Published in final edited form as:

Biomaterials. 2014 February ; 35(6): 1977–1988. doi:10.1016/j.biomaterials.2013.10.069.

Discovery of Antibiotics-derived Polymers for Gene Delivery using Combinatorial Synthesis and Cheminformatics Modeling

Thrimoorthy Potta^{1,#}, Zhuo Zhen^{2,#}, Taraka Sai Pavan Grandhi³, Matthew D. Christensen¹, James Ramos³, Curt M. Breneman², and Kaushal Rege^{1,3,*}

¹Chemical Engineering, Arizona State University, Tempe, AZ 85287-6106 USA

²Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, NY 12180 USA

³Harrington Biomedical Engineering, Arizona State University, Tempe, AZ 85287 USA

Abstract

We describe the combinatorial synthesis and cheminformatics modeling of aminoglycoside antibiotics-derived polymers for transgene delivery and expression. Fifty-six polymers were synthesized by polymerizing aminoglycosides with diglycidyl ether cross-linkers. Parallel screening resulted in identification of several lead polymers that resulted in high transgene expression levels in cells. The role of polymer physicochemical properties in determining efficacy of transgene expression was investigated using Quantitative Structure-Activity Relationship (QSAR) cheminformatics models based on Support Vector Regression (SVR) and ‘building block’ polymer structures. The QSAR model exhibited high predictive ability, and investigation of descriptors in the model, using molecular visualization and correlation plots, indicated that physicochemical attributes related to both, aminoglycosides and diglycidyl ethers facilitated transgene expression. This work synergistically combines combinatorial synthesis and parallel screening with cheminformatics-based QSAR models for discovery and physicochemical elucidation of effective antibiotics-derived polymers for transgene delivery in medicine and biotechnology.

© 2013 Elsevier Ltd. All rights reserved.

*To Whom All Correspondence is to be addressed. Kaushal Rege, Chemical Engineering, 501 E. Tyler Mall, ECG 301, Arizona State University, Tempe, AZ 85287-6106, kaushal.rege@asu.edu, Phone: 480-727-8616, Fax: 480-727-9321.

#Both authors contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

AUTHOR CONTRIBUTIONS

KR, CMB, TP, and ZZ designed research, TP, ZZ, TSPG, MC, and JR performed research, all authors contributed to data analysis, and TP, ZZ, CMB and KR wrote the manuscript.

SUPPORTING INFORMATION. Details on polymer synthesis and characterization, cell culture, and QSAR modeling methods are available in the Supporting Information section.

Keywords

Gene Delivery; Rational Design; Quantitative Structure-Activity Relationship (QSAR) models; Aminoglycosides

1. INTRODUCTION

Effective administration of transgenes for expression of exogenous proteins in target cells has diverse applications in biotechnology and medicine[1–4]. The success of this approach depends on the biocompatibility and efficacy of the delivery vector employed for administering the genetic material into cells. Polymers have received increasing attention for transgene delivery, primarily due to concerns relating to safety, immunogenicity, and potentially high production costs associated with viral vectors[5–9]. Poly(ethyleneimine) or pEI[10, 11], poly- β -amino esters[12–14], poly(vinyl pyrrolidone)[15], dendrimers[16, 17], and cyclodextrin-based polymers[18, 19] have been extensively investigated for transgene delivery.

Despite the promise, discovery of effective polymers for transgene delivery and expression remains largely empirical. Although some rules of thumb (e.g. availability of cationic charges) exist for polymer design, the absence of a fundamental framework that can enable correlation of polymer physicochemical properties with transgene expression efficacies is a critical void, which limits effective and rational discovery. Quantitative Structure-Activity Relationship (QSAR) modeling is a powerful cheminformatics approach for correlating molecular efficacies with their structural attributes (physicochemical properties or descriptors). This approach is popular in drug discovery, particularly for optimizing lead candidates[20–24], and has been adopted for several applications in biotechnology[25–30].

Combinatorial / parallel syntheses are commonly employed for generating libraries of small molecules as drug candidates[31, 32], and have been extended to facilitate the rapid discovery of polymers for transgene delivery and expression[33, 34]. In this work, we investigated aminoglycoside antibiotics as monomers in the parallel synthesis of polymers for transgene delivery to mammalian cells. Parallel screening of combinatorial libraries results in large amounts of information on the efficacy of polymer-mediated transgene expression. We therefore employed Support Vector Machine (SVM) regression-based Quantitative Structure-Activity Relationship (QSAR) modeling in order to correlate information on transgene expression efficacy with physicochemical properties (descriptors), determined from ‘building-block’ approximations of polymer structures. The predictive ability of the model was investigated using polymers not included in training the QSAR model, and molecular descriptors selected in the QSAR model were employed to gain insights into physicochemical factors influencing polymer-mediated transgene expression.

2. EXPERIMENTAL

2.1. Synthesis of the Aminoglycoside-based Polymer Library

A library of 56 aminoglycoside-based polymers was synthesized using ringopening polymerization between amines of seven aminoglycosides and epoxides of eight diglycidyl

ethers as shown in Scheme S1; structures of aminoglycosides and diglycidyl ethers are shown in Figure S1 and Figure S2 (Supporting Information section), respectively. Prior to polymerization, aminoglycosides were converted to their respective amine hydrochloride forms by incubating with Amberlite[®] anion exchange resin in order to remove associated sulfates[35]. Sulfate-free aminoglycosides (chloride form) were reacted with diglycidyl ethers in varying mole fractions (1:1 – 1:3) molar ratios in a mixture of water and N,N-dimethylformamide (DMF) (1.5:1) for 5 hours at 60°C. A ratio of 1:1 aminoglycoside:diglycidyl ethers was employed only in the case of amikacin, since a 1:2 ratio resulted in the formation of insoluble products. The crude reaction mixture was allowed to cool to room temperature and precipitated using acetone. The precipitated product was washed twice with acetone in order to remove unreacted diglycidyl ethers and dried. The product was further purified by dialysis using a 3500 molecular weight cutoff (MWCO) membrane to remove unreacted aminoglycosides. The dialyzed material was lyophilized to obtain the polymer product.

