# nature portfolio

### Peer Review File

### A Molecular Video-derived Foundation Model for Scientific Drug Discovery



**Open Access** This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to

the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <u>http://creativecommons.org/licenses/by/4.0/</u>.

### **REVIEWER COMMENTS**

#### Reviewer #2 (Remarks to the Author):

#### Summary:

The authors propose a molecular video-based foundation model, VideoMol, targeted at representation learning for 3D conformers downstream tasks. To learn useful representations of conformers, VideoMol is trained with three self-supervised learning strategies: Direction-Aware Pretraining Video-Aware Pretraining and Chemical-Aware pretraining, Then subsequently fine-tuned on different prediction tasks. An advantage the proposed method over similar deep learning approaches, is the provided interpretability through denoting key chemical substructures related to 3D conformational changes that according to the manuscript overlap with previous domain knowledge. The experiments show promising results of VideoMol in diverse drug discovery datasets for predicting molecular targets and properties.

The manuscript is well written and self contained. The authors made the effort to include diverse set of baselines including foundation models learnt on sequence, structure and image which gives a nice comparison across the current landscape of available methods. All experiments and results are described in detail and highlight multiple advantages of video-based features compared to current SOTA. It is encouraging that the inhibitors suggested by VideoMol overlap, to some extent, to results from other published domain studies.

#### Questions:

1. As the authors mentioned themselves, using videos increases complexity of conformer prediction models, which is an already challenging setup because of the 3D structure learning. Are the 60 frames the minimal or optimal number of frames to be used? Has this been explored?

2. Could the authors elaborate on how helpful will the features learned by VideoMol without fine tuning for a new dataset? Would it be beneficial for practilioners to use VideoMol as pretrained models to extract features and then use those for downstream tasks (without fine tuning)?

3. Why is there no standard error/deviation for any of the results? Have the experiments been repeated multiple times? Having some indication about the uncertainty intervals around the results would be appreciated.

4. It will help if the authors include some intuition on why videos make more sense and contribute to learning better features. It is not so easy to understands for readers that are less experienced with conformers. Some running/motivating example could help the presentation.

5. How sensitive is the framework to the source for video generation, in this case RDKit. I assume it is quite dependent on this platform, and thus, VideoMol probably does not allow for mixing videos from different sources. Does this pose any kind of limitation in real-world applications?

#### Reviewer #2 (Remarks on code availability):

The provided code is clean and well organised. The authors also included a docker image for setting up the environment. The code for training the model as well as reproducing the results is included. I believe only the code for VideoMol is provided, and not for the baseline methods. Not necessary, but if it is easy to include the related methods, the community might appreciate a full testbed.

### Reviewer #3 (Remarks to the Author):

#### Noteworthy results:

The manuscript presents a video-based pretrained model that can be used to make downstream task predictions across multiple tasks with finetuning. The results are an improvement upon authors' previous work, ImageMol. The improvement is tied to use of video, which can be considered as an augmentation to static images, as well as the use of more comprehensive fingerprints that provide chemical, pharmacological and physicochemistry information. The authors show, through extensive testing, that the new model outperforms the previous and is at least as good or as better as some SOTA models for different tasks.

### Impact to field:

The work is valuable to the field in multiple areas: it is a demonstration of technology transfer from video representation learning. It shows new self supervision tasks that are meaningful for molecule structure videos. It identifies a large set of benchmark cases. However I see major drawbacks or open questions that would limit its use beyond limited academic interest:

- Unlike stated, the model does not capture a dynamic conformation of the molecule. The videos are not generated to represent any physical dynamics, or conformer change, or changes to torsion angles etc. They are movies with standardized rotations around given axis. As such, they are only augmentations to enrich the model input about the 3d structure of the molecule. Authors should consider another wording than dynamics to prevent misleading the reader.

- if the manuscript's main aim was to inject more information about the 3d nature of the molecules, they could have considered an equivariant graph neural network or transformer. An equivariant neural network would remove the need to perform augmentation for different rotations.

- This is where it gets interesting: if the success of the model was truly due to better representation of 3d structure, we would expect the model to be sensitive to different conformers, especially on tasks that provide a binding affinity proxy. While, in multiple places in the manuscript the opposite is claimed, that model is robust to molecule conformer choice. Perhaps authors can devise an experiment to understand why video information does not lead to sensitivity to conformer.

-Which makes me think the success of the model is not due to better 3d representation but one of the several other changes: 1-working with video frames and the new self-supervision tasks have

expanded the effective size of the data and complexity the network processes each molecule (perhaps this is why the number of frames seem to change the prediction accuracy) 2-the large number of domain information that is crafted into the fingerprint may be impactful in several tasks in this work. Further understanding from where exactly the accuracy improvement comes from, can be considered for future work. In the teamtime, the claims of impact of 3D could be dialed down.

-Feedback to the methodology: the splits used in this work would not stop data leaking from train to validation sets and scaffold balancing might not be enough. Indeed we see a hint of the issue in the COX examples where training data from ChEMBL in 8:1:1 split gave high ROC-AU >0.9, but when the model was tested against MedChemExpress data, only less than 40% of inhibitors are successfully identified. This difference may be due to data leak in the high ROC-AUC train-test data, inflating the apparent generalizability. In general if authors would like to claim generalizability, more attention to the split stragety, overlap between data points is needed according to certain similarity metric will be needed.

-Some of the tasks in the work are for high-throughout applications (e.g. virtual screening). In such cases the trade-off between accuracy and compute becomes important. The proposed method should clearly state the compute needs for pretraining and various downstream tasks. Because it works with video, compared to much smaller atomic position files, memory needs should be highlighted too.

-minor typos in text, highlighting one that is on figure in case it escapes proofreading: angel -> angle Fig 1b

#### Reviewer #3 (Remarks on code availability):

Lightly reviewed code. Checked the pretraining tools and the base encoder model definition. Looks rather standard, didnt see any weird libraries or so. Didn't run but I didnt see a reason why it wouldnt. Also there are links to pretraining data and to pretrained model to reproduce inference results in the paper.

### **Point-by-Point Response Letter**

### Manuscript #: NCOMMS-23-62188

We are grateful to the reviewers for their insightful and constructive feedback on our manuscript. In response to the feedback, we provide the detailed responses to address each reviewer's concerns point by point as follows.

### **Responses to the Reviewer #1**

**Overall Summary** – "The authors made the effort to include diverse set of baselines, a **nice comparison** across the current landscape of available methods. All experiments and results are described in detail and highlight **multiple advantages** of video-based features. The manuscript is **well written**" –

Reviewer Comment	The authors propose a molecular video-based foundation model, VideoMol, targeted at representation learning for 3D conformers downstream tasks. To learn useful representations of conformers, VideoMol is trained with three self-supervised learning strategies: Direction-Aware Pretraining Video-Aware Pretraining and Chemical-Aware pretraining, Then subsequently fine-tuned on different prediction tasks. An advantage the proposed method over similar deep learning approaches, is the provided interpretability through denoting key chemical substructures related to 3D conformational changes that according to the manuscript overlap with previous domain knowledge. The experiments show promising results of VideoMol in diverse drug discovery datasets for predicting molecular targets and properties.
	The manuscript is well written and self contained. The authors made the effort to include diverse set of baselines including foundation models learnt on sequence, structure and image which gives a nice comparison across the current landscape of available methods. All experiments and results are described in detail and highlight multiple advantages of video-based features compared to current SOTA. It is encouraging that the inhibitors suggested by VideoMol overlap, to some extent, to results from other published domain studies.
Author Response	We thank the Reviewer for comprehensive summary and his/her positive support on the manuscript. We have made extensive revision to address the reviewer's critiques as below.

## Ref 1.1 – "Are the 60 frames the minimal or optimal number of frames to be used? Has this been explored?" –

Reviewer Comment	As the a conform of the 3I of frame	As the authors mentioned themselves, using videos increases complexity of conformer prediction models, which is an already challenging setup because of the 3D structure learning. Are the 60 frames the minimal or optimal number of frames to be used? Has this been explored?					
Author Response	frames. We used the 60 frames after balancing both the optimized number of frames. We used the 60 frames after balancing both the optimized mode performance and the computational cost. We explored the impact of frames number on the model performance in <b>Supplementary Table 26</b> , including 5 frames, 10 frames, 20 frames, 30 frames and 60 frames. We found that the increased number of frames improve the performance of the models. After balancing computing time and the optimized model performance by the increased number of frames, we selected 60 frames. We have added these new results and more detailed explanations in the revised manuscript. Supplementary Table 26: Effect of frame number on VideoMol on 6 regression datasets with balanced scaffold split. #frame indicates the number of frames. All means and standard deviations are reported through three independent runs.					I number of nized model act of frame including 5 und that the odels. After nce by the added these cript. oMol on 6 the number rough three	
		5H	Г1А	AA	1R	AA	2AR
	#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
		CN	IR2	DR	RD2	HR	RH3
	#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	$0.910\pm0.010$	0.721±0.011	0.792±0.001	0.600±0.004	0.679±0.019	$0.521\pm0.014$
		0.890±0.013	0.700±0.009	0.709±0.003	0.560±0.005	0.000±0.019	0.531±0.015
		0.004±0.000	0.079±0.010	U.14210.004	0.000±0.000	0.000±0.000	0.00010.002
Excerpt from Revised Manuscript	The impact of the video frame number. To explore the impact of different frame numbers on VideoMol, we sampled 5, 10, 20, 30, and 60 molecular frames from 5HT1A, AA1R, AA2AR, CNR2, DRD2, and HRH3 datasets at equal time intervals. We found that the performance of VideoMol is positively correlated with the number of frames with an average performance improvement of 6.5% (5→10 frames), 3.9% (10→20 frames), 2.0% (20→30 frames), 3.9% (30→60 frames) on RMSE metric and 7.6% (5→10 frames), 4.7% (10→20 frames), 2.4% (20→30 frames), 4.4% (30→60 frames) on MAE metric, which shows that the increase of frame number enriches the 3D information extracted by VideoMol and its performance may be expected to be further increased by expanding the frame number (Supplementary Table 26).						

## Ref 1.2 – "Effectiveness of features learned by VideoMol without fine-tuning on new datasets" –

Reviewer Comment	Could the authors elaborate on how helpful will the features learned by VideoMol without fine tuning for a new dataset? Would it be beneficial for practitioners to use VideoMol as pretrained models to extract features and then use those for downstream tasks (without fine tuning)?						
Author Response	We thank the r and reported n pretrained Vide also evaluated (called Ensemt input VideoMol perceptron (M achieved the bu 19.4% averag kinases datase features extract traditional mole	reviewer for ew results eoMol (call features of oleFP). To lFeat and LP). New est perform e MAE in ets. These octed by V ecular finge	or this great on downs ed Videolv extracted l be fair, w Ensembled experime nance with nprovement new find videoMol a erprinting.	at point. W stream tas loIFeat) in by ensemb e did not f FP into a s ntal result 17.2% ave nt compar ings (without are superi	Ve conduct ks using for <b>Suppleme</b> ole fingerp ine-tune V structurally ts showed erage RMS red with 1 out fine tu or alterna	eed new executives extended new executives executives executive for contract of the second executive compositive compositive compositive compositive represents the per	speriments tracted by <b>ble 25</b> . We omparison nd directly multi-layer eoMolFeat ement and FP on 10 w that the bared with
		1 51		0 EL	ACTL	2 4	A1D
		DMCE		Z. JI		J.A	
	E	RIVISE	IVIAE	RIVISE	IVIAE	RIVISE	1VIAE
	Ensembler P-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
	VIDEOI/IOIF eat-MLP	0.818±0.031	0.653±0.033	0.910±0.014	0.695±0.017	0.838±0.007	0.64/±0.005
				- continue ——	420		NDO
		4. A/		5. A	ASK	6. C	
	E	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			DUIO			
		7. D		8. D	RD3	9. H	KH3
		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. C	NHKW	ł			
	I <u> </u>	RMSE	MAE	ł			
	EnsembleFP-MLP	1.023±0.006	0.815±0.002	ļ			
	VideoMolFeat-MLP	0.875±0.004	0.672±0.003				

### Ref 1.3 - "Standard error/deviation of experimental results repeated multiple times" -

Reviewer Comment	Why is there no standard error/deviation for any of the results? Have the experiments been repeated multiple times? Having some indication about the uncertainty intervals around the results would be appreciated.
Author Response	We repeated all molecular property prediction experiments for 10 times with random seeds 0 to 9 and the remaining experiments were repeated with 3 random seeds 0 to 2. All these new standard deviations have been provided in the <b>Supplementary Tables 3-7</b> . We found that overall standard error/deviation and 95% CI are very small, indicating robustness of VideoMol models.

	BTK	CDK4-cyclinD3	EGER	EGER1	FGF	<b>P</b> 2
	0.556+0.118	0 778+0 171	0.583+0.067	0.695+0.24	19 0.667+0	0.132
	0.602+0.129	0.944+0.039	0.750+0.051	0.619+0.37	78 0.667+0	0.052
	0.611+0.000	0.667+0.000	0.536+0.000	0.013±0.07	0 0.007±0	0.002
	0.694+0.000	0.750+0.000	0.821+0.000	0.7/1±0.00	0 0 704+0	0.000
	0.05410.000	0.833+0.000	0.526±0.020	0.74310.00	50 0.704±0	0.046
	0.530±0.023	0.639+0.039	0.667+0.061	0.040±0.00	31 0.7/11+0	0.000
	0.53710.013	0.03910.039	0.548±0.017	0.04310.00	27 0.685±0	0.055
	0.040±0.013	0.917±0.000	0.040±0.017	0.470±0.02	10 0.000±0	0.000
	0.639±0.039	0.639±0.039	0.607±0.000	0.476±0.01	13 0.556±0	J.030
CHEM-BERT	0.648±0.013	0.583±0.297	0.845±0.094	0.429±0.11	17 0.765±0	J.106
ImagelViol	0.843±0.026	0.917±0.068	0.857±0.000	0.857±0.02	23 0.852±0	J.052
VideoMol	0.861±0.023	0.972±0.039	0.905±0.017	0.848±0.02	27 0.988±0	J.017
		Co	ntinue			
	FGFR3	FGFR4	FLT3	KPCD3	ME	T A
	0.760±0.039	0.773±0.121	0.722±0.091	0.571±0.10	)7 0.611±0	).236 (
	0.792±0.106	0.537±0.013	0.722±0.208	0.505±0.06	57 0.574±0	).052 (
RNN LR	0.646±0.015	0.528±0.000	0.778±0.000	0.457±0.00	)0 0.796±0	).026 (
TRFM_LR	0.812±0.000	0.639±0.000	0.611±0.000	0.438±0.02	27 0.778±0	000.0
RNN MLP	0.469±0.026	0.269±0.035	0.630±0.105	0.410±0.03	36 0.667±0	0.045 (
TRFM MLP	0.802±0.015	0.676±0.035	0.667±0.136	0.219±0.02	27 0.556±0	000.0
RNN RF	0.312±0.000	0.389±0.000	0.519+0.026	0.343±0.00	0.500±0	0.000 0
TREM RE	0 646+0 078	0.602+0.013	0.546+0.035	0 262+0 05	58 0.593+0	0.026 (
CHEM-BERT	0 438+0 077	0.528+0.060	0.574+0.189	0.557+0.09	91 0.944+0	0,000 0
ImageMol	0.854+0.064	0.833+0.045	0 722+0 120	0.762+0.08	38 <b>0 963+</b> 0	0 0 26
VideoMol	0.806+0.030	0.852+0.080	0.081+0.026	0.967+0.03	36 0.063+(	0.026 (
VIGEOIVIOI	0.09010.039	0.052±0.000	0.961±0.020	0.007±0.03	0.9031	J.020 (
Table S4: The RM	ISE and MAE perform	ance of different metho erformance. All compa	ods on 10 GPCR with ared results are obtain	balanced scaffold s ned from ImageMol.	plit. The lower the va	alue, the bette
	1. 5H	T1A	2. 5HT	2A	3. A	A1R
	RMSE	MAE	RMSE	MAE	RMSE	MAE
	0.850±0.021	0.670±0.012	0.853±0.019	0.642±0.014	0.786±0.015	0.588±0
	0.949±0.027	0.764±0.014	0.875±0.008	0.681±0.024	0.856±0.026	0.662±0
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
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
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
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
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
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
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
ImageMol	0.776±0.012	0.620±0.014	0.780±0.017	0.578±0.022	0.711±0.012	0.554±0
VideoMol	0.708±0.017	0.547±0.015	0.775±0.017	0.577±0.009	0.655±0.007	0.496±0
	4. AA	2AR	5. AA3	R	6. C	NR2
	RMSE	MAE	RMSE	MAE	RMSE	MAE
	0.748±0.012	0.588±0.008	0.840±0.014	0.692±0.010	0.926±0.047	0.758±0
MOCLRGCN	0.819±0.011	0.651±0.008	0.855±0.010	0.700±0.011	0.978±0.023	0.803±0
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
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
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
TREM MLP	0.948±0.013	0.744±0.005	0.945±0.010	0.749±0.014	1.144±0.055	0.903±0
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
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
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
ImageMol	0.734+0.015	0.573+0.009	0.793+0.008	0.634+0.001	0.905+0.004	0.717+0
VideoMol	0.704±0.010	0.5/3+0.005	0.786+0.006	0.617+0.004	0.864+0.005	0.679+0
VIGEOIVOI	7 DE	0.04010.000	8 DPI	13	0.004±0.005	DH3
	PMSE	MAE	PMSE	MAE	PMSE	MAE
	0.81/1+0.000	0.591+0.007	0.858+0.017	0.673+0.022	0 73/+0 006	0.591+0
	0.855+0.022	0.63/+0.017	0.01/1+0.02/	0.725+0.025	0.70+0.000	0.576+0
	1 1/2+0 077	0.00+10.017	1 316+0 011	0.02010.020	1 616+0 226	0.070±0
	1.14210.077	0.03910.030	1 210±0.011	0.04210.000	1 160+0 002	0.943±0
	0.805±0.000	0.7 1910.000	1.219±0.000	0.914±0.000	0.971+0.015	0.911±0
	0.03010.000	0.03410.000	1.02120.007	0.01910.007	0.07 110.010	0.702±0
	0.91910.010	0.000±0.010	1.012±0.041	0.19010.023	0.003±0.011	0.070±0
	0.03/ ±0.001	0.012±0.001	0.001±0.001	0.00010.001	0.770+0.002	0.013±0
	0.004±0.001	0.000±0.002	0.00410.001	0.011010000	0.770+0.002	0.002±0
	0.010±0.011	0.00/±0.013	0.003±0.029	0.03 I±0.020	0.710±0.033	0.594±0
VideeMel	0.772±0.014	0.5/310.009	0.10010.010	0.010±0.014	000.01010100	0.00 I±0.
VIDEOIVIOI	0.142IU.UU4	0.000IU.005	0.71520.014	0.55410.012	0.00010.008	U.306±0.
	10.01					
	RMSE	MAE				
	0.856±0.008	0.664±0.016				
	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				
_	0.000.0004	0 672+0 010				
CHEM-BERT	0.893±0.024	0.072±0.019				
CHEM-BERT ImageMol	0.893±0.024 0.849±0.018	0.645±0.015				

