ACS Combinatorial Science

RESEARCH ARTICLE

pubs.acs.org/acscombsci

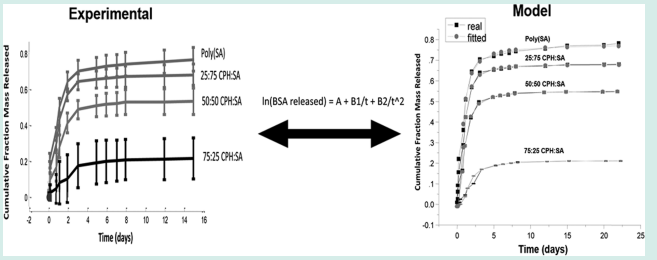
Identifying Factors Controlling Protein Release from Combinatorial Biomaterial Libraries via Hybrid Data Mining Methods

Xue Li,[†,‡] Latrisha Petersen,[†,§] Scott Broderick,[†,‡] Balaji Narasimhan,[†,§] and Krishna Rajan*[,†,‡]

[†]Institute for Combinatorial Discovery, [‡]Department of Materials Science and Engineering, and [§]Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011, United States

Supporting Information

ABSTRACT: Polyanhydrides are a class of degradable biomaterials that have shown much promise for applications in drug and vaccine delivery. Their properties can be tailored for controlled drug release, drug/protein stability, and immune regulation (adjuvant effect). Identifying the relationship between the molecular structures of the polymers and the drug release kinetics profiles would help understand the release mechanism and aid in the accurate prediction of drug release and the rational design of polymer-based drug carrier systems. The molecular structure descriptors that had the most impact on the release kinetics were identified using a prediction/optimization data mining approach. Using this new approach for modeling nonlinear release kinetics behavior, we determined that the descriptors which had the greatest effect on the release kinetics were the number of backbone −COO− nonconjugated bonds, the number of aromatic rings, and the number of $-CH_2-$ bonds.



KEYWORDS: polyanhydrides, degradable biomaterials, drug delivery, vaccine delivery, drug release kinetics

1. INTRODUCTION

The use of biodegradable polymers as controlled delivery devices has a significant advantage over competing delivery systems. Their chemistries can be tailored to stabilize protein drugs and provide controlled, sustained drug release. Their degradation products are biocompatible, nonmutagenic, and inhalable or injectable (i.e., there is no need to surgically remove the device). The release kinetics can be modulated by altering the copolymer composition or by changing the structure of copolymers. Understanding the relationship between the molecular structure of the polymers used in these devices and their drug release kinetics profiles will aid in understanding the release mechanism and the accurate prediction of drug release, which will lead to the rational design of polymer carrier systems. The class of biodegradable polymers considered here are polyanhydrides based on copolymers of sebacic acid (SA) and 1,6-bis(*p*-carboxyphenoxy)-hexane (CPH) and of 1,8-bis(*p*-carboxyphenoxy)3,6-dioxaoctane (CPTEG) and CPH. CPTEG/CPH and CPH/SA copolymers were synthesized in a combinatorial format linearly varying in composition, loaded with a model protein, bovine serum albumin (BSA), and dried to form BSA-loaded polymer film libraries.

Poly(CPH) and poly(SA) are hydrophobic polymers, which can prevent covalent aggregation by reducing the water penetration into the core, whereas the amphiphilic CPTEG is added to enhance the stability of the protein by preventing noncovalent aggregation and acid-induced hydrolysis.[1−4] The drug release profiles from polyanhydride carriers can vary from days to months depending on the polymer chemistry, drug partitioning, and drug loading. The degradation via hydrolysis into their acidic monomers takes a period of months for poly(CPH), weeks for poly(SA),[5,6] and days for poly(CPTEG).[7]

Previous work has shown an important relationship between polymer microstructure and drug release kinetics of the CPH:SA polyanhydride system.[8] The experimental results demonstrated that encapsulated drugs can partition into the microphase separated domains of the copolymers (when present, for example, in 20:80 CPH/SA), which can affect the subsequent drug release kinetics. Additionally, it is known that erosion kinetics are closely related to protein release kinetics in this polyanhydride system because of their surface erosion mechanism. For controlled-release applications, drug release kinetics are controlled by the erosion kinetics, rather than by swelling and diffusion as in some bulk-eroding systems.[9] Previous work in our laboratories focused on developing a high-throughput platform to investigate protein release from this polyanhydride system.[10] This platform enabled the creation of a library of numerous different polymers, which were screened in parallel for protein release with a fluorescensce-based detection method; this screening resulted in the identification of chemistry and pH dependent protein release mechanisms. Several approaches to understand and model the erosion kinetics of biodegradable polymer carriers have been proposed in the literature. Thombre and Himmelstein,[11]

Received: September 27, 2010
Revised: October 11, 2010
Published: November 10, 2010

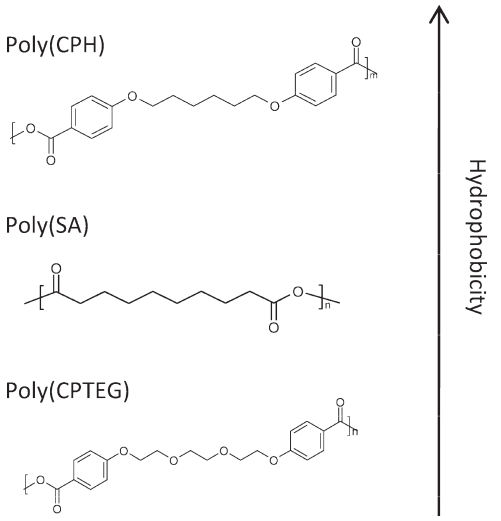


Figure 1. Chemical structure of CPH, SA, and CPTEG repeat units. The monomers vary in hydrophobicity that affects the controlled drug release. The polymers can be synthesized into combinatorial polymer film libraries and loaded with protein. Upon release, the protein kinetics can be quantified and modeled.

Zygourakiset al.,[12,13] and Batyckyet al.[14] developed models for polymers with homogeneous compositions. Larobina et al.[15] and Kipper and Narasimhan[16] developed models that account for microphase separation for 1,3-bis-(*p*-carboxyphenoxy)propane/SA (CPP/SA) and CPH/SA copolymers.

