 26 May 2021; Accepted: 15 February 2022;

Published online: 23 March 2022

References

- Stone, J. E. et al. Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **28**, 2618–2640 (2007).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- This Review article succinctly captures key areas of DL and the most popular architectural paradigms used across domains and modalities.**
- ROCm, a New Era in Open GPU Computing (AMD Corporation, 2021); <https://rocm.github.io/rocmcloc.html>
- Shafie Khorassani, K. et al. Designing a ROCm-aware MPI library for AMD GPUs: early experiences. In *High Performance Computing Lecture Notes in Computer Science* Vol. 12728 (eds. Chamberlain, B. L., Varbanescu, A.-L., Ltaief, H. & Luszczek, P.) 118–136 (Springer, 2021).
- AMD Instinct MI Series Accelerators (AMD Corporation, 2021); <https://www.amd.com/en/graphics/instinct-server-accelerators>
- NVIDIA A100 Tensor Core GPU (NVIDIA Corporation, 2021); <https://www.nvidia.com/en-us/data-center/a100/>
- Vamathevan, J. et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
- Harvey, M. J. & De Fabritiis, G. High-throughput molecular dynamics: the powerful new tool for drug discovery. *Drug Discov. Today* **17**, 1059–1062 (2012).
- Case, D. A. et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
- Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
- Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
- Nyland, L. et al. Achieving scalable parallel molecular dynamics using dynamic spatial domain decomposition techniques. *J. Parallel Distrib. Comput.* **47**, 125–138 (1997).
- Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 44130 (2020).
- Abel, R., Wang, L., Harder, E. D., Berne, B. J. & Friesner, R. A. Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.* **50**, 1625–1632 (2017).
- Yoo, P. et al. Neural network reactive force field for C, H, N, and O systems. *NPJ Comput. Mater.* **7**, 9 (2021).
- Zubatyyuk, R., Smith, J.S., Leszczynski, J. & Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **5**, eaav6490 (2021).
- Wang, L. et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015).
- Devereux, C. et al. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).
- Noé, F., Tkatchenko, A., Müller, K. R. & Clementi, C. Machine learning for molecular simulation. *Ann. Rev. Phys. Chem.* **71**, 361–390 (2020).
- Ribeiro, J. M. L., Bravo, P., Wang, Y. & Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **149**, 72301 (2018).
- Lamim Ribeiro, J. M. & Tiwary, P. Toward achieving efficient and accurate ligand-protein unbinding with deep learning and molecular dynamics through RAVE. *J. Chem. Theory Comput.* **15**, 708–719 (2019).