2.2. Parallel screening of polymer-mediated transgene (luciferase) expression efficacy

The aminoglycoside-based polymer library was screened in parallel for delivering the pGL3 control vector (Promega Corp., Madison, WI), which encodes for the modified firefly luciferase protein, to different cancer cell lines. The pGL3 plasmid DNA (pDNA) was prepared as described previously[36]; plasmid concentration and purity were determined using a NanoDrop Spectrophotometer (ND-1000; NanoDrop Technologies) by measuring absorbance at 260 and 280 nm.

Cells were allowed to attach to the bottom of wells for 16 h after which, fresh serum-free RPMI-1640 media or serum-containing media was added. For the primary screen, pDNA:polymer complexes (polyplexes), prepared at a polymer to plasmid weight ratio of 25:1, were added for 6h; 50 and 75 ng/well pGL3 pDNA were added to MB49 cells, while 200 ng/well were used for all other cell lines. Polymer:pDNA nitrogen:phosphate (N:P) ratios were employed for comparing leads across different doses. After 48 h of further incubation in serum-containing media, cells were lysed, and luciferase protein expression was determined as relative luminescence units (RLU) using the Bright Glo[™] Luciferase assay kit (Promega) with a plate reader (Bio-Tek Synergy 2). Cell lysates were then assayed for total protein content using the BCA Protein Assay kit (Pierce, Rockford, IL, USA). RLU values were normalized by the protein content to yield 'RLU/mg protein' values that were employed for comparing different polymers. Untransfected cells and cells transfected with uncomplexed pDNA were used as controls. Luciferase expression efficacies of polymers from the library were compared to those of 25kDa branched pEI. The pEI solution was prepared fresh right before all transfection experiments, and fresh stocks of commercially available pEI were employed.

2.3. Polymer-mediated EGFP transgene expression and flow cytometry

EGFP (enhanced green fluorescent protein)-encoding pDNA (pEGFP) were delivered as above. pEGFP was prepared and quantified as above, except that 1 mg/mL kanamycin (Research Products International, Corp.) was used in place of 1 mg/mL ampicillin (Research Products International, Corp.) because of the corresponding resistance in the plasmids.

Following transfection, the media was collected into 1.5 mL microcentrifuge tubes (8000 rpm for 10 minutes), each well was washed with 150 μ L, 1 $\times$ PBS, trypsinized with 150 μ L, 0.25 % trypsin-EDTA, and collected. The supernatant was removed and the cell pellet was resuspended in 50 μ L of 1 $\times$ PBS. The percentage of cells exhibiting fluorescence, as well as mean fluorescence of cells in the population, was determined using flow cytometry (Becton Dickinson FACScalibur Flow Cytometer using the FL1 channel); untreated cells were used as controls.

2.4. Statistical Analyses

All cell-based experiments were carried out in triplicate, and the results are expressed as mean $\pm$ one standard deviation. The Student's t-test was used to assess statistical significance of difference between group means; p -values <0.05 , with respect to pEI, are considered statistically significant.

2.5. QSAR Models of Polymer-Mediated Transgene Expression

2.5.1. Selection of Polymer Training and External Test Sets

Selection of the Polymer Set: The $\log_{10}$ (RLU/mg protein) values for all 56 polymers in the library were determined and divided into seven bins or sub-ranges (Table S1, Supporting Information section); $\log_{10}$ (RLU/mg protein) values were employed for QSAR modeling since luciferase expression values (RLU/mg) ranged across four orders of magnitude. The ratio of standard deviation (STD) and (RLU/mg protein) values were calculated for each polymer, and the ones with ratios greater than 1.0 were eliminated. If the bin contained > 6 polymers, only five with the lowest STD/(RLU/mg protein) ratio were picked for model generation.

The logarithm values with base 10 of the luciferase expression i.e. $\log_{10}$ (RLU/mg protein) were used as the response values in the QSAR model. To ensure that both, training and the external test sets were evenly distributed in response, three polymers were selected from bins 3, 5, and 7 (Table S1, Supporting Information section) (one polymer in each bin) as an external test set. The final training set for model construction (30 polymers) and the external test set (3 polymers) for model evaluation are shown in Table S1 (Supporting Information section).

2.5.2 Physicochemical Molecular Descriptors and Feature Selection—Molecular structures of aminoglycosides and diglycidyl ethers, energy minimization methods, and description of 'building block' approximations of the polymers are provided in the Supporting Information section (Part B, Table S2 and Figures S1–S3). The building block structures of our polymers are constructed with only primary amines of the aminoglycoside core reacting with epoxide moieties in the diglycidyl ethers. This is an excellent approximation given that the ratio of kinetic rate constant for primary amine reactions with epoxide moieties is 2-fold over that for secondary amines[37]. The second epoxide group on a diglycidyl linker was also opened in order to represent an infinite environment of the polymers. These are acceptable approximations given that the molecular weights of our polymers range from 4–6kDa in most cases (3–6 monomer repeat units). A set of 139 MOE 2D descriptors and RECON/TAE[38, 39] conformationally independent descriptors was

employed in the current study. MOE 2D descriptors include topological descriptors, atom and bond counts, and partial charge descriptors. These descriptors have been widely used in QSAR studies to provide information on topological connections of atoms, the charge distribution, relative hydrophobicities and other properties. The RECON/TAE method, used as an alternative to *ab initio* calculations, was developed by Breneman *et al.*, [38, 39] to calculate electron density-based properties of molecules in high-throughput QSAR methodologies.

To reduce the dimensionality of the initial descriptor set (Table S3, Supporting Information section) in the QSAR model, selection was limited to those that had low correlations with other descriptors but showed high correlations with the response $\log_{10}(\text{RLU}/\text{mg protein})$. This “objective feature selection” process was done as follows: (1) Descriptors that were highly correlated (Pearson’s correlation coefficient > 0.85) were eliminated from the set, thereby reducing the likelihood of fortuitous correlations within the model. (2) Descriptors that exhibited little influence on the response value (Pearson’s correlation coefficient < 0.15) were also removed. (3) For the remaining descriptors, those that had the largest correlation with the response from each group of highly correlated descriptors were selected for use. A reduced set of 18 descriptors (Table S4, Supporting Information section) was obtained after these three steps.

