A data analytics approach for rational design of nanomedicines with programmable drug release

*Adam S. Mullis[1], Scott R. Broderick[2], Andrea M. Binnebose[3,4], Nathan Peroutka-Bigus[3,4], Bryan H. Bellaire[3,4,5], Krishna Rajan[2,5], and Balaji Narasimhan[1,5]**

[1]Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011, United States

[2]Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY 14260, United States

[3]Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, Iowa 50011, United States

[4]Interdepartmental Microbiology Graduate Program, Iowa State University, Ames, Iowa 50011, United States

[5]Nanovaccine Institute, Iowa State University, Ames, Iowa 50011, United States

KEYWORDS: polyanhydrides, degradable biomaterials, drug delivery, drug release kinetics, informatics, data mining

ABSTRACT

Drug delivery vehicles can improve the functional efficacy of existing antimicrobial therapies by improving biodistribution and targeting. A critical property of such nanomedicine formulations is their ability to control the release kinetics of their payloads. The combination of (and interactions between) polymer, drug, and nanoparticle properties gives rise to nonlinear behavioral relationships and a large data space. These factors complicate both first-principles modeling and screening of nanomedicine formulations. Predictive analytics may offer a more efficient approach toward rational design of nanomedicines by identifying key descriptors and correlating them to nanoparticle release behavior. In this work, antibiotic release kinetics data were generated from polyanhydride nanoparticle formulations with varying copolymer compositions, encapsulated drug type, and drug loading. Four antibiotics, doxycycline, rifampicin, chloramphenicol, and pyrazinamide, were used. Linear manifold learning methods were used to relate drug release properties with polymer, drug, and nanoparticle properties, and key descriptors were identified that are highly correlated with release properties. However, these linear methods could not predict release behavior. Non-linear multivariate modeling based on graph theory was then used to deconvolute the governing relationships between these properties, and predictive models were generated to rapidly screen lead nanomedicine formulations with desirable release properties with minimal nanoparticle characterization. Release kinetics predictions of two drugs containing atoms not included in the model showed good agreement with experimental results, validating the model and indicating its potential to virtually explore new polymer and drug pairs not included in training data set. The models were shown to be robust after inclusion of these new formulations in that the new inclusions did not significantly change model regression. This approach provides the first steps towards development of a

framework that can be used to rationally design nanomedicine formulations by selecting the appropriate carrier for a drug payload to program desirable release kinetics.

1. INTRODUCTION

Intracellular bacterial infections are challenging to treat using traditional antimicrobial therapies due to the difficulty in achieving high enough local drug concentration for antimicrobial activity without inducing host cell toxicity.[1] Elimination of soluble drugs through host metabolism and excretion pathways act to reduce bioavailable amounts of antimicrobials requiring repeated dosing to maintain therapeutic concentrations to mitigate the development of antibiotic resistant in pathogens.[2,3] Drug delivery vehicles can improve the efficacy and potency of antimicrobials by altering the drug biodistribution with improved intracellular localization and delivery of cargo to the pathogen's intracellular niche within host cells.[4,5] Biodegradable polyanhydride nanoparticles show passive targeting and payload stabilization properties that make them uniquely suited for antibiotic delivery for intracellular infections.[4,6] These particles are internalized efficiently by phagocytic cells using multiple mechanisms, and have been used to deliver antibiotics to kill intracellular *Brucella abortus.*[7] Additionally, polyanhydride nanoparticles mediated efficient killing of filarial parasites by co-delivering an antiparasitic with an antibiotic targeting an intracellular endosymbiotic bacterium that supports parasite health and reproduction.[8]

A key feature of the effectiveness of these nanomedicine formulations is their ability to control payload release rate, however rationally designing nanomedicines with programmable release remains elusive. Release kinetics are influenced by drug distribution within a device and/or a particle, which is in turn influenced by polymer-drug thermodynamic interactions.[9–11] These interactions give rise to nonlinear release behavior which is difficult to predict *a priori.*

Screening nanomedicine formulations is challenging as polymer and nanoparticle properties (e.g., polymer chemistry, nanoparticle size, polydispersity, release kinetics, and encapsulation efficiency) yield a large number of additional variables beyond drug-specific properties. This large dataspace, coupled with the multiple length scales at play, poses difficulties for generalizing conclusions to other nanoparticle systems and impedes first principles modeling of nanoparticle behavior.[12,13] Hierarchical modeling may be a more efficient approach for such systems, wherein key descriptors are identified and correlated to performance parameters.

Informatics methods encompass several tools for such hierarchical modeling. Data mining techniques can deconvolute complex behavior, unraveling relationships that lie on non-Euclidian surfaces,[14] which enables pattern recognition and prediction through the development of quantitative structure-property relationships (QSPRs).[15] To this end, previous informatics analyses from our laboratories has enabled identification of polyanhydride chemistry and structural factors that influence protein release from films[16] and enable pathogen-mimicking nanoparticle processing by immune cells.[17–20]

The focus of this work was to develop an informatics-based framework that determines how polymer, drug, and nanoparticle characteristics influence drug encapsulation efficiency and release kinetics. We sought to generate predictive models that can virtually test potential new polymer and drug combinations for desirable release kinetics. Our long-term goal is to develop a predictive analytics framework to enable rational design of nanomedicine formulations for different types of therapeutic and prophylactic applications.

2. MATERIALS AND METHODS

2.1. Materials. Sebacic acid (SA) was purchased from Sigma Aldrich (St. Louis, MO). Triethylene glycol, 4-*p*-hydroxybenzoic acid, 1- methyl-2-pyrrolidinone, and 1,6-dibromohexane

were purchased from Sigma Aldrich for 1,8-bis(p-carboxyphenoxy)-3,6-dioxaoctane (CPTEG) and 1,6-bis(p-carboxyphenoxy)hexane (CPH) monomer synthesis. Potassium carbonate, dimethyl formamide, toluene, acetonitrile, acetic acid, sulfuric acid, N,N-dimethylacetamide, and acetic anhydride were purchased from Fisher Scientific (Fairlawn, NJ) for monomer and polymer synthesis. 4-*p*-fluorobenzonitrile was purchased from Apollo Scientific (Cheshire, UK) for use in monomer synthesis. Methylene chloride, pentane, and hexanes were purchased from Fisher Scientific for polymer purification and nanoparticle synthesis. Doxycycline (DOX), rifampicin (RIF), and pyrazinamide (PZA) were purchased from Sigma Aldrich, and chloramphenicol (CAM) was purchased from Fisher Scientific. Meropenem (MEM) was purchased from Ark Pharm, Inc. (Arlington Heights, IL) and ceftazidime (CAZ) was purchased from Acros Organics (NJ). ^{1}H NMR analysis used deuterated chloroform purchased from Cambridge Isotope Laboratories (Andover, MA). Drug quantification used UV-transparent microplates from Greiner Bio-One (Kremsmünster, Austria), HPLC grade acetonitrile, methanol, and tetrahydrofuran from Fisher Scientific, and phosphoric acid from Sigma Aldrich.

2.2. Polymer and nanoparticle synthesis. CPTEG and CPH diacid were synthesized as described previously.[10,21,22] CPTEG:CPH and CPH:SA copolymers were synthesized as described previously.[10,21] Briefly, monomers were weighed in appropriate molar ratios and added to a round bottom flask. The monomers were acetylated in excess acetic acid at 125°C for 30 min, and rotary evaporation was used to remove excess solvent from the resulting prepolymer. CPTEG:CPH was reacted for six hours at 140°C at <0.1 Torr and CPH:SA was reacted for 30 min at 180°C at <0.3 Torr. Polymers were purified by precipitation in chilled hexanes. Copolymer composition and number average molecular weight (M_n) was confirmed from ^{1}H

NMR spectra acquired on a Varian MR-400 (Varian, Inc. Palo Alto, CA) and thermal properties of the copolymers was characterized by DSC (Q2000, TA Instruments, New Castle, DE).

Antibiotic-loaded nanoparticles were synthesized as described previously.[7,8] Polymer and drug were weighed in separate scintillation vials at appropriate %w/w ratios. Enough methylene chloride to dissolve the polymer at 20 mg/mL was added to the drug vial to dissolve/disperse the drug, then transferred to the polymer vial. The combined drug and polymer solution was poured into a pentane antisolvent bath at room temperature (CPH:SA) or -10°C (CPTEG:CPH) at a solvent:anti-solvent ratio of 1:250 and nanoparticles were recovered by vacuum filtration. CPTEG:CPH nanoprecipitation was carried out in a cold room at 4°C. A total of 68 nanoformulations were tested, spanning drug loadings between 1% and 20% loading (% w/w). All drugs were tested in 20:80 CPH:SA and 20:80 CPTEG:CPH nanoparticles, and rifampicin was additionally tested in 10:90, 30:70, and 50:50 CPTEG:CPH. ^{1}H NMR spectra of empty 20:80 CPTEG:CPH and 20:80 CPH:SA nanoparticles indicated undetectable amounts of methylene chloride and trace amounts of pentane (data not shown).

To validate the informatics analysis, nanoparticles encapsulating meropenem or ceftazidime were synthesized using a high-throughput method adapted from Goodman et al.[23] Briefly, polymer and drugs were dissolved/dispersed in methylene chloride and dispensed via high-throughput robot into 10 mL borosilicate tubes at a final polymer concentration of 20 mg/mL. The robot sonicated and dispensed the combined polymer and drug solution into 50 mL conical polypropylene tubes containing 45 mL pentane (1:18 solvent:anti-solvent ratio) at the temperatures listed above. Multiple particle batches were pooled and recovered by vacuum filtration. Scanning electron microscopy (SEM, FEI Quanta 250, Hillsboro, OR) was used to image all nanoparticles, and size distributions were calculated using Fiji image analysis

software[24] and the ParticleSizer plugin script for Fiji. Nanoparticle zeta potential was measured using a Zetasizer Nano (Malvern Instruments Ltd., Worcester, UK).

2.3. Drug release kinetics. Nanoparticles (9-11 mg) were dispersed in 0.5 mL PBS, pH 7.4 and suspended by sonication (VCX 130 PB, Sonics & Materials, Inc., Newtown, CT). At each time point, the nanoparticles were pelleted by centrifugation and supernatant was collected for drug quantification. Fresh PBS was added to maintain perfect sink conditions and the nanoparticles were dispersed by sonication. At the end of the release experiment, 40 mM sodium hydroxide was added to accelerate polymer degradation and extract the remaining encapsulated drug as described previously.[25]

The drug mass released at each time point was determined by spectrophotometry (SpectraMax M3, Molecular Devices, San Jose, CA) and UV-HPLC (1200 series, Agilent Technologies, Santa Clara, CA). Doxycycline, rifampicin, and chloramphenicol were quantified by absorbance in UV-transparent 96-well plates at 350, 333, and 293 nm, respectively. Pyrazinamide release and base extraction samples were separated using a Phenomenex Kinetex 2.6-micron C18 100Å 100x4.6mm column and a 30:5:65 acetonitrile:methanol:water mobile phase adjusted to pH 5.2 with phosphoric acid.[26] The flow rate was 0.6 mL/min and pyrazinamide was quantified at 268 nm. Meropenem and ceftazidime release and base extraction samples were separated using a Zorbax Eclipse XDB-C8 5-micron 4.6x150 mm column, monitoring at 299 nm and 246 nm, respectively. Meropenem release samples used a mobile phase gradient ramping from 0.1/99.9 (%v/v) methanol/water to 50/50 over 15 min. Meropenem base extraction samples used a gradient ramping from 0.1/99.9 acetonitrile 0.1% trifluoroacetic acid/ water 0.1% trifluoroacetic acid to 50/50 over 15 min. Ceftazidime release samples used a mobile phase protocol with an isocratic step at 0.1/99.9 methanol/water for 5 minutes followed by a gradient ramping to 50/50

over 10 min. Ceftazidime base extraction samples used an isocratic step at 15/85 from 0.1/99.9 acetonitrile 0.1% trifluoroacetic acid/ water 0.1% trifluoroacetic acid for 1 min followed by a gradient ramping to 40/60 over 5 min. All meropenem and ceftazidime HPLC protocols used a flow rate of 1 mL/min.

