Data-driven design of novel polymer excipients for pharmaceutical amorphous solid dispersions

Elena J. Di Mare[‡], Ashish Punia[§], Matthew S. Lamm[§], Timothy A. Rhodes[§], Adam J. Gormley[‡]
[‡]*Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854, USA*
[§]*Preformulation Sciences, MRL, Merck & Co., Inc., Rahway, NJ 07065, USA*

Statement of Purpose: A significant branch of pharmaceutical product formulation is geared towards extending the solid-state stability of active pharmaceutical ingredients (APIs) for oral delivery by preventing API crystallization. APIs lose bioavailability and efficacy upon crystallization, which necessitates maintaining the API in an amorphous state for as long as possible until administration to patients. APIs are commonly combined with polymers as anti-plasticization agents to extend shelf stability. Polymers and other excipients are formulated with APIs into amorphous solid dispersions (ASDs), which maintain APIs in a bioactive, amorphous state for oral delivery. The list of commercially used polymer excipients for ASDs is extremely limited; however, the development of novel polymer excipients is hindered by the lack of quantitative structure-function relationships that define which polymer design features are desirable for extending ASD stability. Higher glass transition temperature (T_g) is associated with extended ASD stability, making T_g a potential target for ASD stability improvement. Here, we report on our recently published work, in which we employ machine learning (ML) to study the structure-function relationships that connect polymer design with T_g by forming ASDs of polymers and probucol, our model API.[1] Namely, we varied polymer design features such as hydrophilic feed fraction, length of monomer side chains, backbone methylation, and side chain linearity to monitor the effect of these parameters on our target parameter, T_g. Experimentally derived T_g data was used to train an ML model to map structure-function relationships and streamline the design of novel polymer excipients for ASDs.

Methods: Monomers included 2-hydroxypropyl acrylate, 2-hydroxypropyl methacrylate, butyl acrylate, butyl methacrylate, *tert*-butyl acrylate, *tert*-butyl methacrylate, hexyl acrylate, cyclo-hexyl acrylate, and methyl acrylate. All monomers were polymerized via photoinduced electron/energy transfer reversible addition-fragmentation chain transfer polymerization to form polymers with degrees of polymerization fixed at 200. Polymers were film-cast with probucol at 20 wt% drug loading, and T_g was determined via differential scanning calorimetry (DSC). T_g data and polymer design features were used to train an ML model with 10-fold cross-validation to predict T_g and inform structure-function relationships.

Results: A dataset of T_g data and polymer design features was compiled from 50 polymers tested with and without 20 wt% probucol loading. Depending on polymer composition and drug loading, T_g ranged from -19ºC to 107ºC. The resulting dataset contained 100 entries, which were used to train the ML model to predict T_g and identify the impact of each design feature on T_g (**Figure 1**). Shapley Additive Explanations (SHAP) analysis determined that methylation of the hydrophilic monomer was the strongest predictor of T_g for the tested polymer library.[2] Methylation was associated with higher T_g, while hydrophilic acrylate monomers with unmethylated backbones were associated with lower T_g. The next strongest predictor was hydrophobic monomer linearity on the side chain, in which case T_g decreased when side chains were linear and increased when side chains were nonlinear (i.e., branched or cyclic). Further, bulkier side groups were predicted to increase T_g. The impact of probucol on T_g was the weakest predictive feature, with probucol loading predicting a slight increase in T_g.

Conclusions: This study underlines the potential for ML to streamline the materials design process and facilitate novel copolymer excipient design for ASD formulation. The ML model was able to rank design features in order of predictive power and inform the effect of individual design features on T_g. We hope this work demonstrates the relative simplicity with which ML can be implemented into the materials design process. Future experiments will leverage the structure-function relationships studied in this work to develop novel polymer excipients for increased stability of ASDs.

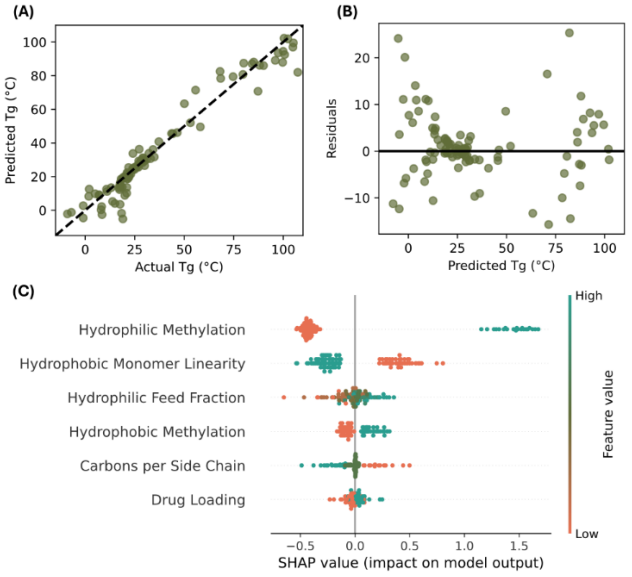

Figure 1. Prediction of T_g. (**A**) Random Forest model T_g predictions plotted against experimental T_g values acquired by DSC (R^2 = 0.88). (**B**) Residuals of prediction values. (**C**) SHAP summary plot ranking feature importance as well as the impact of feature value on model output.

References:

1. Di Mare EJ. Bioconjug. Chem. 2024;*35*:1363-1372.
2. Lundberg SM. NeurIPS. 2017;35.