Similarly, some numerical modeling techniques have been proposed for modeling the release drug kinetics of polymers. Ibric et al.[17] used an artificial neural network (ANN) to model aspirin release from tablets as a function of concentration and compression pressure. Husseini et al.[18] used an ANN to predict acoustic release of doxorubicin from Pluronic P105 micelles. Takayama et al.[19] used an ANN to predict parameters in a developed release equation. The criteria for model selection varied from root mean squared error (RMSE), mean squared error (MSE), and Akaike information criteria (AIC). AIC is the best criterion because it takes both the model complexity and fitting performance into consideration, thereby avoiding an overfitting of the data.[20,21]

The objective of this work was to build a numerical modeling method combining nonlinear machine learning algorithms and evolutionary calculation algorithms for predicting release kinetics behavior. Building off of our previous work,[10] release kinetics of bovine serum albumin (BSA) protein from CPTEG/CPH and CPH/SA copolymer systems linearly varying in copolymer composition were studied by measuring the cumulative mass of protein released over time. Using the modeling method, several key structural descriptors of polymeric carriers that control the release profiles were identified. The approach developed here can provide guidance in selection of polymer molecular structures for predictable controlled release profiles from polymeric devices.

2. EXPERIMENTAL PROCEDURES

2.1. Materials. Monomer synthesis was carried out with the following chemicals: acetic acid, potassium carbonate, toluene, dimethyl formamide, sulfuric acid, and acetonitrile (purchased from Fisher Scientific (Fairlawn, NJ)); 1-methyl-2-pyrrolidinone, 4-*p*-hydroxybenzoic acid, 1,6-dibromohexane, and triethylene glycol (purchased from Sigma Aldrich (St. Louis, MO)); and 4-*p*-fluorobenzonitrile (obtained from Apollo Scientific (Cheshire, U.K.)). The chemicals utilized in polymerization, nanosphere fabrication, and buffer preparation include petroleum ether, pentane, acetic anhydride, chloroform, methylene chloride, dibasic potassium phosphate, and monobasic potassium phosphate (purchased from Fisher Scientific). BSA was purchased from Sigma Aldrich. The microbicinchoninic acid (BCA) protein assay kit was obtained from Pierce Biotechnology, Inc. (Rockford, IL).

Table 1. Number Average Molecular Weight and Polydispersity Index of Combinatorially Synthesized CPH/SA and CPTEG/CPH Copolymers Obtained by GPC

CPH/SA copolymer	M_n (Da)	polydispersity index
0:100	11 400	2.1
25:75	15 200	2.5
50:50	12 900	2.2
75:25	13 500	2.3
100:0	18 200	2.1
CPTEG/CPH Copolymer	**M_n (Da)**	**polydispersity index**
0:100	10 500	2.2
25:75	13 400	3.0
50:50	9600	2.3
75:25	9300	2.6
100:0	11 100	2.4

2.2. High-Throughput Fabrication of Combinatorial BSA-Loaded Polymer Film Libraries. CPTEG and CPH monomers were synthesized as described previously.[7,22] The SA monomer was purchased from Sigma Aldrich. The monomer units of SA, CPH, and CPTEG are shown in Figure 1. CPH/SA and CPTEG/CPH prepolymer libraries, linearly varying in molar composition, were deposited in a multiwell substrate using an automated robotic deposition apparatus. The polymer libraries were then synthesized in high throughput at the respective vacuum and temperature (0.3 Torr and 140 °C (CPTEG:CPH) and 180 °C (CPH:SA)) for 1.5 h as described previously.[10,23−25] Following synthesis of the combinatorial film library, BSA was uniformly dispersed in methylene chloride with sonication and the robotic deposition apparatus then utilized to dissolve the films in the methylene chloride with BSA. Next, the polymer/protein solution was sonicated for uniform dispersion and dried at room temperature under vacuum overnight resulting in 6% BSA loaded polymer film libraries. This high-throughput process for polymer synthesis and protein encapsulation into polyanhydride films was utilized from previously reported methods.[2,10] Four replicates of two different polymer film libraries were investigated for protein release kinetics and included the following CPH/SA chemistries: poly(SA), 25:75 CPH/SA, 50:50 CPH/SA, 75:25 CPH/SA, poly(CPH) and the following CPTEG/CPH chemistries: poly(CPTEG), 75:25 CPTEG/CPH, 50:50 CPTEG/CPH, 25:75 CPTEG/CPH, and poly(CPH).

2.3. Polymer Film Library Characterization. Polymer structure and molecular weight were determined by end group analysis with proton nuclear magnetic resonance spectroscopy (^{1}H NMR) for blank polymer film libraries using a Varian VXR 300 MHz spectrometer (Varian Inc., Palo Alto, CA) (Table 1). Samples were dissolved in deuterated chloroform and their chemical shifts assessed with respect to the chloroform peak

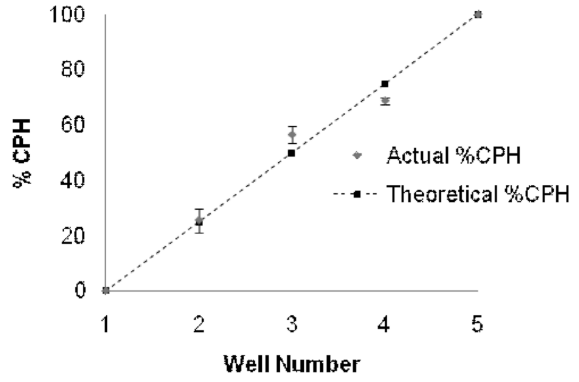


Figure 2. High-throughput FTIR analysis of a CPH/SA copolymer library (*y*-axis percent CPH).[25]

(δ = 7.26 ppm). The polymer molecular weight and polydispersity were measured using gel permeation chromatography (Table 1). A high-throughput evaluation of the surface chemistry (molar composition) and accuracy of the robotic deposition apparatus of the combinatorially synthesized CPH/SA polymer film libraries (no protein) was determined using Fourier transform infrared (FTIR) spectroscopy with a Nicolet 6700 FTIR spectrometer (Thermo Scientific) as described previously (Figure 2).[2,10,23,25] The polymer films libraries were synthesized on an IR transmissive silicon nitride multiwell substrate utilizing the robotic deposition apparatus. FTIR spectroscopy was conducted in an automated fashion with a total of 200 scans per spectrum at a resolution of 4 cm^{-1} by utilizing a programmable mapping software (Atlas). This allowed for specific sample and background maps to be preprogrammed into the sample detection setup, therefore allowing numerous FTIR spectra to be measured without user operation.

