Supporting Information

A Molecular Video-derived Foundation Model for Scientific Drug Discovery

Xiang et al., *Nature Communications* 2024.

Table of Contents

Supplemental Materials and Methods

A. Supplementary Experiment Setup

A.1 Dataset details

Pretraining Dataset

PCQM4Mv2 is a quantum chemistry dataset under the PubChemQC project[1], which include 3,378,606 molecules and their corresponding 3D structures calculated by DFT (Density Functional Theory). Since we generate 60 video frames per molecule, we sample 2 million molecules from them for molecular video generation for efficiency. Ultimately, we obtain a total of 120 million video frames for pre-training.

Drug-Target Binding Activity Prediction

Dataset. Following ImageMol[2], we use 10 kinase targets in compound-kinase binding activity prediction and use 10 G protein coupled receptors (GPCRs) targets in ligand-GPCR binding activity prediction. 10 common biochemical kinases datasets are selected from KinomeScan database (https://lincs.hms.harvard.edu/kinomescan/). 10 GPCR datasets are selected from ChEMBL database (https://www.ebi.ac.uk/chembl/), which are the top 10 most used targets. Kinases and GPCRs are a binary classification task and a regression task, respectively. The statistical details for these datasets are provided in **Supplementary Table 31**.

Molecular Property Prediction

Dataset. Molecular property prediction is a crucial step in drug discovery, which can greatly increase the speed of virtual screening. MoleculeNet[3] is a public and popular benchmark, which include a variety of drug discovery task. Here, we choose 6 classification datasets (BBBP, Tox21, HIV, BACE, SIDER and ToxCast) and 6 regression datasets (FreeSolv, ESOL, Lipophilicity, QM7, QM8 and QM9) from MoleculeNet to evaluate VideoMol. These datasets include various properties of drugs, such as pharmacology, physical chemistry, quantum chemistry and biophysics. Especially, due to the large size of HIV, QM8 and QM9 data, we only sample 5 frames at equal intervals to improve training efficiency. Note that the video format in the property prediction tasks is BGR while the other tasks are RGB, which is due to the fact that OpenCV[4] uses BGR format by default. Empirically, the RGB and BGR formats have only a small impact on the performance of VideoMol. The statistical details for these datasets are provided in **Supplementary Table 32**.

Anti-SARS-CoV-2 activity prediction

Dataset. Following REDIAL-2020[5] and ImageMol, we use 11 anti-SARS-CoV-2 activity datasets, which are originally available from the National Center for the Advancement of Translational Science (NCATS) COVID-19 Portal. All of these datasets are a binary classification task and are evaluated using ROC-

AUC metric. The statistical details for these datasets are provided in **Supplementary Table 33**.

Virtual screening on BACE1, COX-1, COX-2, EP4 targets

Dataset from ChEMBL database. We first collect beta-secretase 1 (BACE1) with 6,860 molecules, cyclooxygenase 1 (COX-1) with 3,423 molecules, COX-2 with 4,988 molecules, prostaglandin E receptor 4 (EP4) with 350 molecules from ChEMBL database[6] for training and evaluating VideoMol. Due to the problem of non-uniform formatting of small molecules collected from the ChEMBL database, we clean the collect molecules according to the following guidelines: (1) we only retain those molecules with IC_{50} values; (2) to obtain positive and negative samples, we use different thresholds to convert IC_{50} to class labels. For molecules with an IC_{50} value below the threshold, we set the label to 1 (active), otherwise we set the label to 0 (inactive). Regarding the selection of the threshold, we follow two criteria: (1) Keep the imbalance rate of positive and negative samples at around 0.6, which is obtained by dividing the number of negative samples by the number of positive samples; (2) The maximum threshold for IC_{50} is 100. The statistical details for these targets are provided in **Supplementary Table 10**.

External validation dataset from MedChemExpress website. We obtain known active drugs from the MedChemExpress (MCE) website (https://www.medchemexpress.com/) for further virtual screening. The

MedChemExpress provides a wide range of high-quality research chemicals and biochemicals, including novel life science reagents, reference compounds and natural compounds for laboratory and scientific purposes[7-9]. In details, we collect drugs known to be active against BACE1 (16 drugs), COX-1 (22 drugs), COX-2 (35 drugs), EP4 (8 drugs) targets and with IC_{50} values below the threshold as external validation set. The details for these active drugs are provided in **Supplementary Table 11**, **Supplementary Table 12**, **Supplementary Table 13** and **Supplementary Table 14**.

External validation dataset from DrugBank database. We virtually screened potential anti-BACE1 inhibitors from 2,500 FDA-approved drugs from the DrugBank database[10]. These drugs are available online from the DrugBank website (https://go.drugbank.com/). The data details can be found in **Supplementary Table 19**.

A.2 Hyperparameters of pre-training

We describe the pre-training hyperparameters of VideoMol in **Supplementary Table 30**. We use a temperature of 0.1 for video-aware pretraining and set a learning rate of 0.01, a batch size of 256, a momentum of 0.9, and a weight decay of 1e-4 for pre-training. To evaluate the effectiveness of the pre-training task, we split 10% of the data for validation. Finally, we declare the training platform. We pre-trained VideoMol over 400k steps on a server with an Intel 6248R 48C@3.0GHz CPU and 4 NVIDIA TESLA A100 (40G) GPUs, which

took about 7 days. See **Supplementary Section C.3** for details of computational requirements in pre-training.

A.3 Hyperparameters of finetuning

General framework for finetuning. We spliced a multi-layer perceptron (MLP) after the video encoder for fine-tuning downstream tasks, and its general architecture can be abstracted as shown in **Supplementary Table 34**. In detail, we use the pre-trained video encoder as the backbone to extract the features of molecular videos and input them into an MLP (dropout1→linear1→activator→dropout2→linear2) for prediction.

Hyperparameter search space. We use grid search to find the best combination of hyperparameters with different random seeds on downstream tasks. In compound-kinase interaction and anti-SARS-CoV-2 viral activity prediction tasks, we tune parameters independently for each seed and report the mean and variance, which means the hyperparameters of different seeds may be different. The specific search spaces are shown in **Supplementary Table 35**. In addition to the data augmentation used by VideoMol, we also try data augmentation from ImageMol in evaluation on 10 kinases. We run all fine-tuning experiments on a single Tesla A100 GPU or a single GTX 4090Ti. See **Supplementary Section C.3** for details of computational requirements in fine-tuning.

B. Supplementary Methods

B.1 Molecular fingerprint details

Considering the insufficient semantics of a single molecular fingerprint, we selected a total of 21 fingerprints of 6 different types for comprehensive information. We summarize all used fingerprints in **Supplementary Table 29**:

- Circular-based fingerprints: Circular-based fingerprints include ECFPx (Extended Connectivity Fingerprints) and FCFPx (Functional-Class Fingerprints), where x represents the bond length or diameter centered on the atom. Here, we set x to 0,2,4,6 and 2,4,6 for ECFP and FCFP, respectively. The ECFP is constructed based on the number of atomic connections, the number of non-hydrogen chemical bonds, the atomic number, the positive and negative of atomic charges, the absolute value of atomic charges, and the number of connected hydrogen atoms. The FCFP is constructed based on the information of pharmacophore, such as hydrogen bond acceptors, hydrogen bond donors, negatively ionizable, positively ionizable, aromatic atoms and halogens.
- Path-based fingerprints: Path-based fingerprints include RDKx (x is 5, 6 or 7), HashTT (Topological Torsion) and HashAP (Atom Pair). HashTT and HashAP are similar, which includes information in three dimensions: atomic number, electron number, and adjacent atom number.
- Substructure-based fingerprints: Substructure-based fingerprints include MACCS (Molecular ACCess System) and Avalon. MACCS fingerprints

adopts the substructure encoded by SMARTS and obtains the molecular fingerprint with a length of 167 according to the type of substructure. Avalon fingerprints include features such as atomic symbol path, atom count, augmented symbol path, augmented atom.

- Longer version-based fingerprints: Longer version-based fingerprints enrich the fingerprint information of ECFP4, ECFP6, FCFP4, FCFP6 and LAvalon to 16384 dimensions.
- Pharmacophore-based fingerprints: We use TPATF (Topological Pharmacophore Atomic Triplets Fingerprints) as pharmacophore-based fingerprints, which were obtained using Mayachemtools[11]. TPATF describes the ligand sites necessary for molecular recognition of macromolecules or ligands by using pharmacophore atom types such as hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), cationizable (PI), Anionizable (NI), Hydrophobic (H) and Aromatic (Ar).
- Physicochemistry-based fingerprints: We use RDKDes (RDKit Descriptors) as physicochemistry-based fingerprints, which contains experimental properties or theoretical descriptors, such as molar refractivity, logP, number of heavy atoms, number of bonds, molecular weight, topological polar surface area.

B.2 Model details of pre-training

The **Supplementary Table 30** shows the model details in pre-training phase. We use a 12-layer vision transformer as the video encoder of VideoMol. For the Axis classifier, rotation classifier, angle classifier and chemical classifier, we used the same neural network architecture, which consists of a fully connected layer with a Softplus activator and a task-dependent fully connected layer.

C. Supplementary Results

C.1 Results on pre-training

The **Supplementary Figure 3** and **Supplementary Figure 4** describe the loss curve and the corresponding accuracy curve of VideoMol in the pre-training phase, respectively. We can clearly see that the loss of each pre-training task in VideoMol is gradually decreasing, which shows that each task can work well together (**Supplementary Figure 3**). At the same time, each pre-training task achieves good accuracy on the validation set, indicating the effectiveness of VideoMol (**Supplementary Figure 4**).

C.2 Results of different representations on 8 basic attributes

To fairly compare the effects of different representations, we evaluated the representation without using any self-supervised tasks. It is well known that

the development of drug discovery depends on accurately capturing chemical and biological representations of molecules. Here, we used several commonly used representative methods (such as GCN, GIN, EGNN, and the image representation used by ImageMol) to inspect the model's ability to understand the 8 basic attributes of molecules, including molecular weight, MolLogP, MolMR, BalabanJ, NumHAcceptors, NumHDonors, NumValenceElectrons and TPSA.

We collected the first 10,000 molecules from the pre-training dataset and used exactly the same experimental setup (batch size of 8, learning rate of 0.005, training epoch of 30, and only training task-relevant predictor) for fair comparison. In detail, we split the training set, validation set, and test set using a ratio of 8:1:1 and reported the results on the test set based on the best validation set score. As shown in **Supplementary Table 1**, we found that VideoMol using only one frame outperformed that of the 2D graph-based methods, the 3D-based graph method and the 2D image-based method, revealing the advantage of the proposed 3D representation. Specifically, compared with the second-place ImageMol without pre-training, the performance of video-1frame improved by 9.9%. When we utilized all video frames (video-60frame), the performance is further significantly improved from 12.47 to 7.51 with a 39.8% improvement rate.

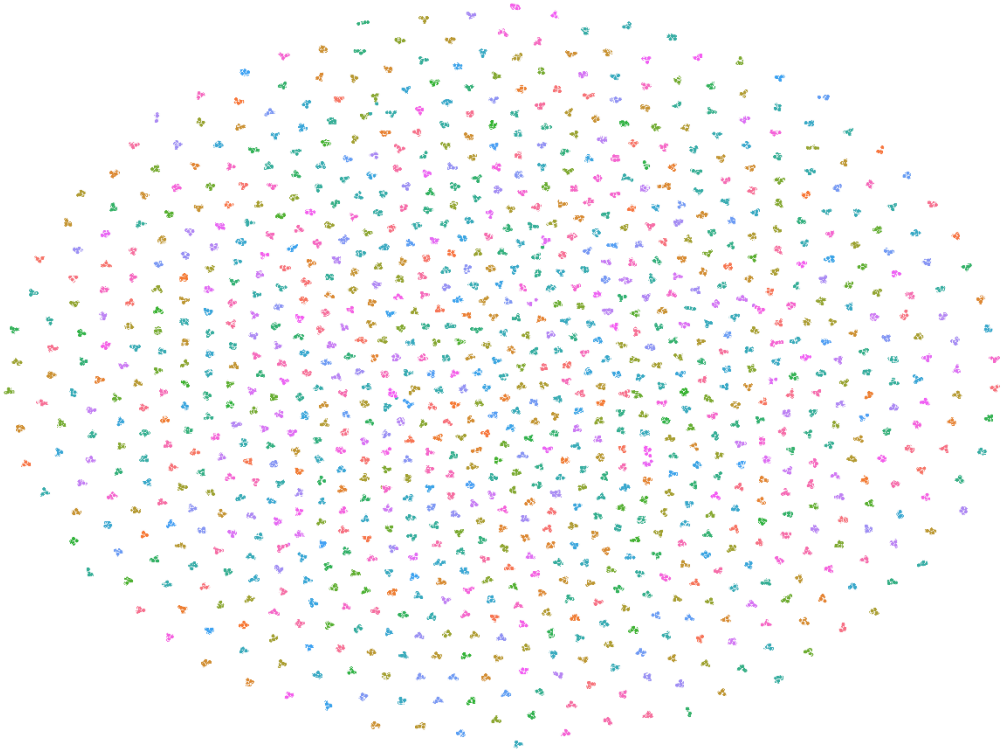
In summary, the proposed 3D representation (whether based on a single frame image or a 60-frame video) has advantages compared to existing

molecular representation approaches. We will further improve our VideoMol framework by increasing the number of 3D frames and integrating other types of 3D representation (such as AlphaFold3[12]) in the near future.