The small mass of drug and large volume of antisolvent used in nanoparticle synthesis render the non-encapsulated drug concentration below the limit of detection of the analytical methods used in this study. Therefore encapsulation efficiency (EE) was calculated from the cumulative sum of detected drug mass released in PBS and base extraction samples using Eq. 1.[25] In a minority of formulations >100% EE was observed, which could arise from the presence of drug nanocrystals[27] (which was not detected on nanoparticle surfaces by SEM), gravimetric inaccuracies due to static charge of the nanoparticles, or residual error in the drug concentration quantification assays. Drug release kinetics are presented as fraction released, where the cumulative drug mass release is normalized by the total encapsulated drug mass. Prism 7 (Graphpad Software, La Jolla, CA) was used to generate release kinetics figures.

Eq. 1)

$$\frac{EE}{100\%} = \frac{cumulative\ drug\ mass\ from\ release\ \&\ base\ extraction}{nanoparticle\ mass * drug\ loading\ fraction\left(\frac{wt}{wt}\right)}$$

2.4. Informatics analysis. Release behavior parameters, along with polymer, drug, and nanoparticle properties, were normalized and mean-centered. Three different informatics approaches were integrated and applied to analyze the data in this work. Linear manifold learning approaches, such as principal component analysis (PCA),[28–30] permit us to identify the right projection of data from which meaningful features associated with the input data can be identified. PCA performs an eigenvector decomposition and defines a new set of linear

combinations of descriptors which maximize the amount of unique information in a minimal set of orthogonal axes, termed principal components (PCs). The original data is decomposed into two matrices of interest for this work: the scores and loadings. The scores describe the different conditions (i.e., nanoparticle and drug chemistry), while the loadings describe the different descriptors and properties. The interpretation of these matrices is provided here with the relevant results, and an additional term called the variable importance projection (VIP) is calculated from the loadings matrix using Eq. 2.

Eq. 2)

$$VIP = \frac{PC_x^T * PC_x^i}{\sum PC_x^T * PC_x^i}$$

In this case, x = 5 because 5 PCs captured >90% of the variance in the data. The analysis was performed for T = encapsulation efficiency, drug released @2 h, and fraction released/day (Figure 6). Partial Least Squares (PLS) is a multi-linear regression approach which accounts for co-linearity in the data, and therefore limits bias and develops more robust quantitative relationships.[31–34] PLS performs separate PC analyses on the predictor variables (i.e., descriptors) and the predicted variables (i.e., properties). These therefore represent linear manifold learning approaches which provide qualitative and quantitative design relationships.

In order to model the drug release properties accurately and robustly, we found that non-linearity needed to be accounted for in the modeling. Therefore, we first developed non-linear parameterization of the data through non-linear manifold learning, based on graph theory, using the Isomap algorithm.[14,35] This approach generates a graph connecting data points on a high dimensional space to their nearest neighbors, mapped out in the high dimensional space, and then

fit to a low dimensional manifold. The assumption here is that the graph Euclidean distance between the points in high dimensions closely approximates the curvilinear distances along the low dimensional manifold. Through dimensionality reduction the manifold unravels into two or three dimensions allowing it to be visualized. The result of such dimensionality reduction is a weighted graph of the original data points where the edges are weighted according to the geodesic distances. Like in PCA, we develop a set of parameters for each set of conditions, although in this case the parameters are based on a non-linear combination of descriptors.

3. RESULTS

3.1. Building descriptor library. To generate the data set, we focused on nanoparticles composed of CPTEG, CPH, and SA copolymers (Figure 1 a,b). Nanoparticles synthesized from these polyanhydride copolymers have been shown to kill intracellular bacteria because of their high internalization rates by phagocytic cells,[7,36] localization in intracellular compartments that harbor these bacteria,[17,36] and improved antimicrobial activity of encapsulated drugs.[6–8] In addition to the structural descriptors defined by Li et al.,[16] we included molecular weight and compositional data from ^{1}H NMR and thermal characterization from DSC analysis. The release kinetics of four antibiotics, doxycycline (Figure 1c), rifampicin (Figure 1d), chloramphenicol (Figure 1e), and pyrazinamide (Figure 1f), were studied. The choice of the drug library was motivated by multiple factors. These drugs were selected due to their diversity of molecular weight, chemical structure, and hydrophobicity, among other physicochemical differences. All are FDA-approved drugs and belong to separate antibiotic classes, target distinct bacterial structures, and have well-characterized pharmacokinetics/pharmacodynamics. Experimental and predicted physicochemical properties for each of these drugs were gathered from the Drugbank

database.[37] Predicted drug properties from this database were calculated by ALGOPS and ChemAxon methodologies.

These drugs were encapsulated in polyanhydride nanoparticles by flash nanoprecipitation, and zeta potential, size distributions, and polydispersity index were obtained. Release profiles and encapsulation efficiencies were obtained from *in vitro* experiments in PBS, pH 7.4 (Figure 2). Figure 2 shows representative release kinetics data for multiple drugs, selected from a total of 68 nanoformulations that were tested. The formulations depicted in Figure 2 were selected to show the diversity of release behavior in the data set. CPH:SA-doxycycline nanoformulations tended to show a higher burst than CPTEG:CPH nanoformulations, and the lower loading in the CPH:SA nanoformulations tended to have a greater sustained release slope (Figure 2a). The chemistry trend was reversed in the rifampicin nanoformulations, where the CPTEG:CPH chemistries tended to show a higher burst release than the CPH:SA chemistries, and increasing the loading increased the burst (Figure 2b). For chloramphenicol, both 20:80 CPH:SA and CPTEG:CPH nanoformulations tended to generate a large burst release followed by a slow rate of drug release (Figure 2c). Pyrazinamide formulations generated a large burst from the 20:80 CPH:SA nanoparticles followed by a steady rate of drug release (Figure 2d). In contrast, the 20:80 CPTEG:CPH nanoparticles encapsulating pyrazinamide showed a small burst and slow rate of drug release and did not release more than 20% of the payload in one week. These results add to the body of literature[9–11,38] that indicates that copolymer chemistry, drug type, and drug loading influence drug release kinetics from biodegradable particles and other devices.

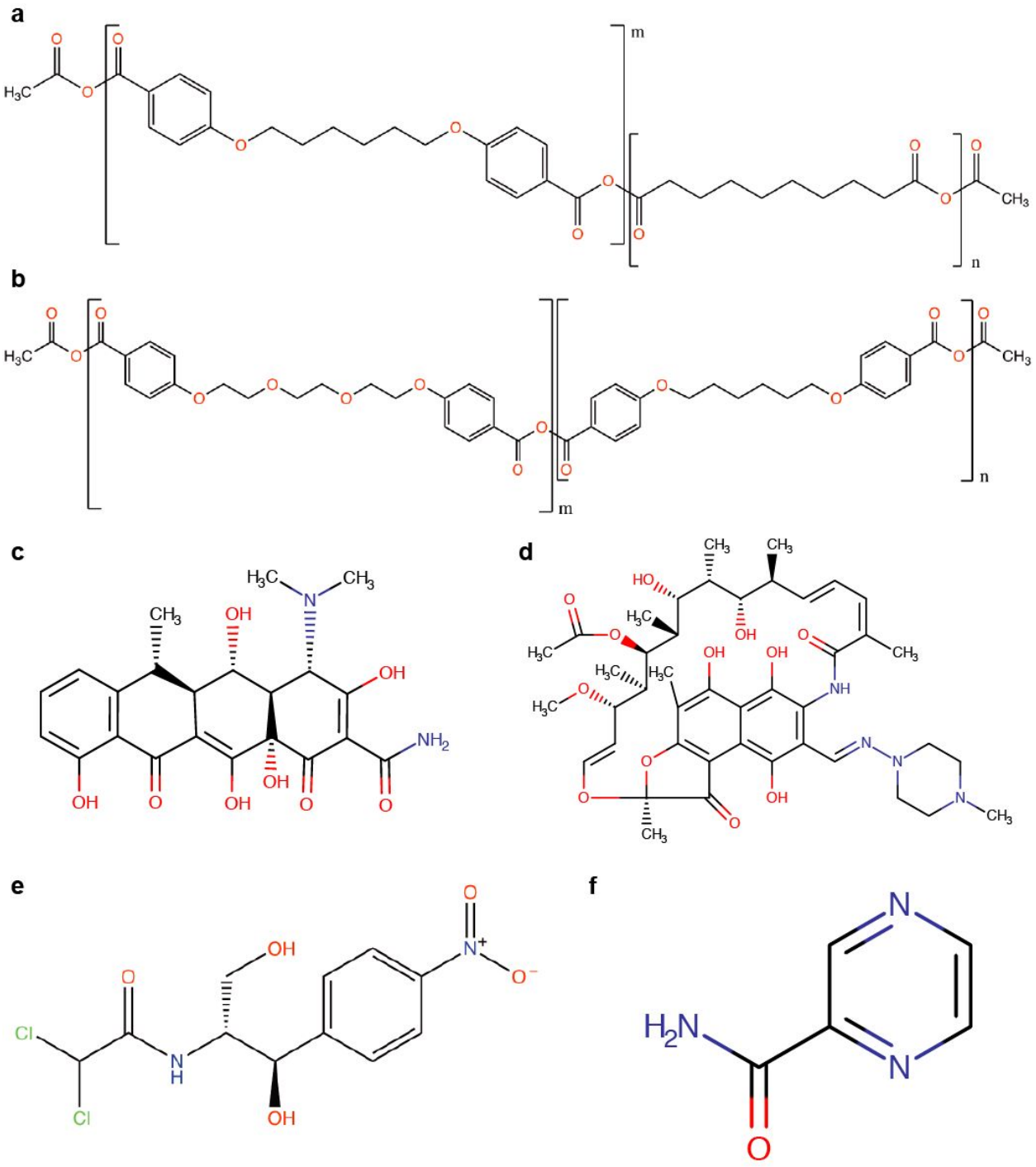


Figure 1. Polymer and antibiotic chemical structures. (a,b) Structures of CPH:SA (a) and CPTEG:CPH (b) copolymers, where m and n are the number of repeats for each unit. (c-f)

Structure of doxycycline (DOX, a), rifampicin (RIF, b), chloramphenicol (CAM, c), and pyrazinamide (PZA, d).

