Supporting Information for:

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

Elena Di Mare[1], Ashish Punia[2,†], Matthew S. Lamm[2], Timothy Rhodes[2], Adam J. Gormley[1,†]

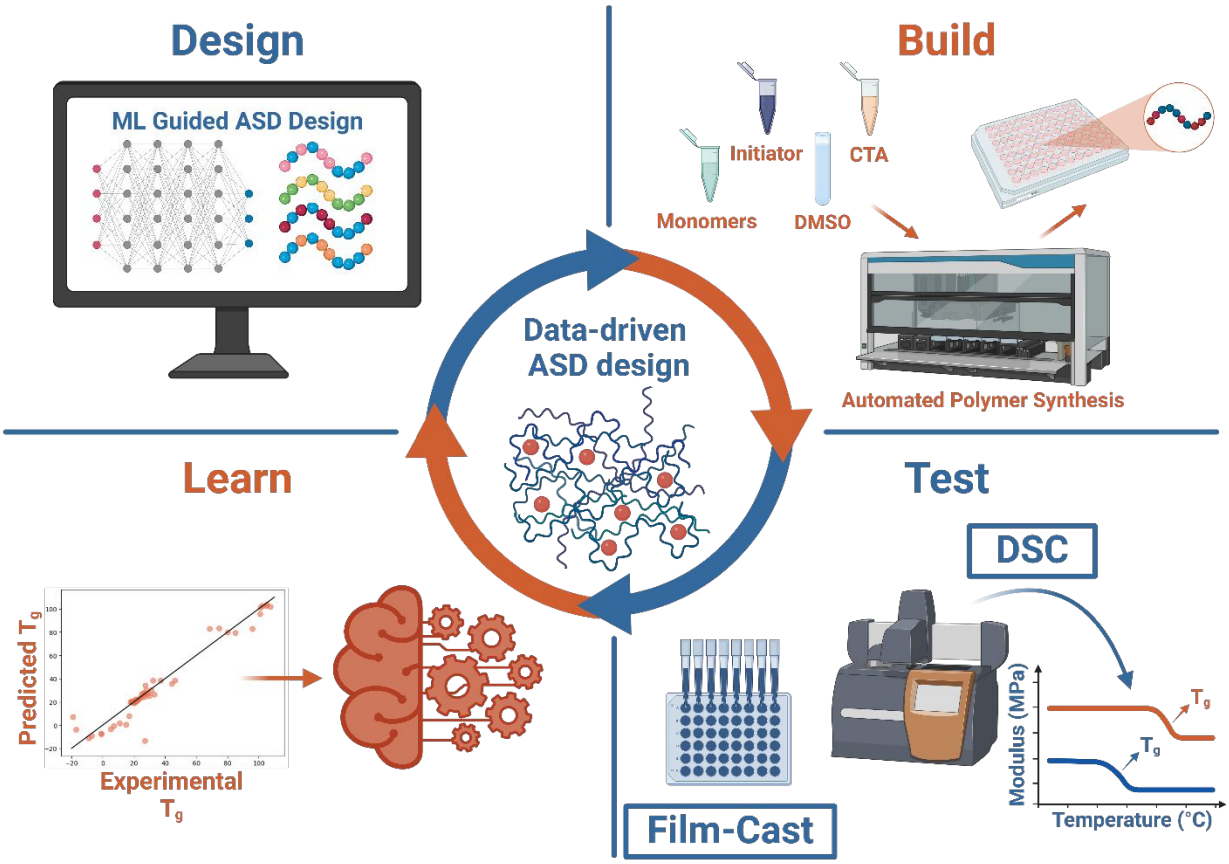
[1]Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

[2]Analytical Research and Development, MRL, Merck & Co., Inc., Rahway, NJ 07065, USA

[†]Correspondence to:

