Investigation of DNA-Binding Properties of an Aminoglycoside-Polyamine Library Using Quantitative Structure−Activity Relationship (QSAR) Models

Kaushal Rege,[†] Asif Ladiwala,[†] Shanghui Hu,[†,||] Curt M. Breneman,[‡] Jonathan S. Dordick,[†,§] and Steven M. Cramer*[,†]

Department of Chemical and Biological Engineering, Department of Chemistry and Chemical Biology, and Department of Biology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180

Received March 10, 2005

We have recently developed a novel multivalent cationic library based on the derivatization of aminoglycosides by linear polyamines. In the current study, we describe the DNA-binding activity of this library. Screening results indicated that several candidates from the library showed high DNA-binding activities with some approaching those of cationic polymers. Quantitative Structure−Activity Relationship (QSAR) models of the screening data were employed to investigate the physicochemical effects governing polyamine−DNA binding. The utility of these models for the a priori prediction of polyamine−DNA-binding affinity was also demonstrated. Molecular descriptors selected in the QSAR modeling indicated that molecular size, basicity, methylene group spacing between amine centers, and hydrogen-bond donor groups of the polyamine ligands were important contributors to their DNA-binding efficacy. The research described in this paper has led to the development of new multivalent ligands with high DNA-binding activity and improved our understanding of structure−activity relationships involved in polyamine−DNA binding. These results have implications for the discovery of novel polyamine ligands for nonviral gene delivery, plasmid DNA purification, and anticancer therapeutics.

INTRODUCTION

Natural polyamines, such as spermine, exist in millimolar (mM) concentrations inside eukaryotic cells and influence a variety of biologically significant reactions with DNA and RNA.[1] In addition, their activity as radioprotective,[2] antioxidant,[3−5] and potential antitumor agents[6−8] is of significant interest. Multivalent cations, such as polyamines, reduce the electrostatic repulsion between DNA molecules by charge neutralization ultimately leading to collapse or condensation of DNA.[9−17] Among the naturally occurring polyamines, spermine is known to be the most efficacious in condensing DNA with a higher efficacy than either spermidine or putrescine. Pelta et al.[12] studied the precipitation of short DNA molecules by spermidine, cobalthexamine, and spermine; while the tetravalent spermine was the most efficient in condensing DNA, cobalthexamine, with three positive charges, was approximately four times more efficient in condensing DNA as compared to the trivalent spermidine. Basu and Marton[13] studied the effects of spermine, two naturally occurring pentamines isolated from the thermophile, *Thermus thermophilus*, and one synthetic pentamine on the aggregation and melting temperature of calf-thymus DNA and on the B-to-Z transition of poly (dG-me5dC). Pentamines caused aggregation of DNA at much lower concentrations than that of spermine suggesting that both the total charge and the distance separating the charge are important for the induction of conformational changes in DNA. Raspaud et. al.[14] analyzed the condensation of DNA, nucleosome, and chromatin in the presence of spermine using electrophoretic measurements. It was observed that the precipitation conditions were independent of the degree of polymerization, the linear charge density, and shape of the DNA. Deng et al.[15] employed Raman spectroscopy to indicate that polyamine−DNA interaction is largely nonspecific in nature. As opposed to alkaline earth metals which disrupt base pairing, polyamine binding to the phosphate groups on DNA did not significantly alter the base pairing interactions of B-DNA. It was also found that the binding of spermine and spermidine molecules to B-form DNA had no effect on the DNA structure. Saminathan et. al.[16] and Vijayanathan et. al.[17] observed significant differences in the ability of various spermine analogues to aggregate oligonucleotides and genomic DNA using static and dynamic light scattering techniques. It was found that although all polyamines caused DNA condensation, their efficacy was related to the structural geometry of the polyamine. The methylene spacing between nitrogen atoms had a significant effect on the condensation of DNA; spermine homologues with methylene spacing from 4 to 8 had comparable condensation properties, while lower and higher homologues caused condensation at higher concentrations.

Despite several reports of polyamine−DNA binding in the literature, limited information exists on relating the structure of different polyamines to their DNA binding, condensation, and aggregation properties. Particularly limiting is the fact that most polyamine−DNA-binding information currently available in the literature is limited to the investigation of linear polyamines. Thus, there is a need to explore a more diverse chemical space in order to better elucidate structure−activity relationships of polyamine DNA binding.

* Corresponding author e-mail: crames@rpi.edu.
[†] Department of Chemical and Biological Engineering.
[‡] Department of Chemistry and Chemical Biology.
[§] Department of Biology.
[||] Present address: Chemical R & D Technologies, La Jolla Laboratories, Pfizer Inc., San Diego, CA 91212.

10.1021/ci050082g CCC: $30.25
Published on Web 10/05/2005

Ethidium bromide (EtBr) is a cationic dye that interacts with both double stranded DNA and RNA by intercalation between the base pairs.[18−21] The molecule has two binding sites with DNA, the primary intercalating site and the secondary electrostatic binding site with phosphate groups on the DNA backbone. A large increase in fluorescence is observed when the phenanthridium moiety of ethidium bromide intercalates with DNA making it a useful probe for measuring drug−DNA interactions. Free ethidium bromide is strongly quenched in aqueous solution and exhibits only weak fluorescence. Measurement of the ability of a drug to displace ethidium bromide from DNA is an established technique for the measurement of DNA-binding ability for both intercalative and nonintercalative molecules. It is hypothesized that polyamine binding above a critical concentration causes conformational changes within the double helix leading to the release of bound ethidium bromide.[18]

Property modeling techniques have been generalized into a broader field known as Quantitative Structure−Property Relationship (QSPR) Analyses. QSPR modeling can be a valuable source of knowledge on the mechanisms operating in a wide variety of experimental systems. These interactions can be identified with the help of multivariate regression equations, relating the molecular structure of analytes to their experimentally determined responses. Quantitative Structure−Activity Relationship (QSAR) approaches have been extensively used to accelerate the development of new and improved therapeutics.[22−24] Subsequently, several researchers have reported correlations for a wide variety of chemical properties including equilibrium constants,[25] drug absorption,[26] toxicity,[27] UV spectral absorbance,[28] and solubility.[29] Quantitative Structure−Retention Relationship (QSRR) models, a subset of QSPR models, have also been reported for various modes of chromatography.[30,31] Recently, we have employed QSRR models to predict protein binding affinity and selectivity[32−34] and Quantitative Structure−Efficacy Relationship (QSER) models to predict displacer efficacy[35,36] in ion-exchange chromatographic systems. While several successful property modeling methods are known, the current work is based on the use of the robust Support Vector Machine (SVM) approach due to its utility in handling large sets of nonorthogonal descriptors and its inherent resistance to overfitting.[37−40]

We have recently developed a novel multivalent cationic library based on the derivatization of aminoglycosides by linear polyamines.[41] In the current manuscript, we investigate the DNA-binding affinity of this unique polyamine-aminoglycoside library. The ethidium bromide displacement assay was employed as a parallel screen to determine the DNA-binding efficacy of the aminoglycoside-polyamine library. The 'percentage fluorescence decreased' value from the ethidium bromide screen was further used to generate Quantitative Structure−Activity Relationship (QSAR) models of DNA-binding efficacy. 2D, 3D Molecular Operating Environment (MOE, Chemical Computing Group Inc., Montreal, Canada) descriptors[32−36] were calculated from the energy-minimized structures of the polyamines. A *training set* of polyamines was selected to generate a QSAR model using SVM regression and bootstrapping techniques,[33,34,36] and the predictions of the test-set ligands were compared with their experimental values. Finally, the molecular descriptors selected during model generation were examined to provide insight into the important structural and physicochemical factors influencing polyamine−DNA binding.

