Original Research Article

Parallel screening and cheminformatics modeling of flavonoid activated aptasensors

Yu Xiu [a,b,c,1], Ni Zhang [d,1], Pranesha Prabhakaran [e,f], Sungho Jang [g,h,i], Qipeng Yuan [c], Curt M. Breneman [d], Gyoo Yeol Jung [j,k], Wanwipa Vongsangnak [e,f,l,**], Mattheos A.G. Koffas [b,e,f,*]

[a] *Beijing Key Laboratory of Bioactive Substances and Functional Food, Beijing Union University, Beijing, 100029, China*
[b] *Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA*
[c] *State Key Laboratory of Chemical Resource Engineering, Beijing University of Chemical Technology, Beijing, 100029, China*
[d] *Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA*
[e] *Interdisciplinary Graduate Programs in Bioscience, Faculty of Science, Kasetsart University, Bangkok, 10900, Thailand*
[f] *Omics Center for Agriculture, Bioresources, Food, and Health, Kasetsart University (OmiKU), Bangkok, Thailand*
[g] *Division of Bioengineering, College of Life Sciences and Bioengineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, 22012, South Korea*
[h] *Department of Bioengineering and Nano-Bioengineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, 22012, South Korea*
[i] *Research Center for Bio Materials & Process Development, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, 22012, South Korea*
[j] *Department of Chemical Engineering, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-gu, Pohang, Gyeongbuk, 37673, South Korea*
[k] *School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-gu, Pohang, Gyeongbuk, 37673, South Korea*
[l] *Department of Zoology, Faculty of Science, Kasetsart University, Bangkok, 10900, Thailand*

ARTICLE INFO

ABSTRACT

A parallel screening of 27 different flavonoids and chalcones was conducted using 6 artificial naringenin-activated riboswitches (M1, M2, M3, O, L and H). A quantitative structure-property relationship approach was applied to understand the physicochemical properties of the flavonoid structures resulting in specificity differences relied on the fluorescence intensity of a green fluorescent protein reporter. Robust models of riboswitches M1, M2 and O that had good predictive power were constructed with descriptors selected for their high correlation. Increased electronegativity and hydrophilicity of the flavonoids structures were identified as two properties that increased binding affinity to RNA riboswitches. Hydroxyl groups at the C-3′ and C-4' positions of the flavonoid molecule were strictly required for ligand-activation with riboswitches M1 and M2. Riboswitches O and L preferred multi-hydroxylated flavones as ligands. Substitutions on the A ring of the flavonoid molecule were not important in the molecular recognition process. *O*-glycosylated derivatives were not recognized by any of the riboswitches, presumably due to steric hindrances. Despite the challenges of detecting RNA conformational change after ligand binding, the resulting models elucidate important physicochemical features in the ligands for conformational structural studies of artificial aptamer complexes and for design of ligands having higher binding specificity.

1. Introduction

Natural flavonoids are the most abundant group of polyphenolic compounds; they are phytochemicals with distinct biological benefits to human health [1,2]. Although plant sources can synthesize thousands of these compounds, there are limitations in commercial scale-up due to

their low bioavailability, extraction difficulties, product mixtures, and time-consuming plant growth [3]. Consequently, the microbial production of plant natural products has been developed as a promising alternative method for producing flavonoids and many other value-added natural products. Microbial sourcing is also useful for bio-synthesizing novel flavonoids that do not exist in nature, which are now routinely produced through advances in metabolic engineering [1,4–10], synthetic biology [11–17], and computational approaches [18–23].

The utilization of riboswitches for reprograming cellular behavior has become increasingly popular in recent years, ever since the mechanism of action of riboswitches became better understood [24–29]. Because the number of natural riboswitches is limited, new artificial riboswitches are now being developed [23,27,30–34]. The screening of new aptamers targeting to either natural or non-natural ligands could be achieved with systematic evolution of ligands by exponential enrichment (SELEX). Crucial challenges in converting aptamers into functional riboswitch include dynamic range, operational range, ligand specificity and low false positive rate.

We recently developed a novel aptasensor, a particular class of biosensor where the biological recognition element is an RNA aptamer, and demonstrated a high correlation between the specific fluorescence of the microbial strain containing the biosensor and naringenin titer [35,

36]. In this study, we continued investigated the key substituents and mechanism for the activation of these artificial riboswitches (M1, M2, M3, O, L, H) by a parallel screening of 27 different flavonoids, flavonoid glycosides and chalcones. Quantitative structure-property relationship (QSPR) techniques, a useful computational approach in initial drug discovery process and for predicting bioactivities and properties of new molecules [37,38], was employed to identify the physiochemical properties that affect the response value to study the binding specificity of M1-, M2-and O-riboswitch. Descriptors selected were interpreted with the information from star plots, correlation plots and descriptor visualization techniques for each RNA riboswitch.

2. Methods

2.1. Media and reagents

M9 salts were purchased from Becton, Dickinson and Company (Franklin Lakes, NJ). Luria broth (LB) lennox modification, dimethyl sulfoxide (DMSO), naringenin, flavanone, (+)-catechin hydrate, (−)-epicatechin were purchased from Sigma-Aldrich (St. Louis, MO). Eriodictyol, apogenin, chrysin, genkwanin, 2′, 4′-dihydroxychalcone, isoliquiritigenin were purchased from Indofine Chemical Company, Inc (Hillsborough, NJ). Cyanidin chloride, cyanidin 3-O-glucoside chloride,

Table 1
The chemical structures of 28 flavonoids. Omg: (2-O-(6-deoxy-α-L-mannopyranosyl)-β-D-glucopyranosyloxy); Orha: O-α-L-rhamnopyranosyl; Oglc: O-glucopyranosyl; Ogr: O-β-D-glucopyranoside-7-O-α-L- rhamnopyranosyl.

Flavanone		Parent skeleton	R_5	R_7	$R_{3'}$	$R_{4'}$	–
1	Naringenin		OH	OH	H	OH	–
2	Flavanone		H	H	H	H	–
3	Eriodictyol		OH	OH	OH	OH	–
4	Naringin		OH	Omg	H	OH	–
5	Hesperidin		OH	Orha	OH	OMe	–
Flavan-3-ol			**R_5**	**R_7**	**$R_{4'}$**	**$R_{5'}$**	**–**
6	(+)-Catechin		OH	OH	OH	OH	–
7	(−)-Epicatechin		OH	OH	OH	OH	–
Flavone			**R_5**	**R_7**	**$R_{3'}$**	**$R_{4'}$**	**$R_{5'}$**
8	Chrysin		OH	OH	H	H	H
9	Apigenin		OH	OH	H	OH	H
10	Luteolin		OH	OH	OH	OH	H
11	Genkwanin		OH	OMe	H	OH	H
12	Acacetin		OH	OH	H	OMe	H
13	3′,4′-Dimethoxyluteolin		OH	OH	OMe	OMe	H
14	Apigenin-7-O-rutinoside		OH	Ogr	H	OH	H
15	Tricin		OH	OH	OMe	OH	OMe
Flavonol			**R_3**	**R_5**	**R_7**	**$R_{3'}$**	**$R_{4'}$**
16	Kaempferol		OH	OH	OH	H	OH
17	Quercetin		OH	OH	OH	OH	OH
18	Astragalin		Oglc	OH	OH	H	OH
19	Nicotiflorin		Ogr	OH	OH	H	OH
20	Isoquercetin		Oglc	OH	OH	OH	OH
Chalcone			**$R_{2'}$**	**$R_{3'}$**	**$R_{4'}$**	**R_3**	**R_4**
21	2′,4′-Dihydroxychalcone		OH	H	OH	H	H
22	Isoliquiritigenin		OH	H	OH	H	OH
23	Lanceoletin		OH	OMe	OH	OH	OH
Anthocyanidin			**R_3**	**R_5**	**R_7**	**$R_{3'}$**	**$R_{4'}$**
24	Cyanidin		OH	OH	OH	OH	OH
25	Pelargonidin		OH	OH	OH	H	OH
26	Cyanidin-3-O-glucoside		Oglc	OH	OH	OH	OH
27	Peonidin-3-O-glucoside		Oglc	OH	OH	OMe	OH
Aurone			**R_6**	**R_7**	**$R_{3'}$**	**$R_{4'}$**	**–**
28	Leptosidin		OH	OMe	OH	OH	–

pelargonidin chloride, peonidin-3-*O*-glucoside were purchased from Extrasynthese (Genay, France). All the other flavonoids were synthesized and were generously provided as a gift from Professor Mamoru Koketsu of Gifu University, Japan (Table 1).

2.2. Aptasensors

The naringenin-responsive RNA aptamer [35,36] was isolated from combinatorically prepared nucleic acid libraries using SELEX and cloned upstream of reporter, superfolder green fluorescent protein (sGFP) (Fig. S1). Aptasensors were designed with different artificial sensing domains possessing conserved features such as randomized linker, primer binding sites, and the ribosome-binding site (RBS) (Table S1). The vector pACYCDuet-1was used as the basis for aptasensor construction while the constitutive promoter BBa_J23113 was utilized.

