Classification of seed members of five riboswitch families as short sequences based on the features extracted by Block Location-Based Feature Extraction (BLBFE) method

Faegheh Golabi1,2*, Elnaz Mehdizadeh Aghdam1,4, Mousa Shamsi1*, Mohammad Hossein Sedaaggi3,4, Abolfazl Barzegar6, Mohammad Saeid Hejazi3,4*

1 Genomic Signal Processing Laboratory, Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran
2 Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran
3 Molecular Medicine Research Center, Biomedicine Institute, Tabriz University of Medical Sciences, Tabriz, Iran
4 Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
5 Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
6 Research Institute of Bioscience and Biotechnology, University of Tabriz, Tabriz, Iran

Abstract

Introduction: Riboswitches are short regulatory elements generally found in the untranslated regions of prokaryotes' mRNAs and classified into several families. Due to the binding possibility between riboswitches and antibiotics, their usage as engineered regulatory elements and also their evolutionary contribution, the need for bioinformatics tools of riboswitch detection is increasing. We have previously introduced an alignment independent algorithm for the identification of frequent sequential blocks in the families of riboswitches. Herein, we report the application of block location-based feature extraction strategy (BLBFE), which uses the locations of detected blocks on riboswitch sequences as features for classification of seed sequences. Besides, monoand dinucleotide frequencies, k-mer, DAC, DCC, DACC, PC-PseDNC-General and SC-PseDNC-General methods as some feature extraction strategies were investigated.

Methods: The classifiers of the Decision tree, KNN, LDA, and Naïve Bayes, as well as k-fold cross-validation, were employed for all methods of feature extraction to compare their performances based on the criteria of accuracy, sensitivity, specificity, and f-score performance measures.

Results: The outcome of the study showed that the BLBFE strategy classified the riboswitches indicating 87.65% average correct classification rate (CCR). Moreover, the performance of the proposed feature extraction method was confirmed with average values of 94.31%, 85.01%, 95.45% and 85.38% for accuracy, sensitivity, specificity, and f-score, respectively.

Conclusion: Our result approved the performance of the BLBFE strategy in the classification and discrimination of the riboswitch groups showing remarkable higher values of CCR, accuracy, sensitivity, specificity and f-score relative to previously studied feature extraction methods.

Article Info

Article Type: Original Article
Article History:
Received: 10 Nov. 2019
Revised: 12 Mar. 2020
Accepted: 20 Mar. 2020
ePublished: 17 Apr. 2020
Keywords: Riboswitches Feature extraction Block-finding algorithm BLBFE Classification

Introduction

Riboswitches, usually found in the 5’-UTR of mRNA, act as specific and selective sensors of various intracellular ligands. In the mechanistic view, the riboswitches' aptamer domain ligand binding triggers conformational changes of the expression platform leading to the downstream gene regulation.1

The rise in antibiotic resistance, as well as limitations in available effective antibiotics, motivate the scientists in the field to seek novel antibacterial compounds to battle these life-threatening phenomena.2 Since the discovery of riboswitches, they were proposed as promising antibacterial targets for the following reasons.3,4 First, the high selectivity and specificity of riboswitches propose them as suitable RNA receptors in antibacterial compound design.5 Second, limited distribution of riboswitches...
mostly in prokaryotes, theoretically prevents cross-reactivity of candidate bacterial riboswitch sensing compounds with nonbacterial, especially eukaryotic cells. Third, riboswitches are present in the mRNA of survival or/and virulence genes, as a result, targeting of such riboswitches enhances the possibility of developing bactericidal or bacteriostatic compounds with a different mechanism. Accordingly, several types of riboswitches have been applied for antibiotic development. However, few successful compounds have been introduced, so far. One of the latest and promising compounds, “Ribocil,” was developed as a bacteriostatic compound and a highly robust riboflavin riboswitches inhibitor by Merck company.

