Novel Method for Structure–Activity Relationship of Aptamer Sequences for Human Prostate Cancer

Xinliang Yu,†‡ Huiqiong Yang,† and Xianwei Huang†

†College of Chemistry and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China
‡State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan University, Changsha, Hunan 410082, China

Supporting Information

ABSTRACT: Prostate cancer (PCa) is one of the most common malignancies in men and seriously threatens men’s health. Developing aptamer probes for PCa cells is of great significance for early diagnosis and treatment of PCa. This paper reports a classification model for SELEX-based aptamers, which were obtained with PCa cell line PCa-3M-1E8 (highly metastatic tumor cell) as target cells and PCa cell line PCa-3M-2B4 (low metastatic tumor cell) as control cells. On the basis of the SELEX principle, 100 oligonucleotide sequences from the 3rd round of SELEX were defined as low affinity and specificity aptamers, and 100 sequences from the 11th round were set as high affinity and specificity aptamers. Seven molecular descriptors were used for the classification model, which were calculated from amino acid sequences translated from DNA aptamer sequences with DNAMAN software. The classification model based on binary logical regression analysis has prediction accuracies, sensitivity, and specificity of about 80% for both the training set and test set. Therefore, it is feasible to calculate molecular descriptors from amino acid sequence translated from DNA aptamer sequences and develop a classification model for PCa cell line PCa-3M-1E8.

INTRODUCTION

Aptamers are short single-stranded DNA (ssDNA) (or RNA) oligonucleotides. They can fold into distinct two-dimensional (2D) and three-dimensional (3D) structures and are capable of binding targets, such as proteins, cells, and viruses, with high affinity and specificity.1 Compared to natural antibodies, aptamers possess several advantages in synthesis and modification. In addition, aptamers are small in sizes and display good binding affinity to targets, which make them easily penetrate tissue and tumor, and possess fast blood clearance.2–3

Aptamers are promising in therapeutic applications, such as cancer cell detection and diagnostics, as well as in molecular tools in the areas of anti-infective, anticoagulation, anti-inflammation, antiangiogenesis antiproliferation, and immune therapies.4–5 DNA aptamers have been developed for the virus-infected cells, mesenchymal stem cells, porcine endothelial precursor cells, and live bacterial cells.6–7 At present, an aptamer (brand name Macugen) has been used for the treatment of neovascular age-related macular degeneration. In addition, some aptamers are currently in clinical trials.8

Aptamers can be acquired through an in vitro iterative selection process designated as the systematic evolution of ligands by an exponential enrichment (SELEX) technique.1,3 This is a reiterative process as follows: the starting oligonucleotide library with approximately 1014–1015 unique sequences is incubated with the targets of interest by an affinity binding method, the bound aptamer–target complexes are partitioned from the unbound oligonucleotides, the unbound oligonucleotides are discarded, the bound aptamers are eluted in from the targets and amplified by polymerase chain reaction (PCR), and the amplified aptamers are incubated with the targets in the next round of SELEX. After several rounds of selection, aptamers with high affinity and specificity will be obtained.

However, the above selection processes are lengthy and need relatively higher consumption in reagents. The quantitative structure–activity relationship (QSAR) (e.g., pattern recognition model or classification model) method can reduce aptamer development costs and accelerate the process of screening candidate sequences. The main objective of the QSAR methodology is to find a mathematical relationship between chemical structures and the activity of interest. To accomplish this purpose, one usually employs a mathematical model (F)

$$activity = F(\text{molecular descriptors}) + \text{error}$$  

(Equation 1) correlates experimental activity values with a set of molecular descriptors, including physiochemical properties and structural properties derived from the molecular structures. Once a reliable QSAR model is developed, this model can be used to predict the activities of new chemicals. Thus, QSAR can be used to understand the mechanism, design new compounds, and screen candidates.9 Then, the most-promising candidate may be synthesized for laboratory testing.