Ashish Punia: ashish.punia@merck.com

Adam J. Gormley: adam.gormley@rutgers.edu



S1. Polymer Structure

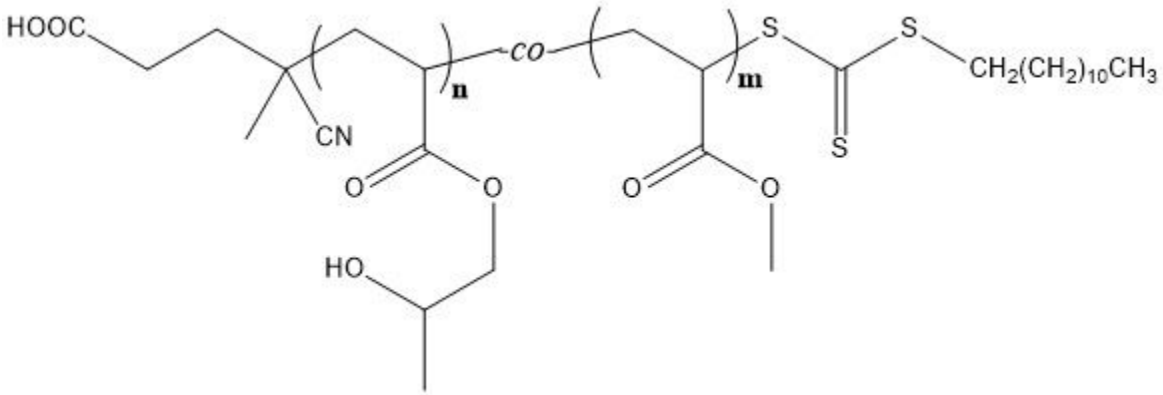


Figure S1.1. Example statistical copolymer synthesized via PET-RAFT polymerization, accounting for the presence of CTA end groups. The displayed chemical structure represents the statistical copolymer formed from HPA and MA monomers via PET-RAFT using the CTA 4-cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl] pentanoic acid.

S2 Gel Permeation Chromatography (GPC) Data

Table S2.1: MW and Dispersity (Đ) of Synthesized Polymers

Polymer ID	$MW_{Theoretical}$ (Da)	MW_{GPC} (Da)	$Đ_{GPC}$
HPA	26028	22703	1.10
HPA-10BA	25989	22077	1.10
HPA-20BA	25949	22291	1.10
HPA-30BA	25910	20998	1.10
HPA-40BA	25870	19770	1.09
HPA-50BA	25831	17040	1.09
HPA-10TBA	25989	21278	1.09
HPA-20TBA	25949	18858	1.08
HPA-30TBA	25910	21523	1.10
HPA-40TBA	25870	20536	1.10
HPA-50TBA	25831	17109	1.08
HPA-10BMA	26269	24174	1.12
HPA-20BMA	26510	20299	1.16
HPA-30BMA	26751	20988	1.21
HPA-40BMA	26992	26397	1.29
HPA-50BMA	27234	38178	2.26
HPA-10TBMA	26269	22781	1.14
HPA-20TBMA	26510	21799	1.13
HPA-30TBMA	26751	24970	1.57

HPA-40TBMA	26992		
HPA-50TBMA	27234		
HPMA	28834	24629	1.45
HPMA-10BMA	28795	22683	1.48
HPMA-20BMA	28755	22764	1.54
HPMA-30BMA	28716	22502	1.60
HPMA-40BMA	28676	18532	1.62
HPMA-50BMA	28637	17689	1.70
HPMA-10TBMA	28795	22240	1.61
HPMA-20TBMA	28755	21295	1.48
HPMA-30TBMA	28716	21964	1.59
HPMA-40TBMA	28676	22487	1.92
HPMA-50TBMA	28637	22170	1.91
HPA-10MA	25147	18048	1.18
HPA-20MA	24266	16451	1.14
HPA-30MA	23385	16110	1.14
HPA-40MA	22504	14356	1.16
HPA-50MA	21623	12969	1.14
HPA-60MA	20742	12302	1.08
HPA-60BA	25792	14633	1.10
HPA-10HA	26549	18993	1.17
HPA-20HA	27070	18311	1.18
HPA-30HA	27592	17607	1.16
HPA-40HA	28113	16878	1.14
HPA-50HA	28634	13740	1.16
HPA-10CHA	26790	17579	1.26
HPA-20CHA	27552	18999	1.18
HPA-30CHA	28313	17374	1.17
HPA-40CHA	29075	16061	1.16
HPA-50CHA	29837	15405	1.16
HPA-60CHA	30599	12541	1.19

S3. Data Preparation and Machine Learning

The Python script used throughout this work was developed in a Google CoLab environment, available here.