On the other side, mutations in riboswitches could give rise to ligand and antibiotic resistance. It is reported that some well-known antibiotics exert their effects via specific or unspecified binding to different classes of riboswitches. For example, roseoflavin and pyrithiamine, are specific binders of FMN and TPP riboswitches, respectively, and conformational changing mutations in the riboswitches could trigger the antibiotic resistance occurrences in bacteria.

Overall, riboswitches are important platforms regarding drug design researches. Furthermore, they could be applied to design robust artificial regulatory elements in constructs using for gene therapy, biosensors, etc. Besides, riboswitches are important sequences from the evolutionary point of view and one of the evidence of RNA world hypothesis. As a result, any tools to find and introduce new representatives of known/unknown riboswitches could be a significant aid in the field.

Nowadays, more than 40 different classes of riboswitches have been reported. Each class is usually nominated based on the ligand that binds to the riboswitches specifically and selectively. Some other riboswitches are assigned as “orphan riboswitches” as their ligands are unknown, however, their sequence and structure highly resemble a riboswitch pattern. As, the aptamer domain of the riboswitches could be a significant aid in the field.

Accordingly, various methods including hidden Markov models (HMM), context-sensitive HMMs (cs-HMM), profile-HMM, and CM or covariance model have been developed to classify the riboswitches. However, such methods have limitations mainly outliers that do not fit with consensus sequences or secondary structures, and also rely on a large database for alignment and homology search.

This is why we have recently developed an alignment-free method of block finding which detects common sequential blocks in riboswitch classes. The method was able to spot 21 particular blocks with the size of 3 to 9 nucleotides to verify riboswitch families of FMN, PreQ1, Purine, SAM and TPP. Using the positions of detected blocks, a new feature extraction method called “block location-based feature extraction” (BLBFE) was designed and used for the classification of 400 random “full” sequences from 5 mentioned riboswitch families. Moreover, seven other families of riboswitches including cobalamin, cyclic-di-GMP-I, glycine, lysine, SAH, SAM-alpha, and SAM-IV riboswitches were studied using the block finding algorithm and BLBFE. In the present study, we used the BLBFE method to classify the seed members of FMN, PreQ1, Purine, SAM, and TPP riboswitch families, based on their family-specific blocks. Mono- and dinucleotide frequencies, k-mer, DAC, DCC, DACC, PC-PseDNC-General and SC-PseDNC-General, as seven separate feature extraction methods, were also considered for their comparative performances against the BLBFE method. Decision tree, k-nearest neighbors (KNN), linear discriminant analysis (LDA) and Naïve Bayes classifiers accompanied by k-fold cross-validation were utilized to detect the classes of riboswitches with all feature extraction methods. Applying the confusion matrices, the calculation of the performance measures including accuracy, sensitivity, specificity and f-score were carried out to compare the performance efficacy and validity of BLBFE and the other feature extraction methods in the categorization of the riboswitches.

Materials and Methods

Datasets

The seed members of riboswitches families of FMN, PreQ1, Purine, SAM, and TPP, containing 144, 41, 133, 433, and 115 members in each class, respectively, were included in this study (Table 1). Datasets were acquired from the Rfam 12.0 database. We applied a new feature extraction strategy for the classification of riboswitches. In this approach, we used the positions of family-specific sequential blocks on different riboswitches as features. The sequential blocks were detected using our former described sequential block finding technique. The method is specific for being an alignment-free based method, searching for the presence of conserved motifs in the members of each riboswitch family and determines the sequential blocks characterization of each family. The family-specific sequential blocks for 5 riboswitch classes used in the present study are listed in Table 2.