2.3. RNA structure prediction

The secondary structures of RNA aptamers were predicted by the m-fold web server (http://unafold.rna.albany.edu/?q=mfold) [39,40]. The three-dimensional (3D) structures of aptamers were predicted by SimRNA web server (http://genesilico.pl/SimRNAweb) [41,42] and visualized with SYBYL software. Structures were all predicted according to free energy minimization algorithm. The molecular docking module surflex-dock in SYBYL package was used for molecular-recognition and specific-recognition-sites prediction.

2.4. Cultivation protocol

Propagation of plasmid was done using *E. coli* DH5α. BL21star™(DE3) was used as the host for naringenin aptasensor. Briefly, single colonies of each strain were inoculated separately into 10 mL of LB in a 125 mL non-baffled shake flask, added with corresponding antibiotics (chloramphenicol 25 μg/mL) and grown overnight at 37 °C. After 14 h incubation, the overnight culture was inoculated at 5% (100 μL) into 2 mL of M9 minimal media (4 g/L glucose). Flavonoids were dissolved in DMSO and added to 0.2 mM final concentration at 7 h after inoculation and grown for 12 h at 37 °C prior to analysis [35]. All experiments were performed using polypropylene 48-well plates (5 mL, VWR). A control sample without flavonoids was used as a negative control to subtracts the specific background fluorescence value to ensure it was comparable between each aptasensor. A control sample with 0.2 mM naringenin was performed in all experiments as a positive control. All experiments were performed in triplicates.

2.5. Fluorescence assay

Flow cytometry was performed on LSRII flow cytometer (BD Biosciences, San Jose, USA) with 488 nm excitation from a blue solid-state laser for all the co-culture experiments. Fluorescence was detected using a 505 nm long-pass and a 530/30 nm band-pass filter set for *sgfp*. Cells were collected with centrifugation (2 min, 20,000×*g*) followed by washing and dilution in cold phosphate buffered saline (PBS) and stored on ice until evaluation [35]. . To avoid debris and clumped cells, gating was performed on forward and side scatter by using BD FACSDiVa 7.0 software. The specific fluorescence was defined as the value of geometric fluorescence mean of 100,000 cells (given in a.u.).

2.6. QSPR models of naringenin-mediated riboswitch

For each RNA riboswitch, computed chemical features (descriptors) and their relative weights were identified by modeling each of the flavonoid sets with fluorescence intensity (FI) of each designed RNA device (Fig. S1). The descriptors are calculated for the 28 flavonoids and chalcones in their single low energy forms. According to the nature of these small molecules to bind with RNA riboswitches for subsequent

triggering of conformational transitions, two classes of descriptors were selected to construct the original descriptor set. A total of 206 connectivity-based two-dimensional (2D) descriptors were calculated to study the properties like partial charges and surface area. Three-dimensional (3D) descriptors were comprised of 128 internal 3D descriptors that depends on internal coordinates only. The molecular operating environment (MOE, Chemical Computing Group, Inc., Montreal, Canada) software was employed to obtain the chemical descriptors discussed above.

Feature selection was done on our in-house support vector regression (SVR)-based online learning equipment (SOLE; http://reccr.chem.rpi.edu/SOLE/index.html). Objective feature selection including correlation with response (FI) and mutual correlation was used to remove descriptors that are of low significance based on the dataset [43]. Besides, subjective feature selection "recursive feature elimination" is used for removing descriptors less important in constructing the model. The model is built with features selected in the above procedure using an in-house web tool. Least square support vector regression (LS-SVR) algorithm is chosen after comparing with partial least square (PLS), standard SVR and fussy SVR results.

3. Results and discussion

3.1. Parallel screening of flavonoids specificity

The naringenin-recognized aptamers were selected *in vitro* and constructed as functional devices *in vivo* [35,36]. In nature, naringenin is a precursor of the vast majority of flavonoids. Our goal was to understand the specificity of each riboswitch so as to exploit these naringenin-responsive riboswitches as dynamic regulators of metabolic pathways. Thus, a parallel specificity screening study was undertaken using 24 flavonoids. These contained the basic C6–C3–C6 carbon skeleton and were diversified through different biotransformations, including hydroxylations, *O*-methylations, *O*-glycosylations, dehydrogenations, hydrogenations and cyclizations. Three different chalcones were also examined since chalcone represents a common core unit in nearly all flavonoid biosynthesis. A reporter gene for sGFP expression quantified activation of the riboswitches.

The tested flavonoids having a 15-carbon skeleton similar to naringenin showed varying efficiency due to the presence of different substituents. In our parallel screening study of flavonoids, the artificial naringenin-mediated riboswitches demonstrated varied molecular-recognition and activation of the downstream reporter (Fig. 2). Riboswitch M1 was much more responsive to naringenin **1**, eriodictyol **3** and chalcones **21**–**23** than the other tested flavonoids. Riboswitch M2 was much more responsive to chalcones, especially 2′,4′-dihydroxychalcone **21**, than to naringenin. Riboswitch M3 was only responsive to naringenin. Riboswitch O showed a strong specific response to quercetin **17** and kaempferol **16**, although the activation with naringenin and chalcones was still relatively high.

The fluorescence intensities in the presence of different flavonoids were normalized to their observed response in the presence of naringenin **1**, which is a flavanone with three hydroxy groups at the 4′, 5, and 7 carbons and their activation ratios were calculated (Fig. 1). Apigenin **9**, a flavone with three hydroxyl groups at the 4′, 5, and 7 carbons, showed a low activation (<0.4) of the riboswitches, except for riboswitches O and L. The flavonol kaempferol **16**, with four hydroxyl groups at the 4′, 5, 7 and 3 carbons, showed a similar result. This could be explained by one of the key structural recognition elements being present at the C-3 position of the C ring. Complementary evidence was found for flavan-3-ols **6** and **7** with no activation and with anthocyanidins **24** and **25** with a low activation (<0.3), except for riboswitches M3 and L. Thus, flavonoid substrates, having dehydrogenation and hydroxylation at their C-3 position, are less potent ligands for riboswitch M1 and M2.

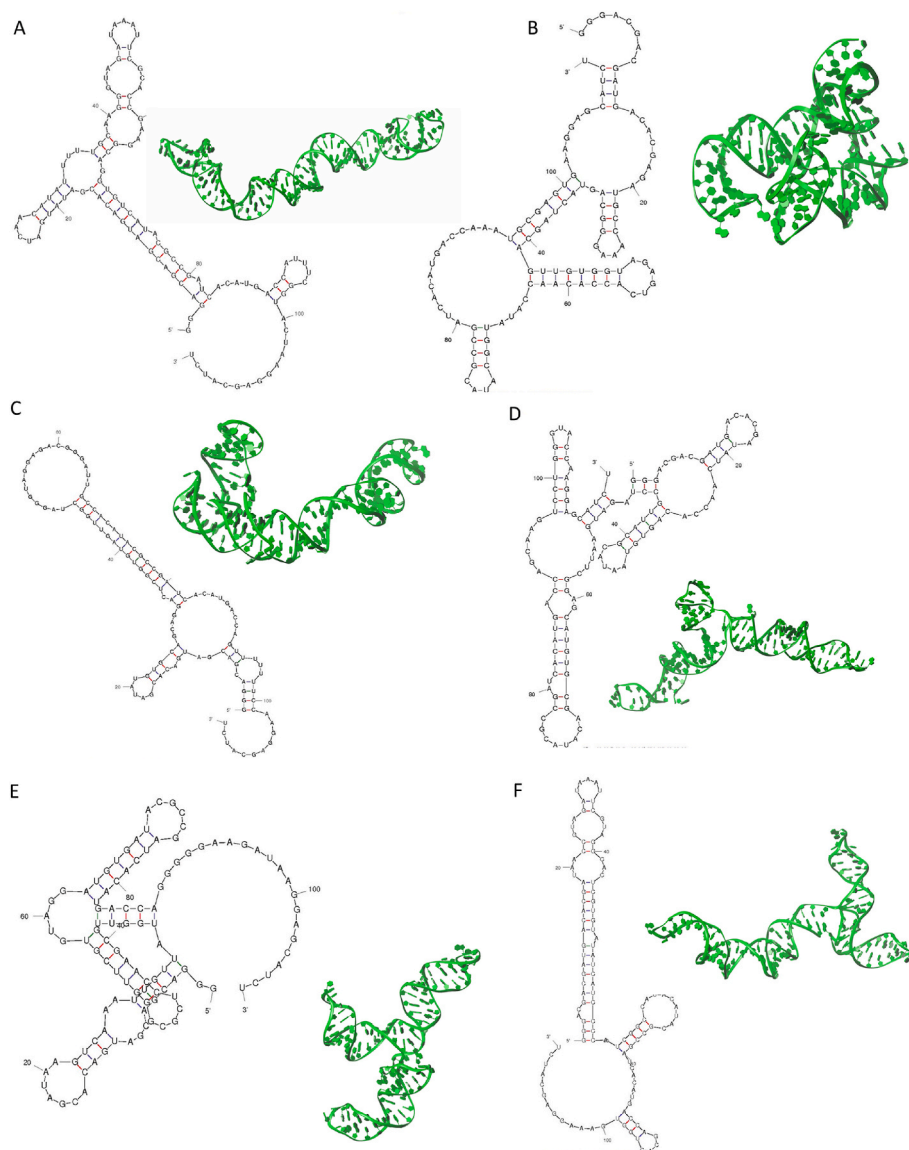Natural flavonoids reveal that the multiple hydroxyl groups,