Received: June 26, 2018
Accepted: August 20, 2018
Published: August 28, 2018
A few groups have carried out QSAR studies for aptamers. Yu et al. introduced a classification model for predicting streptavidin-binding aptamers that may be “low”- or “high”-affinity aptamers. The model possesses a classification accuracy of 97.99% for the training set. Furthermore, the prediction fractions of winning aptamers from the 1st round to the 10th round of SELEX consist of the enrichment characteristics of aptamer based on SELEX selection. In addition, Yu et al. developed another classification model for the binding affinity of aptamers of hepatocarcinoma cell line SMMC-7721. The classification model has good statistical qualities with specificity and sensitivity greater than 80%. For these two classification models, some molecular descriptors used in the models were calculated from the loop structures of center sequences. Li et al. analyzed aptamer−target pairs by applying sequence information from DNA (or RNA) aptamers and their target proteins. The predictive accuracies of the model for the training set and test set are, respectively, 81.34 and 77.41%. Musa et al. introduced a novel QSAR model for the binding affinity of anti-influenza aptamers. The model is accurate with regression coefficients $R^2$ being 0.702 for the training set and 0.66 for the test set, though the descriptors are based on topological structures, which cannot describe three-dimensional information, such as the electronic and geometric properties.

Despite aptamers having found an increasingly wide utilization, only a few QSAR studies have been reported. There are two reasons for this phenomenon. One is that experimental data of aptamers are insufficient. The other is that accurately calculating molecular descriptors from aptamers is difficult because of their larger molecular weight. In this work, we adopt a novel method to calculate molecular descriptors from aptamers against prostate cancer (PCa) cells that were selected by our group with the cell-SELEX method and develop a pattern recognition model for PCa aptamers. PCa is one of the most common cancers nowadays in men and, as the second leading cause of cancer death, seriously threatens men’s health. A deficiency of early clinical diagnosis of PCa can lead to an estimated 270 thousand men dying from PCa in one year in the United States alone. The development of aptamer probes for PCa is urgently required. This work will help the SEXLEX selection of aptamer sequences for PCa.

## RESULTS AND DISCUSSION

Table S1 in the Supporting Information shows 200 center sequences of candidate aptamer (5′-AGAAGGAAGAGGAGGAGGAGGACAC-N₄₀-TATCAGTGGTCGTCGTCAT-3′), with PCa cell line PCa-3M-1E8 acting as target cells and PCa cell line PCa-3M-2B4 being the negative control cells in SELEX selection. Sequence nos. 1–67 and 135–167 were selected from the 3rd round of SELEX, while sequence nos. 68–134 and 168–200 were obtained from the 11th round of selection. These sequences were randomly divided into two sets: a training set and a test set. Sequence nos. 1–134 from the training set were used to build classification models. The remaining sequences (nos. 135–200) in the test set were used for evaluating the predictive power of classification models. Candidate aptamer sequences in Table 1 were selected from the 12th round of SELEX selection. These sequences obtained from the 3rd round of SELEX have low affinity and specificity for PCa-3M-1E8 target. Their class labels were defined as 2.

After calculating 1000 molecular descriptors for each optimized amino acid sequence that was translated from DNA aptamer sequences with DNAMAN software, binary logical regression analysis was performed for selecting the optimal subset of descriptors, by applying Forward Wald method in IBM SPSS Statistics 19 for the total data set (200 sequences).
sequences in Table S1. Classification table, variables in the classification models, and their definitions of molecular descriptors are listed in Tables 2, 3, and 4, respectively. Table 2 shows that only classification model 7 has accuracies above 80%. Table 3 suggests that all the variables in the classification model 7 possess sig. values smaller than 0.05, which indicate that these independent variables have meaningful effect on the dependent variable. Thus, the seven molecular descriptors in classification model 7 were taken as the optimal descriptor subset and used to develop logical regression eq 2 from the training set in Table S1.

