Bioconjugate Chemistry

pubs.acs.org/bc

Article

Data-Driven Design of Novel Polymer Excipients for Pharmaceutical Amorphous Solid Dispersions

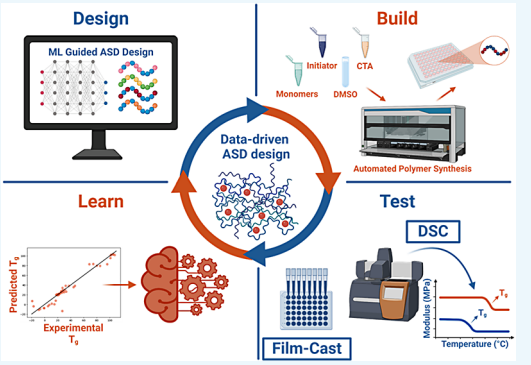
Published as part of Bioconjugate Chemistry special issue "Computational Methods in Drug Delivery".

Elena J. Di Mare, Ashish Punia,* Matthew S. Lamm, Timothy A. Rhodes, and Adam J. Gormley*

Cite This: *Bioconjugate Chem.* 2024, 35, 1363−1372

Read Online

ABSTRACT: About 90% of active pharmaceutical ingredients (APIs) in the oral drug delivery system pipeline have poor aqueous solubility and low bioavailability. To address this problem, amorphous solid dispersions (ASDs) embed hydrophobic APIs within polymer excipients to prevent drug crystallization, improve solubility, and increase bioavailability. There are a limited number of commercial polymer excipients, and the structure−function relationships which lead to successful ASD formulations are not well-documented. There are, however, certain solid-state ASD characteristics that inform ASD performance. One characteristic shared by successful ASDs is a high glass transition temperature (T_g), which correlates with higher shelf stability and decreased drug crystallization. We aim to identify how polymer features such as side chain geometry, backbone methylation, and hydrophilic−lipophilic balance impact T_g to design copolymers capable of forming high-T_g ASDs. We tested a library of 50 ASD formulations (18 previously studied and 32 newly synthesized) of the model drug probucol with copolymers synthesized through automated photoinduced electron/energy transfer-reversible addition−fragmentation chain-transfer (PET-RAFT) polymerization. A machine learning (ML) algorithm was trained on the T_g data to identify the major factors influencing T_g, including backbone methylation and nonlinear side chain geometry. In both polymer alone and probucol-loaded ASDs, a Random Forest Regressor captured structure−function trends in the data set and accurately predicted T_g with an average R^2 > 0.83 across a 10-fold cross validation. This ML model will be used to predict novel copolymers to design ASDs with high T_g, a crucial factor in predicting ASD success.



INTRODUCTION

A significant branch of pharmaceutical product formulation is geared toward enhancing the bioavailability of drug molecules, or active pharmaceutical ingredients (APIs), with poor aqueous solubility.[1,2] Amorphous solid dispersion (ASD) technology is widely used to obtain a molecular dispersion of amorphous drug to increase the aqueous solubility of such poorly soluble APIs. Higher energy amorphous drug has a thermodynamic drive to convert to its energetically favorable crystalline form. Upon crystallization, an API loses both bioavailability and therapeutic efficacy due to reduced aqueous solubility.[3] We are therefore driven to maintain the API in an amorphous state for as long as possible until it can be orally administered to patients.[4] To extend shelf stability, APIs are commonly combined with polymers as antiplasticization agents to prevent crystallization.[1] Polymers, along with other excipients, are formulated with APIs into amorphous solid dispersions (ASDs), which can be developed into shelf stable pills or tablets for oral delivery.[5]

The quality of an ASD is measured in terms of its ability to remain stable in shelf storage conditions and consistently deliver precise doses of bioactive API throughout the lifetime of the ASD.[5] Specifically, a successful ASD will disrupt crystal lattice formation of the API and retain a molecular dispersion in a homogeneous amorphous phase.[6−9] ASDs must maintain performance for at least 12 months at 25 °C and 60% relative humidity to meet official stability standards for approval.[10] One material property, glass transition temperature (T_g), has been identified as a key factor influencing the stability of ASDs.[5,11] T_g represents a transition point from a hard, glassy material to a pliable, rubbery material. Generally, a higher T_g is associated with increased shelf-stability and prolonged time to API crystallization because an ASD stored far below its T_g exists in a hard, glassy state that inhibits crystal formation.[3,12] In practice, ASDs must be stored at least 50 °C below their T_g to ensure there is limited molecular mobility that would drive API

Received: June 27, 2024
Revised: August 10, 2024
Accepted: August 12, 2024
Published: August 16, 2024

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

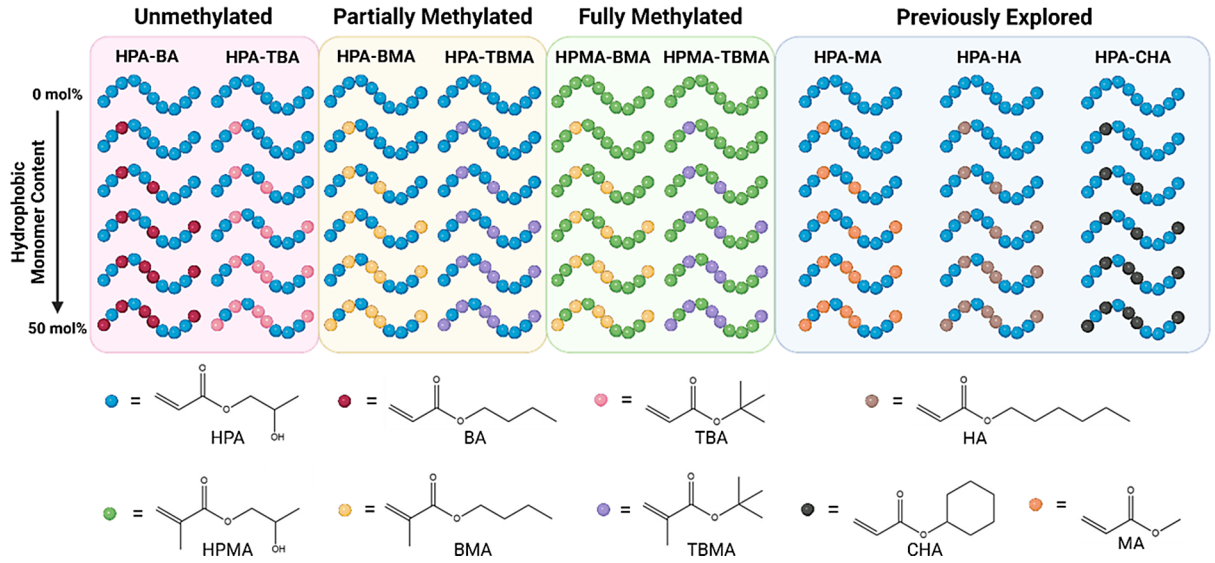


Figure 1. Polymer library and constituent monomers. Polymer synthesis was carried out through automated PET-RAFT polymerization. HPA or HPMA was copolymerized with a hydrophobic monomer such that hydrophobic monomer content varied between 0 and 60 mol %. Hydrophobic monomers had varied side chain geometry, being either linear (BA, BMA, HA, MA), branched (TBA, TBMA), or cyclic (CHA). Monomers were either methacrylates (HPMA, BMA, TBMA) or acrylates (HPA, BA, TBA, HA, CHA, MA), and were combined to form unmethylated, partially methylated, or fully methylated copolymers. Created in Biorender.