2.4. Protein Release from the Polymer Film Libraries. Following fabrication and characterization of the protein-loaded polymer film libraries, 1 mL of phosphate buffer saline (PBS) at pH 7.4 was added to each release sample well. The well plates were sealed to prevent evaporation and incubated in a horizontal shaker at 37 °C and 100 rpm for the duration of the experiment. Samples containing released BSA were taken at incremental time points throughout the study, which was terminated after one month. The BSA release samples were quantified using the micro-BCA assay and the protein release data presented as a cumulative mass fraction of protein released, in which the amount of protein released at each specific time point was normalized by the total amount of protein encapsulated into the films. All samples were run in triplicate, as described by the manufacturer (Pierce) and fresh buffer was added back to the sample well to maintain constant sink conditions. This entire procedure of protein release from polymer films was followed as described previously.[10] Figure 2 shows the cumulative mass fraction released from the polymer film libraries for the CPTEG:CPH and the CPH:SA systems. The polymer release kinetics demonstrated chemistry dependent trends decreasing in protein release with increasing polymer hydrophobicity (i.e., CPH content). This is concurrent with previous studies.[3,10,24]

3. RESULTS

Polymer libraries were synthesized at high throughput consisting of 10 different polymer chemistries with four replicates of each chemistry, resulting in a 40-sample library. The properties of polymers synthesized in these libraries were similar to those of conventionally synthesized polymers.[1,3,4,8,10,22−24] The polymer molecular weights were in the range of 9000−16 000 g/mol with a polydispersity ranging between 1.5 and 3.0 (Table 1) as reported previously. Additionally, ^{1}H NMR chemical peaks were located in the correct positions and were of appropriate relative areas, indicating that the copolymers were pure, and were of the desired copolymer compositions. High throughput FTIR analysis demonstrated that the robotic deposition apparatus was successful at achieving the intended molar ratios of each monomer in the polymer library (Figure 2).

Following polymer characterization, the release kinetics of libraries encapsulating BSA was monitored over one month. As observed in Figure 3, the release kinetics are chemistry dependent with the least hydrophobic polymers (i.e., poly(SA) and poly(CPTEG)) releasing the BSA the most rapidly. This is in agreement with previous conventional and combinatorial studies in which release kinetics of other proteins were investigated.[1,3,10,24]

4. MODELING

4.1. Building Molecular Structure Descriptor Library. To identify which aspects of the molecular structure had the most impact on the release kinetics, the components of the molecular structure must be fully described by a comprehensive library of molecular structure descriptors (Table 2). The library was built based partially on the descriptors defined by Bicerano.[26] On the basis of graph theory, Bicerano defined topological connectivity indices and used them to predict more than 70 polymer properties. As numerous polymer properties can be calculated as a function of these descriptors, it follows that kinetics behavior may also be possible to model, although more complex mathematics are required.

The descriptors defined by Bicerano are as follows. The first four descriptors in Table 2 describe the electronic structure of each non-H atom (for example, the valence shell hybridization, inner-shell electrons, and lone pairs along vertex set) and electronic structure of each bonded set of atoms (including σ and π electrons for molecular graph edge set). BB_index1, BB_index2, N_K, and SG_index, which depict the nature of the connectivity and conformations of the chain backbone and the relative size of the side group portion of the hydrogen-suppressed graph of the repeat unit are defined and used as fitting variables for the steric hindrance parameter. The definitions of other structural descriptors are defined in Table 2.

4.2. Attribute Selection Approach. Genetic algorithm (GA) and support vector regression (SVR) are used to determine which of the structural descriptors has the most impact on the release behavior. The overall logic of the approach is presented in Figure 4. Attribute selection is a nondeterministic polynomial time problem, meaning that a computer cannot solve the problem in polynomial time. To counter this problem, GA was applied because it can exploit the most promising solutions without performing an exhaustive search, thereby overcoming issues with other widely used heuristic algorithms such as sequential forward search and sequential backward search

4.2.1. Genetic Algorithm. GAs are a class of heuristic optimization algorithms that are inspired by evolutionary biology. GAs have been applied to multiple applications, including drug design,[27−31] job scheduling,[32] robot behavior,[33,34] and decision making.[35−37] GAs find a best set of values of parameters that maximize (or minimize) a function by generating a population of

