
ATOM3D: Tasks On Molecules in Three Dimensions

Raphael J. L. Townshend*
Computer Science
Stanford University

Martin Vögele†
Computer Science
Stanford University

Patricia Suriana†
Computer Science
Stanford University

Alexander Derry†
Biomedical Informatics
Stanford University

Alexander S. Powers
Chemistry
Stanford University

Yianni Laloudakis
Computer Science
Stanford University

Sidhika Balachandar
Computer Science
Stanford University

Bowen Jing
Computer Science
Stanford University

Brandon Anderson
Computer Science
University of Chicago

Stephan Eismann
Applied Physics
Stanford University

Risi Kondor
Computer Science, Statistics
University of Chicago

Russ B. Altman
Bioengineering, Genetics, Medicine
Stanford University

Ron O. Dror*
Computer Science
Stanford University

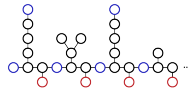

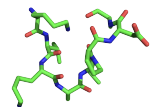

Abstract

Computational methods that operate on three-dimensional (3D) molecular structure have the potential to solve important problems in biology and chemistry. Deep neural networks have gained significant attention, but their widespread adoption in the biomolecular domain has been limited by a lack of either systematic performance benchmarks or a unified toolkit for interacting with 3D molecular data. To address this, we present ATOM3D, a collection of both novel and existing benchmark datasets spanning several key classes of biomolecules. We implement several types of 3D molecular learning methods for each of these tasks and show that they consistently improve performance relative to methods based on one- and two-dimensional representations. The choice of architecture proves to be important for performance, with 3D convolutional networks excelling at tasks involving complex geometries, graph networks performing well on systems requiring detailed positional information, and the more recently developed equivariant networks showing significant promise. Our results indicate that many molecular problems stand to gain from 3D molecular learning, and that there is potential for substantial further improvement on many tasks. To lower the barrier to entry and facilitate further developments in the field, we also provide a comprehensive suite of tools for dataset processing, model training, and evaluation in our open-source `atom3d` Python package. All datasets are available for download from www.atom3d.ai.

*Address correspondence to: raphael@cs.stanford.edu, rondror@cs.stanford.edu

†Equal contribution.

Table 1: Representation choice for molecules. Adding in 3D information consistently improves performance. The depicted 1D representations are the amino acid sequences and SMILES strings [Weininger, 1988] for proteins and small molecules, respectively.

Dimension	Representation	Examples	
		Proteins	Small Molecules
1D	linear sequence	KVKALPDA	CC(C)CC(C)NO
2D	chemical bond graph		
3D	atomistic geometry		

1 Introduction

A molecule’s three-dimensional (3D) shape is critical to understanding its physical mechanisms of action, and can be used to answer a number of questions relating to drug discovery, molecular design, and fundamental biology. While we can represent molecules using lower-dimensional representations such as linear sequences (1D) or chemical bond graphs (2D), considering the 3D positions of the component atoms—the atomistic geometry—allows for better modeling of 3D shape (Table 1). While previous benchmarking efforts, such as MoleculeNet [Wu et al., 2018] and TAPE [Rao et al., 2019], have examined diverse molecular tasks, they focus on these lower-dimensional representations. In this work, we demonstrate the benefit yielded by learning on 3D atomistic geometry and promote the development of 3D molecular learning by providing a collection of datasets leveraging this representation.

Furthermore, the atom is emerging as a “machine learning datatype” in its own right, deserving focused study much like the pixels that make up images in computer vision or the characters that make up text in natural language processing. All molecules, including proteins, DNA, RNA, and drugs, can be represented as atoms in 3D space. These atoms can only belong to a fixed class of element types (carbon, nitrogen, oxygen, etc.), and all molecules are governed by the same underlying laws of physics that impose rotational, translational, and permutational symmetries. These systems also contain higher-level patterns that are poorly characterized, creating a ripe opportunity for learning them from data: though certain basic components are well understood (e.g. amino acids, nucleotides, functional groups), many others can not easily be defined. These patterns are in turn composed in a hierarchy that itself is only partially elucidated.

While deep learning methods such as graph neural networks (GNNs) and convolutional neural networks (CNNs) seem well suited to atomistic geometry, to date there has been no systematic evaluation of such methods on molecular tasks. Additionally, despite the growing number of 3D structures available in databases such as the Protein Data Bank (PDB) [Berman et al., 2000], they require significant processing before they are useful for machine learning tasks. Inspired by the success of accessible databases such as ImageNet [Deng et al., 2009] and SQuAD [Rajpurkar et al., 2016] in sparking progress in their respective fields, we create and curate benchmark datasets for atomistic tasks, process them into a simple and standardized format, systematically benchmark 3D molecular learning methods, and present a set of best practices for other machine learning researchers interested in entering the field of 3D molecular learning (see Section C). We reveal a number of insights related to 3D molecular learning, including the relative strengths and weaknesses of different methods and the identification of several tasks that provide great opportunities for 3D molecular learning. These are all integrated into the `atom3d` Python package to lower the barrier to entry and facilitate reproducible research in 3D molecular learning for machine learning practitioners and structural biologists alike.

2 Related Work

While three dimensional molecular data have long been pursued as an attractive source of information in molecular learning and chemoinformatics [Swamidass et al., 2005, Azencott et al., 2007], their utility has become increasingly clear in the last couple years. Powered by increases in data availability and methodological advances, 3D molecular learning methods have demonstrated significant impact on tasks such as protein structure prediction [Senior et al., 2020, Jumper et al., 2021, Baek et al., 2021], equilibrium state sampling [Noé et al., 2019], and RNA structure prediction [Townshend et al., 2021]. At the same time, broader assessments of tasks involving molecular data have focused on either 1D or 2D representations [Wu et al., 2018, Rao et al., 2019]. Through ATOM3D, we aim to provide a first benchmark for learning on 3D molecular data. There are a few major classes of algorithms that exist for data in this form.