$$P(Y) = \frac{1}{1 + e^{-\left(40.368 - 29.927E_{\text{ED}} + 0.736S_{\text{VV}} + 95.782C_{\text{WP}} + 0.407F_{\text{CO}} - 1.699F_{\text{OS}} + 1.060F_{\text{S}} + 0.352F_{\text{NN}}\right)}}$$  

(2)

Here, we take the calculation of probable value for sequence no. 200 in Table S1 as an example. Its descriptor values for $E_{\text{ED}}$, $S_{\text{VV}}$, $C_{\text{WP}}$, $F_{\text{CO}}$, $F_{\text{OS}}$ and $F_{\text{NN}}$ are, respectively, 1.963, −13.483, 0.139, 28, 2, 3, and 8. From eq 2, we can obtain $P(Y) = 1/(1 + e^{-0.994}) = 0.730$. Because the $P(Y)$ is above 0.5, the class label of sequence no. 200 is “2”.

Subsequently, we adopted eq 2 to predict the class labels of sequences in the test set in Table S1. The class labels calculated are listed in Table S1. Besides the statistical parameter accuracy (=true prediction/(true prediction + false prediction)), sensitivity (=true positives/(true positive + false negative)) and specificity (=true negatives/(true negative + false positives)) were used to describe the characteristics of prediction. The statistical results from eq 2 are listed in Table 5, which shows that the accuracy, sensitivity, and specificity of the training set and test set are about 80%. Further, eq 2 was used to predict class labels of sequences in Table 1 from the 12th round of SELEX. The calculation class labels are listed in Table 1. The prediction accuracy is 83.3%. Therefore, the prediction results are accurate.

The first descriptor $E_{\text{ED}}$ (eigenvalue n (=7) from the edge adjacency matrix weighted by dipole moment) is an edge adjacency index. $E_{\text{ED}}$ belongs to topological indices in nature and is calculated from the edge adjacency matrix of an H-depleted molecule. The square matrix is symmetric and has a dimension of $nBO \times nBO$ (here $nBO$ is the number of nonhydrogen atom pairs), whose entries equal one if the bonds are adjacent and zero for nonadjacent. $E_{\text{ED}}$ is calculated by applying the dipole moment for edge weighting schemes. Thus, $E_{\text{ED}}$ is related to the symmetry of molecules and polarity and possesses good discrimination among isomers.\textsuperscript{18}

The second descriptor $S_{\text{VV}}$ signal 05/weighted by van der Waals volume, belongs to 3D molecule representation of structures based on electron diffraction (3D-MoRSE) descriptors. They can be calculated with the following expression

$$M_{\text{sw}} = \sum_{i=1}^{nAT-1} \sum_{j=i+1}^{nAT} \sum_{\text{sr}} w_{ij} \sin\left(\frac{\text{sr}}{r_i}\right)$$  

(3)

where $M_{\text{sw}}$ denotes the scattered electron intensity, the parameter $s$ means the scattered intensity of atoms in the different directions varying from 0 to 31, $r_i$ represents the distance between atoms $i$ and $j$, $nAT$ denotes the number of atoms, and $w$ denotes an atomic property, including unweighted schemes and weighting properties, such as van der Waals volume and atomic polarizability. The descriptor $S_{\text{VV}}$ describes the scattering of electron intensity along the van der Waals volume. It can describe molecular similarity/diversity from three-dimensional structures.\textsuperscript{18}

The third descriptor, $C_{\text{WP}}$, is the 3rd component symmetry directional WHIM index/weighted by polarizability. WHIM descriptors are based on the configuration of atoms in the 3D space, which is defined by the Cartesian axes ($x$, $y$, and $z$). Molecular principal axes are evaluated to obtain a unique reference frame. Besides atomic projections being carried out along their respective principal axes, atomic dispersion and distribution around the geometric center are calculated. Similarly, six different weighting schemes (including unweighted schemes, atomic mass, the van der Waals volume, the Sanderson atomic electronegativity, the atomic polarizability, and the electrotopological state indices of Kier and Hall) are taken into account to obtain various WHIM descriptors. The descriptor $C_{\text{WP}}$ not only reflects molecular size and shape, but also describes molecular symmetry and atom distribution with respect to invariant reference frames.\textsuperscript{18}