Fig. 1. Structure prediction of naringenin-mediated riboswitches. Predicted structure of the sensing domain of riboswitches M1, M2, M3, O, L, H are labeled from A to F separately. The secondary structures were shown on right, while the tertiary structures were on left in each group.

particularly in the B ring, show enhanced antioxidant potential [44]. Consequently, the multiple hydroxyl groups were utilized to immobilize naringenin on Sepharose matrix in SELEX, as well as recognition and binding to RNA [36]. Moreover, the skeleton structure of a flavanone is also considered to be a specific aptamer-recognition structure. The activation ratio to flavanone **2** showed that only riboswitch M1 and L were strongly responsive to the flavanone skeleton structure, showing a response of 0.68 and 0.49 respectively (Fig. 1). Furthermore, a group of flavones, chrysin **8**, apigenin **9** and luteolin **10** and the flavonol quercetin **17**, in which hydroxyl groups were increased one at a time, interacted with the six different riboswitches. A dramatic increase in the activation was observed for riboswitch O, from 0.59 to 1.91, and similarly in riboswitch L, from 0.45 to 1.23, for all of these compounds except for luteolin **10**. Riboswitch M2 activation positively correlated with an increase in B ring hydroxyl groups but no activation was detected when the hydroxylation was present at C-3 position (quercetin **17**), which was consistent with previous results. The activation of the other riboswitches varied only slightly with changes in hydroxyl groups.

Complementarity was obtained with regards to the *O*-methylatation of hydroxyl groups (Fig. 1). Tricin **15**, 7-*O*-methylated flavone genkwanin **11**, and 4′-O-methylated flavone acacetin **12** which can be derived from apigenin **9**, and 3′, 4′-dimethoxyluteolin **13**, which can result from luteolin **10** were chosen to study. Riboswitch M1 and M2 can hardly recognize ligands with the *O*-methylation on the C-4′ and C-3′ hydroxyl groups, but works well with *O*-methylation group of the hydroxyl group at the C-7 position. Riboswitch L gave no response for all the *O*-methylated flavonoids tested. However, riboswitch M3 was more responsive to *O*-methylated flavonoids than to apigenin **9** and luteolin **10**.

Interestingly, all riboswitches responded well to the three chalcones **21**–**23**. Riboswitch M2 displayed even more activation by these chalcones than it did for the ligand it was designed to respond to, naringenin. Similar results were obtained for riboswitch O activated with isoliquiritigenin **22**. This result could possibly be because natural chalcones can spontaneously cyclize in solution producing an enantiomeric mixture of flavanones (Fig. S2); moreover, chalcone isomerase (CHI) guarantees the formation of biologically active (*2S*)-flavanones [45,46] The results again confirm that the essential for riboswitch response structural features occur at the C-3 position, and may also suggest that a hydroxyl group at C-5 does not greatly contribute to activation.
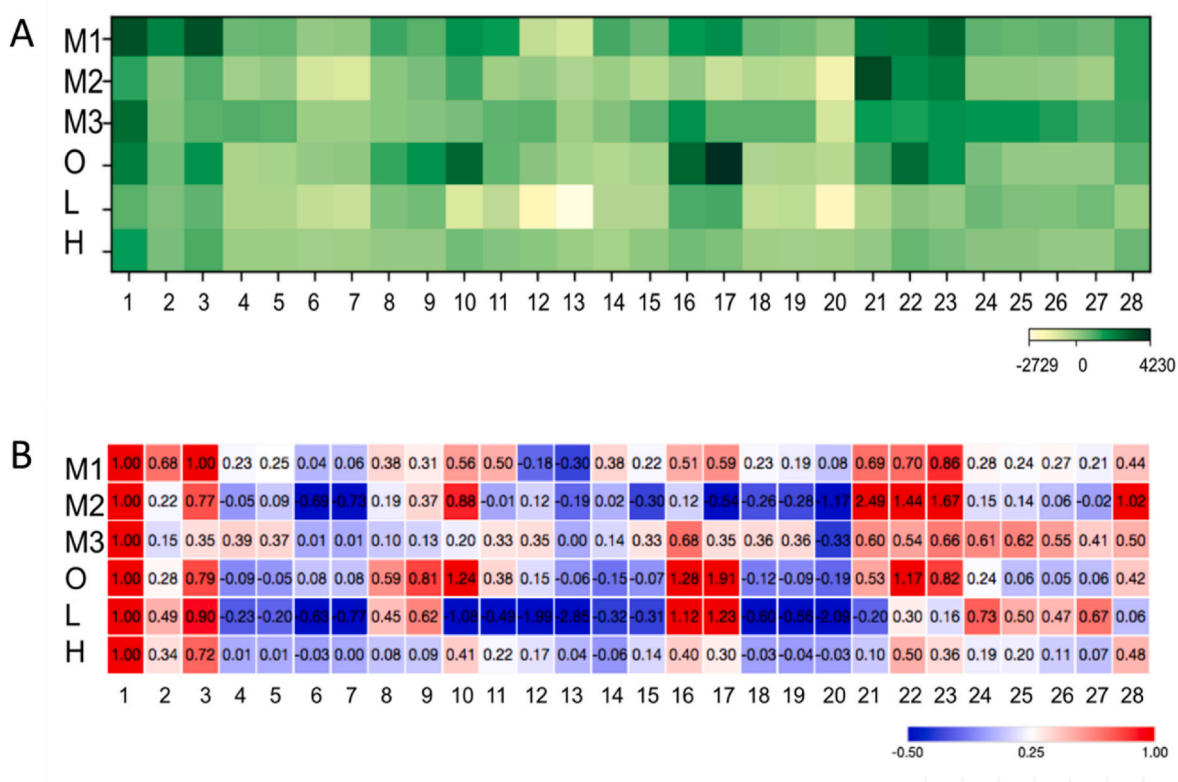
Fig. 2. Ligands spectrum of constructed aptasensors. A) The specific fluorescence intensity of *E. coli* cells harboring aptasensor M1, M2, M3, O, L, H were collected after cultivation in the presence of 0.2 mM of various flavonoids. A control sample for each mentioned apatasenor without ligands addition was detected with the results of FI (a.u.) 1485 ± 63, 1269 ± 68, 2437 ± 106, 1138 ± 5, 3133 ± 54, 425 ± 19 separately. The control sample was included as a background blank which was subtracted here to ensure it was comparable between each apatasensor. B) The values were normalized to the naringenin-activation reporter expression to calculate the activation ratios. The labeled number on axis represents the supplemented flavonoids reference in Table 1.

3.2. Conformational structure study of riboswitch

Contrasting with ribonucleic acids derived from biological sources, which are optimized with regard to several aspects of their cellular functions, aptamers do not balance specificity in ligand binding with additional cellular functions. The functionality, more important the dynamics, of an artificial riboswitch depends on structure-function-dynamics mechanism [22,23,30,34]. The sequence results (Table S1) indicated a homologous tendency between the six aptamers used in this study. The stem-loop motifs predicted in the secondary structures (Fig. 2) are considered as the most likely ligand-binding sites (Table S2), which has also been reported in other RNA regulator studies [27–30]. This 3D structure is a highly optimized scaffold for specific ligand recognition and conformational change, thus, the ligand becomes an essential part of the nucleic acid structure.

The accuracy of structural prediction needs to be validated, but the experimental determination of RNA 3D structures is quite laborious and challenging, typically requiring x-ray crystallography. An indirect factor validation method relying on the docking score and FI results was applied to determine the ligand-recognition site of aptamer complexes. Most plants accumulate flavonoids as their *O*-glycosylated derivatives, since glycosylation of flavonoid aglycones enhances their stability [47]. Thus, several 7-*O*-glycosides: such as naringin **4**, hesperidin **5** and apigenin-7-*O*-rutinoside **14**, and several 3-*O*-glycosides: astragalin **18**, nicotiflorin **19**, isoquercetin **20**, cyanidin-3-*O*-glucoside **26** and peonidin-3-*O*-glucoside **27** were also examined. However, no activation or relatively weak activation was observed for the *O*-glycosylated flavonoids in any of the naringenin-responsive riboswitches (Fig. 1), which implies the possibility of adverse steric interactions between these glycosides and the riboswitch in the aptamer-ligand association.