Graph neural networks (GNNs) have grown to be a major area of study, providing a natural way of learning from data with complex spatial structure. Many GNN implementations have been motivated by applications to atomic systems, including molecular fingerprinting [Duvinaud et al., 2015], property prediction [Schütt et al., 2017, Gilmer et al., 2017, Liu et al., 2019], protein interface prediction [Fout et al., 2017], and protein design [Ingraham et al., 2019]. Instead of encoding points in Euclidean space, GNNs encode their pairwise connectivity, capturing a structured representation of atomistic data. We note that some developed GNNs operate only on the chemical bond graph (i.e., 2D GNNs), with their edges representing covalent bonds, whereas others (including the ones we develop) operate on the 3D atomistic geometry (3D GNNs), with their edges representing distances between nearby pairs of atoms.

Three-dimensional CNNs (3DCNNs) have also become popular as a way to capture these complex 3D geometries. They have been applied to a number of biomolecular applications such as protein model quality assessment [Pagès et al., 2019, Derevyanko et al., 2018], protein sequence design [Anand et al., 2020], protein interface prediction [Townshend et al., 2019], and structure-based drug discovery [Wallach et al., 2015, Torng and Altman, 2017, Ragoza et al., 2017, Jiménez et al., 2018]. These 3DCNNs can encode translational and permutational symmetries, but incur significant computational expense and cannot capture rotational symmetries without data augmentation.

In an attempt to address many of the problems of representing atomistic geometries, equivariant neural networks (ENNs) have emerged as a new class of methods for learning from molecular systems. These networks are built such that geometric transformations of their inputs lead to well-defined transformations of their outputs. This setup leads to the neurons of the network learning rules that resemble physical interactions. Tensor field networks [Thomas et al., 2018] and Cormorant [Kondor, 2018, Anderson et al., 2019] have applied these principles to atomic systems and begun to demonstrate promise on larger systems such as proteins and RNA [Eismann et al., 2020, Weiler et al., 2018, Townshend et al., 2021].


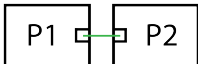

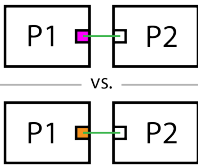

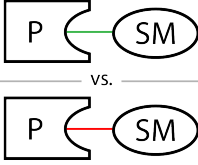


3 Datasets for 3D Molecular Learning

We select 3D molecular learning tasks from structural biophysics and medicinal chemistry that span a variety of molecule types. Several of these datasets are novel, while others are extracted from existing sources (Table 2). We note that these datasets are intended for benchmarking machine learning representations, and do not always correspond to problem settings that would be seen in real-world scenarios. Below, we give a short description of each dataset’s impact and source, as well as the metrics used to evaluate them and the splits. The splits were selected to minimize data leakage concerns and ensure generalizability and reproducibility. These datasets are all provided in a standardized format that requires no specialized libraries. Alongside these datasets, we present corresponding best practices (Appendix C) and further dataset-specific details (Appendix D). Taken together, we hope these efforts will lower the barrier to entry for machine learning researchers interested in developing methods for 3D molecular learning and encourage rapid progress in the field.

3.1 Small Molecule Properties (SMP)

Impact – Predicting physico-chemical properties of small molecules is a common task in medicinal chemistry and materials design. Quantum-chemical calculations can determine certain physico-chemical properties but are computationally expensive.

Table 2: Tasks included in the ATOM3D datasets, along with schematic representations of their inputs. P indicates protein, SM indicates small molecule, R indicates RNA. Lines indicate interaction and a small square within a protein indicates an individual amino acid. New datasets are in bold.

Name (Task Code)	Schematic	Objective	Source
Small Molecule Properties (SMP)		Properties	QM9 [Ruddigkeit et al., 2012]
Protein Interface Prediction (PIP)		Amino Acid Interaction	DIPS [Townshend et al., 2019] DB5 [Vreven et al., 2015]
Residue Identity (RES)		Amino Acid Identity	New, created from PDB [Berman et al., 2000]
Mutation Stability Prediction (MSP)		Effect of Mutation	New, created from SKEMPI [Jankauskaitė et al., 2019]
Ligand Binding Affinity (LBA)		Binding Strength	PDBBind [Wang et al., 2004]
Ligand Efficacy Prediction (LEP)		Ligand Efficacy	New, created from PDB [Berman et al., 2000]
Protein Structure Ranking (PSR)		Ranking	CASP-QA [Kryshtafovych et al., 2019]
RNA Structure Ranking (RSR)		Ranking	FARFAR2-Puzzles [Watkins et al., 2020]

Source – The QM9 dataset [Ruddigkeit et al., 2012, Ramakrishnan et al., 2014b] contains the results of quantum-chemical calculations for 134,000 stable small organic molecules, each made up C, O, N, F, and H and including no more than nine non-hydrogen atoms. For each molecule, the dataset contains the calculated geometry of the ground-state conformation as well as calculated energetic, electronic, and thermodynamic properties.

Targets – We predict the molecular properties from the ground-state structure.

Split – We split molecules randomly.

3.2 Protein Interface Prediction (PIP)

Impact – Proteins interact with each other in many scenarios—for example, antibody proteins recognize diseases by binding to antigens. A critical problem in understanding these interactions is to identify which amino acids of two given proteins will interact upon binding.