The remaining topological molecular descriptors, $F_{\text{CO}}$ (frequency of C--O with the topological distance being 6), $F_{\text{OS}}$ (frequency of O--S with the distance being 7), $F_{\text{CS}}$ (frequency of C--S with the distance being 9), and $F_{\text{NN}}$ (frequency of N--N with the distance being 10), are binary and frequency atom pairs. In binary atom pairs, the value (1) or (0), respectively, expresses the presence or absence of a certain pair of atoms at a specific bond distance. As for the frequency of atom pairs, each molecular descriptor means the number of atom pairs that satisfy the above conditions of topological distance.

The main aim of QSAR studies is to develop correlations of molecular structures and their properties. However, aptamer
structures are related to the structure of amino acid sequences translated from their corresponding DNA aptamer sequences. This is the reason why the seven molecular descriptors present in eq 2 from amino acid sequences can describe the aptamer structures. In general, the QSAR prediction strategies depend on molecular descriptors calculated directly from molecular structures. However, our classification model is related to descriptors through an indirect route. Thus, our paper is novel and provides a new approach for the QSAR investigations of aptamers. Especially, this paper indicates that there is a correlation between the chemical structures of DNA aptamer sequences and amino acid sequences translated from DNA aptamer sequences.

Table 3. Variables in the Classification Models

| model | descriptor | $B$ | SE | Waals | df | sig. | exp($B$) |
|-------|------------|-----|----|-------|----|------|----------|
| model 1 | $F_{CO}$ | -0.125 | 0.027 | 20.919 | 1 | 0.000 | 0.883 |
| | constant | 4.242 | 0.936 | 20.538 | 1 | 0.000 | 69.544 |
| model 2 | $S_{NV}$ | -0.526 | 0.117 | 20.319 | 1 | 0.000 | 0.591 |
| | $F_{CO}$ | -0.221 | 0.038 | 34.229 | 1 | 0.000 | 0.802 |
| | constant | 0.169 | 1.267 | 0.018 | 1 | 0.894 | 1.184 |
| model 3 | $S_{NV}$ | -0.555 | 0.123 | 20.356 | 1 | 0.000 | 0.574 |
| | $F_{CO}$ | -0.228 | 0.039 | 34.630 | 1 | 0.000 | 0.796 |
| | $F_{CS}$ | -0.350 | 0.104 | 11.399 | 1 | 0.001 | 0.705 |
| | constant | 0.442 | 1.334 | 0.110 | 1 | 0.740 | 1.556 |
| model 4 | $E_{ED}$ | 23.734 | 13.191 | 3.237 | 1 | 0.072 | $2.031 \times 10^{10}$ |
| | $S_{NV}$ | -0.622 | 0.128 | 23.566 | 1 | 0.000 | 0.537 |
| | $F_{CO}$ | -0.280 | 0.046 | 37.076 | 1 | 0.000 | 0.756 |
| | $F_{CS}$ | -0.426 | 0.115 | 13.750 | 1 | 0.000 | 0.653 |
| | constant | -45.591 | 25.627 | 3.165 | 1 | 0.075 | 0.000 |
| model 5 | $E_{ED}$ | 42.590 | 15.877 | 7.196 | 1 | 0.007 | $3.139 \times 10^{10}$ |
| | $S_{NV}$ | -0.684 | 0.136 | 25.155 | 1 | 0.000 | 0.505 |
| | $F_{CO}$ | -0.0326 | 0.052 | 39.040 | 1 | 0.000 | 0.722 |
| | $F_{CS}$ | 1.196 | 0.381 | 9.863 | 1 | 0.002 | 3.306 |
| | $F_{NN}$ | -0.855 | 0.194 | 19.404 | 1 | 0.000 | 0.425 |
| | constant | -82.136 | 30.791 | 7.116 | 1 | 0.008 | 0.000 |
| model 6 | $E_{ED}$ | 36.783 | 16.405 | 5.027 | 1 | 0.025 | $9.431 \times 10^{15}$ |
| | $S_{NV}$ | -0.853 | 0.157 | 29.673 | 1 | 0.000 | 0.426 |
| | $F_{CO}$ | -0.362 | 0.058 | 39.332 | 1 | 0.000 | 0.696 |
| | $F_{CS}$ | 1.245 | 0.396 | 9.894 | 1 | 0.002 | 3.473 |
| | $F_{NN}$ | -0.901 | 0.201 | 20.087 | 1 | 0.000 | 0.406 |
| | constant | -70.938 | 31.822 | 7.211 | 1 | 0.007 | 0.787 |
| model 7 | $E_{ED}$ | 36.306 | 16.785 | 4.679 | 1 | 0.031 | $5.852 \times 10^{15}$ |
| | $S_{NV}$ | -0.838 | 0.161 | 27.062 | 1 | 0.000 | 0.433 |
| | $C_{WP}$ | 3.609 | 28.598 | 10.074 | 1 | 0.002 | 0.000 |
| | $F_{CO}$ | -0.404 | 0.064 | 30.162 | 1 | 0.000 | 0.667 |
| | $F_{CS}$ | 1.306 | 0.424 | 19.488 | 1 | 0.002 | 3.691 |
| | $F_{NN}$ | -0.995 | 0.216 | 21.164 | 1 | 0.000 | 0.370 |
| | constant | -55.171 | 32.756 | 7.837 | 1 | 0.092 | 0.000 |