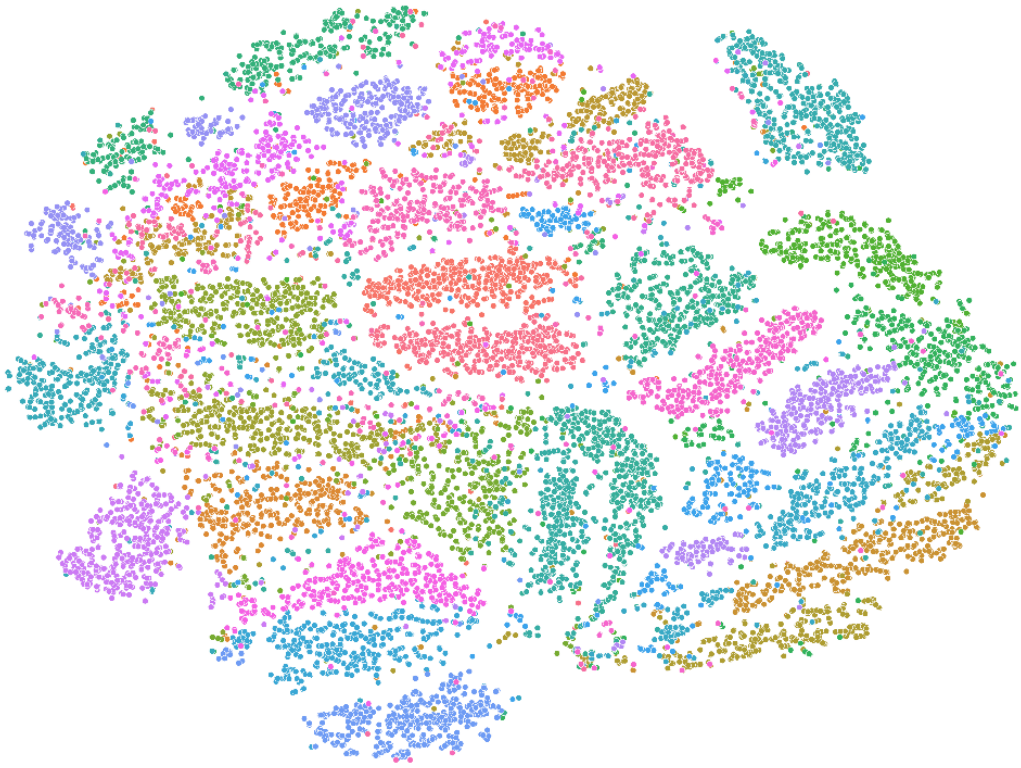
C.3 Computational requirements of VideoMol

Here, we detail the computational requirements of the pre-training, fine-tuning, and screening stages in **Supplementary Table 36**. In the pre-training phase, as shown in **Supplementary Table 36**(a), VideoMol uses 256 frames in each batch for training, which requires 37G of GPU memory and takes about 9 hours to complete 1 epoch on 2 million molecular videos with 60 frames. Next, we use 10,000 molecules and study the impact of different batch sizes (#frame/batch) on memory and training speed in **Supplementary Table 36**(b). We find that fine-tuning does not occupy a large amount of memory, and only requires at least 2.3G of GPU memory to complete. Finally, we evaluate the computational requirements when performing virtual screening on 1 million molecules in **Supplementary Table 36**(c). We find that even using all frames during virtual screening, it only took about 9 hours. In addition, for ever-expanding data sizes, screening can be accelerated by reducing the number of frames. Therefore, VideoMol can complete large-scale virtual screening in a faster time.

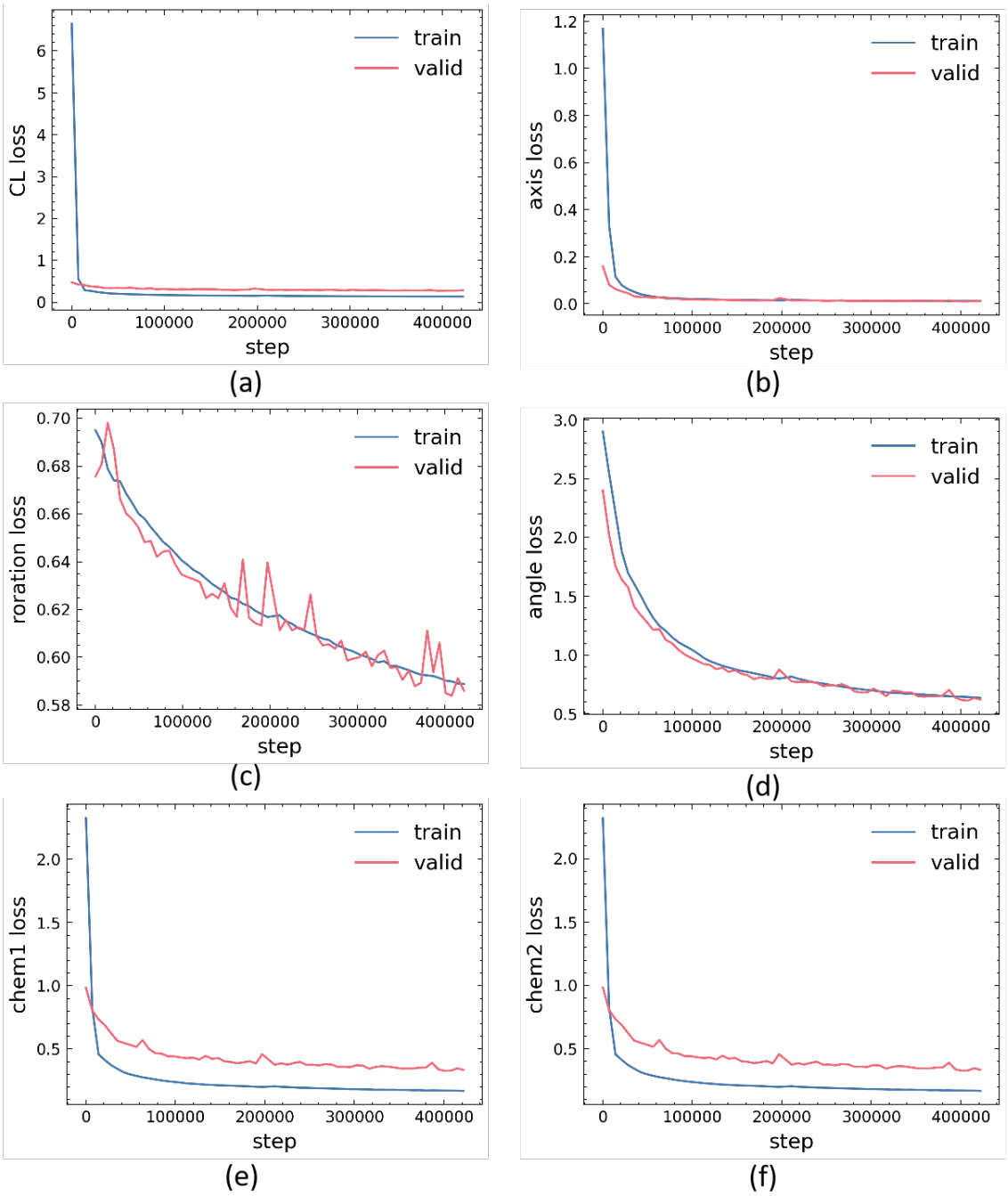
Supplementary Figures



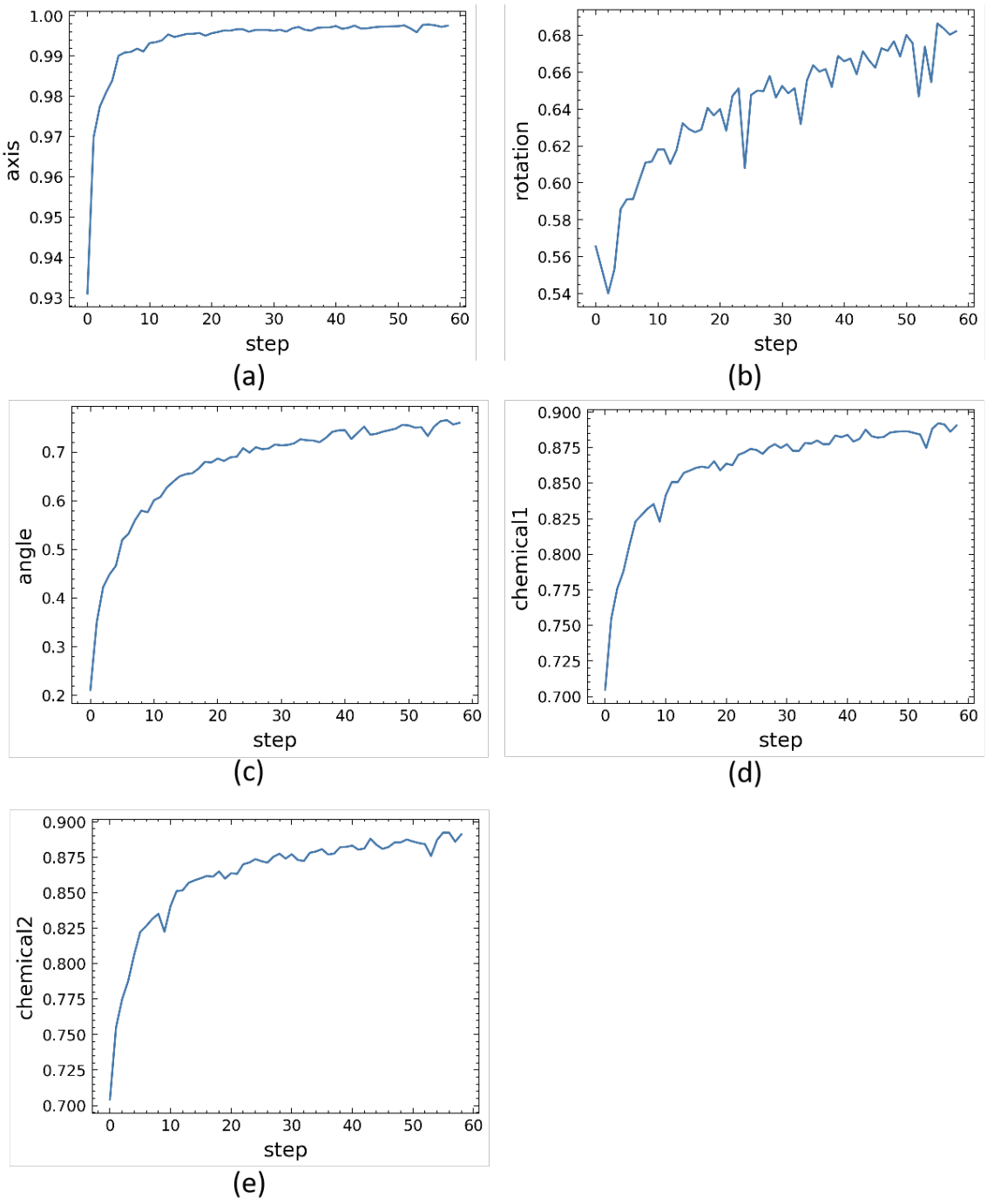
Supplementary Figure 1: t-SNE visualization of 1,000 molecular videos. Different colors represent frames in different cluster videos. Different colors represent different molecules video. DB index (=0.182) is a metric to evaluate the clustering quality, and the larger the value, the better the clustering performance.