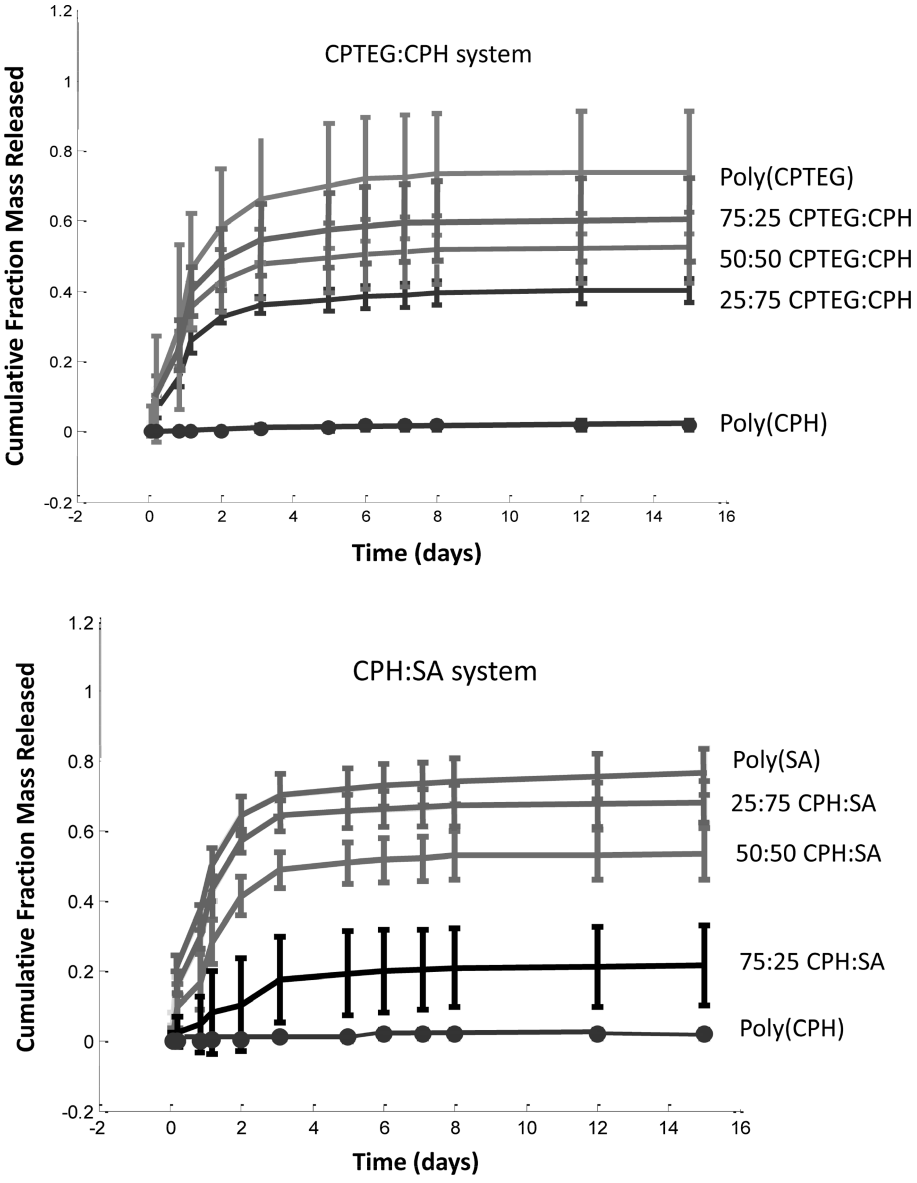


Figure 3. Cumulative mass fraction of released BSA from polymer film libraries: (top) CPTEG/CPH system and (bottom) CPH/SA system. The polymer release kinetics demonstrated chemistry dependent trends of decreasing protein release with increasing polymer hydrophobicity (CPH content).

solutions at one iteration (generation), and then calculating the value of the goal function (fitness) for each solution. Following the logic of evolutionary biology, the solutions with highest fitness pass into the next generation, with crossover (exchange of two solutions) and mutation (alteration of some values) occurring. The solutions with highest fitness functions are most likely to survive, with the best traits of the solutions maintained via crossover (e.g., offspring inheriting best traits of parents), while mutation can serve to avoid being caught in a local minima/maxima. Crossover points are generally randomly selected, with two individuals exchanging some characteristic. The newly generated two individuals are then passed into the next generation. For the individuals selected for mutation, parts of the individuals will be mutated into other values. If the individuals are binary-encoded, zero will be mutated into one, and vice versa. If the individuals are real-number encoded, a user defined mutation method may be applied.

GA randomly generates a population of solutions, which can be expressed as a matrix X_{mn}. M is the population size, which means in each generation (each iteration of GA) m solutions are generated. N is the number of parameters to be assessed in the goal function. Each row of X is called an individual, which is in fact a set of initial values of $\{X_1, X_2, ..., X_n\}$. Each row of matrix X is substituted into the function to be optimized $[y = f(X_1, X_2, ..., X_n)]$ and to calculate individual fitness y_i (i = 1, 2, ..., m). The individuals with highest fitness y_i pass directly into the next generation, with the increased likelihood that an individual is the true solution with increasing fitness.

4.2.2. Support Vector Regression. SVR is an extension of SVM (support vector machine) and is a kernel-based supervised learning algorithm. SVM and SVR are widely used in bioinformatics,[38−40] image processing,[41−43] and control systems.[44] SVMs are a general class of supervised learning methods that can perform classification or regression through mapping data into a

dx.doi.org/10.1021/co100019d |*ACS Comb. Sci.* 2011, 13, 50–58

Table 2. Twenty-Five Molecular Descriptors for the Repeat Unit of the Copolymers Used in This Study[a]

ID	structure descriptor	descriptions
1	${}^0X$	atomic indices
2	${}^0X^v$	atomic indices
3	${}^1X$	connectivity indices
4	${}^1X^v$	connectivity indices
5	N	total non-hydrogen atom number in one repeat unit
6	BB_index1	backbone index1 (12.19)
7	N_SP	number of atoms in the shortest path across the backbone of a polymeric repeat unit, N_SP ≤ N_BB (2.11)
8	N_C	number of carbon atoms in a polymeric repeat unit
9	N_H	number of hydrogen atoms in a polymeric repeat unit
10	N_ester_n	number of backbone −COO− (nonconjugated)
11	N_ester_c	number of backbone −COO− (one-sided conjugation with aromatic ring)
12	N_aromaticRing	number of aromatic rings in a polymeric repeat unit
13	N_CH2	number of CH_2 in a polymeric repeat unit
14	N_ether	numer of −O− in a polymeric repeat unit
15	Nmv	Nmv = 2*N_ester + 3*N_ether (3.14)
16	N_K	N_K = −3*N_ether_quote-3*N_acrylic_ester (12.27)
17	N_rot	the total number of rotational degrees of freedom parameter
18	N_backbone_O	number of backbone oxygen atoms in a polymeric repeat unit
19	N_dc	N_dc = 7*N_backbone_O + 12*N_sideGroup_O (9.11)
20	N_O	number of oxygen atoms in a polymeric repeat unit
21	M	mole weight of the repeat unit
22	BB_index2	backbone index 2 (12.20)
23	SG_index	backbone index 4 (12.22)
24	N_alkyl_ether	the number of ether (R−O−R′) linkages between two units R and R′ both of which are connected to the alkyl carbon atom
25	N_group	N_group = −N_alkyl_ether + 7*N_CO + 2*N_otherCO (5.10)

[a] The equation numbers refer to the corresponding equations in Bicerno.[24] These 25 descriptors were calculated for each composition and organized in a data set so as to be compared with the release profiles.

higher dimension kernel space. On the basis of structural risk minimization, a separating hyperplane is obtained, which maximizes the separation between the classes. The SVM is optimized for structural risk minimization by minimizing the generalization error, as opposed to the empirical error as done with other methods, such as neural networks. SVM is a quadratic problem but with linear constraints, while neural networks involve a nonconvex and unconstrained minimization problem. When used for classification the maximum distance between separated hyperplanes is sought, while when used for regression, the generalization bounds of the hyperplanes are optimized. With the use of kernels, the data is mapped onto a higher dimensional space to account for nonlinear data.