crystallization.[1] Because stability is a vital metric of success for ASDs, novel formulations must prioritize designing ASDs with high T_g and therefore increased shelf stability. For probucol-based ASDs, utilization of high T_g polymers such as polyvinylpyrrolidone (PVP) has been described in literature.[13] PVP-K30 based ASDs show T_g of approximately 135 °C.[13] Higher T_g polymers are preferred for probucol-based ASDs due to low T_g of amorphous probucol.[6]

Accurately predicting the T_g of an ASD informs performance, shelf stability, and optimal storage conditions. Because T_g is impacted by each material present in an ASD, selection of polymer excipients with high T_g can beneficially increase the overall T_g of the ASD system. However, it can be difficult to design or select polymer excipients with high T_g without previously established information on the excipient's phase transition temperature. Molecular structure, packing ability, hydrophobicity, and drug-polymer interactions can each impact T_g and complicate T_g prediction.[12] The standard empirical methods for predicting T_g, such as the Fox equation, Gordon−Taylor equation and the Couchman-Karasz equation, fail to accurately model T_g in a number of complex systems due to the empirical models' assumption of ideal mixing and no interaction between drug and polymer.[12,14] In systems where drug-polymer interactions do take place, over-reliance on empirical models for T_g may lead to misleading predictions.[14] To address the shortcomings of these empirical models, several machine learning (ML) models have been employed to predict T_g based on copolymer properties.[15−17] These models strive to capture structure−function relationships for a range of copolymer designs; however, less work has been done to apply ML toward drug-polymer systems, especially with a focus on T_g as the target parameter for improved ASD formulation.[18−20]

The development of novel polymer excipients is severely limited by a lack of understanding of how polymer selection and polymer-drug interactions dictate T_g. This gap in knowledge is compounded by the narrow range of commercially available polymer excipients for ASDs. Commonly used excipients tend to be either cellulose derivatives (HPMC or HPMCAS), PVP, or polyethylene glycol (PEG) derivatives.[6,21] Cellulose derivatives dominate the ASD market and are used in approximately two-thirds of all formulations.[9] Because the list of polymer excipients used in ASD development is so limited, it is unclear which structural features make a polymer excipient especially well-suited for ASDs. Designing custom excipients tailored to increase stability of a specific API is particularly difficult without established structure−function relationships.

An obvious solution is to quantitatively determine structure−function relationships and fully map the polymer excipient design space, effectively closing the knowledge gap. This approach ensures there is sufficient information to tailor novel polymer excipients to optimize shelf stability and performance of ASDs with a variety of APIs. However, polymer design is a multidimensional problem characterized by complex interactions between all design features. The lengthy list of tunable polymer properties includes (but is certainly not limited to) hydrophobicity, monomer content, chemical structure, degree of polymerization, charge, and molecular weight. Each of these interdependent features impacts T_g, drug-polymer interactions, and overall ASD performance. Fully understanding the breadth of the design space would require very large data sets of ASDs formulated with diverse polymer excipients.

Recent advancements in ML and automated, open-air polymer synthesis enable researchers to begin mapping this complex design space.[22] Namely, photoinduced electron/energy transfer-reversible addition−fragmentation chain-transfer (PET-RAFT) polymerization allows us to synthesize a library of copolymers for ASD formulations with the goal of training an ML model to inform the impact of polymer design on ASD solid-state stability.[6] PET-RAFT is characterized by being air tolerant and offering excellent control over polymer properties, enabling the synthesis of diverse polymer libraries in an open-air, automated setup with a liquid handling robot.[23−25] Automated PET-RAFT is therefore the first step

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

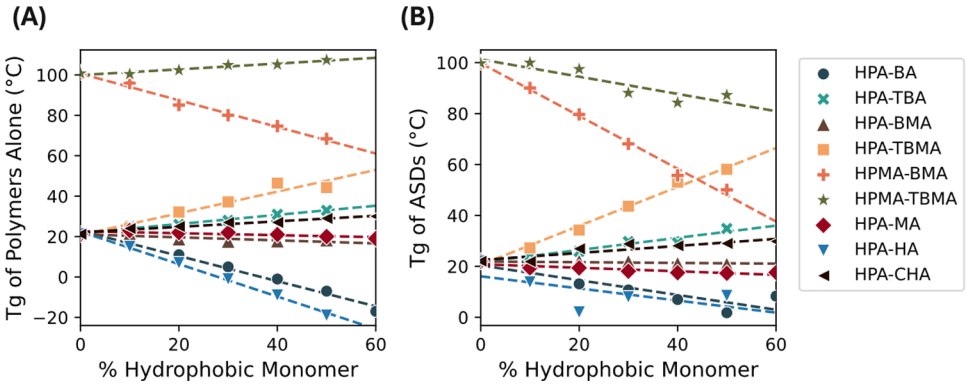


Figure 2. T_g as a function of hydrophobic monomer content for (A) polymers alone and (B) ASDs loaded with 20 wt % probucol. T_g for each polymer alone and each ASD was determined through mDSC using the second heat cycle. Samples were grouped by the two monomers present in the copolymer of each formulation (monomer pairing), then T_g was plotted against increasing hydrophobic monomer content.

in a streamlined pipeline for polymer synthesis, characterization, and subsequent training of an ML model to inform structure−function relationships which dictate ASD behaviors. We characterized a library of 50 unique copolymers and modulated hydrophobic monomer content, monomer side chain geometry, and polymer backbone methylation to study the impact of polymer design on the T_g of ASDs. We hypothesize that ML can be applied to inform structure−function relationships and aid in the development of novel polymer excipients for ASD design.

The copolymers of this library were combined with the model drug probucol to form ASDs. Probucol was chosen as the system's model drug, as it has served as an example hydrophobic API in an extended library of work studying polymer excipients for ASD formulation.[6,26−29] T_gs were acquired through modulated differential scanning calorimetry (mDSC) for polymer alone and for 20 wt % probucol loaded ASDs. These T_gs were compiled into a data set, along with the copolymers' compositional information, and were used to train a Random Forest Regression model to predict T_g and inform feature importance. The results of this study suggest that polymer composition data provides sufficient information for predicting T_g with relatively high accuracy. Although the data set is small by ML standards, highly linear correlations in the data suggest that a larger data set may not be required to accurately capture general trends within this particular polymer library. This study serves as a first step toward mapping the polymer excipient design space to begin identifying excipient characteristics that may increase ASD stability.

Successful excipients typically have high T_g, which is associated with lowered drug mobility and, subsequently, extended shelf life and inhibited crystal formation.[6,9] The scope of this work focuses on identifying trends between copolymer design and T_g in film-cast ASDs. We modulated copolymer hydrophilic/lipophilic balance, side chain geometry, and backbone methylation to explore how design criteria impact T_g.