Supplementary Figure 2: t-SNE visualization of the representations of VideoMol. The total number of samples is 15,000 (including 30 clusters with 500 samples per cluster). Different colors represent different clusters. Davies Bouldin (DB) index (=3.66) represents the intra-cluster distance divided by the inter-cluster distance, where the smaller the value, the better the clustering performance.



Supplementary Figure 3: VideoMol loss details during the pre-training phase. The x-axis and y-axis represent the training step and loss value, respectively. Subfigures (a)-(d) represent $\mathcal{L}_V$, $\mathcal{L}_{axis}$, $\mathcal{L}_{rotation}$, $\mathcal{L}_{angle}$, respectively. The mean of the losses for subfigures (e) and (f) represent $\mathcal{L}_C$.



Supplementary Figure 4: The performance of VideoMol on validation set during the pre-training phase. The x-axis and y-axis represent the training step divided by 7,000 and accuracy, respectively. Subfigures (a)-(c) represent the performance of VideoMol on axis prediction, rotation prediction and angle prediction, respectively. The mean of the accuracy for subfigures (d) and (e)

represent performance of chemical-aware pre-training task. We evaluate the performance of VideoMol on the validation set approximately 7000 steps.

Supplementary Tables

Supplementary Table 1: The RMSE results of different molecular representations on 8 basic attributes (molecular weight, MolLogP, MolMR, BalabanJ, NumHAcceptors, NumHDonors, NumValenceElectrons, TPSA). (.xlsx)

	modality	model	use conformer?	**prop**
graph-based	2D graph	GCN	×	62.304
		GIN	×	62.980
		EGNN	×	17.418
	3D graph	EGNN	√	16.684
image-based	image from imagemol	ResNet18	×	12.469
	video-1frame	ResNet18	√	11.237
	video-5frame	ResNet18	√	8.088
	video-60frame	ViT	√	**7.511**

Supplementary Table 2: Summary of state-of-the-art methods. (.xlsx)

Representation Category	Model	Input	Backbone	Pre-training database
sequence	{RNN, TRFM}_{LR, MLP, RF}	SMILES	RNN or Transformer	ChEMBL24 (0.861M)
	CHEM-BERT	SMILES	BERT	PubChem (10M)
2D graph	InfoGraph	Graph	GIN	-
	GPT-GNN	Graph	HGT	OAG (179 million nodes & 2 billion edges) and Amazon (113 million nodes)
	ContextPred	Graph	5-layer GIN	ZINC15 (2M)
	GraphLoG	Graph	5-layer GIN	ZINC15 (2M)
	G-Contextual	Graph	Grover	ZINC15 & Chembl (11 million)
	G-Motif	Graph	Grover	ZINC15 & Chembl (11 million)
	AD-GCL	Graph	5-layer GIN	ZINC15 (2M)
	JOAO	Graph	5-layer GIN	ZINC15 (2M)
	SimGRACE	Graph	5-layer GIN	ZINC15 (2M)
	GraphCL	Graph	5-layer	ZINC15 (2M)

	GraphMAE	Graph	GIN 5-layer GIN	ZINC15 (2M)
	MGSSL	Graph	5-layer GIN	ZINC15 (250K)
	AttrMask	Graph	5-layer GIN	ZINC15 (2M)
	MolCLR	Graph	GCN+GIN	PubChem (10M)
	Mole-BERT	Graph	5-layer GIN	ZINC15 (2M)
	GraphCL	Graph	5-layer GIN	ZINC15 (2M)
3D graph	3D InfoMax	Geometry+ Graph	PNA+Net3D	QM9(50k)+GEOM-Drugs(140k)+Qmugs(620k)
	GraphMVP	Geometry+ Graph	5-layer GIN+SchNet	GEOM (50K)
	Uni-Mol	Geometry+ Protein Pockets	Transformer	ZINC/ChemBL+PDB (209 M)
image	ImageMol	Image	ResNet18	PubChem (10M)

Supplementary Table 3: The ROC-AUC performance of different methods on 10 main types of biochemical kinases from KinomeScan datasets with balanced scaffold split. All compared results are obtained from ImageMol.

(.xlsx)

	BTK	CDK4-cyclinD3	EGFR	FGFR1	FGFR2
MoCLRGIN	0.556±0.118	0.778±0.171	0.583±0.067	0.695±0.249	0.667±0.132
MoCLRGCN	0.602±0.129	0.944±0.039	0.750±0.051	0.619±0.378	0.667±0.052
RNN_LR	0.611±0.000	0.667±0.000	0.536±0.000	0.771±0.000	0.741±0.000
TRFM_LR	0.694±0.000	0.750±0.000	0.821±0.000	0.743±0.000	0.704±0.000
RNN_MLP	0.556±0.023	0.833±0.000	0.536±0.029	0.848±0.059	0.716±0.046
TRFM_MLP	0.537±0.013	0.639±0.039	0.667±0.061	0.643±0.031	0.741±0.000
RNN_RF	0.546±0.013	0.917±0.000	0.548±0.017	0.476±0.027	0.685±0.055
TRFM_RF	0.639±0.039	0.639±0.039	0.607±0.000	0.476±0.013	0.556±0.030
CHEM-BERT	0.648±0.013	0.583±0.297	0.845±0.094	0.429±0.117	0.765±0.106
ImageMol	0.843±0.026	0.917±0.068	0.857±0.000	**0.857±0.023**	0.852±0.052
VideoMol	**0.861±0.023**	**0.972±0.039**	**0.905±0.017**	0.848±0.027	**0.988±0.017**

------ Continue ------

	FGFR3	FGFR4	FLT3	KPCD3	MET	Average
MoCLRGIN	0.760±0.039	0.773±0.121	0.722±0.091	0.571±0.107	0.611±0.236	0.6716
MoCLRGCN	0.792±0.106	0.537±0.013	0.722±0.208	0.505±0.067	0.574±0.052	0.6712
RNN_LR	0.646±0.015	0.528±0.000	0.778±0.000	0.457±0.000	0.796±0.026	0.6531

TRFM_LR	0.812±0.000	0.639±0.000	0.611±0.000	0.438±0.027	0.778±0.000	0.6990
RNN_MLP	0.469±0.026	0.269±0.035	0.630±0.105	0.410±0.036	0.667±0.045	0.5934
TRFM_MLP	0.802±0.015	0.676±0.035	0.667±0.136	0.219±0.027	0.556±0.000	0.6147
RNN_RF	0.312±0.000	0.389±0.000	0.519±0.026	0.343±0.000	0.500±0.000	0.5235
TRFM_RF	0.646±0.078	0.602±0.013	0.546±0.035	0.262±0.058	0.593±0.026	0.5566
CHEM-BERT	0.438±0.077	0.528±0.060	0.574±0.189	0.557±0.091	0.944±0.000	0.6311
ImageMol	0.854±0.064	0.833±0.045	0.722±0.120	0.762±0.088	**0.963±0.026**	0.8460
VideoMol	**0.896±0.039**	**0.852±0.080**	**0.981±0.026**	**0.867±0.036**	**0.963±0.026**	**0.9133**

Supplementary Table 4: The RMSE and MAE performance of different methods on 10 GPCR with balanced scaffold split. The lower the value, the better the performance. All compared results are obtained from ImageMol. (.xlsx)

	1. 5HT1A		**2. 5HT2A**		**3. AA1R**	
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
MoCLRGIN	0.850±0.021	0.670±0.012	0.853±0.019	0.642±0.014	0.786±0.015	0.588±0.009
MoCLRGCN	0.949±0.027	0.764±0.014	0.875±0.008	0.681±0.024	0.856±0.026	0.662±0.026
RNN_LR	1.574±0.091	0.937±0.019	1.602±0.245	1.103±0.151	1.073±0.087	0.762±0.057
TRFM_LR	1.636±0.004	1.109±0.001	1.389±0.000	0.999±0.001	1.060±0.003	0.810±0.001
RNN_MLP	0.957±0.013	0.768±0.010	1.167±0.010	0.890±0.003	0.848±0.004	0.662±0.008
TRFM_MLP	0.939±0.034	0.730±0.025	1.013±0.026	0.728±0.021	0.878±0.051	0.657±0.031
RNN_RF	0.788±0.004	0.617±0.004	1.001±0.001	0.747±0.001	0.717±0.003	0.554±0.002
TRFM_RF	0.855±0.001	0.672±0.001	1.011±0.002	0.777±0.002	0.740±0.001	0.568±0.001
CHEM-BERT	0.876±0.018	0.706±0.012	0.909±0.057	0.682±0.056	0.734±0.038	0.544±0.027
ImageMol	0.776±0.012	0.620±0.014	0.780±0.017	**0.578±0.022**	0.711±0.012	0.554±0.009
VideoMol	**0.708±0.017**	**0.547±0.015**	**0.775±0.017**	**0.578±0.009**	**0.655±0.007**	**0.496±0.006**
	4. AA2AR		**5. AA3R**		**6. CNR2**	
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
MoCLRGIN	0.748±0.012	0.588±0.008	0.840±0.014	0.692±0.010	0.926±0.047	0.758±0.036
MoCLRGCN	0.819±0.011	0.651±0.008	0.855±0.010	0.700±0.011	0.978±0.023	0.803±0.021
RNN_LR	1.801±0.600	1.193±0.335	2.295±0.463	1.190±0.155	5.505±0.093	1.611±0.032
TRFM_LR	1.130±0.000	0.906±0.000	1.155±0.001	0.919±0.001	1.700±0.001	1.213±0.000
RNN_MLP	0.967±0.002	0.773±0.005	0.883±0.010	0.707±0.012	1.091±0.015	0.881±0.013
TRFM_MLP	0.948±0.013	0.744±0.005	0.945±0.010	0.749±0.014	1.144±0.055	0.903±0.038
RNN_RF	0.887±0.002	0.692±0.001	0.796±0.009	0.624±0.007	0.965±0.002	0.766±0.001
TRFM_RF	0.926±0.003	0.735±0.004	0.856±0.001	0.701±0.002	0.965±0.002	0.800±0.002
CHEM-BERT	0.862±0.071	0.674±0.058	0.861±0.058	0.684±0.047	0.925±0.051	0.727±0.041
ImageMol	0.734±0.015	0.573±0.009	0.793±0.008	0.634±0.001	0.905±0.004	0.717±0.015
VideoMol	**0.712±0.011**	**0.543±0.005**	**0.786±0.006**	**0.617±0.004**	**0.864±0.005**	**0.679±0.010**
	7. DRD2		**8. DRD3**		**9. HRH3**	
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
MoCLRGIN	0.814±0.009	0.591±0.007	0.858±0.017	0.673±0.022	0.734±0.006	0.581±0.004
MoCLRGCN	0.855±0.022	0.634±0.017	0.914±0.024	0.725±0.025	0.740±0.016	0.576±0.006

RNN_LR	1.142±0.077	0.839±0.038	1.316±0.011	0.942±0.005	1.616±0.236	0.943±0.070
TRFM_LR	1.000±0.000	0.719±0.000	1.219±0.000	0.914±0.000	1.169±0.002	0.911±0.002
RNN_MLP	0.895±0.006	0.694±0.005	1.021±0.007	0.819±0.007	0.871±0.015	0.702±0.011
TRFM_MLP	0.919±0.016	0.686±0.016	1.012±0.041	0.790±0.023	0.863±0.011	0.676±0.009
RNN_RF	0.837±0.001	0.612±0.001	0.861±0.001	0.685±0.001	0.771±0.002	0.613±0.002
TRFM_RF	0.864±0.001	0.636±0.002	0.904±0.001	0.717±0.001	0.770±0.002	0.602±0.001
CHEM-BERT	0.816±0.011	0.587±0.013	0.803±0.029	0.631±0.026	0.770±0.033	0.594±0.022
ImageMol	0.772±0.014	0.573±0.009	0.735±0.018	0.576±0.014	0.710±0.006	0.561±0.006
VideoMol	**0.742±0.003**	**0.556±0.005**	**0.715±0.014**	**0.554±0.012**	**0.668±0.008**	**0.506±0.002**