4.2.3. GA/SVR Hybrid Attribute Selection Algorithm. SVR can be embedded within GA to provide a fitness value. The process operates by GA randomly generating a population of individuals, with each individual a subset of structural descriptors. Then for each individual, one SVR model is built based on the selected structural descriptors. A polynomial kernel $K(x,y) = \langle x,y\rangle^p$ is used and the exponent p is optimized by minimizing AIC. AIC is a measure of the fitness of a regression model.

$$\mathrm{AIC} = 2N_w + N\ln(\mathrm{RSS}/N) \quad (1)$$

Here N_w is the total number of weights/support vectors, N is the total number of data points, and RSS is the residual sum of squares. Individuals (subsets of attributes) are ranked based on the AIC value, with a lower AIC value desired. The GA operations of selection, crossover, and mutation are applied, with the best individuals as ranked by AIC maintained and new individuals generated.

5. PARAMETERIZING NONLINEAR RELEASE KINETICS PROFILES

As shown in Figure 4, the logic of this work is to connect the descriptors of the molecular structure with the release kinetics profiles of different combinatorially fabricated polyanhydride films, thus allowing the identification of the aspects of the molecular structure that primarily dictate the release kinetics behavior. However, the release profiles are nonlinear, and SVR fails to build a good model that addresses the nonlinearity between the transition phase and the steady state phase. The challenge then becomes to create a parametrization of these nonlinear profiles that still reflects all of the information in the release curve.

As a first step to parametrize the profiles, exponential functions were developed to fit the profiles, based on the observation that exponential functions are widely used in physics to explain processes that are related with time. The following two equations were found to represent all of the measured profiles.

$$\text{cumulative fraction BSA released} = \exp(A + B/t) \quad (2)$$

$$\text{cumulative fraction BSA released} = \exp\left(A + \frac{B_1}{t} + \frac{B_2}{t^2}\right) \quad (3)$$

On the basis of these two equations, five parameters were identified which can be used to represent the nonlinear profiles

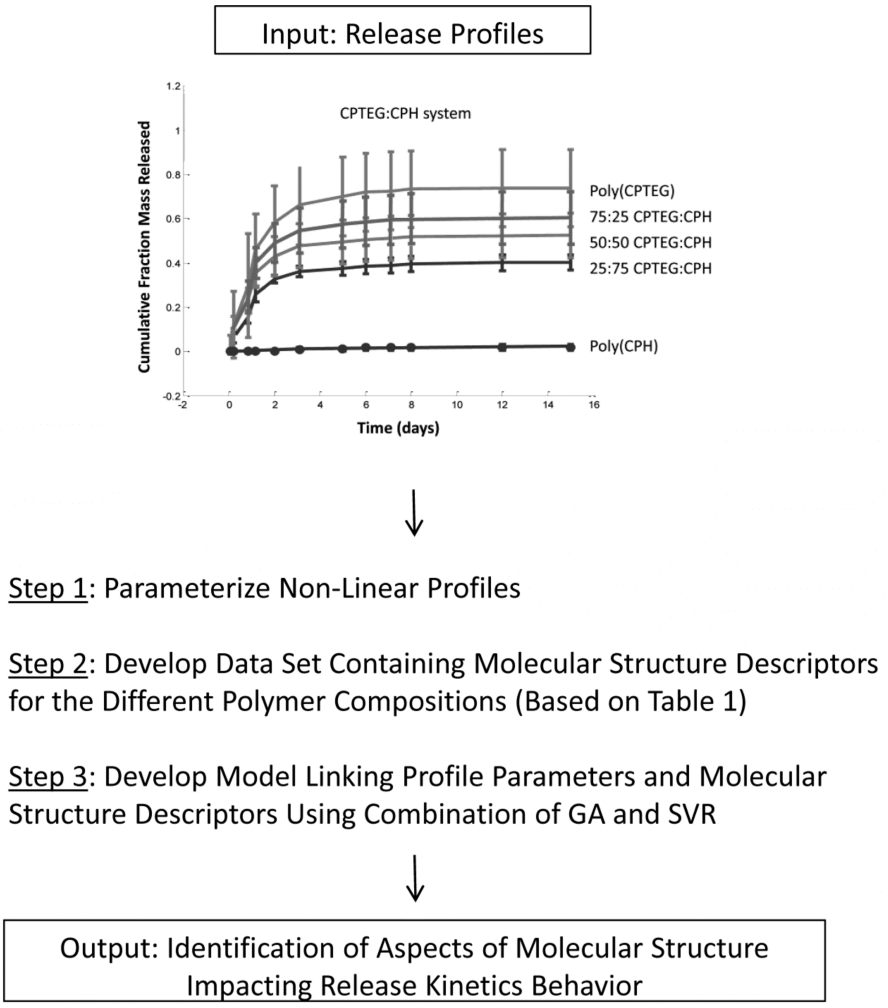


Figure 4. Outline of the method proposed in this study. To more accurately model the release profiles based on the molecular structure descriptors, the nonlinear profiles should be parametrized. Then using the newly developed mathematical model for attribute selection, a prediction model linking the release profile and the molecular structures can be built. This model can then be used to identify which aspects of the molecular structure most impact the release kinetics behavior.

(A and B from eq 2 and A, B_1, and B_2 from eq 3). The values of these parameters for the polyanhydride systems were calculated. An example of the modeled profiles and parameters for the CPTEG/CPH system is shown in Figure 5. Using these two models, all the nonlinear profiles were successfully parametrized.

6. MOLECULAR STRUCTURE IMPACT ON RELEASE KINETICS BEHAVIOR

In section 2, the release kinetics behavior of polyanhydride films was discussed, and in section 5, a model for parametrizing the nonlinear release profiles was developed. In section 4.1, a descriptor library containing information on the molecular structure was established, and in section 4.2, a mathematical approach based on data mining logics for linking the parameters of the release profiles with the molecular structure descriptors was designed. In this section, findings are presented on the basis of the combination of the results from these previous sections.

