International Journal of Biology Research
ISSN: 2455-6548
Impact Factor: RJIF 5.22
www.biologyjournal.in
Volume 3; Issue 1; January 2018; Page No. 185-187

MFold: Machine learning approach for predicting secondary structure of the trnP tRNA

Sonu Mishra, Virendra S Gomase
Department of Biotechnology, Mewar University, Chittorgarh, Rajasthan, India

Abstract
The protein secondary structure prediction is one of the first and most important issues for almost a quarter of century to construct the effective tool with the highest accuracy. In this present study we predicted the secondary structure of the trnP tRNA through machine learning approach that is m-fold (through mfold server). In this investigation, we also predicted the thermodynamics of the structure with secondary structure prediction. This gene structural and functional prediction will play a major role in drug designing or synthetic vaccine development or in the disease better understanding.

Keywords: Mfold, *D.medinensis*, secondary structure, RNA, tRNA, trnP

1. Introduction
RNA secondary prediction methods rely on free energy minimization using nearest-neighbor parameters for predicting the stability of an RNA secondary structure, in terms of Gibbs free energy at 36^0 C [1, 2, 3, 4].The stability prediction rule is the nearest-neighbor parameters. The stability of each base pair depends only on the most adjacent pairs and the sum of each contribution is the total free energy. In the total conformational free energy, the major contributions are the loop initiation, and unpaired nucleotide stacking. It is also found that the favorable free energy increments are less than zero. The base pair free energy increments are counted as stack of adjacent pairs. For example: The consecutive CG base pairs are worth -3.3kcal/mol [4]. It is to be noted that the loop regions have unfavorable increments called loop initiation energies that largely reflect an entropic cost for constraining the nucleotides in the loop. For an example, the hairpin loop of four nucleotides has an initiation of 5.6 kcal/mol [1]. In the stacked nucleotides or as mismatched pairs, the unpaired nucleotides in loops can provide the favorable energy increment. The 3'- most G, called a dangling end, stacks on the terminal base pair and provides -1.3 kcal/mol of stability. The Gibbs free enrgy of formation for an RNA structure (ΔG^o) quantifies the equilibrium stability of that structure at a specific temperature. The RNA secondary structure accuracy can be assessed by predicting structures for RNA sequences with known secondary structures, as determined by comparative sequence analysis. The study suggest that the prediction accuracy can be improved by constraining secondary structure prediction with enzymatic constraints.

1.1 RNA Seconadary Structure Prediction VIA MFold
Mfold is an RNA secondary structure prediction package available through a Web fronted and as code for compilation on Unix and Linux machines[1,5].It uses the current set of nearest neighbor parameters for free energies at 37^0 C [1]. Minimum free energy and suboptimal secondary structures, sampled heuristically [6], are predicted. Predicted suboptimal structures represent alternative structures to the lowest free energy structure and reflect both the possibility that an RNA sequence may have more than a single structure [7] and the fact that the energy rules contain some uncertainty [1].M fold also predicts the energy dot plots, which display the lowest free energy conformation possible for each possible base pair [8]. These plots conveniently demonstrate all possible base pairs within a user –specified increment of the lowest free energy structure, and predicted structures can be color annotated to demonstrate regions in the structure for which many folding alternatives exits [10]. In this investigation we have analyzed the trnP tRNA gene from the *D.medinensis* .

2. Methgodology
A sequence provided to the mfold server (unafold.rna.albany.edu/?q=mfold) and the RNA Folding Form (version 2.3 energies) [5, 9, 10] is selected for the study. In this server the sequence name is provided and used the FASTA format of sequence for the analysis and the other parameter are kept as defaulter and run the RNA fold. The obtained output is recorded and interpreted for the provided sequence is trnP(NCBI Reference Sequence: NC_016019.1). The DNA mfold server predict the secondary structure of the trnP tRNA DNA sequence through DNA folding free energies [7] on the folding temperature is fixed at 37 degree The percentage sub-optimality number is, 5 by default, is the maximum percent difference in free energy from the lowest free energy structure that is allowed when generating suboptimal secondary structure to be predicted .The upper bound on the computed folding (default =50) is the maximum number of suboptimal secondary structure to be predicted the tRNA used here is 53 bp with linear conformation of DNA (LOCUS: NC_016019) of Organism: Dracunculus medinensis [Figure1].

Dracunculus medinensis mitochondrion, complete genome
NCBI Reference Sequence: NC_016019.1
>gi|347600379:14522-14574 Dracunculus medinensis mitochondrion, complete genome
CGATCTTGAGTTTTTTAGAATATTGAGTTTGGGTCTT
AAAGGTTTTTGATCGA

Fig 1: Dracunculus medinensis mitochondrion, complete genome sequence of the trnP