InfoGraph         73.3 (0.6)         61.8 (0.4)         58.7 (0.6)         74.2 (0.9)         68.7 (0.6)         74.3 ()           GPT-GNN         74.9 (0.3)         62.5 (0.4)         58.1 (0.3)         65.2 (2.1)         64.5 (1.4)         77.9 ()           ContextPred         73.6 (0.3)         62.6 (0.6)         59.7 (1.8)         75.6 (1.0)         70.6 (1.5)         78.8 ()           G-Contextual         75.0 (0.6)         62.4 (0.7)         58.7 (1.0)         76.3 (1.5)         68.9 (2.1)         79.3 ()           G-Motif         73.6 (0.7)         62.3 (0.6)         61.0 (1.5)         73.8 (1.2)         66.9 (3.1)         73.6 (0.7)           AD-CGL         74.4 (0.4)         62.8 (0.7)         60.4 (1.5)         76.7 (0.2)         70.7 (0.3)         76.6 ()           JOAO         74.8 (0.6)         62.8 (0.7)         60.2 (0.9)         75.0 (0.6)         71.2 (1.1)         74.9 ()           GraphMAE         75.2 (0.9)         63.3 (0.5)         66.6 (2.1)         76.1 (1.3)         69.1 (1.2)         78.6 ()           MGSSL         75.2 (0.6)         63.3 (0.5)         66.6 (2.1)         75.3 (1.5)         65.2 (1.4)         77.8 ()           MolCLR         75.5 (0.5)         66.3 (0.5)         66.3 (1.3)         77.1 (2.1)         <	#Task	12	617	27	1	1	1
GPT-GNN         74.9 (0.3)         62.5 (0.4)         58.1 (0.3)         65.2 (2.1)         64.5 (1.4)         77.9 ()           ContextPred         73.6 (0.3)         62.6 (0.6)         59.7 (1.8)         75.6 (1.0)         70.6 (1.5)         78.8 ()           GraphLoG         75.0 (0.6)         63.4 (0.6)         59.6 (1.9)         76.1 (0.8)         68.7 (1.6)         78.6 (0.7)           G-Motif         73.6 (0.7)         62.3 (0.6)         61.0 (1.5)         73.8 (1.2)         66.9 (3.1)         73.0 ()           AD-GCL         74.9 (0.4)         63.4 (0.7)         61.5 (0.9)         76.7 (0.2)         70.7 (0.3)         76.6 ()           JUOAO         74.8 (0.6)         62.2 (0.7)         60.2 (0.9)         75.9 (0.7)         66.4 (1.0)         73.2 ()           GraphCL         75.1 (0.7)         63.0 (0.4)         59.8 (7.1)         67.8 (2.4)         74.6 ()           GraphMAE         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 ()           MGSSL         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 ()           MolCLR         75.5 (0.5)         65.9 (0.9)         60.3 (1.3)         77.1 (2.1)         69.9 (1.4)         79	InfoGraph	73.3 (0.6)	61.8 (0.4)	58.7 (0.6)	74.2 (0.9)	68.7 (0.6)	74.3 (2
ContextPred         73.6 (0.3)         62.6 (0.6)         59.7 (1.8)         75.6 (1.0)         70.6 (1.5)         78.8 ()           GraphLoG         75.0 (0.6)         63.4 (0.6)         59.6 (1.9)         76.1 (0.8)         68.7 (1.6)         78.6 (.6)           G-MoUT         73.8 (0.7)         62.2 (0.7)         58.7 (1.0)         76.3 (1.5)         69.9 (2.1)         73.0 (.2)           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)           JOAO         74.8 (0.6)         62.8 (0.7)         60.2 (0.9)         75.0 (0.6)         71.2 (1.1)         74.9 (.4)           GraphCL         75.1 (0.7)         63.0 (0.4)         69.8 (1.3)         75.1 (0.7)         67.8 (2.4)         74.6 (.2)           GraphMAE         75.2 (0.9)         63.6 (0.3)         60.5 (1.2)         76.8 (0.6)         71.2 (1.0)         78.2 (.2)           MGSSL         75.2 (0.9)         63.3 (0.5)         61.6 (1.0)         75.8 (.4)         68.8 (0.6)         78.8 (.4)           MoICLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         77.1 (.2)         69.9 (1.4)         77.6 (.3)           ImageMoI         75.5 (1.0)         65.6 (0.9)         64.9 (1.3)         77.7 (.2)	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
GraphLoG         75.0 (0.6)         63.4 (0.6)         59.6 (1.9)         76.1 (0.8)         68.7 (1.6)         78.6 (1.6)           G-Contextual         75.0 (0.6)         62.8 (0.7)         68.7 (1.0)         76.3 (1.2)         66.9 (3.1)         73.0 (7.3)           G-Motif         73.6 (0.7)         62.3 (0.6)         61.0 (1.5)         73.8 (1.2)         66.9 (3.1)         73.0 (7.3)           JOAO         74.8 (0.6)         62.8 (0.7)         60.4 (1.5)         76.7 (1.2)         70.7 (0.3)         76.6 (1.0)           JOAO         74.8 (0.6)         62.8 (0.7)         60.4 (1.5)         76.7 (1.2)         70.7 (0.3)         76.6 (1.0)         75.2 (1.0)         75.2 (1.1)         74.9 (7.3)           GraphMAE         75.2 (0.9)         63.3 (0.5)         61.6 (1.0)         75.8 (0.6)         71.2 (1.1)         78.8 (1.2)           GraphMAE         75.2 (0.6)         63.3 (0.5)         66.8 (2.1)         76.1 (1.3)         69.1 (1.2)         78.6 (1.3)           MolCLR         75.5 (0.5)         63.3 (0.5)         66.8 (2.1)         76.4 (1.3)         70.5 (1.3)         78.1 (1.2)           IUni-Mol (1 conf)         78.8 (0.7)         60.3 (0.6)         60.5 (0.3)         77.1 (2.1)         69.9 (2.7)         81.1 (1.2)         11.3         79.2 (1.3)	ContextPred	73.6 (0.3)	62.6 (0.6)	59.7 (1.8)	75.6 (1.0)	70.6 (1.5)	78.8 (1
G-Contextual         75.0 (0.6)         62.8 (0.7)         58.7 (1.0)         76.3 (1.5)         69.9 (2.1)         79.3 ( 79.3 (1.2)           G-Motif         73.6 (0.7)         62.3 (0.6)         61.0 (1.5)         73.8 (1.2)         60.9 (3.1)         73.0 ( 70.7 (0.3)         76.6 ( 71.2 (1.1)         74.9 (0.4)         63.4 (0.7)         60.2 (0.9)         75.0 (0.6)         71.2 (1.1)         74.9 (0.4)           GraphCL         75.1 (0.7)         63.3 (0.0)         60.5 (1.2)         76.8 (0.6)         71.2 (1.1)         78.2 (2.4)         74.6 (0.7)           GaraphMAE         75.2 (0.9)         63.3 (0.5)         66.8 (2.1)         76.1 (1.3)         69.1 (1.2)         78.6 (1.3)           ModSL         75.2 (0.6)         63.3 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1.5)         65.2 (1.4)         77.8 (2.4)           MolCLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1.5)         66.2 (1.4)         77.8 (2.1)         69.6 (2.0)         81.0           Uni-Mol (10 conf)         78.3 (0.7)         65.0 (0.5)         63.7 (1.3)         79.2 (0.0)<	GraphLoG	75.0 (0.6)	63.4 (0.6)	59.6 (1.9)	76.1 (0.8)	68.7 (1.6)	78.6 (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 ( 73.0 (0.3)           AD-GCL         74.9 (0.4)         63.4 (0.7)         61.5 (0.9)         76.7 (1.2)         70.7 (0.3)         76.6 (1.0)           JAOA         74.8 (0.6)         62.8 (0.7)         60.4 (1.5)         76.9 (0.7)         66.8 (1.0)         73.2 (1.7)           GraphCL         75.1 (0.7)         63.0 (0.4)         59.8 (1.3)         75.1 (0.7)         67.8 (2.4)         74.6 (0.6)           GraphMAE         75.2 (0.9)         63.3 (0.5)         66.8 (2.1)         76.1 (1.3)         69.1 (1.2)         77.8 (0.6)           GraphMAE         75.2 (0.9)         63.3 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1.5)         65.2 (1.4)         77.8 (1.5)           MolCLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1.5)         65.2 (1.4)         77.8 (1.0)           MolCLR         75.5 (0.5)         63.9 (0.5)         63.3 (1.3)         76.4 (1.3)         76.8 (1.3)         70.5 (1.3)         77.8 (1.0)           Uni-Mol (1 conf)         78.8 (0.7)         69.0 (0.5)         63.6 (1.4)         79.2 (0.9)         69.9 (2.7)         81.1 (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
AD-GCL         74.9 (0.4)         63.4 (0.7)         61.5 (0.9)         76.7 (1.2)         70.7 (0.3)         76.6 ( JOAO           JOAO         74.8 (0.6)         62.8 (0.7)         60.4 (1.5)         76.9 (0.7)         66.4 (1.0)         73.2 ( 73.2 ( 73.2 (1.1))         74.4 (0.3)         62.6 (0.7)         60.2 (0.9)         75.0 (0.6)         71.2 (1.1)         74.9 ( 74.4 (0.3)           GraphCL         75.1 (0.7)         63.3 (0.4)         59.8 (1.3)         75.1 (0.7)         67.8 (0.6)         71.2 (1.0)         78.2 (0.6)           GraphMAE         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 (0.6)           MGSSL         75.2 (0.6)         63.3 (0.6)         60.5 (0.9)         75.3 (1.5)         66.5 (2.1)         76.6 (1.3)         77.4 (1.3)         69.1 (1.4)         77.8 (1.4)           MolCLR         75.5 (1.0)         66.6 (0.9)         64.9 (1.3)         77.4 (1.3)         66.8 (2.0)         61.7 (1.3)           Uni-Mol (10 conf)         78.8 (0.7)         69.0 (0.5)         63.3 (1.4)         79.2 (0.9)         69.9 (2.7)         81.7 (1.0)           Uni-Mol (10 conf)         78.8 (0.5)         66.7 (0.5)         63.3 (0.7)         77.7 (0.7)         65.7 (2.3)         60.2 (1.4)           VideoMol </td <td>G-Motif</td> <td>73.6 (0.7)</td> <td>62.3 (0.6)</td> <td>61.0 (1.5)</td> <td>73.8 (1.2)</td> <td>66.9 (3.1)</td> <td>73.0 (3</td>	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
JOAO         74.8 (0.6)         62.8 (0.7)         60.4 (1.5)         76.9 (0.7)         66.4 (1.0)         73.2 (1)           SimGRACE         74.4 (0.3)         62.6 (0.7)         60.2 (0.9)         75.0 (0.6)         71.2 (1.1)         74.9 (1)           GraphALE         75.2 (0.9)         63.3 (0.0,4)         59.8 (1.3)         75.1 (0.7)         67.8 (2.4)         74.6 (1)           GraphALE         75.2 (0.9)         63.3 (0.0,5)         66.1 (1.0)         75.8 (0.6)         71.2 (1.0)         78.6 (1)           MGSSL         75.2 (0.6)         63.3 (0.5)         661.6 (1.0)         75.8 (0.4)         668.8 (0.6)         78.8 (1)           MoICLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         77.4 (1.13)         66.8 (3.4)         77.8 (1)           GraphMVP-C         74.6 (0.4)         63.4 (0.6)         60.6 (1.3)         77.1 (2.1)         69.9 (1.4)         79.8 (1)           Uni-MoI (1 conf)         78.3 (0.7)         69.0 (0.5)         63.3 (1.4)         79.2 (1.0)         69.9 (2.7)         81.7 (1)           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.2)         80.2 (1)           VideoMol         78.8 (0.7)         66.7 (0.5)         663.3 (0.9)         794	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
SimGRACE         74.4 (0.3)         62.6 (0.7)         60.2 (0.9)         75.0 (0.6)         71.2 (1.1)         74.9 (           GraphCL         75.1 (0.7)         63.0 (0.4)         59.8 (1.3)         75.1 (0.7)         67.8 (2.4)         74.4 (0.7)           GraphMAE         75.2 (0.9)         63.6 (0.3)         60.5 (1.2)         76.8 (0.6)         71.2 (1.0)         78.2 (1.3)           GraphMAE         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 (0.7)           MGSSL         75.2 (0.6)         63.3 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1.6)           MolCLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         77.4 (1.1)         69.9 (1.4)         79.6 (1.0)           ImageMol         75.5 (1.0)         65.6 (0.9)         64.9 (1.3)         79.5 (1.0)         69.9 (1.4)         79.6 (1.0)           Uni-Mol (1 conf)         78.8 (0.7)         69.0 (0.5)         63.7 (1.3)         79.2 (1.0)         69.9 (2.7)         81.7 (1.7)           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (1.7)           VideoMol         78.8 (0.5)         65.7 (0.5)         66.3 (0.9)         <	JOAO	74.8 (0.6)	62.8 (0.7)	60.4 (1.5)	76.9 (0.7)	66.4 (1.0)	73.2 (1
GraphCL         75.1 (0.7)         63.0 (0.4)         59.8 (1.3)         75.1 (0.7)         67.8 (2.4)         74.6 (           GraphMAE         75.2 (0.9)         63.5 (0.3)         60.5 (1.2)         76.8 (0.6)         71.2 (1.0)         78.2 (           3D InfoMax         74.5 (0.7)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 (           MGSSL         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.6 (           MoICLR         75.5 (0.5)         63.3 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (           GraphMVP-C         74.6 (.0.4)         63.4 (0.6)         60.6 (1.3)         77.1 (2.1)         69.9 (1.4)         79.6 (           Uni-Mol (10 conf)         78.3 (0.4)         65.6 (0.9)         63.7 (1.3)         79.2 (1.0)         69.6 (2.0)         81.0 (           Uni-Mol (10 conf)         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (           Rank         1         2         1         1         3         1           Uni-Mol (12 conf)         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4	SimGRACE	74.4 (0.3)	62.6 (0.7)	60.2 (0.9)	75.0 (0.6)	71.2 (1.1)	74.9 (2
GraphMAE         75.2 (0.9)         63.6 (0.3)         60.5 (1.2)         76.8 (0.6)         71.2 (1.0)         78.2 (1.3)           3D InfoMax         74.5 (0.7)         63.5 (0.8)         56.8 (2.1)         76.1 (1.3)         69.1 (1.2)         78.6 (1.3)           MGSSL         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 (0.4)           MolCLR         75.5 (0.5)         63.3 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1.5)           GraphMVP-C         74.6 (0.4)         63.4 (0.6)         60.6 (1.3)         77.1 (2.1)         69.9 (1.4)         79.6 (1.0)           ImageMol         75.5 (1.0)         65.6 (0.9)         64.9 (1.3)         76.8 (1.3)         70.5 (1.3)         78.1 (1.0)           Uni-Mol (10 conf)         78.8 (0.7)         69.0 (0.5)         63.7 (1.3)         79.2 (1.0)         69.6 (2.0)         81.0 (1.0)           Uni-Mol (10 conf)         78.8 (0.7)         69.0 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         80.2 (1.0)           VideoMol         78.8 (0.7)         69.0 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         80.2 (1.0)           Table S6: The RMSE or MAE performance of different methods on 6 molecular property predic	GraphCL	75.1 (0.7)	63.0 (0.4)	59.8 (1.3)	75.1 (0.7)	67.8 (2.4)	74.6 (2
3D InfoMax         74.5 (0.7)         63.5 (0.8)         56.8 (2.1)         76.1 (1.3)         69.1 (1.2)         78.6 (           MGSSL         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 (           AttrMask         75.1 (0.9)         63.3 (0.5)         60.5 (0.9)         75.3 (1.5)         66.2 (1.4)         77.8 (           MoICLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (           GraphMVP-C         74.6 (0.4)         63.4 (0.6)         66.6 (1.3)         77.1 (2.1)         69.9 (1.4)         78.8 (           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 (2.0)           Uni-Mol (1 conf)         78.8 (0.7)         69.0 (0.5)         63.6 (1.4)         79.2 (0.9)         69.9 (2.7)         81.7 (2.0)           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (2.0)           VideoMol         78.8 (0.5) <u>66.7 (0.5)</u> 66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (2.7)           Rank         1         2         1         1         3	GraphMAE	75.2 (0.9)	63.6 (0.3)	60.5 (1.2)	76.8 (0.6)	71.2 (1.0)	78.2 (1
MGSSL         75.2 (0.6)         63.3 (0.5)         61.6 (1.0)         75.8 (0.4)         68.8 (0.6)         78.8 (           AttrMask         75.1 (0.9)         63.3 (0.6)         60.5 (0.9)         75.3 (1.5)         65.2 (1.4)         77.8 (           MolCLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.8 (           GraphMVP-C         74.6 (0.4)         63.4 (0.6)         60.6 (1.3)         77.1 (2.1)         69.9 (1.4)         79.6 (           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 (           Uni-Mol (1 conf)         78.8 (0.7)         69.0 (0.5)         63.6 (1.4)         79.2 (0.9)         69.9 (2.7)         81.7 (           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (           VideoMol         78.8 (0.5)         65.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (           Rank         1         2         1         1         3         1           VideoMol         78.8 (0.5)         ESOL         Lpo         OM7         OM8 (0M8 datasets, respectively, Weregat Respectively, Weregat R	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
AttrMask         75.1 (0.9)         63.3 (0.6)         60.5 (0.9)         75.3 (1.5)         65.2 (1.4)         77.8 (           MolCLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (3.4)           GraphMVP-C         74.6 (0.4)         63.4 (0.6)         60.6 (1.3)         77.1 (2.1)         69.9 (1.4)         79.6 (1.4)           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 (1.0)           Uni-Mol (1 conf)         78.8 (0.7)         69.0 (0.5)         63.3 (1.4)         79.2 (0.9)         69.9 (2.7)         81.7 (1.0)           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (2.7)           VideoMol         78.8 (0.5) <u>66.7 (0.5)</u> 66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (1.0)           Rank         1         2         1         1         3         1           Segen table sector MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run using random seeds from 0 to 9. We report RMSE for FreeSolv. ESOL and Upo datasets and MAE for OM and OMS OMB (datasets, respectively. We repart all comparison methods using the same settings. We use G1N backhone fo	MGSSL	75.2 (0.6)	63.3 (0.5)	61.6 (1.0)	75.8 (0.4)	68.8 (0.6)	78.8 (0
