

# Scoreformer: A Surrogate Model For Large-Scale Prediction of Docking Scores

Álvaro Ciudad, Adrián Morales-Pastor, Laura Malo, Isaac Filella-Mercè, Victor Guallar, Alexis Molina



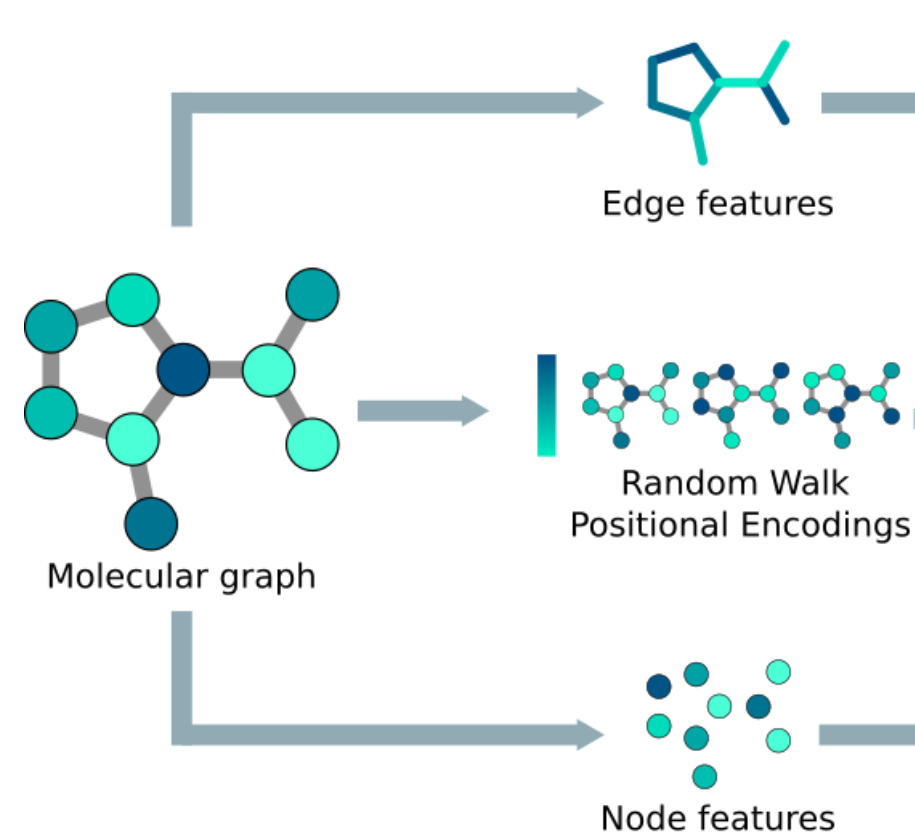
**TLDR:** The current size of molecular databases demands faster methods to screen them. The use of **PNA** and **LRWPE** in a graph transformer allows us to capture molecular representations while **increasing inference speed**.