Complementary evidence was also provided by a molecular docking study. The potential recognition sites in these riboswitches were predicted with the multichannel molecular docking module surflex-dock in the SYBYL package and the potential sites of riboswitch M1, for example, are found scattered across its entire tertiary structure (Fig. S3A). The crash value was examined as it provides information on the degree of inappropriate ligand penetration into the receptor and the degree of internal self-clashing within the ligand. The crash score, which was graded in molecular docking between *O*-glycosylated flavonoid derivatives and riboswitch, should be far below zero (Fig. S3B). In contrast, it could also be assumed that the predicted ligands-binding sites, which have crash values for *O*-glycosylated derivatives far below zero, site 1 and site 2, are quite accurate. The predicted recognition site 1 and site 2 also agreed with the predicted stem-loop motifs in the secondary structure (Table S2).

3.3. QSPR modeling of flavonoids-mediated sGFP expression

In this study, the QSPR approach was able to distinguish the leading physiochemical properties (chemical descriptors) of chemical structures in flavonoid structures that correlate with the observed specificity differences. Feature selection was performed both subjectively and objectively using an in-house support vector machine (SVM)-based learning web tool.

Mutual correlations of the chemical features were calculated first and the descriptors with Pearson correlation above a threshold were deleted from the set to decrease the complexity of the algorithm and the risk of errors from the model. The threshold was determined based on the experimentation results of different models which was 0.95 of the models for M1, M2 dataset and 0.85 of the models for O dataset.

Mutually dependent descriptors with less correlation ($<$50%) with the FI response were also eliminated to improve the efficiency of the modeling [48]. For example, 141 features of model M1 were retained because their mutual correlation was less than 0.95. A test on correlation with response was next performed and descriptors with less than 40% correlation with the response were eliminated. Next, 50% of the original descriptors were retained after a single iteration of ordinary least squares (OLS) recursive feature elimination. Feature dimensions were decreased to reduce overfitting and improve predictive accuracy. The final model is comprised of 8 descriptors of Model M1, 6 descriptors of Model M2, and 5 descriptors of Model O (Table 2).

Least square-support vector regression (LS-SVR) was shown to perform the best with r^2 value of 0.7882, R^2 value of 0.7882 and RMSD of 470.03 with randomly split 90% of training set in the construction of a model of riboswitch M1 (Fig. 3A). By randomly selecting one sample in the training set as validation set to estimate the prediction error, a 20-times of leave-one-out (LOO) cross validation was performed, which helps to avoid overfitting and select proper hyper-parameters. The predictive power of the model is demonstrated in Fig. 3B, which shows the prediction accuracy in response value region through 600 to 1800 is r^2 of 0.9640, R^2 of 0.8707 and RMSE of 163.6539.

The validation of prediction accuracy is crucial in the model building process since it establishes the basis for reliable prediction. By randomly splitting the whole data set of Model riboswitch M1 into 90% training set and 10% test set 10 times, using the same descriptors and modeling method, the prediction r^2, R^2 and RMSE were obtained and are listed in Table S3. The method used to create the model was consistent when trained using the same meta parameters but using different subsets of the original data. The average training r^2 of 10 times training experiments of Model riboswitch M1 was 0.781 and the average testing r^2 is 0.807. The performance of the manual random split validation is consistent for there are only two over 10 times of prediction r^2 on unknown test set is below 0.6.

The model obtained for riboswitch M2, shown in Fig. 4, has the Pearson correlation score r^2 of 0.8114, coefficient of determination R^2 of

0.8114 and root mean square value of 463.96. The best model obtained for riboswitch O with r^2 of 0.7759, R^2 of 0.7747 and RMSD of 571.5046 for 90% split training set was constructed (Fig. 5). In Figs. 4B and 5B, randomly selected test set of three molecules demonstrated the prediction power of the model was acceptable in the relative chemical space. The consistency of models built for M2 and O is proven by the manual split validation with average r^2 of 0.813 for M2 and 0.775 for O (Table S4, Table S5).

3.4. QSPR model interpretation

Using the selected chemical descriptors, a robust model could be constructed as described in the previous section, which suggests that these descriptors are related to the chemical features that determine RNA binding specificity. Star plots of the descriptors arranged in the ranking of correlation with the response values (Fig. 3C, Figs. S4 and S5) show that there is a consistency of positive or negative influence within each bootstrap model. The analysis of descriptors selected for each RNA based riboswitch provides an implicit connection with the physiochemical properties on the flavonoids' binding specificity with the RNA aptamers.

3.5. Riboswitch M1 model descriptors

Q_VSA_FNEG, FASA- and PEOE_VSA_POS are three descriptors that link the shape and electronic information to characterize molecules and encode feature accountability for polar interactions. The first two descriptors relate with the fractional negative surface area of the flavonoids molecules and both correlate positively with the activation. (Fig. 3Da and 3Db) The former one is fractional negative van der Waals surface area with the latter being fractional negative water accessible surface area with a probe radius of 1.4 Å.

Q_VSA_FNEG computes the sum of vi (vdW surface area of atom i) in which qi (partial charge of atom i) is negative divided by the total surface area. A connection table approximation is used to calculate the vi. Descriptors prefixed with Q_ use the partial charges stored with each structure in the database. FASA-provides an indication of the fraction of the solvent accessible molecular surface area bearing a negative electrostatic potential, which would be expected to correlate with Brønsted or Lewis basicity, at the same time providing an indication of molecular dipolarity [49]. The fractional negative charge on the molecular surface ought to be higher for better RNA binding specificity. The calculated electrostatic map of naringenin and eriodictyol that demonstrated best response values have a larger portion of negative surface than the 3′, 4′-dimethoxyluteolin with very low response value (Fig. 6).

This suggests that the active site of the RNA may have positively charged groups or it may also have hydrophilic interaction sites. The conclusion has also been confirmed by the descriptor PEOE_VSA_POS, referring to the positive surface area, which has a negative influence on the response. (Fig. 3De).

Vsurf_CW1 Capacity factors indicate the ratio of the hydrophilic surface over the total molecular surface; in simple terms, it is the hydrophilic surface per surface unit [50]. A total of eight different energy levels were used to calculate the capacity factors and hydrophilic descriptors. CW1 refers to the energy level at −0.2. The positive contribution of this descriptor on the response value indicates that to achieve the higher binding affinity with riboswitch-M1, hydrophilic surface should be of higher ratio in the flavonoids structure (Fig. 7).

Reactive This descriptor is an indicator of the presence of reactive groups, where molecules with a reactive group show a non-zero value. The table of reactive groups is based on the Oprea [51] and includes phospho-, metals, N/O/S–N/O/S single bonds, acyl halides, thiols, Michael Acceptors, azides, esters, etc. According to the correlation plot (Fig. 3Dd), the flavonoids with a reactive group has a potency to react better with riboswitch-M1. For example, the comparison of eriodictyol with luteolin shows that the presence of aliphatic ketone as a reactive
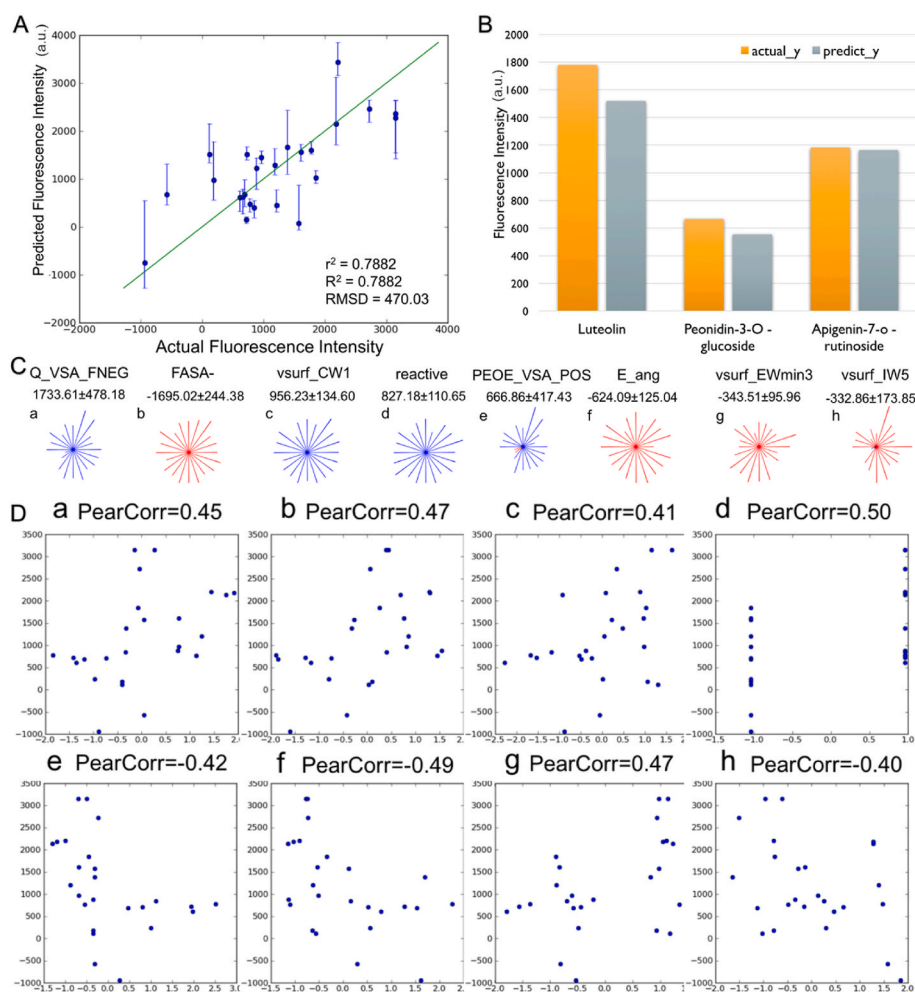
Table 2
Definitions of descriptors selected following subjective and objective feature election in the QSAR model for riboswitch activation.