When the GA/SVR algorithm is carried out to a limited number of generations, multiple results are possible because the initial population of GA is randomly generated. To address this issue, eight different GA/SVR algorithms were run, and the results across all of these models were assessed. The GA/SVR algorithm was carried out to 100 generations and with a population size of 20. The poly kernel exponent p was chosen to optimize the value of AIC. The result provided from the GA/SVR analysis is a numerical model predicting the release profile as a function of the molecular structure descriptors. For example, the most accurate model identified number of carbon atoms in a polymeric repeat unit, number of backbone −COO− nonconjugated bonds, number of aromatic rings in a polymeric repeat unit, number of CH_2 bonds in a polymeric repeat unit, and molar weight of the repeat unit as being sufficient to model the release kinetics profiles, based on the definitions of (A, B, B_1, B_2).

The accuracy of the model for the release kinetics behavior is inversely related to the AIC value (i.e., AIC should be minimized), and on the basis of this criterion, three models were highly accurate. In each of these three models the number of nonconjugated −COO_- backbone units (descriptor 10 in Table 2), the number of aromatic rings in the polymeric repeat unit (descriptor 12 in Table 2), and the number of CH_2 bonds in the polymeric repeat unit (descriptor 13 in Table 2) were used to model the release profile. Additionally, if the number of times each descriptor is used in one of the eight models is counted

dx.doi.org/10.1021/co100019d |ACS Comb. Sci. 2011, 13, 50–58

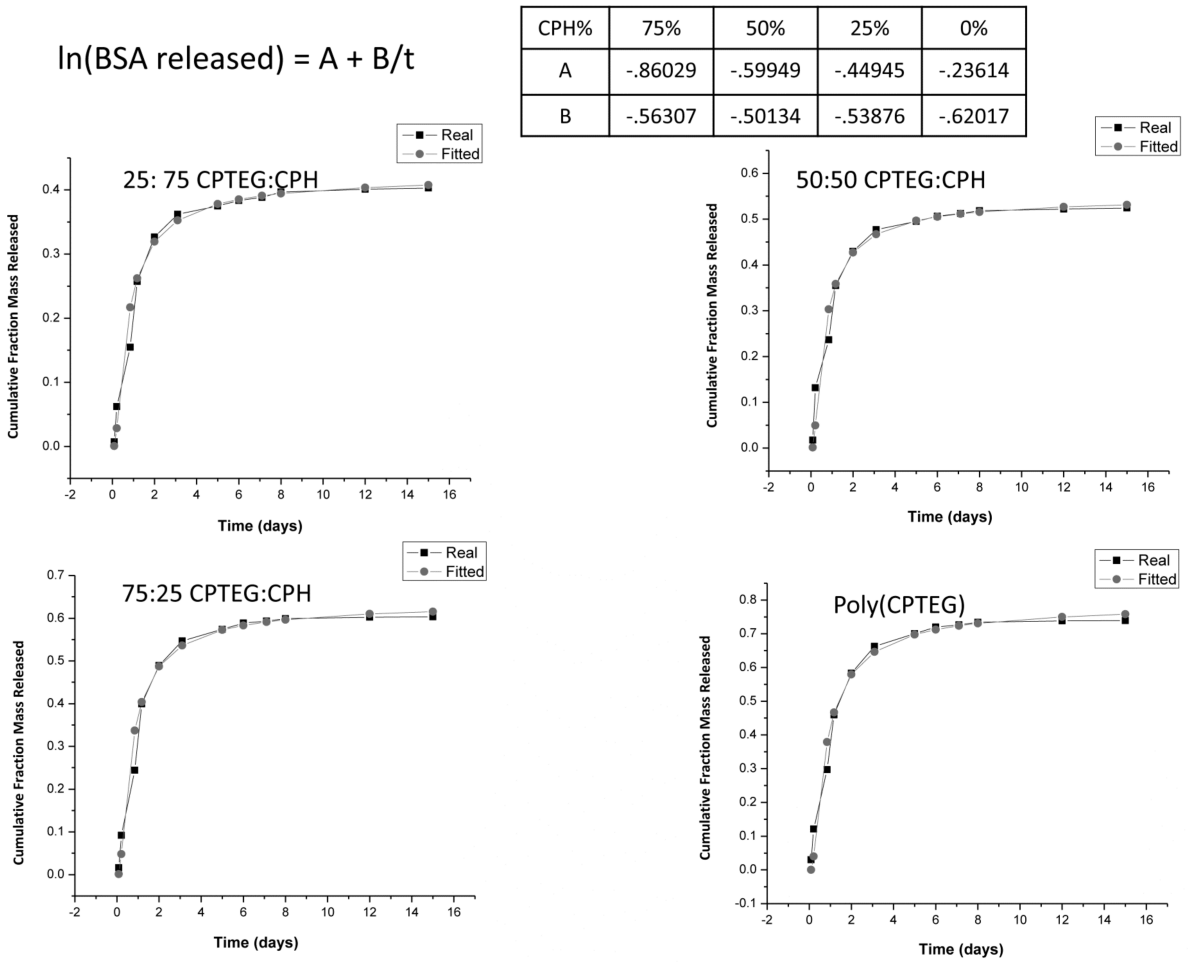


Figure 5. Fitting results for the CPTEG/CPH system using ln(BSA released) = $A + B/t$ and the values found for two of the possible parameters (A and B from eq 2).

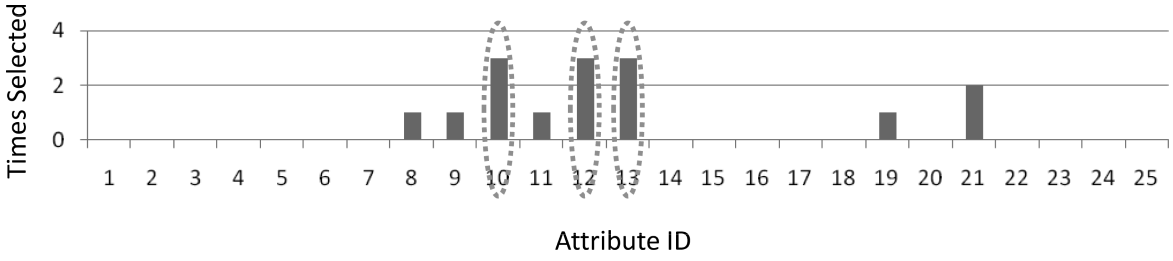


Figure 6. Number of times each molecular structure descriptor was used to model the release kinetics. The ID corresponds with the labels of Table 2. The attributes that are found to be most significant in modeling the release kinetics of the SA and CPTEG systems are nonconjugated −COO− bonds, aromatic rings, and $-CH_2-$ bonds (encircled by the dotted ovals).