3. Result and the Interpretation

The trnp gene sequence were analyzed through mfold DNA server which is 53 nucleotide long and have a default window of 2.A smaller window allows size allowed is zero .the maximum number of unpaired nucleotide in bulge or internal loops is limited to 30, by default.The maximum asymmetry in internal loops (the difference in length is unpaired nucleotide on ech strand) is also 30 by default.The maximum distance allowed between the paired nucleotides defaults to no limit. These values can be modified, as appropriate. The mfold server output form for the secondary structure prediction of the trnPtRNA. Folding trnP at 37° C. (3.5) [Na+] = 1.0 M, [Mg++] = 0.0 M, oligo correction - Computed for 163.53.215.84(Figure 2). The colour annotation on the energy dot plot of the trnp tRNA (Energy dot plot for trnP with magnification 1)- The computed folding contains 11 base pairs out of 18 (61.1%) in the energy dot plot (Figure 3). Structure 1 Folding bases 1 to 53 of trnP(Figure 4). The secondary structure predicted for the tRNA, trnP(mfold Web server , version 3.1)(Figure 5). Table 1 shows the thermodynamics of trnP of tRNA.

```
Linear DNA folding at 5%, window = 2, max folds = 50
12 A's, 4 C's, 13 G's, 24 U/T's and 0 N's.
         10         20         30         40         50
 CGATCTTGAG TTTTTTAGAA TATTGAGTTT GGGTCTTAAA GGTTTTTGAT
         60
 CGA
```

Fig 2: The mfold server output form for the secondary structure prediction of the trnPtRNA . Folding trnP at 37° C. (3.5) [Na+] = 1.0 M, [Mg++] = 0.0 M, oligo correction - Computed for 163.53.215.84

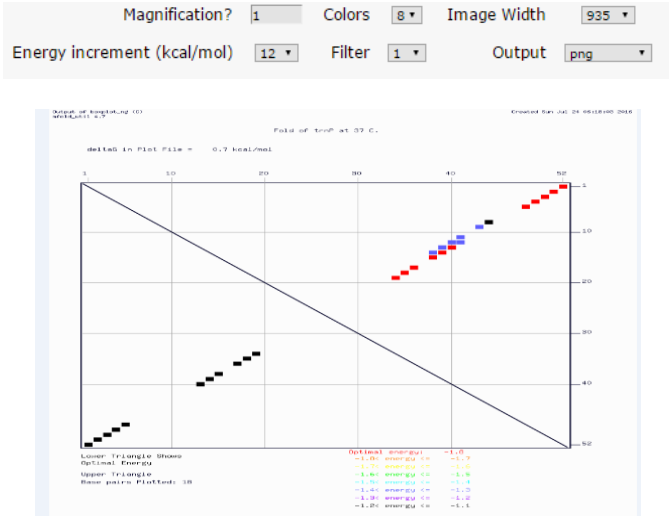

Fig 3: The colour annotation on the energy dot plot of the trnp tRNA(Energy dot plot for trnP with magnification 1)- The computed folding contains 11 base pairs out of 18 (61.1%) in the energy dot plot.

```
Structure 1 Folding bases 1 to 53 of trnP
dG = -1.89  dH = -78.20  dS = -246.04  Tm = 44.7 ℃
      10        20
-|  TTGAGTT   T   ATATTG
 CGATC      TTT AGA     A
 GCTAG      AAA TCT     G
A^   TTTTTGG   T   GGGTTT
 50        40        30
```

Fig 4: Structure 1 Folding bases 1 to 53 of trnP

Structure 5 : ΔG = -1.89 kcal/mol, (Thermodynamic Details).

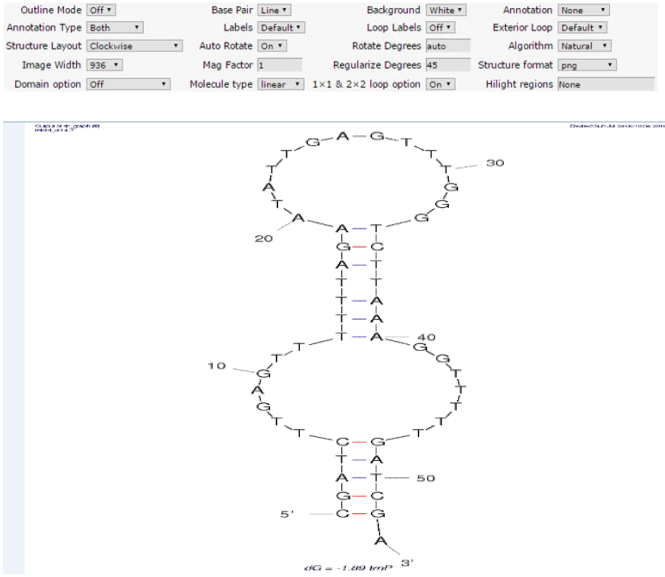

Fig 5: The secondary structure predicted for the tRNA ,trnP(mfold Web server , version 3.1)

Thermodynamics of Folding: $\Delta G = \Delta H - T\Delta S$

- ΔG = -1.89 kcal/mol at 37 °C
- ΔH = -78.20 kcal/mol
- ΔS = -246 cal/(K·mol)
- T_m = 44.6 °C assuming a 2 state model.
- linear DNA folding.
- Ionic conditions: $[Na^+]$ = 1.0 M, $[Mg^{++}]$ = 0.0 M.
- Standard errors are roughly ±5%, ±10%, ±11% and 2-4 °C for free energy, enthalpy, entropy and T_m, respectively.
  Structure5trnP ΔG = -1.89