Table S3.1: Cleaned Database for ML Training (Columns A-M)

Polymer ID	% Hydrophobic Monomer	HPA	HPMA	BA	BMA	TBA	TBMA	MA	HA	CHA	T_g (°C)	T_g (with 20% DL) (°C)
HPA	0	1	0	0	0	0	0	0	0	0	21	22
HPA-10BA	10	0.9	0	0.1	0	0	0	0	0	0	17	16
HPA-20BA	20	0.8	0	0.2	0	0	0	0	0	0	11	13
HPA-30BA	30	0.7	0	0.3	0	0	0	0	0	0	5	11
HPA-40BA	40	0.6	0	0.4	0	0	0	0	0	0	-1	7
HPA-50BA	50	0.5	0	0.5	0	0	0	0	0	0	-7	2
HPA-10TBA	10	0.9	0	0	0	0.1	0	0	0	0	25	23
HPA-20TBA	20	0.8	0	0	0	0.2	0	0	0	0	26	26
HPA-30TBA	30	0.7	0	0	0	0.3	0	0	0	0	28	30
HPA-40TBA	40	0.6	0	0	0	0.4	0	0	0	0	31	29
HPA-50TBA	50	0.5	0	0	0	0.5	0	0	0	0	33	35
HPA-10BMA	10	0.9	0	0	0.1	0	0	0	0	0	21	21
HPA-20BMA	20	0.8	0	0	0.2	0	0	0	0	0	20	22
HPA-30BMA	30	0.7	0	0	0.3	0	0	0	0	0	18	22
HPA-40BMA	40	0.6	0	0	0.4	0	0	0	0	0	18	21
HPA-50BMA	50	0.5	0	0	0.5	0	0	0	0	0	18	21
HPA-10TBMA	10	0.9	0	0	0	0	0.1	0	0	0	27	27
HPA-20TBMA	20	0.8	0	0	0	0	0.2	0	0	0	32	34
HPA-30TBMA	30	0.7	0	0	0	0	0.3	0	0	0	37	44
HPA-40TBMA	40	0.6	0	0	0	0	0.4	0	0	0	46	53
HPA-50TBMA	50	0.5	0	0	0	0	0.5	0	0	0	44	58
HPMA	0	0	1	0	0	0	0	0	0	0	101	100
HPMA-10BMA	10	0	0.9	0	0.1	0	0	0	0	0	96	90
HPMA-20BMA	20	0	0.8	0	0.2	0	0	0	0	0	85	80
HPMA-30BMA	30	0	0.7	0	0.3	0	0	0	0	0	80	68