RESULTS

Differential Scanning Calorimetry. To train an ML model to map structure−function relationships of copolymer and ASD systems, a suitable data set is required. Of the copolymer library explored in the scope of this work, 32 copolymers were newly synthesized, and 18 copolymers were included from a previously explored library (Figure 1).[6] The T_g of the library was assessed both with and without 20 wt % probucol loading to determine how presence of API impacts the T_g of the system. The T_g values reported in this work (Figure 2, Table S3.1) are the result of the second heat cycle to mitigate any thermal history present in polymer or ASD samples. The experimentally derived T_g values were compared to T_g predicted by the Fox equation, a common empirical model used to approximate T_g for copolymers. As expected, the Fox equation predicts T_g accurately for polymers alone; however, the equation becomes less accurate for ASDs, as it fails to account for the impact of drug loading on T_g (Table S4.2).

Polymers without drug loading exhibited highly linear trends in T_g as a function of hydrophobic monomer content when grouped by monomer pairing (Figure 2A). In these samples, unmethylated HPA and methylated HPMA served as hydrophilic monomers. These hydrophilic monomers were copolymerized with a hydrophobic monomer with feed ratios of 0−60% hydrophobic content. The hydrophobic monomers differed from one another in backbone methylation and side chain geometry. Figure 2 clearly demonstrates the contrast between samples with a fully methylated backbone (HPMA-BMA or HPMA-TBMA) and samples with a partially methylated or unmethylated backbone. Copolymers containing one or more methacrylate monomer (HPMA, BMA, and/or TBMA) also exhibit higher T_g than unmethylated backbones, which further supports the theory that increased copolymer rigidity is associated with higher T_g. Backbone methylation is therefore a design feature with the potential to significantly increase the T_g of an ASD.

Side chain geometry of the hydrophobic monomer also appears to heavily influence T_g trends. In hydrophobic monomers with linear side chains (BA, BMA, MA, or HA), T_g decreases with increasing hydrophobic monomer content. Conversely, copolymers with branched or cyclic side chains on the hydrophobic monomer (TBA, TBMA, or CHA) display increasing T_g with increasing hydrophobic monomer content. This behavior agrees with Gibbs and DiMarzio's thermodynamic theory, wherein bulkier and stiffer side chain groups contribute to increased T_g, while longer, more flexible linear alkyl side chains result in lowered T_g.[30−32] There are a cluster of monomer pairings whose T_gs do not appear to vary significantly as a function of hydrophobic monomer content. In the T_g data for HPA-MA, HPA-BMA, HPA-CHA, and HPA-TBA, T_g remains relatively constant near 20 °C in both

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

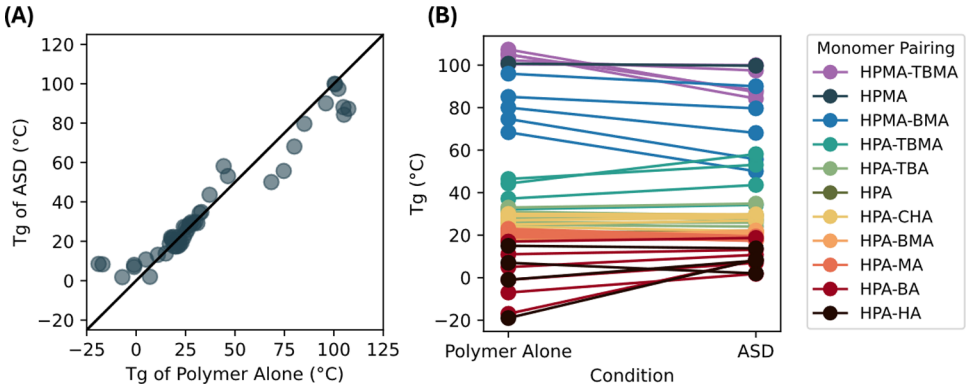


Figure 3. Impact of drug loading on T_g. (A) T_g of ASD formulation is plotted against the T_g of the associated polymer alone. The farther the polymer strays from the diagonal line, the greater the difference between T_g of the polymer alone and T_g of the corresponding ASD. (B) T_g is plotted for samples without ("Polymer Alone") and with ("ASD") drug loading. Importantly, drug loading does not uniformly increase or decrease T_g, making it difficult to predict T_g based on drug loading status alone. Samples are color coded by monomer pairing present in the polymer.

polymer alone samples and ASDs loaded with 20 wt % probucol. Linear versus nonlinear side chain geometry appears to play an important role in T_g and may therefore have predictive power as a training feature for an ML model.

In ASDs loaded with 20 wt % probucol, similar trends are observed compared to polymers alone, albeit with slightly less linearity (Figure 2B). Methylation still greatly increases T_g, as does a branched or cyclic side chain on the hydrophobic monomer. In some cases, the presence of probucol altered the slope of T_g as a function of hydrophobic monomer content. For example, HPMA-TBMA copolymers prior to drug loading exhibited a positive slope with increasing hydrophobic monomer content (increasing TBMA content), while addition of drug resulted in T_g decreasing as hydrophobic monomer content increases. The change in slope from positive to negative with the addition of drug suggests that there are potentially complex interactions at play between drug and copolymer, perhaps hinging on hydrophobic monomer content or side chain geometry. HPMA-BMA copolymers do not exhibit the same shift in slope from positive to negative, which may indicate that side chain geometry (branched TBMA vs linear BMA) impacts how a copolymer system is affected by the addition of an API. These unexpected trends in the data point to the potential usefulness of an ML model capable of untangling the impact of chemical structure, hydrophobic monomer content, and drug loading on the shelf stability of ASDs.

T_g trends vary widely depending on monomer selection, with copolymer composition and chemistry impacting not only the T_g of the polymer alone but also impacting how these polymers behave when formulated into a drug-loaded ASD. Drug loading significantly alters the T_g of several of the tested copolymers (Figure 3). Interestingly, drug loading did not always cause the effect that would be predicted by traditional empirical models. The presence of probucol does not uniformly increase or decrease T_g in ASDs. Instead, we see several copolymers without drug loading whose T_g is around 40 °C, and the addition of probucol increases the T_g of the ASD to be higher than that of the copolymer alone. Conversely, there exist several samples in which the addition of probucol causes a decrease in T_g to below the T_g of the polymer alone. These trends, which are not predicted by empirical models, are precisely the types of unexpected results that point to the usefulness of an appropriately trained ML model.