## Introduction

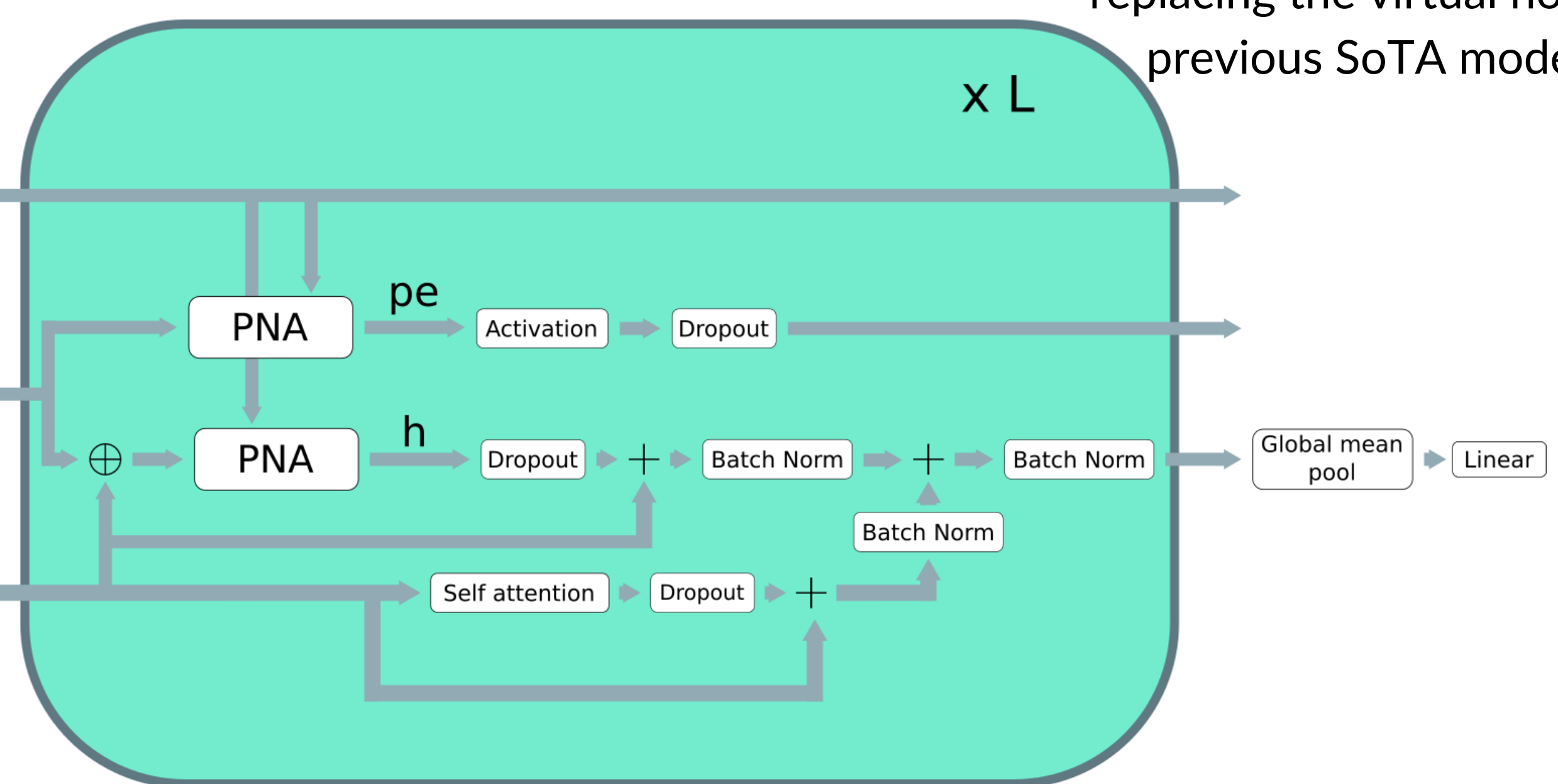
Molecular databases are growing fast due to the advances in combinatorial chemistry. This increases the need for faster tools to efficiently explore such databases to find promising drug candidates. Graph Neural Networks represent a great alternative to classical docking engines due to their ability to learn molecular patterns and their inference speed.

## Methods

PNA message passing layers allow **more expressive integration** of information from neighboring nodes

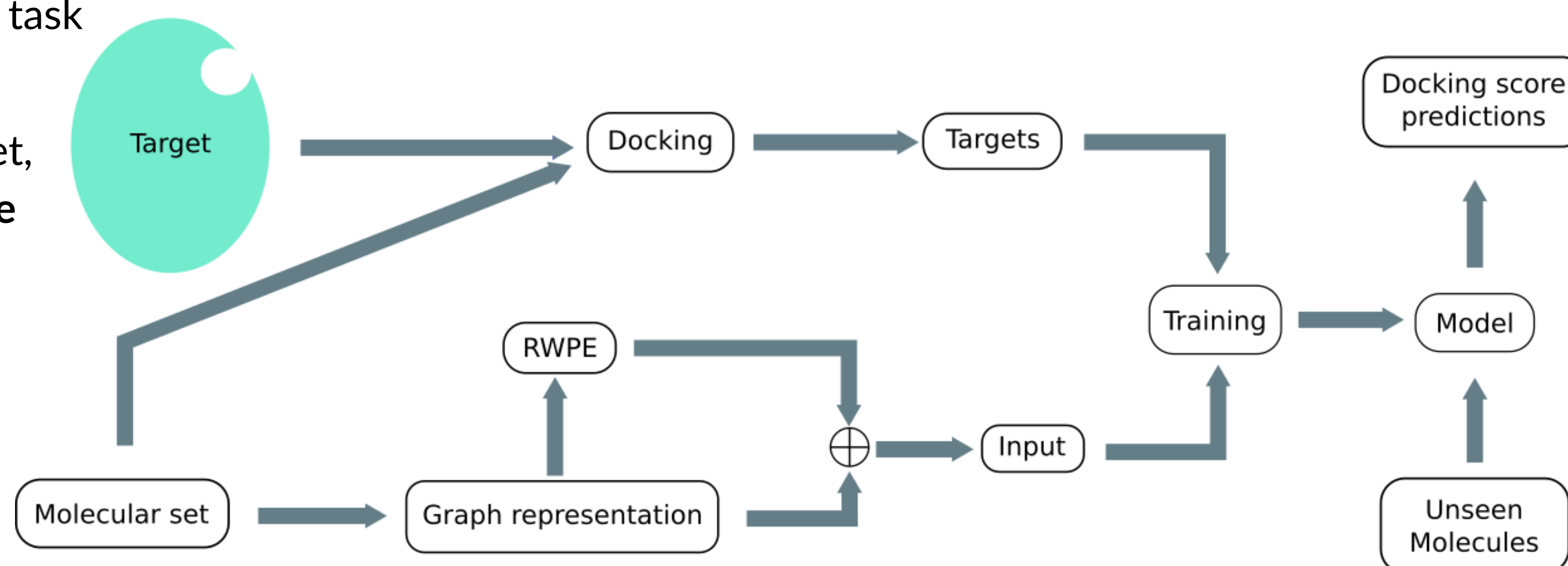


Graph transformers account for long-range interactions, replacing the virtual node in previous SoTA models