22. Smith, Z., Ravindra, P., Wang, Y., Cooley, R. & Tiwary, P. Discovering protein conformational flexibility through artificial-intelligence-aided molecular dynamics. *J. Phys. Chem. B* **124**, 8221–8229 (2020).
23. Drew Bennett, W. F. et al. Predicting small molecule transfer free energies by combining molecular dynamics simulations and deep learning. *J. Chem. Inf. Model.* **60**, 5375–5381 (2020).
24. von Lilienfeld, O. A. Quantum machine learning in chemical compound space. *Angew. Chem. Int. Ed.* **57**, 4164–4169 (2018).
25. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S. & Roitberg, A. E. TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* **60**, 3408–3415 (2020).
26. Doerr, S. et al. TorchMD: a deep learning framework for molecular simulations. *J. Chem. Theory Comput.* **17**, 2355–2363 (2021).
27. Lahey, S. L. J. & Rowley, C. N. Simulating protein-ligand binding with neural network potentials. *Chem. Sci.* **11**, 2362–2368 (2020).
28. Vingelmann, P. & Fitzek, F. H. P. CUDA release 10.2.89 (NVIDIA, 2020).
29. Stone, J. E., Gohara, D. & Shi, G. OpenCL: a parallel programming standard for heterogeneous computing systems. *Comput. Sci. Eng.* **12**, 66–72 (2010).
30. Ufimtsev, I. S. & Martínez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *J. Chem. Theory Comput.* **4**, 222–231 (2008).
31. Asadchev, A. & Gordon, M. S. New multithreaded hybrid CPU/GPU approach to Hartree–Fock. *J. Chem. Theory Comput.* **8**, 4166–4176 (2012).
32. Seritan, S. et al. TeraChem: a graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, e1494 (2021).
33. Yu, J. K., Liang, R., Liu, F. & Martínez, T. J. First-principles characterization of the elusive I fluorescent state and the structural evolution of retinal protonated Schiff base in bacteriorhodopsin. *J. Am. Chem. Soc.* **141**, 18193–18203 (2019).
34. Ehrlich, S., Göller, A. H. & Grimme, S. Towards full quantum-mechanics-based protein-ligand binding affinities. *ChemPhysChem* **18**, 898–905 (2017).
35. Kowalski, K. et al. From NWChem to NWChemEx: evolving with the computational chemistry landscape. *Chem. Rev.* **121**, 4962–4998 (2021).
36. Banerjee, S. et al. 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871–875 (2016).
37. Matsumoto, S. et al. Extraction of protein dynamics information from cryo-EM maps using deep learning. *Nat. Mach. Intell.* **3**, 153–160 (2021).
38. Al-Azzawi, A. et al. DeepCryoPicker: fully automated deep neural network for single protein particle picking in cryo-EM. *BMC Bioinform.* **21**, 509 (2020).
39. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
40. Markoff, J. Scientists see advances in deep learning a part of artificial intelligence. *New York Times* (23 November 2012).
41. Dahl, G. E., Jaitly, N. & Salakhutdinov, R. Multi-task neural networks for QSAR predictions. Preprint at <https://arxiv.org/abs/1406.1231> (2014).
42. **Inspired by the winning solution of the Merck QSAR competition, this work used neural networks to predict activities of compounds for multiple assays. This was a pivotal work in popularizing DL in drug discovery.**
43. Yang, M. et al. Linking drug target and pathway activation for effective therapy using multi-task learning. *Sci. Rep.* **8**, 18322 (2018).
44. Lee, K. & Kim, D. In-silico molecular binding prediction for human drug targets using deep neural multi-task learning. *Genes* **10**, 906 (2019).
45. Tan, M. Prediction of anti-cancer drug response by kernelized multi-task learning. *Artif. Intell. Med.* **73**, 70–77 (2016).
46. Yuan, H., Paskov, I., Paskov, H., González, A. J. & Leslie, C. S. Multitask learning improves prediction of cancer drug sensitivity. *Sci. Rep.* **6**, 31619 (2016).
47. Simões, R. S., Maltarollo, V. G., Oliveira, P. R. & Honório, K. M. Transfer and multi-task learning in QSAR modeling: advances and challenges. *Front. Pharmacol.* **9**, 74 (2018).
48. Burki, T. A new paradigm for drug development. *Lancet Digit. Heal.* **2**, e226–e227 (2020).
49. Al-discovered novel antifibrotic drug goes first-in-human. *Insilico Medicine* <https://insilico.com/blog/fih> (30 November 2021).
50. Richardson, P. et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* **395**, e30 (2020).
51. Ruppert, D. The elements of statistical learning: data mining, inference, and prediction. *J. Am. Stat. Assoc.* **99**, 567–567 (2004).
52. Aoyama, T., Suzuki, Y. & Ichikawa, H. Neural networks applied to structure-activity relationships. *J. Med. Chem.* **33**, 905–908 (1990).
53. Bertoni, M. et al. Bioactivity descriptors for uncharacterized chemical compounds. *Nat. Commun.* **12**, 3932 (2021).