Table 1: Thermodynamics of trnP of tRNA

Structural element	δG	Information
External loop	-0.92	1 ss bases & 1 closing helices.
Stack	-2.17	External closing pair is C^{1}-G^{52}
Stack	-1.30	External closing pair is G^{2}-C^{51}
Stack	-0.88	External closing pair is A^{3}-T^{50}
Stack	-1.30	External closing pair is T^{4}-A^{49}
Helix	-5.65	5 base pairs.
Interior loop	3.80	External closing pair is C^{5}-G^{48}
Stack	-1.00	External closing pair is T^{13}-A^{40}
Stack	-1.00	External closing pair is T^{14}-A^{39}

Helix	-2.00	3 base pairs.
Interior loop	1.36	External closing pair is T [15]-A [38]
Stack	-1.28	External closing pair is A [17]-T [36]
Stack	-1.30	External closing pair is G [18]-C [35]
Helix	-2.58	3 base pairs.
Hairpin loop	4.10	Closing pair is A [19]-T [34]

The base pairs struture of the trnP can be viewed in a circle also . It is a first step to represents a formal structure in which overlap of bases is usually small. Circle Graphs generated from this page correspond to structures drawn with Auto Rotate set to off, and Format set to Default or simple, and no Rotation, and Exterior Loop set to Default. G-C arcs are drawn in red, A-U, A-T arcs are drawn in blue, G-U, G-T arcs are drawn in green, other arcs, if present, are in yellow. Other options are to draw the arcs in all red, or to use p-num or ss-count values to find an average value for each base pair and color the arc accordingly. The bases are drawn clockwise around the circle and lead to a structure in which bases are drawn clockwise around each loop. This is a common way of producing a structure, but a counter-clockwise circle graph and resulting structure can be produced by placing the bases around the circle in a counter-clockwise fashion(Figure 6).

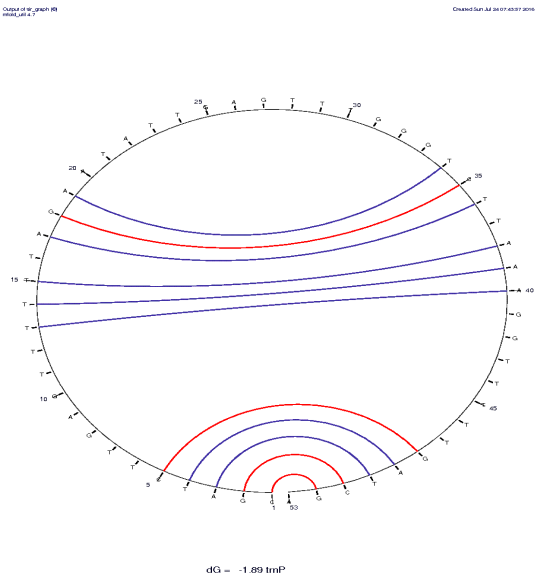

Fig 6: A circle graph is one way to display base pairs of a structure. It also represents a first step to producing a formal structure in which overlap of bases is usually small.

4. Conclusion

This method provides the prediction of trnP -tRNA gene and the secondary structure. Users can use either a complete chromosome or sequence fragments to predict the locations of tRNA gene and tRNA secondary structure. According to our analysis, this method provides an opportunity for biologists and researchers to locate tRNA gene with more efficiency. The structure prediction may provide important markers for refining the phylogenetic relations and improve the disease and molecular level understanding of gene and its evolutions.

Conflicts of Interest

The authors declare no conflict of interest.

5. References

1. Mathews, DH, Sabina J, Zuker M, Turner, DH. Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNAsecondary structure. J. Mol. Biol. 1999b; 317:191-203.
2. Turner DH. Conformational changes .In Nucleic Acids (Bloomfield, V., Crothers, D and Tinoco,I., eds), (University Science Books, Sausalito,CA). 2000, P. 259-334.
3. Xia T. Mathews, D.H., and Turner,D.H.(199). Thermodynamics of RNA secondary structure formation.In Prebiotic chemistry, Molecular Fossils, Nucleosides, and RNA (Soil, D.G., Nishimura, S., and Moore, P.B., ends.), (Elsevier, New York), 21-47.
4. Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry. 1998; 20;37(42):14719-35.
5. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003; 1;31(13):3406-15.
6. Zuker M. On finding all suboptimal foldings of an RNA molecule. Science. 1989; 7;244(4900):48-52.
7. Schultes EA, Bartel DP. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science. 2000; 21;289(5478):448-52.
8. Zuker M, Jacobson AB. "Well- determined" regions in RNA secondary structure predictions:applications to small and large subunit rRNA . Nucl. Acids Res. 1995; 23:2791-2798.
9. Waugh A, Gendron P, Altman R, Brown JW, Case D, Gautheret D, *et al.* RNAML: A standard syntax for exchanging RNA information. RNA. 2002; 8(6):707-717.
10. Zuker M, Jacobson AB. Using Reliability Information to Annotate RNA Secondary Structures. RNA. 1998; 4:669-679.