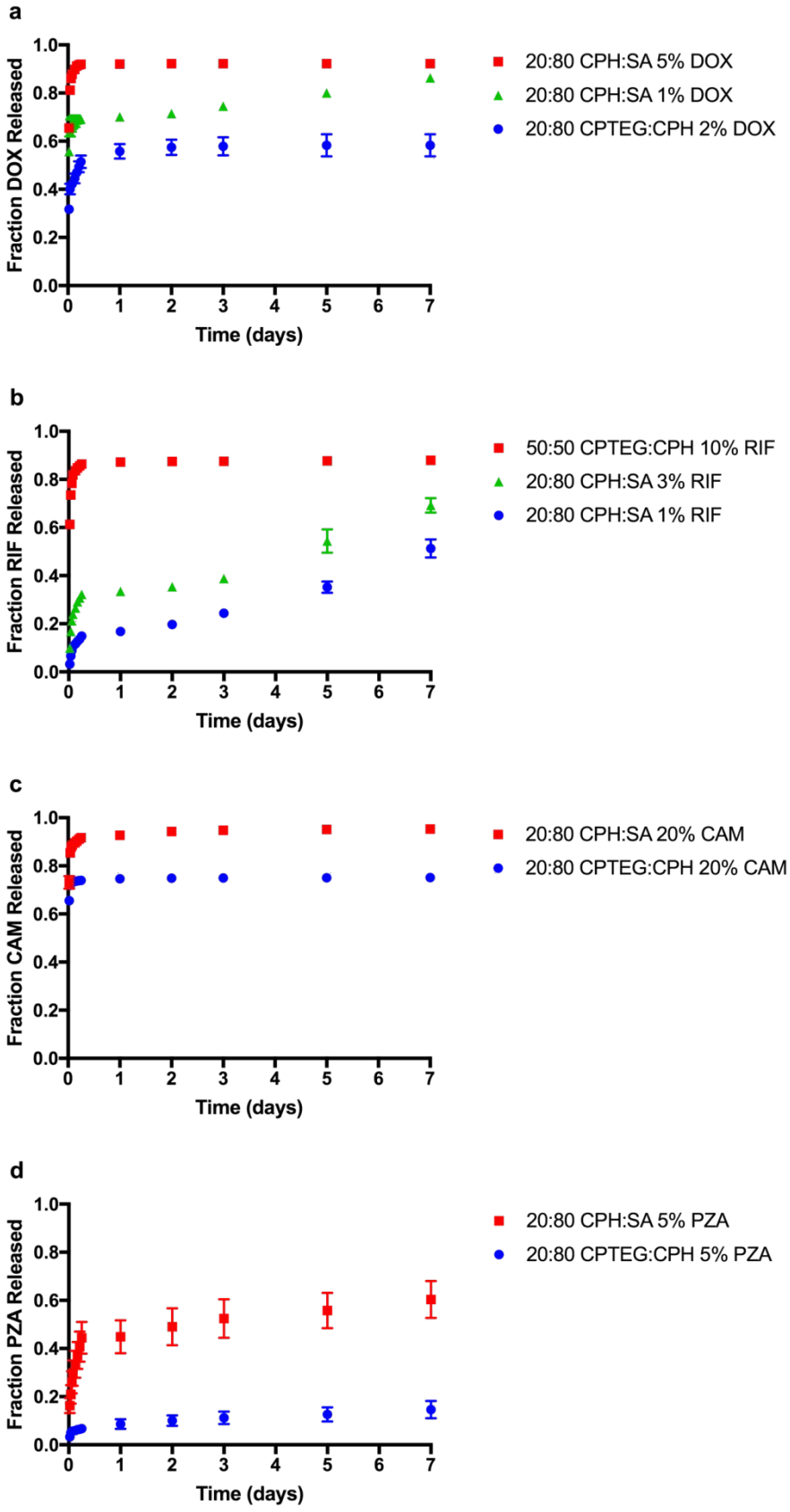

a
Fraction DOX Released
Time (days)
20:80 CPH:SA 5% DOX
20:80 CPH:SA 1% DOX
20:80 CPTEG:CPH 2% DOX
b
Fraction RIF Released
Time (days)
50:50 CPTEG:CPH 10% RIF
20:80 CPH:SA 3% RIF
20:80 CPH:SA 1% RIF
c
Fraction CAM Released
Time (days)
20:80 CPH:SA 20% CAM
20:80 CPTEG:CPH 20% CAM
d
Fraction PZA Released
Time (days)
20:80 CPH:SA 5% PZA
20:80 CPTEG:CPH 5% PZA

Figure 2. Representative antibiotic release kinetics from nanoparticles encapsulating doxycycline (DOX, a), rifampicin (RIF, b), chloramphenicol (CAM, c), and pyrazinamide (PZA, d). The depicted nanoformulations represent a subset of the 68 formulations tested and were selected to display the diversity of release behavior in the data set. Data are presented as mean ± SD. Error bars are not depicted in cases where the error bar height is smaller than the symbol. Release profiles were parameterized into two-hour burst, one day burst, and 2-7 day sustained release slope.

Table 1. Representative antibiotic release properties. FR (2h) fraction released in two-hour burst, FR (24h) fraction released in 24-hour burst. Data are presented as mean ± SD.

Nano-formulation	EE	FR (2h)	FR (24h)	FR slope (FR/day)
20:80 CPH:SA 5% DOX	64.9% ± 5.8%	0.878 ± 0.005	0.921 ± 0.007	0.00003 ± 0.00016
20:80 CPH:SA 1% DOX	159.0% ± 9.3%	0.675 ± 0.025	0.702 ± 0.010	0.02963 ± 0.97909
20:80 CPTEG:CPH 1% DOX	71.6% ± 5.8%	0.329 ± 0.023	0.462 ± 0.159	0.00082 ± 0.00014
50:50 CPTEG:CPH 10% RIF	86.7% ± 2.6%	0.818 ± 0.008	0.872 ± 0.004	0.00087 ± 0.20502
20:80 CPH:SA 3% RIF	29.8% ± 2.4%	0.240 ± 0.013	0.335 ± 0.009	0.06993 ± 0.96020
20:80 CPH:SA 1% RIF	49.3% ± 2.8%	0.106 ± 0.011	0.168 ± 0.010	0.06285 ± 0.96084
20:80 CPH:SA 20% CAM	59.8% ± 0.4%	0.886 ± 0.009	0.927 ± 0.006	0.00192 ± 0.47169
20:80 CPTEG:CPH 20% CAM	123.5% ± 13.0%	0.766 ± 0.020	0.782 ± 0.018	0.00055 ± 0.00821
20:80 CPH:SA 5% PZA	34.9% ± 4.8%	0.298 ± 0.053	0.449 ± 0.069	0.02168 ± 0.30458
20:80 CPTEG:CPH 5% PZA	89.7% ± 36.6%	0.057 ± 0.013	0.086 ± 0.020	0.00879 ± 0.33933

3.2. Identifying Factors that Influence Drug Release. The drug release profiles were parameterized using three attributes: (i) fraction released at two hours (FR (2h)); (ii) fraction released in one day (FR (24h)), both of which characterized the burst effect; and (iii) the slope of the release profile between 2 and 7 days to characterize the sustained release (Table 1). The normalized and mean-centered data are represented in the form of a heat map to provide an overview and to ensure that no outliers are biasing the results (Figure 3, Table S1-2). In this step,

no data specific to particle chemistry was included so as to not bias the analysis. A clustering analysis, based on Euclidian distance, was used to visualize broad trends in the data set between descriptors and nanoformulations, and is represented in Figure 3 by dendrograms which define the correlative indices. The clustering along the y-axis of Figure 3 can be visualized as plotting each nanoformulation in multidimensional space, where each dimension is a different descriptor. Encompassing n-dimensional "spheres" are defined at the locations of the nanoformulations, and as the radii of the spheres increase additional nanoformulations are encompassed. The relative sphere size needed to encompass multiple descriptors is comparable to the height of the branch in the dendrogram. Nanoformulations or descriptors grouped lower in the dendrograms are likely to show relatively strong, positive correlations. Branches higher in the dendrograms are more likely to show weak, positive correlations or inverse correlations.

From Figure 3, we find that the primary difference is between CPH:SA and CPTEG:CPH, given that the two chemistries branch off at the lowest correlation node. Therefore, particle chemistry is the key discriminator for nanoformulation behavior. Within each node, the compounds then group based on drug type, and then finally branch off based on theoretical drug loading and molar monomer ratios within the copolymer. This defines the order of importance on release properties with CPH:SA versus CPTEG:CPH as the most important and the theoretical drug loading having less importance. For CPH:SA nanoparticles, rifampicin and pyrazinamide grouped together strongly, whereas doxycycline and chloramphenicol grouped together within the CPTEG:CPH chemistries. The CPTEG:CPH-chloramphenicol and -pyrazinamide nanoformulations clustered together, and diverged from the CPTEG:CPH-rifampicin and -doxycycline nanoformulations. Considering correlations to the release properties, the fraction released at two hours and 24 hours are strongly correlated with the polymer melting point (T_m)

and zeta potential. The fraction released/day clustered with nanoparticle diameter and PDI polymer DOP and M_n. The relatively low branching of these properties in the dendrogram indicates moderate to strong correlation. Encapsulation efficiency (EE) was most strongly correlated with water solubility, followed by fraction released/day. This (weak) correlation to water solubility is expected, as incompatibility between polymer and drug hydrophobicity/hydrophilicity can result in drug partitioning more strongly in the antisolvent than the polymer matrix. The drug release properties (burst release, slope of release, and encapsulation efficiency) appeared relatively isolated from each other within the dendrogram, suggesting potential for independent control of these properties in designing nanoformulations.