EXPERIMENTAL PROCEDURES

Materials. Apramycin, bekanamycin, butirosin, DAB-$(Am)_4$ Dendrimer generation 1, ethidium bromide, N^8-acetylspermidine, neomycin sulfate, paromomycin sulfate, spermidine trihydrochloride, spermine (free base), streptomycin, tris acid, and tris base were purchased from Sigma (St. Louis, MO). Ethylenediamine, diethylenetriamine, and pentaethylenehexamine were purchased from Aldrich (Milwaukee, WI). Fluorescence analyses were carried out using a Perkin-Elmer plate reader, and the results were analyzed using the software HTSoft 2.0. Fluorescence analyses were carried out using a Costar 96-well plate purchased from Fisher.

Synthesis of the Aminoglycoside-Polyamine Library. The synthesis, purification, and spectroscopic characterization of the polyamine-aminoglycoside library employed in this study is described elsewhere.[41]

Parallel Screening of DNA-Binding Ligands. The components of the polyamine-aminoglycoside library were screened for their DNA-binding efficacy using an ethidium bromide displacement assay. Parallel screening of DNA-binding activity was carried out in the following manner: 1.5 mL of 6 μg/mL calf thymus DNA were equilibrated with 15 μL of 0.5 mg/mL ethidium bromide. After equilibration, 25 μL of each potential DNA-binding compound (0.3 mM) were added to the DNA−ethidium bromide mixture and equilibrated for 20 min. One hundred fifty microliters of the resulting solution was transferred into a 96-well microtiter plate, and the percent fluorescence decrease was monitored using excitation at 260 nm and emission at 595 nm. All solutions were prepared in 20 mM Tris buffer, pH 8.0.

The percent fluorescence decreased values from the ethidium bromide screen were normalized by the 'charge ratio' of the polyamine ligands. The charge ratio of the library candidates was calculated as the total number of ionized amines from the polycationic ligand divided by the total number of phosphates on the DNA backbone (it was assumed that the phosphates carried a charge of −1 at pH 8.0). The ionization of the amine centers on the molecules was calculated as follows: the pK_a value of each amine center on the molecule was calculated using the ACD/pK_a DB software (Advanced Chemistry Development Inc., Toronto, ON, Canada). The degree of ionization of the charge center was calculated using the Henderson-Hasselbach equation. The degree of ionization was added across all charge centers, and the resulting value was reported as the charge content of the molecule. The total number of positive charges was calculated based on the concentration of the ligand, and the resulting value was divided by the number of phosphate groups in solution to yield the charge ratio.

QSAR MODELING PROCEDURE

The goal of QSAR modeling is to correlate the observed experimental data (e.g., percent fluorescence decreased data) with specific physicochemical properties (i.e., molecular descriptors or features) of the molecules. Accordingly, molecular property descriptors were calculated based on molecular structure. A set of 2D and 3D descriptors were computed using the Molecular Operating Environment

(MOE) software package from Chemical Computing Group, Inc. (described in the following section). Using this combined set of descriptors, a Support Vector Machine (SVM) regression algorithm was applied in a feature selection mode to determine a subset of relevant descriptors (i.e., descriptors that showed high correlations with the experimental data). Subsequently, nonlinear SVM models were constructed based on the selected descriptors and were examined for their quality (cross-validated r^2 and RMSE for the training set) and predictive ability for an external test set of molecules not used in the generation of the model. Finally, star plots (described below) were employed as visualization tools to facilitate the interpretation of these models.

Molecular Descriptors. The Molecular Operating Environment (MOE, Chemical Computing Group, Inc., Montreal, Canada) software was used to obtain a set of several types of molecular property descriptors, including connectivity-based topological 2D descriptors, physicochemical property descriptors, shape-independent 3D molecular features, and some pharmacophoric descriptors. The structures of the aminoglycoside-polyamine molecules were drawn in MOE, and formal charges were assigned to the different atoms/groups within the molecules based on the pK_a values calculated by the ACD/pK_a software. Subsequently, the structures were energy minimized using the MMFF94 force field, and MOE descriptors were computed for these molecules using the QuaSAR-descriptors module.

Support Vector Machine (SVM) Regression. An in-house SVM regression program developed by Professor Kristin Bennett and Jinbo Bi in the Department of Mathematics at Rensselaer Polytechnic Institute was used in the present study. The algorithm underlying this program consists of two main steps: feature selection and model building. During the feature selection step, the descriptors which are the most highly correlated with the experimental data are identified, using sparse l_1-norm SVR. Subsequently, in the model building step, the selected descriptors are employed to train the model based on a training set of polyamines using nonlinear SVR.[39,40]

Both the feature selection and model building modes of the SVM implementation employ extensive "*bootstrapping*" where the original training set of polyamines is randomly subdivided into a validation set, with the remaining polyamines used as members of training sets. Each different set of training polyamines is called a "*fold*", and the model created using this set is used to make predictions on the validation set of polyamines left out of the training set for that particular fold. This procedure is repeated 40 times and results in 40 different training and validation subsets and the construction of 40 distinct, but similar models. The predictive quality of the models is initially determined by their performance on the validation sets. Once the models are established, the predictive power of the models is determined based on predictions for true unknowns i.e., a test set of DNA-binding ligands not included in the model training. To facilitate interpretation of these models, star plots are generated to evaluate the relative importance of each of these descriptors selected throughout each of the 40 bootstrap "folds" used for creating the composite model set. A star plot representation of these results consists of a set of radial graphics that represent a weight matrix. In these plots, each star corresponds to a specific descriptor in the weight matrix generated

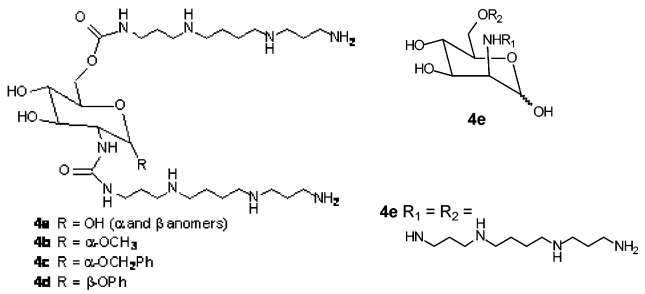

Figure 1. (a) Glucosamine and (b) mannosamine-based dispermine derivatives.

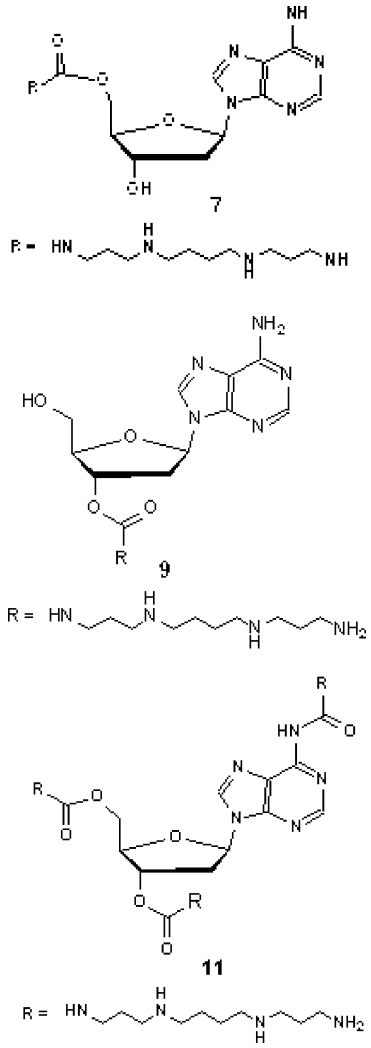

Figure 2. 2′-Deoxyadenosine-based derivatives.