In BLBFE, the observations related to riboswitch family members were produced according to the incidence and the location of the blocks in the sequences. The location of the blocks in the sequence was determined by the start
nucleotide of each block. The location was set to zero and for
the absent blocks in the sequence. For example, in the
sample sequence of:
“CCGCAUUCAGGGCAGGGUGAAAUUC-CUACUGCCGUAACCCGGAGCGUUUGUUU-UAAGG”
the sample blocks of ‘GCCC’, ‘UAUA’ and ‘GGUG’ are
present in locations of 43, 59 and 18 from the start of
the sequence, respectively. This is while the sample blocks
of ‘GGUUC’, ‘GCAACC’, ‘GUGC’ and ‘ACCG’ are absent and
their locations are considered as zero. This means that
considering the template of [GGUUC, GCAACC, GUGC, UAUA, GGUU] and using the proposed method
of BLBFE, the sample sequence above is converted to an
array with the length equal to the number of blocks: [0, 43,
0, 0, 59, 18, 0]
Accordingly, each riboswitch in each class was
represented by an observation. As 21 blocks are detected
for 5 families, the produced observations were 1 by
21 arrays and the value of observations equaled to 5 in this study.

### Cross-validation

The generalization of the classifiers was validated using
k-fold cross-validation (CV), which is the most popular
cross-validation procedure due to its mild computational
cost. It confirms the difference between training and
validation sets. In the standard k-fold cross-validation,
the initial training data is randomly split into k partitions
(folds) of almost equal cardinality. The classification
algorithm is iteratively trained on k-1 folds while using the
remaining fold (called the “holdout fold”) as the validation
set. The mean of the correct classification levels of k stages
shows the overall correct classification rate (CCR). Here,
k=10 folds was used because of suitably low computational
cost in addition to the good error estimation.

### The classifiers

To compare the performances of the proposed feature
extraction methodology and previously studied feature
extraction approaches, four classifiers were employed.

#### Decision tree classifier

Indecision tree learning, a common data mining method,
a predictive model is created to produce an output variable
using several input variables. For this, the observations are
represented in the branches while the conclusions about
the class label of any data are shown in the leaves.

#### K-nearest neighbors (KNN) classifier

In this type of classification, an input data is categorized
into a certain class, based on a greater number of votes of
the class’s K nearest neighbors. K is typically equivalent
to a positive number of classes. Accordingly, the K was
equaled to 5 in this study.

#### Linear discriminant analysis (LDA) classifier

This is a method for pattern recognition, machine
learning, and statistics, to separates or characterizes two

---

**Table 1.** The seed data related to five families of riboswitches employed in this study, obtained from the Rfam 12.0 database.

| Rfam accession number | Riboswitch family name | Number of seed data |
|-----------------------|------------------------|---------------------|
| RF00050               | FMN                    | 144                 |
| RF00522               | PreQ1                  | 41                  |
| RF00167               | Purine                 | 133                 |
| RF00162               | SAM                    | 433                 |
| RF00059               | TPP                    | 115                 |

**Table 2.** The identified frequent blocks for 5 riboswitch families using block finding algorithm.

| Blocks                  | Riboswitch family name |
|-------------------------|------------------------|
| ACCG, CCGAC, CGGU, GGAUG, GGGC, GUGG, UCAC | FMN                   |
| AAAAAACUA, CCC, GGUC    | PreQ1                 |
| UAUA, UCUACC            | Purine                 |
| AGA, AUUC, GAGGGA, GCAACC, GCCC, GUGC | SAM                   |
| ACCUG, CUGAGA, GGG      | TPP                    |
or more classes of items using a linear combination of features.\(^{33}\)

**Naïve Bayesian classifier**

This is a simple probabilistic classifier used in machine learning. Moreover, this classifier is assuming strong (naïve) independence between the features, using Bayes’ theorem. It means the independence of the value of a particular feature from the value of other features.\(^{34-36}\)

**Evaluation of classifiers’ performance**

Based on the confusion matrices, the equations (1) to (4) were used to calculate the accuracy, sensitivity, specificity, and f-score as important performance factors\(^{49-51}\):

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)
\]

\[
\text{Specificity} = \frac{TN}{FP + TN} \quad (3)
\]

\[
F - \text{score} = \frac{2TP}{2TP + FP + FN} \quad (4)
\]

The number of members of each class that are classified into their true class is denoted by TP. On the other hand, FP shows the number of false positives, the sequences which are falsely annotated to a class. Also, TN and FN are the number of true negatives and false negatives, respectively.