MoiCLR         75.5 (0.5)         63.9 (0.5)         60.3 (1.3)         74.4 (1.3)         66.8 (3.4)         75.3 (1)           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)           ImageMol         75.5 (1.0)         65.6 (0.9)         64.9 (1.3)         76.8 (1.3)         70.5 (1.3)         78.1 (0)           Uni-Mol (10 conf)         78.8 (0.7)         69.0 (0.5)         63.6 (1.4)         79.2 (1.0)         69.9 (2.7)         81.7 (1)           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (1)           VideoMol         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (1)           Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold spill. All experiments are run using random seeds from 0 to 9. We report TMSE for FreeSolv. ESOL and Upo datasets and MAE for OM and OM8, OM9 datasets, respectively. We represt all comparison methods using the same settings. We use GIN backbone for MoiCLR because it achieves the best results.           RMSE         FreeSolv         ESOL ESOL and Upo datasets and MAE for OM and OM8, OM9 datasets and OM9 OM OM (2)	AttrMask	75.1 (0.9)	63.3 (0.6)	60.5 (0.9)	75.3 (1.5)	65.2 (1.4)	77.8 (1
GraphMVP-C         74.6(0.4)         63.4 (0.6)         60.6 (1.3)         77.1 (2.1)         69.9 (1.4)         79.6 (           ImageMol         75.5 (1.0)         65.6 (0.9)         64.9 (1.3)         76.8 (1.3)         70.5 (1.3)         78.1 (           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 (           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 (           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (           VideoMol         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (           Rank         1         2         1         1         3         1           Table S8: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold spit. All experiments are nu using random seeds from 0.9. We report RMSE for FreeSolv. ESOL and Upo datasets and MAE for OM7 and OM8, OM6 datasets, respectively. We repart RMSE or FreeSolv. ESOL and Upo datasets and MAE for OM7 and OM8, OM6 datasets, respectively. Use repart RMSE databaset is a stabaset and is a stabaset achieves the best results.         0.0002440         0.00024940.00020         0.00089144         EdgePrMV	MolCLR	75.5 (0.5)	63.9 (0.5)	60.3 (1.3)	74.4 (1.3)	66.8 (3.4)	75.3 (2
ImageMol         75.5 (1.0)         65.6 (0.9)         64.9 (1.3)         76.8 (1.3)         70.5 (1.3)         78.1 (           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 (           Uni-Mol (10 conf) <b>78.8 (0.7) 69.0 (0.5)</b> 63.6 (1.4)         79.2 (0.9)         69.9 (2.7)         81.7 (           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (           VideoMol <b>78.8 (0.5)</b> <u>66.7 (0.5)</u> <b>66.3 (0.9) 79.4 (0.5)</b> <u>70.7 (2.2)</u> <b>82.4</b> (           Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold spit. All experiments are run using random seed from 0.9. We report RMSE for FreeSolv. ESOL and Upo datasets and MAE for OM7 and OM8. OM9 datasets, respectively. We reparation methods using the same settings. We use GIN backbone for MoICLR because it achieves the best results.         CMMSE         CMMSE OR OM8 OM9 datasets, respectively. We reparation and the spit statistic statistat statistat statistic statistic statistatistic statistic statist	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
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 (           Uni-Mol (10 conf) <b>78.8 (0.7) 69.0 (0.5)</b> 63.6 (1.4)         79.2 (0.9)         69.9 (2.7)         81.7 (           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (           VideoMol <b>78.8 (0.5)</b> <u>66.7 (0.5)</u> <b>66.3 (0.9) 79.4 (0.5) 70.7 (2.2) 82.4</b> (           Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run using random seeds from 0 to 9. We report RMSE for FreeSolv, ESOL and Lipo datasets and MAE for CMV and OM8, OM9 datasets, respectively. We regative all comparison methods using the same settings. We use GIN backbone for MolCL Result at hieres the best results.           RMSE         FreeSolv         ESOL         Lipo         OM7         OM8         OM8           GraphMVP         2.559+0.158         1.322-0.062         0.773±0.016         121.022±5.699         0.0022±0.00032         0.0029±4           Mocic R         3.112±0.638         1.462±0.068         0.799±0.018         144.42±6±6.591 <td< td=""><td>ImageMol</td><td>75.5 (1.0)</td><td>65.6 (0.9)</td><td>64.9 (1.3)</td><td>76.8 (1.3)</td><td>70.5 (1.3)</td><td>78.1 (3</td></td<>	ImageMol	75.5 (1.0)	65.6 (0.9)	64.9 (1.3)	76.8 (1.3)	70.5 (1.3)	78.1 (3
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 (           Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (           VideoMol         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (           Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run using random seeds from 0 to 9. We report RMSE for FreeSolv. ESOL and Lipo datasets and MAE for QM8 datasets, respectively. We regall comparison methods using the same settings. We use GIN backbone for MoICLR beause 1 achieves the best results.           RMSE         FreeSolv         ESOL         Lipo         QM7         QM8         QM8           GraphMVP         2.5594.0158         1.3224.0662         0.7734.016         120.3446.637         0.02094940.00032         0.000941           EdgePred         2.8434.0091         1.334.0055         0.7684.013         121.0225.699         0.0022540.00061         0.002941           MolCLR         3.11240.638         1.4624.068         0.7994.018         144.4266.691         0.0359640.0065         0.0	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
Mole-BERT         77.0 (0.3)         64.4 (0.2)         63.2 (0.7)         77.7 (0.7)         65.7 (2.3)         80.2 (           VideoMol         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (           Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run 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 repart and comparison methods using the same settings. We use GIN backbone for MolCLR because it achieves the best results.           RMSE         FreeSolv         ESOL         Lipo         QM7         QM8         QM8           GraphWP-C         2.559e.0158         1.3224.0062         0.773e.0.013         104.337s.3292         0.02024b.000032         0.009294           MolcLR         3.1122.038         1.462±0.068         0.799e.0.018         144.426e.6391         0.03398e.00085         0.01488et           MolcLR         3.1122.038         0.462±0.067         0.7722.0060         116.384z8.445         0.0224b.00033         0.000914f           MolcLR         3.1122.0383         0.964±0.067         0.722±0.006         101.922±.331         0.02073±0.0033	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
VideoMol         78.8 (0.5)         66.7 (0.5)         66.3 (0.9)         79.4 (0.5)         70.7 (2.2)         82.4 (           Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run 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 repail comparison methods using the same settings. We use GIN backbone for MoICLR because it achieves the best results.           RMSE         FreeSolv         ESOL         Lipo         QM7         QM8         QM8           GraphMVP         2.559±0.156         1.322±0.062         0.773±0.016         120.344±6.237         0.02049±0.00032         0.00099±4           EdgePred         2.843±0.091         1.337±0.041         0.778±0.013         121.022±5.699         0.0202±0.00047         0.0029±64           MoICLR         3.112±0.633         1.462±0.068         0.799±0.018         144.426±6.501         0.0358±0.00033         0.0201±0.0003         0.0201±0.0003         0.0201±0.0003         0.0201±0.00033         0.0201±0.0003         0.0201±0.0003         0.0201±0.0003         0.0201±0.0003         0.0201±0.000         0.0038±4           VideoMol         1.73±0.053         0.866±0.017         0	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
Rank         1         2         1         1         3         1           Table S6: The RMSE or MAE performance of different methods on 6 molecular property prediction benchmarks with scaffold split. All experiments are run 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 repair and comparison methods using the same settings. We use GIN backbone for MolCLR because it achieves the best results.           RMSE         FreeSolv         ESOL         Lipo         QM7         QM8         QM8           GraphMVP         2.559±0.158         1.322±0.062         0.773±0.016         120.344±6.237         0.02049±0.00032         0.0089±42           GraphMVP         2.559±0.158         1.322±0.062         0.773±0.016         120.344±6.237         0.0202±0.00061         0.00329±42           GraphMVP-C         2.766±0.199         1.333±0.055         0.768±0.013         121.0225.699         0.0202±2.00071         0.0202±2.00071         0.0202±1.00033         0.00081±42           MolcLR         3.112±0.638         1.462±0.068         0.799±0.018         144.428±6.591         0.0329±4.00033         0.00081±42           Mole-BET         2.988±0.155         1.118±0.017         0.722±0.006         101.922±2.331         0.0207±0.0003         0.0208±42            0.366±0.017         0	VideoMol	78.8 (0.5)	<u>66.7 (0.5)</u>	66.3 (0.9)	79.4 (0.5)	<u>70.7 (2.2)</u>	82.4 (0
RMSE         FreeSolv         ESOL         Lipo         OM7         QM8         QM           GraphMVP         2.55940.158         1.322±0.062         0.773±0.016         120.344±6.237         0.02049±0.00032         0.00891±t           EdgePred         2.843±0.091         1.387±0.041         0.778±0.013         104.387±3.292         0.02068±0.0061         0.00929±t           GraphMVP-C         2.766±0.199         1.333±0.055         0.768±0.013         121.02±5.699         0.02022±0.00047 <u>0.00896±t</u> MoICLR         3.112±0.638         1.462±0.068         0.799±0.018         144.426±5.91         0.0358±0.00085         0.01488±t           ImageMol         2.113±0.235         0.964±0.067         0.702±0.006         116.384±8.445         0.02419±0.00033         0.02061±t           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020 <u>0.00896±t</u> 1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020 <u>0.00896±t</u> Stepsize           OK           ACE2         hCYTOX         MERS-PPE_cs         MERS-PPE         CoV1-PI           REDIAL-2020<	Rank	1	2	1	1	3	1
RMSE         FreeSolv         ESOL         Lipo         QM7         QM8         QM           GraphMVP         2.559±0.158         1.32±0.062         0.773±0.016         120.344±6.237         0.02049±0.00032         0.0089±td           GraphMVP-C         2.643±0.091         1.33±0.062         0.773±0.013         104.387±3.292         0.0202±0.00047         0.0089±td           GraphMVP-C         2.766±0.199         1.333±0.055         0.768±0.013         121.02±6.699         0.0202±0.00047         0.0089±td           MolCLR         3.112±0.638         1.462±0.068         0.799±0.018         144.426±6.591         0.0359±0.00065         0.148±td           ImageMol         2.113±0.235         0.964±0.067         0.702±0.006         116.34±8.445         0.02419±0.00033         0.0091±td           Mole-BERT         2.988±0.155         1.115±0.017         0.727±0.006         101.922±2.331         0.02073±0.00033         0.0091±td           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.0089±td           VideoMol         1.728±0.057         0.753         0.710         0.773±0.011         0.775±           REDIAL-2020         0.713         0.753±0.025         0.765±0.003         0.828±0.027<	Rank Table S6: The RMSE or using random seeds from	1 MAE performance of diff n 0 to 9. We report RMSI	erent methods on 6 m E for FreeSolv, ESOL	olecular property predia and Lipo datasets and	1 ction benchmarks with s MAE for QM7 and QM8	3 caffold split. All experin , QM9 datasets, respective	nents are run 1 titvely. We repr
Table S7: The ROC-AUC performance of different methods on 11 SARS-CoV-2 datasets with balanced scaffold split.           Compared 23CL         ACE2         hCYTOX         MERS-PPE_cs         MERS-PPE         CoV1-PI           Table S7: The ROC-AUC performance of different methods on 11 SARS-CoV-2 datasets with balanced scaffold split.         0.003984.0.007         0.003964.0.007         0.003964.0.007         0.003964.0.007         0.003964.0.007         0.003964.0.007         0.02024.0.00047         0.003964.0.007         0.021496.0.0035         0.0148840           MolcLR         3.112±0.638         1.462±0.068         0.799±0.018         144.426±6.591         0.03598±0.00085         0.0148840         0.022419±0.00033         0.020141           Mole-BERT         2.988±0.155         1.115±0.017         0.727±0.006         101.922±2.331         0.02073±0.00033         0.009144           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.00395±0.0020           VideoMol         1.728±0.053         0.366±0.010         0.727±0.009         0.771±0.009         0.773±0.011         0.775±0.011         0.775±0.011         0.775±0.011         0.773±0.011         0.775±0.003         0.828±0.027         0.814±0.004         0.836±           E continue =====         E continue ===== <td>Rank Table S6: The RMSE or using random seeds fron all cr</td> <td>A MAE performance of diff n 0 to 9. We report RMSI omparison methods using</td> <td>erent methods on 6 m E for FreeSolv, ESOL g the same settings. V</td> <td>olecular property predia and Lipo datasets and l /e use GIN backbone fe</td> <td>1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a</td> <td>3 acaffold split. All experint , QM9 datasets, respectives the best results</td> <td>nents are run 1 titvely. We repr</td>	Rank Table S6: The RMSE or using random seeds fron all cr	A MAE performance of diff n 0 to 9. We report RMSI omparison methods using	erent methods on 6 m E for FreeSolv, ESOL g the same settings. V	olecular property predia and Lipo datasets and l /e use GIN backbone fe	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a	3 acaffold split. All experint , QM9 datasets, respectives the best results	nents are run 1 titvely. We repr
GraphMVP-C         2.766±0.199         1.333±0.055         0.768±0.013         121.0225.699         0.02022±0.00047         0.00896±0           MolCLR         3.112±0.638         1.46±0.068         0.799±0.018         144.426±6.591         0.0328±0.00085         0.01488±0           ImageMol         2.113±0.235         0.964±0.067         0.702±0.060         116.384±8.445         0.0217±0.00033         0.02061±0           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           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.00896±0           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.00896±0           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.00896±0           MERS-PPE         CoV1-PI         CoV1-PI         CoV1-PI         CoV1-PI         CoV1-PI         0.703         0.696         0.661           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol	Rank Table S6: The RMSE or using random seeds fron all or RMSE GranhMVP	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2 55040 158	erent methods on 6 m E for FreeSolv, ESOL g the same settings. V ESOL 1 32240.062	1 olecular property predia and Lipo datasets and I /e use GIN backbone fr Lipo 0.773+0.016	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120 344+6 237	3 ccaffold split. All experin , QM9 datasets, respec chieves the best result	1 nents are run 1 titvely. We repr s QM9
ImageMol         2.113±0.235         0.964±0.067         0.702±0.060         116.384±8.445         0.02419±0.00033         0.02011±0           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           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.00896±0           Table S7: The ROC-AUC performance of different methods on 11 SARS-CoV-2 datasets with balanced scaffold split. compared results are obtained from ImageMol.           SCL         ACE2         hCYTOX         MERS-PPE_Cs         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.66           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           E====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Meas           REDIAL-2020         0.665         0.651         0.688         0.79         0.	Rank Table S6: The RMSE or using random seeds fron all cr RMSE GraphMVP EdgePred	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091	erent methods on 6 m E for FreeSolv, ESOL g the same settings. V ESOL 1.322±0.062 1.326±0.041	0lecular property predia and Lipo datasets and le use GIN backbone for Lipo 0.773±0.016 0.778±0.013	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344:fb.237 104.387±3.292	3 caffold split. All experin QM9 datasets, respec chieves the best result QM8 0.02049t0.00032 0.02058±0.00061	1 nents are run 11 titvely. We repr 3. 0.00891±0.0 0.00929±0.0