by all 40 linear SVM regression bootstrap folds, and the length of each ray represents the weight of this descriptor in one of the bootstrap iterations. The relative contribution of a descriptor to the aggregate model is quantified by the sum of the weights of that descriptor in all bootstrap folds, relative to the sum of the weights of all descriptors in the model. Additional details about the SVM regression algorithm and the star plot visualization strategy can be found elsewhere.[33]

RESULTS AND DISCUSSION

Figures 1−4 show constituents of the recently developed cationic library based on the derivatization of aminoglycosides by linear polyamines.[41] It was hypothesized that the

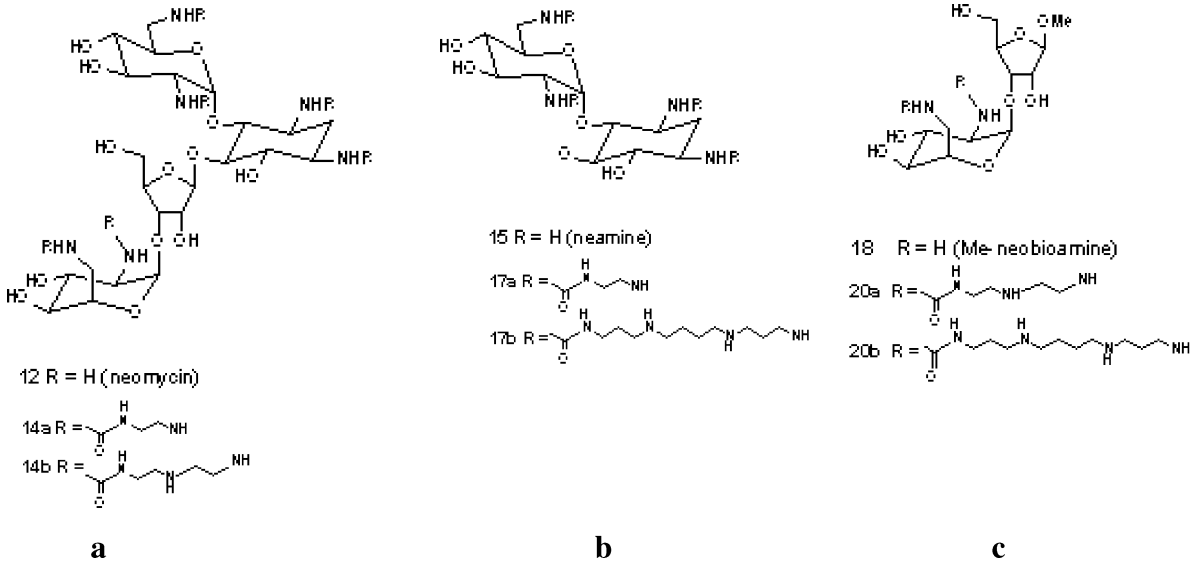

Figure 3. Neomycin (a), neamine (b), and methyl-neobiosamine (c) based polyamine derivatives.

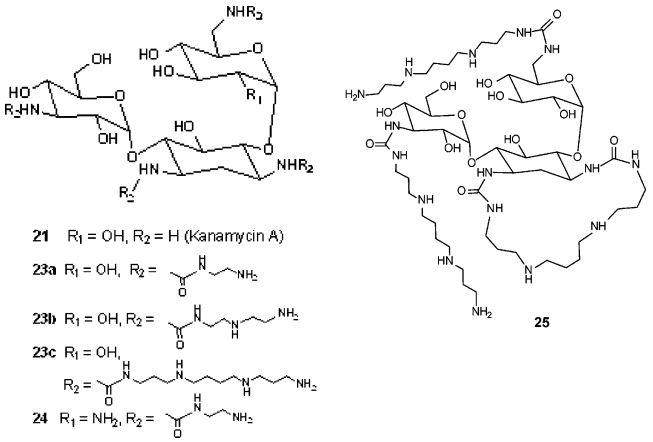

Figure 4. Kanamycin-based derivatives.

constituents of this library could demonstrate high DNA-binding activity due to their high cationic content. Initial work indicated that these polyamines indeed exhibited relatively high DNA binding.[41] In the current study we have extended our initial work on the DNA-binding properties of these molecules by (i) screening more molecules in order to explore a wider chemical space and (ii) employing the screening data to investigate the physicochemical factors influencing polyamine−DNA binding using Quantitative Structure−Activity Relationship (QSAR) models.

Parallel Screening of the Aminoglycoside-Polyamine Library. An ethidium bromide displacement assay was employed to evaluate the DNA-binding affinity of this library in a 96-well format. The percent fluorescence decreased value was used as a parameter to rank the DNA-binding efficacy of the library constituents. As seen in Table 1, naturally occurring polyamines such as spermidine and spermine and commercially available aminoglycosides such as bekanamycin and neomycin demonstrated low percent fluorescence decreased values (0−15%) indicating low DNA-binding activity. Conversely, a significant number of the library constituents demonstrated high percent fluorescence decreased values, indicating high DNA-binding activity. Particularly significant were kanamyin-tetraspermine (**23c**, Figure 4), neamine-tetraspermine (**17b**, Figure 3b), glucosamine-dispermine (**4a**, Figure 1), and kanamycin-trispermine (**25**, Figure 4), which resulted in percent fluorescence decreased values of 82%, 69%, 66%, and 66%, respectively. In addition, these molecules were able to significantly retard calf thymus DNA on agarose gels (not shown). It is particularly encouraging to note that the percent fluorescence decreased value of 25 kDa poly(ethyleneimine) (pEI) was 96% (data not shown), indicating that the DNA-binding

Table 1. DNA-Binding Efficacy Data of the Aminoglycoside-Polyamine Library[a]

no.	polyamine	% fl. dec. (%F)	±	no. of amines/ molecule	net +ve charge	charge ratio	% F/ charge ratio
1	spermidine	0.24	(na)	3	2.81	0.79	0.31
2	ethylenediamine	1.13	1.27	2	1.04	0.29	3.86
3	N8-acetyl spermidine	2.05	0.71	2	1.87	0.53	3.89
4	diethylenetriamine	3.31	4.14	2	1.90	0.54	6.18
5	butirosin	4.06	0.30	4	2.61	0.73	5.53
6	paromomycin	6.21	0.65	5	3.09	0.87	7.12
7	bekanamycin	6.21	6.52	4	2.91	0.82	7.58
8	spermine	7.31	8.75	4	3.66	1.03	7.08
9	apramycin	7.87	1.42	4	2.55	0.72	10.97
10	neamine	8.73	4.27	4	2.48	0.70	12.49
11	streptomycin	9.04	2.88	4	2.66	0.75	12.08
12	pentaethylene hexamine	11.15	1.10	6	2.97	0.84	13.34
13	7	12.56	6.04	4	2.85	0.80	15.64
14	neomycin	14.68	6.82	6	3.91	1.10	13.32
15	9	17.48	4.92	4	2.85	0.80	21.77
16	20a	19.34	2.11	4	2.06	0.58	33.28
17	17a	29.85	5.13	4	3.90	1.10	27.14
18	11	32.26	2.92	9	8.49	2.39	13.49
19	24	32.26	5.94	4	4.59	1.29	24.97
20	$DAB(Am)_4$ dendrimer	39.06	2.78	4	4.05	1.14	34.24
21	4e	46.92	6.67	6	5.70	1.61	29.21
22	23b	49.00	2.32	8	4.24	1.20	40.99
23	4c	51.63	2.95	6	5.70	1.61	32.14
24	14a	54.15	7.59	6	5.83	1.64	32.96
25	20b	56.11	1.33	6	5.66	1.60	35.18
26	4b	59.51	2.00	6	5.70	1.61	37.05
27	17c	64.62	1.97	9	8.55	2.41	26.82
28	25	65.96	6.07	8	7.70	2.17	30.42
29	4a	66.18	5.97	6	5.70	1.61	41.20
30	17b	69.74	1.15	12	11.23	3.15	22.17
31	23c	82.49	2.15	12	11.23	3.16	26.07

[a] The ± values represent one standard deviation of the percent fluorescence decreased data.

