Investigation of Multi-Objective Optimization Criteria for RNA Design

David J. D. Hampson*, Sinem Sav†, and Herbert H. Tsang‡

*‡Applied Research Lab, Trinity Western University, Langley, British Columbia, Canada

†Faculty of Engineering & Computer Science, Bilkent University, Ankara, Turkey

Email: herbert.tsang@twu.ca‡

Abstract—**RNA design is the inverse of RNA folding and it appears to be NP-hard. In RNA design, a secondary structure is given and the goal is to find a nucleotide sequence that will fold into this structure. To find such sequence(s) involves exploring the exponentially large sequence space. In literature, heuristic algorithms are the standard technique for tackling the RNA design. Heuristic algorithms enable effective and efficient exploration of the high-dimensional sequence-structure space when searching for candidates that fold into a given target structure. The main goal of this paper is to investigate the use of multi-objective criteria in SIMARD and Quality Pre-selection Strategy (QPS). The objectives that we optimize are Hamming distance (between designed structure and target structure) and thermodynamic free energy. We examine the different combinations of optimization criteria, and attempt to draw conclusions about the relationships between them. We find that energy is a poor primary objective but makes an excellent secondary objective. We also find that using multi-objective pre-selection produces viable solutions in far fewer steps than was previously possible with SIMARD.**

I. INTRODUCTION

The RNA design problem is one of the NP-Hard problems [1] in the field of bioinformatics and it refers to the procedure of determining an RNA primary sequence given its secondary structure. Thus, it is the reverse of RNA secondary structure prediction. As the function of RNA is determined by its secondary structure, researchers in the field are interested in this problem to pave the way for new biotechnology and medicine researches such as customized drug design.

RNA secondary structure prediction is a well studied computational problem [2], [3], [4], [5], [6], [7], [8], [9]. In general, there are two types of approaches to study RNA secondary structures: 1) the single sequence approaches which predict the secondary structure based on experimentally determined energy parameters and 2) comparative sequence analysis approaches that try to improve their results by using functionally related sequences. The most popular single sequence approach to structure prediction methodology is the minimum free energy method. Among the different approaches, heuristic algorithms are a very popular and successful approach when compared to other deterministic approach. Our lab has developed a heuristic algorithm to predict RNA secondary structure and it has shown good prediction result even in pseudoknotted structures [10] [11] [12] [13].

Most of the RNA design and prediction problems are structured as optimization problems. In general, they require a long run-time. The obvious goal for improvement will be to decrease the run-time and to increase the quality of the solution at the same time.

Recently, we have introduced SIMARD, a RNA prediction algorithm based on simulated annealing [14]. SIMARD is structured as a single-objective optimization problem and it minimized over Hamming distance between the designed structure and desired structure. We have seen promising results by using Hamming distance only.

In this paper, we utilized the simulated annealing framework in SIMARD, and additionally incorporate a multi-objective optimization approach. We examined the result of employing both Hamming distance and free energy objectives. In the subsequent sections, the method, experimental setup, data used and results are expanded on in detail.

II. MULTI-OBJECTIVE OPTIMIZATION

A. Computational Intelligence for RNA Prediction and Design

Current algorithms in the field for RNA prediction and design are generally using global/local sampling methods, dynamic programming, stochastic searches, evolutionary algorithms and context free grammars. Most of them lack a robust optimization methodology as our previous experiments suggest [14]. Thus, they generally suffer from lack of reasonable run-time or better results.

Most RNA design algorithms are based on the heuristic approach. Some of the well-known single objective optimization RNA design packages are RNA-SSD [15] and INFO-RNA [16] using local stochastic searches, RNA-ensign [17] and improved version of RNA-ensign called IncaRNAtion [18] to decrease time complexity using global sampling methodology, MODENA [19], Frnakestein [20], GGI-FOLD [21], ERD [22] and our algorithm SIMARD [14] using evolutionary algorithms, incaRNAfbinv [23] using fragment-based design, and lastly antaRNA [24] which uses the ant colony optimization technique. Recently, some researchers has developed an online game to help with the RNA design problem [25]. They had summarized the design efforts of tens of thousands of human participants and three automated algorithms (RNAInverse, INFO-RNA and RNA-SSD) to generate some design principles.

B. Using MOO for RNA Design

Multi-objective optimization (MOO) is a paradigm which optimizes more than one objective at the same time [26]. When

the number of objectives increases or there are conflicting objectives, optimizing them gets more difficult.