Mole-BERT         2.98840.155         1.11520.017         0.72720.005         101.92222.331         0.0207340.00033         0.0091044           VideoMol         1.728±0.053         0.866±0.017         0.743±0.009         76.736±1.561         0.01890±0.00020         0.00896±0           Table S7: The ROC-AUC performance of different methods on 11 SARS-CoV-2 datasets with balanced scaffold split. compared results are obtained from ImageMol.         3CL         ACE2         hCYTOX         MERS-PPE_Cs         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.66           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           E==== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70	Rank Table S6: The RMSE or using random seeds fron all cr RMSE GraphMVP EdgePred GraphMVP-C MolCLR	1           MAE performance of diff           0 to 9. We report RMSI           omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638	2 erent methods on 6 m f for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.367±0.041 1.333±0.055 1.462±0.068	lecular property predia and Lipo datasets and l le use GIN backbone fr Lipo 0.773±0.016 0.778±0.013 0.768±0.013 0.769±0.018	1 ction benchmarks with s MAE for QM7 and QM8 0007 120.344±6.237 104.387±3.292 121.022±5.699 124.4262±6.591	3 acaffold split. All experin QM9 datasets, respect Chieves the best result 0.020494.0.00032 0.02058±0.00061 0.02022±0.00045	1 nents are run 10 titvely. We repr 3 0.00891±0.0 0.00929±0.0 0.01488±0.0
Table S7: The ROC-AUC performance of different methods on 11 SARS-CoV-2 datasets with balanced scaffold split. compared results are obtained from ImageMol.           3CL         ACE2         hCYTOX         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.661           mageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           E====         CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Meg           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Meg           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Meg           CoV1-PPE          0.734	Rank Table S6: The RMSE or using random seeds fron all or RMSE GraphMVP EdgePred GraphMVP-C MolCLR ImageMol ImageMol	1           MAE performance of diff           0 to 9. We report RMSI           omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235	2 erent methods on 6 m E for FreeSolv, ESOL for SeeSolv, ESOL 1.322t0.062 1.367±0.041 1.333±0.055 0.964±0.068 0.964±0.067	1  olecular property predix and Lipo  Lipo  0.773±0.016  0.778±0.013  0.768±0.013  0.799±0.018  0.792±0.018  0.792±0.018  0.702±0.060	1 ction benchmarks with s MAE for QM7 and QM8 07 MolCLR because it a 04 120.344±6.237 104.387±3.292 121.022±5.699 114.4.26±6.591 116.384±8.445	3 caffold split. All experin QM9 datasets, respec Chieves the best results QM8 0.020494.0.00032 0.020584.0.00061 0.022224.0.00045 0.035984.0.00065 0.024194.0.00035	1 nents are run 10 titvely. We repr s 0.00891±0.0 0.00896±0.0 0.01488±0.0 0.01488±0.0
3CL         ACE2         hCYTOX         MERS-PPE_cs         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.666           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Meez           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMol         0.703±0.008         0.669±0.011         0.728±0.001         0.728±0.007         0.865±0.017	Rank       Table S6: The RMSE or using random seeds fron all cr       RMSE       GraphMVP       EdgePred       GraphMVP-C       MolCLR       ImageMol       Mole-BERT       VideoMol	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL t.322±0.062 1.322±0.062 1.367±0.041 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017	1 olecular property predia and Lipo datasets and l/e use GIN backbone fr Lipo 0.773±0.016 0.778±0.013 0.768±0.013 0.768±0.013 0.709±0.018 0.702±0.060 0.727±0.066 0.724±0.009 0.743±0.00 0.743±0.00 0.7	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 104.387±3.292 121.022±5.699 144.426±6.591 116.384±8.445 101.922±2.331 <b>76.736±1.561</b>	3 ccaffold split. All experin ,QM9 datasets, respec chieves the best result 0.02084.0.00061 0.020824.0.00061 0.0202240.00047 0.035984.0.00083 0.024194.0.00033 0.024194.0.00033 0.021990.00033	1 nents are run 1 titvely. We repr 3. 0.00891±0. 0.00892±0. 0.00896±0. 0.02061±0. 0.00910±0. 0.00896±0.
REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.666           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.6651         0.688         0.79         0.734         0.70	Rank         Table S6: The RMSE or using random seeds from all or or using random seeds from all or or or all or or or all or or or all or or all or or all or or all or or or all or or all or or all or or or all or or or all or or or all or or all or or or all or or or all or or or all or or or or or all or	1           MAE performance of diff 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.367±0.041 1.333±0.055 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu	1      Olecular property predix and Lipo datasets and lipo datasets and lipo datasets and lipo     0.773±0.016     0.773±0.016     0.773±0.013     0.769±0.013     0.722±0.060     0.722±0.006     0.743±0.009      nethods on 11 SAF Its are obtained froc	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 104.387±3.292 121.022±5.699 144.426t.591 116.384±6.591 116.384±6.591 76.736±1.561 RS-CoV-2 datasets om ImageMol.	3 ccaffold split. All experim QM9 datasets, respect chieves the best result 0.0204940.00032 0.0204940.00032 0.02058±0.00061 0.0259±0.00045 0.02419±0.00033 0.02173±0.00033 0.02189±0.00020	1 nents are run 11 titvely. We repr 3 QM9 0.00891±0.0 0.00391±0.0 0.00396±0.0 0.00910±0.0 0.00910±0.0 0.00896±0.0 0.00896±0.0
Integration         0.70220.001         0.72120.003         0.7120.009         0.71320.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           ===== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           mageMol         0.703±0.008         0.666±0.011         0.723±0.007         0.86±0.006         0.77	Rank         Table S6: The RMSE or using random seeds from all cr         RMSE         GraphMVP         GraphMVP-C         MolcLR         ImageMol         Mole-BERT         VideoMol	1           MAE performance of diff           0 to 9. We report RMSI           omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL for FreeSolv, ESOL 1.32240.062 1.33340.055 1.462240.068 0.96440.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2	1 olecular property predix and Lipo datasets and lipo datasets and 0.773±0.016 0.778±0.013 0.768±0.013 0.798±0.018 0.722±0.006 0.722±0.006 0.743±0.009 nethods on 11 SAF ts are obtained froc hCYTOX	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 116.384±8.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol. MERS-PPE_cs	3 acaffold split. All experin QM9 datasets, respec CM8 0.0204940.00032 0.0205840.00061 0.0202240.00085 0.0241940.00033 0.0207340.00085 0.0241940.00033 0.027340.00033 0.01890±0.00020 with balanced sca MERS-PPE	1 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.00896±0.0 0.00910±0.0 0.00910±0.0 0.00896±0.0 0.00895±0.0 0.0085±0
==== continue ====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMod         0.703±0.008         0.665±0.011         0.723±0.007         0.865±0.016         0.723±0.007         0.865±0.015	Rank Table S6: The RMSE or using random seeds fron all cr RMSE GraphMVP EdgePred GraphMVP-C MolCLR Mole-BERT VideoMol Table S7: The REDIAL-2020 ImageMol	1           MAE performance of diffind to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753 0.720±0.001	1 0lecular property predia and Lipo datasets and I ve use GIN backbone fi 0.773±0.016 0.773±0.013 0.769±0.013 0.792±0.018 0.722±0.006 0.727±0.006 0.743±0.009 nethods on 11 SAF its are obtained fro hCYTOX 0.7710 0.727±0.009	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 116.384±8.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol. MERS-PPE_cs 0.703 0.711±0.009	3 acaffold split. All experim QM9 datasets, resplet CM8 0.0204940.00032 0.0205840.00065 0.0201940.00033 0.0207340.00033 0.0207340.00033 0.0211940.00032 0.0211940.00033 0.0211940.00033 0.0211940.00033 0.0211940.00032 0.0211940.00033 0.0211940.00032 0.0111940 0.0211940.0003 0.0111940 0.0	1 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.00896±0.0 0.0091±0.0091±0.0 0.0091±0.0 0.00
CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mee           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMod         0.703±0.008         0.665±0.011         0.723±0.007         0.865±0.016         0.77	Rank         Table S6: The RMSE or using random seeds from all cr         RMSE         GraphMVP         EdgePred         GraphMVP-C         MolCLR         ImageMol         Mole.BERT         VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performat 3CL 0.713 0.762±0.007 0.709±0.006	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753 0.720±0.001 0.759±0.025	1 0lecular property prediamonal lipo datasets and lipo datasets an	1 2010 benchmarks with MAE for QM7 and QM8 or MolCLR because it a 0M7 120.344±6.237 104.387±3.292 121.022±5.699 114.4.426±5.591 116.384±8.445 101.922±2.331 76.736±1.561 RS-COV-2 datasets m ImageMol. MERS-PPE_cs 0.703 0.771±0.009 0.828±0.027	3 ccaffold split. All experin QM9 datasets, respec- chieves the best results 0.02049±0.00032 0.02058±0.00085 0.0221+0.00033 0.02073±0.00033 0.02189±0.00020 s with balanced sca MERS-PPE 0.696 0.773±0.011 0.814±0.004	1 0.00891±0.0 0.00891±0.0 0.00929±0.0 0.00939±0.0 0.00939±0.0 0.01488±0.0 0.02061±0.0 0.00910±0.0 0.00910±0.0 0.00939±0.0 0.00896±0.0 0.00895±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0.0085±0.0 0
MagaMal 0.703+0.000 0.001 0.728+0.001 0.733+0.007 0.806+0.006 0.7	Rank         Table S6: The RMSE or using random seeds from all cr         RMSE         GraphMVP         EdgePred         GraphMVP-C         MolCLR         ImageMol         Mole.BERT         VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performation 3CL 0.713 0.762±0.007 0.709±0.006	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.322±0.062 1.332±0.065 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different n compared resu ACE2 0.753 0.720±0.001 0.759±0.025 ===	1 olecular property prediatation fe use GIN backbone for 0.773±0.016 0.773±0.013 0.768±0.013 0.792±0.060 0.722±0.006 0.722±0.009 0.743±0.009 0.743±0.009 0.743±0.009 0.727±0.009 0.727±0.009 0.727±0.009 0.727±0.009 0.725±0.003 == continue ====	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	3 ccaffold split. All experint , QM9 datasets, respec- chieves the best result 0.02049±0.00032 0.02058±0.00065 0.02419±0.00033 0.02073±0.00033 0.02073±0.00033 0.01890±0.00020 with balanced sca MERS-PPE 0.696 0.773±0.011 0.814±0.004	1 hents are run 1 tively. We repr 3. 0.00891±0.1 0.00929±0.0 0.00896±0.1 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00996±0.0 0.00895±0.0 0.00895±0.0 0.0
Inageivor 0.70510.000 0.00910.011 0.72010.001 0.79510.007 0.00010.000 0.7	Rank Table S6: The RMSE or using random seeds fron all c: RMSE GraphMVP EdgePred GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The REDIAL-2020 ImageMol VideoMol EDEDIAL-2020 ImageMol VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.443±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performat 3CL 0.713 0.762±0.007 0.709±0.006 CoV1-PPE 0.665	2 erent methods on 6 m f or FreeSolv, ESOL the same settings. V ESOL t.322±0.062 t.337±0.041 t.333±0.065 t.1462±0.068 0.964±0.067 t.1115±0.017 0.866±0.017 ance of different n compared resu ACE2 0.753 0.720±0.001 0.753720±0.001 0.759±0.025 === CPE 0.651	1 olecular property prediatation lipo datasets and lipo datasets	1 2010 2010 2010 2010 2010 2010 2010 20	3 ccaffold split. All experiin , QM9 datasets, respect Chieves the best result 0.02049±0.00032 0.02058±0.00085 0.022419±0.00033 0.02073±0.00033 0.02073±0.00033 0.01890±0.00020 with balanced sca MERS-PPE 0.696 0.773±0.011 0.814±0.004	1 hents are run 1 tively. We repr 3. 0.00891±0.1 0.00929±0.0 0.00896±0.0 0.02061±0.0 0.00910±0.0 0.00910±0.0 0.00996±0.0 0.00895±0.0 0.0085±0.0 0.0085±0.0085±0.0 0.0085±0.0085±0.0085±0.0085±0.0000000000
VideoMol 0.737±0.007 0.747±0.013 0.761±0.002 0.841±0.004 0.862±0.002 0.78	Rank Table S6: The RMSE or using random seeds fron all c: RMSE GraphMVP-C GraphMVP-C GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The REDIAL-2020 ImageMol VideoMol REDIAL-2020 ImageMol Image	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv           2.559±0.158           2.549±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053           ROC-AUC performation           3CL           0.713           0.762±0.007           0.709±0.006           CoV1-PPE           0.665           0.703±0.008	2 erent methods on 6 m E for FreeSolv, ESOL 1.322±0.062 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different n compared resu ACE2 0.753 0.720±0.001 0.759±0.025 === CPE 0.651 0.669±0.011	1 olecular property prediatation lipo datasets and lipo datase	1 Ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 144.426±6.591 116.384±8.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol. MERS-PPE_cs 0.703 0.771±0.009 0.828±0.027 = AlphaLISA 0.79 0.793±0.007	3 ccaffold split. All experin QM9 datasets, respec chieves the best result 0.02049±0.00032 0.02058±0.00061 0.02022±0.00047 0.03598±0.00085 0.02419±0.00033 0.02419±0.00033 0.02419±0.00033 0.02419±0.00020 a with balanced sca MERS-PPE 0.696 0.773±0.011 0.814±0.004 TruHit 0.734 0.806±0.006	1 hents are run 1 titvely. We repr 3 0.00891±0.1 0.00929±0.0 0.00896±0.1 0.00906±0.1 0.002061±0.1 0.002061±0.1 0.002061±0.1 0.00896±0.1 affold split. A CoV1-PP 0.666 0.775±0 0.836±0 Mear 0.706 0.746
3CL         ACE2         hCYTOX         MERS-PPE_cs         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.60           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.773±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           ===== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Meg           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.77           ImageMol         0.703±0.008         0.66540.001         0.723±0.007         0.868±0.007         0.868±0.007         0.868±0.007	Rank       Table S6: The RMSE or using random seeds from all cr       RMSE       GraphMVP       EdgePred       GraphMVP-C       MolCLR       ImageMol       Mole-BERT	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155	2 erent methods on 6 m f or FreeSolv, ESOL t.322±0.062 1.322±0.062 1.367±0.041 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017	1 olecular property predia and Lipo datasets and I //e use GIN backbone fr Lipo 0.773±0.016 0.778±0.013 0.768±0.013 0.769±0.018 0.709±0.018 0.709±0.060 0.727±0.006 0.727±0.006	1 Ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 104.387±3.292 121.022±5.699 144.42646.591 116.384±8.445 101.922±2.331	3 ccaffold split. All experin QM9 datasets, respec chieves the best result 0.02054.0.00061 0.0202240.00081 0.02419.0.00083 0.02419.0.00033 0.02419.0.00033	1 nents are run titvely. We rep 0.00891±0 0.00929±0 0.00896±0 0.01488±0 0.02261±0 0.02261±0 0.02261±0 0.02061±0
Table Sr: The ROC-AUC performance or different methods on 11 SARS-CoV-2 datasets with balanced scatfold split. compared results are obtained from ImageMol.           Compared results are obtained from ImageMol.           ACE2         hCYTOX         MERS-PPE_cs         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.66           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           E==== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.77	Rank       Table S6: The RMSE or using random seeds from all cr       RMSE       GraphMVP       EdgePred       GraphMVP-C       MolcLR       ImageMol       Mole-BERT       VideoMol	1           MAE performance of diff           0 to 9. We report RMSI           omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m f for FreeSolv, ESOL for FreeSolv, ESOL 1.322±0.062 1.332±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017	1 Olecular property predix and Lipo I/73±0.016 0.773±0.016 0.778±0.013 0.768±0.013 0.769±0.018 0.722±0.060 0.722±0.006 0.743±0.009	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 104.387±3.292 121.022±5.699 116.384±8.445 101.922±2.331 76.736±1.561	3 ccaffold split. All experin QM9 datasets, respect QM8 0.0204940.00032 0.02058±0.00061 0.02022±0.00047 0.03598±0.00085 0.02419±0.00033 0.02073±0.00033 0.01890±0.00020	1 Annents are run 11 titvely. We repr S QM9 0.0089±0.0 0.0089±0.0 0.01488±0.0 0.02061±0.0 0.00910±0.0 0.00896±0.0
REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.661           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           E==== continue ====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mere           REDIAL-2020         0.665         0.6651         0.688         0.79         0.734         0.70           maceMol         0.703±0.008         0.665±0.011         0.723±0.007         0.86±0.006         0.773	Rank       Table S6: The RMSE or using random seeds from all cr       RMSE       GraphMVP       EdgePred       GraphMVP-C       MolCLR       ImageMol       Mole-BERT       VideoMol	1 MAE performance of diff 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.367±0.041 1.333±0.055 0.964±0.067 1.115±0.017 0.866±0.017	1 Olecular property predia and Lipo datasets and l /e use GIN backbone fr Lipo 0.773±0.016 0.778±0.013 0.768±0.013 0.799±0.018 0.729±0.008 0.722±0.006 0.722±0.006 0.743±0.009	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 104.387±3.292 121.022±5.699 144.4266.591 116.384±8.445 101.922±2.331 <b>76.736±1.561</b>	3 ccaffold split. All experim QM9 datasets, respec chieves the best result QM8 0.0204940.00032 0.0204940.00032 0.025940.00061 0.0359840.00085 0.02419±0.00033 0.02073±0.00033 0.01890±0.00020	1 nents are run 1 titvely. We repr 3 0.009891±0. 0.00896±0. 0.0091±0. 0.0091±0. 0.0091±0. 0.00996±0.