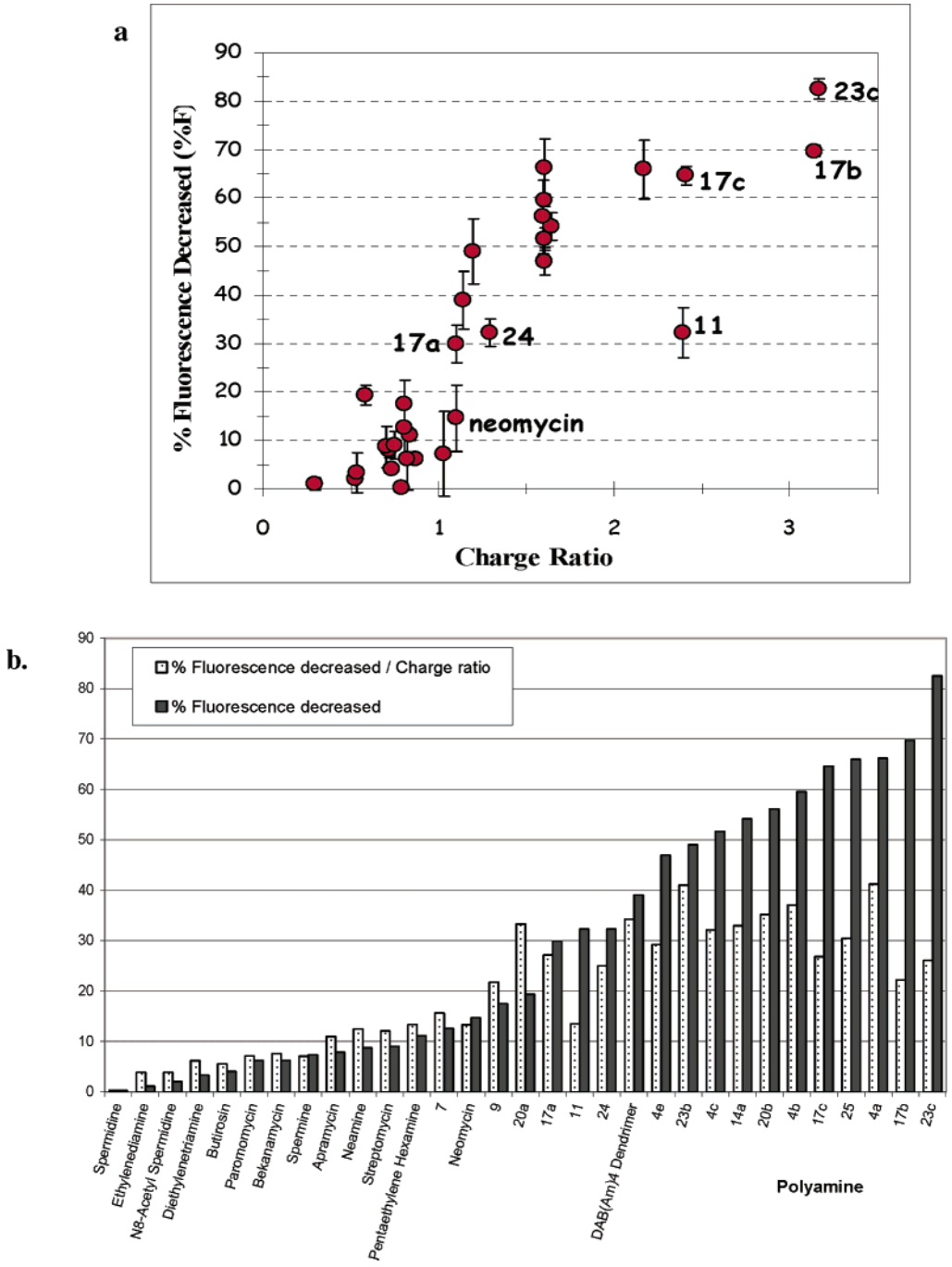

Figure 5. Effect of charge ratio on DNA-binding affinity of the aminoglycoside-polyamine library. One standard deviation is reported by the error bars in the plot. Effect of amine and charge content on DNA-binding efficacy of aminoglycoside-polyamine library. Percent fluorescence decreased values were normalized by the amine and charge contents and compared across the library.

affinity of kanamycin-tetrasepermine (**17b**) was comparable to that of polymeric agents.

To investigate the effects of charge on polyamine−DNA binding, the percent fluorescence decreased data was compared across the library candidates as a function of the charge ratio (Figure 5a). Clearly, if the cationic content alone were responsible for high DNA-binding efficacies, the trend in Figure 5a would have shown a linear increase with charge ratio. However, it can be seen that there is variability in the DNA-binding efficacy across different charge ratios. For example, adenosine-trispermine (**11**, Figure 2) showed a percent fluorescence decreased value of only 32.3% at a charge ratio of 2.39. On the other hand, neamine-trispermine (**17c**) showed a percent fluorescence decreased value of 64.6% at a similar charge ratio (charge ratio = 2.41). Furthermore, kanamycin-trispermine (**25**, Figure 4) showed a percent fluorescence decreased value of 66% at a slightly lower charge ratio (charge ratio = 2.17). Thus, although these molecules had the same number of spermine chains (three), their DNA-binding affinities were significantly different. In addition, neomycin (charge ratio = 1.10) had a percent fluorescence decreased value of 14.7%, while kanamycin-triethylenediamine (**24**, Figure 4) had a percent fluorescence decreased value of 32.2% at a comparable charge ratio (1.29). It is presumed that the flexibility of the amine groups due to the ethylenediamine derivatization on the kanamycin core led to enhanced affinity of this molecule.