The rationale behind using MOO criteria for RNA design comes from our previous findings as a result of SIMARD experiments. We found that there are two separate objectives to optimize in RNA Design problem: Hamming distance and free energy. However, when we used simulated annealing (SA) technique for optimization, we had to choose one of them to optimize. Also, our experiments showed that when we choose Hamming distance as an objective, free energy increases as we ignore it in the optimization step. Similarly, when we use energy as an optimization objective, Hamming distance increases. These results brought using MOO as a new criteria for RNA Design to optimize both Hamming distance and free energy to our mind.

The objectives of this paper are as follows:

- To explain the methodology behind SIMARD with and without Quality Pre-selection Strategy (QPS).
- To give experiment setup and data together with the importance of two criteria: Hamming distance and free energy.
- To show the correlations
- To observe the impact of using Multi-Objective Optimization (MOO).

III. METHOD

The main RNA design framework for our algorithm is SIMARD (Simulated Annealing RNA Design) [14]. SIMARD uses an optimization strategy to find the optimal solution. In the following sections, we will describe SIMARD framework with Quality Pre-selection Strategy (QPS). This setup was described in detail in our other paper [27]. The contribution of this paper will be to show modification of SIMARD to accommodate multi-objective modification, including optimization based on thermodynamic energies and Hamming distance.

A. SIMARD with QPS

SIMARD is a heuristic algorithm to design RNA secondary structure. It uses Simulated Annealing in order to find an optimal solution before the terminal condition is met. Algorithm 1 shows pseudo-code for SIMARD. After a starting solution is generated, a mutation is made to it and its fitness is evaluated based on given criteria. The solution is accepted or rejected based on its fitness and the current temperature of the algorithm. The temperature is reduced over time, moving the nature of the algorithm from exploration to exploitation.

Quality Pre-selection Strategy (QPS) is a mutation operator for SIMARD that generates a given number of sequences and returns the best based on given optimization criteria, discarding the dominated sequences. Algorithm 2 shows pseudo-code for QPS. QPS is flexible in two main areas: pool size and optimization criteria. Pool size refers to the number of sequences considered per step. The more sequences we generate the more expensive the mutation operator is, but the quality of the best solution is generally higher due to a larger sample size. Optimization criteria refers to the standard by

---

Algorithm 1 Secondary structure design with SIMARD

1: Sequence = InitialSequence;
2: Temperature = InitialTemperature;
3: Fitness = FitnessEvaluation(Sequence);
4: **while** (Temperature > FinalTemperature) **do**
5:     **for** ($i = 1$ to NumberOfIterations) **do**
6:         NewSequence = Mutate(Sequence);
7:         Fitness = FitnessEvaluation(NewSequence);
8:         Δ Fitness = NewFitness - Fitness;
9:         **if** (Δ Fitness $\geq$ 0) OR (with Probability[Accept] $= e^{\frac{-\Delta Distance}{Temperature}}$) **then**
10:            Fitness = NewFitness;
11:            Sequence = NewSequence;
12:        **end if**
13:    **end for**
14:    decrease Temperature;
15: **end while**
16: HammingDistance = HammingDistance(NewSequence, Structure);
17: FreeEnergy = FoldAndEvaluate(Sequence);

---

which each sequence is judged. The two main criteria that could be considered are Hamming distance to target structure and free energy.

Algorithm 2 High quality mutation selection with QPS

1: **function** QPS($Sequence$)
2:     BestSequence = Sequence;
3:     initialize BestQuality;
4:     **for** ($i = 1$ to $SequencesToGenerate$) **do**
5:         NewSequence = Modify(Sequence);
6:         NewQuality = Evaluate(NewSequence);
7:         **if** (NewQuality > BestQuality) **then**
8:             BestSequence = NewSequence;
9:             BestQuality = NewQuality;
10:        **end if**
11:        Reset Sequence;
12:    **end for**
13:    return BestSequence
14: **end function**

B. Optimization Combinations

When performing the optimization there are two criteria that we examine under the Multi-objective optimization (MOO) paradigm. These two criteria are: Hamming distance to target structure and free energy. There are a number of different combinations of optimization that we can test with our combination of pre-selection and SA acceptance.

1) **Hamming distance**: Hamming distance is a measure of the difference between two strings. Equation 1 specified the calculation of the hamming between two sequences A_i and B_i, where N is the total length of the strings.