Model	Rank	Descriptor	Class	Descriptor Definition
M1	1	Q_VSA_FNEG	2D	Fractional negative vdw surface area
	2	FASA-	i3D	Fractional negative accessible surface area
	3	vsurf_CW1	i3D	Capacity factor at −0.2
	4	reactive	2D	Reactivity
	5	PEOE_VSA_POS	2D	Fractional positive van der Waals surface area, calculated by Partial Equalization of Orbital Electronegativities (PEOE) method
	6	E_ang	i3D	Angle bend potential energy
	7	vsurf_EWmin3	i3D	3rd lowest hydrophilic energy
	8	vsurf_IW5	i3D	Hydrophilic integy moment at −3.0
M2	1	GCUT_SLOGP_3	2D	LogP GCUT (3/3)
	2	GCUT_SMR_3	2D	Molar Refractivity GCUT (3/3)
	3	reactive	2D	Reactivity
	4	PEOE_VSA_FNEG	2D	Fractional negative vdw surface area
	5	opr_nring	2D	Oprea Ring Count
	6	Q_VSA_FNEG	2D	Fractional negative vdw surface area
O	1	BCUT_SLOGP_0	2D	Log of the octanol/water partition coefficient
	2	GCUT_SMR_2	2D	Molar refractivity GCUT(2/3)
	3	E_ang	i3D	Angle bend potential energy.
	4	PEOE_VSA_POS	2D	Fractional positive van der Waals surface area, calculated by Partial Equalization of Orbital Electronegativities (PEOE) method
	5	Vsurf_R	i3D	Surface roughness

Fig. 3. LS-SVR based QSAR model of naringenin-riboswitch M1 activation. A) QSAR model of training set; B) QSAR model prediction for external test set; C) star plots of descriptors in model M1. The mean and standard deviation of importance values are shown. (Red, negative contribution to riboswitch activation; Blue, positive contribution); D) Correlation plots between selected normalized descriptors (x axis) and response (the specific fluorescence intensity, y axis). The respective Pearson Correlation values are labeled. (a) Q_VSA_FNEG (b) FASA- (c) vsurf_CW1 (d) reactive (e) PEOE_VSA_POS (f) E_ang (g) vsurf_EWmin3 (h) vsurf_IW5.
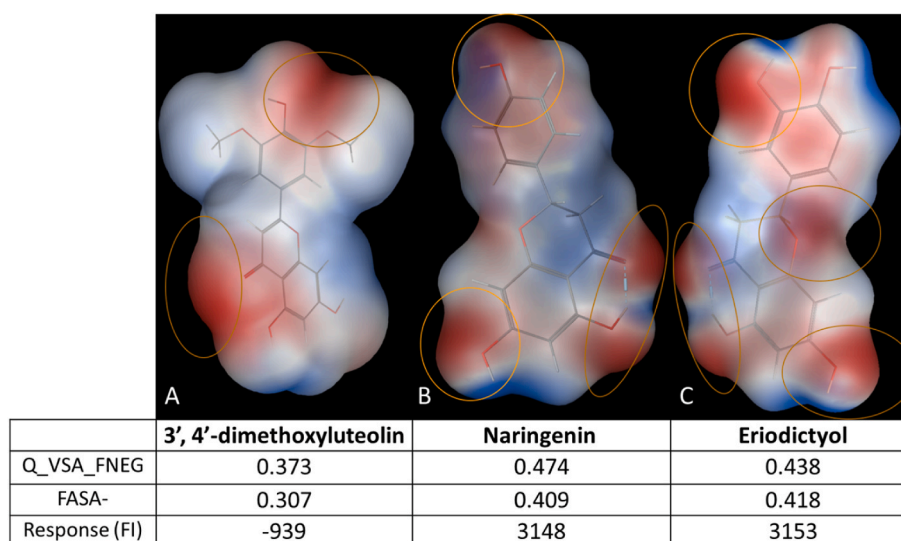
Fig. 4. Electrostatic map of the 3′, 4′-dimethoxyluteolin (A), naringenin (B), eriodictyol (C). Red presents negative; white presents neutral; blue presents positive. The values of descriptors calculated and the positive influence on response fluorescence intensity are listed for comparison in the table.

	3′, 4′-dimethoxyluteolin	Naringenin	Eriodictyol
Q_VSA_FNEG	0.373	0.474	0.438
FASA-	0.307	0.409	0.418
Response (FI)	-939	3148	3153

group works positively on the response.

E_ang Angle bend potential energy. This descriptor contributes negatively in the model (Fig. 3Df) and elucidate the reason behind the need of low angle bend energy for better specificity. This could be achieved if any molecule carries a flexible group in its structure, which are oriented towards the active binding site with less bending energy.

Vsurf_EWmin3 Local interaction energy minima represent the energy of interaction (in kcal/mol) of the best three local energy minima

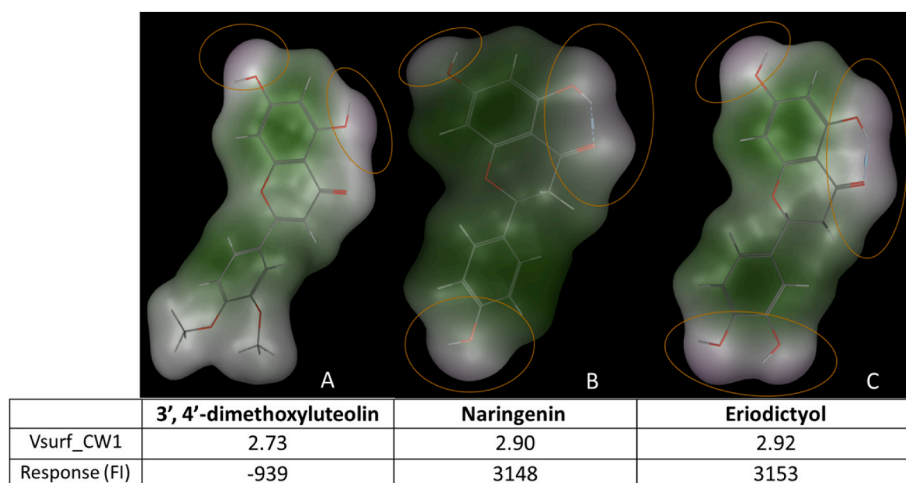|	3', 4'-dimethoxyluteolin	Naringenin	Eriodictyol
Vsurf_CW1	2.73	2.90	2.92
Response (FI)	-939	3148	3153