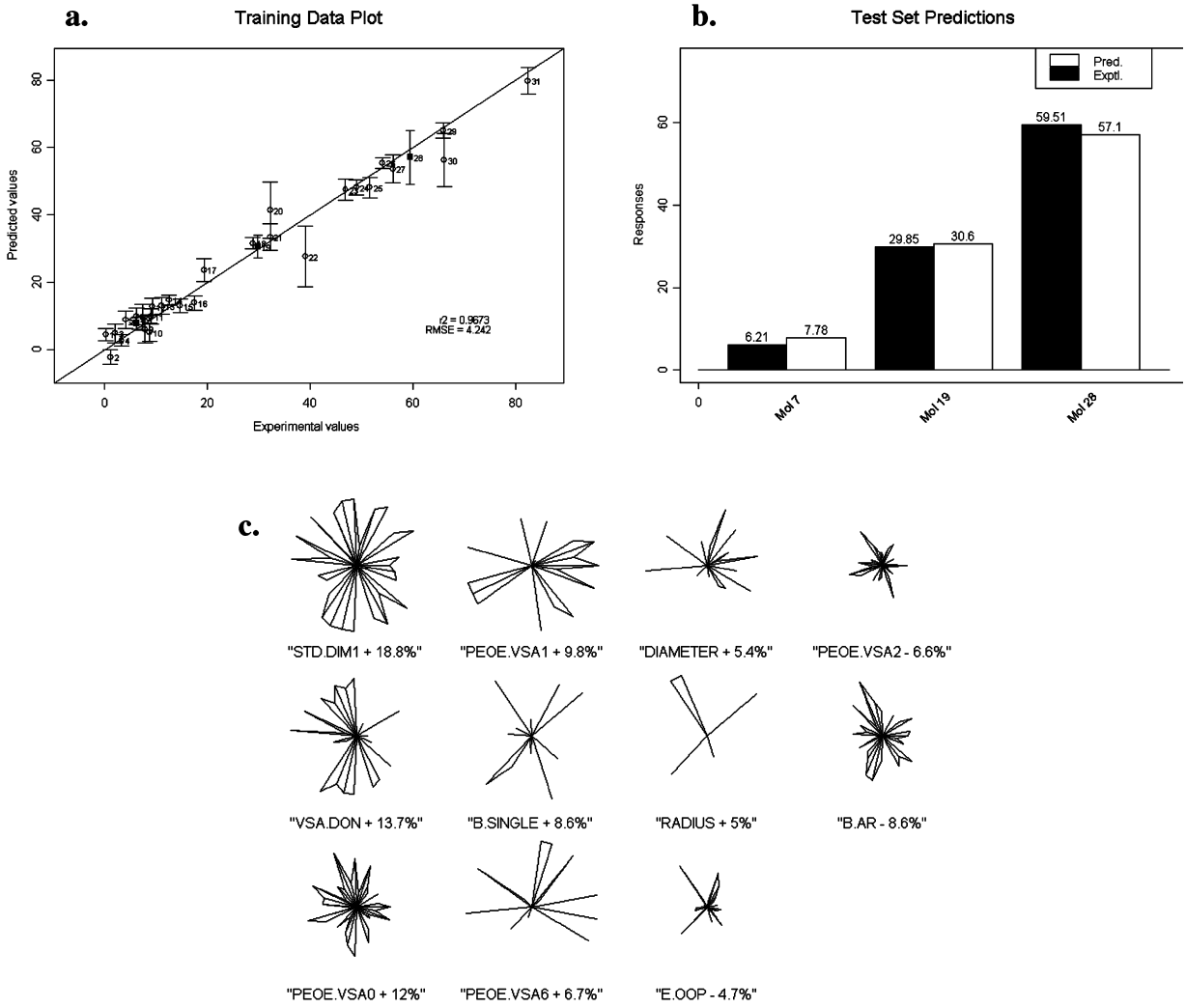

Figure 6. SVM models of percent fluorescence decreased. (a) Training model, (b) test set predictions, (c) star plot visualization of important descriptors. In the figure, mol 7: bekanamycin (Figure 4, molecule **21**), mol 19: **17a** (Figure 3), mol 26: **4b** (Figure 1).

To further investigate the effect of molecular structure on the DNA-binding ability of the library components, the percent fluorescence decreased values of the library candidates were normalized by the charge ratio of the respective candidates. Figure 5b shows a comparison of these normalized values. If the cationic content alone were responsible for the enhanced affinity of the library molecules, similar values would be obtained for all molecules upon normalization with charge ratio. The higher normalized percent fluorescence decreased values for some of the library candidates indicate that DNA-binding efficacy is not governed by charge content alone. This further demonstrates that differences in the chemistry and location of the substituted moieties can result in different DNA-binding efficacies of polyamines.

Quantitative Structure−Activity Relationship Modeling. *QSAR Model Generation.* QSAR models were generated to provide further insight into the physicochemical effects governing polyamine−DNA binding and to enable a priori prediction of the binding affinity of new polyamines using the SVM modeling procedure described above. Figure 6 shows the SVM model for percent fluorescence decreased. As seen in the figure, the cross-validated r^2 value of the training set was 0.97 (Figure 6a), and the test-set predictions were in good agreement with the experimentally observed values (Figure 6b). Furthermore, to decouple the effects of size and charge, QSAR models were also developed for the percent fluorescence decreased normalized with respect to the charge ratio. Figure 7 shows the SVM modeling results for the normalized response. As seen in the figure, the training model had a high cross-validated r^2 value of 0.95 (Figure 7a), and the test-set predictions agreed well with the experimentally determined values (Figure 7b). The results presented in Figures 6 and 7 are significant in that the library covers a wide chemical space, with molecules increasing in complexity from ethylenediamine to neamine-tetraspermine (**17b**). To our knowledge, this is the first report of QSAR models that enable the a priori prediction of DNA-binding efficacy for a diverse library of polyamines.

QSAR Model Interpretation. As described in the QSAR Modeling Procedure section, feature selection was carried out to determine a subset of molecular descriptors that showed high correlations with the experimental data. The important molecular descriptors/features selected in the QSAR models are described in Table 2. As seen from the star plot obtained for the QSAR models of the percent fluorescence decreased (Figure 6c), descriptors associated with the size of the molecule (STD.DIM1, B.SINGLE, and DIAMETER) were the highest contributors to the DNA-binding efficacy. STD.DIM1, which is defined as the square