Source – For training, we use the Database of Interacting Protein Structures (DIPS), a comprehensive dataset of protein complexes mined from the PDB [Townshend et al., 2019]. We predict on the Docking Benchmark 5 [Vreven et al., 2015], a smaller gold standard dataset.

Targets – We predict whether two amino acids will contact when their respective proteins bind.

Split – We split protein complexes such that no protein in the training dataset has more than 30% sequence identity with any protein in the DIPS test set or the DB5 dataset.

3.3 Residue Identity (RES)

Impact – Understanding the structural role of individual amino acids is important for engineering new proteins. We can understand this role by predicting the propensity for different amino acids at a given protein site based on the surrounding structural environment [Torng and Altman, 2017].

Source – We generate a novel dataset consisting of local atomic environments centered around individual residues extracted from non-redundant structures in the PDB.

Targets – We formulate this as a classification task where we predict the identity of the amino acid in the center of the environment based on all other atoms.

Split – We split environments by protein topology class according to the CATH 4.2 [Dawson et al., 2017], such that all environments from proteins in the same class are in the same split dataset.

3.4 Mutation Stability Prediction (MSP)

Impact – Identifying mutations that stabilize a protein’s interactions is important to the design of new proteins. Experimental techniques for probing such mutations are labor-intensive [Antikainen and Martin, 2005, Lefèvre et al., 1997], motivating the development of efficient computational methods.

Source – We derive a novel dataset by collecting single-point mutations from the SKEMPI database [Jankauskaitė et al., 2019] and model each mutation into the structure to produce a mutated structure.

Targets – We formulate this as a binary classification task where we predict whether the stability of the complex increases as a result of the mutation.

Split – We split protein complexes such that no protein in the test dataset has more than 30% sequence identity with any protein in the training dataset.

3.5 Ligand Binding Affinity (LBA)

Impact – Predicting the strength (affinity) of a candidate drug molecule’s interaction with a target protein is a challenging but crucial task for drug discovery applications.

Source – We use the PDBBind database [Wang et al., 2004, Liu et al., 2015], a curated database containing protein-ligand complexes from the PDB and their corresponding binding strengths (affinities).

Targets – We predict $pK = -\log_{10}(K)$, where K is the binding affinity in Molar units.

Split – We split protein-ligand complexes such that no protein in the test dataset has more than 30% sequence identity with any protein in the training dataset.

3.6 Ligand Efficacy Prediction (LEP)

Impact – Many proteins switch on or off their function by changing shape. Predicting which shape a drug will favor is thus an important task in drug design.

Source – We develop a novel dataset by curating proteins from several families with both "active" and "inactive" state structures, and model in 527 small molecules with known activating or inactivating function using the program Glide [Friesner et al., 2004].

Targets – We formulate this as a binary classification task where we predict whether a molecule bound to the structures will be an activator of the protein’s function or not.

Split – We split complex pairs by protein target.

3.7 Protein Structure Ranking (PSR)

Impact – Proteins are one of the primary workhorses of the cell, and knowing their structure is often critical to understanding (and engineering) their function.

Source – We use the structural models submitted to the Critical Assessment of Structure Prediction (CASP) [Kryshtafovych et al., 2019], a blind protein structure prediction competition, over the last 18 years.

Targets – We formulate this as a regression task, where we predict the global distance test (GDT_TS) of each structural model from the experimentally determined structure.

Split – We split structures temporally by competition year.

3.8 RNA Structure Ranking (RSR)

Impact – Similar to proteins, RNA plays major functional roles (e.g., gene regulation) and can adopt well-defined 3D shapes. Yet the problem is data-poor, with only a few hundred known structures.

Source – We use the FARFAR2-Puzzles dataset, which consists of structural models generated by FARFAR2 [Watkins et al., 2020] for 20 RNAs from RNA Puzzles, a blind structure prediction competition for RNA [Cruz et al., 2012].

Targets – We predict the root-mean-squared deviation (RMSD) of each structural model from the experimentally determined structure.

Split – We split structures temporally by competition year.

4 Benchmarking Setup

To assess the benefits of 3D molecular learning, we use a combination of existing and novel 3D molecular learning methods, and implement a number of robust baselines. Our 3D molecular learning methods belong to one of each of the major classes of deep learning algorithms that have been applied to atomistic systems: graph networks, three-dimensional convolutional networks, and equivariant networks. Here we describe the main principles of the core networks used in these models. See Appendix E for task-specific details and hyperparameters.

For GNNs, we represent molecular systems as graphs in which each node is an atom. Edges are defined between all atoms separated by less than 4.5 Å, and weighted by the distance between the atoms using an edge weight defined by $w_{i,j} = \frac{1}{d_{i,j} + \epsilon}$, where $\epsilon = 10^{-5}$ is a small factor added for numerical stability. Node features are one-hot-encoded by atom type. Our core model uses five layers of graph convolutions as defined by Kipf and Welling [2016], each followed by batch normalization and ReLU activation, and finally two fully-connected layers with dropout. For tasks requiring a single output for the entire molecular system, we use global pooling to aggregate over nodes. For tasks requiring predictions for single atoms or amino acids, we extract the relevant node embeddings from each graph after all convolutional layers (see Appendix E).

For 3DCNNs, we represent our data as a cube of fixed size (different per task due to the different molecular sizes) in 3D space that is discretized into voxels with resolution of 1 Å to form a grid (for PSR and RSR, we decrease the grid resolution to 1.3 Å in order to fit in the GPU memory). Each voxel is associated with a one-hot-encoded vector that denotes the presence or absence of each atom type. Our core model consists of four 3D-convolutional layers, each followed by ReLU activation and max-pooling (for every other convolution layer). Two fully connected layers are applied after the convolutional layers to produce the final prediction.

