

MoBiTDE: A Molecular Bilinear Two-Tower Network for Drug-Excipient Nanoparticle Co-Assembly Prediction

Zeyu Ming^{1, a}, Mingsheng Xu^{2, *}

¹ Beijing Dewei British International School, No. 89 Capital Airport Road, Shunyi District, Beijing

² Suzhou Institute of Nano-Tech and Nano-Bionics (SINANO), Chinese Academy of Sciences

* Corresponding Author Email: msxu2021@sinano.ac.cn, ^a happytomato183@126.com

Abstract. Up to 40% of clinical candidates are poorly-soluble drugs, which rely heavily on nanoparticle formulations for oral and targeted delivery. Successful formulation largely depends on the co-assembly of drug and excipients into stable nanostructures. However, traditional wet-lab trial-and-error approaches require months, even years for discoveries to be made, with a substantial cost in raw materials. From research, there are 1440 high-throughput screening pairs of drug combination, but few AI algorithms have been applied to this task, especially with the rapidly evolving chemistry large language model (LLM), which hasn't been evaluated on this task yet. Herein, I introduce the novel application of chemistry foundation model representations to drug-excipient nanoparticle co-assembly prediction. Specifically, this essay introduces MoBiTDE, a molecular bilinear two-tower network designed to learn pairwise compatibility between drugs and excipients. MoBiTDE employs frozen SMI-TED embeddings as pre-trained molecular representations, complemented by 24 RDKit-derived physicochemical descriptors. Drugs and excipient representations are processed through two independent molecular towers and coupled by a rank-32 bilinear interaction module to explicitly capture cross-molecular interactions. The resulting pairwise representation is subsequently fed into a single positive-weighted binary classification head for co-assembly prediction. Experimental results demonstrate that MoBiTDE outperforms other methods under different evaluation settings. Notably, its predictive performance further improves as the dataset expands, highlighting the potential of pre-trained molecular representations to benefit from increasing experimental data. In summary, MoBiTDE provides an effective framework for learning drug–excipient compatibility and offers a promising computational approach to facilitate the discovery of self-assembling nanoparticle formulations. It can improve drug discovery efficiency by reducing the number of trials, focusing only wet-lab testing on promising combinations.

Keywords: co-assembly; SMI-TED; MoBiTDE; Chemical Large Language Model; Insoluble Drugs.

1. Introduction

Complex diseases are one of the most severe challenges to public health, but multiple effective therapeutics have been established for these conditions. For example, the drug sorafenib has been shown to be effective in treating cancer, specifically Hepatocellular carcinoma (HCC), Advanced renal cell carcinoma (RCC), and Progressive, locally advanced or metastatic differentiated thyroid carcinoma [1]. Another example is the drug lumefantrine, which is widely used to treat Malaria. Artemether/ lumefantrine is an ACT approved by the WHO for its safety and efficacy [2]. Lastly, the drug rapamycin is an immunosuppressant, which shows marked effects on autoimmune diseases and also aids organ transplants, reducing the risk of immune rejection, especially in obese recipients [3]. However, these drugs all have a problem to solve: they are effective, but the majority of drug candidates suffer from poor solubility and have a low absorption rate in humans [4]. This hinders the intake of drugs, even if passing the first barrier of acidity in the stomach. Therefore, chemicals cannot be circulated in the blood and will not easily reach the target. According to the Biopharmaceutics Classification System [5], we confirmed that Solubility and Permeability are the major barrier to the drug acting inside human bodies. Research have shown that these two classes have issues on bioavailability and patient compliance[6]. Classes II and IV are both characterised by low solubility. In other words, the rate-determining step in drug action is their dissolution. For example, Ibuprofen

(II) and Vancomycin (IV) are poorly soluble in water, so only a small proportion of them will circulate in the blood. Many excipients are water-soluble. Therefore, if we combine the two, the drug can increase in solubility.

According to Whitesides and Grzybowski [7], co-assembly is the autonomous organization of components into patterns or structures without human intervention. In this context, it is the ordered aggregation of carrier molecules and drugs in an aqueous environment, without external energy input. Nanoparticles are formed in this process. Hydrophobic interactions, hydrogen bonds, van der Waals forces and electrostatic attractions drive the formation of these structures. Excipients are substances that do not treat the disease on their own but perform various functions, such as stabilising the drug, coating it to make it easier to swallow, and controlling the rate of drug release. The advantage of this is that, without altering chemical structures to improve hydrophilic interactions, the drug can remain intact, preserving its active site and allowing it to perform its normal function [8]. On top of that, co-assembly requires minimal energy input, which is realistic in the human body, a biological system that is difficult to intervene in. Additionally, this will be less costly, as established drugs can be utilised in combination with their excipients without requiring extensive research.

Despite these advantages, several economic drawbacks limit their use in pharmaceutical development [9] [10]. For example, there have been about 2000 Class II and IV drugs recorded in the BCS, with nearly 400 types of excipients. Even with a one-to-one combination of drugs and excipients, the theoretical number of pairs has already exceeded 8 million. This will be very costly when wet-lab experiments are performed, especially when repeated. As a result, it would be a great challenge for the discovery to be done mainly by humans.