HPMA-40BMA	40	0	0.6	0	0.4	0	0	0	0	0	75	56
HPMA-50BMA	50	0	0.5	0	0.5	0	0	0	0	0	68	50
HPMA-10TBMA	10	0	0.9	0	0	0	0.1	0	0	0	101	100
HPMA-20TBMA	20	0	0.8	0	0	0	0.2	0	0	0	102	97
HPMA-30TBMA	30	0	0.7	0	0	0	0.3	0	0	0	105	88
HPMA-40TBMA	40	0	0.6	0	0	0	0.4	0	0	0	105	84
HPMA-50TBMA	50	0	0.5	0	0	0	0.5	0	0	0	107	87
HPA-10MA	10	0.9	0	0	0	0	0	0.1	0	0	23	19
HPA-20MA	20	0.8	0	0	0	0	0	0.2	0	0	22	19
HPA-30MA	30	0.7	0	0	0	0	0	0.3	0	0	22	18
HPA-40MA	40	0.6	0	0	0	0	0	0.4	0	0	21	18
HPA-50MA	50	0.5	0	0	0	0	0	0.5	0	0	20	17
HPA-60MA	60	0.4	0	0	0	0	0	0.6	0	0	19	18
HPA-60BA	60	0.4	0	0.6	0	0	0	0	0	0	-17	8
HPA-10HA	10	0.9	0	0	0	0	0	0	0.1	0	15	14
HPA-20HA	20	0.8	0	0	0	0	0	0	0.2	0	7	2
HPA-30HA	30	0.7	0	0	0	0	0	0	0.3	0	-1	8
HPA-40HA	40	0.6	0	0	0	0	0	0	0.4	0	-9	6
HPA-50HA	50	0.5	0	0	0	0	0	0	0.5	0	-19	9
HPA-10CHA	10	0.9	0	0	0	0	0	0	0	0.1	24	22
HPA-20CHA	20	0.8	0	0	0	0	0	0	0	0.2	25	27
HPA-30CHA	30	0.7	0	0	0	0	0	0	0	0.3	27	29
HPA-40CHA	40	0.6	0	0	0	0	0	0	0	0.4	27	28
HPA-50CHA	50	0.5	0	0	0	0	0	0	0	0.5	29	29
HPA-60CHA	60	0.4	0	0	0	0	0	0	0	0.6	30	30

Table S3.2: Cleaned Database for ML Training (Columns N-T)

Theoretical MW	Hydrophilic Methylation	Hydrophobic Methylation	Hydrophilic Feed Fraction	Hydrophobic Monomer Linearity	Carbons per Side Chain	Monomer Pairing
26028	0	0	1	1	0	HPA
25989	0	0	0.9	1	4	HPA-BA
25949	0	0	0.8	1	4	HPA-BA
25910	0	0	0.7	1	4	HPA-BA
25870	0	0	0.6	1	4	HPA-BA
25831	0	0	0.5	1	4	HPA-BA
25989	0	0	0.9	0	4	HPA-TBA
25949	0	0	0.8	0	4	HPA-TBA
25910	0	0	0.7	0	4	HPA-TBA
25870	0	0	0.6	0	4	HPA-TBA
25831	0	0	0.5	0	4	HPA-TBA
26269	0	1	0.9	1	4	HPA-BMA
26510	0	1	0.8	1	4	HPA-BMA
26751	0	1	0.7	1	4	HPA-BMA
26992	0	1	0.6	1	4	HPA-BMA
27234	0	1	0.5	1	4	HPA-BMA
26269	0	1	0.9	0	4	HPA-TBMA
26510	0	1	0.8	0	4	HPA-TBMA
26751	0	1	0.7	0	4	HPA-TBMA
26992	0	1	0.6	0	4	HPA-TBMA
27234	0	1	0.5	0	4	HPA-TBMA
28834	1	0	1	1	0	HPMA
28795	1	1	0.9	1	4	HPMA-BMA
28755	1	1	0.8	1	4	HPMA-BMA
28716	1	1	0.7	1	4	HPMA-BMA
28676	1	1	0.6	1	4	HPMA-BMA
28637	1	1	0.5	1	4	HPMA-BMA
28795	1	1	0.9	0	4	HPMA-TBMA
28755	1	1	0.8	0	4	HPMA-TBMA
28716	1	1	0.7	0	4	HPMA-TBMA

28676	1	1	0.6	0	4	HPMA-TBMA
28637	1	1	0.5	0	4	HPMA-TBMA
25147	0	0	0.9	1	1	HPA-MA
24266	0	0	0.8	1	1	HPA-MA
23385	0	0	0.7	1	1	HPA-MA
22504	0	0	0.6	1	1	HPA-MA
21623	0	0	0.5	1	1	HPA-MA
20742	0	0	0.4	1	1	HPA-MA
25792	0	0	0.4	1	4	HPA-BA
26549	0	0	0.9	1	6	HPA-HA
27070	0	0	0.8	1	6	HPA-HA
27592	0	0	0.7	1	6	HPA-HA
28113	0	0	0.6	1	6	HPA-HA
28634	0	0	0.5	1	6	HPA-HA
26790	0	0	0.9	0	6	HPA-CHA
27552	0	0	0.8	0	6	HPA-CHA
28313	0	0	0.7	0	6	HPA-CHA
29075	0	0	0.6	0	6	HPA-CHA
29837	0	0	0.5	0	6	HPA-CHA
30599	0	0	0.4	0	6	HPA-CHA