For ENNs, we use SE(3)-equivariant networks that represent each atom of a structure by its coordinates in 3D space and by a one-hot encoding of its atom type. No rotational augmentation is needed due to the rotational symmetry of the network. The core of all architectures in this work is Cormorant, a network of four layers of covariant neurons that use the Clebsch–Gordan transform as nonlinearity, as described and implemented by Anderson et al. [2019].

5 Benchmarking Results

To assess the utility of 3D molecular learning, we evaluate our methods on the ATOM3D datasets and compare performance to state-of-the-art methods using 1D or 2D representations (for a comparison to the overall state-of-the-art, see Table 7). We note that in many cases, 3D molecular learning methods have not been applied to the proposed tasks, and that several of the tasks are novel. In the following sections, we describe the results of our benchmarking and some key insights that can be derived from them. We also aggregate these results along with additional metrics and standard deviations over three replicates in Table 8. For each metric, we bold the best-performing method as well as those within one standard deviation of the best-performing method.

5.1 3D representations consistently improve performance

Our evaluation of 3D methods on the tasks in ATOM3D reveals that incorporating atomistic geometry leads to consistently superior performance compared to 1D and 2D methods. For small molecules,

Table 3: Small molecule results. Metric is mean absolute error (MAE).

Task	Target	3D			Non-3D	
		3DCNN	GNN	ENN	[Tsubaki et al., 2019]	[Liu et al., 2019]
SMP	μ [D]	0.754	0.501	0.052	0.496	0.520
	ε_{gap} [eV]	0.580	0.137	0.095	0.154	0.184
	U_0^{at} [eV]	3.862	1.424	0.025	0.182	0.218

Table 4: Biopolymer results. AUROC is the area under the receiver operating characteristic curve. Asterisks (*) indicate that the exact training data differed (though splitting criteria were the same).

Task	Metric	3D			Non-3D	
		3DCNN	GNN	ENN	[Sanchez-Garcia et al., 2018]	[Rao et al., 2019]
PIP	AUROC	0.844	*0.669	—	0.841	
RES	accuracy	0.451	0.082	*0.072	*0.30	
MSP	AUROC	0.574	0.609	0.574	0.554	

state-of-the-art methods do not use 1D representations, so we focus instead on comparing to representations at the 2D level, i.e. the chemical bond graph. This is the approach taken by the 2D GNN introduced by [Tsubaki et al., 2019] and the N-gram graph method by [Liu et al., 2019], which both obtain similar results (Table 3) on the small-molecule-only dataset SMP. When we add 3D coordinate information as in our ENN implementation, performance improves across all targets in SMP.

For tasks involving biopolymers (proteins and RNA), state-of-the-art methods do not use 2D representations, primarily because most of the chemical bond graph can be re-derived from the 1D representation, i.e. the linear sequence that makes up the biopolymer. We thus compare to representations at the 1D level (Table 4). For MSP and RES, both new datasets, we evaluate against the TAPE model [Rao et al., 2019], a transformer architecture that operates on protein sequence and is state-of-the-art amongst 1D methods for many tasks. For PIP, we compare to the sequence-only version of BIPSPI [Sanchez-Garcia et al., 2018], a state-of-the-art boosted decision tree method for protein interaction prediction. We find that 3D methods outperform these 1D methods on all biopolymer-only datasets (PIP, RES, MSP).

For tasks involving both biopolymers and small molecules, we compare to DeepDTA [Öztürk et al., 2018]. This network uses a 1D representation via a 1DCNN for both the biopolymer and small molecules. For LBA, we additionally compare to DeepAffinity [Karimi et al., 2019] which uses pairs of ligand SMILES strings and structurally annotated protein sequences. Using a 3D representation for both the ligand and protein leads to improved or comparable performance for the joint protein–small molecule datasets (LBA and LEP, see Table 5).

The biopolymer structure ranking tasks (PSR and RSR) are inherently 3D in nature, as they involve evaluating the correctness of different 3D shapes taken on by the same biopolymer. Thus, critically,

Table 5: Joint small molecule/biopolymer results. R_S is Spearman correlation, R_P is Pearson correlation, AUROC is area under the receiver operating characteristic curve, and RMSE is root-mean-squared error.

Task	Metric	3D			Non-3D	
		3DCNN	GNN	ENN	[Öztürk et al., 2018]	[Karimi et al., 2019]
LBA	RMSE	1.416	1.601	1.568	1.565	1.893
	glob. R_P	0.550	0.545	0.389	0.573	0.415
	glob. R_S	0.553	0.533	0.408	0.574	0.426
LEP	AUROC	0.589	0.681	0.663	0.696	—

Table 6: Structure ranking results. R_S is Spearman correlation. Mean measures the correlation for structures corresponding to the same biopolymer, whereas global measures the correlation across all biopolymers.

Task	Metric	3D		
		3DCNN	GNN	SotA
PSR	mean R_S	0.431	0.411	0.432 [Pagès et al., 2019]
	glob. R_S	0.789	0.750	0.796 [Pagès et al., 2019]
RSR	mean R_S	0.264	0.234	0.173 [Watkins et al., 2020]
	glob. R_S	0.372	0.512	0.304 [Watkins et al., 2020]

a 1D or 2D representation would not be able to differentiate between these different shapes since the linear sequence and chemical bond graph would remain the same. We therefore compare to state-of-the-art 3D methods as shown in Table 6, finding competitive or better results.