Table 4. Definitions of Molecular Descriptors in Classification Models

| no. | symbol | definition | class |
|-----|--------|------------|-------|
| 1   | $E_{ED}$ | eigenvalue n (≥7) from edge adjacency matrix weighted by dipole moment | edge adjacency indices |
| 2   | $S_{NV}$ | signal OS/weighted by van der Waals volume | 3D-MoRSE descriptors |
| 3   | $C_{WP}$ | 3rd component symmetry directional WHIM index/weighted by polarizability | WHIM descriptors |
| 4   | $F_{CO}$ | frequency of C–O at topological distance 6 | 2D atom pairs |
| 5   | $F_{CS}$ | frequency of O–S at topological distance 7 | 2D atom pairs |
| 6   | $F_{NN}$ | frequency of C–S at topological distance 9 | 2D atom pairs |
| 7   | $F_{NN}$ | frequency of N–N at topological distance 10 | 2D atom pairs |

Table 5. Statistical Results from Logical Regression Equation 2

| data set | label | experiment | prediction | accuracy (%) |
|----------|-------|------------|------------|--------------|
| training set | class 1 | 67 | 53 | 14 | 79.1 |
| | class 2 | 67 | 12 | 55 | 82.1 |
| test set | class 1 | 33 | 27 | 6 | 81.8 |
| | class 2 | 33 | 6 | 27 | 81.8 |
developing binary logical regression equation for aptamers of PCa cell line PCa-3M-1E8 has been demonstrated. However, it should be noted that our method is based on the DNA sequences encoding protein’s information to realize the prediction for aptamer sequences. However, the method cannot be used to predict their specific confirmations, which are correlated with aptamer binding specificity. The reason is that the specific confirmations of aptamers depend on target properties. That is to say, an aptamer sequence may possess different confirmations when it is applied in various fields.