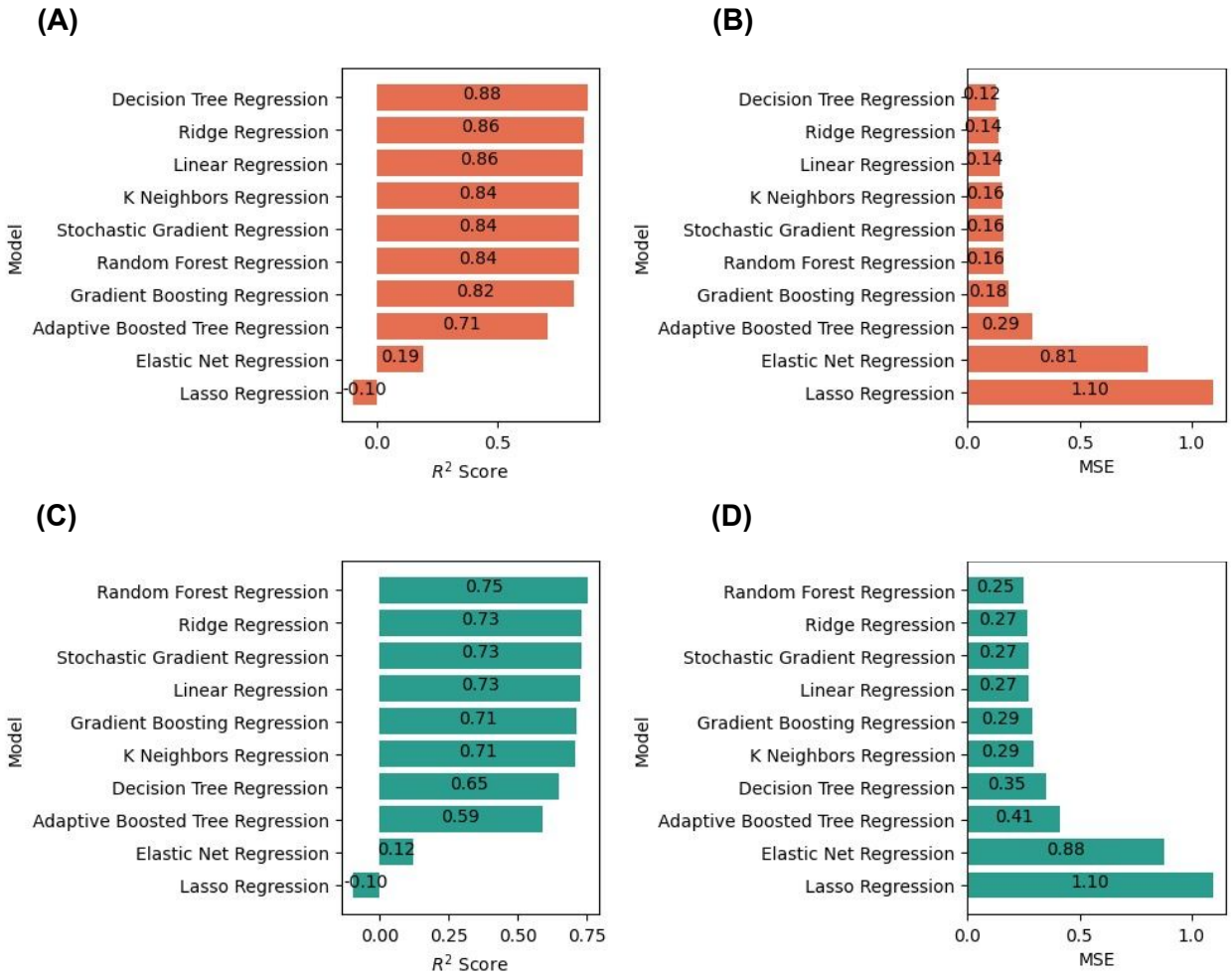


Figure S3.3. Determination of suitable ML model. We tested the accuracy of 10 commonly used ML models to assess which model would be most appropriate for T_g prediction in polymers alone (**a**, **b**) and in 20 wt% probucol loaded ASDs (**c**, **d**). Models were scored by R^2 (**a**, **c**) and MSE (**b**, **d**). Random Forest performed well for both polymers alone and ASDs, and as such Random Forest was chosen as the model used throughout the remainder of the work.