More generally, we find that learning methods that leverage the 3D geometry of molecules hold state-of-the-art on the majority of tasks on our benchmark (Table 7). These results demonstrate the potential of 3D molecular learning to address a wide range of problems involving molecular structure, and we anticipate that continued development of such models will aid progress in biological and chemical research.

5.2 Different tasks benefit from different architectures

While 3D molecular learning methods outperform their non-3D counterparts and provide a systematic way of representing molecular data, our results also provide evidence that architecture selection plays an important role in performance.

For tasks focused on biopolymers and with large amounts of training data (PIP and RES, Figure 1) we observe that 3DCNNs generally outperform standard GNNs. We hypothesize this is due to the ability of 3DCNNs to learn many-body patterns within a single filter, as opposed to GNNs that operate on one-body (node) and two-body (edge) features. Such many-body patterns are especially present in biopolymers, which generally adopt complex 3D geometries. This *many-body representation* hypothesis implies that 3DCNNs have specific advantages in terms of representational power.

However, as the size of the datasets decrease (Table 9), we see more even performance when comparing 3DCNNs and GNNs. In particular, performance is quite similar on the intermediate-sized PSR and RSR datasets, and GNNs nearly fully supplant 3DCNNs on the small-sized MSP and LEP datasets. On these datasets, ENNs also equal or exceed 3DCNN performance. This is in line with the many-body representation hypothesis, as the increased representational power of 3DCNNs becomes less important in these data-poor regimes.

A notable exception to this trend is the large SMP dataset, where we see improved performance from GNNs and ENNs. We note, however, that this is the sole dataset involving only small molecules. These molecules generally do not contain as complex of 3D geometries as biopolymers, and therefore do not contain large numbers of many-body patterns. In addition, many of the prediction targets depend instead on the exact positions of each atom relative to its neighbors. While particle-based methods such as GNNs and ENNs can precisely record these distances, volumetric 3DCNNs must instead approximate these positions. While increasing spatial resolution increases precision, it also leads to cubic scaling of complexity that prevents the same level of precision.

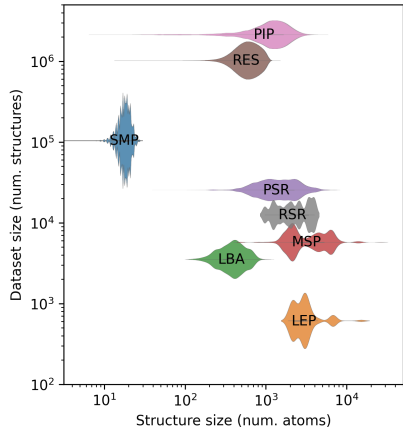


Figure 1: Datasets plotted by their size, as well as the number of atoms in each one of their molecules.

Finally, equivariant networks show significant promise, despite being a recent innovation. One motivation for their use is that they fill a “happy medium” where they both represent atom positions precisely and capture the many-body patterns present in complex geometries. On SMP, the only dataset on which we tested ENNs without any limitations, we observed state-of-the-art performance. For other tasks, the performance of the ENN implementation we used limited us to training on a fraction of the data ($< 1\%$ for RES) or on a portion of the entire atomic structure (LBA, LEP, MSP), or did not permit us to apply it at all (PIP, PSR, RSR). Faster implementations are now available to allow scaling of ENNs to larger systems [Geiger et al., 2020, Kondor and Thiede, 2021].

6 Conclusion

In this work we present a vision of the atom as a new “machine learning datatype” deserving focused study, as 3D molecular learning has the potential to address many unsolved problems in biology and chemistry. In particular, systems of atoms are well-suited to machine learning as they contain several underlying symmetries as well as poorly understood higher-level patterns. With ATOM3D, we take a first step towards this vision by providing a comprehensive suite of benchmark datasets and computational tools for building machine learning models for 3D molecular data.

We provide several benchmark datasets and compare the performance of different types of 3D molecular learning models across these tasks. We demonstrate that, for nearly all tasks that can be formulated in lower dimensions, 3D molecular learning yields gains in performance over 1D and 2D methods. We also show that selection of an appropriate architecture is critical for optimal performance on a given task; depending on the structure of the underlying data, a 3DCNN, GNN, or ENN may be most appropriate, especially in light of our many-body representation hypothesis. Equivariant networks in particular are continuing to improve in efficiency and stability, and we expect these to prove effective due to their ability to concisely model physical laws.

While this work demonstrates the potential of 3D structures and provides an initial set of benchmarks, there are some limitations to consider when using these resources. First, the datasets and tasks represented in ATOM3D are inherently biased towards biomolecules with solved structures. Certain classes of molecules (e.g. intrinsically disordered or transmembrane proteins) may therefore be underrepresented or absent, and the performance on these benchmarks will not necessarily generalize to such structures. Second, the benchmark models we report here are designed to be competitive but simple baselines. A bespoke architecture designed specifically for a certain task or molecule class and with comprehensive hyperparameter tuning is expected to outperform many of these baselines, and we encourage the exploration of novel and innovative approaches even within model classes that appear to underperform in these benchmarks (e.g. GNNs for the PIP task).

Third, several of the benchmark tasks are formulated differently from those that biologists, chemists, and drug designers typically wish to solve. For example, when predicting the binding affinity of a ligand, one would rarely have access to a 3D structure of the ligand bound to the target (as in the LBA benchmark), because determining this structure experimentally would typically be far more expensive and time-consuming than measuring the ligand binding affinity. Likewise, when using machine learning methods to predict small-molecule properties more efficiently than through quantum chemical calculations (as in the SMP benchmark), one would not typically have access to the ground-state structure, because determining that structure requires equally expensive quantum chemical calculations. Although formulated in a somewhat artificial manner for convenience, such benchmarks have proven useful in evaluating general machine learning representations and methods.