In order to reduce the potential for over-fitting, and to increase the predictive ability of the model, a sensitivity analysis-based “subjective” feature (descriptor) selection process was then carried out using an in-house method described previously by us [40]. Sensitivity analysis determines the influence of each descriptor on the output of a constructed model, enabling us to eliminate the least relevant descriptors from the modeling process. Our software computes numerical derivatives for each descriptor with respect to the model output by incrementally increasing or decreasing the value of each descriptor while holding others at their mean values. The differences between the predicted values based on the original model and the incremented-descriptor model were then used as metrics of descriptor sensitivity. In a typical run, sensitivity analysis was carried out over eight iterations, with 85% of the top-ranking descriptors from each iteration being selected for the final descriptor set (Table 1) that was used to generate the final Support Vector Regression (SVR) QSAR model.

2.5.3. Support Vector Machines (SVMs)—A schematic of the overall SVR-based QSAR approach is shown in Figure S4 (Supporting Information section). Support vector machines (SVM) are a powerful and increasingly popular mathematical algorithm used to build both, classification (SVM) and regression models (SVR). The utility of the support vector approach stems from its low sensitivity to data overfitting [41]. SVR models are created using training sets in the same way as any other regression model, and have been successfully employed for a variety of applications [28, 41, 42]. In the current study, an SVR-based QSAR model was constructed using our in-house software [40] to quantitatively relate the selected descriptor patterns to transgene expression efficacies. Model cross-validation, Yscrambling, and model evaluation methods (including metrics), are described in the Supporting Information section (Part B2 and Table S5).

3. RESULTS AND DISCUSSION

3.1. Combinatorial Synthesis of Aminoglycoside-based Polymer Library for Transgene Delivery and Expression

Aminoglycosides are a class of small-molecule antibiotics that are clinically used for treating bacterial infections[43, 44]. Aminoglycosides exhibit antibacterial properties by different mechanisms, including irreversible inhibition of protein synthesis upon binding to the 30S subunit of bacterial ribosome. The multi-cationic nature of aminoglycosides is key to binding prokaryotic ribosomal RNA. Studies have also shown that aminoglycosides and their derivatives can bind DNA[28, 45], leading to the exciting possibility of using these molecules for transgene delivery.

In general, small molecular weight polyamines possess poor efficacies for delivering nucleic acids to cells, while polymers show enhanced efficacies[46]. In addition to several amines which facilitate DNA binding, the presence of multiple sugar groups in aminoglycosides was hypothesized to result in polymers with low toxicities. We employed aminoglycosides as monomers for generating a library of 56 polymers, using diglycidyl ethers[36] as cross-linkers (Scheme S1, Supporting Information section). While linear polyamines have been previously employed for generating polymer libraries for transgene expression[36], we describe generation of aminoglycoside-based polymer libraries in this work. Seven different aminoglycosides with diverse molecular structural attributes were employed for generating the polymer library. Diglycidyl ethers were used as cross-linkers, because of their rapid ability to react with amines, resulting in polymer formation. The polymerization reaction employed for generation of the library is shown in Scheme S1.

3.2. Parallel Screening of Aminoglycoside Polymers for Transgene Expression Efficacy

The aminoglycoside-based polymer library was first screened in parallel in order to identify candidates that resulted in the highest transgene expression levels in PC3 human prostate cancer cells (Figure 1a) and Mia PaCa-2, pancreatic cancer cells (Figure 1b). This primary screen led to the identification of several candidates that exhibited significantly higher luciferase protein expression than 25 kDa poly(ethyleneimine) or pEI at the same polymer:pDNA weight ratio (25:1) in PC3 cells. Several polymers from the library also demonstrated high levels of transgene expression in Mia-PaCa-2 cells, albeit lesser numbers of polymer leads candidates were found compared to those with PC3 cells. Seven polymers, apramycin-RDE (**P1**), paromomycin-RDE (**P2**), paromomycin-GDE (**P3**), sisomicin-RDE (**P4**), sisomicin-GDE (**P5**), amikacin-EGDE (**P6**), and neomycin-GDE (**P7**), demonstrated the highest levels of luciferase (RLU/mg) expression, and were characterized further for molecular weights, ¹H-NMR spectra, amine concentrations and cytotoxicity (Figures S5–S12 and Table S6, Supporting Information section). Physicochemical properties of polymer:pDNA complexes are summarized in Table S6 (Supporting Information section).

Lead polymers were evaluated at different polymer:pDNA N/P ratios in several cell lines, and also for delivering a different plasmid. Although the primary screen was carried out with a polymer:pDNA weight ratio of 25:1 for convenience, we employed N/P (number of nitrogens in polymers to phosphates in pDNA) or ‘charge ratios’ in these extended studies in

order to facilitate a more comprehensive comparison. Figure 2 shows luciferase expression following pDNA delivery using lead aminoglycoside polymers and pEI at different polymer to pDNA (N/P) ratios. Dose response studies revealed that pEI exhibited highest luciferase expression at an N/P ratio of 7.6 in PC3 cells; under these conditions, the pEI concentration was 0.4 ng/μl, and the cell viability was 94%. However, significantly lesser levels of luciferase expression were found at higher concentrations of pEI, in part due to significant cytotoxicity of higher doses of the polymer. It is important to note that these are optimized results with pEI. We have consistently observed similar activities with newly purchased stocks of pEI, whose aqueous solutions are prepared freshly prepared before transfection experiments. Empirical observations in our hands indicate that pEI efficacy varies significantly with age of the pEI stock and its storage (not shown). While a more systematic investigation is underway in our laboratory, we wish to alert the reader that the efficacy of pEI is likely dependent on storage.

Most lead polymers exhibited highest levels of luciferase expression efficacies when the N/P (nitrogen to phosphate) ratios were between 50–120 (Figure 2). As the polymer concentration increased, some polymers exhibited lower luciferase expression levels, presumably due to toxicity under these conditions. For example, polymers **P1**, **P3** and **P5** exhibited highest luciferase expression at an N/P ratio of ~100, while **P2**, **P4** and **P7** exhibited the highest luciferase expression at an N/P ratio of ~50. Luciferase expression dropped only ~2-fold in presence of 10% serum relative to that in absence of serum, indicating that serum has only modest effects (Figure S13a, Supporting Information section). Decrease in transgene expression in the presence of serum is likely due to the interference of serum proteins with polyplex uptake[47, 48]. Lead polymers **P1**, **P2**, **P3** and **P4** demonstrated minimal decrease in transgene expression in 22Rv1 cells in presence of serum-containing media, while **P5**, **P6**, and **P7** demonstrated a two-fold decrease (Figure S13b, Supporting Information section). Transgene expression efficacy of lead polymers in MB49 murine bladder cancer cells was also higher for the lead polymers compared to that observed with pEI (Figure S14, Supporting Information section).