54. **The generation of bioactivity signatures or fingerprints is reported using a collection of DNNs derived from broadly released bioactivity data that are relevant to capturing known biological properties, showing a substantial improvement in performance across a series of biophysics and physiology activity prediction benchmarks.**
55. Pandey, M. et al. Extraction of radiographic findings from unstructured thoracoabdominal computed tomography reports using convolutional neural network based natural language processing. *PLoS ONE* **15**, e0236827 (2020).
56. Fernandez, M. et al. Toxic colors: the use of deep learning for predicting toxicity of compounds merely from their graphic images. *J. Chem. Inf. Model.* **58**, 1533–1543 (2018).
57. Goh, G. B., Siegel, C., Vishnu, A., Hodas, N. & Baker, N. How much chemistry does a deep neural network need to know to make accurate predictions? In *2018 IEEE Winter Conference on Applications of Computer Vision* 1340–1349 (IEEE, 2018).
58. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
59. Sahoo, S., Adhikari, C., Kuanar, M. & Mishra, B. A short review of the generation of molecular descriptors and their applications in quantitative structure property/activity relationships. *Curr. Comput. Aid. Drug Des.* **12**, 181–205 (2016).
60. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
61. Goh, G. B., Hodas, N. O., Siegel, C. & Vishnu, A. SMILES2vec: an interpretable general-purpose deep neural network for predicting chemical properties. Preprint at <https://arxiv.org/abs/1712.02034> (2017).
62. Jaeger, S., Fulle, S. & Turk, S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J. Chem. Inf. Model.* **58**, 27–35 (2018).
63. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at <https://doi.org/10.48550/arXiv.1312.6114> (2013).
64. Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J. & Chen, H. Application of generative autoencoder in de novo molecular design. *Mol. Inf.* **37**, 1700123 (2018).
65. Sattarov, B. et al. De novo molecular design by combining deep autoencoder recurrent neural networks with generative topographic mapping. *J. Chem. Inf. Model.* **59**, 1182–1196 (2019).
66. Samanta, B. et al. NEVAE: a deep generative model for molecular graphs. *J. Mach. Learn. Res.* <https://www.jmlr.org/papers/volume21/19-671/19-671.pdf> (2020).
67. Gui, J., Sun, Z., Wen, Y., Tao, D. & Ye, J. A review on generative adversarial networks: algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* <https://doi.org/10.1109/TKDE.2021.3130191> (2022).
68. Lin, E., Lin, C.-H. & Lane, H.-Y. Relevant applications of generative adversarial networks in drug design and discovery: molecular de novo design, dimensionality reduction, and de novo peptide and protein design. *Molecules* **25**, 3250 (2020).
69. Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouquié, D. & Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* **11**, 10 (2020).
70. Blanchard, A. E., Stanley, C. & Bhowmik, D. Using GANs with adaptive training data to search for new molecules. *J. Cheminform.* **13**, 14 (2021).
71. Ben-Baruch, E. et al. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **16**, 687–694 (2019).
72. Shin, B., Park, S., Kang, K. & Ho, J. C. Self-attention based molecule representation for predicting drug-target interaction. In *Proc. 4th Machine Learning for Healthcare Conference* **106**, 230–248 (2019).
73. Huang, K., Xiao, C., Glass, L. M. & Sun, J. MolTrans: molecular interaction transformer for drug-target interaction prediction. *Bioinformatics* **37**, 830–836 (2021).
74. **A molecular interaction transformer (MolTrans) was developed that uses knowledge-inspired sub-structural pattern mining to better extract substructure semantic relations from massive unlabelled biomedical data to improve prediction of ligand–target interactions.**
75. Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Sci. Rep.* **11**, 31619 (2021).
76. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. Preprint at <https://arxiv.org/abs/1903.02428> (2019).
77. Wang, M. et al. Deep Graph Library: a graph-centric, highly-performant package for graph neural networks. Preprint at <https://doi.org/10.48550/arXiv.1909.01315> (2019).
78. Duvenaud, D. et al. Convolutional networks on graphs for learning molecular fingerprints. *Adv. Neural Inf. Process. Syst.* <https://proceedings.neurips.cc/paper/2015/file/f9be311e65d81a9ad8150a60844bb94c-Paper.pdf> (2015).
79. Feinberg, E. N. et al. PotentialNet for molecular property prediction. *ACS Cent. Sci.* **4**, 1520–1530 (2018).
80. Klicpera, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. Preprint at <https://arxiv.org/abs/2003.03123> (2020).