3CL         ACE2         hCYTOX         MERS-PPE_cs         MERS-PPE         CoV1-PI           REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.66           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.755±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           E==== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.77           ImageMol         0.703±0.008         0.6964±0.011         0.728±0.007         0.86±0.006         0.73	Rank       Table S6: The RMSE or using random seeds from all cr       RMSE       GraphMVP       EdgePred       GraphMVP-C       MolCLR       ImageMol       Mole-BERT       VideoMol	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.769±0.158           2.113±0.235           2.138±0.053	2 erent methods on 6 m f or FreeSolv, ESOL g the same settings. V ESOL 1.367±0.041 1.3340.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017  Dance of different n	1 olecular property predia and Lipo datasets and lipo datasets and live os GIN backbone fr Lipo 0.778±0.016 0.778±0.013 0.768±0.013 0.768±0.013 0.709±0.018 0.727±0.060 0.727±0.060 0.724±0.009 nethods on 11 SAF	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 144.426±6.591 116.384±8.445 101.922±2.331 76.736±1.561 35-CoV-2 datasets	3 ccaffold split. All experin QM9 datasets, respect chieves the best result 0.02049±0.00032 0.02058±0.00061 0.02022±0.00047 0.03598±0.00085 0.02419±0.00033 0.02419±0.00033 0.02419±0.00033	1 nents are run 1 titvely. We repr 3 QM9 0.00891±0. 0.00999±0. 0.00896±0. 0.02061±0. 0.002061±0. 0.00910±0. 0.00910±0. 0.00896±0. affold split 6
REDIAL-2020         0.713         0.753         0.710         0.703         0.696         0.66           ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMol         0.703±0.008         0.668±0.011         0.728±0.001         0.723±0.007         0.865±0.016         0.73	Rank         Table S6: The RMSE or using random seeds from all or         RMSE         GraphMVP         EdgePred         GraphMVP-C         MolCLR         ImageMol         Mole-BERT         VideoMol	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL g the same settings. V ESOL 1.367±0.041 1.332±0.062 1.367±0.041 1.333±0.055 1.462±0.067 1.115±0.017 0.866±0.017 ence of different m compared resu	1      olecular property predia and Lipo datasets and live of a constraint of the set of the s	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 144.426±6.591 116.384±8.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol.	3 ccaffold split. All experin QM9 datasets, respec chieves the best result 0.02049±0.00032 0.02058±0.00061 0.02022±0.00047 0.03598±0.00085 0.02419±0.00033 0.02419±0.00033 0.02419±0.00033 0.02193±0.00020	1 nents are run 1 tively. We repr 3 QM9 0.00891±0. 0.00995±0. 0.00295±0. 0.00295±0. 0.00295±0. 0.00910±0. 0.00936±0. 0.00896±0. 0.00896±0.
ImageMol         0.762±0.007         0.720±0.001         0.727±0.009         0.771±0.009         0.773±0.011         0.775±           VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           Emergina colspan="4">Emergina colspan="4">Emergina colspan="4">Emergina colspan="4">Colspan= 4">Colspan= 4"	Table S6: The RMSE or using random seeds from all cr         RMSE         GraphMVP         EdgePred         GraphMVP-C         MolCLR         ImageMol         Mole.BERT         VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performation	2 erent methods on 6 m f for FreeSolv, ESOL g the same settings. V ESOL 1.322±0.062 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different n compared resu	1 olecular property predia and Lipo datasets and live use GIN backbone fr Lipo 0.773±0.016 0.778±0.013 0.768±0.013 0.768±0.013 0.709±0.018 0.722±0.006 0.722±0.006 0.723±0.009 nethods on 11 SAF	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 121.022±5.699 144.426±6.591 116.324±8.445 101.922±2.331 <b>76.736±1.561</b> RS-CoV-2 datasets m ImageMol.	3 ccaffold split. All experin , QM9 datasets, respec chieves the best result: 0.02049±0.00032 0.02058±0.00061 0.02022±0.00047 0.0359±0.00033 0.02419±0.00033 0.02419±0.00033 0.02199±0.00020 with balanced sca	1 nents are run 1 tively. We repr 3 QM9 0.00891±0. 0.00929±0. 0.00896±0. 0.02061±0. 0.00910±0. 0.00936±0. 0.00896±0. 0.00896±0. 0.00896±0. 0.00896±0.
Indgeneric         C.10120.001         C.12120.001         C.112120.001         C.110120.011         C.110120.011 <td>Rank Table S6: The RMSE or using random seeds fron all or RMSE GraphMVP GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The</td> <td>1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performation 3CL</td> <td>2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ence of different m compared resu ACE2</td> <td>1 0lecular property prediated and Lipo datasets and J le use GIN backbone fi 0.773±0.016 0.773±0.018 0.792±0.018 0.792±0.018 0.792±0.018 0.722±0.006 0.743±0.009 0.743±0.009 nethods on 11 SAF</td> <td>1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.34446.237 104.38743.292 121.022±5.699 116.38448.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol. MERS-PPE_cs</td> <td>3 ccaffold split. All experim QM9 datasets, respec- chieves the best results 0.0204940.00032 0.0205840.00065 0.0201940.00033 0.0207340.00033 0.0207340.00033 0.021890±0.00020 with balanced sca MERS-PPE</td> <td>1 0.00929±0. 0.00939±0. 0.00939±0. 0.00939±0. 0.00939±0. 0.00939±0. 0.00930±0. 0.00910±0. 0.00930±0. 0.00030±0. 0.00000±0.00000±0. 0.00000±0. 0.0000±</td>	Rank Table S6: The RMSE or using random seeds fron all or RMSE GraphMVP GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performation 3CL	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ence of different m compared resu ACE2	1 0lecular property prediated and Lipo datasets and J le use GIN backbone fi 0.773±0.016 0.773±0.018 0.792±0.018 0.792±0.018 0.792±0.018 0.722±0.006 0.743±0.009 0.743±0.009 nethods on 11 SAF	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.34446.237 104.38743.292 121.022±5.699 116.38448.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol. MERS-PPE_cs	3 ccaffold split. All experim QM9 datasets, respec- chieves the best results 0.0204940.00032 0.0205840.00065 0.0201940.00033 0.0207340.00033 0.0207340.00033 0.021890±0.00020 with balanced sca MERS-PPE	1 0.00929±0. 0.00939±0. 0.00939±0. 0.00939±0. 0.00939±0. 0.00939±0. 0.00930±0. 0.00910±0. 0.00930±0. 0.00030±0. 0.00000±0.00000±0. 0.00000±0. 0.0000±
VideoMol         0.709±0.006         0.759±0.025         0.765±0.003         0.828±0.027         0.814±0.004         0.836±           ===== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMol         0.703±0.008         0.669±0.011         0.723±0.007         0.865±0.006         0.77	Rank         Table S6: The RMSE or using random seeds from all cr         RMSE         GraphMVP         EdgePred         GraphMVP-C         MolCLR         ImageMol         Mole-BERT         VideoMol	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053	2 erent methods on 6 m for FreeSolv, ESOL t = 50L 1.367±0.062 1.367±0.041 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753	1      olecular property predia and Lipo datasets and 1 //e use GIN backbone for      Lipo     0.773±0.016     0.779±0.013     0.769±0.013     0.799±0.018     0.727±0.066     0.727±0.066     0.727±0.069     0.743±0.009      hethods on 11 SAF Its are obtained froc     hCYTOX     0.710	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 144.426±6.591 116.384±8.445 101.922±2.331 76.736±1.561 RS-CoV-2 datasets m ImageMol. MERS-PPE_cs 0.703	3 ccaffold split. All experin ,QM9 datasets, respec chieves the best result 0.02049±0.00032 0.02058±0.00061 0.02022±0.00047 0.03598±0.00085 0.02419±0.00033 0.02419±0.00033 0.02419±0.00033 0.02199±0.00020 c with balanced sca MERS-PPE 0.696	1 nents are run 1 titvely. We repr 3 0.00929±0 0.00896±0 0.002061±0 0.002061±0 0.002061±0 0.009010±0 0.00906±0 0.00896±0 0.0085±0 0.0086±0 0.0085±0 0.0085±0 0.0085±0 0.0085±0 0.0085±0 0.0085±0
===== continue =====           CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMol         0.70340         0.08         0.69040         0.11         0.72340         0.07         0.86640         0.65	Rank         Table S6: The RMSE or using random seeds from all cr         RMSE         GraphMVP         EdgePred         GraphMVP-C         MolCLR         ImageMol         Mole-BERT         VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performat 3CL 0.713 0.762±0.007	2 erent methods on 6 m E for FreeSolv, ESOL g the same settings. V ESOL 1.322±0.062 1.332±0.062 1.333±0.055 1.362±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different n compared resu ACE2 0.753 0.720±0.001	1 0lecular property prediatation of the second seco	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a <u>QM7</u> 120.344±6.237 121.022±5.699 144.426±6.591 145.344±6.445 101.922±2.331 <b>76.736±1.561</b> RS-CoV-2 datasets sm ImageMol. MERS-PPE_cs 0.771±0.009	3 ccaffold split. All experin , QM9 datasets, respec chieves the best result: 0.02049±0.00032 0.02058±0.00061 0.02022±0.00047 0.0396±0.00033 0.02419±0.00033 0.02419±0.00033 0.02193±0.00033 0.01890±0.00020 cwith balanced sca	1 hents are run 1 tively. We repr 3. 0.00891±0. 0.00892±0. 0.00896±0. 0.02061±0. 0.00910±0. 0.00910±0. 0.00896±0. 0.00895±0. 0
CoV1-PPE         CPE         Cytotox         AlphaLISA         TruHit         Mea           REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMed         0.703±0.008         0.669±0.011         0.728±0.001         0.793±0.007         0.866±0.006         0.77	Rank Table S6: The RMSE or using random seeds fron all co RMSE GraphMVP EdgePred GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol REDIAL-2020 ImageMol VideoMol VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.843±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performation 3CL 0.713 0.762±0.007 0.709±0.006	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753 0.720±0.001 0.759±0.025	1 0lecular property prediated and Lipo datasets and J le use GIN backbone fi 0.773±0.016 0.773±0.013 0.768±0.013 0.762±0.060 0.722±0.006 0.743±0.009 0.743±0.009 nethods on 11 SAF lts are obtained frc hCYTOX 0.710 0.727±0.009 0.765±0.003	1 2010 benchmarks with s WAE for QM7 and QM8 or MolCLR because it a 2017 104.387±3.292 121.022±5.699 1143.426£.591 116.384±8.445 101.922±2.331 76.736±1.561 276.736±1.561 276.736±1.561 20.771±0.009 0.828±0.027	3 ccaffold split. All experin QM9 datasets, respec- chieves the best results 0.02049±0.00032 0.02058±0.00065 0.02219±0.00033 0.02073±0.00033 0.0219±0.00033 0.01890±0.00020 with balanced sc: MERS-PPE 0.696 0.773±0.011 0.814±0.004	1 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.00896±0.0 0.01488±0.0 0.02061±0.0 0.00910±0.0 0.00910±0.0 0.009306±0.0 0.009306±0.0 0.00836±0.0 0.0306±0.0 0.0306±0.0 0.
REDIAL-2020         0.665         0.651         0.688         0.79         0.734         0.70           ImageMol         0.703±0.008         0.669±0.011         0.728±0.001         0.793±0.007         0.806±0.006         0.72	Rank         Table S6: The RMSE or using random seeds from all considered and the seeds from all constraints of the seeds from all constrates of the seed	1           MAE performance of diffind to be with the second state of the sec	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753 0.720±0.001 0.759±0.025 ====	1 0lecular property prediated and Lipo datasets and J le use GIN backbone fi 0.773±0.016 0.773±0.013 0.769±0.013 0.792±0.018 0.702±0.060 0.727±0.006 0.743±0.009 0.743±0.009 nethods on 11 SAF lts are obtained frc hCYTOX 0.710 0.727±0.009 0.765±0.003 == continue ====	1 ction benchmarks with s MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 116.384±8.445 101.922±2.331 76.736±1.561 76.736±1.561 MERS-PPE_cs 0.703 0.771±0.009 0.828±0.027	3 acaffold split. All experim QM9 datasets, respec chieves the best results 0.02049±0.00032 0.02058±0.00085 0.022±0.00047 0.03598±0.00085 0.02419±0.00033 0.02073±0.00033 0.01890±0.00020 active the balanced sca MERS-PPE 0.696 0.773±0.0111 0.814±0.004	1 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.0091±0.00000000000000000000000000
	Rank Table S6: The RMSE or using random seeds from all cr RMSE GraphMVP EdgePred GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The REDIAL-2020 ImageMol VideoMol	1 MAE performance of diff n 0 to 9. We report RMSI omparison methods using FreeSolv 2.559±0.158 2.43±0.091 2.766±0.199 3.112±0.638 2.113±0.235 2.988±0.155 1.728±0.053 ROC-AUC performat 3CL 0.713 0.762±0.007 0.709±0.006 CoV1-PPE	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different n compared resu ACE2 0.753 0.720±0.001 0.759±0.025 === CPE 0.551	1 olecular property prediated and Lipo datasets and I fe use GIN backbone fi 0.773±0.016 0.773±0.013 0.768±0.013 0.768±0.013 0.792±0.060 0.727±0.006 0.743±0.009 0.743±0.009 0.743±0.009 0.727±0.009 0.7	1 2010 benchmarks with MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 114.4.26±6.591 116.384±8.445 101.922±2.331 76.736±1.561 76.736±1.561 RS-COV-2 datasets m ImageMol. MERS-PPE_cs 0.703 0.771±0.009 0.828±0.027 = AlphaLISA	3 ccaffold split. All experin QM9 datasets, respec- chieves the best results 0.02049±0.00032 0.02058±0.00085 0.0221+0.00033 0.02073±0.00033 0.02189±0.00020 s with balanced sca MERS-PPE 0.696 0.773±0.011 0.814±0.004 TruHit	1 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.00896±0.0 0.02061±0.0 0.00910±0.0 0.00910±0.0 0.00930±0.0 0.00896±0.0 0.0080±0.0080±0.0 0.0080±0.0 0.0080±0.0 0.0080±0.0 0.0080±0.0 0.0080±
VideoMol 0.737±0.007 0.747±0.013 0.761±0.002 0.841±0.004 0.862±0.002 0.74	Rank Table S6: The RMSE or using random seeds fron all co RMSE GraphMVP EdgePred GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The REDIAL-2020 ImageMol VideoMol REDIAL-2020 ImageMol REDIAL-2020 ImageMol	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053           ROC-AUC performation           3CL           0.713           0.762±0.007           0.709±0.006           CoV1-PPE           0.665           0.703±0.008	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753 0.720±0.001 0.759±0.025 === CPE 0.6651 0.669±0.011	1 olecular property prediated and Lipo datasets and I fe use GIN backbone fi 0.773±0.016 0.773±0.013 0.768±0.013 0.768±0.013 0.702±0.060 0.722±0.006 0.743±0.009 0.743±0.009 0.743±0.009 0.727±0.009 0.727±0.009 0.727±0.009 0.727±0.009 0.727±0.009 0.728±0.001	1 2010 benchmarks with a MAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 114.4.426±5.591 116.384±8.445 101.922±2.331 76.736±1.561 76.736±1.561 RS-COV-2 datasets m ImageMol. MERS-PPE_cs 0.703 0.771±0.009 0.828±0.027 2 AlphaLISA 0.79 0.7340.007	3 ccaffold split. All experin QM9 datasets, respec- chieves the best results 0.02049±0.00032 0.02058±0.00065 0.0221+0.00033 0.02073±0.00033 0.02189±0.00020 s with balanced sca MERS-PPE 0.696 0.773±0.011 0.814±0.004 TruHit 0.734 0.806±0.006	1 0.00891±0.0 0.00891±0.0 0.00891±0.0 0.00931±0.0 0.0091±0.0 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00910±0.0 0.00896±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00895±0.0 0.00805±0.0 0.0
Mindguide         0.7001000         0.5001001         0.7210.001         0.7010100         0.7010000         0.701000         0.70100000         0.70100000         0.70100000         0.70100000         0.70100000         0.70100000         0.70100000         0.70100000         0.70100000	Rank Table S6: The RMSE or using random seeds fron all co RMSE GraphMVP EdgePred GraphMVP-C MolCLR ImageMol Mole-BERT VideoMol Table S7: The REDIAL-2020 ImageMol VideoMol REDIAL-2020 ImageMol NoteeMol REDIAL-2020 ImageMol NoteeMol REDIAL-2020 ImageMol NoteeMol REDIAL-2020 ImageMol	1           MAE performance of diff n 0 to 9. We report RMSI omparison methods using           FreeSolv           2.559±0.158           2.843±0.091           2.766±0.199           3.112±0.638           2.113±0.235           2.988±0.155           1.728±0.053           ROC-AUC performation           3CL           0.713           0.762±0.007           0.709±0.006           CoV1-PPE           0.665           0.703±0.008	2 erent methods on 6 m for FreeSolv, ESOL the same settings. V ESOL 1.322±0.062 1.333±0.055 1.462±0.068 0.964±0.067 1.115±0.017 0.866±0.017 ance of different m compared resu ACE2 0.753 0.720±0.001 0.759±0.025 === CPE 0.6651 0.669±0.011	1 olecular property prediated and Lipo datasets and I fe use GIN backbone fi 0.773±0.016 0.773±0.013 0.768±0.013 0.768±0.013 0.762±0.060 0.722±0.006 0.743±0.009 0.743±0.009 0.743±0.009 0.722±0.009 0.722±0.009 0.722±0.009 0.722±0.009 0.722±0.009 0.722±0.009 0.722±0.009 0.728±0.001	1 2010 benchmarks with a WAE for QM7 and QM8 or MolCLR because it a QM7 120.344±6.237 104.387±3.292 121.022±5.699 114.4.426±5.591 116.384±8.445 101.922±2.331 76.736±1.561 76.736±1.561 RS-COV-2 datasets m ImageMol. MERS-PPE_cs 0.703 0.771±0.009 0.828±0.027 2 AlphaLISA 0.79 0.7340.007	3 ccaffold split. All experin QM9 datasets, respec- chieves the best results 0.02049±0.00032 0.02058±0.00065 0.02219.00033 0.02073±0.00033 0.02189±0.00020 s with balanced sc: MERS-PPE 0.696 0.773±0.011 0.814±0.004 TruHit 0.734 0.806±0.006	1 0.00891±0. 0.00991±0. 0.00991±0. 0.00991±0. 0.0091±0. 0.00910±0. 0.0000±0. 0.0000±0. 0.0000±0. 0.0000±0. 0.0