Fig. 5. Molecular surface of 3′, 4′-dimethoxyluteolin (A), naringenin (B) and eriodictyol (C). Purple presents hydrophilic; translucent presents neutral; green presents hydrophobic. The values of descriptors calculated and the positive influence on response fluorescence intensity are listed for comparison in the table.
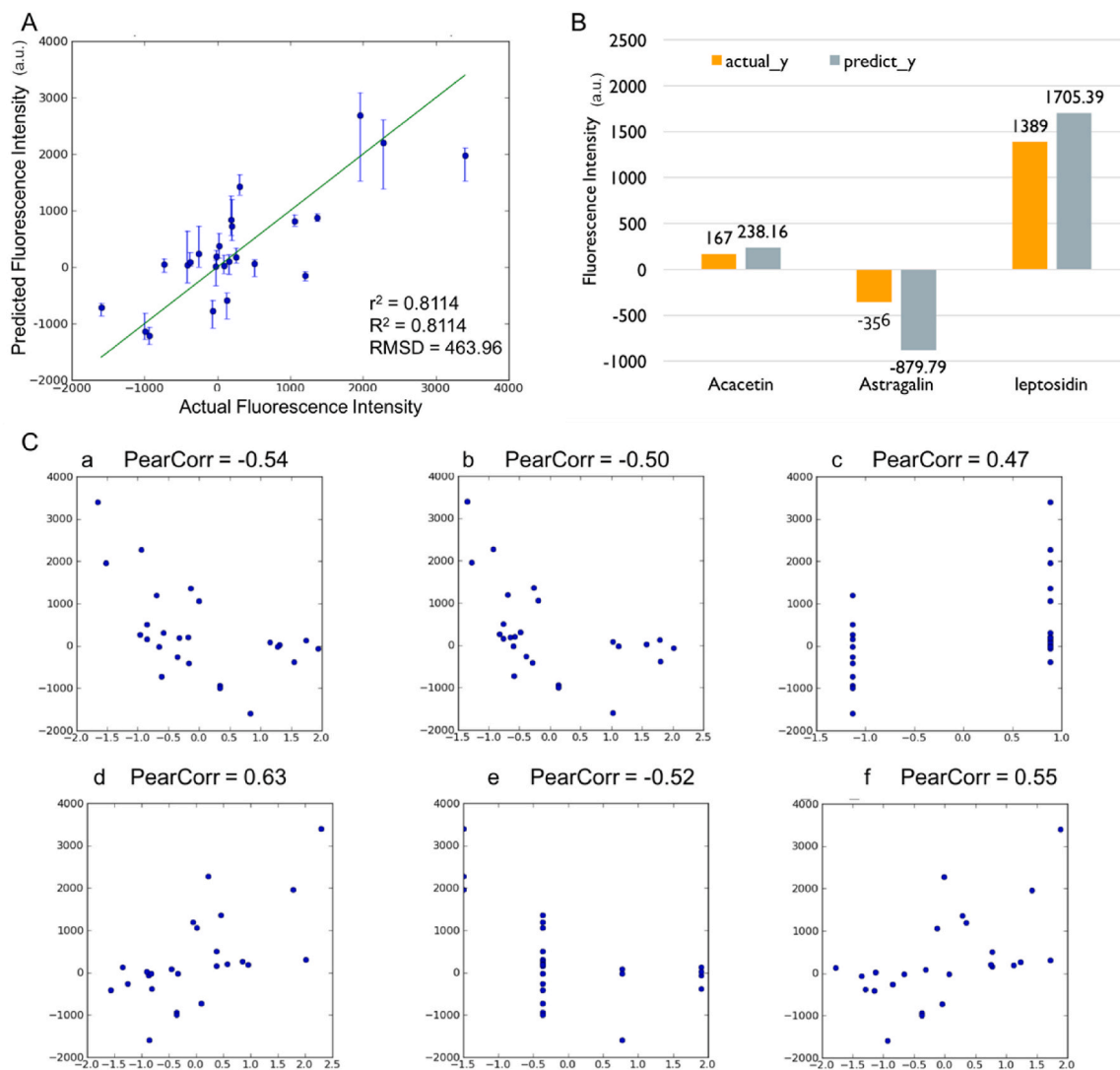
Fig. 6. LS-SVR based QSAR model of naringenin-riboswitch M2 activation. A) QSAR model of training set; B) QSAR model prediction for external test set; C) Correlation plots between selected descriptors and response (the specific fluorescence intensity). The respective Pearson Correlation values are labeled. (a) GCUT_SLOGP_3 (b) GCUT_SMR_3 (c) reactive (d) PEOE_VSA_FNEG (e) opr_nring (f) Q_VSA_FNEG.
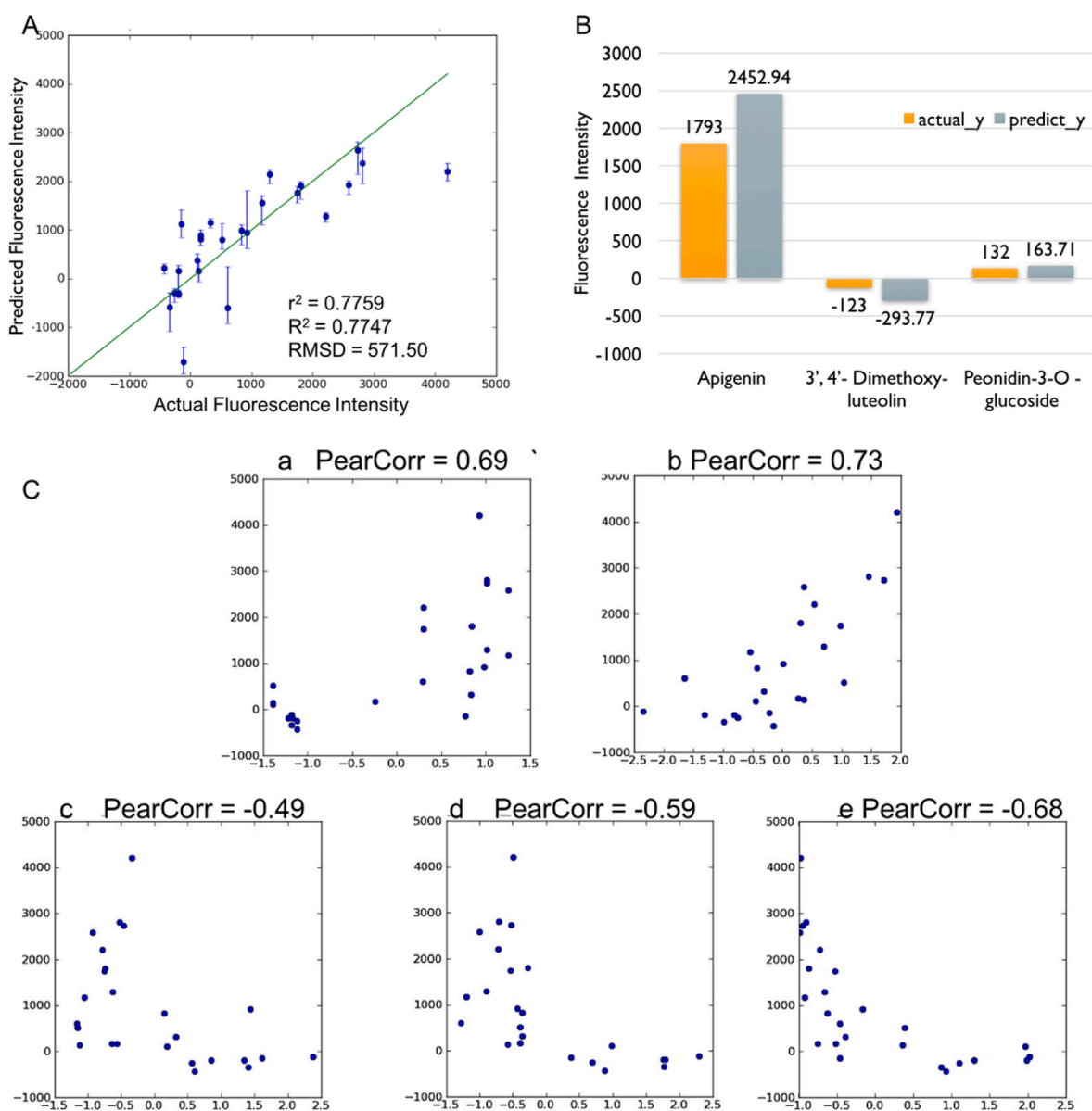
Fig. 7. LS-SVR based QSAR model of naringenin-riboswitch O activation. A) QSAR model of training set; B) QSAR model prediction for external test set; C) Correlation plots between selected descriptors and response (the specific fluorescence intensity). The respective Pearson Correlation values are labeled. (a) BCUT_SLOGP_0 (b) BCUT_SMR_2 (c) E_ang (d) PEOE_VSA_POS (e) Vsurf_R.

between the water probe and the target molecule. This minima refers to the third deepest local minima in the molecular electrostatic potential, which is also known as the third lowest hydrophilic energy [50]. A positive correlation (Fig. 3Dg) is observed in the descriptor in this model, indicating the interaction energies (local minima) when the hydrophilic region has contact with the target should be minimum. The interaction energy will be low if the active site of the RNA has hydrophobic properties, as there is less water present in the active site region.

Vsurf_IW5 Like dipole moments, interaction energy moments express the lack of balance between the center of mass of a molecule and the barycenter of its hydrophilic regions [50]. In case of referring to hydrophilic region, the interaction energy moments, are vectors pointing from the center of mass to the center of the hydrophilic regions. Negative contribution (Fig. 3Dh) shows that hydrophilic moieties should be either close to the center of the molecule or at opposite ends of the molecule for better binding specificity with riboswitch M1.

3.6. Riboswitch M2 model descriptors

GCUT_SLOGP_3 and GCUT_SMR_3 GCUT descriptors were developed initially for study of molecular similarity [52]. GCUT descriptors are calculated from the eigenvalues of a modified adjacent matrix. Each ij entry of the adjacency matrix takes the value $1/sqrt (dij)$ where dij is the modified graph distance between atoms i and j. The resulting eigenvalues are then sorted.