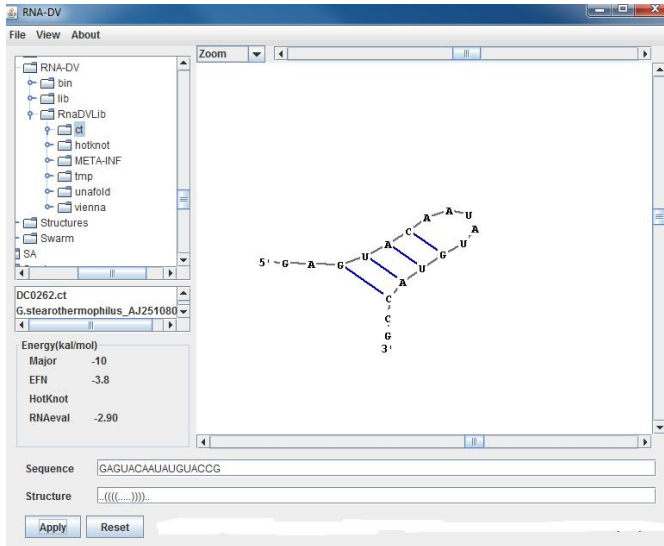$$Hamming\_Distance = \sum_{i=o}^{N-1} |A_i - B_i| \qquad (1)$$

Fig. 1. RNA secondary structure expressed visually with dot bracket notation shown above. Image was generated using *RNA-DV* software [28]

Figure 1 is showing the visualization of RNA secondary structure in both 2D structure and dot bracket format. The dot bracket format is a string composed of three symbols: dots, opening parentheses, and closing parentheses. This makes computing the Hamming distance of two structures very straight forward. Using this metrics, we can get a measure of how similar our designed structure is to the target structure.

In our case, we want to minimize Hamming distance to target structure, as a lower Hamming distance means that our working structure is closer to folding into the right shape. A Hamming distance to target structure of zero means that our working structure is identical to our target structure. This is one of our terminal conditions.

2) **Free energy**: Free energy is a measure of how stable a structure is. The lower the free energy, the more stable the structure. Most RNA secondary structure prediction and design algorithms are based on free energy minimization techniques, SIMARD can heuristically search for the structure with a free energy close to the minimum free energy ΔG for a strand of RNA, within given constraints.

SIMARD uses the thermodynamic energy model from Vienna RNA package [29]. For the purposes of this paper, the process of calculating the energy can be considered as a black box. The most important thing to realize is that we want a structure with as low free energy as possible. It is important to note that there are many ways any given RNA primary sequence can fold but we only consider the lowest possible free energy structure.

3) **Correlation and problems**: In our previous studies, we generally examined thermodynamic energy and Hamming distance separately as the optimization parameters [14] [27]. However, according to our trials, Hamming distance to target structure and free energy have a negative correlation. We calculated a correlation coefficient of -0.699 from 235,226

analyzed sequences produced by SIMARD. The sequences were all generated within the first 10,000 steps of the algorithm, where its behaviour is still fairly unbiased due to its exploratory nature. This negative correlation is easy to see when shown a plot of Hamming distance to target structure and free energy over the course of a SIMARD run. Figure 2 shows this. Without QPS, whichever objective is being optimized gradually decreases, while the other gradually increases.

In summary, because of this unexpected relationship between thermodynamic energy and Hamming distance, we propose to optimize based on both criteria. Table I shows the options of the different combination of the two criteria that we explored for this paper. While there are many other possibilities to explore, these are the ones we felt had the most promise, and would give us the best idea of the relationships between these values.

TABLE I
OPTIMIZATION COMBINATIONS BETWEEN THE TWO CRITERIA (HAMMING DISTANCE AND THERMODYNAMIC ENERGY).

SIMARD optimization	QPS optimization	QPS pool size
Hamming distance	Energy	2
Energy	Hamming distance	2
Hamming distance	Multi-optimization method (Table II)	2
Hamming distance	Energy	3
Energy	Hamming distance	3
Hamming distance	Multi-optimization method (Table II)	3

C. Experiments Setup

The primary goal of SIMARD is to produce a sequence with Hamming distance of zero to the target structure. However, without energy optimization from QPS, the system often returns a sequence with poor free energy. This is why our first experimental SIMARD configuration was to preselect an optimal energy sequence but accept it based on Hamming distance to target structure. Two variations of this were run: with QPS pool size 2, and with QPS pool size 3.

For our next experimental configuration, we preselected an optimal Hamming distance structure and accepted it based on energy. While we realize that it is not likely going to be a better solution, testing Hamming distance to target structure optimization in pre-selection (QPS) and SA optimization of energy can give us an idea of the relationship between the two objectives. We also ran two variations of this configuration: with QPS pool size 2, and with QPS pool size 3.