## CONCLUSIONS

Although directly calculating molecular descriptors from aptamers is difficult, the seven molecular descriptors $E_{ED}$, $S_{VV}$, $F_{CD}$, $C_{VPP}$, $F_{CG}$, $F_{CS}$, and $F_{SN}$ can describe the aptamer structures, which were calculated from amino acid sequence translated by DNA aptamer sequences with DNAMAN software. The classification model can be used to recognize candidate aptamers with high or low affinity and specificity to PCa cell line PCa-3M-1E8. The prediction accuracies, sensitivity, and specificity obtained from the training and test sets are about 80%. Therefore, the prediction results are accurate and acceptable. The investigation provides a novel approach to calculate descriptors for aptamer molecules.

## MATERIALS AND METHODS

### Cell-SELEX Experimental

Candidate aptamer sequences used in this paper were selected in our laboratory with the cell-SELEX technique. The forward primer (FP) labeled with 6-carboxyfluorescein (FAM) and the biotinylated reverse primer (RP) were, respectively, 5′-FAM-AAGAAGAAGAGACAC-GACAC-3′ and 5′-biotin-ATGACGACCGACCAGT-GATA-3′, which were used for the synthesis of the random ssDNA library, 5′-TTGACTTGGCCACTGACCTACCC-(randomized region with 40 ± 2 nucleotides)-GAAAGTCAGTCGGTCGT-CATC-3′. In addition, the primers (FP, 5′-AGAAGAAGAGACAC-GACAC-3′) and (RP, 5′-ATGACGACCGACCAGT-GATA-3′) were applied to clone and sequence the product sequences. These primer products and the initial ssDNA pool completely dissolved in 900 μL of binding buffer with the concentrations of 1 mg/mL for bovine serum albumin and 0.1 mg/mL for yeast tRNA) and 0.2 mol/L sodium hydroxide. The enriched ssDNA aptamer pools from each round of SELEX were used for the next round. After the fourth round of SELEX selection, the counterselection was introduced to increase the specificity of aptamers, which may differentiate similar PCa cells. In addition, the incubation and washing conditions were made stringent gradually to improve aptamer performances and SELEX screening efficiency. Other SELEX procedure and materials can be found in References 5,15 After 12 rounds of SELEX selection, the selection library reached a platform, which was monitored by using the flow cytometer (FACScalibur, BD Bioscience). The ssDNA pools from the 3rd and 11th rounds were sequenced by a high-throughput sequencing device. The resulting pool from the 12th round was sequenced with the classic method of cloning and sequencing.

According to aptamer evolutionary principles of SELEX selection, the binding ability of aptamers increases exponentially during the first few rounds. Finally, the binding affinity does not increase obviously or reaches a saturation point. Although a candidate aptamer appearing in one of the first few rounds does not mean it does not bind, the sequences from 3rd round and 11th round (see Table S1) are not duplicated in our experiment. Thus, it is reasonable to define the class labels of ssDNA sequences obtained from the first few rounds (e.g., from the 1st to 3rd rounds) of SELEX as 1 and the class labels of sequences from the final rounds (such as the 11th and 12th rounds) as 2 (here, class labels 1 and 2, respectively, denote aptamer sequences having low and high affinity). The experimental affinities from aptamer candidates (nos. 1, 2, 4, and 5 in Table 1) with PCa cell line PCa-3M-1E8 were tested with a flow cytometer. PCa-3M-1E8 cells (1 × 10⁶) and candidate aptamers in diverse concentrations were incubated in binding buffer (200 mL) on ice for 30 min. The equilibrium dissociation constant $K_d$ used to describe the affinities was calculated with eq 4.

$$B = \frac{(B_{\text{max}} \times C)}{(K_d + C)}$$

where the parameter C denotes the concentration of aptamer sequences, the parameter $B$ denotes the aptamer adsorption, and the parameter $B_{\text{max}}$ expresses the saturated adsorption. The experimental $K_d$ values are 3.24, 2.0, 46.8, and 7.1 nM which are in the nanomolar range and acceptable.