Reker et al. (2021) demonstrated that machine learning could enable the prediction of drug-excipient nanoparticle co-assembly and established an important public benchmark for this task [11]. However, the task remains understudied. In particular, it remains unclear whether complex deep learning models and pre-trained LLMs can improve predictive performance over conventional molecular descriptors and how they can be effectively used to model compatibility between a drug and an excipient [12]. To address these challenges, I introduce MoBiTDE, a molecular bilinear two-tower network for drug-excipient nanoparticle co-assembly prediction. MoBiTDE combines pre-trained molecular representations with physicochemical descriptors and processes drugs and excipients through two independent molecular towers [13]. A bilinear interaction module is then used to explicitly learn pairwise compatibility between the two molecules. I further evaluate representative chemistry foundation models, including SMI-TED, MolFormer-XL, ChemBERTa-2, and Uni-Mol, and benchmark different molecular representations and machine-learning algorithms for this task [14]. My experiments on two drug-excipient co-assembly datasets show that MoBiTDE outperforms conventional machine-learning baselines across different evaluation settings. I further observe that predictive performance improves as the available training data increase, suggesting that pre-trained molecular representations and pair-aware learning may benefit from the continued expansion of experimental co-assembly data. In summary, this work provides a systematic evaluation of modern molecular representations for drug-excipient co-assembly and establishes MoBiTDE as an effective computational framework for drug discovery.

2. Materials & Methods

2.1. Data

2.1.1. Data Overview

I evaluated my method using two drug-excipient nanoparticle co-assembly datasets, denoted as D1 and D2. D1 was originally reported by Reker et al. (2021) and contained 1,440 drug-excipient pairs made up of 16 drugs and 90 excipients [11]. Among these pairs, 94 were experimentally identified as positive co-assembly examples, giving a positive rate of 6.5%. To evaluate my model on a larger and more chemically diverse dataset, the extended dataset D2, released by the Reker Lab together with Xiang et al, was also adopted[15]. After removing one duplicated drug-excipient pair from the

original release, D2 contained 1,521 unique pairs involving 75 drugs and 95 excipients, of which 139 (9.1%) are positive examples. D2 includes all pairs from D1 together with 81 additional drug–excipient combinations, which notably expands the chemical diversity of the benchmark, particularly on the drug side, as shown in Figure 1.

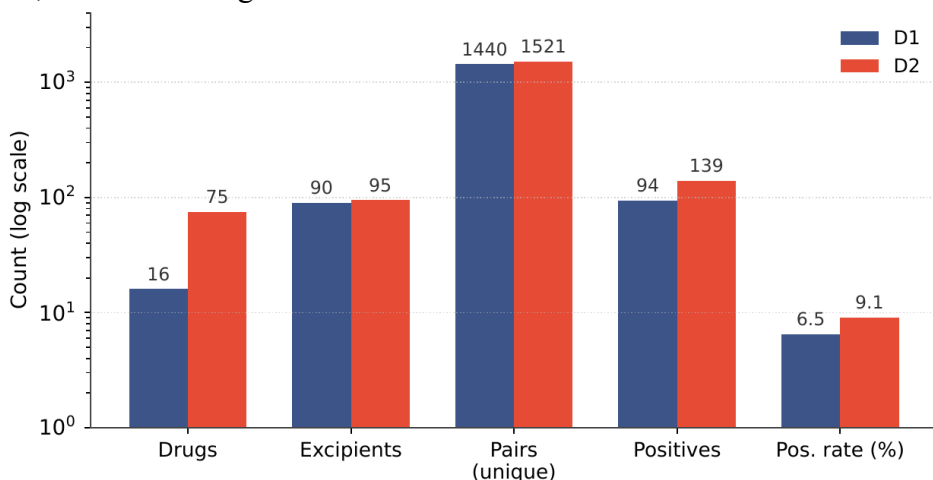


Figure 1. The overview of the dataset

2.1.2. Dataset splitting

All models were assessed using three 5-fold cross-validation protocols with increasing levels of generalization difficulty: Random, Group-Drug, and Group-both. For Random splitting, stratified sampling was applied over drug-excipient pairs to allow both molecules in a test pair to appear during training and provide direct comparability with previous studies. For Group-Drug splitting, I assigned drugs to non-overlapping folds, ensuring that all test drugs were unseen during training while excipients may have been observed. Lastly, Group-both gave the strictest evaluation by partitioning both drugs and excipients into non-overlapping groups, so that neither molecule in a test pair appeared in the training set. These three protocols allowed me to examine model performance from standard random-split prediction to increasingly challenging out-of-distribution generalization scenarios.