Finally, we present a novel multi-objective technique for pre-selection with QPS, described in Table II. In essence, we choose the best overall sequence - the one with the lowest relative Hamming distance to target structure and free energy. In a situation where sequence A has the lowest Hamming distance and the second lowest energy, and sequence B has the lowest energy and the second lowest Hamming distance, sequence A will be selected as Hamming distance is prioritized
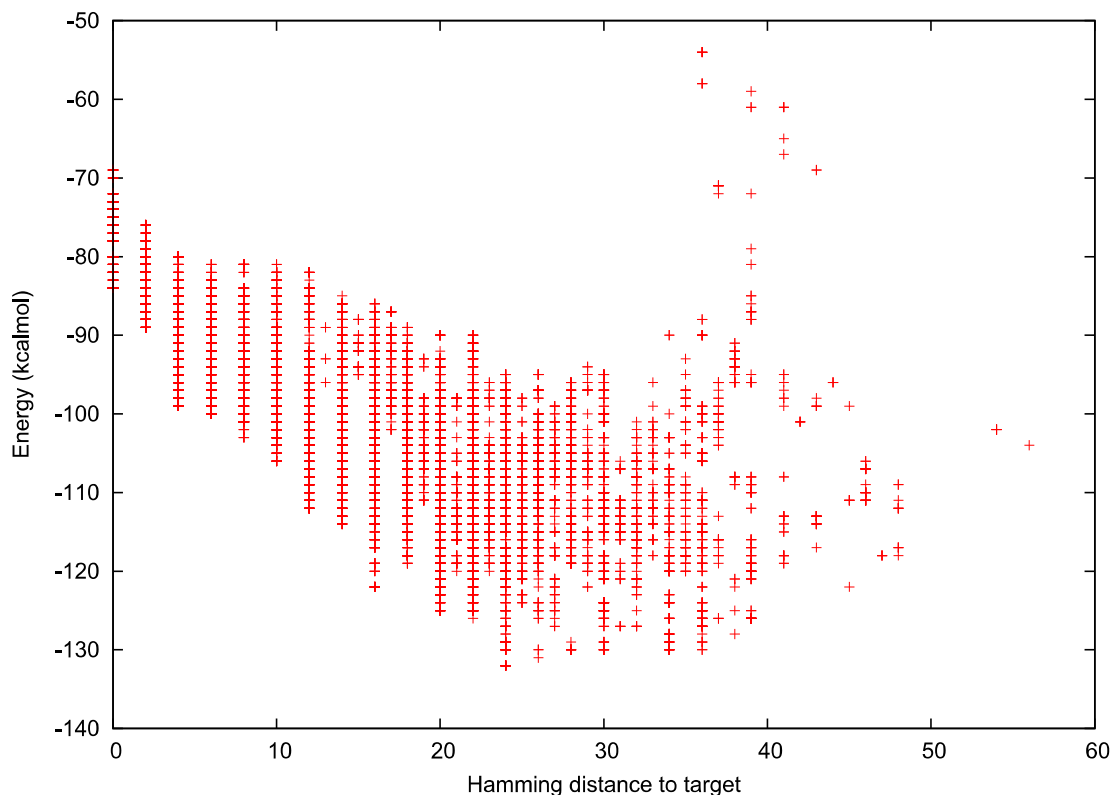
Fig. 2. Solutions generated in a run of SIMARD using energy as pre-selection criteria and Hamming distance to target as SA criteria. The horizontal axis represents Hamming distance and the vertical axis represents energy.

over energy. Once again, we also ran two variations of this configuration: with QPS pool size 2, and with QPS pool size 3.

TABLE II
MINIMIZING MULTIPLE OBJECTIVES

Rank	Hamming distance	Free energy	Ranking
1	A	C	A (1+2=3)
2	B	A	C (3+1=4)
3	C	B	B (2+3=5)

D. Data

We ran the experiments on sequences 1-30 from the Rfam dataset, excluding sequence RF00023 [30]. This is because RF00023 has lots of pseudo-knotted base pairs [20]. The Rfam dataset contains sequences with lengths varying between $54 - 451$ nt inclusive.

IV. RESULTS AND DISCUSSION

Table III summarized the three sets of results.

TABLE III
THREE SETS OF EXPERIMENTS RESULTS.

Result	Pre-selection	SA optimization
1	Energy	Hamming distance
2	Hamming distance	Energy
3	Multi-objective strategy outlined in Table II	Hamming distance

A. Energy optimized in pre-selection, Hamming distance optimized in SA

In addition to having good Hamming distance, solutions produced by this technique have large energy improvements over Vanilla SIMARD. A pool size of 3 sees a greater improvement of energy than a pool size of 2.