GCUT_SLOGP_3 uses the atomic contribution to log P (octanol-water partition coefficient) as the values on the diagonal of the adjacency matrix and largest eigenvalues calculated is reported here to have a negative influence on the fluorescence intensity response. The log P is the most commonly used measure of lipophilicity of the molecules [53]. The negative influence of this descriptor on FI, demonstrated in Fig. 4Ca, suggests the lipophilicity surface portion of the designed ligand molecules, on the contrary to riboswitch O, should be relatively small to obtain a high binding affinity to riboswitch M2.

GCUT_SMR_3 describes an importance of molar refractivity of the

ligands in determining the binding affinity to M2 aptamer. As shown in Fig. 4Cb, the molar refractivity of the molecules is relatively small in the high FI response region. This could be interpreted in the way that molecules of smaller molecular size and low polarizability are prone to bind with riboswitch M2.

Reactive and Opr_nring As described in M1 model, reactive is a descriptor of the reactive groups and those ligands with reactive groups such as 2′, 4′-dihydroxychalcone that has a Michael acceptor in the middle of the structure tends to have a higher binding tendency towards riboswitch M2. Designed together in the paper by Oprea et al. [51], Opr_nring is a descriptor that counts the number of rings in the structure. With the four categories shown in Fig. 4Cc, the values of the descriptors, which stand for the number of rings, are 2, 3, 4, 5 from left to right. The ligands with higher FI response are clustered in 2 and 3 rings when binding with riboswitch M2.

PEOE_VSA_FNEG and Q_VSA_FNEG calculate the fractional negative surface of the ligand molecules as mentioned in M1 model. They both exhibit a positive correlation with the FI response value in Fig. 4Cd and 4Cf. The ligands that have larger portion of negative surface are prone to have better binding affinity with riboswitch M2.

3.7. Riboswitch O model descriptors

BCUT_SLOGP_0, BCUT_SMR_2 BCUT descriptors are similar descriptors to GCUT [52]. The only difference lies in the utilization of modified distance adjacency matrix 1/sqrt(bij) where bij is the formal bond order between bonded atoms i and j.

BCUT_SLOGP_0 is similar to GCUT_SLOGP_3 as described in the model of riboswitch M2. The meaning of selecting this descriptor shows the difference of lipophilicity could influence the response value. Fig. 5Cb shows there exists a positive influence of this descriptor on the response. A larger portion of lipophilic surface could be designed to obtain a larger response value when binding with riboswitch O.

BCUT_SMR_2 in contrast, uses the atomic contribution to molar refractivity, which contains information of both the molecular size and polarizability [54], as the diagonal values of the adjacency matrix for calculation. In this case, the positive influence of this descriptor (Fig. 5Cb) suggests the binding force between the polar portions of the RNA aptamer and ligands can be increased for a better response value.

E_ang refers to angle bend potential energy as in model riboswitch M1. As seen from the correlation plot Fig. 5C, molecules with relatively low angle binding energy have a higher FI, which is plausible since the ligands could bind with RNA aptamer with weaker resistance with low angle binding energy. Comparing quercetin, which exhibits the highest FI of 4203, with hesperidin with FI of −111, shows the former molecule is more flexible while the structure of the latter molecule has 3 internal hydrogen bonds that would cause a higher angle binding energy. The message for the ligand design is that relative flexible structures with fewer internal hydrogen bonds.

PEOE_VSA_POS and *vsurf_R* Correlation plot Fig. 5Cd and 5Ce suggest a negative influence of positive surface area on the FI response. Thus, smaller area of negative surface is preferred for a better affinity to riboswitch O. In contrast, the glucoside of flavonoids tends to have high *PEOE_VSA_POS* values at around 300 and the response FI when binding with riboswitch O tends to be low. *Vsurf_R* is a measure of molecular wrinkled surface (rugosity) [50] made by calculating the ratio of volume and surface. With the negative correlation on FI response, suggests that new molecules could be designed with smaller volume/surface ratio for better binding affinity with riboswitch O.

Artificial naringenin riboswitches represent diverse three-dimensional motifs, which display specific high responsiveness to flavanones, as well as varied responses to derivatives substituted with hydroxyl groups at different positions. Although flavonoids are mostly *O*-glycosylated, these *O*-glycosylated derivatives were not recognizable by any of the riboswitches prepared in this study. The spatial steric hindrance of glycosides in the aptamer complexes can possibly help in

identifying the ligand recognition site using computational prediction tools. The negative surface areas of the molecules were found to have a positive influence on response value of all three riboswitches. The presence of reactive groups like Michael acceptors could result in better binding specificity of riboswitch M1 and M2. Suggestions of the ligand structure modification and complementary RNA structures are provided in terms of hydrophilicity ratio and location of the hydrophilic moieties. Cheminformatics studies represented a powerful and useful technique in the field of the synthetic biology of device design.

CRediT authorship contribution statement

Yu Xiu: designed the study, performed all the experiments, analyzed the data and wrote the manuscript, Formal analysis, Data curation. **Ni Zhang:** finished the QSPR models, Formal analysis, Data curation. **Pranesha Prabhakaran:** Formal analysis, Data curation. **Sungho Jang:** developed the riboswitches. **Qipeng Yuan:** Formal analysis, Data curation. **Curt M. Breneman:** finished the QSPR models. **Gyoo Yeol Jung:** designed the study. **Wanwipa Vongsangnak:** Formal analysis, Data curation. **Mattheos A.G. Koffas:** designed the study, Formal analysis, Data curation.

Declaration of competing interest

We declare that we have no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2022.07.006.

References

[1] Pandey RP, Parajuli P, Koffas MAG, Sohng JK. Microbial production of natural and non-natural flavonoids: pathway engineering, directed evolution and systems/synthetic biology. Biotechnol Adv 2016;34:634–62. https://doi.org/10.1016/J.BIOTECHADV.2016.02.012.

[2] Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. J Nutr Sci 2016;5:1–15.

[3] Wang J, Guleria S, Koffas MAG, Yan Y. Microbial production of value-added nutraceuticals. Curr Opin Biotechnol 2016;37:97–104. https://doi.org/10.1016/J.COPBIO.2015.11.003.

[4] He Fu L, Li G, Andrew Jones J, Linhardt RJ, Koffas M. Production of chondroitin in metabolically engineered E. coli. Metab Eng 2015;27:92–100. https://doi.org/10.1016/j.ymben.2014.11.003.

[5] Zhao S, Jones JA, Lachance DM, Bhan N, Khalidi O, Venkataraman S, et al. Improvement of catechin production in Escherichia coli through combinatorial metabolic engineering. Metab Eng 2015;28:43–53. https://doi.org/10.1016/j.ymben.2014.12.002.

[6] Jones JA, Collins SM, Vernacchio VR, Lachance DM, Koffas MAG. Optimization of naringenin and p-coumaric acid hydroxylation using the native E. coli hydroxylase

complex. HpaBC. Biotechnology Progress 2016;32:21–5. https://doi.org/10.1002/btpr.2185.

[7] Chen X, Xiao J. Microbial biotransformation of bioactive flavonoids. Biotechnol Adv 2015;33:214–23. https://doi.org/10.1016/j.biotechadv.2014.10.012.

[8] Xiao J, Muzashvili TS, Georgiev MI. Advances in the biotechnological glycosylation of valuable flavonoids. Biotechnol Adv 2014;32:1145–56. https://doi.org/10.1016/J.BIOTECHADV.2014.04.006.

[9] Raman S, Rogers JK, Taylor ND, Church GM. Evolution-guided optimization of biosynthetic pathways. Proc Natl Acad Sci USA 2014;111:201409523. https://doi.org/10.1073/pnas.1409523111.

[10] Cress BF, Leitz QD, Kim DC, Amore TD, Suzuki JY, Linhardt RJ, et al. CRISPRi-mediated metabolic engineering of E. coli for O-methylated anthocyanin production. Microb Cell Factories 2017;16. https://doi.org/10.1186/S12934-016-0623-3.

[11] Cress BF, Jones JA, Kim DC, Leitz QD, Englaender JA, Collins SM, et al. Rapid generation of CRISPR/dCas9-regulated, orthogonally repressible hybrid T7-lac promoters for modular, tuneable control of metabolic pathway fluxes in Escherichia coli. Nucleic Acids Res 2016;44:4472. https://doi.org/10.1093/NAR/GKW231.

[12] Cress BF, Trantas EA, Ververidis F, Linhardt RJ, Koffas MAG. Sensitive cells: enabling tools for static and dynamic control of microbial metabolic pathways. Curr Opin Biotechnol 2015;36:205–14. https://doi.org/10.1016/J.COPBIO.2015.09.007.

[13] Tera W, Felipe A, Segura A, Rojas A, Ramos J, Gallegos M-T. Antibiotic-dependent induction of Pseudomonas putida DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. Antimicrob. Agents Chrmother. 2003;47:3067–72. https://doi.org/10.1128/AAC.47.10.3067.