LRWPE allows to the model **learn a local positional encoding** for each node that optimizes the regression task

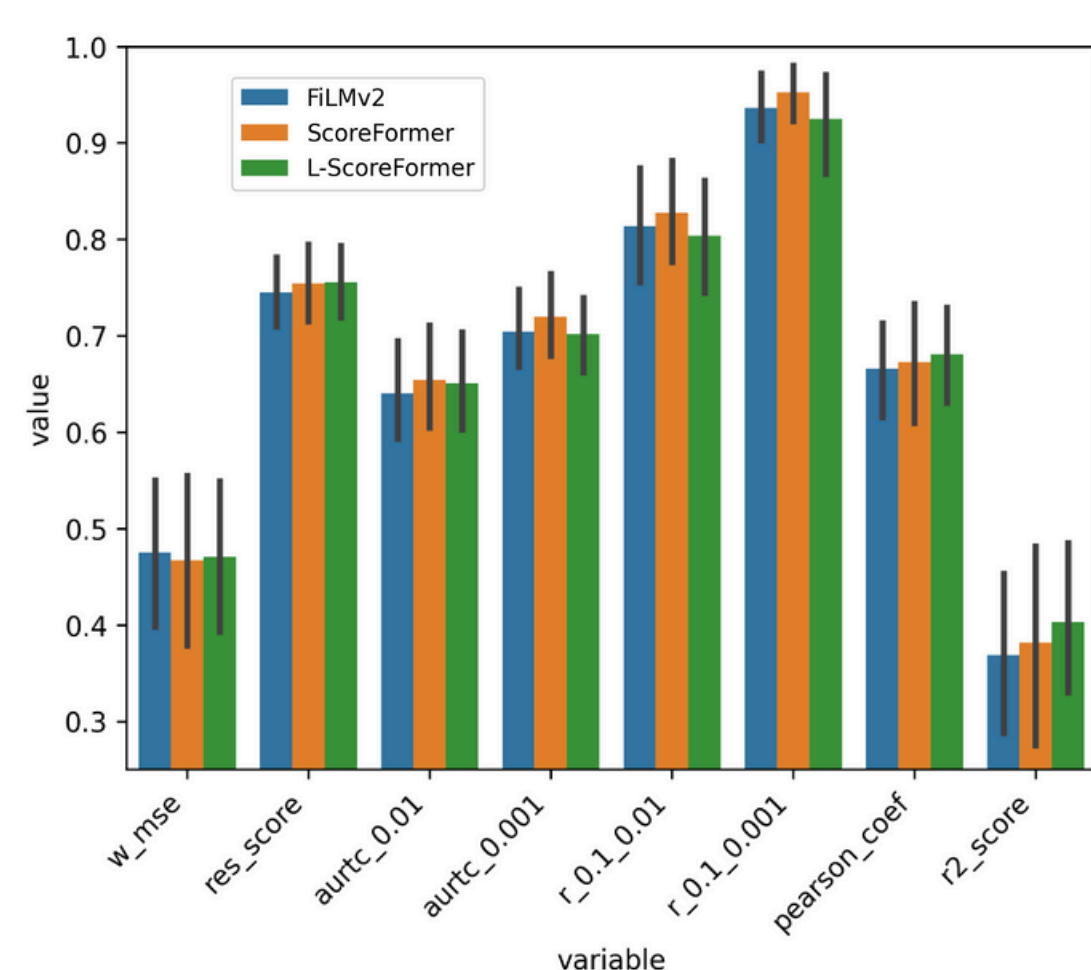
For each new target, **docking scores are computed**, and molecules are represented as graphs



After the training, the model is able to process unseen molecules and **infer the docking score on the target of interest**.

## Results

### Regression results



Scoreformer architectures achieve state-of-the-art performance in regression and hit recovery metrics

### Inference speed

MODEL	SAMPLES/S	128M TIME (H)
FiLMv2	1323.942	26.850
SCOREFORMER	2186.828	16.259
L-SCOREFORMER	2468.404	14.404

MODEL	SPEEDUP	# PARAMETERS
FiLMv2	1.000	102977
SCOREFORMER	1.652	5398273
L-SCOREFORMER	1.864	147457

Despite having more parameters, Scoreformers increase the inference speed of baseline models by 60 and 80 percent

### Generalization

CONFIGURATION	RES	AURTC <sub>0.01</sub>	AURTC <sub>0.001</sub>
SCOREFORMER	<b>0.458</b>	0.344	<b>0.359</b>
L-SCOREFORMER	0.449	<b>0.358</b>	0.336
FiLMv2	0.431	0.333	0.314
REFERENCE	0.746	0.653	0.735

Scoreformer also demonstrates better generalization with out-of-distribution molecules compared to previous SoTA models