However, as the algorithm runs, the higher the pool size, the more the energy increases near the end of the run (relative to where it was before). This is surprising behavior that may be caused by the strict pre-selection and exploratory SA optimization early on in the algorithm, forcing the search
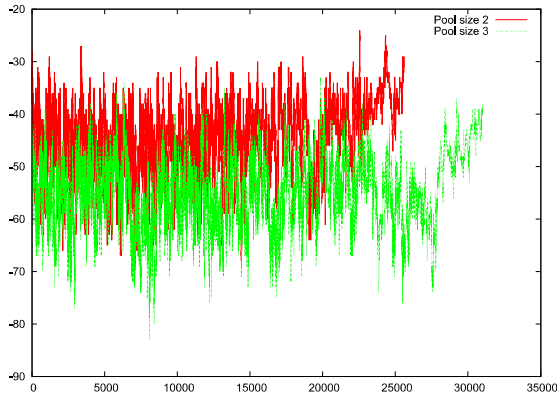
Fig. 3. Energy over two runs of the same sequence, RF00020, the 20th sequence in the RF dataset

space into a poor Hamming distance area, resolved by the exploitative nature of the late stage of SA optimization. If this is a case, a high pool size past a certain point is redundant as it just drives the algorithm from the optimal space at early stages, only to be overpowered by the strict SA optimization at late stages of the algorithm.

A segmented run may be the answer to this, as even though the run is driven back into a higher energy state, it is not driven all the way back up to its initial state. With this in mind, we could preselect energy for the early stages of the run, and then stop pre-selecting when the run got to a certain point.

It should be noted that the rise in energy is accompanied by a sharp fall in Hamming distance. This shows the negative correlation between the objectives in the case of this problem. Figure 3 shows the differences in early algorithm energy and the rise at the end of the algorithm.

Note that the sudden rise in energy is accompanied by a drop in Hamming distance. Also, the run with pool size 3 takes more steps. This is likely due to the negative correlation between Hamming distance to target structure and free energy, as the more greedy pre-selection tends to return sub-optimal Hamming distance solutions, making a longer road to one with Hamming distance of zero.

B. Hamming distance optimized in pre-selection, energy optimized in SA

While it is clear that this technique is not optimal for designing sequences, due to the larger final Hamming distance to target structure, it can be helpful in determining the connection between Hamming distance and energy. Our tests confirmed our hypothesis: that energy and Hamming distance don't have a very meaningful connection. There are a theoretically infinite amount of low energy structures, but they will not necessarily be the same structure as the target.

On the other hand, when paired with Hamming distance optimization, as is the case in the early phases of this algorithm

(when SA is in an exploratory state and accepts almost every sequence), Hamming distance is optimized, the algorithm is guided towards the correct structure's low energy state. However, as the algorithm gets more and more greedy for low energy, it abandons the target structure and moves towards the closest free energy structure to its current state. Firure 5 shows this phenomenon around step 1,500. It is unlikely that an energy focused approach will ever be the optimal solution to the RNA design problem. Energy guided Hamming distance optimization, on the other hand, shows great promise. It is further explored in the following section.

C. Multi Objective pre-selection, Hamming distance optimized in SA

This technique found sequences with Hamming distance to target structure of zero surprisingly fast. Table V shows a comparison between the number of steps taken to terminate between the pool size 2 and 3 runs of this technique and the pool size 2 and 3 runs of the energy pre-selection technique(see Section IV-A). Fewer steps are required likely due to the fact that Hamming distance was optimized in pre-selection as well as in SA acceptance for this method, as opposed to only in SA acceptance for the other.

Table IV shows that free energy of solutions was also very low. A likely reason for energy being lower when the pool size was bigger could be that a bigger pool size skewed the overall optimization bias towards energy (since the pool of multi-objective optimization was growing but the SA Hamming optimization was not becoming more strict).

Because of how quickly this run terminates, we could try letting it run longer and increasing the weight of energy in the pre-selection phase. This would allow us to run it for however much time is available to us, continuing to optimize the solution further after a viable solution is found.

V. CONCLUSION

In this paper, we explored different SIMARD configurations using QPS. We looked at energy optimized in pre-selection and Hamming distance optimized with SA, as well as the opposite: Hamming distance optimized in pre-selection and energy optimized with SA.

From these two methods we learned that Hamming distance is a far better primary objective than energy as there are many low energy states that are not the same as a given target structure. We also confirmed that there is a negative correlation between those objectives in terms of this problem. Finally, we looked at a new multi-objective pre-selection technique, which proved to find a solution much faster than any other explored technique. It also outperformed vanilla SIMARD (no QPS) in terms of free energy.