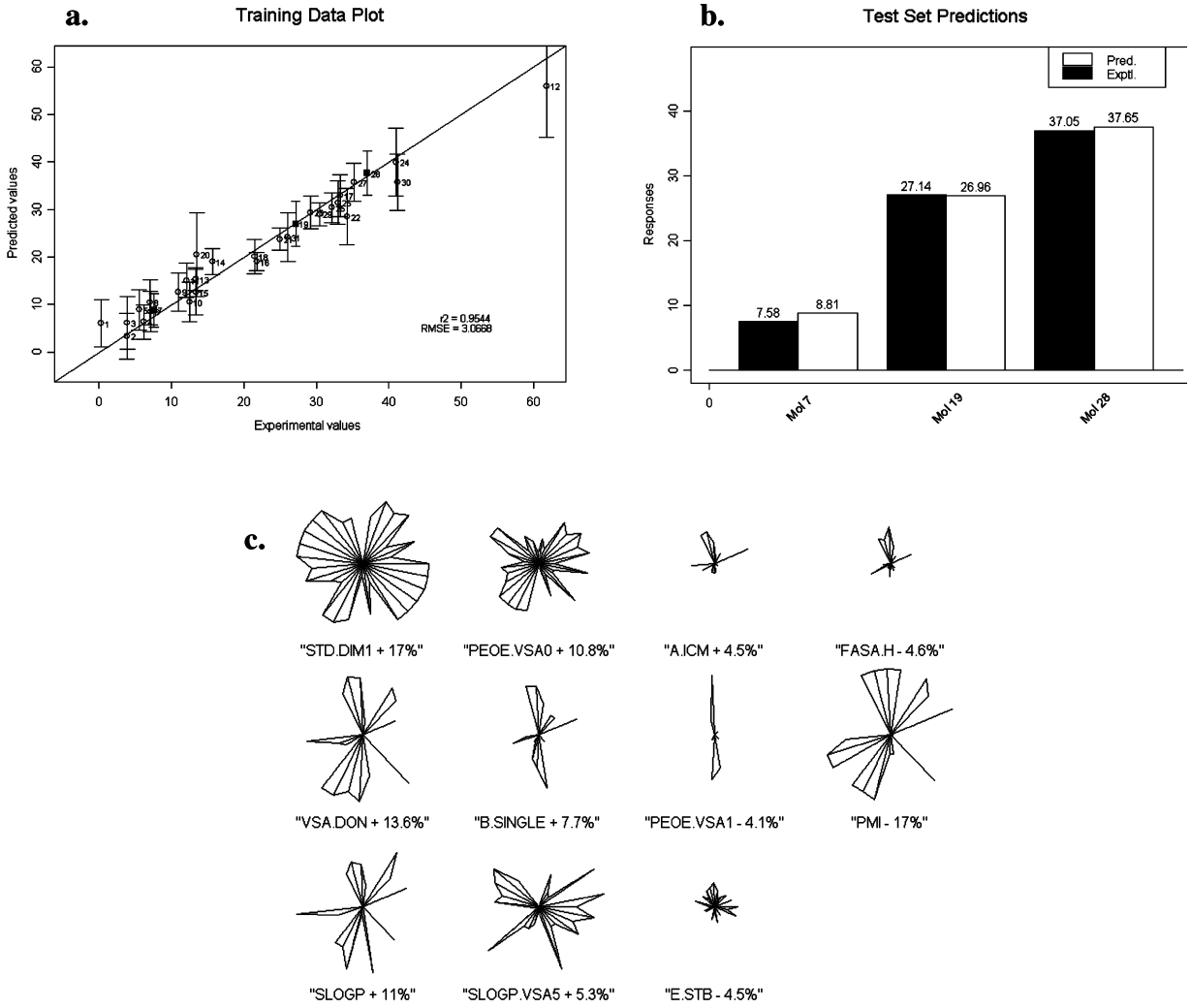

Figure 7. SVM models of percent fluorescence decreased/charge ratio. (a) Training model, (b) test set predictions, (c) star plot visualization of important descriptors. In the figure, mol 7: bekanamycin (Figure 4, molecule **21**), mol 19: **17a** (Figure 3), mol 26: **4b** (Figure 1).

Table 2. Definition of the Descriptors Used for Modeling Polyamine−DNA Binding

descriptor name	physicochemical information encoded in these descriptors/definition
A.ICM	atom information content (mean)—This is the entropy of the element distribution in the molecule (including implicit hydrogens but not lone pair pseudoatoms).
B.AR	number of aromatic bonds.
B.SINGLE	number of single bonds (including implicit hydrogens)—Aromatic bonds are not considered to be single bonds.
DIAMETER	largest value in the distance matrix[43]
E.OOP	out-of-plane potential energy
E.STB	bond stretch−bend cross-term potential energy
FASA.H	fractional water accessible surface area of all hydrophobic atoms
PEOE.VSA0, PEOE.VSA1	Sum of VDW surface area where partial charge of atoms is in the range [0.00,0.05) and [0.05,0.10), respectively.
PEOE.VSA2	Sum of VDW surface area where partial charge of atoms is in the range [0.10,0.15).
PEOE.VSA6	Sum of VDW surface area where partial charge of atoms is greater than 0.3.
PMI	principal moment of inertia
RADIUS	If r_i is the largest matrix entry in row i of the distance matrix **D**, then the radius is defined as the smallest of the r_i.[43]
RGYR	radius of gyration
SLOGP	log of the octanol/water partition coefficient (including implicit hydrogens)
SLOGP.VSA5	Sum of VDW surface area such that the contribution to logP(o/w) of the atom(s) is in the range (0.15,0.2].[44]
STD.DIM1	The square root of the largest eigenvalue of the covariance matrix of the atomic coordinates—A standard dimension is equivalent to the standard deviation along a principal component axis. This is the general shape descriptor.
VSA.DON	sum of VDW surface areas of pure hydrogen bond donors (not counting basic atoms and atoms that are both hydrogen bond donors and acceptors such as −OH)

root of the largest eigenvalue of the covariance matrix of the atomic coordinates, provides a representation of the distribution of atoms along the principal coordinate axis of the molecule. An interpretation of the information carried by this descriptor is that it represents the degree of concentration of atomic coordinates near the center of a molecule and consequently encodes both molecular size and shape. For example, a high value of STD.DIM1 would be

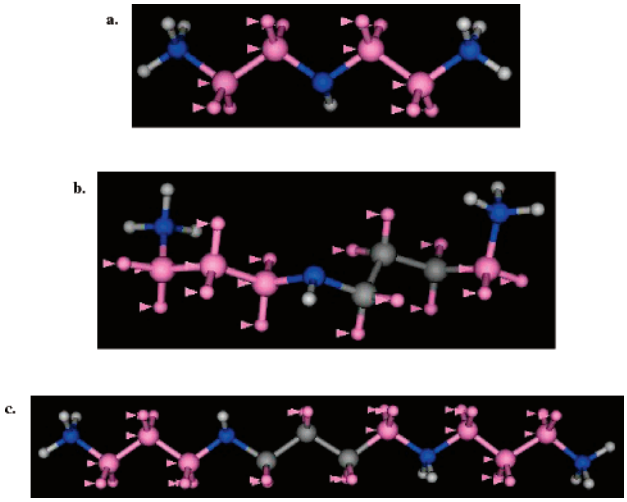

Figure 8. Molecular visualization of regions corresponding to the descriptors PEOE.VSA0 (shown by arrows) on (a) diethylenetriamine, (b) spermidine, and (c) spermine.

seen in an elongated or bar-shaped molecule, whereas a lower value would be found for a spheroid structure. This descriptor emerged as the most important contributor in the QSAR models for both the percent fluorescence decreased and the normalized response. In addition, the number of single bonds in the molecule (B.SINGLE) was a high positive contributor in both models. This descriptor is thought to encode size information, in that the number of bonds in a molecule increases monotonically as its size increases. DIAMETER, which is a measure of the diameter of the molecule at its widest point, was also observed to be an important descriptor in the model for percent fluorescence decreased. The positive contribution of these size-based descriptors suggests that as the size of the polyamine increases, its DNA-binding efficacy increases—a trend that is to be expected for polyamine DNA-binding ligands.

Interestingly, the descriptors PEOE.VSA0 and PEOE.VSA1 that represent the van der Waals surface area of a molecule associated with very low positive partial charge values were found to be high positive contributors in both models (either collectively or individually). Molecular surface visualization revealed that this range of positive partial charge was associated mainly with the methylene groups between amine moieties (representative depictions are shown for spermine, spermidine, and diethylenetriamine in Figure 8). As seen in the figure, although both spermidine and diethylenetriamine have three amines and two positive charges each, the values of the PEOE.VSA0 and PEOE.VSA1 descriptors were found to be significantly higher for spermidine than for diethylenetriamine due to higher methylene spacing between the amine groups (three-methylene spacing in spermidine vs two-methylene spacing in diethylenetriamine). Methylene group spacing between the amine moieties is apparently critical for regulating the charges on the amines through the screening of inductive effects between adjacent nitrogen atoms. This may explain why molecules with a three- and four-methylene spacing between the amine centers (e.g., spermine and its derivatives) have a greater DNA-binding efficacy than molecules with a two-methylene spacing between the amine centers (e.g., diethylenetriamine and its derivatives). This result is particularly significant because it is consistent with previous findings in the literature on the role of the methylene spacing between the amine centers on polyamine−DNA binding (for example, ref 42).