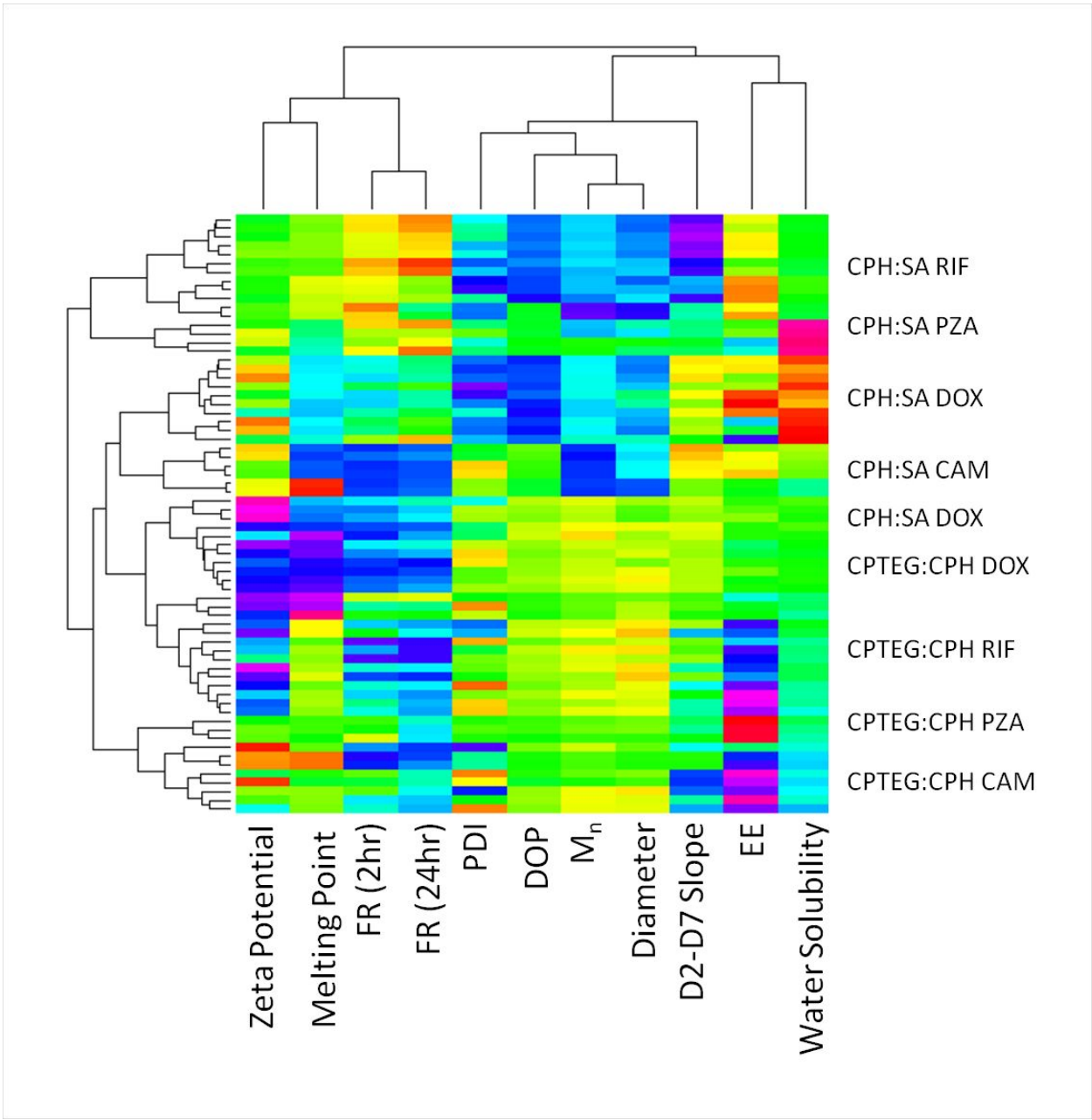


Figure 3. Representation of correlations in data using Euclidian distance-based clustering, with the dendrograms defining the degree of correlation (i.e., branches at the bottom of the dendrogram have high correlation and correlation decreases as moving along the branches). From the dendrograms, the key discriminators among nanoformulations (vertical axis) in the order of importance are carrier chemistry (CPTEG:CPH versus CPH:SA), drug type, and

theoretical drug loading. Concerning correlations between drug release properties and descriptors (horizontal axis), nanoparticle burst release (FR (2hr) and FR (24 hr)) was most strongly correlated with zeta potential and drug melting point. The release slope was most strongly correlated with the nanoparticle and polymer size properties. Encapsulation efficiency was most strongly correlated with the water solubility of the drug. That these data fall within comparable ranges demonstrates the robustness of the method and data set, enabling interrogation of nanoformulation behavior. Drug abbreviations: doxycycline (DOX), rifampicin (RIF), chloramphenicol (CAM), and pyrazinamide (PZA). Raw and processed input data are included in Tables S1-2.

A dimensionality reduction analysis, specifically principal component analysis (PCA), was then applied to the data of Figure 3, with descriptors specific to the particle and drug chemistry added to the data set (Table S3). Plots of formulation mapping and descriptor mapping within the dimensionally-reduced space are shown in Figures 4 and 5, respectively. In these figures, the principal components (PC) are ordered in terms of decreasing variability captured. PC1, the most important PC, captured particle chemistry properties (41.3%); therefore, differences in particle chemistry descriptors explain more variance in release behavior than other descriptor sets. The next most important PC, PC2, captured differences in drug-specific descriptors (27.9%). The scores plot (Figure 4), which maps individual nanoparticle formulations onto these PC's (which, between them, allow us to reliably capture correlations in those two dimensions), shows a clear separation between CPTEG:CPH and CPH:SA particle chemistries. Within each polymer, doxycycline, chloramphenicol, and pyrazinamide clustered together, whereas rifampicin formulations formed a cluster isolated from the other drugs, indicating potentially different types of interactions with the particle carriers.

The loadings plot (Figure 5), which maps the descriptor variables onto the PC's shows that the role of the particle and the drug bank descriptors have been isolated (i.e., particle data lies along the PC1 axis and drug bank data is along the PC2 axis). Given that PC1 is the most important axis, we are capturing that the particle chemistry is the critical characteristic for predicting particle release behavior. The drug release properties do not adhere exclusively to either PC1 or PC2 axes, indicating that they are influenced by both polymer and drug characteristics. The ability to isolate the different controls allows us to assess, model, and design by the material characteristics.

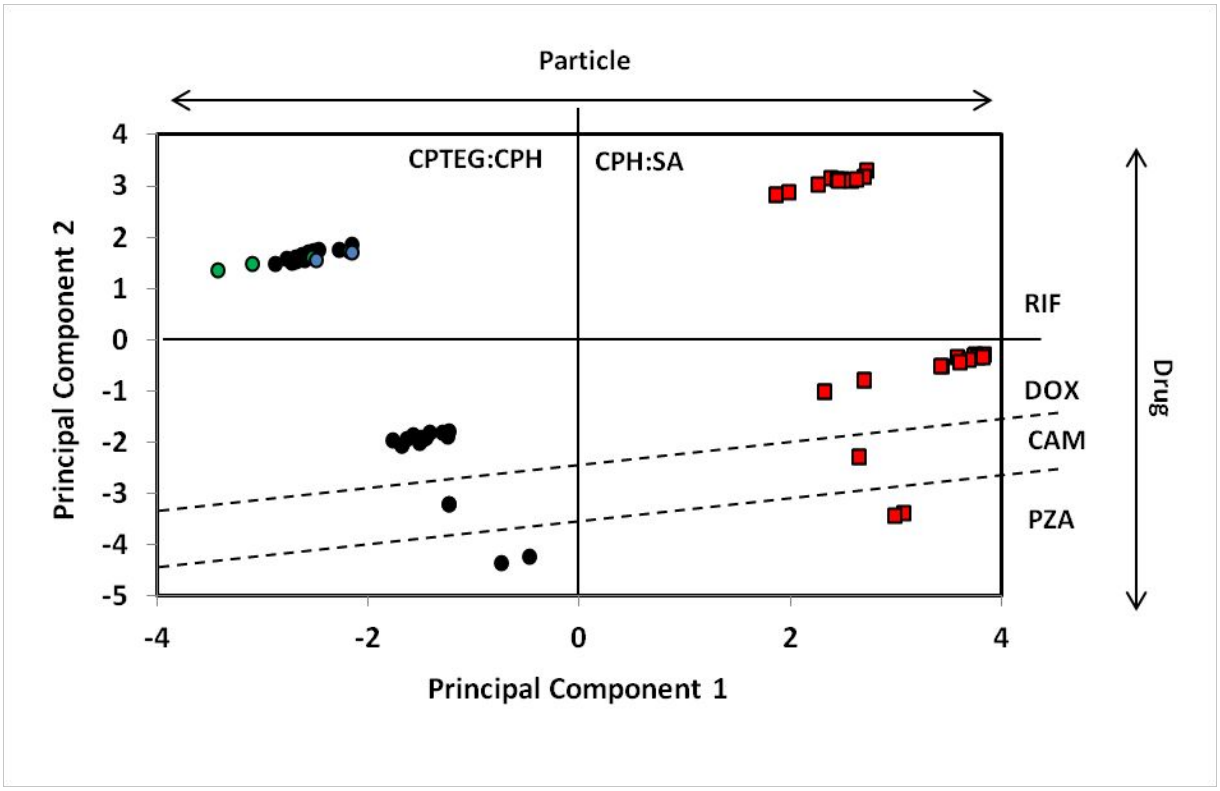


Figure 4. PCA scores plot. PC1 captures differences due to particle chemistry (10:90 CPTEG:CPH in blue, 20:80 CPTEG:CPH in black, and >30:<70 CPTEG:CPH in green) and PC2 captures the differences due to drug. There is a clear separation of formulations due to different

chemistries, with a demonstrated capability to isolate the effects of particle chemistry from drug properties. PC1 and PC2 captured 43.1% and 27.9% of variability, respectively. Drug abbreviations: doxycycline (DOX), rifampicin (RIF), chloramphenicol (CAM), and pyrazinamide (PZA). Raw input data are provided in Table S3.

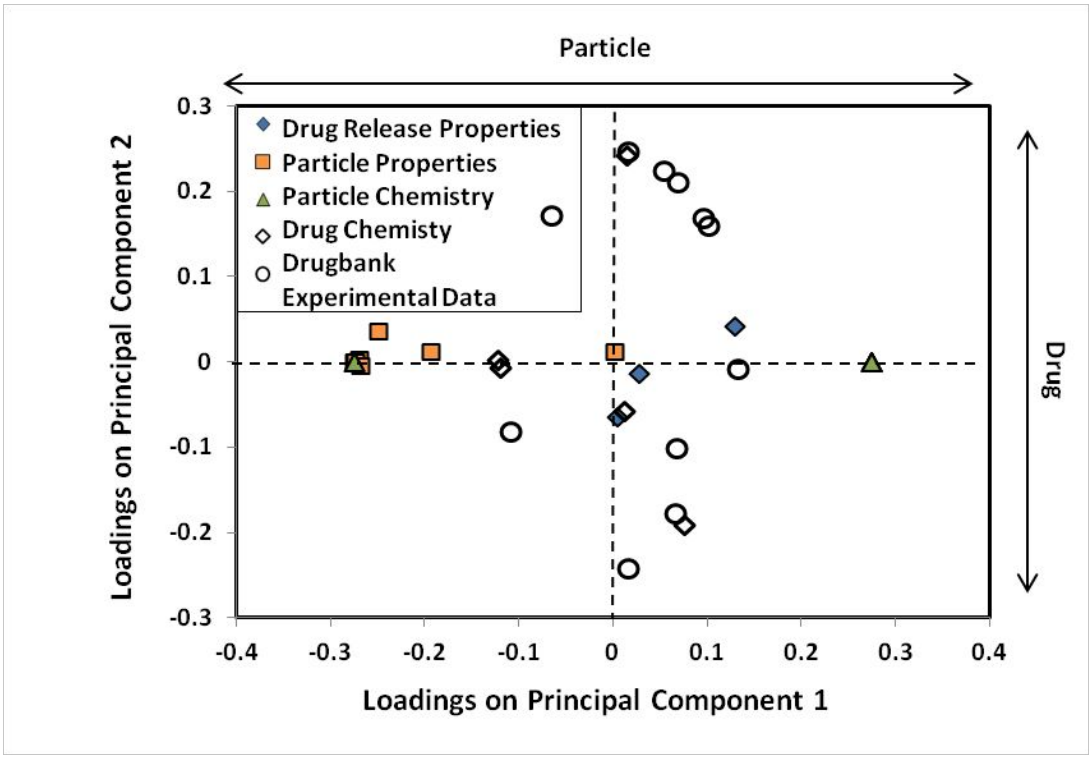


Figure 5. PCA loadings plot. Particle descriptors lie along the PC1 axis and drug descriptors lie along the PC2 axis. Drug release properties lie along both axes, indicating some dependence on both particle and drug descriptors. PC1 and PC2 captured 43.1% and 27.9% of variability, respectively. Raw input data are provided in Table S3.