10. OPRM

	RMSE	**MAE**
MoCLRGIN	0.856±0.008	0.664±0.016
MoCLRGCN	0.853±0.009	0.653±0.020
RNN_LR	2.649±1.024	1.744±0.614
TRFM_LR	1.694±0.001	1.282±0.000
RNN_MLP	1.022±0.014	0.781±0.008
TRFM_MLP	1.084±0.009	0.849±0.007
RNN_RF	0.876±0.010	0.671±0.009
TRFM_RF	0.852±0.002	0.660±0.003
CHEM-BERT	0.893±0.024	0.672±0.019
ImageMol	0.849±0.018	0.645±0.015
VideoMol	**0.795±0.015**	**0.579±0.011**

Supplementary Table 5: The ROC-AUC performance (%) of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run 10 times using random seeds from 0 to 9. GraphMVP-C, Mole-BERT, Uni-Mol and ImageMol are reproduced from their source code and other results from Mole-BERT. (.xlsx)

	Tox21	ToxCast	Sider	HIV	BBBP	BACE
#Molecules	7831	8576	1427	41127	2039	1513
#Task	12	617	27	1	1	1
InfoGraph	73.3 (0.6)	61.8 (0.4)	58.7 (0.6)	74.2 (0.9)	68.7 (0.6)	74.3 (2.6)
GPT-GNN	74.9 (0.3)	62.5 (0.4)	58.1 (0.3)	65.2 (2.1)	64.5 (1.4)	77.9 (3.2)
ContextPred	73.6 (0.3)	62.6 (0.6)	59.7 (1.8)	75.6 (1.0)	70.6 (1.5)	78.8 (1.2)
GraphLoG	75.0 (0.6)	63.4 (0.6)	59.6 (1.9)	76.1 (0.8)	68.7 (1.6)	78.6 (1.0)
G-Contextual	75.0 (0.6)	62.8 (0.7)	58.7 (1.0)	76.3 (1.5)	69.9 (2.1)	79.3 (1.1)
G-Motif	73.6 (0.7)	62.3 (0.6)	61.0 (1.5)	73.8 (1.2)	66.9 (3.1)	73.0 (3.3)
AD-GCL	74.9 (0.4)	63.4 (0.7)	61.5 (0.9)	76.7 (1.2)	70.7 (0.3)	76.6 (1.5)
JOAO	74.8 (0.6)	62.8 (0.7)	60.4 (1.5)	76.9 (0.7)	66.4 (1.0)	73.2 (1.6)
SimGRACE	74.4 (0.3)	62.6 (0.7)	60.2 (0.9)	75.0 (0.6)	**71.2 (1.1)**	74.9 (2.0)

GraphCL	75.1 (0.7)	63.0 (0.4)	59.8 (1.3)	75.1 (0.7)	67.8 (2.4)	74.6 (2.1)
GraphMAE	75.2 (0.9)	63.6 (0.3)	60.5 (1.2)	76.8 (0.6)	**71.2 (1.0)**	78.2 (1.5)
3D InfoMax	74.5 (0.7)	63.5 (0.8)	56.8 (2.1)	76.1 (1.3)	69.1 (1.2)	78.6 (1.9)
MGSSL	75.2 (0.6)	63.3 (0.5)	61.6 (1.0)	75.8 (0.4)	68.8 (0.6)	78.8 (0.9)
AttrMask	75.1 (0.9)	63.3 (0.6)	60.5 (0.9)	75.3 (1.5)	65.2 (1.4)	77.8 (1.8)
MolCLR	75.5 (0.5)	63.9 (0.5)	60.3 (1.3)	74.4 (1.3)	66.8 (3.4)	75.3 (2.9)
GraphMVP-C	74.6(0.4)	63.4 (0.6)	60.6 (1.3)	77.1 (2.1)	69.9 (1.4)	79.6 (1.7)
ImageMol	75.5 (1.0)	65.6 (0.9)	64.9 (1.3)	76.8 (1.3)	70.5 (1.3)	78.1 (3.5)
Uni-Mol (1 conf)	78.3 (0.4)	68.7 (0.5)	63.7 (1.3)	79.2 (1.0)	69.6 (2.0)	81.0 (3.9)
Uni-Mol (10 conf)	**78.8 (0.7)**	**69.0 (0.5)**	63.6 (1.4)	79.2 (0.9)	69.9 (2.7)	81.7 (3.4)
Mole-BERT	77.0 (0.3)	64.4 (0.2)	63.2 (0.7)	77.7 (0.7)	65.7 (2.3)	80.2 (0.9)
VideoMol	**78.8 (0.4)**	66.7 (0.5)	**66.3 (0.9)**	**79.4 (0.5)**	70.7 (1.5)	**82.4 (0.9)**
Rank	1	2	1	1	3	1

Supplementary Table 6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run 10 times using random seeds from 0 to 9. We report RMSE for FreeSolv, ESOL and Lipo datasets and MAE for QM7 and QM8, QM9 datasets, respectively. We reproduced all comparison methods using the same settings. We use GIN backbone for MolCLR because it achieves the best results. (.xlsx)

RMSE	FreeSolv	ESOL	Lipo	QM7	QM8	QM9
GraphMVP	2.559±0.158	1.322±0.062	0.773±0.016	120.344±6.237	0.02049±0.00032	**0.00891±0.00010**
EdgePred	2.843±0.091	1.367±0.041	0.778±0.013	104.387±3.292	0.02058±0.00061	0.00929±0.00009
GraphMVP-C	2.766±0.199	1.333±0.055	0.768±0.013	121.022±5.699	0.02022±0.00047	0.00896±0.00011
MolCLR	3.112±0.638	1.462±0.068	0.799±0.018	144.426±6.591	0.03598±0.00085	0.01488±0.00020
ImageMol	2.113±0.235	0.964±0.067	**0.702±0.060**	116.384±8.445	0.02419±0.00033	0.02061±0.00019
Mole-BERT	2.988±0.155	1.115±0.017	0.727±0.006	101.922±2.331	0.02073±0.00033	0.00910±0.00010
VideoMol	**1.725±0.053**	**0.866±0.017**	0.743±0.009	**76.436±1.561**	**0.01890±0.00020**	0.00896±0.00003

Supplementary Table 7: The ROC-AUC performance of different methods on 11 SARS-CoV-2 datasets with balanced scaffold split. All compared results are obtained from ImageMol. (.xlsx)

	3CL	ACE2	hCYTOX	MERS-PPE_cs	MERS-PPE	CoV1-PPE_cs	CoV1-PPE	CPE	Cytotox	AlphaLISA	TruHit	Mean
REDIAL-2020	0.713	0.753	0.710	0.703	0.696	0.661	0.665	0.651	0.688	0.79	0.734	0.706
ImageMol	**0.762±0.007**	0.720±0.001	0.727±0.009	0.771±0.009	0.773±0.011	0.775±0.005	0.703±0.008	0.669±0.011	0.728±0.001	0.793±0.007	0.806±0.006	0.748
VideoMol	0.709±0.006	**0.759±0.025**	**0.765±0.003**	**0.828±0.027**	**0.814±0.004**	**0.836±0.029**	**0.737±0.007**	**0.747±0.013**	**0.761±0.002**	**0.841±0.004**	**0.862±0.002**	**0.787**

Supplementary Table 8: The uncertainty intervals with 95% confidence intervals of ImageMol and VideoMol on 10 ligand-GPCR binding activity prediction datasets. UI(·) represents the uncertainty intervals and "Improvement" represents the relative performance improvement of VideoMol compared to ImageMol. (.xlsx)

Dataset	UI(RMSE)			UI(MAE)		
	ImageMol	VideoMol	Improvement	ImageMol	VideoMol	Improvement
5HT1A	0.782±0.057	0.719±0.059	8.06%	0.629±0.048	0.550±0.046	12.56%
5HT2A	0.816±0.109	0.810±0.102	0.74%	0.587±0.059	0.583±0.059	0.68%
AA1R	0.718±0.062	0.662±0.068	7.80%	0.559±0.045	0.499±0.046	10.73%
AA2AR	0.739±0.055	0.714±0.056	3.38%	0.575±0.045	0.544±0.045	5.39%
AA3R	0.796±0.056	0.795±0.065	0.13%	0.632±0.051	0.622±0.053	1.58%
CNR2	0.916±0.073	0.878±0.072	4.15%	0.722±0.060	0.686±0.064	4.99%
DRD2	0.779±0.060	0.749±0.053	3.85%	0.574±0.041	0.559±0.040	2.61%
DRD3	0.738±0.053	0.704±0.054	4.61%	0.580±0.044	0.548±0.042	5.52%
HRH3	0.747±0.067	0.669±0.061	10.44%	0.582±0.050	0.507±0.047	12.89%
OPRM	0.898±0.089	0.797±0.075	11.25%	0.667±0.065	0.584±0.062	12.44%

Supplementary Table 9: The uncertainty intervals with 95% confidence intervals of ImageMol and VideoMol on 11 SARS-CoV-2 viral activity prediction datasets. UI(·) represents the uncertainty intervals and

“Improvement” represents the relative performance improvement of VideoMol compared to ImageMol. (.xlsx)

Dataset	UI(AUC)		
	ImageMol	VideoMol	Improvement
3CL	0.685±0.117	0.710±0.110	3.65%
ACE2	0.658±0.133	0.763±0.112	15.96%
hCYTOX	0.736±0.087	0.760±0.087	3.26%
MERS-PPE_cs	0.727±0.119	0.817±0.103	12.38%
MERS-PPE	0.720±0.082	0.799±0.0743	10.97%
CoV1-PPE_cs	0.688±0.115	0.832±0.098	20.93%
CoV1-PPE	0.701±0.063	0.736±0.060	4.99%
CPE	0.646±0.084	0.736±0.077	13.93%
Cytotox	0.729±0.051	0.760±0.054	4.25%
AlphaLISA	0.762±0.062	0.836±0.049	9.71%
TruHit	0.772±0.059	0.855±0.046	10.75%

Supplementary Table 10: Basic statistical information of 4 common targets (beta-secretase 1 [BACE1], cyclooxygenase 1 [COX-1], COX-2, prostaglandin E receptor 4 [EP4]) collected from ChEMBL database. #Train, #Val and #Test represent the numbers of train set, valid set, and test set, respectively. IC50 thresh (nM) represents the dividing value between positive and negative samples, where those below the threshold are positive samples, and those above the threshold are negative samples. #Pos and #Neg represent the number of positive samples and the number of negative samples, respectively. Imbalanced ratio reflects the degree of class imbalance, which is calculated by dividing the number of negative samples by the number of positive samples. #Tasks represents the number of binary prediction task.

	BACE1	COX-1	COX-2	EP4
#Train	5488	2737	3990	280

#Val	686	343	499	35
#Test	686	343	499	35
IC50 thresh (nM)	80	100	100	30
#Pos/#Neg	2563/4297	142/3281	1006/3982	133/217
Imbalanced ratio	0.597	0.043	0.253	0.613
#Tasks	1			
Metric	ROC-AUC			

Supplementary Table 11: The anti-BACE1 drugs with an IC50 threshold of 80 from MedChemExpress website. The cas is the unique identifier for a drug.