bias and skewness of the bootstrap parameter estimates by incorporating a bias-correction factor and an acceleration factor. The results of the uncertainty interval are reported in the Extended Table 1 (the Revised Supplemental Table 8) and Extended Table 2 (the Revised Supplemental Table 9) below, which shows the effectiveness of VideoMol with an average improvement ranging from 5.44% to 10.07%.

In summary, these new experiments highlight the robustness of our VideoMol models. We have added these new experiments and more detailed explanations in the revised manuscript.

**Extended Table 1** (the Revised Supplemental Table 8). The uncertainty intervals with 95% confidence intervals of ImageMoI and VideoMoI on 10 compound-kinase interaction datasets.  $UI(\cdot)$  represents the uncertainty intervals and "Improvement" represents the relative performance improvement of VideoMoI compared to ImageMoI.

		UI(RMSE)			UI(MAE)	
Dataset	ImageMol	VideoMol	Improvem ent	ImageMol	VideoMol	Improvem ent
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%

**Extended Table 2** (the Revised Supplemental Table 9). The uncertainty intervals with 95% confidence intervals of ImageMol and VideoMol on 11 SARS-CoV-2 viral activity prediction datasets.  $UI(\cdot)$  represents the uncertainty intervals and "Improvement" represents the relative performance improvement of VideoMol compared to ImageMol.

Detect		UI(AUC)	
Dataset	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±0.074	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 TruHit	0.762±0.062 0.772±0.059	0.836±0.049 0.855±0.046	9.71% 10.75%
	References [1] Efron B. Better <i>statistical Associat</i> [2] Efron B, Tibshir Hall/CRC, 1994.	bootstrap confidenc tion, 1987, 82(397): rani R J. An introduc	e intervals[J]. <i>Jouri</i> 171-185. tion to the bootstrap	nal of the American p[M]. Chapman and
Excerpt from Revised Manuscript	For 10 compound- performance than cyclinD3 (AUC=0. 0.848±0.027), FG FGFR4 (AUC=0 (AUC=0.867±0.030 performance impro Supplementary Ta art methods of Ima and 20.6%.	kinase interaction da other methods acr 972±0.039), EGFR FR2 (AUC=0.988± 0.852±0.080), FL 6) and MET (AU ovement of 5.9% rar ble 3). In particular, ageMol and MolCLF	atasets, VideoMol at oss BTK (AUC=0.8 (AUC=0.905±0.01 0.017), FGFR3 (A T3 (AUC=0.981: JC=0.963±0.026) nging from 1.8% to 2 VideoMol outperform R with average imp	chieves better AUC 361±0.023), CDK4- 7), FGFR1 (AUC= AUC=0.896±0.039), ±0.026), KPCD3 with an average 20.3% ( <b>Fig. 2a</b> and ms the state-of-the- rovements of 6.7%
	In classification characteristic (RO across BBBP (AUC=79.4%±0.5) ToxCast (AUC=66 Supplementary Ta values across Free Lipo (RMSE=0 (MAE=0.01890±0. other methods ( <b>Fig</b>	task, using the a C) curve (AUC), Vid (AUC=70.7%±1.5), , BACE (AUC=82. 5.7%±0.5), outperfo able 5). In regressio eSolv (RMSE=1.728 .743±0.009), QM 0020) and QM9 (M g. 2d and Suppleme	area under the i eoMol achieves ele Tox21 (AUC= 4%±0.9), SIDER orming other meth on task, VideoMol ±0.053), ESOL (RM M7 (MAE=76.73 AE=0.00896±0.000 ntary Table 6).	receiver operating vated performance 78.8%±0.4), HIV (AUC=66.3%±0.9), ods ( <b>Fig. 2c</b> and achieves low error <i>I</i> SE=0.866±0.017), 6±1.561), QM8 03), outperforming
	We found that (3CL=0.709±0.006 PPE_cs=0.828±0.0 PPE_cs=0.836±0.0 AlphaLISA=0.841± improvement rang average 8.1% imp REDIAL-2020 (Fig	VideoMol achieved 5, ACE2=0.759±0.0 027, MERS-PPE=0. 029, CoV1-PPE= 0.004, TruHit=0.8 ing from 3.3% to 7 provement ranging <b>1. 2e</b> and Supplement	d elevated ROC- 20, hCYTOX=0.7 814±0.004, CPE=0 0.737±0.007, Cyt 62±0.002) with a .8% compared with from 0.6% to 17.5 ntary Table 7).	AUC performance 65±0.003, MERS- .747±0.013, CoV1- totox=0.761±0.002, in average 3.9% ImageMol and an 5% compared with
	Furthermore, we of intervals (CI) of interaction dataset details, we used th intervals <sup>41,42</sup> to ca intervals (CI), which parameter estimate acceleration factor effectiveness of Vi	calculated the unce ImageMol and Vi s and 11 SARS-Cov the popular bias-corre- alculate the uncert ch corrects for both ted by incorporation or. The results of deoMol with an aver-	rtainty intervals wit deoMol using 10 V-2 viral activity pre ected and accelerat ainty intervals with bias and skewnes ng a bias-correction the uncertainty rage improvement r	th 95% confidence compound-kinase diction datasets. In ed (BCa) bootstrap n 95% confidence as of the bootstrap on factor and an interval show the ranging from 5.44%

to 10.07% (Supplementary Tables 8-9).
References [41] Efron, B. Better bootstrap confidence intervals. Journal of the American statistical Association 82, 171-185 (1987). [42] Efron, B. & Tibshirani, R.J. An introduction to the bootstrap. (Chapman and Hall/CRC, 1994).

## Ref 1.4 – "Intuition on why videos make more sense and contribute to learning better features" –

Reviewer Comment	It will help if the authors include some intuition on why videos make more sense and contribute to learning better features. It is not so easy to understands for readers that are less experienced with conformers. Some running/motivating examples could help the presentation.					
Author Response	We thank the reviewer for this great point and we have added more intuition and examples on why videos make more sense and contribute to learning better features in the revised Results and Discussion.					
Excerpt from Revised Manuscript	Results: Framework of VideoMol					
	Molecules exist in nature and are constantly conformational dynamics, making video the most direct representation method. The molecular 3D information can be directly observed from the video without the help of manual feature extraction, such as the distance between pairs of atoms and the angle formed between multiple atoms and so on. In addition, we evaluated the advantages of different representations in feature extraction capabilities and found that our proposed video representation has obvious advantages over existing representations with a 66% improvement rate on 8 basic attributes (Supplementary Methods and Supplementary Table 1). Therefore, these significant differences motivate us to develop VideoMol for accurately predicting the targets and properties of molecules in the form of videos derived from molecules.					
		modality	model	use conformer?	prop	
			GCN	×	62.304	
	anont to	2D graph	GIN	×	62.980	
	graph-based		EGNN	×	17.418	
		3D graph	EGNN	$\checkmark$	16.684	
		image from imagemol	ResNet18	×	12.469	
	income to a set	video-1frame	ResNet18	$\checkmark$	11.237	
	image-based	video-5frame	ResNet18		8.088	
		video-60frame	ViT		7.511	
C.2 Results of different representations on 8 basic attributes To fairly compare the effects of different representations, we evaluate					tes e 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 representation used by ImageMoI) 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 randomly collected 10,000 molecules from the pre-training dataset and used exactly the same experimental setup 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 3D representation. Specifically, compared with the second-place ImageMol without pre-training, the performance of video-1frame improved by 11%. When we utilized all video frames (video-60frame), the performance is further significantly improved from 12.47 to 7.55 with a 66% 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 inceasing the number of 3D frames and integrating other types of 3D representation (such as AlphaFold3 <sup>11</sup> ) in the near future.
	[11] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. Nature, 2024: 1-3.

Reviewer Comment	How sensitive is the framework to the source for video generation, in this case RDKit. I assume it is quite dependent on this platform, and thus, VideoMol probably does not allow for mixing videos from different sources. Does this pose any kind of limitation in real-world applications?
Author Response	<ul> <li>We thank the reviewer for this critique. To evaluate the sensitivity of VideoMol to video generation sources, we utilized two additional methods to generate molecular videos as below:</li> <li>1. OpenBabel<sup>[1]</sup>: It is a chemical toolbox designed to code many languages of chemical data, which generates 3D conformer by four steps: (1) Use the OBBuilder to create a 3D structure using rules and fragment templates; (2) Use 250 steps of a steepest descent geometry optimization with the MMFF94 forcefield; (3) Use 200 iterations of a Weighted Rotor conformational search (optimizing each conformer with 25 steps of a steepest descent); (4) Use 250 steps of a conjugate gradient geometry optimization.</li> <li>2. DeepChem<sup>[2]</sup>: It aims to provide a high quality open-source toolchain that</li> </ul>

democratizes the use of deep-learning in drug discovery, materials science, quantum chemistry, and biology. It uses three steps to generate molecular conformer: (1) Generate a pool of conformers using UFF force field; (2) Minimize conformers; (3) Prune conformers using an RMSD threshold.

As shown in the revised **Extended Table 1** (the Revised Supplemental Table 27), we found that the video generation source has no significant impact on VideoMol with an average performance of 0.755±0.068 (Openbabel), 0.755±0.072 (DeepChem), 0.742±0.064 (RDKit) in RMSE metric and 0.581±0.057 (Openbabel), 0.576±0.060 (DeepChem), 0.565±0.053 (RDKit) performance in MAE metric. Therefore, VideoMol has low sensitivity to video generation sources.

**Extended Table 1** (the Revised Supplemental Table 27). The performance of different video generation source on 10 kinases datasets with balanced scaffold split.

	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±0.0
DeepChem	0.718±0.006	0.557±0.009	0.790±0.016	0.595±0.019	0.677±0.014	0.504±0.0
RDKit	0.708±0.017	0.547±0.015	0.775±0.017	0.577±0.009	0.655±0.007	0.496±0.0
		-	—— continue —			
	4. AA	A2AR	5. A	A3R	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±0.0
DeepChem	0.703±0.005	0.558±0.004	0.798±0.019	0.620±0.015	0.896±0.013	0.708±0.0
RDKit	0.712±0.011	0.543±0.005	0.786±0.006	0.617±0.004	0.864±0.005	0.679±0.0
		-	—— continue —			
	7. D	RD2	8. D	RD3	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±0.0
DeepChem	0.745±0.005	0.551±0.002	0.728±0.014	0.556±0.008	0.666±0.005	0.505±0.0
RDKit	0.742±0.004	0.556±0.005	0.715±0.014	0.554±0.012	0.668±0.008	0.506±0.0
		-	— continue —			
	10. C	PRM	Mean			
	RMSE	MAE	RMSE	MAE		
Openbabel	0.776±0.008	0.590±0.001	0.755±0.068	0.581±0.057		
DeenChom	0.825±0.008	0.607±0.009	0.755±0.072	0.576±0.060		
Deeponem			0 742+0 064	0 565+0 053		

[2] Altae-Tran H, Ramsundar B, Pappu A S, et al. Low data drug discovery

	with one-shot learning[J]. <i>ACS Central Science</i> , 2017, 3(4): 283-293. URL: <u>https://github.com/deepchem/deepchem</u>
Excerpt from Revised Manuscript	Ablation study: Sensitivity of VideoMol for video generation source. To verify the sensitivity of VideoMol to video generation sources, we used two additional platforms to generate molecular videos, which are OpenBabel <sup>58</sup> and DeepChem <sup>59</sup> . We found that the video generation source of different platforms has no significant impact on VideoMol with an average performance of 0.755±0.068 (Openbabel), 0.755±0.072 (DeepChem), 0.750±0.065 (RDKit) in RMSE metric and 0.581±0.057 (Openbabel), 0.576±0.060 (DeepChem), 0.572±0.056 (RDKit) in MAE metric (Supplementary Table 27). Therefore, VideoMol has low sensitivity to video generation sources from different platforms. References
	<ul> <li>[58] O'Boyle, N.M. et al. Open Babel: An open chemical toolbox. 3, 1-14 (2011).</li> <li>[59] Altae-Tran, H., Ramsundar, B., Pappu, A.S. &amp; Pande, V. Low data drug discovery with one-shot learning. ACS central science 3, 283-293 (2017).</li> </ul>

### Comments on code availability - "The provided code is clean and well organized" -

Reviewer Comment	The provided code is clean and well organized. The authors also included a docker image for setting up the environment. The code for training the model as well as reproducing the results is included. I believe only the code for VideoMol is provided, and not for the baseline methods. Not necessary, but if it is easy to include the related methods, the community might appreciate a full testbed.
Author Response	We thank the reviewer for checking our codes. We have provided codes for the baseline methods and related methods/models as well in this link (https://1drv.ms/f/s!Atau0ecyBQNTgTd736-8RPWEXSVt?e=DkOyw2). Of course, you can also access it via our github repository (https://github.com/ChengF-Lab/VideoMoI), as shown below: The code for other comparison methods can be accessed through this link. <b>Reference</b> [1] Landrum G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling[J]. Greg Landrum, 2013, 8: 31. [2] DeLano W L. Pymol: An open-source molecular graphics tool[J]. CCP4 Newsl. Protein Crystallogr, 2002, 40(1): 82-92. [3] Hu W, Fey M, Ren H, et al. Ogb-lsc: A large-scale challenge for machine learning on graphs[J]. arXiv preprint arXiv:2103.09430, 2021.