To further quantify the correlation between descriptors and release properties, we calculated the VIP. In all, a total of 36 descriptors were used in the VIP analysis (as shown in Table 2), describing nanoparticle (1-3), polymer (4-13), and drug properties (14-36). The encapsulation

efficiency was most strongly correlated with zeta potential (-), % Cl (drug) (+), % O (drug) (+), $T_{m,drug}$ (-), predicted water solubility (-), pKa (strongest base) (-), and drug rotatable bond count (+). As seen in Figure 3, the two-hour burst and slope of release were highly correlated with each other, and showed similar correlations with the descriptors. Both the two-hour burst and slope of release were most strongly correlated with % Cl (drug) (+), % O (drug) (+), water solubility (-), % N (drug) (-), pKa (strongest acid) (-), and rotatable bond count (+). The identification of several highly correlated descriptors allows for reduction of the descriptor space to a minimum number and defines the number of descriptors necessary for performing high throughput calculations. This minimization is an important objective in computational modeling to improve model robustness. The purpose of VIP analysis is to assess the descriptors that contribute significant information as well as to identify correlated descriptors. While we identify the drug-related descriptors as having the highest individual impact, the particle-related descriptors collectively contribute the largest amount of information as seen in Figure 4.

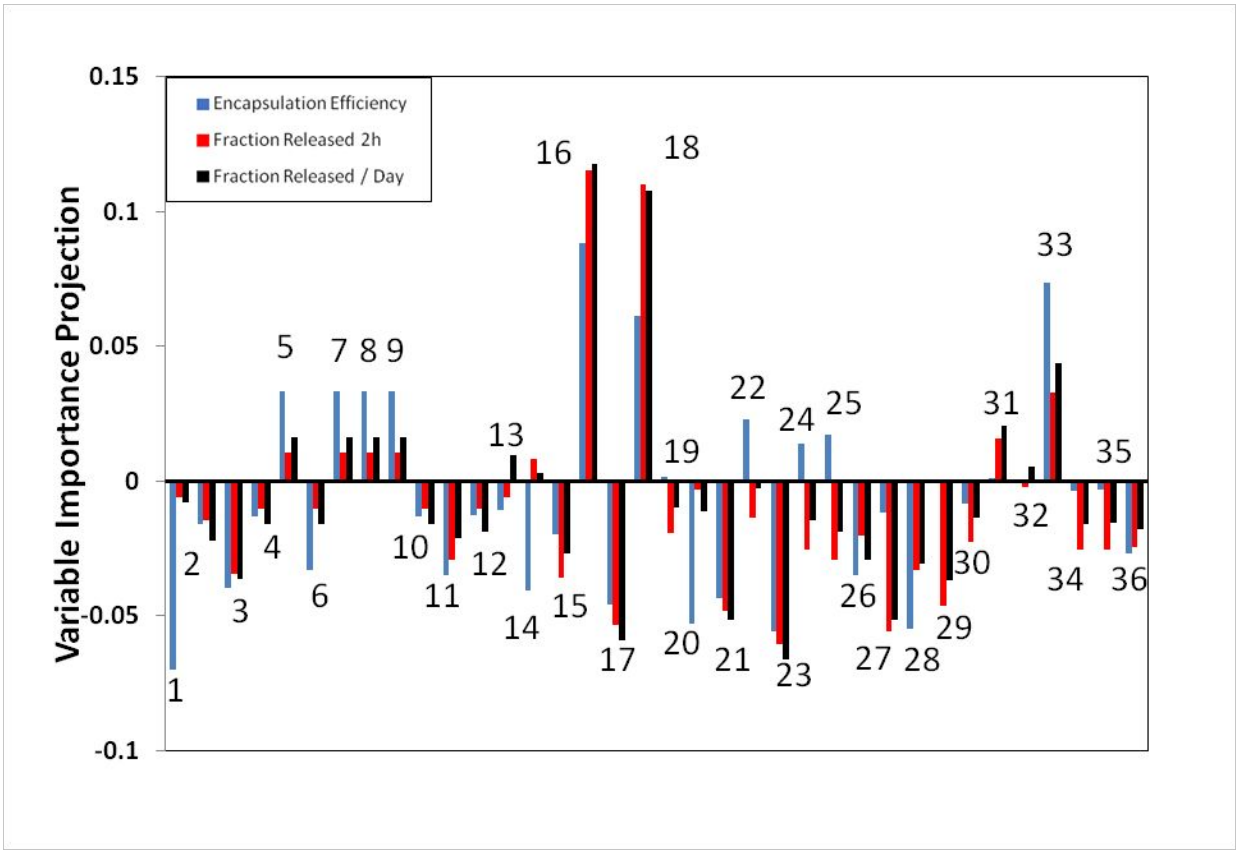


Figure 6. Variable importance projection of descriptors with respect to drug release properties. Descriptors are listed in Table 2. Positive VIP values correspond to positive correlation, and negative values to inverse correlation.

Table 2. List of descriptors used in VIP analysis.

1	Zeta potential	13	Tg (°C)	25	logP (predicted, ChemAxon)
2	Diameter	14	% C (Drug)	26	logS (predicted, ALGOPS)
3	PDI	15	% H (Drug)	27	pKa (Strongest Acidic)
4	Water contact angle	16	% Cl (Drug)	28	pKa (Strongest Basic)
5	Backbone O	17	% N (Drug)	29	Physiological charge
6	Aliphatic C	18	% O (Drug)	30	Hydrogen acceptor count
7	Aromatic C	19	molar mass	31	Hydrogen donor count
8	% O	20	Tm (°C, Drug)	32	Polar surface area
9	% H	21	water solubility (experimental)	33	Rotatable bond count
10	% C	22	logP (experimental)	34	Refractivity
11	DOP	23	water solubility (predicted, ALGOPS)	35	Polarizability
12	Mn (Da)	24	logP (predicted, ALGOPS)	36	Number of rings

3.3. Modeling release behavior. Beyond only observing the correlation of data, we wanted to identify similarities and design pathways between the various nanoformulations. This connectivity defines samples which have the most similar behavior and can provide information on potential replacements and design. In order to accomplish this, we performed a graph theory analysis (Figure 7). For the CPH:SA particle chemistries, there is a high connectivity (illustrated by black lines) and tight clustering within individual drugs. For the CPTEG:CPH particle chemistries, chloramphenicol, doxycycline, and pyrazinamide showed a high internal connectivity, but rifampicin branched out significantly. Each drug showed some degree of connectivity between the CPH:SA and CPTEG:CPH particle chemistries and doxycycline appeared to be the most interconnected across particle chemistries.

The degree of similarity can be defined by the number of connections required to connect two points. The distance along the two-dimensional projection also indicates the similarity of formulations. Pyrazinamide and chloramphenicol generated the least similarity in release behavior, as they required 4-6 connections, and lie far from each other along the projection.

Within each particle chemistry, doxycycline showed the most similarity to rifampicin and pyrazinamide, and rifampicin showed the most similarity to doxycycline and chloramphenicol. The branched region of the CPTEG:CPH-rifampicin nanoformulations indicates some dissimilarity from the other rifampicin nanoformulations and some unique behavior that will need to be explored more systematically using experiments. Of note, the rifampicin formulations with altered molar composition of CPTEG:CPH (from the 20:80 that makes up most of the data set) showed high similarity to the 20:80 CPTEG:CPH nanoformulations within the cluster. This would suggest that nano-carrier copolymer compositions can be interchanged within these rifampicin-loaded formulations without major impact.

This graph theory mapping in Figure 7 yielded notably different drug clustering within each nano-carrier chemistry compared to PCA (Figure 4). Rifampicin and chloramphenicol formulations are closely related in this map, while they were distant from each other in the PCA scores plot. Strikingly, chloramphenicol and pyrazinamide are most distant in the graph theory map, while they were clustered closely in the PCA scores plot. These clustering differences are likely due to PCA's limited ability to capture non-linear relationships. Non-linear modeling techniques like graph theory are better equipped to capture the non-linear release behavior arising from interactions between polymer and drug properties. In summary, the graph theory mapping defined similarity and connectivity between different nanoparticle formulations, while capturing non-linearity in relationships that can be lost in linear analysis.

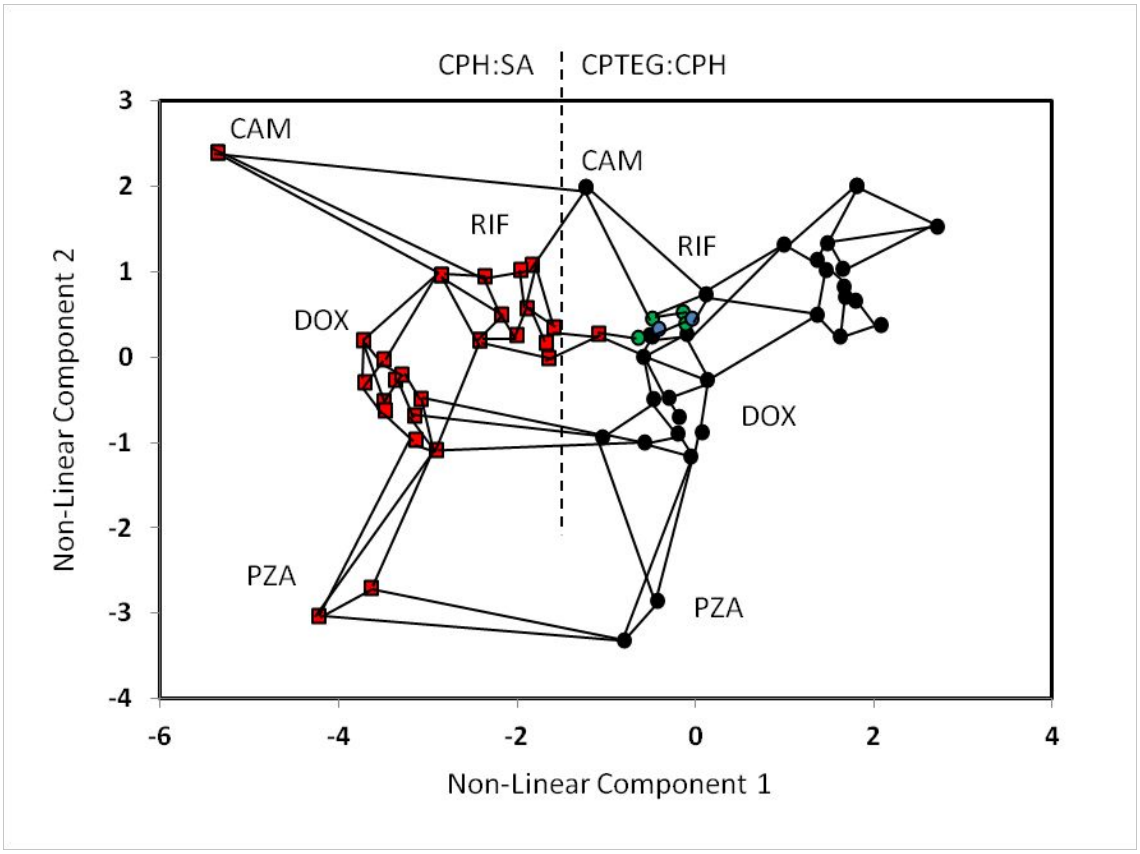


Figure 7. Graph theory map of formulation connectivity of release properties. Similarity between points is defined as the number of connections (solid lines) required to connect points. CPH:SA chemistries are represented by orange squares, while CPTEG:CPH chemistries are represented by circles (10:90 CPTEG:CPH in blue, 20:80 CPTEG:CPH in black, and >30:<70 CPTEG:CPH in green) This represents an approach for building a set of non-linearly derived parameters for performing high-throughput predictions. This approach was applied to a reduced descriptor set in order to develop a parameterization of the data, which ensures robustness by minimizing the number of input parameters, while incorporating non-linear relationships and maximizing variance in the data. Drug abbreviations: doxycycline (DOX), rifampicin (RIF), chloramphenicol (CAM), and pyrazinamide (PZA).