78. Torng, W. & Altman, R. B. Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* **59**, 4131–4149 (2019).
79. Gao, K. Y. et al. Interpretable drug target prediction using deep neural representation. *Proc. 27th International Joint Conference on Artificial Intelligence* **2018**, 3371–3377 (2018).
80. Yang, M., Luo, H., Li, Y. & Wang, J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* **35**, i455–i463 (2019).
81. Gaudelet, T. et al. Utilizing graph machine learning within drug discovery and development. *Brief. Bioinform.* <https://doi.org/10.1093/bib/bbab159> (2021).
This is an informed review of the applications of GNNs and their variants in various components of drug discovery.
82. Olivecrona, M., Blaschke, T., Engkvist, O. & Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminform.* **9**, 48 (2017).
83. Putin, E. et al. Reinforced adversarial neural computer for de novo molecular design. *J. Chem. Inf. Model.* **58**, 1194–1204 (2018).
84. You, J., Liu, B., Ying, R., Pande, V. S. & Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. Preprint at <https://doi.org/10.48550/arXiv.1806.02473> (2018).
85. Zhou, Z., Kearnes, S., Li, L., Zare, R. N. & Riley, P. Optimization of molecules via deep reinforcement learning. *Sci. Rep.* **9**, 10752 (2019).
86. Grygorenko, O. O. et al. Generating multibillion chemical space of readily accessible screening compounds. *iScience* **23**, 101681 (2020).
87. Gorgulla, C. et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).
88. Acharya, A. et al. Supercomputer-based ensemble docking drug discovery pipeline with application to Covid-19. *J. Chem. Inf. Model.* <https://doi.org/10.1021/acs.jcim.0c01010> (2020).
89. McIntosh-Smith, S., Price, J., Sessions, R. B. & Ibarra, A. A. High performance in silico virtual drug screening on many-core processors. *Int. J. High Perform. Comput. Appl.* **29**, 119–134 (2015).
90. Pérez-Serrano, J., Imbernón, B., Cecilia, J. M. & Ujaldón, M. Energy-based tuning of metaheuristics for molecular docking on multi-GPUs. *Concurr. Comput.* **30**, e4684 (2018).
91. LeGrand, S. et al. GPU-accelerated drug discovery with docking on the summit supercomputer: porting, optimization, and application to COVID-19 research. In *Proc. 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* <https://doi.org/10.1145/3388440.3412472> (2020).
92. Darne, P. et al. Amide v2: high-throughput screening based on AutoDock-GPU and improved workflow leading to better performance and reliability. *Int. J. Mol. Sci.* **22**, 7489 (2021).
93. Gentile, F. et al. Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS Cent. Sci.* **6**, 939–949 (2020).
The authors propose a novel DL-inspired paradigm that helps accelerate docking, enabling them to screen larger libraries.
94. Ton, A.-T., Gentile, F., Hsing, M., Ban, F. & Cherkasov, A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol. Inf.* **39**, 2000028 (2020).
95. Gentile, F. et al. Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus deep docking of 40 billion small molecules. *Chem. Sci.* <https://doi.org/10.1039/D1SC05579H> (2021).
About 40 billion molecules were computationally screened against SARS-CoV-2 main protease, returning a large number of experimentally confirmed inhibitors using a fully automated end-to-end drug discovery protocol that integrates machine learning and human expertise.
96. Graff, D. E., Shakhnovich, E. I. & Coley, C. W. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* **12**, 7866–7881 (2021).
97. Yang, Y. et al. Efficient exploration of chemical space with docking and deep learning. *J. Chem. Theory Comput.* **17**, 7106–7119 (2021).
98. Hofmarcher, M. et al. Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3561442> (2020).
99. Ramsundar, B., Eastman, P., Walters, P. & Pande, V. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More* (O'Reilly Media, Inc., 2019).
This work attempts to democratize DL for life sciences and drug discovery by providing tools for representing data in DL-suitable formats for subsequent modelling.
100. Minnich, A. J. et al. AMPL: a data-driven modeling pipeline for drug discovery. *J. Chem. Inf. Model.* **60**, 1955–1968 (2020).
101. Korshunova, M., Ginsburg, B., Tropsha, A. & Isayev, O. OpenChem: a deep learning toolkit for computational chemistry and drug design. *J. Chem. Inf. Model.* **61**, 7–13 (2021).
102. Mercado, R. et al. Graph networks for molecular design. *Mach. Learn. Sci. Technol.* **2**, 25023 (2021).
103. Polykovskiy, D. et al. Molecular Sets (MOSES): a benchmarking platform for molecular generation models. *Front. Pharmacol.* **11**, 1931 (2020).
104. Cherkasov, A. et al. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **57**, 4977–5010 (2014).
105. Dixon, S. L. et al. Medicinal chemistry AutoQSAR: an automated machine learning tool for best-practice QSAR modeling. *Future Med. Chem.* **8**, 1825–1839 (2016).
106. Ekins, S. et al. Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **18**, 435–441 (2019).
107. Al Saadi, A. et al. IMPECCABLE: Integrated Modeling Pipeline for COVID Cure by Assessing Better LEads. In *50th International Conference on Parallel Processing* **20**, 1–12 (ACM, 2021).