### Molecular Descriptor Derivation

An important step in classical QSAR studies is to calculate molecular descriptors. Generally, the QSAR models are based on molecular descriptors calculated directly from their respective molecular structure. However, this approach of calculating molecular descriptors is only appropriate to small molecules, rather than macromolecules or larger molecules. For instance, the commerce soft, Dragon, which has been widely used in molecular descriptor calculation, is not appropriate to aptamers (typical with 15–45 nucleotides).

Since the nucleotide sequence of a genome segment encodes the amino acid sequence, the amino acid sequence indicates the structure information of the corresponding nucleotide sequence. Thus, the structure descriptors of amino acid sequence are correlated with that of their corresponding nucleotide sequence. In this paper, DNA aptamer sequences were translated into amino acid sequence with DNAMAN software (version 6.0). Then, molecular structures of amino acid sequences were constructed by ChemBioDraw Ultra 11.0 in ChemOffice 2008, which were subsequently converted to 3D structures with ChemBio3D Ultra 11.0. The 3D structures of amino acid sequences were optimized by applying the molecular mechanics (MM2 force field) method with the convergence criterion of minimum root-mean-square error of...
the optimized amino acid sequences saved as Sybyl MOL2 files (*.mol2) which can calculate 4885 descriptors for each molecule. After removing redundant and nonuseful structure descriptors, which are equal to a constant or have a pair of correlation coefficients greater than 0.95, 1000 variables remained for descriptor selection.

**ASSOCIATED CONTENT**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b01464.

Candidate aptamer sequences and their class labels (PDF)

**AUTHOR INFORMATION**

Corresponding Author

*E-mail: yxl@hnie.edu.cn. Phone: +86 731 58680049. Fax: +86 731 68680125.

**ORCID**

Xinliang Yu: 0000-0002-0952-3567

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The work was financially supported by the National Natural Science Foundation of China (nos. 21190044 and 21190041), Scientific Research Fund of Hunan Provincial Education Department (no. 16A047), and the Open Project Program of State Key Laboratory of Chemo/Biosensing and Chemo-metrics (Hunan University) (no. 2016013).

**REFERENCES**

(1) Tan, W.; Donovan, M. J.; Jiang, J.-H. Aptamers from Cell-Based Selection for Bioanalytical Applications. *Chem. Rev.* 2013, 113, 2842–2862.

(2) Kashefi-Kheyrabadi, L.; Mehrgardi, M. A.; Wiede, E.; Turner, A. P.; Tiwari, A. Ultrasensitive detection of human liver hepatocellular carcinoma cells using a label-free aptasensor. *Anal. Chem.* 2014, 86, 4956–4960.

(3) Shi, H.; Cui, W.; He, X.; Guo, Q.; Wang, K.; Ye, S.; Tang, J. Whole cell-SELEX aptamers for highly specific fluorescence molecular imaging of carcinomas in vivo. *PLoS One* 2013, 8, No. e70476.

(4) Hong, H.; Goel, S.; Zhang, Y.; Cai, W. Molecular imaging with nucleic acid aptamers. *Curr. Med. Chem.* 2011, 18, 4195–4205.

(5) Guo, Q.-P.; Liu, X.-D.; Tan, Y.-Y.; Wang, K.-M.; Yang, X.-H.; Zhou, Y.; Ye, L.; Zhao, X.-Y. Selection of aptamers for human hepatocellular carcinoma with high specificity. *Chin. Sci. Bull.* 2013, 58, 2745–2750.

(6) Musafia, B.; Oren-Banaraya, R.; Noiman, S. Designing anti-influenza aptamers: novel quantitative structure activity relationship approach gives insights into aptamer–virus interaction. *PLoS One* 2014, 9, No. e97696.

(7) Binning, J. M.; Leung, D. W.; Amarasinghe, G. K. Aptamers in virology: recent advances and challenges. *Front. Microbiol.* 2012, 3, No. 29.

(8) Bouchard, P. R.; Hutabarat, R. M.; Thompson, K. M. Discovery and development of therapeutic aptamers. *Annu. Rev. Pharmacol. Toxicol.* 2010, 50, 237–257.