Prediction of T_g. The T_g data set obtained experimentally was used to train an ML model to understand the impact of chemical composition on ASD behaviors. The ML pipeline used in this work is largely informed by the steps outlined in a user's guide to machine learning previously published by our group.[33] We explored ten different ML models on the T_g of drug-loaded ASDs and polymers alone to determine which model was most appropriate for predicting T_g. A Random Forest Regressor model emerged as one of the top predictors of T_g in polymers alone and ASDs loaded with 20 wt % probucol (Figure S2). We proceeded to train a Random Forest Regressor with a 10-fold cross validation, which predicted copolymer T_g with an R^2 of 0.89 and predicted ASD T_g with an R^2 of 0.83 (Figure 4). Importantly, this model was trained using only compositional information and polymer design features. Using relatively sparse training data and a

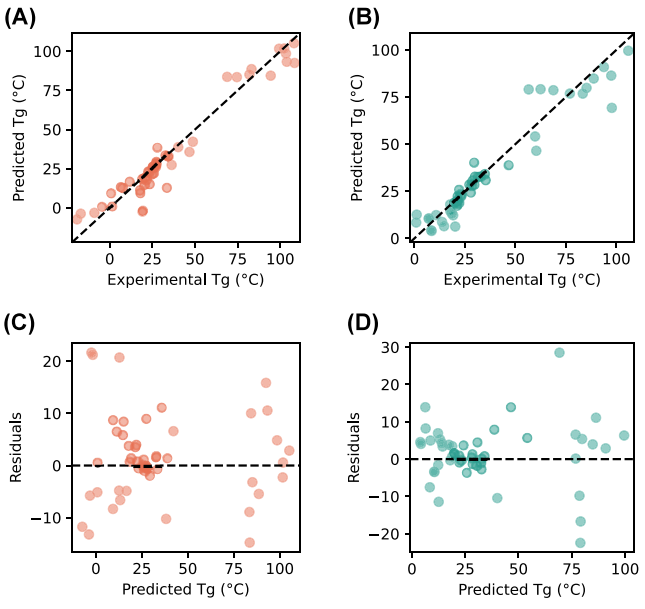


Figure 4. Prediction of T_g in (A) polymers alone and (B) drug loaded ASDs. A Random Forest Regressor with 400 estimators and 10-fold cross validation predicted T_g with $R^2 > 0.83$. Residual plots are included for both (C) polymer alone and (D) ASD prediction plots.

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

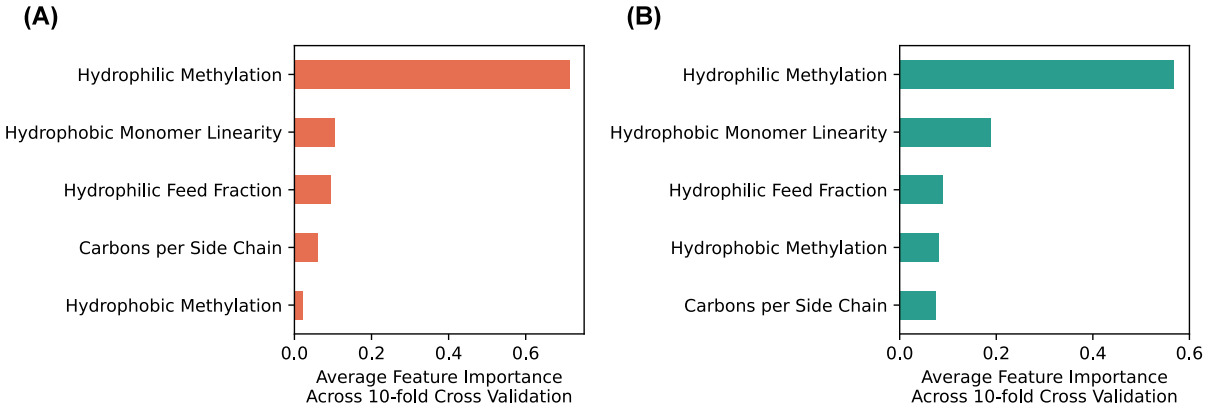


Figure 5. Analysis of feature importance on predicted T_g. Feature importance was calculated while predicting T_g of samples without (A) and with (B) 20 wt % probucol loading. In each fold of the 10-fold cross validation, importance values were calculated for every feature and averaged post cross validation. With and without drug loading, hydrophilic backbone methylation is ranked highest in importance, followed by linear vs nonlinear side chain geometry. In polymers alone, carbons per side chain was ranked of higher predictive importance than hydrophobic methylation, while the opposite was true for ASDs.

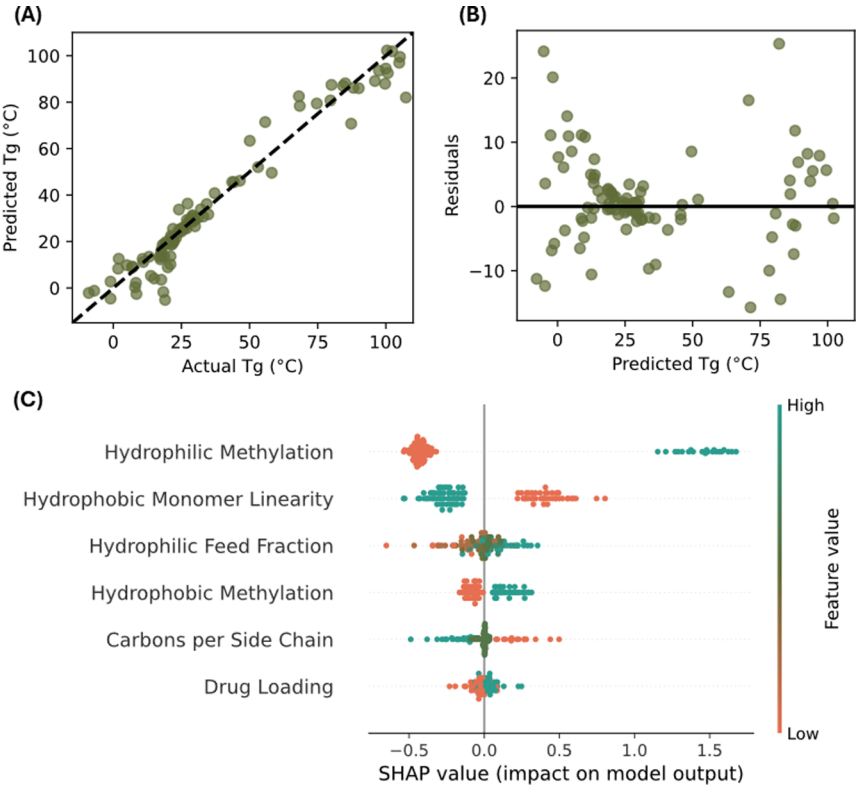


Figure 6. Prediction of T_g with drug loading as a feature.(A) Random Forest model T_g predictions plotted against experimental T_g values acquired by mDSC (R^2 = 0.88). (B) Residuals of prediction values. (C) SHAP plot ranking feature importance as well as impact of feature value on model output.

comparatively small data set of 50 copolymers, the simple Random Forest model is robust enough to predict T_g of polymer alone and of drug loaded ASDs.

These results indicate that T_g can be predicted with relatively high confidence using only theoretical copolymer structural information. Within this data set, the highly linear nature of T_g as a function of hydrophobic monomer content (Figure 2) may be a determining factor that facilitates clear mapping of structure−function relationships with limited experimental data. A relatively small library of experimentally tested copolymer designs may sufficiently train a model to predict T_g values for novel copolymer designs within the scope of the data set's design space. For example, these 50 copolymer compositions represent a design space that spans nine monomers which could be combined in myriad feed ratios to form novel designs. Researchers may therefore train an ML model on a small library of copolymers and go on to predict the T_g of immense theoretical copolymer libraries to develop potentially high-performing ASDs prior to synthesizing or characterizing any copolymers. This addresses a major obstacle in the study of copolymer/ASD structure−function relationships, which is the difficulty associated with designing, synthesizing, and characterizing in high throughput a library large enough to map the immense copolymer design space.