2.2. Methods

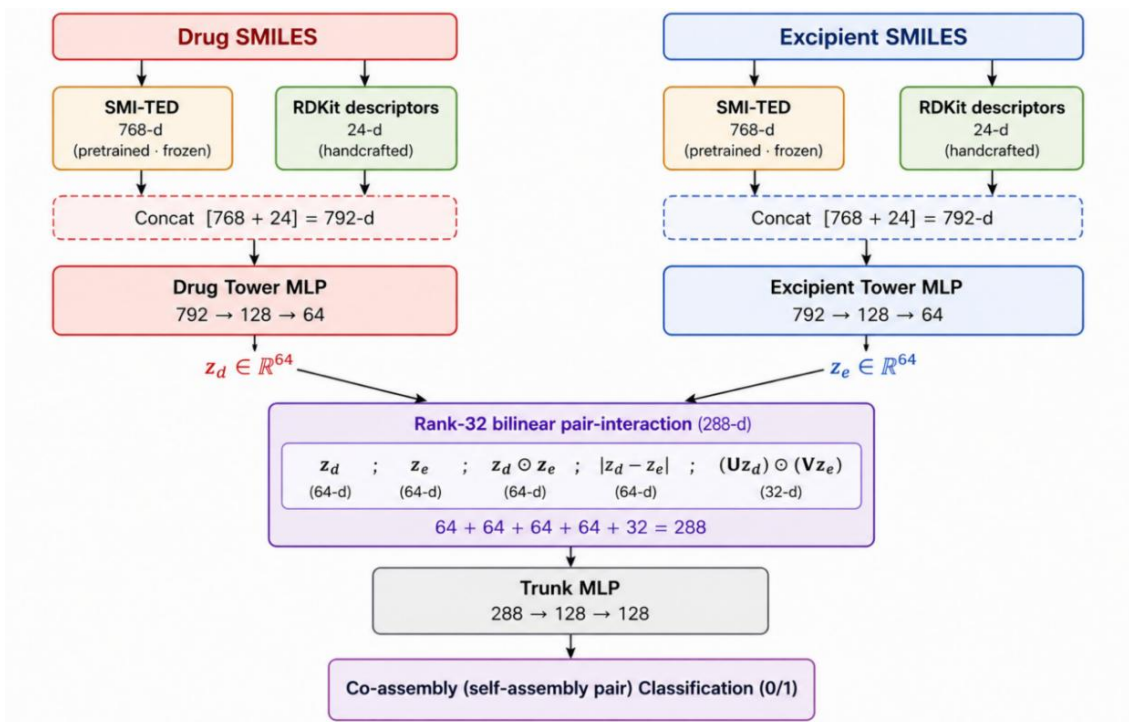


Figure 2. Model Structure

I propose MoBiTDE (Molecular Bilinear Two-Tower for Drug–Excipient Co-assembly), a pair-aware neural network for predicting drug-excipient nanoparticle co-assembly. As illustrated in Figure 2, MoBiTDE consists of three main components: pre-trained molecular representation, role-specific two-tower encoding, and explicit pairwise interaction learning.

Molecular Representation. Each drug and excipient is represented using two complementary sources of molecular information. First, the molecular SMILES is encoded using the pre-trained SMI-TED chemistry model to obtain a 768-dimensional molecular embedding [16]. The SMI-TED backbone remains frozen throughout training and serves solely as a pre-trained molecular feature extractor. The embeddings are precomputed and cached before model training, reducing computational cost and avoiding direct fine-tuning of the large pre-trained model on the relatively small co-assembly dataset. Also, 24 physicochemical descriptors are calculated for each molecule using RDKit. These descriptors characterize molecular properties such as molecular size, lipophilicity, polarity, hydrogen-bonding capacity, flexibility, aromaticity, molecular connectivity, surface area, and drug-likeness. The 768-dimensional SMI-TED embedding and 24-dimensional RDKit descriptor vector are concatenated to form a 792-dimensional molecular representation. In this way, MoBiTDE combines molecular patterns learned from large-scale chemical pre-training with explicit physicochemical information.

Two-Tower Encoder. The drug and excipient are processed by two separate molecular towers with the same architecture but different parameters. Each tower maps the 792-dimensional input to a 64-dimensional representation through two layers ($792 \rightarrow 128 \rightarrow 64$). Using separate towers allows the model to learn different representations for drugs and excipients, reflecting their different roles in nanoparticle co-assembly.

Bilinear Pair-Interaction Module [17] [18]. The two molecular representations are then combined to model drug–excipient compatibility. The interaction module includes the drug and excipient representations, their element-wise product, absolute difference, and a rank-32 bilinear interaction. These features capture complementary information about the two molecules and their pairwise relationship, resulting in a 288-dimensional representation. The low-rank bilinear interaction provides an efficient way to learn interactions between drug and excipient features while keeping the number of trainable parameters small. This is particularly suitable for the limited size of the co-assembly dataset.

Classification and Training. The 288-dimensional pair representation is passed through a two-layer MLP ($288 \rightarrow 128 \rightarrow 128$), followed by a binary classification head to predict drug–excipient co-assembly. To address class imbalance, I used positive-weighted binary cross-entropy and weighted sampling during training [19] [20]. The SMI-TED model remains frozen throughout training, and only the downstream MoBiTDE network is optimized.

3. Results

This section presents the experimental results of MoBiTDE for drug-excipient nanoparticle co-assembly prediction. I first compared MoBiTDE with eight baseline models on the extended D2 dataset. Then, I examined model performance under three different data-splitting methods: Random, Group-Drug, and Group-both. After that, I compared MoBiTDE against a single-tower neural network to investigate the effect of the two-tower bilinear design. Finally, the results on D1 and D2 were compared to examine whether increasing the available training data improves prediction performance.

3.1. Overall Performance

I first compared MoBiTDE with eight baseline models on the D2 dataset. The baselines covered three commonly used machine-learning algorithms, including Logistic Regression (LR), Random Forest (RF), and XGBoost (XGB), as well as a neural-network baseline based on a single-tower MLP. Different molecular representations were also included in the comparison, including RDKit physicochemical descriptors, SMI-TED embeddings, and their combination. All models were

evaluated using the same data-processing and cross-validation procedures. AUPRC was used as the main evaluation metric because positive co-assembly pairs account for only a small part of the dataset. A higher AUPRC indicates that the model is better at identifying positive drug–excipient pairs. Table 1 presents the main results. MoBiTDE achieved the highest AUPRC under all three data-splitting methods. Its AUPRC values were 0.666 for Random, 0.593 for Group-Drug, and 0.580 for Group-both.