As we have discussed in previous papers, a direction that we could go in would be to try a new secondary prediction structure (instead of Vienna Package). We can also look at trying to do a segmented run, or change the weight of the objectives in our multi-objective pre-selection.
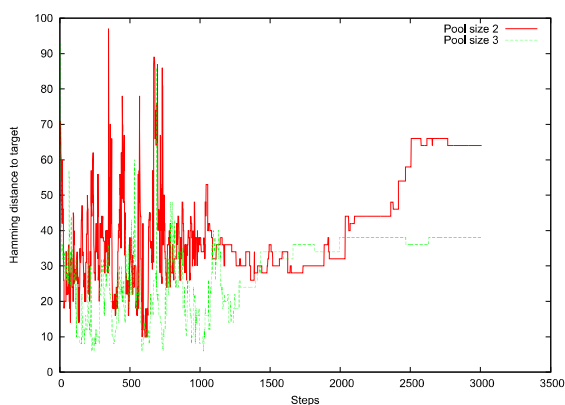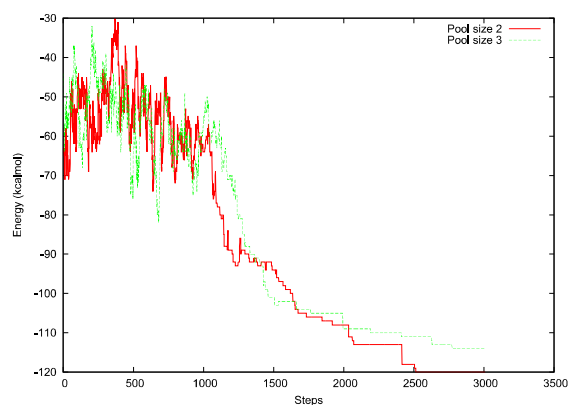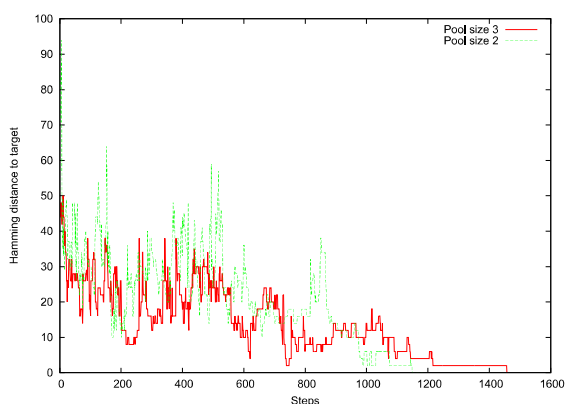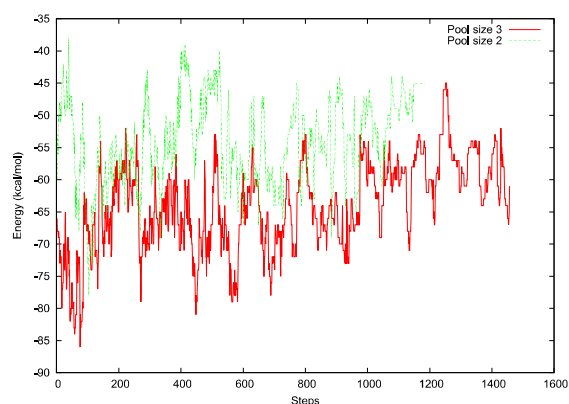
(a)   (b)

Fig. 4. SIMARD run with Hamming distance pre-selection and Energy SA optimization of sequence from RF dataset RF00025. The horizontal axis represents algorithmic steps and the vertical axis represents (a) Hamming distance to target structure (b) Free energy

(a)   (b)

Fig. 5. SIMARD run with multi-objective pre-selection of sequence from RF dataset RF00025. The horizontal access represents algorithmic steps and the vertical axis represents (a) Hamming distance to target structure (b) Free energy

ACKNOWLEDGMENT

REFERENCES

[1] M. B. Schnall-Levin, "RNA : algorithms, evolution and design," Ph.D. dissertation, Massachusetts Institute of Technology, 2011.

[2] S. Washietl, I. L. Hofacker, and P. F. Stadler, "From The Cover: Fast and reliable prediction of noncoding RNAs," *Proceedings of the National Academy of Sciences*, vol. 102, no. 7, pp. 2454–2459, 2005.

[3] M. Zuker, "Prediction of RNA secondary structure by energy minimization," *Methods in Molecular Biology*, pp. 267–294, 1994.

[4] A. O. Harmanci, G. Sharma, and D. H. Mathews, "Efficient pairwise rna structure prediction using probabilistic alignment constraints in dynalign," *BMC Bioinformatics*, vol. 8:130, 2007.

[5] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, vol. 31, pp. 3406–3415, 2003.

[6] F. H. D. V. Batenburg, A. P. Gultyaev, and C. W. A. Pleij, "An APL-programmed genetic algorithm for the prediction of RNA secondary structure," *Journal of Theoretical Biology*, vol. 174, no. 3, pp. 269–280, 1995.

[7] I. Tinoco, O. C. Uhlenbeck, and M. D. Levine, "Estimation of secondary structure in ribonucleic acids," *Nature*, vol. 230, pp. 362–267, 1971.