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

Structure−Function Relationships. The model's ability to accurately predict T_g suggests that structure−functional trends within the data set are being accurately captured by the model. Random Forest models include built-in methods to calculate and rank feature importance, which allows us to identify which design criteria were identified by the model as having the strongest impact on T_g. To determine the importance of each feature in the model, we tracked each feature's importance for every fold of the cross validation and averaged the importance across all folds (Figure 5A,B). We observed similar trends in importance for ASDs and polymers alone. Regardless of drug loading status, the most important structural feature appears to be the methylation of the hydrophilic monomer's backbone (presence of unmethylated HPA or methylated HPMA). The placement of backbone methylation as a leading predictor of T_g agrees with the experimental results, where T_g jumped from around 20 °C to over 100 °C when the monomers shifted from acrylates to methacrylates. The linearity or nonlinearity of the hydrophobic monomer follows as the next strongest predictor of T_g, indicating that side chain geometry can be an important consideration for excipient design. The hydrophilic feed fraction, which is a representation of the hydrophilic/lipophilic balance of the polymer, was ranked as the third strongest predictor of T_g. In ASDs, methylation of the hydrophobic monomer's backbone and the number of carbons per side chain on the hydrophobic monomer were ranked almost equally important, while number of carbons per side chain outranks hydrophobic methylation in predictive importance for polymers alone.

Tracking feature importance enables us to determine which polymer design criteria are most influential on functional characteristics of the tested ASDs. While feature importance ranks design criteria by predictive power, it does not inform the effect each feature has on measured T_g. These values can inform future copolymer design, as future libraries can be designed with these structure/function relationships in mind to maximize ASD T_g. For a more in-depth study on the impact of each of these design features on T_g, we extended our analysis to include drug loading as a feature and delved more deeply into both feature importance and each feature impact's on the model's predictions.

Drug Loading as a Predictive Feature. The difference in behavior between polymers alone and probucol-loaded ASDs led us to question the role of probucol in determining the T_g of the system. We trained a model on the complete data set, reconfigured to include drug loading status as a sixth training feature to explore this concept. Following the same model selection as performed for polymers alone and for ASDs, we trained the data set on 10 potential ML models and identified Random Forest as a top performer (Figure S3).

We moved forward with the Random Forest model to predict T_g of all polymers alone and all ASD samples. The inclusion of drug loading as a feature did not significantly alter the model's performance, which had an average R^2 of 0.88 across a 10-fold cross validation (Figure 6A,B). We proceeded to identify feature importances and Shapley Additive Explanations (SHAP) values for this reconfigured data set (Figure 6C).[36] The model ranked features by importance in the same order as the model trained on ASD data, now with drug loading ranked as having the least predictive power. Importantly, SHAP values also illustrate the predicted impact a design feature may have on T_g. For example, backbone methylation dominated T_g predictions, which again agrees with the experimental results of Figure 2; however, Figure 6C provides further context by illustrating that samples with low values for hydrophilic methylation are predicted to exhibit lower T_gs, while higher values for hydrophilic methylation are predicted to increase T_g. Presence of linear side chain geometry is predicted to decrease T_g, while presence of nonlinear side chain geometry is predicted to increase T_g. SHAP predicts that samples with higher hydrophilic feed fractions and methylated hydrophilic monomers will exhibit higher T_gs. While drug loading does not drastically impact the T_g of most copolymers tested within the scope of this study, we were able to identify that in this data set, drug loading is predicted to slightly increase T_g, while no drug loading is predicted to slightly decrease T_g. This weak positive correlation between drug loading and T_g is only representative of the data set explored in this work, and cannot be used to draw firm conclusions concerning general relationships between drug loading and T_g. Works by Rieneke and others have explored larger drug loading ranges and found that T_g does not uniformly increase or decrease with increased drug loading, indicating that further study is necessary to determine the explicit impact of API on the T_g of an ASD.[12,34,35] In order to clearly identify the relationship between drug loading and T_g, future studies will examine a larger range of drug loading conditions, from 0−50 wt % API. The current results indicate the potential usefulness of including drug loading as a feature to train a ML model, and point to the added benefit of SHAP analysis to demonstrate how the model weighs each design feature's contribution to predict T_g.

DISCUSSION

The scope of this work does not intend to provide a complete map of the polymer excipient design space, but rather to act as a starting point for the development of predictive tools for ASD excipient design. We attempted to predict T_g of copolymers loaded with the model drug probucol using a relatively small data set of 50 unique copolymers, of which 18 were previously studied and 32 were newly synthesized.[6] This data set explores only a fraction of the available polymer design space, leaving ample room for further evaluation. The monomer families explored in this study are limited to methacrylates and acrylates, and the side chain geometries are simply labeled as either "linear" or "nonlinear." Future work in this area could explore other monomer families, such as acrylamides, or delve deeper into how specific side chain geometries impact T_g, as well as examining a broader range of drug loading conditions. The experimental process was largely automated from polymer synthesis on the LHR to auto sampling mDSC, facilitating a design-build-test-learn loop that is easily amenable to future studies. The experimental results from the build-test portion of the loop were used in the learn-design portion of the loop to train an ML model to predict T_g and inform important structure/function relationships driving ASD behavior. Design-build-test-learn loops build upon all previous data to iteratively improve the model's predictive ability.

Random Forest Regressor models were trained to predict T_g of polymers alone, polymer/drug ASDs formulated with 20 wt % probucol loading, and finally, a compiled data set of 99 samples that included drug loaded ASDs and polymers alone. Importantly, these models were trained on polymer design features such as monomer methylation, side chain geometry,

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

and hydrophilic monomer feed fraction. No experimental data is needed to predict T_g, meaning this model is able to predict the T_g of polymer excipient designs prior to any synthesis or characterization. The model predicted T_g in polymers alone with an R^2 of 0.89 and predicted T_g of ASDs with an R^2 of 0.83. When drug loading was included as a feature in the data set, T_g was predicted with an average R^2 of 0.88.

While good model performance is an important aspect of this work, the more clinically relevant discussion centers on the polymer design features that are predictors of successful ASD development. Feature importance and SHAP values indicate that higher feed ratios of methylated monomers and monomers with nonlinear side chains are associated with higher T_g. Namely, the model predicts that T_g will be higher for a polymer that is designed with methylation on the backbones of both the hydrophilic and hydrophobic monomers, a larger hydrophilic monomer feed fraction, and a bulky, nonlinear side chain on the hydrophobic monomer. With these design features in mind, the development of novel polymer excipients for high T_g ASD development is more straightforward than the existing approach, which is essentially an educated guess-and-check method for polymer excipient design. We were able to correlate structural features to the functional output of T_g, a predictor of ASD success, with a small data set and limited training information for the model, which speaks to the potential for more sophisticated models to map this design space more effectively and greatly facilitate ASD excipient development.