Moreover, the CCR for each classification was calculated by dividing the number of total truly classified sequences (sum of TPs for all families) by the total number of classified sequences and was reported as a percentage. The illustration of the current study’s program flow is shown in Fig. 1.

**Results and Discussion**

**Classification results**

In our previous study, the successful application of a sequential based block finding algorithm was shown to be used for the detection of family-specific blocks of the riboswitches.\(^{26}\) In this method, the detection and determination of 21 specific blocks [GGGC, GGUG, UCCC, ACCG, CGGAC, GGAUG, GGUUC, CCC, AAAAAACUA, UAUA, UCACC, AUC, AGA, GAGGGA, GCCC, GCAACC, GUGC, GGG, CUGAGA, ACCUG] were carried out for 5 riboswitch families. Later, we proposed a BLBFE method which creates annotations for riboswitch classes.\(^{27,28}\) Using the same method in the present study, the locations of the 21 detected blocks were taken as family members’ features which led to 1 by 21 arrays of observations. Every array represents each riboswitch sequence for the developed classifier. Consequently, as 866 members were present in the five studied riboswitch classes, 866 arrays were generated of 1 by 21 observations.

The frequency of 20 mono- and dinucleotides including A, C, G, U, AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, and UU in all of 866 members, were considered as the second set of features, resulting in the generation of 866 arrays of 1 by 20 as observations. In addition, 866 observation arrays of 1 by 16 were obtained using the k-mer technique of feature extraction with k=2 to produce pseudo components for 866 riboswitches. The DAC approach of feature extraction with lag=4 resulted in 24 features and generated 866 arrays of 1 by 24 observations. Also, the feature extraction methods of DCC and DACC with lag=1, produced 866 observation arrays of 1 by 30 and 866 observation arrays of 1 by 36, respectively. The "slide", "roll", "rise", "shift", "tilt" and "twist" were chosen as the physicochemical features for DAC, DCC and DACC methods.

Furthermore, 26 and 22 features were generated considering lambda=10 for the PC-PseDNC-General and lambda=1 for the SC-PseDNC-General methods, respectively, with weight=0.5 for both of them. Therefore, PC-PseDNC-General and SC-PseDNC-General methods generate 866 observation arrays of 1 by 26 and 866 observation arrays of 1 by 22, respectively. Similarly, the "slide", "roll", "rise", "shift", "tilt" and "twist" were chosen as physicochemical features for the last described methods of feature extraction.

---

**Fig. 1.** The program flow of this study is illustrated. The BLBFE method, as well as 7 other feature extraction methods, are used for the riboswitches and their classification performances were evaluated and compared.
When using any of the 8 methods of feature extraction, 144, 41, 133, 433, and 115 observations out of the 866 ones belonged to FMN, PreQ1, Purine, SAM, and TPP families, respectively. Decision tree, KNN, LDA, and Naïve Bayes classifiers along with the 10-fold cross-validation were utilized for each set of observations and correct and incorrect classification of riboswitches for every set were considered to calculate the corresponding CCRs.

The CCRs of the classifications are shown in Fig. 2. Using the BLBFE method, the highest CCR with 96.88% belonged to the KNN classifier while the CCR of 63.97% for the Naïve Bayes classifier indicated a poor performance. Among the eight studied feature extraction methods, the proposed method of BLBFE showed the highest level of 89.45% for the mean CCR of four classifiers. Using the mono- and dinucleotide frequency-based method, the mean CCR level decreased to 76.03%.