[8] D. H. Mathews, T. C. Andre, J. Kim, D. H. Turner, and M. Zuker, "An updated recursive algorithm for rna secondary structure prediction with improved thermodynamic parameters," *ACS Symposium Series*, vol. 682, pp. 246–257, 1997.

[9] S. Bellaousov, J. S. Reuter, M. G. Seetin, and D. H. Mathews, "Rnastructure: Web servers for rna secondary structure prediction and analysis,"

Nucleic Acids Res., vol. 41, 2013.

[10] H. H. Tsang and K. C. Wiese, "SARNA-Predict: Accuracy improvement of RNA secondary structure prediction using permutation based simulated annealing," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 99, no. 1, pp. 727–740, 2010.

[11] P. Grypma, J. Babbitt, and H. H. Tsang, "A study on the effect of different thermodynamic models for predicting pseudoknotted RNA secondary structures," in *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2013, pp. 52–59.

[12] P. Grypma and H. H. Tsang, "SARNA-Predict: Using adaptive annealing schedule and inversion mutation operator for RNA secondary structure prediction," in *IEEE Symposium Series on Computational Intelligence*, 2014, pp. 150–156.

[13] H. H. Tsang and K. C. Wiese, "A Permutation Based Simulated Annealing Algorithm to Predict Pseudoknotted RNA Secondary Structures," *Int. J. Bioinformatics Res. Appl.*, vol. 11, no. 5, pp. 375–396, Sep. 2015.

[14] H. E. Erhan, S. Sav, S. Kalashnikov, and H. H. Tsang, "Examining the Annealing Schedules for RNA Design Algorithm," in *Proceedings of the IEEE Congress on Evolutionary Computation*, 2016, pp. 1295–1302.

[15] M. Andronescu, A. P. Fejes, F. Hutter, H. H. Hoos, and A. Condon, "A new algorithm for rna secondary structure design." *Journal of Molecular Biology*, vol. 336, no. 3, pp. 607–624, February 2004.

[16] A. Busch and R. Backofen, "INFO-RNAa server for fast inverse RNA folding satisfying sequence constraints," *Nucleic Acids Research*, vol. 35, pp. 310–313, 2007.

[17] A. Levin, M. Lis, Y. Ponty, C. W. O'Donnel, S. Devadas, B. Berger, and J. Waldispuhl, "A global sampling approach to designing and reengineering rna secondary structures," *Nucleic Acids Research*, vol. 40, 2012.

[18] V. Reinharz, Y. Ponty, and J. Waldispuhl, "A weighted sampling algorithm for the design of RNA sequences with targeted secondary structure and nucleotide distribution," *ICMB/ECCB*, vol. 29, pp. i308–i315, 2013.

[19] A. Taneda, "Multi-objective genetic algorithm for pseudoknotted RNA sequence design," *Frontiers in Genetics*, vol. 3, pp. 1–9, 2012.

[20] R. B. Lyngs, J. W. Anderson, E. Sizikova, A. Badugu, T. Hyland, and J. Hein, "Frnakenstein: multiple target inverse RNA folding," *Frontiers in Genetics*, vol. 3, pp. 1–12, 2012. [Online]. Available: http://www.biomedcentral.com/1471-2105/13/260

[21] M. Ganjtabesh, F. Zare-Mirakabad, and A. Nowzari-Dalini, "Inverse RNA Folding Solution Based on Multi-Objective Genetic Algorithm and Gibbs Sampling Method," *EXCLI*, vol. 2013, pp. 546–555, 2013.

[22] A. Esmaili-Taheri, M. Ganjtabesh, and M. Mohammad-Noori, "Evolutionary Solution for the RNA Design Problem," *Bioinformatics*, vol. 30, no. 9, pp. 1250–1258, 2014.

[23] M. Retwitzer, V. Reinharz, Y. Ponty, J. Waldispuhl, and D. Barash, "incaRNAfbinv: a web server for the fragment-based design of RNA sequences," *Nucleic Acids Research*, vol. 44, pp. 304–314, May 2016.

[24] K. Robert, M. Martin, and B. Rolf, "antaRNA - Ant Colony Based RNA Sequence Design," *Bioinformatics Advance Access*, vol. 31, pp. 3114–3121, May 2015.

[25] J. Anderson-Lee, E. Fisker, V. Kosaraju, M. Wu, J. Kong, J. Lee, M. Lee, M. Zada, A. Treuille, and R. Das, "Principles for predicting {RNA} secondary structure design difficulty," *Journal of Molecular Biology*, vol. 428, no. 5, Part A, pp. 748 – 757, 2016.

[26] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001.

[27] S. Sav, D. J. Hampson, and H. H. Tsang, "SIMARD: A Simulated Annealing Based RNA Design Algorithm with Quality Pre-Selection Strategies," in *IEEE Symposium Series on Computational Intelligence*, 2016, p. submitted.