CONCLUSIONS

It is difficult to design novel polymer excipients due to the limited number of commercially available polymer excipients used in ASDs. High T_g is a feature shared among these few commercial excipients; therefore, T_g has become an important predictor of an ASD's solid-state stability and ability to prevent drug crystallization. Predicting T_g in novel polymer excipients is complex, partly due to the experimental burden associated with testing a large enough library of copolymer designs to effectively map relationships between polymer design criteria and ASD T_g. This study underlines the potential for ML to streamline the materials design process and facilitate novel copolymer excipient design for ASD formulation. Following an existing user's guide to produce ML for biomaterials design, we employed a simple but effective approach toward novel polymer excipient research for pharmaceutical ASD design.[33] The model predicts T_g with high accuracy ($R^2 > 0.83$) for samples with and without drug loading, performing equally well when drug loading is included as a training feature. The model is amenable to SHAP analysis, which offers insights into the importance and impact each design feature has on the target parameter, T_g. Increasing the size and diversity of the data set will undoubtedly provide a more detailed map of the design space; however, we hope this work serves as a demonstration of the relative simplicity with which structure−function relationships may be explored and evaluated. To further assess the predictive power of this model, future studies will implement active learning to predict and test the T_g of a novel copolymer library composed of the monomers tested in this work, combined in more complicated designs (i.e., varied chain length, 3+ monomer systems, diversified drug loading conditions, etc.).

EXPERIMENTAL SECTION

Materials. 2-Hydroxypropyl acrylate (HPA) monomer (purity 99%) was purchased from Polysciences. Butyl methacrylate (BMA) monomer (purity 99%) was purchased from VWR. Monomers 2-hydroxypropyl methacrylate (HPMA), butyl acrylate (BA), *tert*-butyl acrylate (TBA) and *tert*-butyl methacrylate (TBMA) were purchased from Sigma-Aldrich, with purities of 97, 99, 98, and 98%, respectively. Chain transfer agent (CTA) 4-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid (purity 97%) was purchased from Sigma-Aldrich. Zinc(II) tetraphenylporphyrin (ZnTPP) (purity 98%) initiator was purchased from Fisher Scientific. Dimethyl sulfoxide (DMSO) was purchased from Sigma-Aldrich. Prior to handling, monomers were disinhibited as needed by passing through a column of inhibitor remover beads purchased from Sigma-Aldrich. Monomer aliquots were prepared at 2 M, CTA aliquots were prepared at 50 mM, and ZnTPP aliquots were prepared at 2 mM in DMSO. Probucol (purity 98%) was purchased from Combi-Blocks.

Automated PET-RAFT Polymerization. Polymers were prepared in an automated PET-RAFT polymerization setup using a Hamilton Microlab STARlet liquid handling robot (LHR).[6,24,25] The LHR was loaded with 1 mL aliquots of 2 M monomers, 50 mM CTA, and 2 mM ZnTPP, along with approximately 50 mL of DMSO and a 96-well polypropylene plate. Polymer feed ratios (reported in Table S3.1), reagent volumes, well positions, and general procedure for the LHR to follow were detailed in an Excel workbook and executed by a Python script controlling the LHR. Degree of polymerization (DP) was fixed at 200, hydrophilic monomer content ranged from 40−100 mol %, hydrophobic monomer content ranged from 0−60 mol %, and reaction volumes were prepared in triplicate to ensure polymers were synthesized in sufficient quantity. Monomer to CTA ratios were fixed at 200:1, and CTA to ZnTPP ratios were fixed at 100:1. Once reagents were dispensed into the 96-well plate, the reagents were mixed well by pipetting. The plate was then sealed with tape and irradiated with a 560 nm LED light (5 mW/cm^2) for 16 h. Postpolymerization, 30 μL was taken for gel permeation chromatography (GPC) characterization.

Polymer Purification. Postpolymerization, samples were diluted 3× in DMSO, followed by a 10x dilution in ultrapure water. Samples were transferred to 3.5 kDa MWCO dialysis tubing, sealed, and dialyzed in 4 L of ultrapure water for 48 h with four overall water changes. After dialysis, samples were lyophilized in 50 mL tubes.

Gel Permeation Chromatography (GPC). Polymers were diluted to 2 mg mL^{-1} in DMF, passed through 0.5 μm PTFE filters and loaded into GPC vials. An Agilent 1260 Infinity II system with a DMF/LiBr mobile phase was used to identify molecular weights (M_w and M_n) and dispersity ($Đ$) against PMMA standards. GPC results are reported in SI.

Film-Cast ASD Formation. Probucol was dissolved in acetone to 6.67 mg mL^{-1} and added to preweighed polymers to form polymer/drug solutions at 20 wt % probucol loading (polymer concentration 26.67 mg mL^{-1}). Solutions were vortexed to mix. In 10 μL increments, polymer/drug solutions were added to preweighed Tzero DSC pans (TA Instruments) to a total volume of 60 μL. Pans were dried overnight at ambient temperature in an Isotemp vacuum oven to form film

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

casts. The DSC pans were weighed after film casting and hermetically sealed with Tzero lids (TA Instruments).

Modulated Differential Scanning Calorimetry. Probucol, polymers, and polymer/drug film casts were run on a TA Discovery DSC with slightly varied methods, as reported below. All T_g values are reported as analyzed from the second heat cycle. The second heat cycle was used in this work, as there were substantial residual solvent effects in the first cycle thermograms of samples with low T_g. In the second heat cycle, the plasticization effect from residual solvent was removed and T_g of polymer and API could be accurately assessed. A table of T_g values is reported in SI.

Probucol. Powder probucol was weighed into a Tzero DSC pan to approximately 5.0 mg. The pan was weighed and hermetically sealed with a Tzero lid punctured by two pinholes. The mDSC system was equilibrated to 10.0 °C and isothermally held for 5 min. A temperature ramp was applied at 5.0 °C min^{-1} to 150.0 °C, at which point the temperature was ramped to 10.0 °C at a rate of 5.0 °C min^{-1}.

Polymers and Polymer/Drug Film Casts. Lyophilized polymer was weighed into Tzero DSC pans to 2.0−5.0 mg, weighed, and hermetically sealed with Tzero lids punctured by two pinholes. Film-cast samples were prepared as reported above. The mDSC instrument was equilibrated to −50.0 °C and held for 5 min. The temperature was modulated with an amplitude of 0.5 °C and a period of 60 s and the system ramped to 170.0 °C at a rate of 2.0 °C min^{-1}. For the cooling cycle, a 10.0 °C min^{-1} ramp was applied to bring the system down to −50.0 °C, followed by a 2 min isothermal hold. For the second heat cycle, a modulated temperature ramp was applied with an amplitude of 0.5 °C, a period of 60 s, and a heating rate of 2.0 °C min^{-1} to a final temperature of 160.0 °C.

Data Preparation and Machine Learning. The polymer database used for this work included 32 polymers newly synthesized for this study and 18 polymers from a previously published study.[6] This database of characterization and composition data of 50 unique polymers was trimmed to include only columns relevant to training the ML model to predict T_g with and without 20 wt % probucol loading. There was a single T_g value missing for an ASD composed of HPA and 40% HA, and this sample was removed from the data set. As a result, there were 50 polymer T_gs recorded without probucol loading and 49 T_gs recorded of drug-loaded polymer ASDs. When the model was trained to predict T_g without drug loading, 50 samples were used. When the model was trained to predict T_g in drug-loaded ASDs, 49 samples were used. When the model was trained with drug loading as a training feature, 99 samples were used, a combination of the polymers alone and the ASD data.