(9) Yu, X.; Deng, J.; Guo, Q. Computer-aided design of aptamers for SMMC-7721 liver carcinoma cells. *Turk. J. Biochem.* 2017, 42, 565–570.

(10) Yu, X.; Yu, Y.; Zeng, Q. Support Vector Machine Classification of Streptavidin-Binding Aptamers. *PLoS One* 2014, 9, No. e99964.

(11) Yu, X.; Yu, R.; Tang, L.; Guo, Q.; Zhang, Y.; Zhou, Y.; Yang, Q.; He, X.; Yang, X.; Wang, K. Recognition of candidate aptamer sequences for human hepatocellular carcinoma in SELEX screening using structure–activity relationships. *Chemom. Intell. Lab. Syst.* 2014, 136, 4–14.

(12) Li, B.-Q.; Zhang, Y.; Huang, G.; Cai, W.; Zhang, N.; Cai, Y. D. Prediction of aptamer-target interacting pairs with pseudo-amino acid composition. *PLoS One* 2014, 9, No. e86729.

(13) Thorne, H.; Willems, A. J.; Niedermayer, E.; Hoh, I. M.; Li, J.; Clouston, D.; Mitchell, G.; Fox, S.; Hopper, J. L.; et al. Kathleen Cunningham Consortium for Research in Familial Breast Cancer Consortium, Bolton, D. Decreased prostate cancer-specific survival of men with BRCA2 mutations from multiple breast cancer families. *Cancer Prev. Res.* 2011, 4, 1002–1010.

(14) Gian, F.; Sharma, N. V.; Moran, J. D.; Moreno, C. S. The biology of castration-resistant prostate cancer. *Curr. Probl. Cancer* 2015, 39, 17–28.

(15) Huang, Z.-X.; Xie, Q.; Guo, Q.-P.; Wang, K.-M.; Meng, X.-X.; Yuan, B.-Y.; Wan, J.; Chen, Y.-Y. DNA aptamer selected for specific recognition of prostate cancer cells and clinical tissues. *Chin. Chem. Lett.* 2017, 28, 1252–1257.

(16) Duan, M.; Long, Y.; Yang, C.; Wu, X.; Sun, Y.; Li, J.; Hu, X.; Lin, W.; Han, D.; Zhao, Y.; Liu, J.; Ye, M.; Tan, W. Selection and characterization of DNA aptamer for metastatic prostate cancer recognition and tissue imaging. *Oncotarget* 2016, 7, 36436–36446.

(17) Wang, Y.; Lu, Y.; Bing, T.; Chen, Z.; Lu, M.; Zhang, N.; Shangguan, D.; Gao, X. DNA Aptamer Evolved by Cell-SELEX for Recognition of Prostate Cancer. *PLoS One* 2014, 9, No. e100243.

(18) Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. DRAGON-Software for the Calculation of Molecular Descriptors, revision 6.0 for Windows; Talete s.r.l.: Milan, 2010.

(19) Deeb, O.; Goodarzi, M.; Khadikar, P. V. Quantum Chemical QSAR Models to Distinguish Between Inhibitory Activities of Sulfonylamides Against Human Carbonic Anhydrases I and II and Bovine IV Isozymes. *Chem. Biol. Drug Des.* 2012, 79, 514–522.

(20) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* 2012, 112, 2889–2919.

(21) Shahlaei, M. Descriptor Selection Methods in Quantitative Structure–Activity Relationship Studies: A Review Study. *Chem. Rev.* 2013, 113, 8093–8103.

(22) Singh, N.; Chaudhury, S.; Liu, R.; Abdulhameed, M. D.; Tawa, G.; Wallqvist, A. QSAR Classification Model for Antibacterial Compounds and Its Use in Virtual Screening. *J. Chem. Inf. Model.* 2012, 52, 2559–2569.