Parallel screening and cheminformatics modeling of flavonoid activated aptasensors

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

Keywords: activated aptasensors

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 flavonoid structures were identified as two properties that increased binding affinity to RNA riboswitches. Hydroxyl groups at the C3′ and C4′ positions of the flavonoid molecule were strictly required for ligand-activation with riboswitches M1 and M2. Riboswitch 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 biosynthesizing 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 reprogramming 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 riboswitches are now being addressed [23,27,30]. The utilization of riboswitches for reprogramming 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 riboswitches are now being addressed [23,27,30].

The utilization of riboswitches for reprogramming 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 riboswitches are now being addressed [23,27,30].

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). Empodoxy, apogenin, chrysin, genkwanin, 2′, 4′-dihydroxychalcone, isoliquiritigenin were purchased from Indofine Chemical Company, Inc (Hillsborough, NJ). Cyanidin chloride, cyanidin 3-O-glucoside chloride, synthetic and systems biotechnology 7 (2022) 1148–1158.

### Table 1

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

| Flavonane                | Parent skeleton | R5  | R7  | R3  | R4  | R5  |
|-------------------------|----------------|-----|-----|-----|-----|-----|
| 1 Naringenin            |                | OH  | OH  | H   | OH  | –   |
| 2 Flavanone             |                | H   | H   | H   | H   | –   |
| 3 Erinodictylose        |                | OH  | OH  | OH  | –   | –   |
| 4 Naringin              |                | OH  | Omg | H   | OH  | –   |
| 5 Hesperidin            |                | OH  | Orha| OH  | OMe | –   |
| 6 Flavan-3-ol           | (+)–Catechin   | OH  | OH  | OH  | OH  | –   |
| 7 (−)–Epicatechin       |                | OH  | OH  | OH  | OH  | –   |

| Flavone                 | R5  | R7  | R3  | R4  | R5  |
|-------------------------|-----|-----|-----|-----|-----|
| 8 Chrysin               | OH  | OH  | H   | H   | H   |
| 9 Apigenin              | OH  | OH  | H   | OH  | H   |
| 10 Luteolin             | OH  | OH  | H   | OH  | H   |
| 11 Genkwanin            | OH  | OMe | H   | OMe | H   |
| 12 Acacetin             | OH  | OH  | OMe | OMe | H   |
| 13 2′,4′-Dimethoxyluteolin | OH  | OH  | Ogr | H   | OH  |
| 15 Tricin               | OH  | OH  | Ogr | H   | OMe |
| 16 Kaempferol           | OH  | OH  | H   | OH  | OH  |
| 17 Quercetin            | OH  | OH  | OH  | OH  | OH  |
| 18 Astraquinin          | Oglc| OH  | OH  | H   | OH  |
| 19 Nicotifolin          | Oglc| OH  | OH  | H   | OH  |
| 20 Isoquercetin         | Oglc| OH  | OH  | OH  | OH  |

| Flavonol                | R5  | R7  | R3  | R4  | R5  |
|-------------------------|-----|-----|-----|-----|-----|
| 21 2′,4′-Dihydroxychalcone | OH  | H   | OH  | H   | H   |
| 22 Isoliquiritigenin     | OH  | H   | OH  | H   | OH  |
| 23 Lanceoletin          | OH  | OMe | OH  | OH  | OH  |

| Anthocyanidin           | R5  | R7  | R3  | R4  | R5  |
|-------------------------|-----|-----|-----|-----|-----|
| 24 Cyanidin             | OH  | OH  | OH  | OH  | OH  |
| 25 Pelargonidin         | OH  | OH  | OH  | OH  | OH  |
| 26 Cyanidin-3-O-glucoside| Oglc| OH  | OH  | OH  | OH  |
| 27 Peonidin-3-O-glucoside| Oglc| OH  | OH  | OMe | OH  |

| Aurocone                | R5  | R7  | R3  | R4  | R5  |
|-------------------------|-----|-----|-----|-----|-----|
| 28 Leptosidin           | Oglc| OH  | OMe | OH  | OMe | 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 combinatorially 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 pACYCDueto-1 was 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 mfold web server (http://unofold.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 analysis [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. QSAR 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 “recessive 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 hydroxylation, O-methylations, O-glycosylations, de-hydrogenations, dehydrogenations 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, 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 hydroxyl 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,
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 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, apigenin and luteolin and the flavonol quercetin, 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. 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), 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-methylation of hydroxyl groups (Fig. 1). Tricin, 7-O-methylated flavone genkwanin, and 4’-O-methylated flavone acacetin which can be derived from apigenin, and 3’, 4’-dimethoxy luteolin, which can result from luteolin 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 and luteolin.

Interestingly, all riboswitches responded well to the three chalcones. 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. 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. 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.
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 (Fig. 2) as considered 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, nicotifolin 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 F1 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 Mod O (Table 2).

Least square-support vector regression (LS-SVR) was shown to perform the best with $r^2$ value of 0.7982, $R^2$ value of 0.7862 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 0 with $r^2$ of 0.7759, $R^2$ of 0.7747 and R M S D 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. QSAR 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 physicochemical properties on the flavonoids’ binding specificity with the RNA aptamers.