In order to investigate if lead polymers could result in expression of a different transgene, PC3 and 22Rv1 cells were transfected with pEGFP plasmid that encodes for enhanced green fluorescent protein or EGFP. Transgene expression efficacies were visualized by fluorescence microscopy, and quantified using flow cytometry. Figure S15a shows fluorescence microscopy images of PC3 cells transfected with pEGFP both, in the presence and absence of serum; Figure S16 (Supporting Information section) shows images of 22Rv1 cells under similar conditions. Higher EGFP expression can be visualized in both cell lines for aminoglycoside-based polymers compared to that observed with pEI. In these studies, aminoglycoside polymers and pEI were evaluated at their optimal polymer:pDNA ratios (Figure 2). EGFP expression decreased in the presence of serum, although the decrease was significantly more drastic for pEI.

Figure S15b shows the flow cytometry analysis of PC3 cells expressing EGFP in the presence and absence of serum. Up to 23% cells expressed EGFP when polymer **P3** was used, while **P1**, **P2** and **P7** resulted in EGFP expression in up to 19%, 15% and 20% cells, respectively. Interestingly, use of pEI resulted in EGFP expression in only 1.7% cells. The

percentage of cells expressing EGFP decreased in serum-containing media for all polymers, including pEI. Polymers **P3**, **P2**, **P2**, and **P7** resulted in EGFP expression in up to 13%, 12%, 11% and 11% cells, respectively in the presence of serum, while pEI demonstrated EGFP expression in only 1% cells. The amount of pDNA delivered did not significantly influence transgene (EGFP) expression; the polymer (**P3**) amount was kept constant at 5 μ g, while the pEGFP amount was varied from 100 ng to 10 μ g. Transgene expression remained largely invariant from 100 – 1000 ng pEGFP. However, EGFP expression was significantly lower for higher pEGFP amounts, presumably due to reduction in the polymer:pEGFP weight, and therefore, charge ratios (Figure S17, Supporting Information section), and reduced availability of the polymer for uptake by cells.

3.3 Cheminformatics Studies: QSAR Models of Aminoglycoside-Polymer Mediated Transgene Expression

Physicochemical properties of polymers play a significant role in determining the intracellular fate and transgene expression efficacy of plasmid DNA in target cells. We therefore generated cross-validated SVR-based QSAR models in order to correlate relationships between chemical attributes of aminoglycoside polymers and their transgene expression efficacies. QSAR modeling relies on identifying important correlations between the statistical variances of a set of molecular physicochemical “descriptors” and a response variable, i.e. polymer-mediated luciferase expression. Descriptors are numerical representations of the relevant features of a molecular system that may play causative roles in its behavior. The function of machine learning in QSAR model development is to train a statistical learning model to associate key variation patterns of descriptors with known system responses. This is accomplished through use of training data and a model validation workflow[49]. A schematic of a typical QSAR modeling approach is shown in Figure S4 (Supporting Information section).

We first processed the training and test data as described in the Experimental and Supporting Information sections (Part B), in order to ensure balanced and representative data sets. The set of 33 polymers selected for cheminformatics studies was representative of all seven aminoglycosides and eight diglycidyl ethers employed for polymer synthesis. This process also yielded balanced response values i.e. $\log_{10}(\text{RLU}/\text{mg protein})$, which ensured an unbiased QSAR model. The model was built based on precisely tested transgene expression efficacies by selecting data points with relatively small STD/(RLU/mg protein) values. The selected test set was external to the model training procedures, and was not involved in any part of the model training / generation processes.

We developed a ‘building block structure’ of aminoglycoside-based polymers in order to generate a simplified structure of the aminoglycoside polymers (Figures S3aa–bg) as described in the Experimental section. The building block structures of our polymers are constructed with only primary amines of the aminoglycoside core reacting with epoxide moieties in the diglycidyl ethers. This is an excellent approximation given that the ratio of kinetic rate constant for primary amine reactions with epoxide moieties is twice that of secondary amines[37]. The second epoxide group on a diglycidyl linker was also opened in order to represent an infinite environment of the polymers. These are acceptable

approximations given that the molecular weights of our polymers range from 4–6kDa in most cases (3–6 monomer repeat units). These simplified polymer structures were employed for calculating molecular descriptors, and further applied for model construction. As described on the Methods section, we reduced the descriptor set to 5 from an initial set of 139 in order to reduce descriptor redundancy and overfitting.

Support Vector Regression—As described in the Methods section, $\log_{10}(\text{RLU}/\text{mg})$ values were employed as the response variable for generating the SVRbased QSAR models of transgene expression. The QSAR model was based on five descriptors, *PEOE_VSA_PPOS*, $\log P(o/w)$, *RECON_SIEPMax*, *BCUT_PEOE_3*, and *RECON_PIPMax* (Table 1), which were identified as part of the feature selection process. Considering the wide range of magnitudes in luciferase expression efficacy values, some of the descriptors demonstrated an approximately logarithmic relationship with RLU/mg values (Figure S18, Supporting Information section). When the response value was modeled using a log scale, the variances of the magnitudes of the values represented a more useful way of representing this relationship. Construction of the SVR-based QSAR model was accomplished using our in-house software[40]; Figure 3a–c shows the SVR model for $\log_{10}(\text{RLU}/\text{mg})$ values. The training model had a squared Pearson's correlation coefficient (r^2) of 0.78 and a coefficient of determination (R^2) of 0.78 (Figure 3a and Table S5, Supporting Information section). The cross-validated model was constructed using a part of the training set, and tested on the remaining polymers as the validation set. The cross-validated model had an r^2 value of 0.65 and an R^2 of 0.65 (Figure 3b). These results indicated that the QSAR model had a robust predictive ability. The decrease of the squared correlation coefficients while using part of the training set in the cross-validated model indicated that the polymers in the complete training set provided more chemical information for model construction compared to when only subsets of the training data were used.

Figure S19 (Supporting Information section) shows the Y-scrambling results with respect to both the squared Pearson's correlation coefficient and the root-mean-square-error (RMSE). As described in the Supporting Information section (Part B), the y-scrambling technique was used to test the overfitting potential of our modeling method. The point on the top right of the Figure S19a shows the r^2 value of the actual model, while all points on the bottom left show r^2 values of the artificial scrambled models. Figure S19b is similar to Figure S19a, but with the bottom right showing the RMSE for the actual model, while the top left shows the RMSE for the scrambled models. Yscrambling indicated that the actual model was easily distinguishable from scrambled models, which provided an indication that the model was robust and not over-trained.