(.xlsx)

index	name	cas	IC50(nM)	canonical smiles
1	BACE-1 inhibitor 2	2563970-92-1	1.5	C[C@H]1C[C@H]2CC(F)(F)C(N)=N[C@@]2(c2cc(NC(=O)c3cnc(OCF)cn3)ccc2F)CO1
2	BACE-1 inhibitor 1	1262858-14-9	56	NC1=N[C@@](c2cc(NC(=O)c3ccc(Br)cn3)ccc2F)(C(F)F)COC1
3	BACE1-IN-13	1397683-26-9	2.9	COc1cnc(C(=O)Nc2ccc(F)c([C@]3(C)Cn4nc(C#N)cc4C(N)=N3)c2)cn1
4	BACE1-IN-1	1310347-50-2	32	C[C@]1(c2cc(NC(=O)c3ccc(C#N)cn3)ccc2F)N=C(N)OCC1(F)F
5	BACE1-IN-6	2079945-75-6	1.5	C[C@]1(c2cc(/C=C(\F)c3ccc(C#N)cn3)ccc2F)N=C(N)S[C@@]2(C(=O)N3CCOCC3)C[C@H]21
6	BACE1-IN-4	2361157-92-6	3.8	C[C@]1(c2cc(NC(=O)c3cnc(OCF)cn3)ccc2F)N=C(N)SCC12CCS(=O)(=O)CC2
7	BACE1-IN-2	1352416-78-4	22	C[C@]1(c2cc(NC(=O)c3ccc(C#N)cn3)ccc2F)N=C(N)CO[C@H]1C(F)(F)F
8	BACE1/2-	26710	10	C[C@@H]1C[C@H]1COC(C)(C)[C@@H]1C[C@

	IN-1	36-34-1		H]1[C@]12CN(c3ncccn3)C[C@H]1CSC(N)=N2
9	Umibecestat	1387560-01-1	11	C[C@@]1(c2nc(NC(=O)c3ncc(C(F)(F)F)cc3Cl)ccc2F)CO[C@@](C)(C(F)(F)F)C(N)=N1
10	LY2886721	1262036-50-9	20.3	NC1=N[C@@]2(c3cc(NC(=O)c4ccc(F)cn4)ccc3F)COC[C@H]2CS1
11	beta-Secretase Inhibitor IV	797035-11-1	15.6	C[C@@H](NC(=O)c1cc(C(=O)N[C@@H](Cc2ccccc2)[C@H](O)CNC2CC2)cc(N(C)S(C)(=O)=O)c1)c1ccccc1
12	AMG-8718	1215868-94-2	0.7	CC1(C#Cc2cnc3c(c2)[C@]2(COC(N)=N2)c2cc(-c4cccnc4F)ccc2O3)COC1
13	PF-06751979	1818339-66-0	7.3	C[C@H]1C[C@H]2CSC(N)=N[C@@]2(c2nc(NC(=O)c3ccc(OC(F)F)cn3)cs2)CO1
14	JNJ-67569762	2380313-26-6	2.7	CCS(=O)(=O)[C@]1(C)CC[C@@](CF)(c2cc(NC(=O)c3cc4c(cn3)OC(F)(F)O4)ccc2F)N=C1N
15	NB-360	1262857-73-7	5	Cc1cc(C#N)cnc1C(=O)Nc1ccc(F)c([C@]2(C)CO[C@@](C)(C(F)(F)F)C(N)=N2)c1
16	AM-6494	1874232-80-0	0.4	C#CCOc1cnc(C(=O)Nc2cc(F)c(F)c([C@@]3(C)N=C(N)S[C@@]4(COC)C[C@H]43)c2)cn1

Supplementary Table 12: The anti-COX-1 drugs with an IC50 threshold of 100 from MedChemExpress website. The cas is the unique identifier for a drug. (.xlsx)

index	name	cas	IC50 (nM)	canonical smiles
1	Indomethacin	53-86-1	18	COc1ccc2c(c1)c(CC(=O)O)c(C)n2C(=O)c1ccc(Cl)cc1
2	Indomethacin sodium	7681-54-1	18	COc1ccc2c(c1)c(CC(=O)[O-])c(C)n2C(=O)c1ccc(Cl)cc1.[Na+]
3	Indomethacin sodium hydrate	74252-25-8	18	COc1ccc2c(c1)c(CC(=O)OC/C=C(\C)CC/C=C(\C)CCC=C(C)C)c(C)n2C(=O)c1ccc(Cl)cc1
4	Diclofenac	15307-86-5	4	O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl
5	Diclofenac potassium	15307-81-0	4	O=C([O-])Cc1ccccc1Nc1c(Cl)cccc1Cl.[K+]
6	Diclofenac diethylamine	78213-16-8	4	CCNCC.O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl
7	Diclofenac Sodium	15307-79-6	4	O=C([O-])Cc1ccccc1Nc1c(Cl)cccc1Cl.[Na+]
8	Mofezolac	78967	1.44	COc1ccc(-c2noc(CC(=O)O)c2-

		-07-4		c2ccc(OC)cc2)cc1
9	Tenidap	120210-48-2	30	NC(=O)N1C(=O)/C(=C(\O)c2cccs2)c2cc(Cl)ccc21
10	Tenidap-d3	142741-60-4	30	NC(=O)N1C(=O)C(C(=O)c2cccs2)c2cc(Cl)ccc21
11	SC-560	188817-13-2	9	COc1ccc(-n2nc(C(F)(F)F)cc2-c2ccc(Cl)cc2)cc1
12	Lornoxicam	70374-39-9	5	CN1C(C(=O)Nc2ccccn2)=C(O)c2sc(Cl)cc2S1(=O)=O
13	S-(+)-Ketoprofen	22161-81-5	1.9	C[C@H](C(=O)O)c1cccc(C(=O)c2ccccc2)c1
14	FR122047	130717-51-0	28	COc1ccc(-c2nc(C(=O)N3CCN(C)CC3)sc2-c2ccc(OC)cc2)cc1.Cl
15	Ketorolac tromethamine salt	74103-07-4	20	NC(CO)(CO)CO.O=C(c1ccccc1)c1ccc2n1CCCC2C(=O)O
16	Eltenac	72895-88-6	30	O=C(O)Cc1cscc1Nc1c(Cl)cccc1Cl
17	Bromfenac	91714-94-2	5.56	Nc1c(CC(=O)O)cccc1C(=O)c1ccc(Br)cc1
18	Bromfenac sodium	91714-93-1	5.56	Nc1c(CC(=O)[O-])cccc1C(=O)c1ccc(Br)cc1.[Na+]
19	Ketoprofen	22071-15-4	2	CC(C(=O)O)c1cccc(C(=O)c2ccccc2)c1
20	Ketoprofen (RP-19583) lysinate	57469-78-0	2	CC(C(=O)O)c1cccc(C(=O)c2ccccc2)c1.NCCCC[C@H](N)C(=O)O
21	Ketorolac	74103-06-3	20	O=C(c1ccccc1)c1ccc2n1CCC2C(=O)O
22	Ketorolac hemicalcium	167105-81-9	20	O=C(c1ccccc1)c1ccc2n1CCC2C(=O)O.[CaH2]

Supplementary Table 13: The anti-COX-2 drugs with an IC50 threshold of 100 from MedChemExpress website. The cas is the unique identifier for a drug. (.xlsx)

index	name	cas	IC50(nM)	canonical smiles
1	COX-2-IN-28	2413565-18-9	54	CSc1nc2ccccc2n1-c1csc(NCC(C)=NN=c2scc(-c3ccc(C)cc3)n2-c2ccccc2)n1
2	COX-2-IN-26	2413565-19-0	67	CSc1nc2ccccc2n1-c1csc(NCC(C)=NN=C2SCC(=O)N2c2ccccc2)n1
3	COX-2-IN-20	2529451-43-0	17.9	COC(=O)c1nc(CCl)n(-c2ccc(F)cc2)n1
4	COX-2-IN-21	25196	39	COc1cc(C2CC(c3cccc(-

		31-11-7		n4cnnn4)c3)=NN2C(C)=O)cc(OC)c1OC
5	COX-2/15-LOX-IN-1	2413565-15-6	75	CSc1nc2ccccc2n1-c1csc(NCC(C)=NNC(=S)Nc2ccccc2)n1
6	COX-2-IN-17	2411390-10-6	20	COC(=O)[C@@H](N)Cc1cn(-c2nc(Cl)nc(N3CCCCC3)n2)c2ccccc12
7	COX-2/5-LOX-IN-2	2410384-59-5	10	NS(=O)(=O)c1ccc(-n2nc(C(=O)O)cc2-c2cc3ccccc3s2)cc1
8	Diclofenac	15307-86-5	1.3	O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl
9	Diclofenac potassium	15307-81-0	1.3	O=C([O-])Cc1ccccc1Nc1c(Cl)cccc1Cl.[K+]
10	Diclofenac diethylamine	78213-16-8	1.3	CCNCC.O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl
11	Diclofenac Sodium	15307-79-6	1.3	O=C([O-])Cc1ccccc1Nc1c(Cl)cccc1Cl.[Na+]
12	Rofecoxib	162011-90-7	26	CS(=O)(=O)c1ccc(C2=C(c3ccccc3)C(=O)OC2)cc1
13	Valdecoxib	181695-72-7	5	Cc1onc(-c2ccccc2)c1-c1ccc(S(N)(=O)=O)cc1
14	Indomethacin (Indometacin) sodium	7681-54-1	26	COc1ccc2c(c1)c(CC(=O)[O-])c(C)n2C(=O)c1ccc(Cl)cc1.[Na+]
15	Indomethacin	53-86-1	26	COc1ccc2c(c1)c(CC(=O)O)c(C)n2C(=O)c1ccc(Cl)cc1
16	Anti-inflammatory agent 8	1103920-19-9	0.09	Cc1c(C(=O)NNC(=S)Nc2ccccc2)sc2nc3ccccc3n12
17	Indomethacin (Indometacin) sodium hydrateis	74252-25-8	26	COc1ccc2c(c1)c(CC(=O)OC/C=C(\C)CC/C=C(\C)CCC=C(C)C)c(C)n2C(=O)c1ccc(Cl)cc1
18	SC-58125	162054-19-5	40	CS(=O)(=O)c1ccc(-n2nc(C(F)(F)F)cc2-c2ccc(F)cc2)cc1
19	S-2474	158089-95-3	11	CCN1CC/C(=C\c2cc(C(C)(C)C)c(O)c(C(C)(C)C)c2)S1(=O)=O
20	FR-188582	189699-82-9	17	CS(=O)(=O)c1ccc(-c2cc(Cl)nn2-c2ccccc2)cc1
21	DuP-697	88149-94-4	10	CS(=O)(=O)c1ccc(-c2cc(Br)sc2-c2ccc(F)cc2)cc1
22	Celecoxib	169590-42-5	40	Cc1ccc(-c2cc(C(F)(F)F)nn2-c2ccc(S(N)(=O)=O)cc2)cc1
23	Indomethacin farnesil	85801-02-1	26	COc1ccc2c(c1)c(CC(=O)OCC=C(C)CCC=C(C)CCC=C(C)C)c(C)n2C(=O)c1ccc(Cl)cc1
24	Lornoxicam	70374-39-9	8	CN1C(C(=O)Nc2ccccn2)=C(O)c2sc(Cl)cc2S1(=O)=O
25	Tilmacoxib	180200-68-4	85	Cc1nc(C2CCCCC2)c(-c2ccc(S(N)(=O)=O)c(F)c2)o1
26	S-(+)-Ketoprofen	22161-81-5	27	C[C@H](C(=O)O)c1cccc(C(=O)c2ccccc2)c1

27	SC57666	158959-32-1	26	CS(=O)(=O)c1ccc(C2=C(c3ccc(F)cc3)CCC2)cc1
28	Eltenac	72895-88-6	30	O=C(O)Cc1cscc1Nc1c(Cl)cccc1Cl
29	Bromfenac	91714-94-2	7.45	Nc1c(CC(=O)O)cccc1C(=O)c1ccc(Br)cc1
30	Bromfenac sodium	91714-93-1	7.45	Nc1c(CC(=O)[O-])cccc1C(=O)c1ccc(Br)cc1.[Na+]
31	Desmethyl Celecoxib	170569-87-6	32	NS(=O)(=O)c1ccc(-n2nc(C(F)(F)F)cc2-c2ccccc2)cc1
32	Imrecoxib	395683-14-4	18	CCCN1CC(c2ccc(S(C)(=O)=O)cc2)=C(c2ccc(C)cc2)C1=O
33	Fenofibric acid	42017-89-0	48	CC(C)(Oc1ccc(C(=O)c2ccc(Cl)cc2)cc1)C(=O)O
34	Thioflosulide	158205-05-1	2.3	CS(=O)(=O)Nc1cc2c(cc1Sc1ccc(F)cc1F)C(=O)CC2
35	Ketoprofen	22071-15-4	26	CC(C(=O)O)c1cccc(C(=O)c2ccccc2)c1