Table 1. Performance comparison on the D2 dataset

Embedding	Algorithm	Rand. AUPRC	Group-Drug AUPRC	Group-both AUPRC	Group-both AUROC
RDKit	Logistic Regression	0.513	0.411	0.387	0.731
RDKit	Random Forest	0.576	0.475	0.399	0.733
RDKit	XGBoost	0.572	0.503	0.437	0.783
SMI-TED	Logistic Regression	0.541	0.390	0.323	0.667
SMI-TED	Random Forest	0.609	0.491	0.368	0.745
SMI-TED	XGBoost	0.607	0.516	0.426	0.762
SMI-TED + RDKit	XGBoost	0.600	0.501	0.360	0.707
SMI-TED	MLP	0.649	0.564	0.533	0.830
MoBiTDE	Two-tower + Bilinear	0.666	0.593	0.580	0.853

Under the Random setting, most models achieved reasonable performance. Among the traditional machine-learning methods, RF + SMI-TED achieved an AUPRC of 0.609 and XGB + SMI-TED achieved 0.607. Both were higher than their RDKit-based versions. The single-tower MLP further improved the AUPRC to 0.649. MoBiTDE achieved the highest value of 0.666. The results were different when the evaluation became more difficult. Under Group-Drug, the performance of most baseline models decreased. LR + RDKit decreased from 0.513 to 0.411, while RF + SMI-TED decreased from 0.609 to 0.491. XGB + SMI-TED achieved 0.516. In comparison, the single-tower MLP achieved 0.564 and MoBiTDE achieved 0.593. The differences were clearer under Group-both. In this setting, neither the drug nor the excipient in a test pair was observed during training. The traditional machine-learning models achieved AUPRC values between 0.323 and 0.437. The single-tower MLP performed better, reaching 0.533. MoBiTDE achieved the highest AUPRC of 0.580.

Figure 3 also presents the AUPRC results visually. The differences between MoBiTDE and the baseline methods were relatively small under Random splitting but became more visible under Group-Drug and Group-both. This pattern suggested that the main advantage of MoBiTDE appears when the model needs to make predictions for molecules that were not included during training. Overall, MoBiTDE achieved the best results across the three evaluation settings. The results also showed that neural-network models, especially the single-tower MLP and MoBiTDE, performed better than most traditional machine-learning models when the evaluation involved unseen molecules.

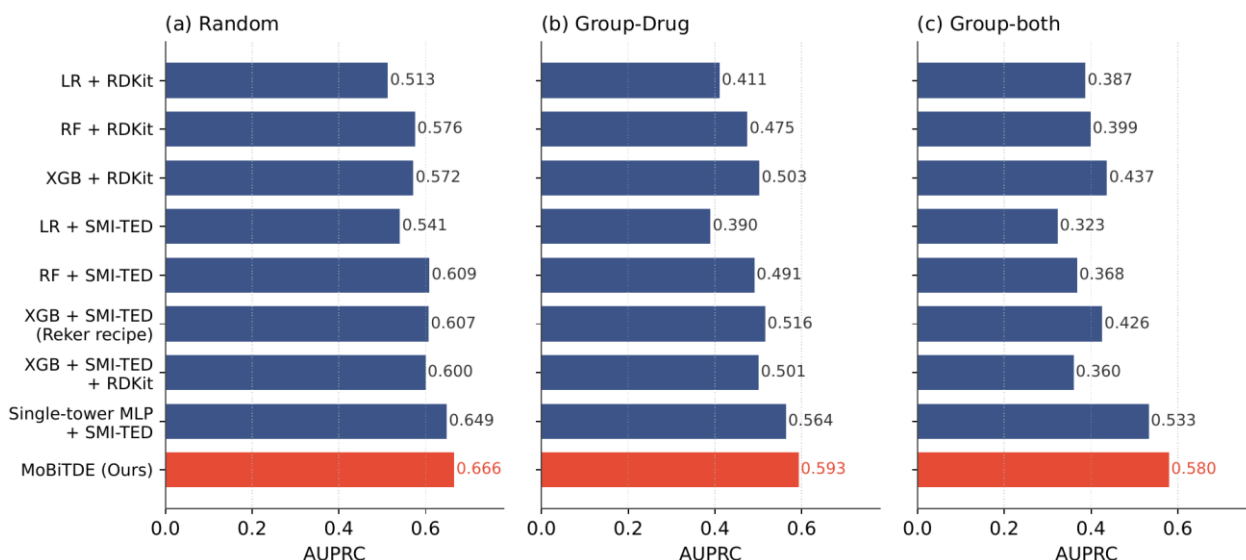


Figure 3. AUPRC performance comparison on the D2 dataset

3.2. Performance under Different Data Splits

The three data-splitting methods were designed to test the models under different levels of difficulty. Random splitting is the least strict setting. In this setting, a drug or excipient appearing in the test set may also have appeared in the training set as part of another pair. Group-Drug is more difficult because all drugs in the test fold are excluded from training. Group-both is the strictest setting because both the drug and the excipient in each test pair are unseen during training.

The results showed a general decrease in performance from Random to Group-Drug and Group-both. However, the size of this decrease was different across models. For example, LR + SMI-TED achieved an AUPRC of 0.541 under Random splitting, but this decreased to 0.390 under Group-Drug and 0.323 under Group-both. RF + SMI-TED decreased from 0.609 to 0.491 and then to 0.368. XGB + SMI-TED also dropped from 0.607 under Random to 0.516 under Group-Drug and 0.426 under Group-both. These results showed that good performance under Random splitting does not always lead to good performance when unseen molecules are tested. For example, RF + SMI-TED was one of the strongest traditional baselines under Random splitting, with an AUPRC of 0.609. However, its AUPRC decreased to 0.368 under Group-both. The neural-network models showed a smaller decrease. The single-tower MLP achieved AUPRC values of 0.649, 0.564, and 0.533 under Random, Group-Drug, and Group-both, respectively. MoBiTDE achieved 0.666, 0.593, and 0.580. In particular, MoBiTDE decreased by only 0.013 from Group-Drug to Group-both. This is notable because Group-both introduces an additional challenge: not only the drug but also the excipient is unseen during training.