### 3.5. Riboswitch M1 model descriptors

| Model | Rank | Descriptor | Class | Descriptor Definition |
|-------|------|------------|-------|-----------------------|
| M1    | 1    | Q_VSA_FNEG | 2D    | Fractional negative vDW surface area |
|       | 2    | FASA-     | 3D    | Fractional negative accessible surface area |
|       | 3    | vsurf_CW1 | 3D    | 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     | 3D    | Angle bend potential energy |
|       | 7    | vsurf_EWmin3 | 3D | 3rd lowest hydrophilic energy |
|       | 8    | vsurf_JWS | 3D    | Hydrophilic integral 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_zring | 2D    | Oprea Ring Count |
|       | 6    | Q_VSA_FNEG | 2D    | Fractional negative vdw surface area |
| O     | 1    | BCU_SLOGP_0 | 2D | Log of the octanol/water partition coefficient |
|       | 2    | GCUT_SMR_2 | 2D    | Molar refractivity GCUT(2/3) |
|       | 3    | E_ang     | 3D    | 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   | 3D    | Surface roughness |

**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 Brunsted 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 eriodyctiol that demonstrated best response values have a larger portion of negative surface than the 3′, 4′-dimethoxy luteolin 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 unit surface [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, thiois, 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 eriodyctiol with luteolin shows that the presence of aliphatic ketone as a reactive...
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.
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.

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

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.
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 hydro-
phobic properties, as there is less water present in the active site region.

Vsurf_IW5 Like dipole moments, interaction energy moments ex-
press 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 point-
ning 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 devel-
oped initially for study of molecular similarity [52]. GCUT descriptors
are calculated from the eigenvalues of a modified adjacent matrix. Each
i j 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 ei-
genvalues 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 mole-
cules, 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 $\mathrm{2, 4'}$-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 number of negative surface are prone to have better binding affinity 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 $1/\sqrt{\text{bij}}$ 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_angle** 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 few 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 flavonones, 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. Pranesh Prabhakaran: Formal analysis, Data curation. Sung-ho Jang: developed the riboswitches. Qi Peng Yuan: Formal analysis, Data curation. Curt M. Breneman: finished the QSPR models. Gyoo Yeol Jung: designed the study. Wannipa 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.

Acknowledgments

This work was supported by the Global Research Laboratory Program (NRF 2016K1A1A2912829) through the National Research Foundation of Korea (NRF). This work was also supported by the National Natural Science Foundation of China (grant number 21636001) and Research Program of Beijing Municipal Education Commission (grant number KM20191417012). This work was also financially supported by the Office of the Ministry of Higher Education, Science, Research and Innovation; and the Thailand Science Research and Innovation through the Kasetsart University Reinvinting University Program 2021 and Department of Zoology, Faculty of Science, Kasetsart University. P.P. would like to acknowledge the supports provided by the International Affairs under the scholarship program of ASEAN+6 and Interdisciplinary Graduate Program in Bioscience in Kasetsart University.

The authors would like to thank Dr. Mamoru Koketsu for a generously help in flavonoids synthesis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sysbio.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;34:634-642. 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.COBIO.2015.11.003.

[4] He Pu Li, 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.meb.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.meb.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
Synthetic and Systems Biotechnology 7 (2022) 1148–1158

complex. HpaBC. Biotechnology Progress 2016;32:21–5. https://doi.org/10.1002/btp.1955.

[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, Muzahidvi TS, Georgiev ML. 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, Suzuki JY, Linhardt RJ, et al. CRISPRi-mediated metabolic engineering of E. coli for O-methylated anthocyanin production. Microbiol Cell Factories 2017;16. https://doi.org/10.1186/s12934-016-0623-3.

[10] Cress BF, Leitz QD, Kim DG, Amore TD, Suzuki JY, Linhardt RJ, et al. CRISPRi-mediated metabolic engineering of E. coli for O-methylated anthocyanin production. Microbiol Cell Factories 2017;16. https://doi.org/10.1186/s12934-016-0623-3.

[11] Cress BF, Jones KA, Kim DG, Leitz QD, Englaender JA, Collins SM, et al. Rapid generation of CRISPR/dsCas9-regulated, orthogonally represedible TetR-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 metabolic engineering. Metab Eng 2015;31:35. https://doi.org/10.1016/J.COPBIO.2012.01.005.

[14] Jang S, Jia X, Xiu Y, Jones JA, Zill NA, Linhardt RJ. Naringenin-responsive riboswitch-based fluorescence biosensor module for Escherichia coli. Curr Opin Biotechnol 2017. https://doi.org/10.1016/j.copbio.2017.05.005.

[15] Jones JA, Xiu Y, Jang S, Jones JA, Zill NA, Linhardt RJ. Naringenin-responsive riboswitch-based fluorescence biosensor module for Escherichia coli. Curr Opin Biotechnol 2017. https://doi.org/10.1016/j.copbio.2017.05.005.

[16] Jang S, Jia X, Xiu Y, Fang 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/acs.chembio.6b00715.

[17] Raman S, Rogers JK, Taylor ND, Church GM. Evolution-guided optimization of synthetic riboswitches for gene expression control. Nucleic Acids Res 2016;44:4472. https://doi.org/10.1093/NAR/GKW231.

[18] Cress BF, Barbosa FL, Leitz QD, Linhardt RJ, Koffas MAG. Synthetic riboswitches for metabolic engineering. Metab Eng 2015;31:35. https://doi.org/10.1016/J.COPBIO.2012.01.005.

[19] Raman S, Rogers JK, Taylor ND, Suzuki JY, Linhardt RJ, et al. CRISPRi-mediated metabolic engineering of E. coli for O-methylated anthocyanin production. Microbiol Cell Factories 2017;16. https://doi.org/10.1186/s12934-016-0623-3.

[20] Jia X, Xiu Y, Jang S, Jones JA, Zill NA, Linhardt RJ. Naringenin-responsive riboswitch-based fluorescence biosensor module for Escherichia coli. Curr Opin Biotechnol 2017. https://doi.org/10.1016/j.copbio.2017.05.005.

[21] Jia J, Jang S, Xiu Y, Tang YJ, 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/acs.chembio.6b00715.

[22] Y. Xiu et al. Synthetic and Systems Biotechnology 7 (2022) 1148–1158