The methods of feature extraction other than BLBFE also showed relatively less CCR. The value of average CCRs for the different methods of k-mer, DAC, DCC, DACC, PC-PseDNC-General and SC-PseDNC-General were 75.24%, 70.76%, 76.88%, 76.61%, 72.77% and 75.7%, respectively.

\section*{Evaluation results}

The multiclass confusion matrix presentation for the decision tree classifier with the BLBFE method is shown in Table 3. As mentioned before, TP is the number of true positives, TN reflects the number of true negatives, FP shows the number of false positives and FN represents the number of false negatives. The multiclass confusion matrices for KNN, LDA and Naïve Bayes classifiers with the BLBFE method are represented in the supplementary file. The multiclass confusion matrices were also generated for the other seven feature extraction techniques. Consequently, according to the confusion matrices, the measures of accuracy, sensitivity, specificity, and f-score for all classifiers were calculated based on the BLBFE method and presented in Table 4.

The calculated performance measures for the studied classifiers according to the feature extraction methods are presented in Table 4. For each classifier, the highest measure among the feature extraction methods is bolded. As can be seen, the BLBFE method showed the best performance according to all five factors of CCR, accuracy, sensitivity, specificity and f-score for the decision tree as well as KNN classifiers. Also, LDA classifier, showed reasonably good performance based on 3 measures of CCR, accuracy and specificity, while the two factors of sensitivity and f-score were also acceptably high. Only the naïve Bayes classifier performed relatively poor in company with the BLBFE method.

All the studied BLBFE method’s performance parameters are also demonstrated in Fig. 3. Accordingly, the best performance with the accuracy of 98.73 % was shown for the KNN classifier, which is the highest among 4 classifiers. This is while the Naïve Bayes classifier with an accuracy of 82.72% showed the poorest performance. This result was also confirmed by other evaluation measures. The highest sensitivity also belonged to KNN classifier which equals 95.4% and the lowest sensitivity was 59.66% for the Naïve Bayes classifier. The specificity factor of 99.04% was shown for the KNN classifier whereas the lowest specificity of 85.57% was presented for the Naïve Bayes classifier. All the same, the highest f-score of 95.88% is presented for the KNN classifier. Similarly, the Naïve Bayes classifier showed the lowest f-score of 60.69 %.

Fig. 4A-D shows accuracy, sensitivity, specificity, and f-score parameters regarding all the methods of feature extraction in addition to the average performance measures for each method.

Considering the other 7 feature extraction methods, the DCC method has an average accuracy of 92.63% which is the highest, while DACC and SC-PseDNC-General methods are closely behind with average accuracies of 92.38% and 92%, respectively. This is still lower than the
The BLBFE method’s average accuracy of 94.31%. The highest accuracy among the 7 methods is 97.83% which belongs to the LDA classifier with the DACC feature extraction. Only the naïve Bayes classifier shows relatively higher accuracy when using the other methods of feature extraction compared to the BLBFE method.

Again the BLBFE method with the average sensitivity of 85.01% has the best performance. Within the other methods of feature extraction, 79.18% average sensitivity was the highest for the k-mer method. The SC-PseDNC-General and DACC methods were in the following order with the mean sensitivity scores of 78.56% and 78.48%, respectively.

The factor of specificity performance demonstrates similar outcomes. The highest average specificity of 95.45% belonged to the BLBFE method whereas the DCC and DACC methods with values of 94.2% and 94.01%, respectively have the highest average specificities among the other seven feature extraction methods.

At last, the BLBFE method with the f-score of 95.88% has shown the highest level with the KNN classifier. It also presents the highest mean f-score level of 85.38% whereas the average f-score of DCC method was 80.38%.