Figure 3 also showed this pattern. For many traditional models, the bars became clearly shorter as the evaluation moves from Random to Group-both. In contrast, the reduction for MoBiTDE was smaller. These results showed why it is useful to report more than Random cross-validation for this task. Random splitting measured prediction on new combinations, but it did not always require the model to work with completely new molecules. Group-Drug and Group-both provided more difficult tests of model generalization. According to figure 1 and table 1, among the three settings, Group-both provided the strongest test in this study. MoBiTDE achieved the highest AUPRC and AUROC in this setting, showing that its performance remained relatively stable even when both molecules in a pair were unseen during training.

3.3. Effect of the Two-Tower Bilinear Design

The results in Section 3.1 showed that the two neural-network models performed better than most traditional machine-learning baselines. I therefore further compared MoBiTDE with the single-tower MLP to examine whether the proposed two-tower bilinear structure provided an additional

improvement. This comparison is useful because the two models are closer to each other than the traditional baselines. Both use pre-trained SMI-TED molecular representations and both use neural networks for prediction. The main difference is how the drug–excipient pair is processed. The single-tower MLP uses a simpler joint representation, while MoBiTDE first processes the drug and excipient separately and then combines them using the pair-interaction module.

Table 2. AUPRC comparison between the single-tower MLP and MoBiTDE

Data split	Single-tower MLP	MoBiTDE	Improvement
Random	0.649	0.666	+0.017
Group-Drug	0.564	0.593	+0.029
Group-both	0.533	0.580	+0.047

Under Random splitting, MoBiTDE improved AUPRC from 0.649 to 0.666. The improvement was relatively small at 0.017. Under Group-Drug, the difference increased. MoBiTDE achieved 0.593 compared with 0.564 for the single-tower MLP, giving an improvement of 0.029. The largest improvement was found under Group-both. The single-tower MLP achieved an AUPRC of 0.533, while MoBiTDE reached 0.580. The difference was 0.047. MoBiTDE also improved Group-both AUROC from 0.830 to 0.853. Furthermore, we found that a clear pattern could be seen across the three evaluation settings. The improvement over the single-tower model increased from 0.017 under Random, to 0.029 under Group-Drug, and finally to 0.047 under Group-both. The results supported the use of separate drug and excipient encoding followed by pairwise interaction modeling.

3.4. Effect of Increasing Training Data

The available amount of experimental data is another important factor in this task. I therefore compared MoBiTDE on D1 and D2 to study how model performance changes when more drug–excipient data are available. D1 contains 1,440 drug–excipient pairs, while D2 contains 1,521 unique pairs. For this comparison, the same MoBiTDE model and evaluation procedure were used. This allowed us to directly compare the prediction results obtained with the two datasets, as is shown in Table 3.

Table 3. Comparison of MoBiTDE performance on D1 and D2

Dataset	Random AUPRC	Group-both AUPRC
D1	0.471	0.487
D2	0.666	0.580
Improvement	+0.195	+0.093

Under Random splitting, MoBiTDE achieved an AUPRC of 0.471 on D1. On D2, the AUPRC increased to 0.666. This represents an absolute improvement of 0.195. The improvement under these two settings showed that MoBiTDE achieved better prediction results when trained and evaluated using the extended dataset. The increase under Group-both is especially useful because this setting tests the model on unseen drugs and unseen excipients. It is also interesting that the gain was larger under Random splitting than under Group-both. However, both settings showed the same overall direction: performance was higher on D2 than on D1.

These results suggest that increasing the amount of available drug–excipient data can help improve prediction performance. At present, datasets for drug–excipient nanoparticle co-assembly are still small compared with datasets used in many other molecular machine-learning tasks. Therefore, further collection of experimental co-assembly data may provide additional improvements in the future.

3.5. Effect of Pre-trained Molecular Models and Fine-Tuning

I next compared different pre-trained molecular models and examined whether fine-tuning could improve their performance. Three molecular models were tested on D1: SMI-TED, ChemBERTa-2, and MolFormer-XL [21]. For each model, we compared frozen molecular representations with rank-8 LoRA fine-tuning [22]. The results were shown in Table 4.

Among the frozen models, SMI-TED achieved the highest Group-both AUPRC of 0.487, followed by ChemBERTa-2 at 0.479 and MolFormer-XL at 0.421. Based on this comparison, SMI-TED was selected as the main molecular representation model in MoBiTDE. Fine-tuning with LoRA did not clearly improve the results. SMI-TED’s group-both AUPRC rose slightly from 0.487 to 0.492, while MolFormer-XL increased from 0.421 to 0.424. In contrast, ChemBERTa-2 dropped from 0.479 to 0.461. Overall, the differences between frozen and LoRA settings were small. The largest improvement was only 0.005. These results showed that fine-tuning did not provide a clear advantage on the current dataset. I therefore used frozen SMI-TED representations in the final MoBiTDE model.

Table 4. Group-both AUPRC of different pre-trained molecular models with frozen representations and LoRA fine-tuning on D1

Molecular model	Frozen	LoRA	Difference
SMI-TED	0.487	0.492	+0.005
ChemBERTa-2	0.479	0.461	−0.018
MolFormer-XL	0.421	0.424	+0.003

3.6. Effect of Class-Imbalance Strategies

The drug-excipient co-assembly dataset is highly imbalanced, with many more negative pairs than positive pairs. I therefore compared several training strategies for handling class imbalance. Our default method used positive-weighted binary cross-entropy together with weighted sampling. I compared this method with three other approaches: class-balanced loss (CB loss), LDAM with deferred re-weighting (LDAM + DRW), and asymmetric loss (ASL). The same MoBiTDE architecture and molecular representations were used in all experiments. The results were shown in Table 5.