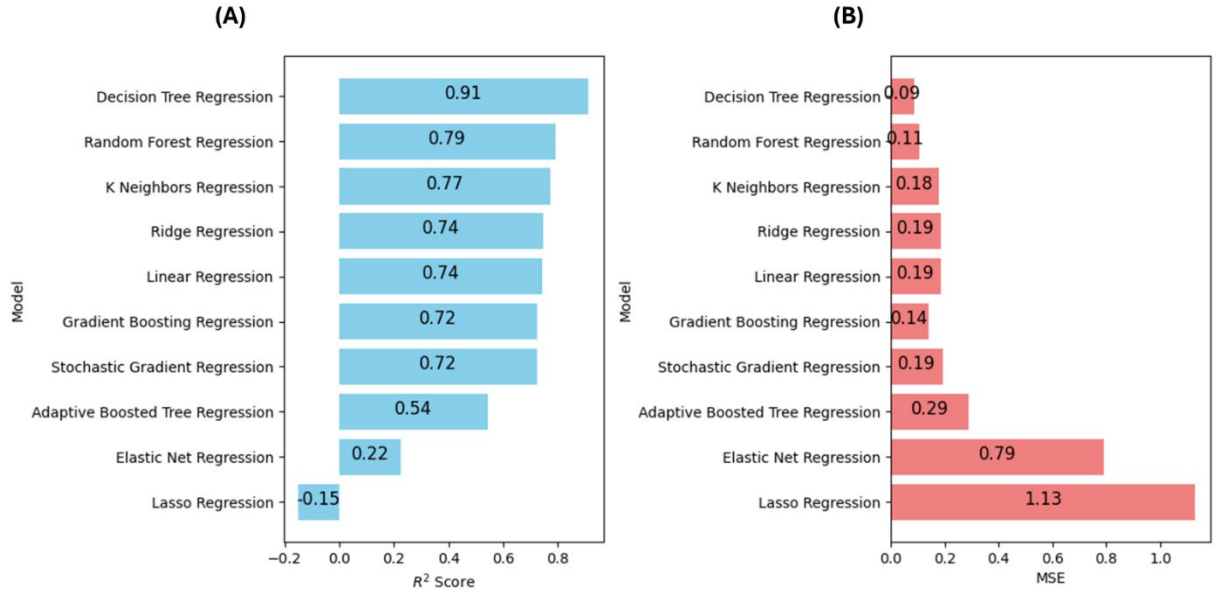


Figure S3.4 Model selection for predicting T_g with drug loading as a training feature. Of 10 potential models, Random Forest performed second best and was chosen as the model for subsequent analyses because it is less prone to overfitting than a Decision Tree Regressor, which was the top performing model. (a) R^2 score of each model. (b) Mean Squared Error (MSE) of each model.