Comparing the methods of feature extraction, the application of BLBFE method has shown a better performance of classifiers. As can be seen, excluding the Naïve Bayes classifier, better measures resulted from the other three classifiers using the BLBFE method. On the other hand, the average levels of 94.31%, 85.01%, 95.45% and 85.38% for the accuracy, sensitivity, specificity and f-score for the BLBFE method with, respectively, showed far better results compared to the other methods such as the DCC method with the maximum average accuracy of 92.63%, the maximum average specificity of 94.2% and the maximum average f-score of 80.38% and the k-mer method with a maximum average sensitivity of 79.18%.

**Conclusion**

We studied the application of the BLBFE method, a developed feature extraction scheme that utilizes the existence and the location of sequential blocks as features to classify the riboswitches. Besides, seven additional
The common methods for riboswitch classification are alignment-based approaches. An alignment-free algorithm was designed for the detection of specific blocks in riboswitch families. BLBFE strategy which was recently introduced, used the locations of the blocks to classify riboswitches. Employing the BLBFE method accompanied by KNN or decision tree classifiers results in better performance in the classification of riboswitches.
Novel riboswitch-binding flavin analog that protects mice against Clostridium difficile infection without inhibiting cecal flora. *Antimicrob Agents Chemother* 2015; 59: 5736-46. https://doi.org/10.1128/aac.01282-15

Serganov A, Huang L, Panke S. Design and antimicrobial action of a purine analogue that binds Guanine riboswitches. *ACS Chem Biol* 2009; 4: 915-27. https://doi.org/10.1021/cb900146k

Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci JB, et al. Selective small-molecule inhibition of an RNA structural element. *Nature* 2015; 526: 672-7. https://doi.org/10.1038/nature15542

Kim JN, Blount KF, Puskarz I, Lim J, Link KH, Breaker RR. Design and antimicrobial action of a purine analogue that binds Guanine riboswitches. *ACS Chem Biol* 2009; 4: 915-27. https://doi.org/10.1021/cb900146k

Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci JB, et al. Selective small-molecule inhibition of an RNA structural element. *Nature* 2015; 526: 672-7. https://doi.org/10.1038/nature15542

Serganov A, Huang L, Patel DJ. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* 2005; 433: 44-9. https://doi.org/10.1038/nature03487

The rise of antibiotic-resistant infections. *FDA Consum* 1995; 29: 11-5.

2. Lewis R. The rise of antibiotic-resistant infections. *FDA Consum* 1995; 29: 11-5.

Blount KF, Breaker RR. Riboswitches as antibacterial drug targets. *Nat Biotechnol* 2006; 24: 1558-64. https://doi.org/10.1038/nbt1268

Blount KF, Wang JX, Lim J, Sudarsan N, Breaker RR. Antibacterial lysine analogs that target lysine riboswitches. *Nat Chem Biol* 2007; 3: 44-9. https://doi.org/10.1038/nchembio842

5. Penchovsky R, Stolovla CC. Riboswitch-based antibacterial drug discovery through high-throughput screening methods. *Expert Opin Drug Discov* 2013; 8: 65-82. https://doi.org/10.1517/17460441.2013.740545

6. Cheah MT, Wachter A, Sudarsan N, Breaker RR. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 2007; 447: 497-500. https://doi.org/10.1038/nature05769

7. Gao F, Luo H, Zhang CT, Zhang R. Gene essentiality analysis based on DEG 10, an updated database of essential genes. *Methods Mol Biol* 2015; 1279: 219-33. https://doi.org/10.1007/978-1-4939-2398-4_14

8. Blount KF, Megyola C, Plummer M, Osterman D, O’Connell T, Aristoff P, et al. Novel riboswitch-binding flavin analog that protects mice against Clostridium difficile infection without inhibiting cecal flora. *Antimicrob Agents Chemother* 2015; 59: 5736-46. https://doi.org/10.1128/aac.01282-15

9. Multibacher J, Brouillette E, Allard M, Fortier LC, Malouin F, Lafontaine DA. Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog* 2010; 6: e1000865. https://doi.org/10.1371/journal.ppat.1000865