Monomer family (acrylate, methacrylate, or combination), side chain geometry (linear, branched, or cyclic), and presence of each individual monomer were one-hot encoded. The complete, cleaned data set contained the following features: Polymer ID, hydrophobic monomer content (%), monomer feed ratio information, T_g of the polymer alone, T_g of the polymer/drug film cast, theoretical molecular weight, hydrophilic monomer methylation status, hydrophobic monomer methylation status, hydrophilic feed fraction, hydrophobic monomer linearity or nonlinearity, and number of carbons per hydrophobic monomer side chain. Theoretical molecular weight here refers to the predicted molecular weight of the polymer based on monomer molecular weight, monomer feed ratios, and target DP of the polymer, and was not ultimately used as a training feature in the data set.

There were well over a dozen features before principal component analysis (PCA) and dimensionality reduction. PCA indicated that nearly 100% of the cumulative explained variance could be captured through seven features, and about 85% of the cumulative explained variance could be captured through five features (Figure S3.5). After dimensionality reduction, the five feature columns used to train the model simply included one-hot encoded variables for hydrophilic methylation, hydrophobic methylation, and hydrophobic monomer linearity, along with features for the hydrophilic feed fraction within the polymer and the number of carbons per hydrophobic monomer side chain. Importantly, these five features are determined by polymer design alone, not through experimentally derived data. The model therefore does not require any characterization or testing to determine a polymer design's impact on the predicted T_g. This feature selection affords us the potential to design novel polymers within the design space of this library and predict their T_gs without needing to synthesize or characterize said polymers.

We tested an ensemble of ML models and determined that a simple Random Forest Regressor performed well to predict T_g in both polymers alone and polymer/drug ASDs, as well as accurately predicting T_g when drug loading was included as a training feature in the data set (Figures S3.3 and S3.4). This Random Forest Regressor model was employed first to predict polymer T_g alone, next to predict polymer/drug film cast T_g, and finally to predict T_g on the entire data set with drug loading as a sixth feature. In this last case, we grouped samples by their polymer content to avoid data leakage. Namely, copolymer compositions were segregated such that a copolymer composition present in the testing set could not appear in the training set as a drug-loaded ASD with an identical copolymer. This latter model allowed insight into how drug loading impacts T_g and the importance of drug loading as a training feature. We used a power transformer scaler along with a 10-fold cross validation. Feature importance and Shapley Additive Explanations (SHAP) values were calculated for the Random Forest model to inform structure−function relationships.[36]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.bioconjchem.4c00294.

Example chemical structure of a PET-RAFT polymer, GPC data, ML data, mDSC data, and comparison between experimental T_g values and empirically derived T_g values based on the Fox equation (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Ashish Punia − *Analytical Research and Development, MRL, Merck & Co., Inc., Rahway, New Jersey 07065, United States*; Email: ashish.punia@merck.com

Adam J. Gormley − *Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States*; orcid.org/0000-0002-2884-725X; Email: adam.gormley@rutgers.edu

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372

Authors

Elena J. Di Mare − *Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States*

Matthew S. Lamm − *Analytical Research and Development, MRL, Merck & Co., Inc., Rahway, New Jersey 07065, United States;* orcid.org/0000-0002-2787-0642

Timothy A. Rhodes − *Analytical Research and Development, MRL, Merck & Co., Inc., Rahway, New Jersey 07065, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.bioconjchem.4c00294

Funding

This work was supported by NSF 2009942 and 2309852 including an INTERN supplement on these awards. Additional funding was provided by NIH NIGMS R35GM138296 and NIH T32 GM135141. E.D. would like to thank Matthew Tamasi and Cesar Ramirez for their guidance in the preparation of the Python script used throughout this work.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

PET-RAFT	photoinduced electron/energy transfer-reversible addition−fragmentation chain-transfer polymerization
ASD	amorphous solid dispersion
API	active pharmaceutical ingredients
T_g	glass transition temperature
ML	machine learning
mDSC	modulated differential scanning calorimetry
GPC	gel permeation chromatography
HPA	2-hydroxypropyl acrylate
HPMA	2-hydroxypropyl methacrylate
BA	butyl acrylate
BMA	butyl methacrylate
TBA	*tert*-butyl acrylate
TBMA	*tert*-butyl methacrylate
HA	hexyl acrylate
CHA	cyclohexyl acrylate
MA	methyl acrylate
CTA	chain transfer agent
ZnTPP	zinc(II) tetraphenylporphyrin
DMSO	dimethyl sulfoxide
LHR	liquid handling robot

■ REFERENCES

(1) Jelić, D. Thermal Stability of Amorphous Solid Dispersions. *Molecules* **2021**, *26* (1), 238.

(2) Organization, W. H. Stability testing of active pharmaceutical ingredients and finished pharmaceutical products. *WHO Technical report series*, 2009, vol *953*; pp 87−123.

(3) Jones, E. C. L.; Bimbo, L. M. Crystallisation Behaviour of Pharmaceutical Compounds Confined within Mesoporous Silicon. *Pharmaceutics* **2020**, *12* (3), 214.

(4) Van den Mooter, G. The use of amorphous solid dispersions: A formulation strategy to overcome poor solubility and dissolution rate. *Drug Discovery Today: Technologies* **2012**, *9* (2), e79−e85.

(5) Shi, Q.; Chen, H.; Wang, Y.; Wang, R.; Xu, J.; Zhang, C. Amorphous Solid Dispersions: Role of the Polymer and Its Importance in Physical Stability and In Vitro Performance. *Pharmaceutics* **2022**, *14* (8), 1747.

(6) Upadhya, R.; Punia, A.; Kanagala, M. J.; Liu, L.; Lamm, M.; Rhodes, T. A.; Gormley, A. J. Automated PET-RAFT Polymerization Towards Pharmaceutical Amorphous Solid Dispersion Development. *ACS Appl. Polym. Mater.* **2021**, *3* (3), 1525−1536.

(7) Ting, J. M.; Porter, W. W., 3rd; Mecca, J. M.; Bates, F. S.; Reineke, T. M. Advances in Polymer Design for Enhancing Oral Drug Solubility and Delivery. *Bioconjug Chem.* **2018**, *29* (4), 939−952.

(8) Qian, F.; Huang, J.; Hussain, M. A. Drug-polymer solubility and miscibility: Stability consideration and practical challenges in amorphous solid dispersion development. *J. Pharm. Sci.* **2010**, *99* (7), 2941−2947.

(9) Sarabu, S.; Kallakunta, V. R.; Bandari, S.; Batra, A.; Bi, V.; Durig, T.; Zhang, F.; Repka, M. A. Hypromellose acetate succinate based amorphous solid dispersions via hot melt extrusion: Effect of drug physicochemical properties. *Carbohydr. Polym.* **2020**, *233*, No. 115828.

(10) Wolbert, F.; Fahrig, I. K.; Gottschalk, T.; Luebbert, C.; Thommes, M.; Sadowski, G. Factors Influencing the Crystallization-Onset Time of Metastable ASDs. *Pharmaceutics* **2022**, *14* (2), 269.

(11) Ma, X.; Williams, R. O. Characterization of amorphous solid dispersions: An update. *Journal of Drug Delivery Science and Technology* **2019**, *50*, 113−124.