As PCA projects data onto a linear manifold, it has difficulty explaining non-linear relationships. To this end, PCA demonstrated insufficient capability to accurately predict release properties in this data set. By contrast, graph theory can be used to project the data onto a non-linear manifold. This provides high-throughput modeling that accounts for non-linearity without requiring so many terms as to reduce the robustness of analysis. Therefore, the input into the predictions defines the graph theory values of Figure 7, which reflects a non-linear combination of descriptors, and a multi-linear regression between these values and the drug release properties was developed. It should be noted that the drug release properties were not included in the non-linear parameterization used for the prediction input, because that would result in predicting a property as a function of itself. The result of the high throughput modeling is shown in Figure 8. This represents a model with non-linear parameters that are a function of nanoparticle chemistry and theoretical drug loading and is defined generally so as to be applied to a wide range of chemistries. These models are fairly accurate, with R^2 values ranging between 70.0% and 75.5%. Cross validation was applied to ensure an even trade-off between robustness and accuracy. Since these methods are based on descriptors that can be generated for potential new nanoparticle formulations, the models provide a method to virtually explore a large search space. This method can guide experimentation by predicting target properties for a desired release profile, suggesting chemistries that match the targeted properties for testing.

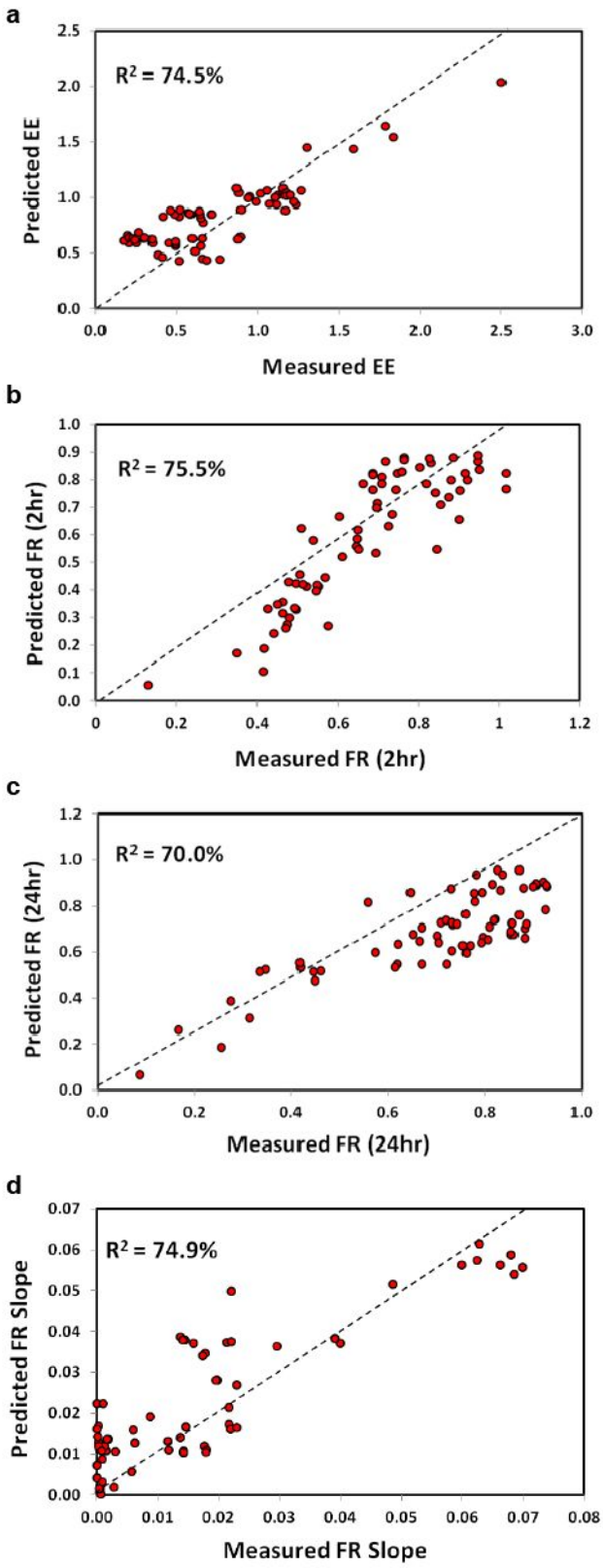

a
R² = 74.5%
Predicted EE
Measured EE
b
R² = 75.5%
Predicted FR (2hr)
Measured FR (2hr)
c
R² = 70.0%
Predicted FR (24hr)
Measured FR (24hr)
d
R² = 74.9%
Predicted FR Slope
Measured FR Slope

Figure 8. Graph theory high-throughput modeling of drug release properties. The horizontal axis is the experimental measurements. The vertical axis is the predicted encapsulation efficiency (a), 2-hour burst release (b), 24-hour burst release (c), and the d2-d7 release profile slope values from our model based on the reduced descriptor set. These calculations are based on a new hybrid informatics approach where non-linear manifold projections serve as the input, thereby accounting for greater complexity in descriptor-property relationships while also increasing the robustness of the models. The models are reasonably accurate for all tested release properties.

3.4. Model validation. To evaluate the robustness and accuracy of the multilinear models, nanoparticles encapsulating two new antibiotic drugs (not included in the training data set), meropenem and ceftazidime, were synthesized and characterized. Importantly, these drugs contain sulfur atoms (Figure S4) which are not present in the four drugs used in the original model training. The models in Figure 8 were used to predict the release properties for these new formulations (Table 3). Based on these predictions, it is expected that all eight formulations would show a high (>80%) burst release at 2h and 24h and minimal sustained release over d2-d7. With the exception of the 20:80 CPH:SA–meropenem formulation, all other nanoformulations are expected to show near-100% encapsulation efficiency.

Strikingly, these predictions match experimental results closely. These new nano-formulations displayed similar release profiles characterized by a >90% burst release within two hours, followed by small amounts of drug released over the following two weeks (Figure S5, Table 3). For this data set, the models tended to under-predict the burst release and over-predict the sustained release behavior of the nano-formulations. The EE model was relatively accurate for 20:80 CPH:SA formulations, within 5%-20% of the measured EE's. The EE model showed more

deviation from measured values for 20:80 CPTEG:CPH formulations, at circa 15-35% differences from the experimental values.

Table 3. Parameterized release properties of nanoparticles encapsulating two drugs (meropenem and ceftazidime) not included in training data set. FR (2h) fraction released in two-hour burst, FR (24h) fraction released in 24-hour burst.

	EE		FR (2h)		FR (24h)		FR slope (FR/day)	
Nano-formulation	predicted	measured	predicted	measured	predicted	measured	predicted	measured
20:80 CPTEG:CPH 5% meropenem	159%	195.0%	0.83	0.9406	0.84	0.9408	0.0033	0.000853
20:80 CPTEG:CPH 10% meropenem	128%	140.9%	0.81	0.9616	0.86	0.9617	0.0026	0.000271
20:80 CPH:SA 5% meropenem	29%	34.5%	0.95	0.9933	0.90	0.9933	0.0011	0.000656
20:80 CPH:SA 10% meropenem	49%	61.7%	0.98	0.9965	0.91	0.9967	0.0014	0.000168
20:80 CPTEG:CPH 5% ceftazidime	175%	156.8%	0.92	0.9997	0.88	1.0000	0.0019	0.000001
20:80 CPTEG:CPH 10% ceftazidime	190%	159.9%	0.91	0.9997	0.87	1.0000	0.0041	0.000004
20:80 CPH:SA 5% ceftazidime	114%	91.8%	0.97	0.9915	0.91	0.9952	0.0031	0.000771
20:80 CPH:SA 10% ceftazidime	129%	112.7%	0.94	0.9910	0.91	0.9941	0.0051	0.000686

To test the robustness of the models when adding new, untrained chemistries, eight nano-formulations were included in the models (compositional percentages were calculated including sulfur atoms, but without a separate descriptor for sulfur) and new regressions were calculated. We found R^2 values for EE, FR (2h), FR (24h) and d2-d7 slope after these inclusions to be 74.3%, 75.5%, 69.9%, and 74.6%, respectively. The small changes in regression from the original model data in Figure 8 indicates that the analytics methodology was able to incorporate new drug chemistries with minor impacts on the models. This confirms the robustness of the model and its capability to screen drug and polymer chemistries not included in the model development.

4. DISCUSSION

Due to the wide diversity of microbial infections, nanomedicines need to be customizable. Infections that are responsive to antibiotics may benefit from sustained release-skewing

formulations by leveraging the dose sparing properties, limiting the risk of off-target effects, reducing the number of administrations, and enhancing patient compliance.[4,8,39] Polyanhydride nanoparticles represent an attractive and adaptable nanomedicine platform by virtue of their tunable degradation and payload release rates,[25,40] high biocompatibility,[41,42] and efficient internalization by phagocytic cells.[7]

Predictive analytics approaches have the potential to accelerate nanomedicine clinical translation, but the application of such informatics and data mining techniques to nanomedicine design has been slow to develop.[13] To date, the majority of such efforts has focused on either linear dimensionality reduction through PCA and regression through PLS,[27] which provides insight into relationships between formulations and variables but has limited capacity to capture nonlinear behavior, or else artificial neural network "black box" models,[43–45] which can capture nonlinear behavior but obscure interpretation of the structure of the model and dataspace. As the long-term goal of this research is to facilitate rational design of nanomedicine formulations, interpretation of the relationships between formulations is important. Accordingly, the dimensionality reduction approach was selected for this research, and paired with graph theory mapping to overcome the linearity limitations of PCA.

A hybrid data mining approach was employed to deconvolute the complex polymer and drug relationships and develop QSPRs that describe release kinetics and encapsulation efficiency. We correlated antibiotic release properties from varying polyanhydride chemistries, encapsulated drug types, and drug loading within the nanoparticles. Through PCA analysis, we showed that release properties are dependent on both copolymer chemistry properties and drug properties, with polymer properties being more important. VIP analysis identified key polymer and drug

descriptors that predicted drug release and encapsulation properties, but PCA was insufficient to predict release behavior from these formulations.