### **Responses to the Reviewer #3**

0.0.0	
Reviewer Comment	The manuscript presents a video-based pretrained model that can be used to make downstream task predictions across multiple tasks with finetuning. The results are an improvement upon authors' previous work, ImageMol. The improvement is tied to use of video, which can be considered as an augmentation to static images, as well as the use of more comprehensive fingerprints that provide chemical, pharmacological and physicochemistry information. The authors show, through extensive testing, that the new model outperforms the previous and is at least as good or as better as some SOTA models for different tasks. Impact to field: The work is <b>valuable to the field in multiple areas</b> : it is a demonstration of technology transfer from video representation learning. It shows new self supervision tasks that are meaningful for molecule structure videos.
Author Response	We thank the Reviewer for great summary and his/her support on the important value of our proposed VideoMol in multiple drug discovery tasks.

### Overall Summary – "The work is valuable to the field in multiple areas" –

### Ref 2.1 – "More explanations about dynamics" –

Reviewer Comment	Unlike stated, the model does not capture a dynamic conformation of the molecule. The videos are not generated to represent any physical dynamics, or conformer change, or changes to torsion angles etc. They are movies with standardized rotations around given axis. As such, they are only augmentations to enrich the model input about the 3D structure of the molecule. Authors should consider another wording than dynamics to prevent misleading the reader.
Author Response	We thank the reviewer for these critiques. We agreed with the reviewer that the current ViodeMol framework cannot capture the physical dynamics or conformational changes of ligand-receptor dynamics. Integrating physical dynamics or conformational changes from 3D ligand-receptor structures or models (i.e., alphaFold3 [1]) may improve performance of ViodeMol in the future. We have changed our original claims and added more explanations in the revised manuscript.
	Reference [1] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. Nature, 2024: 1-3.
Excerpt from	Results:
Manuscript	Framework of VideoMol
	Molecules exist in nature and are constantly conformational dynamics,

making video the most direct representation method. The molecular 3D information can be directly observed from the video without the help of manual feature extraction, such as the distance between pairs of atoms and the angle formed between multiple atoms and so on. In addition, we evaluated the advantages of different representations in feature extraction capabilities and found that our proposed video representation has obvious advantages over existing representations with a 66% improvement rate on 8 basic attributes (Supplementary Section C.2 and Supplementary Table 1). Therefore, these significant differences motivate us to develop VideoMol for accurately predicting the targets and properties of molecules in the form of videos derived from molecules.

### 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 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 randomly collected 10,000 molecules from the pre-training dataset and used exactly the same experimental setup 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 3D representation. Specifically, compared with the second-place ImageMol without pre-training, the performance of video-1frame improved by 11%. When we utilized all video frames (video-60frame), the performance is further significantly improved from 12.47 to 7.55 with a 66% 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 inceasing the number of 3D frames and integrating other types of 3D representation (such as AlphaFold311) in the near future.

[11] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. Nature, 2024: 1-3.

#### Discussion

Using a simple extension to VideoMol, we can allow the model to learn the correlations and variances between different conformations in the same molecule from videos of dynamic changes, thereby further playing an important role in molecular dynamics scenarios.

We believe that it is promising to represent molecules and perform inferences through videos as molecular imaging techniques continue to

advance. In summary, the introduction of VideoMol on the one hand enriches
discovery, and on the other hand inspires people to learn and understand the
molecules from different perspectives.

### Ref 2.2 – "Equivariant graph neural networks or transformers can be used to eliminate the need to perform augmentation for different rotations" –

Reviewer Comment	If the manuscript's main aim was to inject more information about the 3d nature of the molecules, they could have considered an equivariant graph neural network or transformer. An equivariant neural network would remove the need to perform augmentation for different rotations.
Author Response	We thank the reviewer for this valuable point. We agreed with reviewer that equivariant graph neural networks or transformers can indeed inject more information about the 3D nature of molecules by learning on well-characterized rotation-independent features. However, video and graph are complementary and they each have unique advantages in data presentation and analysis. There are several essential differences between VideoMol and equivariant graph neural network in obtaining the 3D of molecules as reflected below:
	(1) Modality representation. Obviously, video and graph are completely different in representation. There are several advantages for choosing video for molecular representation. First, video is a more intuitive representation method and information related to 3D nature can be directly observed from the videos, which allows the model to learn the information of bioactive molecules directly from the video without the help of any manual feature extraction. Secondly, VideoMol attempts the feasibility of learning representations from molecular videos and can be extended to learn a video related to potential ligand-receptor information by integrating with alphaFold3 or other available tools in the future.
	(2) Feature extraction. VideoMol extracts dense pixel-level features, while the graph-based model extracts relaxed node-level features. VideoMol allows the model to perceive 3D information in molecules by learning local textures in videos, such as the distance between pairs of atoms and the angle formed between three atoms and so on (as shown in the below <b>Extended Figure 1</b> ). In contrast, 3D graph-based methods require the intervention of explicit knowledge to guide the model to learn this information.
	These significant differences motivate us to develop a deep learning method based on video representations and we can also see significant advantages of VideoMol as we demonstrated in multiple drug discovery tasks. We have added these new explanations in the revised manuscript (Section Motivation for using video representations).



## Ref 2.3 – "Design experiment to understand why video information does not lead to sensitivity to conformer" –

Reviewer Comment	This is where it gets interesting: if the success of the model was truly due to better representation of 3d structure, we would expect the model to be sensitive to different conformers, especially on tasks that provide a binding affinity proxy. While, in multiple places in the manuscript the opposite is claimed, that model is robust to molecule conformer choice. Perhaps authors can devise an experiment to understand why video information does not lead to sensitivity to conformer.
Author Response	We agreed with the reviewer that the experimental section of " <i>Robustness</i> of <i>VideoMol</i> " is confused because we evaluated the performance of different models on different conformers, which does not reflect the robustness of a single model to different conformations.
	Therefore, we designed new experiments and evaluated the sensitivity of the same VideoMol model to different conformations. Specifically, we directly used pre-trained VideoMol to extract features of molecules with different conformers from 10 kinases datasets of binding affinity profiles and compared the similarities between different videos with different conformers. Since the similarity between conformers is related to their RMSD (Root-Mean-Square Deviation) distance, we also calculated the similarity of features in different RMSD intervals.
	As shown in the <b>Extended Table 1</b> (the Revised Supplemental Table 28) below, we found that VideoMol was discriminative for videos from different conformers. Further, when the RMSD between two conformations is larger, the feature similarity extracted by VideoMol shows a decreasing trend. Especially in the 90-100 percentile range, the feature similarity extracted by VideoMol is always the lowest. Therefore, VideoMol is sensitive to different conformers. We have added these new results and more detailed explanations in the revised manuscript.

	<b>Extended Table 1</b> (the Revise to distinguish different confor RMSD values from small to percentile interval.	ed Supple ormers. Th large and	emental Ta le percenti selecting t	ble 28). Th le interval he value (	ne ability of I refers to correspond	f VideoMol sorting all ding to the
	Percentile interval in RMSD	5HT1A	5HT2A	AA1R	AA2AR	AA3R
	0-10	0.692	0.749	0.799	0.803	0.823
	10-20	0.724	0.726	0.786	0.784	0.771
	20-30	0.741	0.747	0.755	0.764	0.772
	30-40	0.726	0.733	0.749	0.747	0.752
	40-50	0.735	0.745	0.759	0.772	0.758
	50-60	0.703	0.744	0.765	0.761	0.760
	60-70	0.708	0.733	0.742	0.756	0.745
	70-80	0.663	0.748	0.753	0.743	0.766
	80-90	0.631	0.702	0.745	0.753	0.752
	90-100	0.518	0.625	0.654	0.702	0.733
	0-100 (all data)	0.684	0.725	0.751	0.758	0.763
	=:	===== cor	ntinue ====	===		
	Percentile interval in RMSD	CNR2	DRD2	DRD3	HRH3	OPRM
	0-10	0.769	0.729	0.701	0.740	0.730
	10-20	0.747	0.733	0.740	0.721	0.740
	20-30	0.736	0.750	0.779	0.690	0.708
	30-40	0.735	0.748	0.795	0.703	0.716
	40-50	0.723	0.760	0.775	0.709	0.712
	50-60	0.710	0.743	0.785	0.706	0.725
	60-70	0.696	0.748	0.777	0.698	0.727
	70-80	0.712	0.715	0.775	0.700	0.716
	80-90	0.709	0.707	0.723	0.669	0.680
	90-100	0.680	0.615	0.631	0.585	0.658
	0-100 (all data)	0.722	0.725	0.748	0.692	0.711
Excerpt from Revised Manuscript	<b>Results:</b> <i>VideoMol captures conform</i> trained VideoMol to extract from 10 compound-kinase similarities between differe similarity between conforme Deviation) distance, we als RMSD intervals. We found different conformers (Supp between two conformation VideoMol shows a decrea range, the feature similarit Therefore, VideoMol can e molecules.	national di features interactio ent videos ers is rela o calculat that Vide lementary is is large sing trend cy extracte effectively	ifferences of molecu on dataset with diffe ted to the ed the sin eoMol is d Table 28 er, the fea d. Especia ed by Vid capture o	of molec iles with of s and co erent con ir RMSD inilarity of liscriminat 3). Furthe ature sim ally in the eoMol is conformat	ules. We different co ompared to formers. (Root-Mea features in tive for vio r, when to illarity ext e 90-100 always th ional diffe	used pre- onformers he cosine Since the an-Square n different deos from he RMSD racted by percentile ne lowest. rences of

## Ref 2.4 – "Add more discussion about understanding from where exactly the accuracy improvement comes from" –

Г

Reviewer Comment	Which makes me think the success of the model is not due to better 3d representation but one of the several other changes: 1-working with video frames and the new self-supervision tasks have expanded the effective size of the data and complexity the network processes each molecule (perhaps this is why the number of frames seem to change the prediction accuracy) 2-the large number of domain information that is crafted into the fingerprint may be impactful in several tasks in this work. Further understanding from where exactly the accuracy improvement comes from, can be considered for future work. In the teamtime, the claims of impact of 3D could be dialed down.								
Author Response	We thank the achieved by Vi and fingerprints on different sel task can promo	Reviewer f deoMol ma s. In <b>Suppl</b> f-supervise ote the impr	or these ex y be related ementary d pre-traini ovement of	xcellent po d to the intr <b>Table 24</b> , y ng tasks, a f VideoMol	ints. We ag oduction of we perform nd we find performanc	gree that the self-super ablation e that each p ce.	ne success vised tasks xperiments pre-training		
	Supplementary with balanced aware, direction using only vide strategy, respending All means and	Table 24: scaffold spl on-aware, a eo-aware s ctively. & re standard de	Effect of pr lit. w/o pret and chemic trategy, dir epresents the eviations ar	re-training s rain means cal-aware ection-awa ne combina re reported	strategy on no pre-tra represent re strategy tion of mult	6 regression ined Video pre-training , and chen iple pre-traine ree indeper	on datasets Mol. video- VideoMol nical-aware ining tasks.		
		5H <sup>-</sup>	T1A	AA	1R	AA	2AR		
	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		
			$0.000 \pm 0.014$	$0.002 \pm 0.010$	0.494±0.015	<u>0.710±0.001</u>	$0.562 \pm 0.002$		
	chemical-aware	0.719+0.002	0 550+0 004	0 706+0 005	0 516+0 000	0 724+0 009	0 569+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_direction chemical_video	0.718±0.002 0.716±0.010 0.774±0.014	0.559±0.004 0.553±0.010	0.706±0.005 0.670±0.007 0.730±0.004	0.516±0.008 0.511±0.005 0.553±0.002	0.724±0.008 0.719±0.013 0.865±0.011	0.568±0.003 0.553±0.014 0.672±0.005		
	chemical-aware chemical_direction chemical_video direction_video VideoMol	0.718±0.002 0.716±0.010 0.774±0.014 0.708±0.017	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005		
	chemical-aware chemical_direction chemical_video direction_video VideoMol	0.718±0.002 0.716±0.010 0.774±0.014 0.708±0.017 CN	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006	0.724±0.008 0.719±0.013 0.865±0.011 <b>0.712±0.011</b>	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3		
	chemical-aware chemical_direction chemical_video direction_video VideoMol	0.718±0.002 0.716±0.010 0.774±0.014 0.708±0.017 CN RMSE	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 IR2 MAE	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 RD2 MAE	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 HR RMSE	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 IR2 MAE 1.009±0.008	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 RD2 MAE 0.782±0.001	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 HR RMSE 0.819±0.001	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 IR2 MAE 1.009±0.008 0.782±0.010	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001 0.818±0.009	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 RD2 MAE 0.782±0.001 0.602±0.009	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 HR RMSE 0.819±0.001 0.732±0.008	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002 0.556±0.009		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware direction-aware	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025 1.024±0.01	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 IR2 MAE 1.009±0.008 0.782±0.010 0.842±0.008	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001 0.818±0.009 0.903±0.010	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 2D2 MAE 0.782±0.001 0.602±0.009 0.700±0.009	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 HR RMSE 0.819±0.001 0.732±0.008 0.759±0.005	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002 0.556±0.009 0.575±0.007		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware direction-aware chemical-aware	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025 1.024±0.01 0.890±0.004	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 R2 MAE 1.009±0.008 0.782±0.010 0.842±0.008 0.698±0.002	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001 0.818±0.009 0.903±0.010 0.759±0.003	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 202 MAE 0.782±0.001 0.602±0.009 0.700±0.009 0.565±0.003	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 HR RMSE 0.819±0.001 0.732±0.008 0.759±0.005 0.669±0.010	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002 0.556±0.009 0.575±0.007 0.512±0.010		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware direction-aware chemical-aware chemical_direction	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025 1.024±0.01 0.890±0.004 0.899±0.008	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 IR2 MAE 1.009±0.008 0.782±0.010 0.842±0.008 0.698±0.002 0.701±0.005	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001 0.818±0.009 0.903±0.010 0.759±0.003 0.773±0.017	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 RD2 MAE 0.782±0.001 0.602±0.009 0.700±0.009 0.565±0.003 0.579±0.007	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 HR RMSE 0.819±0.001 0.732±0.008 0.759±0.005 0.669±0.010 0.683±0.001	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002 0.556±0.009 0.575±0.007 0.512±0.010 0.514±0.001		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware direction-aware chemical-aware chemical_direction chemical_video	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025 1.024±0.01 0.890±0.004 0.899±0.008 0.874±0.012	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 IR2 MAE 1.009±0.008 0.782±0.010 0.842±0.008 0.698±0.002 0.701±0.005 0.686±0.013	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001 0.818±0.009 0.903±0.010 0.759±0.003 0.773±0.017 0.745±0.012	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 202 MAE 0.782±0.001 0.602±0.009 0.700±0.009 0.565±0.003 0.579±0.007 0.555±0.013	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 RMSE 0.819±0.001 0.732±0.008 0.759±0.005 0.669±0.010 0.683±0.001 0.686±0.003	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002 0.556±0.009 0.575±0.007 0.512±0.010 0.514±0.001 0.526±0.004		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware direction-aware chemical-aware chemical_direction chemical_video direction_video	0.718±0.002 0.718±0.002 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025 1.024±0.01 0.890±0.004 0.899±0.008 <u>0.874±0.012</u> 0.997±0.025	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 R2 MAE 1.009±0.008 0.782±0.010 0.842±0.008 0.698±0.002 0.701±0.005 0.686±0.013 0.789±0.021	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 DR RMSE 0.980±0.001 0.818±0.009 0.903±0.010 0.759±0.003 0.773±0.017 0.745±0.012 0.832±0.005	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 <b>RD2</b> <b>MAE</b> 0.782±0.001 0.602±0.009 0.700±0.009 0.565±0.003 0.579±0.007 <b>0.555±0.013</b> 0.615±0.007	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 RMSE 0.819±0.001 0.732±0.008 0.759±0.005 0.669±0.010 0.683±0.001 0.686±0.003 0.734±0.006	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 0.543±0.005 0.631±0.002 0.556±0.009 0.575±0.007 0.512±0.010 0.514±0.001 0.526±0.004 0.559±0.002		
	chemical-aware chemical_direction chemical_video direction_video VideoMol strategy w/o pretrain video-aware direction-aware chemical-aware chemical-aware chemical_video direction_video VideoMol	0.718±0.002 0.718±0.002 0.774±0.014 0.774±0.014 0.708±0.017 CN RMSE 1.216±0.012 0.978±0.025 1.024±0.01 0.890±0.004 0.899±0.008 <u>0.874±0.012</u> 0.997±0.025 0.864±0.005	0.559±0.004 0.553±0.010 0.603±0.015 0.547±0.015 R2 R2 0.782±0.008 0.782±0.008 0.698±0.002 0.701±0.005 0.686±0.013 0.789±0.021 0.679±0.010	0.706±0.005 0.670±0.007 0.730±0.004 0.655±0.007 <b>RMSE</b> 0.980±0.001 0.818±0.009 0.903±0.010 0.759±0.003 0.773±0.017 <u>0.745±0.012</u> 0.832±0.005 <b>0.742±0.004</b>	0.516±0.008 0.511±0.005 0.553±0.002 0.496±0.006 <b>MAE</b> 0.782±0.001 0.602±0.009 0.700±0.009 0.565±0.003 0.579±0.007 <b>0.555±0.013</b> 0.615±0.007 0.556±0.005	0.724±0.008 0.719±0.013 0.865±0.011 0.712±0.011 RMSE 0.819±0.001 0.732±0.008 0.759±0.005 0.669±0.010 0.683±0.001 0.686±0.003 0.734±0.006 0.668±0.008	0.568±0.003 0.553±0.014 0.672±0.005 0.543±0.005 H3 MAE 0.631±0.002 0.556±0.009 0.575±0.007 0.512±0.010 0.514±0.001 0.526±0.004 0.559±0.002 0.506±0.002		