[28] H. H. Tsang and D. C. Dai, "RNA-DV: An interactive tool for editing and visualizing RNA secondary structures," in *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 2012, pp. 601–603.

[29] I. L. Hofacker, W. Fontana, F. S. Peter, L. S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast Folding and Comparison of RNA Secondary Structures," *Monatshefte fur Chemie*, vol. 125, pp. 167–188, 1994. [Online]. Available: http://fontana.med.harvard.edu/www/Documents/WF/Papers/vienna.rna.pdf

[30] S. W. Burge, J. Daub, R. Eberhardt, J. Tate, L. Barquist, E. P. Nawrocki, S. R. Eddy, P. P. Gardner, and A. Bateman, "Rfam 11.0: 10 years of RNA families," *Nucleic Acids Research*, vol. 41, no. D1, pp. D226–D232, 2013.

TABLE IV

THE THERMODYNAMIC FREE ENERGY Δ G RESULT OF THE FINAL DESIGNED STRUCTURE. THE SEQUENCES ARE ARRANGED IN ORDER OF INCREASING LENGTH. BEST RESULTS ARE DENOTED IN BOLD.

Sequence name	Length	Energy (kcal/mol)		
		Multi objective pool size 2	pre-selection pool size 3	Vanilla SIMARD
RF0008	54	-5	-10	**-26**
RF00029	73	-11	**-16**	-8
RF0005	74	-12	**-14**	-13
RF00027	79	-24	**-29**	-25
RF00019	83	-16	**-18**	-14
RF00014	87	-23	**-27**	-15
RF0006	89	-10	-15	**-63**
RF00026	102	-2	-2	**-1**
RF0001	117	-18	**-23**	-8
RF00021	118	-27	**-38**	-20
RF00020	119	-22	**-24**	-15
RF00016	129	-13	-15	**-31**
RF00015	140	-18	**-20**	-18
RF00022	148	-21	**-30**	-18
RF0002	151	-14	-13	**-31**
RF0007	154	-31	**-36**	-23
RF0003	161	-28	**-34**	-25
RF00013	185	-31	**-47**	-15
RF0004	193	-33	**-37**	-2
RF00025	210	-22	**-30**	-23
RF00012	215	-26	-36	**-67**
RF00017	301	-70	**-99**	-65
RF00030	340	-41	**-49**	-45
RF00028	344	-32	**-41**	-29
RF0009	348	-34	**-39**	-19
RF00010	357	-65	**-79**	-14
RF00018	360	-38	**-49**	-36
RF00011	382	-68	**-75**	-28
RF00024	451	-76	-76	**-55**

TABLE V
THE NUMBER OF STEPS, OR ATTEMPTED SOLUTIONS, BEFORE THE TERMINARY CONDITION IS MET. RESULTS SHOWN ARE THE AVERAGE OF TWO RUNS PER SEQUENCE. THE SEQUENCES ARE ARRANGED IN ORDER OF INCREASING LENGTH. BEST RESULTS ARE DENOTED IN BOLD.

| Sequence name | Length | Algorithm Steps | | | |
| | | Multi objective pre-selection | | Energy pre-selection | |
		pool size 2	pool size 3	pool size 2	pool size 3
RF0008	54	**92**	**92**	14419	30439
RF00029	73	**274**	319	26033	19625
RF0005	74	**92**	**92**	802	802
RF00027	79	**92**	**92**	17222	10414
RF00019	83	**92**	319	25279	19924
RF00014	87	**92**	137	70509	71296
RF0006	89	**137**	183	27235	26834
RF00026	102	92	**46**	802	802
RF0001	117	**1138**	1366	28762	55865
RF00021	118	**92**	**92**	30439	28837
RF00020	119	2185	**2048**	29534	23624
RF00016	129	2139	**2048**	8411	8411
RF00015	140	**365**	547	53343	31640
RF00022	148	**183**	**183**	1202	802
RF0002	151	**1730**	2231	33242	32441
RF0007	154	729	**638**	19807	15739
RF0003	161	**2230**	2412	37247	35169
RF00013	185	**410**	729	38878	32857
RF0004	193	**137**	228	802	802
RF00025	210	**1366**	1411	13217	20026
RF00012	215	**319**	410	36465	30350
RF00017	301	**1320**	1502	22429	18824
RF00030	340	**1821**	2003	48862	30439
RF00028	344	2367	**2139**	61678	51038
RF0009	348	1821	**1639**	8411	4406
RF00010	357	**2641**	3006	65311	65689
RF00018	360	**2823**	2868	25232	22028
RF00011	382	**2549**	2641	17222	15620
RF00024	451	**2321**	2958	36847	30439