[14] Polycultures E, Cress B, Gross R. Complete biosynthesis of anthocyanins using crossm complete biosynthesis of anthocyanins. 2017. https://doi.org/10.1128/mBio.00621-17.

[15] Jones JA, Vernacchio VR, Lachance DM, Lebovich M, Fu L, Shirke AN, et al. ePathOptimize: a combinatorial approach for transcriptional balancing of metabolic pathways. Sci Rep 2015;5. https://doi.org/10.1038/SREP11301.

[16] Cress BF, Toparlak OD, Guleria S, Lebovich M, Stieglitz JT, Englaender JA, et al. CRISPathBrick: modular combinatorial assembly of type II-A CRISPR arrays for dCas9-mediated multiplex transcriptional repression in E. coli. ACS Synth Biol 2015;4:987–1000. https://doi.org/10.1021/ACSSYNBIO.5B00012.

[17] Lim CG, Wong L, Bhan N, Dvora H, Xu P, Venkiteswaran S, et al. Development of a recombinant Escherichia coli strain for overproduction of the plant pigment anthocyanin. Appl Environ Microbiol 2015;81:6276. https://doi.org/10.1128/AEM.01448-15.

[18] Jones JA, Vernacchio VR, Sinkoe AL, Collins SM, Ibrahim MHA, Lachance DM, et al. Experimental and computational optimization of an Escherichia coli co-culture for the efficient production of flavonoids. Metab Eng 2016;35:55–63. https://doi.org/10.1016/J.YMBEN.2016.01.006.

[19] He L, Xiu Y, Jones JA, Baidoo EEK, Keasling JD, Tang YJ, et al. Deciphering flux adjustments of engineered E. coli cells during fermentation with changing growth conditions. Metab Eng 2017;39:247–56. https://doi.org/10.1016/J.YMBEN.2016.12.008.

[20] Alam KK, Chang JL, Burke DH. FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. Mol Ther Nucleic Acids 2015;4:e230. https://doi.org/10.1038/MTNA.2015.4.

[21] Carothers JM, Goler J a, Juminaga D, Keasling JD. Devices to quantitatively program. Science 1979:1716–9. https://doi.org/10.1126/science.1212209. 2011; 36498.

[22] Hoinka J, Berezhnoy A, Dao P, Sauna ZE, Gilboa E, Przytycka TM. Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. Nucleic Acids Res 2015;43:5699. https://doi.org/10.1093/NAR/GKV308.

[23] Espah Borujeni A, Mishler DM, Wang J, Huso W, Salis HM. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. Nucleic Acids Res 2016;44:1–13. https://doi.org/10.1093/NAR/GKV1289.

[24] Chang AL, Wolf JJ, Smolke CD. Synthetic RNA switches as a tool for temporal and spatial control over gene expression. Curr Opin Biotechnol 2012;23:679–88. https://doi.org/10.1016/J.COPBIO.2012.01.005.

[25] Liu D, Evans T, Zhang F. Applications and advances of metabolite biosensors for metabolic engineering. Metab Eng 2015;31:35–43. https://doi.org/10.1016/J.YMBEN.2015.06.008.

[26] Topp S, Gallivan JP. Emerging applications of riboswitches in chemical biology. ACS Chem Biol 2010;5:139–48. https://doi.org/10.1021/CB900278X.

[27] Wachsmuth M, Findeiß S, Weissheimer N, Stadler PF, Mörl M. De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res 2013;41:2541–51. https://doi.org/10.1093/NAR/GKS1330.

[28] McKeague M, Wong RS, Smolke CD. Opportunities in the design and application of RNA for gene expression control. Nucleic Acids Res 2016;44:2987. https://doi.org/10.1093/NAR/GKW151.

[29] Wittmann A, Suess B. Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. FEBS (Fed Eur Biochem Soc) Lett 2012;586: 2076–83. https://doi.org/10.1016/J.FEBSLET.2012.02.038.

[30] Beisel CL, Chen YY, Culler SJ, Hoff KG, Smolke CD. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. Nucleic Acids Res 2011;39:2981–94. https://doi.org/10.1093/NAR/GKQ954.

[31] Michener JK, Smolke CD. High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. Metab Eng 2012;14:306–16. https://doi.org/10.1016/J.YMBEN.2012.04.004.

[32] Dixon N, Duncan JN, Geerlings T, Dunstan MS, McCarthy JE, Leys D, et al. Reengineering orthogonally selective riboswitches. Proc Natl Acad Sci U S A 2010; 107:2830–5. https://doi.org/10.1073/pnas.0911209107.

[33] Lee SW, Oh MK. A synthetic suicide riboswitch for the high-throughput screening of metabolite production in Saccharomyces cerevisiae. Metab Eng 2015;28: 143–50. https://doi.org/10.1016/J.YMBEN.2015.01.004.

[34] Yang J, Seo SW, Jang S, Shin SI, Lim CH, Roh TY, et al. Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. Nat Commun 2013;4. https://doi.org/10.1038/NCOMMS2404.

[35] Jones JA, Xiu Y, Jang S, Jones JA, Zill NA, Linhardt RJ. Naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli Co-cultures naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli Co-cultures. 2017. https://doi.org/10.1002/bit.26340.

[36] Jang S, Jang S, Xiu Y, Kang TJ, Lee H, Koffas MAG, et al. Development of artificial riboswitches for monitoring of naringenin in vivo development of artificial riboswitches for monitoring of naringenin in vivo. 2017. https://doi.org/10.1021/acssynbio.7b00128.

[37] Miryala B, Zhen Z, Potta T, Breneman CM, Rege K. Parallel synthesis and quantitative structure-activity relationship (QSAR) modeling of aminoglycoside-derived lipopolymers for transgene expression. ACS Biomater Sci Eng 2015;1: 656–68. https://doi.org/10.1021/ACSBIOMATERIALS.5B00045/ASSET/IMAGES/LARGE/AB-2015-000459_0007. JPEG.

[38] Deng W, Breneman C, Embrechts MJ. Predicting protein-ligand binding affinities using novel geometrical descriptors and machine-learning methods. J Chem Inf Comput Sci 2004;44:699–703. https://doi.org/10.1021/ci034246+.

[39] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406–15. https://doi.org/10.1093/NAR/GKG595.

[40] Waugh A, Gendron P, Altman R, Brown JW, Case D, Gautheret D, et al. RNAML: a standard syntax for exchanging RNA information. RNA 2002;8:707–17. https://doi.org/10.1017/S1355838202028017.

[41] Magnus M, Boniecki MJ, Dawson W, Bujnicki JM. SimRNAweb : a web server for RNA 3D structure modeling with optional restraints. https://doi.org/10.1093/nar/gkw279; 2016. 1-5.

[42] Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res 2016;44:e63. https://doi.org/10.1093/NAR/GKV1479.

[43] Zhuo Z. QSAR/QSPR modeling on gene delivery polymers and development of property-encoded protein binding site surface partial comparison method. Rensselaer Polytechnic Institute; 2015.

[44] Hanasaki Y, Ogava S, Fukui S. The correlation between active oxygens scavanging and antioxidative effects of flavonoids. Free Radic Biol Med 1994;16:845–50.

[45] Jez JM, Noel JP. Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. J Biol Chem 2002;277:1361–9. https://doi.org/10.1074/jbc.M109224200.

[46] Ngaki MN, Louie GV, Philippe RN, Manning G, Bowman ME, Li L, et al. Evolution of the chalcone isomerase fold from fatty acid-binding to stereospecific enzyme. Nature 2014;485:530–3.

[47] Of J, Chemistry A. Substituent effects on in vitro antioxidizing properties , stability , and solubility in flavonoids. 2014. https://doi.org/10.1021/jf405570u.

[48] Max-dependency C. Feature selection based on mutual information. Analysis 2005; 27:1226–38. https://doi.org/10.1007/11892755_59.

[49] Mazza CB, Sukumar N, Breneman CM, Cramer SM. Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure. Anal Chem 2001;73:5457–61. https://doi.org/10.1021/AC010797S.

[50] Cruciani G, Crivori P, Carrupt PA, Testa B. Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. J Mol Struct: THEOCHEM 2000;503:17–30. https://doi.org/10.1016/S0166-1280(99)00360-7.

[51] Oprea TI. Property distribution of drug-related chemical databases. J Comput Aided Mol Des 2000;14:251–64. https://doi.org/10.1023/A:1008130001697.

[52] Pearlman RS, Smith KM. Metric validation and the receptor-relevant subspace concept. J Chem Inf Comput Sci 1999;39:28–35. https://doi.org/10.1021/CI980137X.

[53] Comer J, Tam K. Lipophilicity profiles: theory and measurement. G: Folkers; 2001.

[54] Hansch C, Leo A, Hoekman D, Livingstone DJ. Exploring QSAR: hydrophobic, electronic, and steric constants. Washington, DC: American Chemical Society; 1995.