Graph theory was used to characterize the multilinear connectedness and similarity of formulations, which can guide selection of replacement formulations with similar release behavior. For example, it is expected that 20:80 CPH:SA rifampicin-loaded nanoparticles (Figure 2b) would demonstrate similar release behavior (including burst release, slope of release between days 2-7, and encapsulation efficiency) as 20:80 CPH:SA doxycycline-loaded nanoparticles (Figure 2a) based on their close connections and proximal distance on the map (Figure 7). Similarly, 20:80 CPTEG:CPH pyrazinamide-loaded nanoparticles (Figure 2d) would be expected to show large differences in release behavior from 20:80 CPH:SA chloramphenicol-loaded nanoparticles (Figure 2c) due to the large number of lines needed to connect them and far distance on the map (Figure 7). The descriptors identified by VIP analysis were paired with the multilinear mapping from graph theory to generate predictive models for *a priori* screening of nanoparticle formulations with desired release kinetics and high encapsulation efficiency.

The physicochemical properties of compounds influence their distribution, either in blood plasma or a polymer matrix. To this effect, VIP analysis (Figure 6) indicated that the descriptors most strongly correlated with release properties were both polymer and drug properties. This is expected, as favorable mixing thermodynamics allows distribution of the drugs inside the polymer device.[10,46] In polyanhydride nanoparticles, such a distribution allows an erosion-controlled release profile, which tends toward sustained release.[9,11] In contrast, poor mixing between the polymer and drug induces thermodynamic partitioning of the drug into polymer microdomains and/or localization at the particle surface, which skews the release profile toward

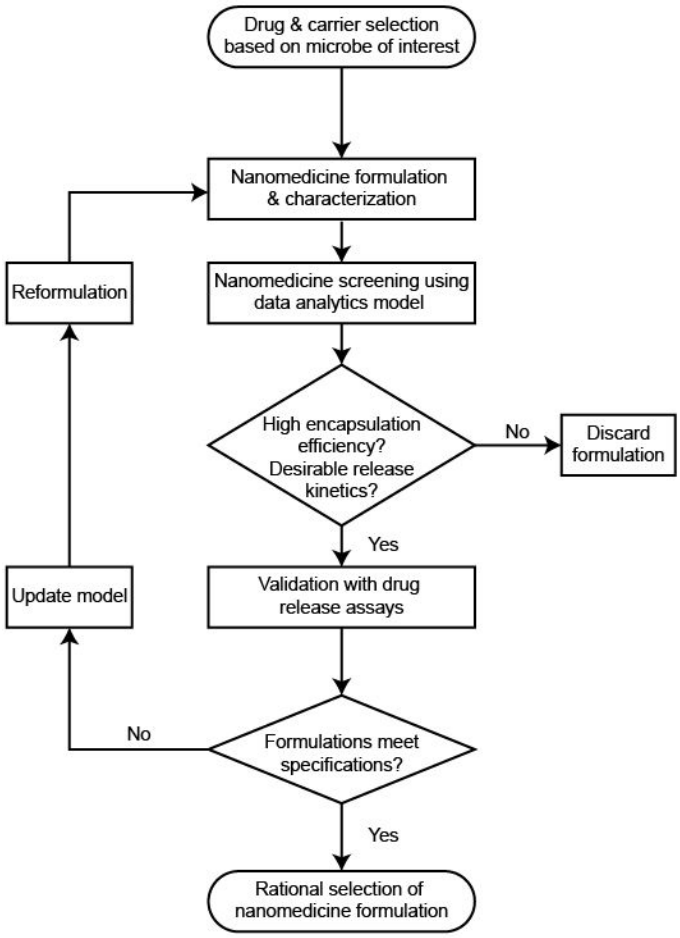
a high-burst, diffusion-dominated regime.[11] Many of these same drug properties were correlated with encapsulation efficiency, supporting the notion that polymer-drug mixing influences the carrying capacity of delivery devices. As empty polyanhydride nanoparticles have a moderately negative zeta potential,[23] the strong negative correlation between zeta potential and encapsulation efficiency could reflect a strong surface localization of positively charged drugs. If this were the case, however, we would expect a strong positive correlation between zeta potential and the two-hour burst release, which was not observed. Regardless, the predictive power of this descriptor could support the use of zeta potential as a quality control metric to ensure consistent encapsulation efficiencies of lead formulations. While it is not surprising that these drug properties affect encapsulation and release kinetics, this informatics analysis provides a sense of their relative impact. Reducing the data space in this way can help guide rational selection of antibiotic and polymer carrier pairs for nanomedicine formulations. These observations underline the complexity of these relationships and provide support for the use of data analytics approaches to enable rational design of nanomedicines.

It should be noted that we can only confidently make quantitative predictions in chemical spaces represented in our training data. While the additional testing of drugs containing sulfur, which was not represented in our training data, resulted in approximately no change in accuracy, materials that have unique behavior but with chemistries outside our training data may not be quantitatively described by this approach. However, even in these cases, our approach has significant impact. While the objective for the systems described by our training data is to predict properties with high accuracy, the objective for systems containing groups and elements not in our training data is to identify polymer and drug combinations which have the most promising

characteristics and identify where additional experiments are needed. This leads to an iterative approach where necessary experiments are identified, thus feeding back to the analysis.

From all of these results, we propose a framework for rational design and rapid testing of nanomedicine formulations (Scheme 1). In the first step, selected antibiotic drug candidates are encapsulated within nanoparticles of various polymer chemistries (potentially using high throughput techniques,[23] as demonstrated in section 3.4), and characterization of size distribution by SEM and zeta potential is obtained. These nanoparticle characteristics, along with polymer properties and drug properties, can then be fed into the multilinear graph theory model to predict encapsulation efficiencies and release kinetics. Nanomedicine candidates that demonstrate insufficient encapsulation and/or undesirable drug release profiles can be discarded. The *in vitro* performance of the lead nanomedicine candidates that emerge from this step can then be validated using drug release kinetics assays. A feedback reformulation loop allows gradual optimization of nanomedicine formulations and iterative updates to the models when release behavior deviates from predictions. In theory, this framework could be expanded to include other performance metrics, including internalization by appropriate cells and biological efficacy. As this methodology uses standard polymer and nanoparticle characterization techniques used in nano-carrier drug delivery research and publicly available drug information, this approach could be expanded to include other types of polymeric materials and other classes of small molecule drugs. This data analytics framework constitutes the first steps toward the rational design of nanomedicine formulations for antimicrobial therapies.

Scheme 1. Data analytics framework for rapid nanomedicine design and screening.



5. CONCLUSIONS

A multivariate data analytics approach was used to correlate drug release profiles from nanomedicine formulations based on different polyanhydride chemistries, encapsulated antibiotic drug type, and varying drug loading. We showed that both drug and polymer properties influence the drug encapsulation efficiency within the nanoparticles, the prevalence of burst in the drug release profile, and the slope of post-burst release. Polymer and drug properties that significantly impacted drug encapsulation efficiency and release kinetics were identified and defined a minimum descriptor set. The informatics analysis captured and preserved non-linear behavior governing relationships between drug type, polymer chemistry, and nanoparticle release properties, enabling interrogation of nanomedicine design pathways. We developed predictive

models for drug release kinetics of untested drugs, using data from the Drugbank database and nano-carrier characterization as inputs. Release kinetics predictions of two drugs containing atoms not included in the model showed good agreement with experimental results, validating the model and indicating its potential to virtually explore new polymer and drug pairs not included in training data set. The models were shown to be robust after inclusion of these new formulations in that there were no significant changes in the model regressions. This multilinear modeling approach provides the first steps towards development of a framework that can be used to rationally design nanomedicine formulations by selecting the appropriate carrier for a drug payload to program desirable release kinetics profiles.

6. SUPPORTING INFORMATION

Dendrogram raw and processed input data; Principal component analysis raw input data; Chemical structures of drugs in validation data set; Release kinetics of drugs in validation data set.

AUTHOR INFORMATION

Corresponding Author

*Email: nbalaji@iastate.edu

Address: 2035 Sweeney Hall, 618 Bissell Road, Ames, IA 50011-2230

ORCID

Balaji Narasimhan: 0000-0002-7955-5353

Adam S. Mullis: 0000-0002-5186-5093

Present Addresses

†If an author's address is different than the one given in the affiliation line, this information may be included here.

Author Contributions

ASM, BN, AMB, NPB, and BHB designed the experiments. ASM performed the experiments. ASM, SB, KR, and BN designed the informatics approach. SB and KR performed the informatics analysis. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

ISU Nanovaccine Institute; Defense Threat Reduction Agency (Contract HDTRA119C0005); National Science Foundation (Grant No. 1640867).

ACKNOWLEDGMENTS

B.N. acknowledges the Vlasta Klima Balloun Faculty Chair.

ABBREVIATIONS

CAM, chloramphenicol; CAZ, ceftazidime; CPH, 1,6-bis(*p*-carboxyphenoxy)hexane; CPTEG, 1,8-bis(*p*-carboxyphenoxy)-3,6-dioxaoctane; DOX, doxycycline; DSC, dynamic scanning calorimetry; EE, encapsulation efficiency; FR, fraction released; MEM, meropenem; NMR, nuclear magnetic resonance; PCA, principal component analysis; PZA, pyrazinamide; RIF, rifampicin; QSPR, quantitative structure-property relationship; SA, sebacic acid; SEM, scanning electron microscopy; T_g, glass transition temperature; T_m, melting point; VIP, variable importance projection

REFERENCES

(1) Arora, D.; Sharma, N.; Sharma, V.; Abrol, V.; Shankar, R.; Jaglan, S. An Update on Polysaccharide-Based Nanomaterials for Antimicrobial Applications. *Appl. Microbiol.*

Biotechnol. **2016**, *100* (6), 2603–2615.

(2) Vorachit, M.; Chongtrakool, P.; Arkomsean, S.; Boonsong, S. Antimicrobial Resistance in Burkholderia Pseudomallei. *Acta Trop* **2000**, *74* (2–3), 139–144.

(3) Seung, K. J.; Keshavjee, S.; Rich, M. L. Drug-Resistant Tuberculosis. *Cold Spring Harb Perspect Med* **2015**, *5*, 1–20.

(4) Narasimhan, B.; Goodman, J. T.; Vela Ramirez, J. E. Rational Design of Targeted Next-Generation Carriers for Drug and Vaccine Delivery. *Annu. Rev. Biomed. Eng* **2016**, *18* (1), 25–49.

(5) Chan, C.-F.; Huang, K.-S.; Lee, M.-Y.; Yang, C.-H.; Wang, C.-Y.; Lin, Y.-S. Applications of Nanoparticles for Antimicrobial Activity and Drug Delivery. *Curr. Org. Chem.* **2014**, *18* (2), 204–215.

(6) Lueth, P.; Haughney, S. L.; Binnebose, A. M.; Mullis, A. S.; Peroutka-Bigus, N.; Narasimhan, B.; Bellaire, B. H. Nanotherapeutic Provides Dose Sparing and Improved Antimicrobial Activity against Brucella Melitensis Infections. *J. Control. Release* **2019**, *294*, 288–297.