The default strategy achieved the highest AUPRC under all three data-splitting settings. Under Random splitting, weighted BCE with weighted sampling reached an AUPRC of 0.666. The other methods reached values between 0.619 and 0.657. The differences were smaller under Group-Drug and Group-both. Under Group-both, the default strategy had an AUPRC of 0.580, while LDAM + DRW had a similar value of 0.577. CB loss and ASL achieved 0.568 and 0.567, respectively. Overall, none of the alternative methods consistently improved the prediction performance. I thus used weighted BCE together with weighted sampling as the default strategy for MoBiTDE.

Table 5. Comparison of different class-imbalance strategies on D2

Strategy	Random AUPRC	Group-Drug AUPRC	Group-both AUPRC
Weighted BCE+ sampling	0.666	0.593	0.580
CB loss	0.619	0.551	0.568
LDAM + DRW	0.648	0.577	0.577
ASL	0.657	0.578	0.567

4. Discussion

This study proposed MoBiTDE to be used in predicting the success of drug-excipient co-assembly into nanostructures. In contrast to traditional machine-learning baselines, MoBiTDE achieved better results under all three data-splitting methods (Random, Group-Drug, and Group-both). Apart from overall prediction performance, this study also compared different model architectures, data sizes,

pre-trained molecular models and class-imbalance training strategies. Below are discussed experimental results on model architecture, pre-trained molecular models and data sizes.

Drug-Excipient pair modelling: A key feature of this study is to view the drug and excipient as two distinct molecular entities, instead of a type of combined input. This is because drugs and excipients often serve different purposes, so even similar structural features can contribute differently to overall co-assembly prediction, depending on whether they belong to the drug or the excipient molecule. If the pair was inputted as a whole, it would be difficult to learn whether the structural contribution is from the drug or the excipient, so the model would struggle to extract key features of individual molecules. Therefore, the model may underperform in tasks which require generalizing to unseen molecular pairs, so it might exhibit over-fitting, in contrast with learning transferable features for individual molecules. Furthermore, certain properties are found very rare among drugs, and more common among excipients (or vice-versa), due to drug-design restrictions. For example, very long saturated alkyl chains (C12-C18) impede drug action and solubility. However, it is common for excipients to retain them as they aid in forming nanoparticles. With a single-tower, the model may hallucinate that drugs with long, saturated alkyl tails will contribute to co-assembly, thus will emphasize parameters related to saturation. (CITATION). Similarly, this effect is common in chemistry LLMs, where a set of features that contribute to a phenomenon may not be applicable to combination other than the specified pair, so still requires wet-lab testing. Thus, our method may further reduce wet-lab experiments, aiming for sharp and accurate experiments, which is vital for finding features that contribute to chemical relationships.

From experimental results, MoBiTDE was validated via a higher AUPRC, suggesting the usefulness of this design. When compared to Single-Tower MLP, MoBiTDE demonstrated greater area under the precision-recall curve under all three data-splitting methods. Another noticeable result is that the differences between the two models (MoBiTDE & Single-Tower MLP) increased as the prediction task became more difficult. Under Random, Group-Drug and Group-Both (from easiest to hardest), the AUPRC for MoBiTDE yielded AUPRC improvements of 0.017, 0.029 and 0.047 respectively. This data is concordant with drug-excipient self assembly nature, as this property depends on both molecules' structures, instead of being governed by a feature of one molecule. This suggests that solely studying each chemical's structure may be insufficient: the model should also learn non-linear interactions (relationships) between the two molecules.

The use of pre-trained molecular models: This study also utilises pre-trained molecular models to gain the molecular representation of drugs and excipients. Unlike training with a small dataset, pre-trained models had learnt basic foundations of chemical structure and interactions from a wider chemical dataset. This suggests its applicability as a molecular feature extractor. The experiment compared various pre-trained molecular models, and these models were shown to have good compatibility with MoBiTDE's two-tower architecture, even though there were slight performance differences. In particular, SMI-TED delivered favourable effects in the testing, so it was selected to be used by MoBiTDE as a chemical feature extractor. I also compared frozen representations and LoRA fine-tuning. In practice, LoRA did not bring a significant increase in performance. For example, SMI-TED (fine-tuned) merely increased the AUPRC from 0.487 to 0.492 under Group-both; other pre-trained models also had minimal enhancements. Therefore, frozen pretrained representations are a simpler and more effective method, while it also reduces cost of training. More importantly, from current results, complex fine-tuning did not bring extra benefits, which also indicates that the drug-excipient co-assembly task cannot be fully solved by fine-tuning existing pre-trained models. Components specifically designed for drug-excipient combinations are required to model their inter-relationships.

Class imbalance: In drug-excipient co-assembly data, pairs that can self-assemble are significantly fewer than those that cannot. Therefore, class-imbalance is also one key problem for model training. I compared weighted BCE + weighted sampling, CB loss, LDAM + DRW and ASL. The experimental results showed that more complex class-imbalance strategies did not bring better overall performance. In contrast, the simpler combination of weighted BCE and weighted sampling achieved the best results across all three data-splitting methods. This result suggests that under the investigated

task and dataset size, complex loss functions are not always needed to deal with class-imbalance. Simple weighting schemes is enough to yield satisfying results. It is worth noting that this conclusion is only applicable to the current dataset and experimental setup. As future samples change in number and distribution, different class-imbalance strategies can also change in performance.