10. Kim JN, Blount KF, Puskarz I, Lim J, Link KH, Breaker RR. Design and antimicrobial action of a purine analogue that binds Guanine riboswitches. *ACS Chem Biol* 2009; 4: 915-27. https://doi.org/10.1021/cb900146k

11. Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci JB, et al. Selective small-molecule inhibition of an RNA structural element. *Nature* 2015; 526: 672-7. https://doi.org/10.1038/nature15542

12. Serganov A, Huang L, Patel DJ. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* 2009; 458: 233-7. https://doi.org/10.1038/nature07642

13. Sudarsan N, Cohen-Chalamin S, Nakamura S, Emilsson GM, Shamsi M, Sedaaghi MH, et al. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

14. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

15. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new oligonucleotide block location-based feature extraction (BLRFE) method for the classification of riboswitches. *Bioimpacts* 2018; 8: 15-24.

16. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new oligonucleotide block location-based feature extraction (BLRFE) method for the classification of riboswitches. *Bioimpacts* 2018; 8: 15-24.

17. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

18. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

19. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

20. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

21. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

22. Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 1998; 26: 544-8.

23. Yoon B, Vaidyanathan P, editors. HMM with auxiliary memory: a new tool for modeling RNA secondary structures. Proc 38th Asilomar Conference on Signals, Systems, and Computers; 2004.

24. Singh P, Bandyopadhyay P, Bhattacharya S, Krishnamachari A, Sengupta S. Riboswitch detection using profile hidden Markov models. *BMC Bioinformatics* 2009; 10: 325. https://doi.org/10.1186/1471-2105-10-325

25. Eddy SR, Durbin R. RNA sequence analysis using covariance models. *Nucleic Acids Res* 1994; 22: 2079-88.

26. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.

27. Golabi F, Shamsi M, Sedaaghi MH, Barzegar A, Hejazi MS. Development of a new sequential block finding strategy for detection of conserved sequences in riboswitches. *Bioimpacts* 2018; 8: 15-24.
Classificaion of five families of riboswitches using BLBFE method

42. Friedel M, Nikolajewa S, Sühnel J, Wilhelm T. DiProDB: a database for dinucleotide properties. *Nucleic Acids Res* 2009; 37: D37-D40. https://doi.org/10.1093/nar/gkn597
43. Dong Q, Zhou S, Guan J. A new taxonomy-based protein fold recognition approach based on autocross-covariance transformation. *Bioinformatics* 2009; 25: 2655-62. https://doi.org/10.1093/bioinformatics/btp500
44. Guo Y, Yu L, Wen Z, Li M. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. *Nucleic Acids Res* 2008; 36: 3025-30. https://doi.org/10.1093/nar/gkn159
45. Chen W, Zhang X, Brooker J, Lin H, Zhang L, Chou KC. PseKNC-General: a cross-platform package for generating various modes of pseudo nucleotide compositions. *Bioinformatics* 2015; 31: 119-20. https://doi.org/10.1093/bioinformatics/btu602
46. Breiman L, Spector P. Submodel Selection and Evaluation in Regression. The X-Random Case. *International Statistical Review* 1992; 60: 291-319. https://doi.org/10.2307/1403680
47. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. Montreal, Quebec, Canada; 1995. p. 1137-43.
48. Braga-Neto UM, Dougherty ER. Is cross-validation valid for small-sample microarray classification? *Bioinformatics* 2004; 20: 374-80. https://doi.org/10.1093/bioinformatics/btg419
49. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett* 2006; 27: 861-74. https://doi.org/10.1016/j.patrec.2005.10.010
50. Sun Y, Kamel MS, Wong AKC, Wang Y. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recogn* 2007; 40: 3358-78. https://doi.org/10.1016/j.patcog.2007.04.009
51. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 2009; 45: 427-37. https://doi.org/10.1016/j.ipm.2009.03.002