(12) Červinka, C.; Fulem, M. Structure and Glass Transition Temperature of Amorphous Dispersions of Model Pharmaceuticals with Nucleobases from Molecular Dynamics. *Pharmaceutics* **2021**, *13* (8), 1253.

(13) Thybo, P.; Pedersen, B. L.; Hovgaard, L.; Holm, R.; Müllertz, A. Characterization and Physical Stability of Spray Dried Solid Dispersions of Probucol and PVP-K30. *Pharm. Dev. Technol.* **2008**, *13* (5), 375−386.

(14) Baird, J. A.; Taylor, L. S. Evaluation of amorphous solid dispersion properties using thermal analysis techniques. *Adv. Drug Delivery Rev.* **2012**, *64* (5), 396−421.

(15) Volgin, I. V.; Batyr, P. A.; Matseevich, A. V.; Dobrovskiy, A. Y.; Andreeva, M. V.; Nazarychev, V. M.; Larin, S. V.; Goikhman, M. Y.; Vizilter, Y. V.; Askadskii, A. A.; et al. Machine Learning with Enormous "Synthetic" Data Sets: Predicting Glass Transition Temperature of Polyimides Using Graph Convolutional Neural Networks. *ACS Omega* **2022**, *7* (48), 43678−43691.

(16) AlFaraj, Y. S.; Mohapatra, S.; Shieh, P.; Husted, K. E. L.; Ivanoff, D. G.; Lloyd, E. M.; Cooper, J. C.; Dai, Y.; Singhal, A. P.; Moore, J. S.; et al. A Model Ensemble Approach Enables Data-Driven Property Prediction for Chemically Deconstructable Thermosets in the Low-Data Regime. *ACS Cent Sci.* **2023**, *9* (9), 1810−1819.

(17) Jiang, Z.; Hu, J.; Marrone, B. L.; Pilania, G.; Yu, X. B. A Deep Neural Network for Accurate and Robust Prediction of the Glass Transition Temperature of Polyhydroxyalkanoate Homo- and Copolymers. *Materials* **2020**, *13* (24), 5701.

(18) Jiang, J.; Lu, A.; Ma, X.; Ouyang, D.; Williams, R. O., 3rd. The applications of machine learning to predict the forming of chemically stable amorphous solid dispersions prepared by hot-melt extrusion. *Int. J. Pharm. X* **2023**, *5*, No. 100164.

(19) Han, R.; Xiong, H.; Ye, Z.; Yang, Y.; Huang, T.; Jing, Q.; Lu, J.; Pan, H.; Ren, F.; Ouyang, D. Predicting physical stability of solid dispersions by machine learning techniques. *J. Controlled Release* **2019**, *311*−*312*, 16−25.

(20) Dong, J.; Gao, H.; Ouyang, D. PharmSD: A novel AI-based computational platform for solid dispersion formulation design. *Int. J. Pharm.* **2021**, *604*, No. 120705.

(21) Walden, D. M.; Bundey, Y.; Jagarapu, A.; Antontsev, V.; Chakravarty, K.; Varshney, J. Molecular Simulation and Statistical Learning Methods toward Predicting Drug−Polymer Amorphous Solid Dispersion Miscibility, Stability, and Formulation Design. *Molecules* **2021**, *26* (1), 182.

(22) Gormley, A. J.; Webb, M. A. Machine learning in combinatorial polymer chemistry. *Nature Reviews Materials* **2021**, *6* (8), 642−644.

(23) Gormley, A. J.; Yeow, J.; Ng, G.; Conway, O.; Boyer, C.; Chapman, R. An Oxygen-Tolerant PET-RAFT Polymerization for

Screening Structure−Activity Relationships. *Angew. Chem., Int. Ed.* **2018**, *57* (6), 1557−1562.
(24) Lee, J.; Mulay, P.; Tamasi, M. J.; Yeow, J.; Stevens, M. M.; Gormley, A. J. A fully automated platform for photoinitiated RAFT polymerization. *Digital Discovery* **2023**, *2* (1), 219−233.
(25) Tamasi, M.; Kosuri, S.; DiStefano, J.; Chapman, R.; Gormley, A. J. Automation of Controlled/Living Radical Polymerization. *Advanced Intelligent Systems* **2020**, *2* (2), No. 1900126.
(26) Dalsin, M. C.; Tale, S.; Reineke, T. M. Solution-state polymer assemblies influence BCS class II drug dissolution and supersaturation maintenance. *Biomacromolecules* **2014**, *15* (2), 500−511.
(27) Tale, S.; Purchel, A. A.; Dalsin, M. C.; Reineke, T. M. Diblock Terpolymers Are Tunable and pH Responsive Vehicles To Increase Hydrophobic Drug Solubility for Oral Administration. *Mol. Pharmaceutics* **2017**, *14* (11), 4121−4127.
(28) Broman, E.; Khoo, C.; Taylor, L. S. A comparison of alternative polymer excipients and processing methods for making solid dispersions of a poorly water soluble drug. *Int. J. Pharm.* **2001**, *222* (1), 139−151.
(29) Ting, J. M.; Navale, T. S.; Jones, S. D.; Bates, F. S.; Reineke, T. M. Deconstructing HPMCAS: Excipient Design to Tailor Polymer-Drug Interactions for Oral Drug Delivery. *ACS Biomater Sci. Eng.* **2015**, *1* (10), 978−990.
(30) Xie, R.; Weisen, A. R.; Lee, Y.; Aplan, M. A.; Fenton, A. M.; Masucci, A. E.; Kempe, F.; Sommer, M.; Pester, C. W.; Colby, R. H.; et al. Glass transition temperature from the chemical structure of conjugated polymers. *Nat. Commun.* **2020**, *11* (1), 893.
(31) Dudowicz, J.; Freed, K. F.; Douglas, J. F. The Glass Transition Temperature of Polymer Melts. *J. Phys. Chem. B* **2005**, *109* (45), 21285−21292.
(32) Hinrichsen, G. Fundamentals of polymer science by Paul C. Painter and Michael M. Coleman, Technomic Publishing Co., Lancaster, PA, 1994, 433 pp., ISBN 1−56676-152-2. *Acta Polym.* **1995**, *46* (2), 183−183.
(33) Meyer, T. A.; Ramirez, C.; Tamasi, M. J.; Gormley, A. J. A User's Guide to Machine Learning for Polymeric Biomaterials. *ACS Polymers Au* **2023**, *3* (2), 141−157.
(34) Ting, J. M.; Navale, T. S.; Bates, F. S.; Reineke, T. M. Design of Tunable Multicomponent Polymers as Modular Vehicles To Solubilize Highly Lipophilic Drugs. *Macromolecules* **2014**, *47* (19), 6554−6565.
(35) Kabedev, A.; Zhuo, X.; Leng, D.; Foderà, V.; Zhao, M.; Larsson, P.; Bergström, C. A. S.; Löbmann, K. Stabilizing Mechanisms of β-Lactoglobulin in Amorphous Solid Dispersions of Indomethacin. *Mol. Pharmaceutics* **2022**, *19* (11), 3922−3933.
(36) Lundberg, S.; Lee, S.-I. *A Unified Approach to Interpreting Model Predictions*, 2017.

https://doi.org/10.1021/acs.bioconjchem.4c00294
Bioconjugate Chem. 2024, 35, 1363−1372