3D representation, we evaluated the representation advantages of 3D-based molecular videos without using any self-supervised tasks and fingerprint information. It is well known that the development of drug discovery depends on the understanding of basic information about molecules. Here, we use several of the most representative methods (such as GCN, GIN, EGNN, and the representation used by ImageMoI) 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 randomly collected 10,000 molecules from the pre-training dataset and use exactly the same experimental setup for a fair comparison. In detail, we split the training set, validation set, and test set using a ratio of 8:1:1 and report the results on the test set based on the best validation set score. As shown in **Extended Table 1** (the Revised Supplemental Table 1) below, we found that VideoMol based on only one frame outperformed traditional 2D graph-based methods, the 3D-based graph method and the 2D image-based method, revealing the advantage of 3D representation. Specifically, compared with the second-place ImageMol without pre-training, the performance of video-1frame improved by 11%. When we utilized all video frames (video-60frame), the performance was further significantly improved from 12.469 to 7.55 with a 66% improvement rate.

Overall, the proposed 3D representation (whether based on a single frame image or a 60-frame video) has obvious advantages over existing representations. In addition, integrating physical dynamics or conformational changes from 3D ligand-receptor structures or models (i.e., alphaFold3 [1]) may improve performance of ViodeMol further. We have added these new results and more detailed explanations in the revised manuscript.

**Extended Table 1** (the Revised Supplemental Table 1). 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.

	modality	model	use conformer?	prop
mark based		GCN	×	62.304
	2D graph	GIN	×	62.980
graph-based		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

Reference

[1] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. Nature, 2024: 1-3.

## Ref 2.5 – "This difference may be due to data leak in the high ROC-AUC train-test data, inflating the apparent generalizability" –

Reviewer Comment	Feedback to the methodology: the splits used in this work would not stop data leaking from train to validation sets and scaffold balancing might not be enough. Indeed, we see a hint of the issue in the COX examples where training data from ChEMBL in 8:1:1 split gave high ROC-AUC>0.9, but when the model was tested against MedChemExpress data, only less than 40% of inhibitors are successfully identified. This difference may be due to data leak in the high ROC-AUC train-test data, inflating the apparent generalizability. In general, if authors would like to claim generalizability, more attention to the split strategy, overlap between data points is needed according to certain similarity metric will be needed.
Author Response	We carefully examined COX-1 and COX-2 datasets from ChEMBL. We confirmed that the training, validation, and test sets did not contain any overlapping molecules, <b>without ant data leaking issue</b> . In the actual virtual screening or new drug development process (termed external validation set), the performance value will not be very high due to several common factors, such as low similarity or activity cliff between the training data and the drugs in the external validation sets to be virtually screened. We confirmed that there was no data leak issues and we have added more explanations in the revised manuscript.

## Ref 2.6 – "Describe the computational requirements for pre-training and various downstream tasks" –

Reviewer Comment	Some of the tasks in the work are for high-throughout applications (e.g. virtual screening). In such cases the trade-off between accuracy and compute becomes important. The proposed method should clearly state the compute needs for pretraining and various downstream tasks. Because it works with video, compared to much smaller atomic position files, memory needs should be highlighted too.
Author Response	We have evaluated the computational requirements of the proposed VideoMol in pre-training, fine-tuning, and virtual screening stages. We found that VideoMol overall is highly cost-effective and time-effective in multiple tasks as discussed as below. We have added more detailed explanations about the computational requirements of the proposed VideoMol in pre- training, fine-tuning, and virtual screening stages in the revised manuscript.
Excerpt from Revised Manuscript	<b>C.3 Computational requirements of VideoMol</b> Here, we detail the computational requirements for the pre-training, fine- tuning, and virtual screening stages in <b>Supplementary Table 36</b> . In the pre- training phase ( <b>Supplementary Table 36a</b> ), VideoMol uses 256 frames in each batch for training and it requires <u>37G of GPU memory</u> and takes about 9 hours to complete 1 epoch on <u>2 million molecular videos with 60 frames</u> .

(#frame/bat We find tha only require computation molecules in when using	sed 10,000 m ch) on memor at fine-tuning o es at least <u>2</u> nal requiremen n <b>Supplemen</b> all frames dui	olecules to f ry and trainin does not occ 2.3G of GP nts when pentary Table ring virtual se	test the impa og speed in <b>S</b> cupy a large <u>U memory</u> . erforming virtu <b>36c</b> . We find creening for	act of differen <b>upplementa</b> amount of n Finally, we ual screening that it only 1 million mole	nt batch sizes <b>ry Table 36b</b> . memory and it evaluate the g on 1 million takes 9 hours ecules.	
<b>Supplemer</b> training, fin number of	n <b>tary Table</b> ( e-tuning, and frames in a	<b>36</b> : The con screening batch. #sa	mputational stages. #fra mples repres	requirements ame/batch re sent the tota	s in the pre- epresents the al number of	
molecules.	#frame/video	indicate hov	v many fram	es of a video	o to select for	
molecules. inference.	#frame/video	indicate hov	ements in the pre-training	es of a video	o to select for	
molecules. inference.	#frame/video	he computational required	ements in the pre-training	es of a video	o to select for	
molecules. inference. #samples 2 million	#frame/video	he computational require GPU memory ~37G	ements in the pre-training Training time -9 hours/epoch	g stage.	b to select for	
molecules. inference. #samples 2 million	#frame/video (a) Tr #frame/batch 256 (b) Th	tindicate hov	ements in the pre-training Training time -9 hours/epoch rements in the fine-tunin	es of a video stage. CPU: Intel 6248R 48C@ g stage.	o to select for server g3.0GHz; GPU: A100 (40G)	
molecules. inference. #samples 2 million #samples	(a) Tr #frame/video (b) Th #frame/batch	tindicate hov	ements in the pre-training Training time 	g stage. ☐ CPU: Intel 6248R 48C@ g stage. ☐ Stage. ☐ Stage. ☐ Stage.	b to select for	
molecules. nference. #samples 2 million #samples	(a) The second s	he computational require GPU memory -37G te computational requir GPU memory 2.3G	ements in the pre-training Training time -9 hours/epoch ements in the fine-tunin Training time -28 minutes/epoch	es of a video g stage. CPU: Intel 6248R 48C@ g stage. S	Server 33.0GHz; GPU: A100 (40G) Server	
molecules. nference. *samples 2 million #samples	#frame/video (a) Tr #frame/batch 256 (b) Th #frame/batch 8 16	he computational require GPU memory -37G e computational requir GPU memory 2.3G 2.6G	ements in the pre-training Training time -0 hours/epoch rements in the fine-tuning Training time -26 minutes/epoch -15 minutes/epoch	es of a video g stage. CPU: Intel 6248R 48C@ g stage. S	o to select for server 33.0GHz; GPU: A100 (40G) ierver	
molecules. inference. #samples 2 million #samples	#frame/video (a) Tr (b) Th (c)	e computational require GPU memory ~37G e computational require GPU memory 2.3G 2.6G 3.2G	ements in the pre-training Training time ~9 nours/epoch rements in the fine-tunin Training time ~26 minutes/epoch ~15 minutes/epoch ~12 minutes/epoch	es of a video	b to select for Berver §3.0GHz; GPU: A100 (40G) Server tel® Core™ i7-13700K	
molecules. nference. #samples 2 million #samples 10,000	(a) Tr (a) Tr #frame/batch 256 (b) Th #frame/batch 8 16 32 64	he computational require GPU memory ~37G te computational requir GPU memory 2.3G 2.6G 3.2G 4.3G	ements in the pre-training Training time ~9 hours/epoch ements in the fine-tunin Training time ~26 minutes/epoch ~15 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch	g stage. CPU: Intel 6248R 48C@ g stage. CPU: 13th Gen In GPU	D to select for           Server           23.0GHz; GPU: A100 (40G)           Server           tel® Core™ i7-13700K           2: 4090 Ti	
molecules. inference. #samples 2 million #samples 10,000	(a) Tr (b) Th (b) Th (b) Th (c) Th	he computational require GPU memory -37G e computational require GPU memory 2.3G 2.6G 3.2G 4.3G 6.5G	ements in the pre-training Training time -0 hours/epoch ements in the fine-tunit Training time -26 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch	es of a video g stage. CPU: Intel 6248R 48C@ g stage. CPU: 13th Gen In GPU:	D to select for           Server           93.0GHz; GPU: A100 (40G)           ierver           tel® Core™ i7-13700K           J: 4090 Ti	
molecules. inference. #samples 2 million #samples 10,000	#frame/video (a) Tr #frame/batch 256 (b) Th #frame/batch 8 16 32 64 128 256	indicate hov recomputational require GPU memory ~37G recomputational require GPU memory 2.3G 2.6G 3.2G 4.3G 6.5G 10.7G	ements in the pre-training Training time ~9 hours/epoch rements in the fine-tunin Training time ~26 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch	es of a video	b to select for ierver §3.0GHz; GPU: A100 (40G) ierver tel® Core™ i7-13700K 1: 4090 Ti	
molecules. inference. #samples 2 million #samples 10,000	(a) Th (b) Th (c) Th (c	he computational require GPU memory ~37G te computational requir GPU memory 2.3G 2.6G 3.2G 4.3G 6.5G 10.7G The computational requi	ements in the pre-training Training time -9 hours/epoch rements in the fine-tunin Training time -26 minutes/epoch -15 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch rements in the screening	es of a video g stage. S CPU: Intel 6248R 48C@ g stage. CPU: 13th Gen In GPL stage.	Server           33.0GHz; GPU: A100 (40G)           Server           tel® Core™ i7-13700K           U: 4090 Ti	
molecules. inference. <u>#samples</u> 2 million #samples 10,000 #samples	(a) Tr #frame/video (b) Th #frame/batch 8 (b) Th #frame/batch 8 16 32 64 128 256 (c) T #frame/video	e computational require GPU memory -37G e computational requir GPU memory 2.3G 2.6G 3.2G 4.3G 4.3G 6.5G 10.7G The computational requir #videos/batch	ements in the pre-training Training time -9 hours/epoch ements in the fine-tuning Training time -26 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch rements in the screening inference time	es of a video	D to select for           erver           g3.0GHz; GPU: A100 (40G)           ierver           tel® Core™ i7-13700K           b: 4090 Ti           Server	
molecules. inference. #samples 2 million #samples 10,000 #samples	(a) Tr (b) Th (c) Th	indicate hov recomputational require GPU memory ~37G recomputational require QPU memory 2.3G 2.6G 3.2G 4.3G 6.5G 10.7G re computational require #videos/batch 480	ements in the pre-training Training time ~9 nours/epoch ements in the fine-tunin Training time ~26 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch rements in the screening inference time ~9 minutes	es of a video	D to select for           Berver           §3.0GHz; GPU: A100 (40G)           Server           tel® Core™ i7-13700K           2: 4090 Ti	
molecules. inference. #samples 2 million #samples 10,000 #samples	(a) Tr (a) Tr #frame/batch 256 (b) Th #frame/batch 8 16 32 64 128 256 (c) T #frame/video 1 5	indicate hov recomputational require GPU memory -37G recomputational requir QPU memory 2.3G 2.6G 3.2G 4.3G 6.5G 10.7G recomputational requir #videos/batch 480 96	ements in the pre-training Training time -9 hours/epoch rements in the fine-tuning -26 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch rements in the screening inference time -9 minutes -48 minutes	es of a video	Description         Server           33.0GHz; GPU: A100 (40G)         3000000000000000000000000000000000000	
molecules. inference. <u>#samples</u> 2 million #samples 10,000 #samples	(a) TF #frame/video (b) Th #frame/batch 256 (b) Th #frame/batch 8 16 32 64 128 256 (c) T #frame/video 1 5 10	Indicate hov     GPU memory         ~37G     e computational require     GPU memory         ~37G     e computational require         3.2G         3.2G         4.3G         6.5G         10.7G         The computational require         #videos/batch         480         96         48	ements in the pre-training Training time ~9 hours/epoch rements in the fine-tunin Training time ~28 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~12 minutes/epoch ~14 minutes/epoch ~14 minutes/epoch ~14 minutes/epoch ~14 minutes/epoch ~14 minutes/epoch ~14 minutes ~48 minutes ~30 minutes	stage. CPU: Intel 6248R 48C@ g stage. CPU: 13th Gen In GPU stage. 17.7.0	D to select for           ierver           g3.0GHz; GPU; A100 (40G)           ierver           tel® Core™ i7-13700K           : 4090 Ti           Core™ i7-13700K           Core™ i7-13700K	
molecules. inference. #samples 2 million #samples 10,000 #samples 1 million	#frame/video           (a) Tr           #frame/batch           256           (b) Th           #frame/batch           8           16           32           64           128           256           (c) T           #frame/video           1           5           10           20	indicate hov	ements in the pre-training Training time ~9 nours/epoch ements in the fine-tunin Training time ~26 minutes/epoch ~12 minutes/epoch ~13 minutes ~48 minutes ~30 minutes ~30 minutes	es of a video	Server           §3.0GHz; GPU: A100 (40G)           Server           tel® Core™ i7-13700K           2: 4090 Ti           CPU: 13th Gen Intel® Core™ i7-13700K           CPU: 13th Gen Intel® Core™ i7-13700K	
molecules. inference. #samples 2 million #samples 10,000 #samples 1 million	#frame/video (a) Tr #frame/batch 256 (b) Th #frame/batch 8 16 32 64 128 256 (c) T #frame/video 1 5 10 20 30	he computational require GPU memory -37G e computational require GPU memory 2.3G 2.6G 3.2G 4.3G 6.5G 10.7G he computational require #videos/batch 480 96 48 24	ements in the pre-training Training time -9 hours/epoch ements in the fine-tunit Training time -26 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch -12 minutes/epoch rements in the screening inference time -9 minutes -3 hours -3 hours -4 hours	es of a video	Server           33.0GHz; GPU: A100 (40G)           Server           tel® Core™ i7-13700K           2: 4090 Ti           CPU: 13th Gen Intel® Core™ i7-13700K           CPU: 13th Gen Intel® Core™ i7-13700K           GPU: 4090 Ti	

Ref 2.7 – "Correct Minor typos" –

Reviewer Comment	minor typos in text, highlighting one that is on figure in case it escapes proofreading: angel -> angle Fig 1b
Author Response	We have fixed this typo and further polish English of the entire manuscript.

### **REVIEWERS' COMMENTS**

#### Reviewer #2 (Remarks to the Author):

I have carefully reviewed the detailed rebuttal and the revised manuscript. I am pleased to see that the authors have addressed all of the concerns I raised in my initial review. The additional experiments and clarifications provided have significantly strengthened the support for the claims. I am satisfied with the revisions and believe that the manuscript is now ready for publishing.

#### Reviewer #2 (Remarks on code availability):

The code is well organised, and the authors made the effort to provide guidelines for setting up an environment and running VideoMol.

### **Reviewer #4 (Remarks to the Author):**

I have assessed the comments from the authors to Reviewers 3 concerns, and I believe the authors has responded well to the comments