108. NVIDIA Clara. <https://developer.nvidia.com/clara> (NVIDIA Corporation, 2021).
109. Baum, Z. J. et al. Artificial intelligence in chemistry: current trends and future directions. *J. Chem. Inf. Model.* <https://doi.org/10.1021/acs.jcim.1c00619> (2021).
110. Artrith, N. et al. Best practices in machine learning for chemistry. *Nat. Chem.* **13**, 505–508 (2021).
111. Feinberg, E. N., Joshi, E., Pande, V. S. & Cheng, A. C. Improvement in ADMET prediction with multitask deep featurization. *J. Med. Chem.* **63**, 8835–8848 (2020).
112. Klicpera, J., Weyshwanth, C. & Günnemann, S. Directional message passing on molecular graphs via synthetic coordinates. Preprint at <https://doi.org/10.48550/arXiv.2111.04718> (2021).
113. Wieder, O. et al. A compact review of molecular property prediction with graph neural networks. *Drug Disc. Today Technol.* <https://doi.org/10.1016/j.ddtec.2020.11.009> (2020).
114. Putin, E. et al. Adversarial threshold neural computer for molecular de novo design. *Mol. Pharm.* **15**, 4386–4397 (2018).
115. Samanta, B. et al. NeVAE: a deep generative model for molecular graphs. *Proc. AAAI Conf. Artif. Intell.* **33**, 1110–1117 (2019).
116. Asgari, E. & Mofrad, M. R. K. ProtVec: a continuous distributed representation of biological sequences. *PLoS ONE* **10**, 141287 (2015).
117. Imrie, F., Bradley, A. R., van der Schaar, M. & Deane, C. M. Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data. *J. Chem. Inf. Model.* **58**, 2319–2330 (2018).
118. Mayr, A., Klambauer, G., Unterthiner, T. & Hochreiter, S. DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* **3**, 80 (2016).
119. Ye, Z., Yang, Y., Li, X., Cao, D. & Ouyang, D. An integrated transfer learning and multitask learning approach for pharmacokinetic parameter prediction. *Mol. Pharm.* **16**, 533–541 (2018).
120. Ashtawy, H. M. & Mahapatra, N. R. Task-specific scoring functions for predicting ligand binding poses and affinity and for screening enrichment. *J. Chem. Inf. Model.* **58**, 119–133 (2017).
121. Chen, S., Xue, D., Chuai, G., Yang, Q. & Liu, Q. FL-QSAR: a federated learning-based QSAR prototype for collaborative drug discovery. *Bioinformatics* **36**, 5492–5498 (2021).
122. Xiong, Z. et al. Facing small and biased data dilemma in drug discovery with enhanced federated learning approaches. *Sci. China Life Sci.* <https://doi.org/10.1007/s11427-021-1946-0> (2021).
123. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).
124. Neil, D. et al. Exploring deep recurrent models with reinforcement learning for molecule design. Preprint at <https://openreview.net/forum?id=HkcTe-bR> (2018).
125. Ståhl, N., Falkman, G., Karlsson, A., Mathiason, G. & Boström, J. Deep reinforcement learning for multiparameter optimization in de novo drug design. *J. Chem. Inf. Model.* **59**, 3166–3176 (2019).
126. Liu, R., Wang, H., Glover, K. P., Feasel, M. G. & Wallqvist, A. Dissecting machine-learning prediction of molecular activity: is an applicability domain needed for quantitative structure-activity relationship models based on deep neural networks? *J. Chem. Inf. Model.* **59**, 117–126 (2018).
127. Schwaller, P. et al. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS Cent. Sci.* **5**, 1572–1583 (2019).
128. Zhang, Y. & Lee, A. A. Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning. *Chem. Sci.* **10**, 8154–8163 (2019).
129. Ryu, S., Lim, J., Hong, S. H. & Kim, W. Y. Deeply learning molecular structure-property relationships using attention- and gate-augmented graph convolutional network. Preprint at <https://doi.org/10.48550/arXiv.1805.10988> (2018).
130. Coley, C. W. et al. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **10**, 370–377 (2019).
131. Rajasekaran, S., Fiondella, L., Ahmed, M. & Ammar, R. A. (eds). *Multicore Computing: Algorithms, Architectures, and Applications* 1st edn (Chapman & Hall/CRC, 2013).
132. Li, M. et al. Scaling distributed machine learning with the parameter server. In *Proc. 11th USENIX Conference on Operating Systems Design and Implementation* 583–598 (USENIX Association, 2014).

Acknowledgements

This work was funded by the Canadian Institutes of Health Research (CIHR), Canadian 2019 Novel Coronavirus (2019-nCoV) Rapid Research grant numbers OV3-170631

and VR3-172639, and generous donations for COVID-19 research from TELUS, Teck Resources, the 625 Powell Street Foundation, the Tai Hung Fai Charitable Foundation and the Vancouver General Hospital Foundation. E.G. is supported by fellowships from the Canadian Institutes for Health Research (MFE-171324), Michael Smith Foundation for Health Research/VCHRI and VGH UBC Hospital Foundation (RT-2020-0408) and the Ermenegildo Zegna Foundation.

Competing interests

A.C.S. is employed by the NVIDIA corporation, a manufacturer of GPU technology. No other authors received funding from NVIDIA for this work.

Additional information

Correspondence should be addressed to Abraham C. Stern or Artem Cherkasov.

Peer review information *Nature Machine Intelligence* thanks Jeremy Smith, Leonardo Solis-Vasquez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2022