As seen in the star plots, VSA.DON (the sum of the van der Waals surface area of pure hydrogen bond donors as opposed to donor/acceptors) was also a high positive contributor in both QSAR models. This suggests that hydrogen-bond donor groups enhance the DNA-binding ability of polyamines. In addition, descriptors associated with positive charge (PEOE.VSA6) or the lack of it thereof (PEOE.VSA2) were observed to be important contributors to the percent fluorescence decreased model. PEOE.VSA6, which represents the van der Waals surface area with high positive partial charge, was found to be associated with charged amine groups on the molecules. The positive contribution of this descriptor is to be expected since a higher positive charge would increase a molecule's efficacy toward binding with the negatively charged DNA. Conversely, the PEOE.VSA2 descriptor associated with uncharged amine moieties emerged as a negative contributor in the model suggesting that the presence of uncharged amine groups in this homologous series of polyamines decreases the DNA-binding efficacy. Finally, the number of aromatic bonds in the molecules (B.AR) had the highest negative contribution in the percent fluorescence decreased model, suggesting that an apparent increase in the aromatic character of the molecules decreases their DNA-binding affinity. Since the aromatic character represented in the data set is mostly associated with low affinity 2′-deoxyadenosine scaffolds (Figure 2), the negative contribution to DNA binding attributed to this source may result from the paucity of sites for amine chain attachment on these scaffolds relative to other scaffolds used in the study (e.g., neomycin).

The star plots for the QSAR model of the normalized response would be expected to contain minimal contributions from charge-related descriptors since the dependent variable was normalized with respect to charge. As seen in Figure 7c, molecular size (represented by STD.DIM1) was again the most important positive contributor followed by VSA.DON and PEOE.VSA0 (discussed earlier). In addition, hydrophobicity-related descriptors (SLOGP and SLOGP.VSA5) also emerged as positive contributors to the model. While SLOGP is the log of the octanol/water partition coefficient, SLOGP.VSA5 is related to the van der Waals surface on the molecule associated with an intermediate range of hydrophobicity. The presence of these hydrophobicity-related descriptors may be a reflection of the fact that as the size of the polyamines increased, the number of carbon atoms and thus the hydrophobicity of the molecules also increased. Therefore, in this case, hydrophobicity may be serving as a surrogate for size-related information. As expected, descriptors associated with positive partial charge/basicity of the molecules were not important contributors in the model for the normalized response. Among the negative contributors to this model, the principal moment of inertia (PMI) had the highest value. This suggests that heavier aminoglycoside cores (e.g., neomycin vs glucosamine) with resulting lower values of PMI are favorable for DNA binding, in part due to their ability to host multiple polyamine substituents, resulting in higher net values of positive charge.

In summary, the star plot for the percent fluorescence decreased model indicated that the size and shape of a

molecule played a major role in its DNA-binding efficacy. An elongated or rod-shaped structure of the polyamines was found to be the most efficient shape for binding DNA. In addition, the presence of hydrogen bond donors and positive charges on the molecules was observed to enhance their DNA-binding affinity, while hydrophobic/aromatic groups on the molecules decreased their binding to DNA. Finally, methylene group spacing between the amine centers also emerged as an important parameter in determining the DNA-binding efficacy of the polyamines examined. From the star plot of the QSAR model for the charge-normalized response, additional insights related to a preferred aminoglycoside core structure, increased binding with larger molecules, and a tendency for higher affinity with elongated structures were obtained. While many of these results may be expected for polyamine DNA-binding systems, it is important to remind the reader that the feature selection process begins with a large set of molecular descriptors. The SVM modeling approach then selects those features which are most highly correlated with the dependent variable (e.g., percent fluorescence decreased). It is significant that the final QSAR models were predictive and were indeed able to capture the appropriate physics of the polyamine−DNA-binding process, thus affirming the validity of this modeling approach.

CONCLUSIONS

In conclusion, we have successfully evaluated an aminoglycoside-polyamine library for potential DNA-binding activity using a parallel screen and QSAR models of the screening data. The percent fluorescence decreased values from the ethidium bromide screen were employed to generate QSAR models based on Support Vector Machine (SVM) regression techniques. The training models had high correlation values ($r^2 > 0.95$), and the predictions of external test-set molecules were in good agreement with experimental values. The molecular descriptors selected in the QSAR models were employed for a systematic investigation into structure−property effects influencing polyamine−DNA binding. Descriptor analysis using star plots indicated that increases in molecular size, hydrogen-bond donor groups, positive charge, methylene group spacing between amine centers, and a rod-shaped geometry of the polyamines enhanced their DNA-binding efficacy. To our knowledge, this is the first report of QSAR models that enables the a priori prediction of DNA-binding efficacy of a diverse polyamine library. It is significant that the QSAR models were indeed able to capture the appropriate physics of the polyamine−DNA-binding process, thus affirming the validity of the modeling approach. Significantly, these results justify the rationale of grafting multiple copies of polyamines on larger aminoglycoside cores for enhanced affinity making the case for multivalent ligands. The research described in this paper has led to the development of new ligands with high DNA-binding activity and improved our understanding of structure−activity relationships involved in polyamine−DNA binding. This work may have implications for the discovery of future generations of polyamine ligands for nonviral gene delivery, anticancer therapeutics, and plasmid DNA purification.

ACKNOWLEDGMENT

The authors acknowledge financial support from NSF (Grant BES-0079436), NIH (Grant GM 47372), and Rensselaer Seed Grant (Grant 151140).