Limitations: This study retains certain limitations. First of all, the current drug-excipient co-assembly dataset is relatively small. Although D2 has more experimental results on the basis of D1, and has improved MoBiTDE's performance, it still has much less training data when compared to common chemistry LLMs' training dataset. As a result, the ability for the model to generalize to a bigger dataset might require more data for it to be improved: decreasing the chance of overfitting. Moreover, this study's aim of predicting the success of drug-excipient co-assembly is a binary classification problem. In actual nanoparticle studies, it is not fully sufficient to solely understand co-assembly of the pair. Particle radius, stability and other experimental properties are equally important. Future studies could embed such information, to aim at more holistic model, with multiple classification heads. Furthermore, current results are based upon prior experimental data. An important step in the future is to use MoBiTDE to predict new drug-excipient combinations, and validate the results through wet-lab experiments. Lastly, drug-excipient co-assembly is dependent on laboratory conditions, so may lose applicability if the conditions are altered (such as pH, pressure and temperature). More data with different experimental conditions can be used as inputs to train the model, widening its applicability.

5. Conclusion

This study proposes MoBiTDE, a two-tower model that can be used to predict drug-excipient co-assembly. The model combines pre-trained molecular representations, RDKit physicochemical descriptors and a bilinear interaction module, to separately represent drugs and excipients, and to further learn the pair relationships between two molecules. Experimental results demonstrate that MoBiTDE achieved promising performance across all three data-splitting methods (Random, Group-Drug and Group-Both), which is higher than classical machine learning methods and single-tower MLP compared in this study. Even in the stricter Group-Both settings, MoBiTDE still retained its high performance. This suggests that learning separately for drugs and excipients to model their relationships is beneficial to this task. Overall, this study demonstrates that pre-trained molecular representations and drug-excipient pairs, when combined, is applicable to predicting the success of drug-excipient co-assembly. MoBiTDE can be used as a primary method to screen potential combinations before experimental validation, to reduce costs and time. Future studies could further increase the dataset on drug-excipient co-assembly, to alter experimental conditions and introducing more labels such as nanoparticle radius and stability. Another application could be to discover important structural features. For example, by analyzing the matrix in the bilinear layer and reducing its dimensionality, we can gain insight into potential novel factors that contribute to co-assembly, and by validating these effects in laboratory settings, mechanistic interaction between molecules can be discovered. At the same time, prospective experimental validations are required to evaluate the model's performance on new drug-excipient combinations.

Acknowledgements

I would like to express my deepest gratitude to all who have contributed to the completion of this research project. Without their consistent support and guidance, this project would not have been possible.

This project originated from a failure in my A-Level EPQ's pre-experiment, and a habit of reviewing academic news. My 96-checker board assay method aiming to quantify synergistic interactions between chemical pairs did not yield satisfying results, because of bacteria death, making results undistinguishable. During reflections on further standardization of the experiment, I realized that with such vast combination of drugs and adjuvants/excipients, screening experiments incur substantial

costs. Recognizing that AI tools, such as alpha-fold, is becoming popular in the field of biochemistry, I thought that AI might be utilised in predicting drug-excipient self-assembly, especially with limited literature on the latest large-language-model implementations. My friend, who specialises in computer science, assisted me in locating machine-learning resources, and helped me standardize code syntax to match the running environment. Apart from this, school ICT technicians helped me set up software-running environments on my computer. Professor Andrew Ng's free university machine learning course provided significant support on building my foundation of knowledge, and provided abilities on model selection, training and error diagnostics which I then adapted to choose learning methods that I used in this project.

Meanwhile, my school tutor, Ms Zhang, provided context regarding AI's development in Chemistry, and suggested practical explanations for its implementation. For example, she explained how the SMILES encoder worked. I am particularly grateful for her encouragement of creative thinking, which has been invaluable in the journey. She showed constant support for me even in times of difficulties, and encouraged me to persist through setbacks, motivating me to reach out for help from forums. I hereby confirm that all guidance and supervision provided by her were entirely pro bono, based on shared passion for scientific research and her commitment to student conducting research.

Furthermore, I would also like to sincerely mention the anonymous users on technology forums for providing me with help at times when I was stuck or encountered technical issues. They offered suggestions regarding potential sources of training failure and recommendations to optimize model performance, which I interpreted and implemented independently, for instance by identifying logical errors within the code. Their technical support and assistance were essential to the successful outcomes of this project.

To summarize, I came up with the topic of this project, and did data analysis on the results given by the AI. I wrote the essay and developed the model by finalizing selections on particular training methods and features. I also made the decision of using the bilinear two-tower model.

Moreover, I declare the use of artificial intelligence assistance (AI) use in this paper. All AI use was limited to literature research and language-polishing of this essay. For example, grammar was checked and edited to ensure it maintained clear information delivery. In addition, technological terms were replaced to ensure standardization of terminology, matching with existing literature. The model used was Doubao AI assistant, based on Doubao-Seed 2.1, ByteDance 2026. It was used during the final part to finalize the essay. AI was also used for researching relevant literature on this topic. The platform Bohrium Science Navigator, from DP Technology, was used. It was mainly used during research for literature retrieval. All outputs generated by AI were manually checked against primary sources. AI was first used on May 5 times, to do literature research. Then, it was used in June to see if the essay was structured well. In addition, it was used three times on September to polish the essay. Finally, it was checked for misconceptions and literature citations once on October.

Lastly, I acknowledge the foundational work of all researchers cited in this paper. Their pioneering studies built the vital context and motivation for this project.