(Figure 6), nonconjugated −COO− backbone bonds, aromatic rings, and CH_2 bonds were again found to be the most relevant for modeling the release kinetics. Therefore, on the basis of the criteria of either accuracy of model or the number of times the descriptor is used for any of the models, these three molecular structure descriptors were found to be the most relevant.

7. CONCLUSIONS

In this paper, a new approach was presented for determining which aspects of the polymeric molecular structure have the most impact on the release kinetics behavior and to use these molecular descriptors to model the release profile of proteins from a combinatorial library of biodegradable polymers. The mathematical approach for modeling the release profile was based on a combination of GA and SVR. The polymers explored from the combinatorial libraries included linearly varying molar compositions of CPTEG/CPH and CPH/SA copolymers, with the protein release rate found to increase with decreasing polymer hydrophobicity. From the new attribute selection approach, it was determined that number of nonconjugated backbone −COO− bonds, number of aromatic rings, and

dx.doi.org/10.1021/co100019d |ACS Comb. Sci. 2011, 13, 50–58

number of $-CH_2-$ bonds were most important for accurately modeling the release kinetics behavior. This work introduces a new approach for understanding which factors control drug release and has important implications for the rational design of new polymeric carriers for drug delivery.

■ ASSOCIATED CONTENT

Supporting Information. Parametrization of the CPH/SA system, additional details on the hybrid GA/SVR approach used in this paper, and the descriptors selected for each model. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*Postal address: 2240 Hoover Hall, Iowa State University, Ames, IA 50011. E-mail: krajan@iastate.edu. Phone: 515-294-2670. Fax: 515-294-5444.

■ ACKNOWLEDGMENT

The authors acknowledge financial support from the Office of Naval Research MURI program for Novel Vaccines: Targeting and Exploiting the Bacterial Quorum Sensing Pathway, award number N00014-06-1-1176, and the National Science Foundation, under award number CCF-AF:0917202 (K.R.), EEC 0552584, and 0851519 (B.N.).

■ REFERENCES

(1) Lopac, S. K.; Torres, M. P.; Wilson-Welder, J. H.; Wannemuehler, M. J.; Narasimhan, B. Effect of polymer chemistry and fabrication method on protein release and stability from polyanhydride microspheres. *J. Biomed. Mater. Res. B* **2009,** *12B*, 938–947.

(2) Petersen, L. K.; Sackett, C. K.; Narasimhan, B. High throughput analysis of protein stability in polyanhydride nanoparticles. *Acta Biomater.* **2010,** *6*, 3873–3881.

(3) Torres, M. P.; Determan, A. S.; Anderson, G. L.; Mallapragada, S. K.; Narasimhan, B. Amphiphilic polyanhydrides for protein stabilization and release. *Biomaterials* **2007,** *28*, 108–116.

(4) Carrillo-Conde, B.; Schiltz, E.; Torres, M. P.; Yu, J.; Phillips, G.; Minion, C.; Wannemuehler, M. J. Encapsulation into amphiphilic polyanhydride microparticles stabilizes *Yersinia pesis* antigens. *Acta Biomater.* **2010,** *6*, 3110–3119.

(5) Katti, D. S.; Lakshmi, S.; Langer, R.; Laurencin, C. T. Toxicity, biodegradation, and elimination of polyanhydrides. *Adv. Drug Delivery Rev.* **2002,** *54*, 933–61.

(6) Rosen, H. B.; Chang, J.; Wnek, G. E.; Linhardt, R. J.; Langer, R. Bioerodible polyanhydrides for controlled drug delivery. *Biomaterials* **1983,** *4*, 131–3.

(7) 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. A* **2006,** 76102–10.

(8) 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. Controlled Release* **2002,** *82*, 115–25.

(9) Zhang, M.; Yang, Z.; Chow, L. L.; Wang, C. H. Simulation of drug release from biodegradable polymeric microspheres with bulk and surface erosions. *J. Pharm. Sci.* **2003,** *92*, 2040–56.

(10) Petersen, L. K.; Sackett, C. K.; Narasimhan, B. A novel high-throughput screening method to study the effect of device geometry, polymer chemistry and pH on in vitro protein release from polyanhydrides. *J. Comb. Chem.* **2010,** *12*, 51–56.

(11) Thombre, A. G.; Himmelstein, K. J. Modelling of drug release kinetics from a laminated device having an erodible drug reservoir. *Biomaterials* **1984,** *5*, 250–4.

(12) Zygourakis, K. Development and temporal evolution of erosion fronts in bioerodible controlled release devices. *Chem. Eng. Sci.* **1990,** *45*, 2359–2366.

(13) Zygourakis, K.; Markenscoff, P. A. Computer-aided design of bioerodible devices with optimal release characteristics: A cellular automata approach. *Biomaterials* **1996,** *17*, 125–35.

(14) Batycky, R. P.; Hanes, J.; Langer, R.; Edwards, D. A. A theoretical model of erosion and macromolecular drug release from biodegrading microspheres. *J. Pharm. Sci.* **1997,** *86*, 1464–77.

(15) Larobina, D.; Mensitieri, G.; Kipper, M. J.; Narasimhan, B. Mechanistic understanding of degradation in bioerodible polymers for drug delivery. *Bioeng., Food, Nat. Prod.* **2004,** *48*, 2960–2970.

(16) Kipper, M. J.; Narasimhan, B. Molecular description of erosion phenomena in biodegradable polymers. *Macromolecules* **2005,** *38*, 1989–1999.

(17) Ibrić, S.; Jovanović, M.; Djurić, Z.; Parojčić, J.; Petrović, S.; Solomun, L.; Stupar, B. Artificial neural networks in the modeling and optimization of aspirin extended release tablets with eudragit L 100 as matrix substance. *AAPS PharmSciTech* **2003,** *4*, 62–70.