REFERENCES AND NOTES

(1) Wallace, H. M.; Fraser, A. V.; Hughes, A. A perspective of polyamine metabolism. *Biochem. J.* **2003**, *376*, No. 1, 1−14.
(2) Newton, G. L.; Aguilera, J. A.; Ward, J. F.; Fahey, R. C. Polyamine-induced compaction and aggregation of DNA-a major factor in radioprotection of chromatin under physiological conditions. *Radiat. Res.* **1996**, *145*, No. 6, 776−780.
(3) Khan, A. U.; Mascio, P. D.; Medeiros, M. H. G.; Wilson, T. Spermine and Spermidine Protection of Plasmid DNA Against Single-Strand Breaks Induced by Singlet Oxygen. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 11428−11430.
(4) Ha, H. C.; Sirisoma, N. S.; Kuppusamy, P.; Zweier, J. L.; Woster, P. M.; Casero, P. M., Jr. The natural polyamine spermine functions directly as a free radical scavenger. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11140−11145.
(5) Chattopadhyay, M. K.; Tabor, C. W.; Tabor, H. Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2261−2265.
(6) Thomas, T. J.; Faaland, C. A.; Gallo, M. A.; Thomas, T. Suppression of c-myc oncogene expression by a polyamine-complexed triplex forming oligonucleotide in MCF-7 breast cancer cells. *Nucleic Acids Res.* **1995**, *23*, 3594−3599.
(7) Thomas, T.; Balabhadrapathruni, S.; Gallo, M. A.; Thomas, T. J. Development of polyamine analogues as cancer therapeutic agents. *Oncol. Res.* **2002**, *13*, No. 3, 123−135.
(8) Criss, W. E. A review of polyamines and cancer. *Turk. J. Med. Sci.* **2003**, *33*, No. 4, 195−205.
(9) Srivenugopal, K. S.; Wemmer, D. E.; Morris, D. R. Aggregation of DNA by analogues of spermidine; enzymatic and structural studies. *Nucleic Acids Res.* **1987**, *15*, 2563−2580.
(10) Osland, A; Kleppe, K. Polyamine induced aggregation of DNA. *Nucleic Acids Res.* **1977**, *4*, 685−695.
(11) Koltover, I.; Wagner, K.; Safinya, C. R. DNA condensation in two dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14046−14051.
(12) Pelta, J.; Livolant, F.; Sikorav, J. L. DNA aggregation Induced by Polyamines and Cobalthexamine. *J. Biol. Chem.* **1996**, *271*, No. 10, 5656−5662.
(13) Basu, H. S.; Marton, L. J. The interaction of spermine and pentamines with DNA. *Biochem. J.* **1987**, *244*, No. 1, 243−246.
(14) Raspaud, E.; Chaperon, I.; Leforestier, A.; Livolant, F. Spermine-Induced Aggregation of DNA, Nucleosome and Chromatin. *Biophys. J.* **1999**, *77*, No. 3, 1547−1555.
(15) Deng, H.; Bloomfield, V. A.; Benevides, J. M.; Thomas, G. J., Jr. Structural basis of polyamine-DNA recognition: spermidine and spermine interactions with genomic B-DNAs of different GC content probed by Raman spectroscopy. *Nucleic Acids Res.* **2000**, *28*, No. 17, 3379−3385.
(16) Saminathan, M.; Antony, T.; Shirahata, A.; Sigal, L H.; Thomas, T.; Thomas, T. J. Ionic and structural specificity effects of natural and synthetic polyamines on the aggregation and resolubilization of single-, double-, and triple-stranded DNA. *Biochemistry* **1999**, *38*, No. 12, 3821−3830.
(17) Vijayanathan, V.; Thomas, T.; Sriharata, A.; Thomas, T. J. DNA condensation of polyamines: A laser light Scattering Study of Structural Effects. *Biochemistry* **2001**, *40*, No. 45, 13644−13651.
(18) Geall, A. J.; Blagbrough, I. S. Rapid and sensitive ethidium fluorescence quenching assay of polyamine conjugate-DNA interactions for the analysis of lipoplex formation in gene therapy. *J. Pharm. Biomed. Anal.* **2000**, *22*, No. 5, 849−859.
(19) Blagbrough, I. S.; Al-Hadithi, D.; Geall, A. J. Cheno-, urso- and deoxycholic acid spermine conjugates: relative binding affinities for Calf thymus DNA. *Tetrahedron* **2000**, *56*, No. 21, 3439−3447.
(20) Geall, A. J.; Taylor, R. J.; Earll, M. E.; Eaton, M. A. W.; Blagbrough, I. S. Synthesis of Cholesteryl Polyamine Carbamates: pK_a studies and condensation of Calf thymus DNA. *Bioconjugate Chem.* **2000**, *11*, 314−326.
(21) Boger, D. L.; Brian, F. E.; Brunette, S. R.; Tse, W. C.; Hedrick, M. P. A simple high resolution method for establishing DNA binding affinity and sequence selectivity. *J. Am. Chem. Soc.* **2001**, *123*, No. 25, 5878−5891.
(22) Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. A Quantitative Structure−Activity Relationship and Molecular Graphics Study of Carbonic-Anhydrase Inhibitors. *Mol. Pharmacol.* **1985**, *27*, No. 5, 493−498.

(23) Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of human intestinal absorption of drug compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, No. 4, 726−735.
(24) Hall, L. H.; Maw, H. H. E-state modeling of HIV-1 protease inhibitor binding independent of 3D information. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (2), 290−298.
(25) Adam, K. R. New density functional and atoms in molecules method of computing relative p*K*(a) values in solution. *J. Phys. Chem. A* **2002**, *106*, No. 49, 11963−11972.
(26) Holm, R.; Hoest, J. Successful in silico predicting of intestinal lymphatic transfer. *Int. J. Pharm.* **2004**, *272*, No. 1−2, 189−193.
(27) Katritzky, A. R.; Oliferenko, P.; Oliferenko, A.; Lomaka, A.; Karelson, M. Nitrobenzene toxicity: QSAR correlations and mechanistic interpretations. *J. Phys. Org. Chem.* **2003**, *16*, No. 10, 811−817.
(28) Karelson, M.; Fitch, W. L.; McGregor, M.; Katritzky, A. R.; Lomaka, A.; Petrukhin, R. Prediction of ultraviolet spectral absorbance using quantitative structure−property relationships. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (4), 830−840.
(29) Guertin, J. A.; Li, J. W.; Masso, J. J. Prediction of drug solubility in an acrylate adhesive based on the drug-polymer interaction parameter and drug solubility in acetonitrile. *J. Contr. Relat.* **2002**, *83* (2), 211−221.
(30) Kaliszan, R. Quantitative Structure-Retention Relationships Applied To Reversed-Phase High-Performance Liquid-Chromatography. *J. Chromatogr. A* **1993**, *656* (1−2), 417−435.
(31) Kaliszan, R.; Baczek, T. Quantitative structure/retention relationships in affinity chromatography. *J. Biochem. Biophys. Methods* **2001**, *49* (1−3), 83−98.
(32) Mazza, C. B.; Sukumar, N.; Breneman, C. M.; Cramer, S. M. Prediction of Protein Retention in Ion-Exchange Systems using Molecular Descriptors Obtained from Crystal Structure. *Anal. Chem.* **2001**, *73*, No. 22, 5457−5461.
(33) Song, M.; Breneman, C. M.; Bi, J.; Sukumar, N.; Bennett, K. P.; Cramer, S. M.; Tugcu, N. Prediction of Protein Retention Times in Anion-Exchange Chromatography Systems Using Support Vector Regression. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1347−1357.
(34) Ladiwala, A.; Rege, K.; Breneman, C. M.; Cramer, S. M. Investigation of Mobile Phase Salt Type Effects on Protein Retention and Selectivity in Cation Exchange Systems using Quantitative Structure-Retention Relationship Models. *Langmuir* **2003**, *19* (20), 8443−8454.
(35) Mazza, C. B.; Rege, K.; Breneman, C. M.; Sukumar, N.; Dordick, J. S.; Cramer, S. M. High Throughput Screening and Quantitative Structure-Efficacy Relationship Models of Potential Displacer Molecules for Ion-Exchange Systems. *Biotechnol. Bioeng.* **2002**, *80*, No. 1, 60−73.
(36) Rege, K.; Ladiwala, A.; Tugcu, N.; Breneman, C. M.; Cramer, S. M. Parallel Screening of Selective and High-Affinity Displacers for Proteins in Ion-Exchange Systems. *J. Chromatogr. A* **2004**, *1033*, 19−28.
(37) Vapnik, V. *The Nature of Statistical Learning Theory*; John Wiley: New York, 1996.
(38) Vapnik, V. N. An overview of statistical learning theory. *IEEE Trans. Neural Networks* **1999**, *10*, No. 5, 988−999.
(39) Bennett, K. P. On Mathematical Programming Methods and Support Vector Machines. In *Advances in Kernel Methods − Support Vector Machines*; Schölkopf, A., Burges, C., Smola, A., Eds., MIT Press: Cambridge, MA, 1998; pp 307−326.
(40) Bi, J.; Bennett, K.; Embrechts, M. J.; Breneman, C. M.; Song, M. Dimensionality Reduction via Sparse Support Vector Machines. *J. Mach. Learn. Res.* **2003**, *3*, 1229−1243.
(41) Rege, K.; Hu, S.; Moore, J. A.; Dordick, J. S.; Cramer, S. M. Chemoenzymatic Synthesis and Parallel Screening of an Aminoglycoside-Polyamine Library: Identification of High Affinity Displacers and DNA-Binding Ligands. *J. Am. Chem. Soc.* **2004**, accepted for publication.
(42) van Dam, L.; Korolev, N.; Nordenskiold, L. Polyamine-Nucleic Acid Interactions and the Effects of Structure in Oriented DNA-fibers. *Nucleic Acids Res.* **2002**, *30* (2), 419−428.
(43) Petitjean, M. Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331−337.
(44) Wildman, S. A.; Crippen, G. M. Prediction of Physiochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, No. 5, 868−873.

CI050082G