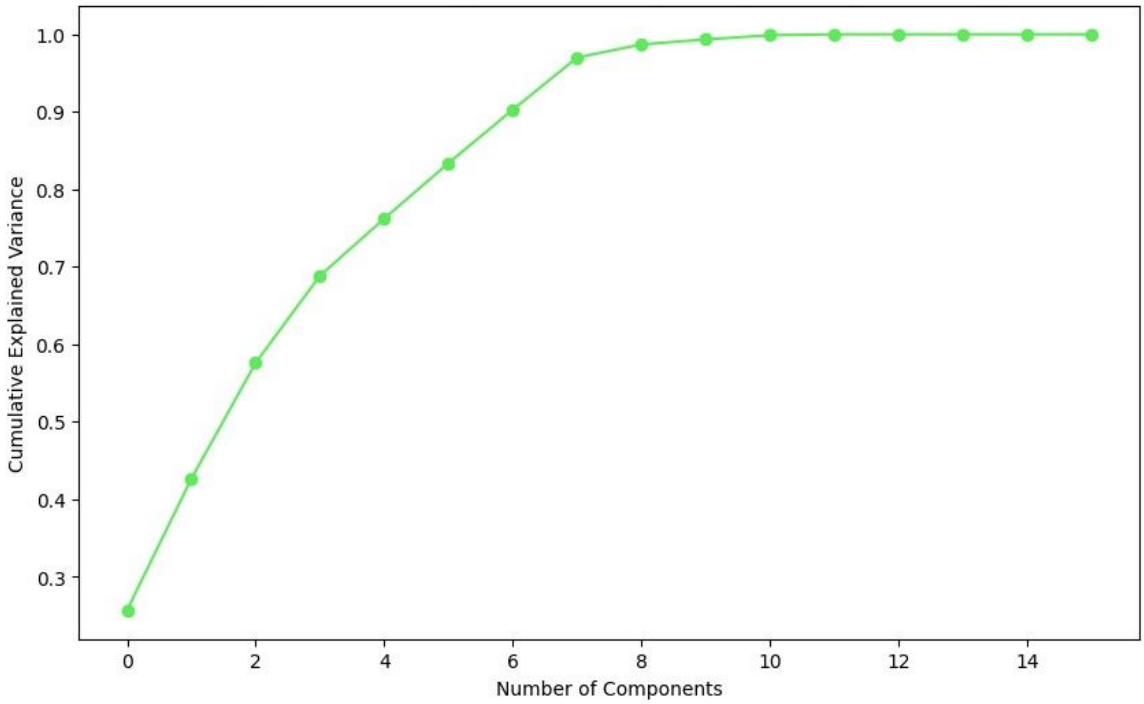


Figure S3.5 Principal Component Analysis (PCA) of design features for dimensionality reduction. PCA indicated that he cumulative explained variance of dataset could be sufficiently explained with less than seven features. Five features (Hydrophilic methylation, hydrophilic feed

fraction, linear/nonlinear side chain geometry, hydrophobic monomer methylation, and number of carbons per hydrophobic monomer side chain) were used to train the model to predict T_g of polymer alone and T_g of 20 wt% probucol loaded ASDs. A sixth feature, drug loading, was added to predict the T_g of the complete dataset.