In keeping with the model validation results, actual predictions of transgene expression efficacy of an external test set of polymers showed excellent agreement with experimentally observed values (Figure 3c and 3d). The $\log_{10}(\text{RLU}/\text{mg})$ values were converted back to RLU/mg values using the anti-logarithm function. Model prediction for RLU/mg values were also in excellent agreement with experimentally determined RLU/mg values. It is important to note that the external test set was not included at any stage of the model generation / descriptor selection process. In summary, the SVR-based QSAR model, based

on only five physicochemical descriptors, was robust, and demonstrated excellent predictive ability for polymers not seen during training of the model.

QSAR Model Interpretation: Polymer Physicochemical Factors Influencing Transgene Expression

—We next examined the relative contributions of the five descriptors represented in the model: *PEOE_VSA_PPOS*, *logP(o/w)*, *RECON_SIEPMax*, *BCUT_PEOE_3*, and *RECON_PIPMax*, in order to examine how different physicochemical properties of polymers influence their transgene expression efficacies. While we recognize the hazards associated with attributing phenomenological causality to specific patterns of descriptor weights, it can nevertheless be instructive to examine the relative importance of each descriptor in the model. To this end, sensitivity analysis of descriptors (Figure 4) was carried out using our in-house software by increasing or decreasing the values in each descriptor column by one standard deviation and performing a new prediction with all other descriptors held fixed. The difference between the predicted values of the new model and those of the original one were then calculated and ranked. Figure 4 shows the sensitivity analysis of the selected descriptors in a ranked order, which indicates the strength of relation between each descriptor and $\log_{10}(\text{RLU/mg})$. Figure 5a–e shows correlation plots between each descriptor and luciferase expression; the plots indicate how these different physicochemical properties influence polymer-mediated transgene expression.

The descriptor *PEOE_VSA_PPOS* is a MOE 2D electrostatic descriptor that represents the total polar positive surface area of each molecule, as described by the empirical PEOE method.[50] As shown in Figure 4, *PEOE_VSA_PPOS* was found to be the most sensitive descriptor in the model. This descriptor provides quantification of the partially positive charged hydrophilic surface in regions of the cationic polymers adjacent to their oxygen atoms. The value of this descriptor appears dependent on the number of oxygen atoms within the polymer, as shown in Figure 6. From the set of cross-linkers employed, glycerol diglycidyl ether or GDE has one extra oxygen atom compared to the others, which is reflected in higher *PEOE_VSA_PPOS* values for polymers with this cross-linker. Higher numbers of oxygen atoms within the aminoglycoside core also resulted in higher values of this descriptor. Thus, presence of higher numbers of oxygen atoms correlate with higher efficacies of transgene expression, likely due to increased hydrogen bonding potential and polarization of the polymers.

The *logP(o/w)* partition coefficient descriptor, an indicator of molecular hydrophobicity, ranked second among the five selected descriptors. The *logP* values for all training set polymers are negative, which indicates the high water solubility associated with our polymers. An investigation into correlation plots indicated that incorporation of some hydrophobicity in the polymers, resulted in higher levels of transgene expression; as seen in Figure 5, polymers with moderately negative values of the descriptor (approximately -5), consistently demonstrated higher levels of transgene expression than those with lower hydrophobicities. Cross-linkers that contain cyclohexyl (i.e. CDDE) or aromatic (i.e. RDE) ring groups demonstrate higher *logP(o/w)* descriptor values as seen using molecular visualization methods (Figure 7), and are associated with polymers that demonstrate higher luciferase expression. These findings indicate that some amount of hydrophobicity is desired

in polymers, likely since this can help them traverse cellular membranes, leading to enhanced delivery efficacies.

The *RECON_SIEPMax* is a RECON/TAE descriptor which correlates with the maximum value of positive electrostatic potential on the surface of a molecule. Figure 8 shows electrostatic maps of molecules generated by MOE; blue regions show the positive electrostatic potential van der Waals surface area of the molecule, while those in red are negative. The descriptor *RECON_SIEPMax* indicates the most highly positive surface area of the polymers, which corresponds to the blue regions in the map. The difference between electrostatic maps calculated using MOE and this RECON/TAE descriptor is that the surfaces in Figure 8 are based on energy minimized 3D building block structures, while the RECON/TAE descriptors are 2D-based electronic descriptors that capture atom/group-based properties. As shown in the figure, most cross-linkers show neutral electrostatic potential (white), but the phenyl group in RDE has a negative electrostatic potential (red) and NPGDE and PPGDE have a small blue-colored region in the middle of the molecule. Since the descriptor indicates only the most positive surface area within a molecule, its values for NPGDE and PPGDE are slightly larger than those for other cross-linkers (e.g. RDE and GDE). This might be due to presence of methyl groups within these two cross-linkers that increase the local electrostatic potential. In addition, this descriptor was found having no high correlation (Pearson's correlation coefficient > 0.80) with any other MOE 2D or RECON/TAE descriptors, but had a correlation coefficient of 0.63 with the number of double bonds within the polymers. The effects of different double bonds on the descriptor value can also be found in the electrostatic maps shown in Figure 8. Streptomycin has a relatively high descriptor value due to its basic guanidine group. In addition, sisomycin also demonstrated high values of this descriptor since the double bond in one of its glycoside six-membered ring affects the basicity of the amine group in its proximity. Amikacin-based polymers contain a high local electrostatic potential around the peptide bond in this aminoglycoside (Figure 8). This descriptor does not demonstrate a clear correlation with $\log_{10}(\text{RLU})$ values, so it is difficult to draw simplistic conclusions on its effect on transgene expression. However, some specific effects offer interesting insights. First, the branched methyl group in the cross-linkers might be unfavorable for condensation of plasmid DNA condensation given the bulk and hydrophobicity of these moieties. However, in case of aminoglycoside monomers, increasing basicity can facilitate DNA binding. This is not obvious in streptomycin-based polymers, probably due to the fact that the two amine groups within the guanidine group are very close to each other.