(18) Husseini, G. A.; Abdel-Jabbar, N. M.; Mjalli, F. S.; Pitt, W. G. Modeling and sensitivity analysis of acoustic release of Doxorubicin from unstabilized pluronic P105 using an artificial neural network model. *Technol. Cancer Res. Treat* **2007,** *6*, 49–56.

(19) Takayama, K.; Fujikawa, M.; Obata, Y.; Morishita, M. Neural network based optimization of drug formulations. *Adv. Drug Delivery Rev.* **2003,** *55*, 1217–31.

(20) Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974,** *19*, 716–723.

(21) Stoica, P.; Selen, Y. Model-order selection: a review of information criterion rules. *IEE Signal Process. Mag.* **2004,** *21*, 36–47.

(22) Vogel, B. M.; Mallapragada, S. K. Synthesis of novel biodegradable polyanhydrides containing aromatic and glycol functionality for tailoring of hydophilicity in controlled drug delivery devices. *Biomaterials* **2005,** *26*, 721–728.

(23) Adler, A. F.; Petersen, L. K.; Wilson, J. H.; Torres, M. P.; Thorstenson, J. B.; Gardner, S. W.; Mallapragada, S. K. High throughput cell-based screening of biodegradable polyanhydride libraries. *Comb. Chem. High Throughput Screening* **2009,** *12*, 634–645.

(24) Determan, A. S.; Trewyn, B. G.; Lin, V. S.; Nilsen-Hamilton, M.; Narasimhan, B. Encapsulation, stabilization, and release of BSA-FITC from polyanhydride microspheres. *J. Controlled Release* **2004,** *100*, 97–109.

(25) Petersen, L. K.; Xue, L.; Wannemuehler, M. J.; Rajan, K.; Narasimhan, B. The simultaneous effect of polymer chemistry and device geometry on the in vitro activation of murine dendritic cells. *Biomaterials* **2009,** *30*, 5131–5142.

(26) Bicerano, J. *Prediction Of Polymer Properties*, 2nd ed.; Marcel Dekker, Inc.: 1996.

(27) Abu Hammad, A. M.; Taha, M. O. Pharmacophore modeling, quantitative structure−activity relationship analysis, and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors. *J. Chem. Inf. Model.* **2009,** *49*, 978–96.

(28) Corbeil, C. R.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J. Chem. Inf. Model.* **2009,** *49*, 997–1009.

(29) Kang, L.; Li, H.; Jiang, H.; Wang, X. An improved adaptive genetic algorithm for protein−ligand docking. *J. Comput.-Aided Mol. Des.* **2009,** *23*, 1–12.

(30) Xi, L.; Du, J.; Li, S.; Li, J.; Liu, H.; Yao, X. A combined molecular modeling study on gelatinases and their potent inhibitors. *J. Comput. Chem.* **2009,** *31*, 24–42.

(31) Zaheer-ul, H.; Uddin, R.; Yuan, H.; Petukhov, P. A.; Choudhary, M. I.; Madura, J. D. Receptor-based modeling and 3D-QSAR for a quantitative production of the butyrylcholinesterase inhibitors based on genetic algorithm. *J. Chem. Inf. Model.* **2008,** *48*, 1092–103.

(32) Li, B. B.; Wang, L. A hybrid quantum-inspired genetic algorithm for multiobjective flow shop scheduling. *IEEE Trans. Syst. Man Cybern., Part B: Cybern.* **2007,** *37*, 576–91.
(33) Alnajjar, F.; Murase, K. Self-organization of spiking neural network that generates autonomous behavior in a real mobile robot. *Int. J. Neural Syst.* **2006,** *16*, 229–39.
(34) Zhou, Y.; Er, M. J. An evolutionary approach toward dynamic self-generated fuzzy inference systems. *IEEE Trans. Syst. Man Cybern., B: Cybern.* **2008,** *38*, 963–9.
(35) Mazurowski, M. A.; Habas, P. A.; Zurada, J. M.; Tourassi, G. D. Decision optimization of case-based computer-aided decision systems using genetic algorithms with application to mammography. *Phys. Med. Biol.* **2008,** *53*, 895–908.
(36) Mazurowski, M. A.; Zurada, J. M.; Tourassi, G. D. Selection of examples in case-based computer-aided decision systems. *Phys. Med. Biol.* **2008,** *53*, 6079–96.
(37) van der Lee, J. H.; Svrcek, W. Y.; Young, B. R. A tuning algorithm for model predictive controllers based on genetic algorithms and fuzzy decision making. *ISA Trans.* **2008,** *47*, 53–59.
(38) Xu, Q.; Hu, D. H.; Xue, H.; Yu, W.; Yang, Q. Semi-supervised protein subcellular localization. *BMC Bioinf.* **2009,** *10* (Suppl 1), S47.
(39) Zervakis, M.; Blazadonakis, M. E.; Tsiliki, G.; Danilatou, V.; Tsiknakis, M.; Kafetzopoulos, D. Outcome prediction based on microarray analysis: a critical perspective on methods. *BMC Bioinf.* **2009,** *10*, 53.
(40) Zou, A. M.; Wu, F. X.; Ding, J. R.; Poirier, G. G. Quality assessment of tandem mass spectra using support vector machine (SVM). *BMC Bioinf.* **2009,** *10* (Suppl 1), S49.
(41) Kakar, M.; Olsen, D. R. Automatic segmentation and recognition of lungs and lesion from CT scans of thorax. *Comput. Med. Imaging Graph.* **2009,** *33*, 72–82.
(42) Lin, T.; Cervino, L. I.; Tang, X.; Vasconcelos, N.; Jiang, S. B. Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy. *Phys. Med. Biol.* **2009,** *54*, 981–92.
(43) Masotti, M.; Lanconelli, N.; Campanini, R. Computer-aided mass detection in mammography: False positive reduction via gray-scale invariant ranklet texture features. *Med. Phys.* **2009,** *36*, 311–6.
(44) Yuan, X. F.; Wang, Y. N.; Wu, L. H. Adaptive Inverse Control of Excitation System with Actuator Uncertainty. *Neural Proc. Lett.* **2007,** *27*, 125–136.