Finally, in addition to the datasets described here, there are many other open areas in biomedical research and molecular science that are ripe for 3D molecular learning, especially as structural data becomes readily available. Such tasks include virtual screening and pose prediction of small molecule drug candidates, as well as the incorporation of conformational ensembles instead of static structures in order to represent more faithfully the entire set of structures a molecule could adopt. Building on our easily extensible framework, we anticipate the addition of new datasets and tasks from across the research community.

Through this work, we hope to lower the entry barrier for machine learning practitioners, encourage the development of algorithms focused on 3D atomistic data, and promote an emerging paradigm within the fields of structural biology and medicinal chemistry.

7 Acknowledgments

We thank Truong-Son Hy, Maria Karelina, David Liu, Lgia Melo, Joseph Paggi, and Erik Thiede for discussions and advice. We also thank Aditi Krishnapriyan and Nicolas Swenson for pointing out an error in earlier GNN performance numbers. This work was supported by the U.S. Department of Energy (DOE), Office of Science, Graduate Student Research (SCGSR) program (RJLT); EMBO Long-Term Fellowship ALTF 235-2019 (MV); an NSF Graduate Research Fellowship (PS); National Library of Medicine training grant LM012409 (AD); a Stanford Bio-X Bowes Fellowship (SE); NIH grants GM102365 and HG010615 (RBA); the Chan Zuckerberg Biohub (RBA); the DOE, Office of Science, Scientific Discovery through Advanced Computing (SciDAC) program (ROD); Intel (ROD); and DARPA Agreement No. HR0011-18-9-0038 (RK). Most of the computing for this project was performed on the Sherlock cluster. We thank Stanford University and the Stanford Research Computing Facility for providing computational resources and support that contributed to these research results.

References

- Rebecca F. Alford, Andrew Leaver-Fay, Jeliuzko R. Jeliuzkov, Matthew J. O’Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J. Gray. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 2017. ISSN 15499626. doi: 10.1021/acs.jctc.7b00125.
- S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–10, 1990. ISSN 0022-2836. doi: 10.1016/S0022-2836(05)80360-2. URL <http://www.sciencedirect.com/science/article/pii/S0022283605803602>.
- Namrata Anand, Raphael R Eguchi, Alexander Derry, Russ B Altman, and Po-Ssu Huang. Protein sequence design with a learned potential. January 2020.
- Brandon Anderson, Truong-Son Hy, and Risi Kondor. Cormorant: Covariant Molecular Neural Networks. 2019. URL <http://arxiv.org/abs/1906.04015>.
- Antonina Andreeva, Dave Howorth, Cyrus Chothia, Eugene Kulesha, and Alexey G. Murzin. SCOP2 prototype: A new approach to protein structure mining. *Nucleic Acids Research*, 42(D1):310–314, 2014. ISSN 03051048. doi: 10.1093/nar/gkt1242.
- Nina M Antikainen and Stephen F Martin. Altering protein specificity: techniques and applications. *Bioorganic & medicinal chemistry*, 13(8):2701–2716, 2005.
- Chloé-Agathe Azencott, Alexandre Ksikes, S Joshua Swamidass, Jonathan H Chen, Liva Ralaivola, and Pierre Baldi. One-to four-dimensional kernels for virtual screening and the prediction of physical, chemical, and biological properties. *Journal of chemical information and modeling*, 47(3):965–974, 2007.
- Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. Van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. ISSN 10959203. doi: 10.1126/science.abj8754.
- Pedro J Ballester and John B O Mitchell. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9):1169–1175, May 2010.

- Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1): 235–242, 2000.
- Wouter Boomsma and Jes Frellsen. Spherical convolutions and their application in molecular modelling. In I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3433–3443. Curran Associates, Inc., 2017.
- Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J Gray. Pyrosetta: a script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics*, 26(5):689–691, 2010.
- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939785. URL <https://doi.org/10.1145/2939672.2939785>.
- Jianlin Cheng, Arlo Randall, Michael Sweredoski, and Pierre Baldi. Scratch: a protein structure and structural feature prediction server. *Nucleic acids research*, 33:W72–6, 08 2005. doi: 10.1093/nar/gki396.
- José Almeida Cruz, Marc-Frédéric Blanchet, Michal Boniecki, Janusz M Bujnicki, Shi-Jie Chen, Song Cao, Rhiju Das, Feng Ding, Nikolay V Dokholyan, Samuel Coulbourn Flores, et al. Rna-puzzles: a casp-like evaluation of rna three-dimensional structure prediction. *Rna*, 18(4):610–625, 2012.
- Natalie L Dawson, Tony E Lewis, Sayoni Das, Jonathan G Lees, David Lee, Paul Ashford, Christine A Orenge, and Ian Sillitoe. Cath: an expanded resource to predict protein function through structure and sequence. *Nucleic acids research*, 45(D1):D289–D295, 2017.
- Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. pages 248–255, 2009. doi: 10.1109/cvprw.2009.5206848.
- Georgy Derevyanko, Sergei Grudin, Yoshua Bengio, and Guillaume Lamoureux. Deep convolutional networks for quality assessment of protein folds. *Bioinformatics*, 34(23):4046–4053, 2018.
- David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in neural information processing systems*, pages 2224–2232, 2015.
- Stephan Eismann, Raphael J L Townshend, Nathaniel Thomas, Milind Jagota, Bowen Jing, and Ron Dror. Hierarchical, rotation-equivariant neural networks to predict the structure of protein complexes. *arXiv*, 2020.
- Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. Protein interface prediction using graph convolutional networks. In I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6530–6539. Curran Associates, Inc., 2017.
- Richard A Friesner, Jay L Banks, Robert B Murphy, Thomas A Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.
- Joffrey Gabel, Jérémy Desaphy, and Didier Rognan. Beware of machine learning-based scoring functions-on the danger of developing black boxes. *J. Chem. Inf. Model.*, 54(10):2807–2815, October 2014.