The descriptor *BCUT_PEOE_3* is based on a combination of connectivity information and partial charge properties of molecules, and belongs to a class of descriptors originally developed for use in determining molecular similarities[51, 52]. We observed that a decrease in the size and increase in the total partial charges of the aminoglycoside core correlated with increased values of this descriptor (Figure 5). Aminoglycosides with small sizes, e.g. sisomicin, amikacin and kanamycin, were favorable for polymer-mediated transgene expression. However, some polymers, including those based on neomycin, paromomycin and apramycin, possess relatively larger *BCUT_PEOE_3* values and demonstrated similar transgene expression efficacies as those with lower values for the

descriptor. Further analysis indicated that aminoglycosides with larger sizes also possessed higher number of amines. The presence of the *BCUT_PEOE_3* descriptor in the model therefore indicates that increasing the number of amines in aminoglycosides contributes to the favorable positive charge of the polycations, which, in turn, favors transgene expression. Although this is along expected lines, it is important that our model captures the role of Coulombic interactions in pDNA-mediated transgene expression. It is likely that some neomycin, paromomycin, and apramycin-derived polymers demonstrated high transgene expression efficacies due to the presence of larger numbers of ionizable amines, even though they had unfavorable core sizes.

The *RECON_PIP* descriptor series describe how tightly held the electron density is at various locations on the van der Waals surface of the molecule. This feature is associated with both nucleophilicity and Lewis basicity, and the *RECON_PIP_Max* descriptor describes the site of the highest local ionization potential of the molecules. While not experimentally observable, *PIP_Max* describes regions of molecules that have tightly held electron densities, which are associated with both chemical “hardness” and electrophilic character. Sisomicin core has a unique value of this descriptor compared to others, most likely due to the carbon-carbon double bond within the molecule that increase the nucleophilicity and basicity of amines in the vicinity of the double bond.

4. CONCLUSIONS

We successfully generated an polymer library based on ring-opening polymerization of aminoglycoside antibiotics with diglycidyl ethers as cross linkers. Parallel screening of this library for transgene delivery led to the identification of several leads which demonstrated high luciferase expression levels. Cell viability studies revealed that aminoglycoside-derived polymers were less toxic than 25 kDa pEI. Polymers in which aminoglycosides were crosslinked with resorcinol diglycidyl ether (RDE) and glycerol diglycidyl ether (GDE) demonstrated among the highest levels of transgene expression. Experimental observations indicate that physicochemical properties of polymers play a key role in determining the efficacy of transgene expression. The role of physicochemical properties on the efficacy of polymer-mediated transgene expression was therefore investigated using QSAR models generated using Support Vector Regression (SVR). The model was robust, not over-trained, and was able to predict the efficacies of polymers not included in the training set. The QSAR model indicated that balance between hydrophobic and hydrophilic moieties, as well as the basicity of the aminoglycoside amine groups, were instrumental in determining the efficacy of polymer-mediated transgene expression. In all, we successfully synergized combinatorial synthesis and parallel screening of a polymer library, along with the development of cheminformatics-based QSAR models that can predict efficacy of transgene expression and report for key molecular attributes governing polymer-mediated transgene expression in a quantitative manner. This combinatorial synthesis and parallel screening approach, combined with QSAR models, can have a transformative impact on the rational design and development of materials for nucleic acid delivery for applications in gene therapy, cellular engineering, and gene silencing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the NIH/NIGMS (Grant 1R01GM093229-01A1) and the Arizona Biomedical Research Commission (ABRC; Grant 1009) for financial support of this study. We also like to thank the RPI Center for Biotechnology and Interdisciplinary Studies for being an excellent cheminformatics work environment.

REFERENCES

1. Mintzer MA, Simanek EE. Nonviral vectors for gene delivery. *Chem Rev.* 2009; 109:259–302. [PubMed: 19053809]
2. Verma IM, Somia N. Gene therapy: promises, problems and prospects. *Nature.* 1997; 389:239–242. [PubMed: 9305836]
3. Zhang S, Zhao Y, Zhao B, Wang B. Hybrids of nonviral vectors for gene delivery. *Bioconjugate Chem.* 2010; 21:1003–1009.
4. Anderson WF. Human gene therapy. *Nature.* 1998; 392:25–30. [PubMed: 9579858]
5. Check E. Gene therapy: A tragic setback. *Nature.* 2002; 420:116–118. [PubMed: 12432357]
6. Check E. Gene therapy put on hold as third child develops cancer. *Nature.* 2005; 433:561.
7. Lehnman S. Virus treatment questioned after gene therapy death. *Nature.* 1999; 401:517–518. [PubMed: 10524611]
8. Nayak S, Herzog RW. Progress and prospects: immune responses to viral vectors. *Gene Therapy.* 2010; 17:295–304. [PubMed: 19907498]
9. Thomas CE, Ehrhardt A, Kay MA. Box 2 | Engineering the main groups of viral vectors. *Nat Rev Genet.* 2003; 4:346–358. [PubMed: 12728277]
10. Abdallah B, Hassan A, Benoist C, Goula D, Behr JP, Demeneix BA. A Powerful nonviral vector for in vivo gene transfer into the adult mammalian brain: polyethylenimine. *Hum Gene Ther.* 1996; 7:1947–1954. [PubMed: 8930654]
11. Zanta M-A, Boussif O, Adib A, Behr J-P. In vitro gene delivery to hepatocytes with galactosylated polyethylenimine. *Bioconjugate Chem.* 1997; 8:839–844.
12. Anderson DG, Lynn DM, Langer R. Semi-automated synthesis and screening of a large library of degradable cationic polymers for gene delivery. *Angew Chem Int Edit.* 2003; 115:3261–3266.
13. Green JJ, Langer R, Anderson DG. A combinatorial polymer library approach yields insight into nonviral gene delivery. *Accounts Chem Res.* 2008; 41:749–759.
14. Lynn DM, Langer R. Degradable Poly(β -amino esters): synthesis, characterization, and self-Assembly with plasmid DNA. *J Am Chem Soc.* 2000; 122:10761–10768.
15. Mendiratta SK, Quezada A, Matar M, Wang J, Hebel HL, Long S, et al. Intratumoral delivery of IL-12 gene by polyvinyl polymeric vector system to murine renal and colon carcinoma results in potent antitumor immunity. *Gene Therapy.* 1999; 6:833–839. [PubMed: 10505108]
16. Lee CC, MacKay JA, Frechet JMJ, Szoka FC. Designing dendrimers for biological applications. *Nat Biotechnol.* 2005; 23:1517–1526. [PubMed: 16333296]
17. Tang MX, Redemann CT, Szoka FC. In vitro gene delivery by degraded polyamidoamine dendrimers. *Bioconjugate Chem.* 1996; 7:703–714.
18. Hwang SJ, Bellocq NC, Davis ME. Effects of structure of β -cyclodextrin-containing polymers on gene delivery. *Bioconjugate Chem.* 2001; 12:280–190.
19. Pun SH, Bellocq NC, Liu A, Jensen G, Machemer T, Quijano E, et al. Cyclodextrin-modified polyethylenimine polymers for gene delivery. *Bioconjugate Chem.* 2004; 15:831–840.
20. Cheng F, Yu Y, Shen J, Li W, Liu G, Lee PW, et al. Classification of cytochrome P450 inhibitors and noninhibitors using combined classifiers. *J Chem Inf Model.* 2011; 51:996–1011. [PubMed: 21491913]