Supplementary Table 14: The anti-EP4 drugs with an IC50 threshold of 30 from MedChemExpress website. The cas is the unique identifier for a drug. (.xlsx)

index	name	cas	IC50 (nM)	canonical smiles
1	EP4 receptor antagonist 1	2287259-07-6	6.1	C/C=C/c1nnn(Cc2ccc(C(F)(F)F)cc2)c1C(=O)N[C@@H](C)c1ccc(C(=O)O)cc1
2	EP4 receptor antagonist 2	1965316-82-8	7.8	CC(C)NC(=O)c1ccc2c(c1)[C@]1(CCO2)C[C@H]1C(=O)Nc1cc(C#N)ccc1CCCC(=O)O
3	BAY-1316957	1613264-40-6	15.3	CCn1c2ccc(C)cc2c2cc(-c3nc4c(C)c(C(=O)O)ccc4n3CCOC)ccc21
4	TP-16	2332972-26-4	2.1	C[C@H](NC(=O)c1c(Cc2ccc(F)cc2)sc2c1CCOC2)c1ccc(C(=O)O)cc1
5	Palupiprant	1369489-71-3	13.5	C[C@H](NC(=O)c1c(C(F)F)nn(C)c1Oc1cccc(C(F)(F)F)c1)c1ccc(C(=O)O)cc1
6	CJ-42794	847728-01-2	10	C[C@H](NC(=O)c1cc(Cl)ccc1Oc1ccc(F)cc1)c1ccc(C(=O)O)cc1
7	MK-2894	1006036-87-8	2.5	Cc1sc(C)c(C(=O)NC2(c3ccc(C(=O)O)cc3)CC2)c1Cc1ccc(C(F)(F)F)cc1
8	MK-2894 sodium salt	1006036-88-9	2.5	Cc1sc(C)c(C(=O)NC2(c3ccc(C(=O)O)cc3)CC2)c1Cc1ccc(C(F)(F)F)cc1.[NaH]

Supplementary Table 15: The virtual screening on external BACE1 valiation set using ImageMol and VideoMol. imagemol_probs and videomol_probs

represent the predicted probabilities of anti-BACE1 given by ImageMol and VideoMol, respectively. (.xlsx)

index	Drug Name	imagemol_probs	videomol_probs
1	BACE-1 inhibitor 2	0.8419358	0.9690067
2	BACE-1 inhibitor 1	0.7290269	0.9868666
3	BACE1-IN-13	0.93444705	0.9687921
4	BACE1-IN-1	0.7387245	0.9473744
5	BACE1-IN-6	0.9090035	0.9966139
6	BACE1-IN-4	0.581801	0.9993352
7	BACE1-IN-2	0.5748341	0.99737775
8	BACE1/2-IN-1	0.41264746	0.9997812
9	Umibecestat	0.7080996	0.9896259
10	LY2886721	0.80451745	0.92877966
11	beta-Secretase Inhibitor IV	0.036601424	0.16687886
12	AMG-8718	0.9256922	0.99652356
13	PF-06751979	0.31388915	0.57966435
14	JNJ-67569762	0.8896324	0.9999232
15	NB-360	0.87558085	0.9982304
16	AM-6494	0.9452854	0.99410605

Supplementary Table 16: The virtual screening on external COX-1 valiation set using ImageMol and VideoMol. imagemol_probs and videomol_probs represent the predicted probabilities of anti-COX-1 given by ImageMol and VideoMol, respectively. (.xlsx)

index	Drug Name	imagemol_probs	videomol_probs
1	Indomethacin	0.01401831	5.10E-06
2	Indomethacin sodium	0.013258358	7.33E-06
3	Indomethacin sodium hydrate	0.016143031	0.000184664
4	Diclofenac	0.01718534	0.9793444
5	Diclofenac potassium	0.03288859	0.001633527
6	Diclofenac diethylamine	0.04155041	0.862361
7	Diclofenac Sodium	0.02927339	0.001424704
8	Mofezolac	0.33693957	0.2581888
9	Tenidap	0.13598277	0.008560411
10	Tenidap-d3	0.109688826	0.00995435
11	SC-560	0.4061602	0.9985505
12	Lornoxicam	0.06643499	0.000148638
13	S-(+)-Ketoprofen	0.16169016	0.9383299
14	FR122047	0.030881904	0.59208053

15	Ketorolac tromethamine salt	0.015761768	0.18289186
16	Eltenac	0.016934287	0.021998327
17	Bromfenac	0.039675813	0.9789971
18	Bromfenac sodium	0.057841644	0.48522508
19	Ketoprofen	0.16769184	0.15350829
20	Ketoprofen (RP-19583) lysinate	0.091650955	0.07088906
21	Ketorolac	0.12183086	0.9987212
22	Ketorolac hemicalcium	0.11995495	0.9982821

Supplementary Table 17: The virtual screening on external COX-2 valiation set using ImageMol and VideoMol. imagemol_probs and videomol_probs represent the predicted probabilities of anti-COX-2 given by ImageMol and VideoMol, respectively. (.xlsx)

index	Drug Name	imagemol_probs	videomol_probs
1	COX-2-IN-28	0.71353495	0.99945635
2	COX-2-IN-26	0.2967862	0.52510965
3	COX-2-IN-20	0.000742534	0.000186789
4	COX-2-IN-21	0.008292489	1.03E-06
5	COX-2/15-LOX-IN-1	0.66978335	0.5747041
6	COX-2-IN-17	0.20885414	0.9999974
7	COX-2/5-LOX-IN-2	0.24981825	0.81430477
8	Diclofenac	0.13238531	0.057505697
9	Diclofenac potassium	0.20137852	0.53897834
10	Diclofenac diethylamine	0.016165184	0.48384362
11	Diclofenac Sodium	0.14288409	0.508286
12	Rofecoxib	0.005311308	0.45926
13	Valdecoxib	0.2744622	0.000237122
14	Indomethacin (Indometacin) sodium	0.09831848	1.81E-05
15	Indomethacin	0.1264606	2.15E-07
16	Anti-inflammatory agent 8	0.016152615	1.40E-05
17	Indomethacin (Indometacin) sodium hydrateis	0.31498575	0.00723425
18	SC-58125	0.10150733	0.000491314
19	S-2474	0.001583099	8.77E-07
20	FR-188582	0.22198991	0.008781442
21	DuP-697	0.20812099	0.0001205
22	Celecoxib	0.14900573	0.000566671
23	Indomethacin farnesil	0.29184505	0.00585067
24	Lornoxicam	0.08456622	8.87E-07
25	Tilmacoxib	0.49571124	0.4181791
26	S-(+)-Ketoprofen	0.002051125	2.42E-06

27	SC57666	0.087090984	0.99999297
28	Eltenac	0.016617168	0.010179955
29	Bromfenac	0.037134465	0.99999905
30	Bromfenac sodium	0.030660748	0.14692196
31	Desmethyl Celecoxib	0.2064141	0.88198847
32	Imrecoxib	0.14230609	0.9998147
33	Fenofibric acid	0.00863223	0.000214783
34	Thioflosulide	0.2516166	0.9999695
35	Ketoprofen	0.001724457	1.78E-06

Supplementary Table 18: The virtual screening on external EP4 valiation set using ImageMol and VideoMol. imagemol_probs and videomol_probs represent the predicted probabilities of anti-EP4 given by ImageMol and VideoMol, respectively. (.xlsx)

index	Drug Name	imagemol_probs	videomol_probs
1	EP4 receptor antagonist 1	0.09598055	0.9960939
2	EP4 receptor antagonist 2	0.14752103	0.000257138
3	BAY-1316957	0.14312284	0.9475602
4	TP-16	0.44609693	0.78389543
5	Palupiprant	0.06736643	0.007245741
6	CJ-42794	0.09034469	0.85260254
7	MK-2894	0.35040438	0.5237833
8	MK-2894 sodium salt	0.39034134	0.99885285

Supplementary Table 19: The 2,500 FDA-approved drugs from DrugBank database. We only show partial information of the first 10 items. For full information, please see the xlsx file. (.xlsx)

index	DrugBank ID	Name
1	DB15444	Elexacaftor
2	DB08949	Inositol nicotinate
3	DB14568	Ivosidenib
4	DB11700	Setmelanotide
5	DB13879	Glecaprevir
6	DB13867	Fluticasone
7	DB11363	Alectinib

8	DB11574	Elbasvir
9	DB11653	Bremelanotide
10	DB15982	Berotralstat

Supplementary Table 20: Virtual screening on 2,500 FDA-approved drugs. The probs represents the probability that the drug is a BACE1 inhibitor, which are predicted by ImageMol. Grid Score calculated by Dock6.10[13] represents the binding force between ligand and receptor. The lower the value, the stronger the binding force. The evidence represents research related to the treatment of Alzheimer's disease. We only show the first 20 items. For full information, please see the xlsx file. (.xlsx)

DrugBank ID	Name	probs	Grid Score	Related?	evidence description	evidence link
DB00582	Voriconazole	0.8937607	-31.719074	No		
DB13345	Dihydroergocristine	0.81123424	-44.298321	Yes	The FDA-approved natural product dihydroergocristine reduces the production of the Alzheimer's disease amyloid-β peptides	https://pubmed.ncbi.nlm.nih.gov/26567970/
DB13878	Pibrentasvir	0.80861706	-70.160301	Yes	A recent study reported that treatment of HCV infection with direct-acting antivirals (e.g., glecaprevir/pibrentasvir, elbasvir/grazoprevir, and ledipasvir/sofosbuvir) significantly reduces mortality risk in patients with Alzheimer's disease (AD) and related dementia	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8959984/
DB06708	Lumefantrine	0.76963395	-44.58	No		

			0559			
DB11273	Dihydroergocornine	0.76337	-42.261795	Yes	Dihydroergocornine (as the component of Ergoloid mesylates) has been used to treat dementia and age-related cognitive impairment (such as in Alzheimer disease), as well as to aid in recovery after stroke.	https://drugs.ncats.io/drug/IK4C1OC8NE
DB11274	Dihydro-alpha-ergocryptine	0.76058465	-44.593708	No		
DB08815	Lurasidone	0.72137386	-41.349476	No		
DB00320	Dihydroergotamine	0.71469796	-41.734951	No		
DB11828	Neratinib	0.71019876	-49.609707	Yes	Our data support further evaluation of AR12 and neratinib in neuronal cells as repurposed treatments for AD.	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8312464/
DB05812	Abiraterone	0.69422716	-30.7535	No		
DB11700	Setmelanotide	0.6869522	-61.032345	No		
DB11104	Sulfur hexafluoride	0.68663824	-11.747492	No		
DB11633	Isavuconazole	0.67825085	-38.032864	No		
DB09099	Somatostatin	0.66924506	nan	-		
DB15982	Berotralstat	0.6565085	-46.143105	No		
DB00106	Abarelix	0.65392923	-101.951157	No		
DB00007	Leuprolide	0.65373343	-90.477882	Yes	The LUCINDA trial: Leuprolide + cholinesterase inhibition to reduce neurologic decline in	https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.

					Alzheimer's	1002/alz.038780
DB11275	Epicriptine	0.64791554	-43.824257	No		
DB06791	Lanreotide	0.63942754	-62.846352	No		
DB00921	Buprenorphine	0.6381763	-35.94207	No		

Supplementary Table 21: Virtual screening on 2,500 FDA-approved drugs. The probs represents the probability that the drug is a BACE1 inhibitor, which are predicted by VideoMol. Grid Score calculated by Dock6.10 represents the binding force between ligand and receptor. The lower the value, the stronger the binding force. The evidence represents research related to the treatment of Alzheimer's disease. We only show the first 20 items. For full information, please see the xlsx file. (.xlsx)

DrugBank ID	Name	probs	Grid Score	Related?	evidence description	evidence link
DB15444	Elexacaftor	0.99302226	-46.733532	No		
DB08949	Inositol nicotinate	0.9899024	-55.276123	Yes	As a result, inositol nicotinate was predicted to act on at least one anti-AD drug target yet act against AD through various mechanisms	https://pesquisa.bvsalud.org/portal/resource/pt/wpr-780222
DB14568	Ivosidenib	0.9597691	-42.924538	No		
DB11700	Setmelanotide	0.95926934	-61.032345	No		
DB13879	Glecaprevir	0.95918256	-45.78	Yes	glecaprevir/pibrentasvir,	https://www.ncbi.