(7) Phanse, Y.; Lueth, P.; Ramer-Tait, A. E.; Carrillo-Conde, B. R.; Wannemuehler, M. J.; Narasimhan, B.; Bellaire, B. H. Cellular Internalization Mechanisms of Polyanhydride Particles: Implications for Rational Design of Drug Delivery Vehicles. *J. Biomed. Nanotechnol.* **2016**, *12* (7), 1544–1552.

(8) Binnebose, A. M.; Haughney, S. L.; Martin, R.; Imerman, P. M.; Narasimhan, B.;

Bellaire, B. H. Polyanhydride Nanoparticle Delivery Platform Dramatically Enhances Killing of Filarial Worms. *PLoS Negl. Trop. Dis.* **2015**, *9* (10), e0004173.

(9) Shen, E.; Kipper, M. J.; Dziadul, B.; Lim, M.-K.; Narasimhan, B. Mechanistic Relationships between Polymer Microstructure and Drug Release Kinetics in Bioerodible Polyanhydrides. *J. Control. Release* **2002**, *82* (1), 115–125.

(10) Shen, E.; Pizsczek, R.; Dziadul, B.; Narasimhan, B. Microphase Separation in Bioerodible Copolymers for Drug Delivery. *Biomaterials* **2001**, *22* (3), 201–210.

(11) Berkland, C.; Kipper, M. J.; Narasimhan, B.; Kim, K.; Pack, D. W. Microsphere Size, Precipitation Kinetics and Drug Distribution Control Drug Release from Biodegradable Polyanhydride Microspheres. *J. Control. Release* **2004**, *94* (1), 129–141.

(12) Barnard, A. S. Challenges in Modelling Nanoparticles for Drug Delivery. *J. Phys. Condens. Matter* **2016**, *28* (2), 023002.

(13) Jones, D. E.; Ghandehari, H.; Facelli, J. C. A Review of the Applications of Data Mining and Machine Learning for the Prediction of Biomedical Properties of Nanoparticles. *Comput. Methods Programs Biomed.* **2016**, *132*, 93–103.

(14) Tenenbaum, J.; Silva, V.; Langford, J. A Global Geometric Framewok for Nonlinear Dimensionality Reduction. *Science (80-.).* **2000**, *290* (December), 2319–2323.

(15) Broderick, S.; Rajan, K. Informatics Derived Materials Databases for Multifunctional Properties. *Sci. Technol. Adv. Mater.* **2015**, *16* (1), 1–8.

(16) Li, X.; Petersen, L.; Broderick, S.; Narasimhan, B.; Rajan, K. Identifying Factors

Controlling Protein Release from Combinatorial Biomaterial Libraries via Hybrid Data Mining Methods. *ACS Comb. Sci.* **2011**, *13* (1), 50–58.

(17) Ulery, B. D.; Petersen, L. K.; Phanse, Y.; Kong, C. S.; Broderick, S. R.; Kumar, D.; Ramer-Tait, A. E.; Carrillo-Conde, B.; Rajan, K.; Wannemuehler, M. J.; et al. Rational Design of Pathogen-Mimicking Amphiphilic Materials as Nanoadjuvants. *Sci. Rep.* **2011**, *1*, 1–9.

(18) Petersen, L. K.; Ramer-Tait, A. E.; Broderick, S. R.; Kong, C. S.; Ulery, B. D.; Rajan, K.; Wannemuehler, M. J.; Narasimhan, B. Activation of Innate Immune Responses in a Pathogen-Mimicking Manner by Amphiphilic Polyanhydride Nanoparticle Adjuvants. *Biomaterials* **2011**, *32* (28), 6815–6822.

(19) Phanse, Y.; Carrillo-Conde, B. R.; Ramer-Tait, A. E.; Roychoudhury, R.; Pohl, N. L. B.; Narasimhan, B.; Wannemuehler, M. J.; Bellaire, B. H. Functionalization of Polyanhydride Microparticles with Di-Mannose Influences Uptake by and Intracellular Fate within Dendritic Cells. *Acta Biomater.* **2013**, *9* (11), 8902–8909.

(20) Phanse, Y.; Carrillo-Conde, B. R.; Ramer-Tait, A. E.; Roychoudhury, R.; Broderick, S.; Pohl, N.; Rajan, K.; Narasimhan, B.; Wannemuehler, M. J.; Bellaire, B. H. Functionalization Promotes Pathogen-Mimicking Characteristics of Polyanhydride Nanoparticle Adjuvants. *J. Biomed. Mater. Res. - Part A* **2017**, *105* (10), 2762–2771.

(21) Torres, M. P.; Vogel, B. M.; Narasimhan, B.; Mallapragada, S. K. Synthesis and Characterization of Novel Polyanhydrides with Tailored Erosion Mechanisms. *J. Biomed. Mater. Res. - Part A* **2006**, *76* (1), 102–110.

(22) Conix, A. Poly[1,3-Bis(p-Carboxyphenoxy)-Propane Anhydride]. *Macromol. Synth.* **1966**, *2*, 95–98.

(23) Goodman, J. T.; Mullis, A. S.; Dunshee, L.; Mitra, A.; Narasimhan, B. Automated High-Throughput Synthesis of Protein-Loaded Polyanhydride Nanoparticle Libraries. *ACS Comb. Sci.* **2018**, *20* (5), 298–307.

(24) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, *9* (7), 676–682.

(25) Carrillo-Conde, B. R.; Darling, R. J.; Seiler, S. J.; Ramer-Tait, A. E.; Wannemuehler, M. J.; Narasimhan, B. Sustained Release and Stabilization of Therapeutic Antibodies Using Amphiphilic Polyanhydride Nanoparticles. *Chem. Eng. Sci.* **2015**, *125*, 98–107.

(26) Prasanthi, B.; Ratna, J. V.; Phani, R. S. C. Development and Validation of RP-HPLC Method for Simultaneous Estimation of Rifampicin, Isoniazid and Pyrazinamide in Human Plasma. *J. Anal. Chem.* **2015**, *70* (8), 1015–1022.

(27) Silva, J.; Mendes, M.; Cova, T.; Sousa, J.; Pais, A.; Vitorino, C. Unstructured Formulation Data Analysis for the Optimization of Lipid Nanoparticle Drug Delivery Vehicles. *AAPS PharmSciTech* **2018**, *19* (5), 1–12.

(28) Broderick, S. R.; Suh, C.; Provine, J.; Roper, C. S.; Maboudian, R.; Howe, R. T.; Rajan, K. Application of Principal Component Analysis to a Full Profile Correlative Analysis of FTIR Spectra. *Surf. Interface Anal.* **2012**, *44* (3), 365–371.

(29) Ashton, M.; Hennig, R. G.; Broderick, S. R.; Rajan, K.; Sinnott, S. B. Computational Discovery of Stable M2AX Phases. *Phys. Rev. B* **2016**, *94* (5), 054116.

(30) Ericksson, L.; Byrne, T.; Johansson, E.; Trygg, J.; Vikstrom, C. *Multi- and Megavariate Data Analysis : Basic Principles and Applications*; Umetrics Ab,: Umea, 2001.

(31) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130.

(32) Nguyen, D. V.; Rocke, D. M. Tumor Classification by Partial Least Squares Using Microarray Gene Expression Data. *Bioinformatics* **2002**, *18* (1), 39–50.

(33) Balachandran, P. V.; Broderick, S. R.; Rajan, K. Identifying the "inorganic Gene" for High-Temperature Piezoelectric Perovskites through Statistical Learning. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2011**, *467* (2132), 2271–2290.

(34) Wodo, O.; Broderick, S.; Rajan, K. Microstructural Informatics for Accelerating the Discovery of Processing-Microstructure-Property Relationships. *MRS Bull.* **2016**, *41* (8), 603–609.

(35) Srinivasan, S.; Broderick, S. R.; Zhang, R.; Mishra, A.; Sinnott, S. B.; Saxena, S. K.; LeBeau, J. M.; Rajan, K. Mapping Chemical Selection Pathways for Designing Multicomponent Alloys: An Informatics Framework for Materials Design. *Sci. Rep.* **2015**, *5*, 1–8.

(36) Ulery, B. D.; Phanse, Y.; Sinha, a.; Wannemuehler, M. J.; Narasimhan, B.; Bellaire, B. H. Polymer Chemistry Influences Monocytic Uptake of Polyanhydride Nanospheres. *Pharm.*

Res. **2009**, *26* (3), 683–690.

(37) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074–D1082.

(38) Kipper, M. J.; Shen, E.; Determan, A.; Narasimhan, B. Design of an Injectable System Based on Bioerodible Polyanhydride Microspheres for Sustained Drug Delivery. *Biomaterials* **2002**, *23* (22), 4405–4412.

(39) Thomas, S. N.; Schudel, A. Overcoming Transport Barriers for Interstitial-, Lymphatic-, and Lymph Node-Targeted Drug Delivery. *Curr. Opin. Chem. Eng.* **2015**, *7*, 65–74.

(40) Petersen, L. K.; Sackett, C. K.; Narasimhan, B. Novel, High Throughput Method to Study in Vitro Protein Release from Polymer Nanospheres. *J. Comb. Chem.* **2010**, *12* (1), 51–56.

(41) Huntimer, L.; Ramer-Tait, A. E.; Petersen, L. K.; Ross, K. a.; Walz, K. a.; Wang, C.; Hostetter, J.; Narasimhan, B.; Wannemuehler, M. J. Evaluation of Biocompatibility and Administration Site Reactogenicity of Polyanhydride-Particle-Based Platform for Vaccine Delivery. *Adv. Healthc. Mater.* **2013**, *2* (2), 369–378.

(42) Adler, A. F.; Petersen, L. K.; Wilson, J. H.; Torres, M. P.; Thorstenson, J. B.; Gardner, S. W.; Mallapragada, S. K.; Wannemuehler, M. J.; Narasimhan, B. High Throughput Cell-Based Screening of Biodegradable Polyanhydride Libraries. *Comb. Chem. High Throughput Screen.* **2009**, *12* (7), 634–645.

(43) Metwally, A. A.; Hathout, R. M. Computer-Assisted Drug Formulation Design: Novel

Approach in Drug Delivery. *Mol. Pharm.* **2015**, *12* (8), 2800–2810.

(44) Shalaby, K. S.; Soliman, M. E.; Casettari, L.; Bonacucina, G.; Cespi, M.; Palmieri, G. F.; Sammour, O. A.; El Shamy, A. A. Determination of Factors Controlling the Particle Size and Entrapment Efficiency of Noscapine in PEG/PLA Nanoparticles Using Artificial Neural Networks. *Int. J. Nanomedicine* **2014**, *9* (1), 4953–4964.

(45) Husseini, G. A.; Mjalli, F. S.; Pitt, W. G.; Abdel-Jabbar, N. M. Using Artificial Neural Networks and Model Predictive Control to Optimize Acoustically Assisted Doxorubicin Release from Polymeric Micelles. *Technol. Cancer Res. Treat.* **2009**, *8* (6), 479–488.

(46) Marsac, P. J.; Shamblin, S. L.; Taylor, L. S. Theoretical and Practical Approaches for Prediction of Drug–Polymer Miscibility and Solubility. *Pharm. Res.* **2006**, *23* (10), 2417–2426.

TABLE OF CONTENTS GRAPHIC