21. Hou T, Wang J, Li Y. ADME evaluation in drug discovery. 8. The prediction of human intestinal absorption by a support vector machine. *J Chem Inf Model*. 2007; 47:2408–2415. [PubMed: 17929911]
22. Judson R, Elloumi F, Setzer RW, Li Z, Shah I. A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. *BMC Bioinformatics*. 2008; 9
23. Lombardo F, Obach RS, DiCapua FM, Bakken GA, Lu J, Potter DM, et al. A hybrid mixture discriminant analysis-random forest computational model for the prediction of volume of distribution of drugs in human. *J Med Chem*. 2006; 49:2262–2267. [PubMed: 16570922]
24. Zaretski J, Bergeron C, Rydberg P, Huang T-w, Bennett KP, Breneman CM. RS-predictor: a new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP3A4. *J Chem Inf Model*. 2011; 51:1667–1689. [PubMed: 21528931]
25. Mazza CB, Rege K, Breneman CM, Sukumar N, Dordick JS, Cramer SM. High-throughput screening and quantitative structure-efficacy relationship models of potential displacer molecules for ion-exchange systems. *Biotechnol Bioeng*. 2002; 80:60–72. [PubMed: 12209787]
26. Ladiwala A, Rege K, Breneman CM, Cramer SM. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proc Natl Acad Sci U S A*. 2005; 102:11710–11715. [PubMed: 16081542]
27. Das S, Kokardekar A, Breneman CM. Rapid comparison of protein binding site surfaces with property encoded shape distribution. *J Chem Inf Model*. 2009; 49:2863–2872. [PubMed: 19919089]
28. Rege K, Ladiwala A, Hu S, Breneman CM, Dordick LS, Cramer SM. Investigation of DNA-Binding Properties of an Aminoglycoside-Polyamine Library Using Quantitative Structure-Activity Relationship (QSAR) Models. *J Chem Inf Model*. 2005; 43:1854–1863. [PubMed: 16309293]
29. Sajda P. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng*. 2006; 8:537–565. [PubMed: 16834566]
30. Rege K, Ladiwala A, Tugcu N, Breneman CM, Cramer SM. Parallel screening of selective and high-affinity displacers for proteins in ion-exchange systems. *J Chromatogr A*. 2004; 1033:19–28. [PubMed: 15072287]
31. Gordon EM, Barrett RW, Dower WJ, Fodor SPA, Gallop MA. Applications of combinatorial technologies to drug discovery. 2. combinatorial organic synthesis, library screening strategies, and future directions. *J Med Chem*. 1994; 37:1385–1401. [PubMed: 8182695]
32. Gallop MA, Barrett RW, Dower WJ, Fodor SPA, Gordon EM. Applications of combinatorial technologies to drug discovery. 1. background and peptide combinatorial libraries. *J Med Chem*. 1994; 37:1233–1251. [PubMed: 8176700]
33. Vu L, Ramos J, Potta T, Rege K. Generation of a focused poly(amino ether) library: polymer-mediated transgene delivery and gold-nanorod based theranostic systems. *Theranostics*. 2012; 2:1160–1173. [PubMed: 23382773]
34. Barua S, Ramos J, Potta T, Taylor D, Huang H-C, Montanez G, et al. Discovery of cationic polymers for non-viral gene delivery using combinatorial approaches. *Comb Chem High Throughput Screen*. 2011; 14:908–924. [PubMed: 21843141]
35. Micheal L, Wang H, Tor Y. Enhanced RNA binding of dimerized aminoglycosides. *Bioorgan Med Chem*. 1999; 7:1361–1371.
36. Barua S, Joshi A, Banerjee A, Matthews D, Sharfstein ST, Cramer SM, et al. Parallel synthesis and screening of polymers for nonviral gene delivery. *Molecular Pharmaceutics*. 2009; 6:86–97. [PubMed: 19102694]
37. Paz-Abuin S, Pellin MP, Paz-Pazos M, Lopez-Quintela A. Influence of the reactivity of amine hydrogens and the evaporation of monomers on the curing kinetics of epoxy-amine: kinetic questions. *Polymer*. 1997; 38:3795–3804.
38. Breneman CM, Thompson TR, Rhem M, Dung M. Electron density modeling of large system using the transferable atom equivalent method. *Computers Chem*. 1995; 19:161–179.