			6842		elbasvir/grazoprevir, and ledipasvir/sofosbuvir) significantly reduces mortality risk in patients with Alzheimer's disease (AD) and related dementia	nlm.nih.gov/pmc/articles/PMC8959984/
DB13867	Fluticasone	0.94764495	-35.311394	Yes	The lower incidence of Alzheimer's dementia in the intranasal fluticasone propionate (Flonase) group compared to the Lipitor group was significant	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6110392/
DB11363	Alectinib	0.9432152	-37.98465	Yes	downregulation of macroautophagy (autophagy) plays an essential role in AD pathogenesis. Therefore, targeting autophagy has drawn considerable attention as a therapeutic approach for the treatment of AD	https://www.sciencedirect.com/science/article/pii/S0163725822000651
DB11574	Elbasvir	0.92173696	-65.377815	Yes	See Table S22	
DB11653	Bremelanotide	0.92082703	-63.477474	No		
DB15982	Berotralstat	0.91874415	-46.143105	No		
DB06636	Isavuconazonium	0.9137141	-68.284996	No		
DB06663	Pasireotide	0.9129874	-63.459614	No		
DB00115	Cyanocobalamin	0.9082298	-75.670044	Yes	See Table S22	
DB11828	Neratinib	0.9022543	-49.609707	Yes	Neratinib as a Potential Therapeutic for	https://www.emjreviews.co

					Mutant RAS and Osimertinib-Resistant Tumours	m/flagship-journal/article/neratinib-as-a-potential-therapeutic-for-mutant-ras-and-osimertinib-resistant-tumours-j190322/
DB11842	Angiotensin II	0.88213295	-84.811951	Yes	See Table S22	
DB12457	Rimegepant	0.8742169	-41.61306	No		
DB00644	Gonadorelin	0.86522996	-93.958443	Yes	See Table S22	
DB15494	Edotreotide gallium Ga-68	0.8631429	-78.598076	Yes	See Table S22	
DB00520	Caspofungin	0.8616219	-56.143566	Yes	In this study, we discovered that an antifungal drug, Caspofungin (CAS) is a potent Aβ aggregation inhibitor that displays significantly reduced toxicity associated with AD.	https://pubmed.ncbi.nlm.nih.gov/33722677/
DB09065	Cobicistat	0.8536151	-67.501389	No		

Supplementary Table 22: Screening results of Top 6 approved drugs with the best grid score from DrugBank. DrugID corresponds to the drug id in DrugBank. Grid score represents binding ability between ligand and BACE1-4IVS receptor calculated by Dock6.10 and the smaller value denotes the

better score. The evidence represents research related to the treatment of Alzheimer's disease. (.xlsx)

No.	DrugID	Name	Grid Score	evidence description	evidence link
1	DB00644	Gonadorelin	-93.958443	Drugs used in intervention or treatment: Drug: GnRH, gonadorelin acetate	https://clinicaltrials.gov/ct2/show/NCT04390646
2	DB11842	Angiotensin II	-84.811951	Recently, emerging preclinical and clinical evidence has associated the brain renin angiotensin system (RAS) to AD pathology. Accumulating evidence has additionally identified antihypertensive medications targeting the RAS, particularly angiotensin receptor blockers (ARBs) and angiotensin-converting enzyme inhibitors (ACEIs), to delay AD onset and progression.	https://pubmed.ncbi.nlm.nih.gov/32700176/
3	DB15494	Edotreotide gallium Ga-68	-78.598076	Imaging tools of β-amyloid (Aβ) plaques are necessary for clinical and neuropsychological characteristics in AD. Gallium-68 is a promising radionuclide for in vivo imaging of β-amyloid plaques because it is easily produced by a generator.	https://openmedscience.com/gallium-68-radiotracers-for-alzheimers-plaque-imaging/
4	DB00115	Cyanocobalamin	-75.670044	Here we demonstrate that BACE (β-secretase), as well as PS1, is regulated by methylation and that the reduction of folate and vitamin B12 in culture medium can cause a reduction of SAM levels with consequent increase in presenilin1 and BACE levels and with increase in Aβproduction.	https://sciencedirect.com/science/article/abs/pii/S1044743104002209
5	DB06636	Isavuconazonium	-68.284996	-	-
6	DB11574	Elbasvir	-65.377815	Several studies have reported that patients with chronic hepatitis C virus (HCV) infection tend to exhibit cognitive impairment and may increase the risk for dementia. A recent study reported that treatment of HCV infection with direct-acting antivirals (e.g., glecaprevir/pibrentasvir,	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8959984/

				elbasvir/grazoprevir, and ledipasvir/sofosbuvir) significantly reduces mortality risk in patients with Alzheimer's disease (AD) and related dementia.	

Supplementary Table 23: Evidence of key structures belonging to BACE-1 inhibitors. (.xlsx)

No.	key structure	evidence title	evidence description	evidence link
1	Fluorine	Development and Structural Modification of BACE1 Inhibitors	In addition, the introduction of a fluorine atom in the side chain on the phenyl could improve BACE1 inhibitory activities.	https://www.mdpi.com/1420-3049/22/1/4
2	1,2,4-Oxadiazole	Design and synthesis of potent β-secretase (BACE1) inhibitors with P1' carboxylic acid bioisosteres	On the other hand, inhibitors 16–21, which contained one or two tetrazole rings, or acidic heterocycles (5-oxo-1,2,4-oxadiazole, 5-oxo-1,2,4-thiadiazole, and 2-thioxo-1,3,4-oxadiazole)18, 19 at P_1^' position, respectively, showed significantly higher BACE1 inhibitory activities than their lead compound 1.	https://www.sciencedirect.com/science/article/pii/S0960894X06001636
3	Chromene	BACE-1 inhibitory activities of new substituted phenyl-piperazine coupled to various heterocycles:Chromene, coumarin and quinoline	There is some latitude with regard to functional group identity (coumarin, chromene and quinoline) that allows further refinement of inhibitory potency relative to b-secretase (BACE-1).	https://www.sciencedirect.com/science/article/pii/S0960894X05016331?via%3Dihub
4	Pyridine	Aminoimidazoles as Potent and Selective Human β-Secretase (BACE1) Inhibitors	The addition of either a pyridine or a pyrimidine group on the scaffold of 3 in a manner such that the group extends and occupies the S3 region of the BACE1 binding pocket, not previously accessed with 3, has	https://pubs.acs.org/doi/full/10.1021/jm9006752

			significantly contributed to the ligand's potency.	
5	Cyclopentane	Stereoselective Synthesis of Constrained Oxacyclic Hydroxyethylene Isosteres of Aspartic Protease Inhibitors: Aldol and Mukaiyama Aldol Methodologies for Branched Tetrahydrofuran 2-Carboxylic Acids	The carbocyclic analogue of 21c in which the tetrahydrofuran ring is replaced by a cyclopentane and a cyclopentanone are low nanomolar inhibitors of BACE1. The substantial loss of activity when a cyclopentane is replaced by an oxacyclic analogue might be the result of repulsive interactions of the ring oxygen atom with amide carbonyls in the active site, thus deviating from optimal conformations for binding.	https://pubs.acs.org/doi/full/10.1021/jo050749y
6	Tetrazole	Design and synthesis of potent β-secretase (BACE1) inhibitors with P1′ carboxylic acid bioisosteres	Among them, tetrazole ring-containing compounds, KMI-570 (IC50 = 4.8 nM) and KMI-684 (IC50 = 1.2 nM), exhibited significantly potent BACE1 inhibitory activities.	https://www.sciencedirect.com/science/article/pii/S0960894X06001636

Supplementary Table 24: Effect of pre-training strategy on 6 regression datasets (GPCRs) with balanced scaffold split. w/o pretrain means no pre-trained VideoMol. video-aware, direction-aware, and chemical-aware represent pre-training VideoMol using only video-aware strategy, direction-aware strategy, and chemical-aware strategy, respectively. & represents the combination of multiple pre-training tasks. All means and standard deviations are reported through three independent runs with random seeds of 0, 1, 2.

	5HT1A		AA1R		AA2AR	
strategy	RMSE	MAE	RMSE	MAE	RMSE	MAE
w/o pretrain	0.993±0.001	0.804±0.002	0.919±0.006	0.755±0.008	1.073±0.01	0.882±0.013
video-aware	0.772±0.011	0.604±0.009	0.709±0.007	0.783±0.010	0.847±0.002	0.651±0.002

direction-aware	0.871±0.002	0.691±0.004	0.821±0.01	0.652±0.013	0.875±0.008	0.701±0.012
chemical-aware	0.736±0.012	0.566±0.014	0.662±0.018	**0.494±0.015**	0.716±0.001	0.562±0.002
chemical_direction	0.718±0.002	0.559±0.004	0.706±0.005	0.516±0.008	0.724±0.008	0.568±0.003
chemical_video	0.716±0.010	0.553±0.010	0.670±0.007	0.511±0.005	0.719±0.013	0.553±0.014
direction_video	0.774±0.014	0.603±0.015	0.730±0.004	0.553±0.002	0.865±0.011	0.672±0.005
VideoMol	**0.708±0.017**	**0.547±0.015**	**0.655±0.007**	0.496±0.006	**0.712±0.011**	**0.543±0.005**

	CNR2		DRD2		HRH3	
strategy	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
w/o pretrain	1.216±0.012	1.009±0.008	0.980±0.001	0.782±0.001	0.819±0.001	0.631±0.002
video-aware	0.978±0.025	0.782±0.010	0.818±0.009	0.602±0.009	0.732±0.008	0.556±0.009
direction-aware	1.024±0.01	0.842±0.008	0.903±0.010	0.700±0.009	0.759±0.005	0.575±0.007
chemical-aware	0.890±0.004	0.698±0.002	0.759±0.003	0.565±0.003	0.669±0.010	0.512±0.010
chemical_direction	0.899±0.008	0.701±0.005	0.773±0.017	0.579±0.007	0.683±0.001	0.514±0.001
chemical_video	0.874±0.012	0.686±0.013	0.745±0.012	**0.555±0.013**	0.686±0.003	0.526±0.004
direction_video	0.997±0.025	0.789±0.021	0.832±0.005	0.615±0.007	0.734±0.006	0.559±0.002
VideoMol	**0.864±0.005**	**0.679±0.010**	**0.742±0.003**	0.556±0.005	**0.668±0.008**	**0.506±0.002**

Supplementary Table 25: The performance of molecular fingerprint on 10 GPCR datasets with balanced scaffold split. EnsembleFP-MLP indicates the performance of integrating traditional molecular fingerprints in CAP and training an MLP. VideoMolFeat-MLP represents the performance of using VideoMol to extract molecular features and train an MLP. (.xlsx)

	1. 5HT1A		2. 5HT2A		3. AA1R	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
EnsembleFP-MLP	1.030±0.000	0.837±0.001	1.207±0.005	0.975±0.006	0.915±0.005	0.734±0.005
VideoMolFeat-MLP	**0.818±0.031**	**0.653±0.033**	**0.910±0.014**	**0.695±0.017**	**0.838±0.007**	**0.647±0.005**
—— continue ——						
	4. AA2AR		5. AA3R		6. CNR2	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
EnsembleFP-MLP	1.076±0.003	0.871±0.002	0.991±0.001	0.791±0.002	1.228±0.004	1.027±0.002
VideoMolFeat-MLP	**0.850±0.005**	**0.691±0.004**	**0.862±0.014**	**0.695±0.012**	**1.003±0.070**	**0.789±0.066**
—— continue ——						
	7. DRD2		8. DRD3		9. HRH3	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
EnsembleFP-MLP	0.979±0.000	0.785±0.001	1.096±0.004	0.933±0.002	0.888±0.003	0.681±0.002
VideoMolFeat-MLP	**0.885±0.003**	**0.660±0.002**	**0.837±0.012**	**0.663±0.011**	**0.741±0.001**	**0.578±0.002**
—— continue ——						
	10. OPRM					
	RMSE	MAE				
EnsembleFP-MLP	1.023±0.006	0.815±0.002				
VideoMolFeat-MLP	**0.875±0.004**	**0.672±0.003**				

Supplementary Table 26: Effect of frame number on VideoMol on 6 regression datasets (GPCRs) with balanced scaffold split. #frame indicates the number of frames. All means and standard deviations are reported through three independent runs with random seeds of 0, 1, 2.