- Mario Geiger, Tess Smidt, Alby M., Benjamin Kurt Miller, Wouter Boomsma, Bradley Dice, Kostiantyn Lapchevskiy, Maurice Weiler, Michał Tyszkiewicz, Simon Batzner, Martin Uhrin, Jes Frellsen, Nuri Jung, Sophia Sanborn, Josh Rackers, and Michael Bailey. Euclidean neural networks: e3nn, 2020. URL <https://doi.org/10.5281/zenodo.5292912>.
- Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry, 2017.
- Jie Hou, Renzhi Cao, and Jianlin Cheng. Deep convolutional neural networks for predicting the quality of single protein structural models. *bioRxiv*, 2019. doi: 10.1101/590620. URL <https://www.biorxiv.org/content/early/2019/03/28/590620>.
- John Ingraham, Vikas K Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for Graph-Based protein design. March 2019.
- Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H Moal. Skempi 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019.
- José Jiménez, Miha Škalič, Gerard Martínez-Rosell, and Gianni De Fabritiis. KDEEP: Protein-Ligand absolute binding affinity prediction via 3D-Convolutional neural networks. *J. Chem. Inf. Model.*, 58(2):287–296, February 2018.
- Robbie Joosten, Tim Beek, Elmar Krieger, Maarten Hekkelman, Rob Hooft, Reinhard Schneider, Chris Sander, and Gert Vriend. A series of pdb related databases for everyday needs. *Nucleic acids research*, 39:D411–9, 11 2010. doi: 10.1093/nar/gkq1105.
- John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A.A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. ISSN 14764687. doi: 10.1038/s41586-021-03819-2. URL <http://dx.doi.org/10.1038/s41586-021-03819-2>.
- Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983. ISSN 1097-0282. doi: 10.1002/bip.360221211. URL <http://dx.doi.org/10.1002/bip.360221211>.
- Mostafa Karimi, Di Wu, Zhangyang Wang, and Yang Shen. Deepaffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, 35(18):3329–3338, 2019.
- Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- Risi Kondor. N-body networks: a covariant hierarchical neural network architecture for learning atomic potentials, 2018.
- Risi Kondor and Erik Henning Thiede. GELib — C++/CUDA library for rotation group operations, 2021. URL <https://github.com/risi-kondor/GELib>.
- Christian Kramer and Peter Gedeck. Leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets. *J. Chem. Inf. Model.*, 50(11):1961–1969, November 2010.
- G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack. Improved prediction of protein side-chain conformations with scwrl4. *Proteins*, 2009. URL <http://dunbrack.fccc.edu/SCWRL3.php/>.
- Andriy Kryshtafovych, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moult. Critical assessment of methods of protein structure prediction (casp)—round xiii. *Proteins: Structure, Function, and Bioinformatics*, 87(12):1011–1020, 2019.

- Andrew Leaver-Fay, Michael Tyka, Steven Lewis, Oliver Lange, James Thompson, Ron Jacak, Kristian Kaufman, P. Renfrew, Colin Smith, Will Sheffler, Ian Davis, Seth Cooper, Adrien Treuille, Daniel Mandell, Florian Richter, Yih-En Ban, Sarel Fleishman, Jacob Corn, David Kim, and Philip Bradley. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology*, 487:545–74, 01 2011.
- Fabrice Lefèvre, Marie-Hélène Rémy, and Jean-Michel Masson. Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function. *Nucleic acids research*, 25(2):447–448, 1997.
- Yan Li, Zhihai Liu, Jie Li, Li Han, Jie Liu, Zhixiong Zhao, and Renxiao Wang. Comparative assessment of scoring functions on an updated benchmark: 1. compilation of the test set. *J. Chem. Inf. Model.*, 54(6):1700–1716, June 2014.
- Yang Li and Jianyi Yang. Structural and sequence similarity makes a significant impact on machine-learning-based scoring functions for protein–ligand interactions. *Journal of Chemical Information and Modeling*, 57(4):1007–1012, 2017.
- Shengchao Liu, Mehmet F Demirel, and Yingyu Liang. N-gram graph: Simple unsupervised representation for graphs, with applications to molecules. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8464–8476. Curran Associates, Inc., 2019.
- Zhihai Liu, Yan Li, Li Han, Jie Li, Jie Liu, Zhixiong Zhao, Wei Nie, Yuchen Liu, and Renxiao Wang. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics*, 31(3):405–412, February 2015.
- Christophe Magnan and Pierre Baldi. Sspro/acpro 5: Almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning, and structural similarity. *Bioinformatics (Oxford, England)*, 30:2592–97, 05 2014. doi: 10.1093/bioinformatics/btu352.
- Simon Mitternacht. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research*, 5(189), 2016. ISSN 1759796X. doi: 10.12688/f1000research.7931.1.
- Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457), 2019. ISSN 0036-8075. doi: 10.1126/science.aaw1147. URL <https://science.sciencemag.org/content/365/6457/eaaw1147>.
- Hakime Öztürk, Elif Ozkirimli, and Arzucan Özgür. DeepDTA: Deep Drug-Target Binding Affinity Prediction. pages 1–17, 2018. URL <http://arxiv.org/abs/1801.10193>.
- Guillaume Pagès, Benoit Charmettant, and Sergei Grudin. Protein model quality assessment using 3d oriented convolutional neural networks. *Bioinformatics*, 35(18):3313–3319, 2019.
- Matthew Ragoza, Joshua Hochuli, Elisa Idrobo, Jocelyn Sunseri, and David Ryan Koes. Protein-Ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.*, 57(4):942–957, April 2017.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, (ii):2383–2392, 2016.
- Raghunathan Ramakrishnan, Pavlo Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules, Jul 2014a. URL https://springernature.figshare.com/collections/Quantum_chemistry_structures_and_properties_of_134_kilo_molecules/978904/5.
- Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1, 2014b. ISSN 20524463. doi: 10.1038/sdata.2014.22.

- Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yan Duan, Peter Chen, John Canny, Pieter Abbeel, and Yun Song. Evaluating protein transfer learning with tape. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 9689–9701. Curran Associates, Inc., 2019. URL <http://papers.nips.cc/paper/9163-evaluating-protein-transfer-learning-with-tape.pdf>.
- Lars Ruddigkeit, Ruud van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *Journal of Chemical Information and Modeling*, 52(11):2864–2875, nov 2012. ISSN 1549-9596. doi: 10.1021/ci300415d. URL <https://doi.org/10.1021/ci300415d>.
- Ruben Sanchez-Garcia, C O S Sorzano, J M Carazo, and Joan Segura. BIPSPi: a method for the prediction of partner-specific protein–protein interfaces. *Bioinformatics*, 35(14):343–353, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty647. URL <https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/bty647/5055586>.
- Soumya Sanyal, Ivan Anishchenko, Anirudh Dagar, David Baker, and Partha Talukdar. Proteingcn: Protein model quality assessment using graph convolutional networks. *bioRxiv*, 2020. doi: 10.1101/2020.04.06.028266. URL <https://www.biorxiv.org/content/early/2020/04/07/2020.04.06.028266>.
- Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.
- Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, 2017.
- Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Židek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 2020.
- Minyi Su, Qifan Yang, Yu Du, Guoqin Feng, Zhihai Liu, Yan Li, and Renxiao Wang. Comparative assessment of scoring functions: The CASF-2016 update. *J. Chem. Inf. Model.*, 59(2):895–913, February 2019.
- S Joshua Swamidass, Jonathan Chen, Jocelyne Bruand, Peter Phung, Liva Ralaivola, and Pierre Baldi. Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics*, 21:i359–i368, 2005.
- Nathaniel Thomas, Tess Smidt, Steven M. Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *CoRR*, abs/1802.08219, 2018. URL <http://arxiv.org/abs/1802.08219>.
- Wen Torng and Russ B Altman. 3D deep convolutional neural networks for amino acid environment similarity analysis. *BMC Bioinformatics*, 18(1):302, June 2017.
- Raphael Townshend, Rishi Bedi, Patricia Suriana, and Ron Dror. End-to-end learning on 3d protein structure for interface prediction. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 15642–15651. 2019.
- Raphael J L Townshend, Stephan Eismann, Andrew M Watkins, Ramya Rangan, Maria Karelina, Rhiju Das, and Ron O Dror. Geometric Deep Learning of RNA Structure. *Science*, 2021.
- Masashi Tsubaki, Kentaro Tomii, and Jun Sese. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, 35(2):309–318, 2019. ISSN 14602059. doi: 10.1093/bioinformatics/bty535.
- Karolis Uziela, David Menéndez Hurtado, Nanjiang Shu, Björn Wallner, and Arne Elofsson. ProQ3d: improved model quality assessments using deep learning. *Bioinformatics*, page btw819, jan 2017. doi: 10.1093/bioinformatics/btw819. URL <https://doi.org/10.1093/bioinformatics/btw819>.

- Thom Vreven, Iain H. Moal, Anna Vangone, Brian G. Pierce, Panagiotis L. Kastiris, Mieczyslaw Torchala, Raphael Chaleil, Brian Jiménez-García, Paul A. Bates, Juan Fernandez-Recio, Alexandre M.J.J. Bonvin, and Zhiping Weng. Updates to the integrated protein–protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2. *Journal of Molecular Biology*, 427(19):3031–3041, 2015. ISSN 0022-2836. doi: <https://doi.org/10.1016/j.jmb.2015.07.016>. URL <http://www.sciencedirect.com/science/article/pii/S0022283615004180>.
- Izhar Wallach, Michael Dzamba, and Abraham Heifets. Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery, 2015.
- Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The pdbname database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *Journal of Medicinal Chemistry*, 47(12):2977–2980, 2004. doi: 10.1021/jm030580l. URL <https://doi.org/10.1021/jm030580l>. PMID: 15163179.
- Andrew Martin Watkins, Ramya Rangan, and Rhiju Das. FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Structure*, 28(8):963–976.e6, aug 2020. ISSN 09692126. doi: 10.1016/j.str.2020.05.011. URL <https://www.biorxiv.org/content/10.1101/764449v2https://linkinghub.elsevier.com/retrieve/pii/S0969212620301805>.
- Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. July 2018.
- David Weininger. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988. ISSN 00952338. doi: 10.1021/ci00057a005.
- Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.*, 9:513–530, 2018. doi: 10.1039/C7SC02664A. URL <http://dx.doi.org/10.1039/C7SC02664A>.
- Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57:702–10, 09 2007. doi: 10.1002/prot.20264. URL <https://zhanglab.ccmb.med.umich.edu/TM-score/>.
- David Zilian and Christoph A Sotriffer. SFCscore(RF): a random forest-based scoring function for improved affinity prediction of protein-ligand complexes. *J. Chem. Inf. Model.*, 53(8):1923–1933, August 2013.