Reference

- [1] G. Gadaleta-Caldarola et al., "Sorafenib: 10 years after the first pivotal trial," *Future Oncol.*, vol. 11, no. 13, pp. 1863–1880, Jul. 2015, doi: 10.2217/fon.15.85.
- [2] M. Makanga and S. Krudsood, "The clinical efficacy of artemether/lumefantrine (Coartem®)," *Malar. J.*, vol. 8, no. S1, p. S5, Dec. 2009, doi: 10.1186/1475-2875-8-S1-S5.
- [3] A. Deißler et al., "Rapamycin delays allograft rejection in obese graft recipients through induction of myeloid-derived suppressor cells," *Immunol. Lett.*, vol. 236, pp. 1–11, Aug. 2021, doi: 10.1016/j.imlet.2021.05.003.
- [4] S. Kalepu and V. Nekkanti, "Insoluble drug delivery strategies: review of recent advances and business prospects," *Acta Pharm. Sin. B*, vol. 5, no. 5, pp. 442–453, Sep. 2015, doi: 10.1016/j.apsb.2015.07.003.
- [5] G. L. Amidon, H. Lennernäs, V. P. Shah, and J. R. Crison, "A Theoretical Basis for a Biopharmaceutic Drug Classification: The Correlation of in Vitro Drug Product Dissolution and in Vivo Bioavailability," *Pharm. Res.*, vol. 12, no. 3, pp. 413–420, Mar. 1995, doi: 10.1023/A:1016212804288.

- [6] A. R. Hashmi, M. Sekar, F. Zahra, N. Molugulu, and L. S. Wong, “Advanced Drug Delivery Strategies to Overcome Solubility and Permeability Challenges: Driving Biopharmaceutical Advancements Toward Commercial Success,” *ACS Omega*, vol. 10, no. 36, pp. 40769–40792, Sep. 2025, doi: 10.1021/acsomega.5c06171.
- [7] G. M. Whitesides and B. Grzybowski, “Self-Assembly at All Scales,” *Science*, vol. 295, no. 5564, pp. 2418–2421, Mar. 2002, doi: 10.1126/science.1070821.
- [8] O. C. Farokhzad and R. Langer, “Impact of Nanotechnology on Drug Delivery,” *ACS Nano*, vol. 3, no. 1, pp. 16–20, Jan. 2009, doi: 10.1021/nn900002m.
- [9] Y. Xu, J. Pei, and L. Lai, “Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction,” *J. Chem. Inf. Model.*, vol. 57, no. 11, pp. 2672–2685, Nov. 2017, doi: 10.1021/acs.jcim.7b00244.
- [10] H. Öztürk, A. Özgür, and E. Ozkirimli, “DeepDTA: deep drug–target binding affinity prediction,” *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018, doi: 10.1093/bioinformatics/bty593.
- [11] D. Reker et al., “Computationally guided high-throughput design of self-assembling drug nanoparticles,” *Nat. Nanotechnol.*, vol. 16, no. 6, pp. 725–733, Jun. 2021, doi: 10.1038/s41565-021-00870-y.
- [12] K. Huang, C. Xiao, L. M. Glass, and J. Sun, “MolTrans: Molecular Interaction Transformer for drug–target interaction prediction,” *Bioinformatics*, vol. 37, no. 6, pp. 830–836, May 2021, doi: 10.1093/bioinformatics/btaa880.
- [13] J.-H. Kim, J. Jun, and B.-T. Zhang, “Bilinear Attention Networks,” Oct. 19, 2018, arXiv: arXiv:1805.07932. doi: 10.48550/arXiv.1805.07932.
- [14] J. Ross, B. Belgodere, V. Chenthamarakshan, I. Padhi, Y. Mroueh, and P. Das, “Large-scale chemical language representations capture molecular structure and properties,” *Nat. Mach. Intell.*, vol. 4, no. 12, pp. 1256–1264, Dec. 2022, doi: 10.1038/s42256-022-00580-7.
- [15] Y. Xiang et al., “Physics-informed design of drug-excipient nanoparticles via free energy calculation and yoked deep learning,” *Npj Comput. Mater.*, Aug. 2026, doi: 10.1038/s41524-026-02262-z.
- [16] E. Soares, E. Vital Brazil, V. Shirasuna, D. Zubarev, R. Cerqueira, and K. Schmidt, “An open-source family of large encoder-decoder foundation models for chemistry,” *Commun. Chem.*, vol. 8, no. 1, p. 193, Jul. 2025, doi: 10.1038/s42004-025-01585-0.
- [17] T.-Y. Lin, A. RoyChowdhury, and S. Maji, “Bilinear CNN Models for Fine-Grained Visual Recognition,” in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile: IEEE, Dec. 2015, pp. 1449–1457. doi: 10.1109/ICCV.2015.170.
- [18] H. Öztürk, A. Özgür, and E. Ozkirimli, “DeepDTA: deep drug–target binding affinity prediction,” *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018, doi: 10.1093/bioinformatics/bty593.
- [19] J.-H. Kim, J. Jun, and B.-T. Zhang, “Bilinear Attention Networks,” Oct. 19, 2018, arXiv: arXiv:1805.07932. doi: 10.48550/arXiv.1805.07932.
- [20] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *J. Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.
- [21] S. Chithrananda, G. Grand, and B. Ramsundar, “ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction,” Oct. 23, 2020, arXiv: arXiv:2010.09885. doi: 10.48550/arXiv.2010.09885.
- [22] E. J. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” Oct. 16, 2021, arXiv: arXiv:2106.09685. doi: 10.48550/arXiv.2106.09685.