S4. Differential Scanning Calorimetry

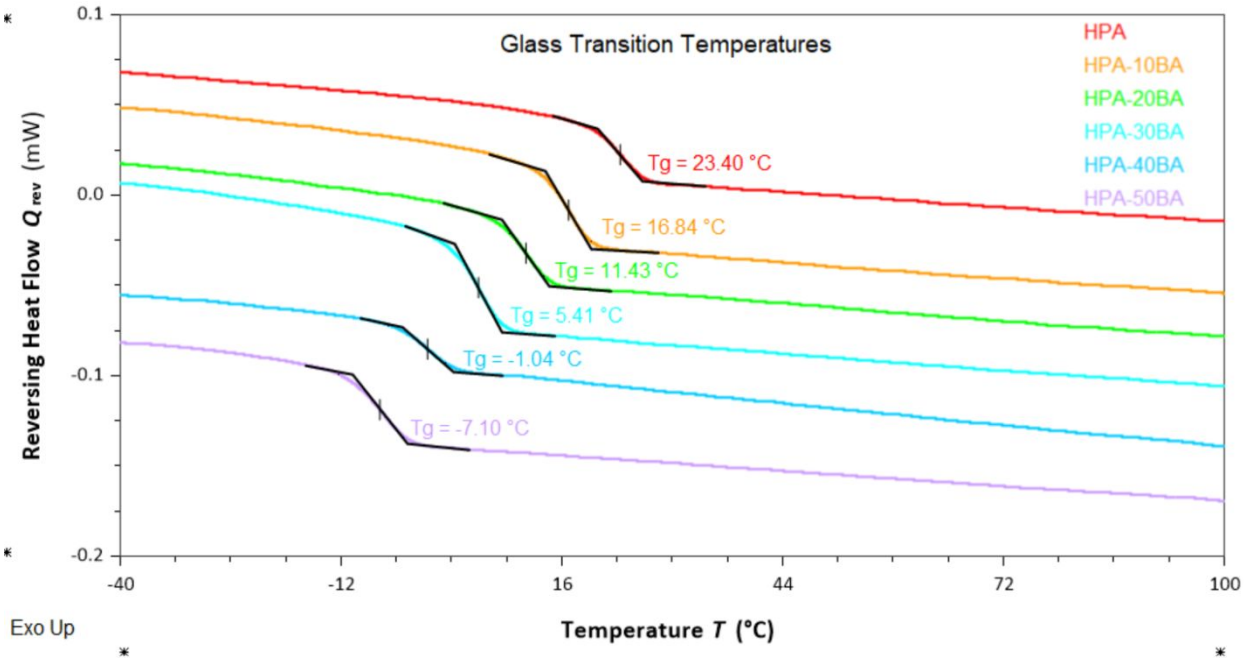


Figure S4.1. DSC thermograms for selected polymers. T_g was calculated as the midpoint of the phase transition present in the second heat cycle.

Table S4.2: Comparison of Experimental T_g and T_g Predicted by Fox Equation

Polymer ID	Experimental T_g of Polymer Alone (°C)	Experimental T_g for ASDs (°C)	Fox-Predicted T_g (°C)	Percent Error for Polymer Alone (%)	Percent error for ASDs (%)
HPA	21	22	21	0	0
HPA-10BA	17	16	11	2	2
HPA-20BA	11	13	2	3	4
HPA-30BA	5	11	-7	4	6
HPA-40BA	-1	7	-15	5	8
HPA-50BA	-7	2	-22	6	9
HPA-10TBA	25	23	23	1	0
HPA-20TBA	26	26	25	0	0
HPA-30TBA	28	30	27	0	1
HPA-40TBA	31	29	29	1	0
HPA-50TBA	33	35	32	0	1
HPA-10BMA	21	21	21	0	0
HPA-20BMA	20	22	21	0	0
HPA-30BMA	18	22	21	1	0
HPA-40BMA	18	21	21	1	0
HPA-50BMA	18	21	20	1	0
HPA-10TBMA	27	27	29	1	1
HPA-20TBMA	32	34	38	2	1
HPA-30TBMA	37	44	46	3	1
HPA-40TBMA	46	53	55	3	1
HPA-50TBMA	44	58	65	6	2
HPMA	101	100			
HPMA-10BMA	96	90	90	2	0
HPMA-20BMA	85	80	81	1	0
HPMA-30BMA	80	68	72	2	1
HPMA-40BMA	75	56	64	3	2
HPMA-50BMA	68	50	55	4	2

HPMA-10TBMA	101	100	102	0	0
HPMA-20TBMA	102	97	103	0	2
HPMA-30TBMA	105	88	105	0	5
HPMA-40TBMA	105	84	107	0	6
HPMA-50TBMA	107	87	109	0	6
HPA-10MA	23	19	20	1	0
HPA-20MA	22	19	19	1	0
HPA-30MA	22	18	18	1	0
HPA-40MA	21	18	18	1	0
HPA-50MA	20	17	17	1	0
HPA-60MA	19	18	15	1	1
HPA-60BA	-17	8	-30	5	13
HPA-10HA	15	14	11	1	1
HPA-20HA	7	2	3	2	0
HPA-30HA	-1	8	-5	2	5
HPA-40HA	-9	6	-13	1	7
HPA-50HA	-19	9	-19	0	10
HPA-10CHA	24	22	21	1	0
HPA-20CHA	25	27	21	1	2
HPA-30CHA	27	29	20	2	3
HPA-40CHA	27	28	20	2	3
HPA-50CHA	29	29	20	3	3
HPA-60CHA	30	30	20	3	3