	5HT1A		**AA1R**		**AA2AR**	
#frame	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
5	0.873±0.023	0.693±0.012	0.847±0.013	0.656±0.025	0.824±0.009	0.659±0.012
10	0.800±0.025	0.624±0.017	0.773±0.010	0.584±0.006	0.766±0.002	0.609±0.005
20	0.765±0.026	0.591±0.010	0.728±0.003	0.546±0.005	0.744±0.005	0.586±0.005
30	0.742±0.014	0.573±0.011	0.704±0.004	0.527±0.001	0.736±0.019	0.570±0.013
60	**0.708±0.017**	**0.547±0.015**	**0.655±0.007**	**0.496±0.006**	**0.712±0.011**	**0.543±0.005**
	CNR2		**DRD2**		**HRH3**	
#frame	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
5	0.991±0.032	0.811±0.027	0.875±0.016	0.660±0.012	0.733±0.017	0.567±0.009
10	0.950±0.032	0.762±0.039	0.822±0.014	0.620±0.003	0.696±0.013	0.540±0.014
20	0.910±0.016	0.721±0.011	0.792±0.001	0.600±0.004	0.679±0.019	0.521±0.014
30	0.890±0.013	0.700±0.009	0.769±0.005	0.580±0.003	0.686±0.019	0.531±0.015
60	**0.864±0.005**	**0.679±0.010**	**0.742±0.003**	**0.556±0.005**	**0.668±0.008**	**0.506±0.002**

Supplementary Table 27: The performance of different video generation source on 10 GPCR datasets with balanced scaffold split. (.xlsx)

	1. 5HT1A		2. 5HT2A		3. AA1R	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Openbabel	0.733±0.009	0.562±0.006	0.796±0.004	0.597±0.003	0.666±0.014	0.510±0.011
DeepChem	0.718±0.006	0.557±0.009	0.790±0.016	0.595±0.019	0.677±0.014	0.504±0.010
RDKit	**0.708±0.017**	**0.547±0.015**	**0.775±0.017**	**0.578±0.009**	**0.655±0.007**	**0.496±0.006**
—— continue ——						
	4. AA2AR		5. AA3R		6. CNR2	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Openbabel	0.705±0.004	0.548±0.008	**0.779±0.015**	0.618±0.009	0.903±0.018	0.714±0.012
DeepChem	**0.703±0.005**	0.558±0.004	0.798±0.019	0.620±0.015	0.896±0.013	0.708±0.012
RDKit	0.712±0.011	**0.543±0.005**	0.786±0.006	**0.617±0.004**	**0.864±0.005**	**0.679±0.010**
—— continue ——						
	7. DRD2		8. DRD3		9. HRH3	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Openbabel	0.763±0.018	0.562±0.009	0.745±0.005	0.580±0.003	0.679±0.004	0.524±0.004
DeepChem	0.745±0.005	**0.551±0.002**	0.728±0.014	0.556±0.008	**0.666±0.005**	**0.505±0.002**
RDKit	**0.742±0.003**	0.556±0.005	**0.715±0.014**	**0.554±0.012**	0.668±0.008	0.506±0.002

—— continue ——					
	10. OPRM		Mean		
	RMSE	MAE	RMSE	MAE	
Openbabel	**0.776±0.008**	0.590±0.001	0.755±0.068	0.581±0.057	
DeepChem	0.825±0.008	0.607±0.009	0.755±0.072	0.576±0.060	
RDKit	0.795±0.015	**0.579±0.011**	**0.742±0.064**	**0.565±0.053**	

Supplementary Table 28: The ability of VideoMol to distinguish different conformers. The percentile interval refers to sorting all RMSD values from small to large and selecting the value corresponding to the percentile interval. (.xlsx)

percentile interval in RMSD	5HT1A	5HT2A	AA1R	AA2AR	AA3R	CNR2	DRD2	DRD3	HRH3	OPRM
0-10	0.692	0.749	0.799	0.803	0.823	0.769	0.729	0.701	0.740	0.730
10-20	0.724	0.726	0.786	0.784	0.771	0.747	0.733	0.740	0.721	0.740
20-30	0.741	0.747	0.755	0.764	0.772	0.736	0.750	0.779	0.690	0.708
30-40	0.726	0.733	0.749	0.747	0.752	0.735	0.748	0.795	0.703	0.716
40-50	0.735	0.745	0.759	0.772	0.758	0.723	0.760	0.775	0.709	0.712
50-60	0.703	0.744	0.765	0.761	0.760	0.710	0.743	0.785	0.706	0.725
60-70	0.708	0.733	0.742	0.756	0.745	0.696	0.748	0.777	0.698	0.727
70-80	0.663	0.748	0.753	0.743	0.766	0.712	0.715	0.775	0.700	0.716
80-90	0.631	0.702	0.745	0.753	0.752	0.709	0.707	0.723	0.669	0.680
90-100	0.518	0.625	0.654	0.702	0.733	0.680	0.615	0.631	0.585	0.658
0-100 (all data)	0.684	0.725	0.751	0.758	0.763	0.722	0.725	0.748	0.692	0.711

Supplementary Table 29: Details of the molecular fingerprints used in chemical-aware pretraining task.

No.	Type of fingerprint	Name	Dimension
1	Circular-based	ECFP0	1024
2		ECFP2	1024
3		ECFP4	1024
4		ECFP6	1024
5		FCFP2	1024
6		FCFP4	1024
7		FCFP6	1024
8	Path-based	RDK5	1024
9		RDK6	1024

10		RDK7	1024
11		HashAP	1024
12		HashTT	1024
13	Substructure-based	MACCS	167
14		Avalon	1024
15	Longer version-based	LECFP4	16384
16		LECFP6	16384
17		LFCFP4	16384
18		LFCFP6	16384
19		LAvalon	16384
20	Pharmacophore-based	TPATF	2692
21	Physicochemistry-based	RDKDes	208

Supplementary Table 30: The parameter details in the pre-training phase.

Model details	
Backbone of VideoMol	12-layer Vision Transformer with 16 patches
Axis Classifier	2 layers of fully connected neural network, where the first layer has a Softplus activation function and the number of output neurons in the last layer is task-dependent
Rotation Classifier	
Angle Classifier	
Chemical Classifier	
Training parameters	
Temperature	0.1
Learning Rate	0.01
Batch Size	256
Momentum	0.9
Weight Decay	1.00E-04
Learning Rate Decay	Linear
Video Size	60×3×224×224
Training Step	>400k
Validation split	random split with 90% training set and 10% validation set
Training Platform	
CPU	Intel 6248R 48C@3.0GHz
GPU	4 * NVIDIA TESLA A100 (40G)

Supplementary Table 31: Basic statistical information of 10 kinases and 10 GPCRs. #Molecules represents the number of molecules. #Tasks represents the number of binary prediction task. Target represents the type of target, such as kinases and GPCRs.

Dataset	#Molecules	#Tasks	Target	Metric	Type
BTK	106	1	Kinases	ROC-AUC	Classification
CDK4-cyclinD3	80				
EGFR	105				
FGFR1	108				
FGFR2	109				
FGFR3	107				
FGFR4	107				
FLT3	110				
KPCD3	109				
MET	105				
AA1R	3408	1	GPCRs	RMSE and MAE	Regression
5HT1A	3568				
5HT2A	3079				
AA2AR	3866				
AA3R	3306				
CNR2	3079				
DRD2	5771				

DRD3	3945				
HRH3	3206				
OPRM	2977				

Supplementary Table 32: Basic statistical information of MoleculeNet benchmark datasets. #Molecules represents the number of molecules. #Tasks represent the number of binary prediction task.

Dataset	#Molecules	#Tasks	Property	Metric	Type
BBBP	2039	1	Pharmacology	ROC-AUC	Classification
Tox21	7831	12	Pharmacology		
SIDER	1427	27	Pharmacology		
HIV	41127	1	Biophysics		
BACE	1513	1			
ToxCast	8575	617	Pharmacology		
FreeSolv	642	1	Physical chemistry	RMSE	Regression
ESOL	1128	1			
Lipo	4200	1			
QM7	6830	1	Quantum chemistry	MAE	
QM8	21786	12			
QM9	133885	8			

Supplementary Table 33: Basic statistical information of 11 SARS-CoV-2 datasets with classification task. #Train, #Val and #Test represent the

numbers of train set, valid set, and test set, respectively. #Effects indicates how the drug influences SARS-CoV-2. #Tasks represents the number of binary prediction task.

Dataset	#Train	#Val	#Test	#Effects	#Tasks	Metric
3CL	404	87	87	Viral replication	1	ROC-AUC
hCYTOX	530	114	114	Human cell toxicity		
MERS-PPE	674	145	145	In vitro infectivity		
MERS-PPE_cs	250	54	54			
CoV1-PPE	1226	263	263			
CoV1-PPE_cs	216	47	47			
CPE	691	148	149	Live virus infectivity		
Cytotox	1569	336	337			
ACE2	228	49	49	Viral entry		
AlphaLISA	1089	233	234			
TruHit	1175	251	252			

Supplementary Table 34: The general framework of VideoMol for fine-tuning. Identity layer represents identity mapping.

name	description
base model	pretrained video encoder
dropout1	Dropout with a dropout ratio of 0 to 1
linear1	Linear layer or Identity layer
activator	Activator
dropout2	Dropout with a dropout ratio of 0 to 1
linear2	Linear layer

Supplementary Table 35: The hyperparameters in fine-tuning of VideoMol.

Hyperparameter	Molecular property prediction	Binding activity prediction
dropout1	0	0
linear1	Linear layer	[Identify layer, Linear layer]
activator	[Gelu, Softplus, None]	None
dropout2	[0.3, 0.5, 0.8]	[0, 0.2]
linear2	[Identify layer, Linear layer]	[Identify layer, Linear layer]
Learning rate	[8e-5, 3e-4, 5e-4, 8e-4, 3e-3, 5e-3, 0.03, 0.05]	[1e-4, 5e-4, 8e-4, 1e-3, 3e-3, 5e-3, 8e-3, 1e-2, 3e-2, 5e-2, 0.1]
Batch size	[8, 16]	[8, 16, 32, 64, 128]
Epochs	[2, 5, 10, 30]	10~60
Warmup ratio	0	0
Seed	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]	[0, 1, 2]

Supplementary Table 36: The computational requirements in the pre-training, fine-tuning, and screening stages. #frame/batch represents the number of frames in a batch. #samples represents the total number of molecules. #frame/video indicates how many frames of a video to select for inference.

(.xlsx)

(a) The computational requirements in the pre-training stage.				
#samples	#frame/batch	GPU memory	Training time	Server
2 million	256	~37G	~9 hours/epoch	CPU: Intel 6248R 48C@3.0GHz; GPU: A100 (40G)
(b) The computational requirements in the fine-tuning stage.				
#samples	#frame/batch	GPU memory	Training time	Server
10,000	8	2.3G	~26 minutes/epoch	CPU: 13th Gen Intel® Core™ i7-13700K GPU: 4090 Ti
	16	2.6G	~15 minutes/epoch	
	32	3.2G	~12 minutes/epoch	
	64	4.3G	~12 minutes/epoc	

			h		
	128	6.5G	~12 minutes/epoch		
	256	10.7G	~12 minutes/epoch		
(c) The computational requirements in the screening stage.					
#samples	#frame/video	#videos/batch	inference time	GPU memory	Server
1 million	1	480	~9 minutes	17.7 G	CPU: 13th Gen Intel® Core™ i7-13700K GPU: 4090 Ti
	5	96	~48 minutes		
	10	48	~90 minutes		
	20	24	~3 hours		
	30	16	~4.5 hours		
	60	8	~9 hours		

Supplementary References

1. Nakata, M. & Shimazaki, T. PubChemQC project: a large-scale first-principles electronic structure database for data-driven chemistry. *J. Chem. Informat. Model.* **57**, 1300-1308 (2017).
2. Zeng, X. et al. Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat. Mach. Intell.* **4**, 1004-1016 (2022).
3. Wu, Z. et al. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **9**, 513-530 (2018).
4. Bradski, G. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* **25**, 120-123 (2000).
5. Bocci, G. et al. A machine learning platform to estimate anti-SARS-CoV-2 activities. *Nat. Mach. Intell.*, 1-9 (2021).
6. Gaulton, A. et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945-D954 (2017).
7. González, A. et al. Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum. *Nature*, 1-8 (2023).
8. Guo, J. et al. A tripartite rheostat controls self-regulated host plant resistance to insects. *Nature*, 1-9 (2023).
9. Fidelle, M. et al. A microbiota-modulated checkpoint directs immunosuppressive intestinal T cells into cancers. *Science* **380**, eabo2296 (2023).
10. Wishart, D.S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074-D1082 (2018).
11. Sud, M. MayaChemTools: an open source package for computational drug discovery. *J. Chem. Informat. Model.* **56**, 2292-2297 (2016).
12. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1-3 (2024).
13. Balius, T.E., Tan, Y.S. & Chakrabarti, M. DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking. *J. Comput. Chem.* **45**, 47-63 (2024).