39. Breneman CM, Rhem M. QSPR analysis of HPLC column capacity factors for a set of high-energy materials using van der Waals surface property descriptors computed by transferable atom equivalent method. *J Computational Chemistry*. 1997; 18:182–197.
40. Krein, M. On the development and use of large chemical similarity networks, informatics best practices and novel chemical descriptors towards materials quantitative structure property relationships. Troy: Rensselaer Polytechnic Institute; 2011.
41. Yap CW, Chen YZ. Prediction of cytochrome P450 3A4, 2D6, 2C9 inhibitors and substrates by using support vector machines. *J Chem Inf Model*. 2005; 45:982–992. [PubMed: 16045292]
42. Liu Y. Active learning with support vector machine applied to gene expression data for cancer classification. *J Chem Inf Comput Sci*. 2004; 44:1936–1941. [PubMed: 15554662]
43. Davies J, Davis BD. Misreading of ribonucleic acid code words induced by aminoglycoside antibiotics. *J Biol Chem*. 1968; 243:3312–3316. [PubMed: 5656371]
44. Moazed D, Noller HF. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*. 1987; 327:389–394. [PubMed: 2953976]
45. Rege K, Hu S, Moore JA, Dordick JS, Cramer SM. Chemoenzymatic synthesis and high-throughput screening of an aminoglycoside–polyamine library: identification of high-affinity displacers and DNA-binding ligands. *J Am Chem Soc*. 2004; 126:12306–12315. [PubMed: 15453764]
46. Dauty E, Remy J-S, Blessing T, Behr H-P. Dimerizable cationic detergents with a low cmc condense plasmid DNA into nanometric particles and transfect cells in culture. *J Am Chem Soc*. 2001; 123:9227–9234. [PubMed: 11562201]
47. Li S, Tseng WC, Stolz DB, Wu SP, Watkins SC, Huang L. Dynamic changes in the characteristics of cationic lipidic vectors after exposure to mouse serum: implications for intravenous lipofection. *Gene Ther*. 1999; 6:585–594. [PubMed: 10476218]
48. Zelphati O, Uyechi LS, Barron LG, Szoka FC Jr. Effect of serum components on the physico-chemical properties of cationic lipid/oligonucleotide complexes and on their interactions with cells. *Biochim Biophys Acta*. 1998; 1390:119–133. [PubMed: 9507083]
49. Tropsha A, Golbraikh A. Predictive QSAR Modeling Workflow, Model applicability domains, and virtual screening. *Current Pharmaceutical Design*. 2007; 13:3494–3504. [PubMed: 18220786]
50. Czodrowski P, Dramburg I, Sottriffer CA, Klebe G. Development, validation, and application of adapted PEOE charges to estimate pK(a) values of functional groups in protein-ligand complexes. *Proteins*. 2006; 65:424–437. [PubMed: 16927370]
51. Pearlman RS, Smith KM. Metric Validation and the receptor-relevant subspace concept. *J Chem Inf Comput Sci*. 1999; 39:28–35.
52. Pearlman RS, Smith KM. Novel software tools for chemical diversity. *Perspect Drug Discovery Des*. 1998; 9–11:339–353.

Figure 1. Luciferase transgene expression following aminoglycoside polymer-mediated delivery of pGL3 plasmid DNA to (a) PC3 human prostate cancer cells and (b) Mia Pa Ca 2 human pancreatic cancer cells at a polymer to plasmid weight ratio of 25:1 in serum-free media. Transgene expression efficacies of polymers were quantified in terms of relative luciferase units (RLU) and normalized to total protein content (mg), resulting in RLU/mg values, which were used for comparison between polymers. Experiments were carried out in triplicate (n=3; error bars not shown).

Figure 2.

Effect of lead aminoglycoside-based polymers and pEI concentration on luciferase transgene expression (RLU/mg) in PC3 cells. Lead polymers and pEI were evaluated at different polymer:pDNA N/P ratios in the absence of serum. Different amounts of lead polymers or pEI and the same amount of the pGL3 plasmid (200 ng) were used in the experiments. Data represent mean $\pm$ one standard deviation of three independent experiments (n=3). Lead polymers demonstrated up to 38-fold higher luciferase expression compared to pEI. * = $p < 0.05$, ** = $p < 0.01$.

0.05; ** = $p < 0.01$. p-values were obtained by comparing data for each lead polymer to pEI under the corresponding conditions.

Figure 3.

Support Vector Regression based QSAR model of transgene expression efficacy. (a) SVR-based QSAR model of the polymer training set; (b) Cross-validation SVR model on the polymer training set; (c) Cross-validation SVR model predictions for an external test set of three polymers using $\ln(RRLU/mg \text{ protein})$ as indicator of transgene expression efficacy; (d) SVR-based QSAR model predictions for three external test set polymers using RLU/mg protein as the response variable.

Figure 4.

SVM sensitivity analysis of the five descriptors in ranked order as they appear in the QSAR model. The blue-colored bars represent the impact of increasing the descriptor by one standard deviation, while the red colored bars represent the impact of decreasing the descriptor by one standard deviation.

a. Correlation plot of RECON_PIPMax **b. Correlation plot of RECON_SIEPMax**

c. Correlation plot of PEOE_VSA_PPOS **d. Correlation plot of logP(o/w)**

e. Correlation plot of BCUT_PEOE_3

Figure 5.
Correlation plots of the five descriptors selected in the QSAR model with luciferase expression.

Figure 6.

Molecular visualization of regions corresponding to the *PEOE_VSA_PPOS* descriptor for paromomycin-GDE in 2D (carbon atoms highlighted in pink). Paromomycin-GDE demonstrates the largest value of this descriptor, given that it contains the most number of oxygen atoms. Oxygen atoms are shown in red color in the figure.

Figure 7. Molecular visualization of regions corresponding to $\log P(o/w)$ descriptor. (a) CDDE; (b) NPGDE; (c) BGDE (d) EGDE; (e) RDE; (f) GDE; (g) PEGDE; (h) PPGDE

Figure 8.

Molecular visualization of the MOE electrostatic molecular surface regions corresponding to *RECON_SIEPMax* descriptor. (A) The electrostatic maps of crosslinkers. (a) CDDE; (b) NPGDE; (c) BDGE; (d) EDGE; (e) RDE; (f) GDE; (g) PEGDE; (h) PPGDE. (B) The electrostatic maps of aminoglycoside cores. (a) Amikacin; (b) Streptomycin; (c) Sisomicin. The corresponding regions of the descriptor is highlighted in yellow circles.

Table 1

Definitions of Descriptors Selected in the QSAR Model for Polymer-mediated Transgene Expression

Descriptor Name	Descriptor Definition
<i>PEOE_VSA_PPOS</i>	Total polar positive van der Waals surface area.
<i>logP(o/w)</i>	log octanol/water partition coefficient
<i>RECON_SIEPMax</i>	Maximum value of surface integral of electrostatic potential
<i>BCUT_PEOE_3</i>	BCUT descriptors are calculated from the eigenvalues of a modified adjacent matrix. Each ij entry of the adjacency matrix takes the value $1/\sqrt{b_{ij}}$ where b_{ij} is the formal bond order between bonded atoms I and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the largest eigenvalues are reported as BCUT_PEOE_3.
<i>RECON_PIPMax</i>	Maximum value of PIP Function over molecule.