

# Supplementary Information for “Determining Parameters for Non-Linear Models of Multi-Loop Free Energy Change”

Max Ward, Hongying Sun, Amitava Datta, Michael Wise, David H. Mathews

---

**Supplementary Algorithm 1** A description of the Iterative Brute Force algorithm including structure seeding and separate RNA families

---

```
 $\theta \leftarrow$  an arbitrary initial parameter vector  
 $S \leftarrow$  a set of sequence-structure pairs to optimize on  
 $S' \leftarrow \{\}$   
for all  $\theta' \in \{\theta_1, \theta_2, \dots, \theta_5\}$  do  $S' \leftarrow S' \cup \{fold(p, \theta') : p \in S\}$   
end for  
loop  
     $S' \leftarrow S' \cup \{fold(p, \theta) : p \in S\}$   
     $next\theta \leftarrow \theta$   
     $bestscore \leftarrow -\infty$   
    for all  $\theta' \in$  candidate parameter vectors do  
         $score \leftarrow 0$   
        for all  $family \in$  RNA families in  $S$  do  
             $famscore \leftarrow 0$   
            for all  $p \in family$  do  
                 $predictedp \leftarrow p$   
                 $predictede \leftarrow \Delta G(p, \theta')$   
                for all  $p' \in S'$  s.t.  $p'$  corresponds to  $p$  do  
                     $e \leftarrow \Delta G(p', \theta')$   
                    if  $e \leq predictede$  then  
                         $predictede \leftarrow e$   
                         $predictedp \leftarrow p'$   
                    end if  
                end for  
                 $predictedsc \leftarrow fscore(p, predictedp)$   
                 $famscore \leftarrow famscore + predictedsc$   
            end for  
             $score \leftarrow score + \frac{famscore}{|family|}$   
        end for  
         $score \leftarrow \frac{score}{\text{the number of RNA families } \in S}$   
        if  $score > bestscore$  then  
             $bestscore \leftarrow score$   
             $next\theta \leftarrow \theta'$   
        end if  
    end for  
    if  $next\theta = \theta$  then return  $\theta$   
    end if  
     $\theta \leftarrow next\theta$   
end loop
```

---

Supplementary Table 1: A summary of the space of parameters optimized over using the IBF algorithm for the linear model. All units are in *kcal/mol*. This summary includes the lower and upper limits on each parameter, as well as the step size used. The parameters used are *lower, lower + step, lower + 2step, ..., upper*.

	Lower	Upper	Step
Initiation ( <i>A</i> )	3	20	0.1
Branch Cost ( <i>B</i> )	-6	3	0.1
Unpaired Cost ( <i>C</i> )	-6	3	0.1

Supplementary Table 2: A summary of the space of parameters optimized over using the IBF algorithm for the logarithmic model. All units are in *kcal/mol* except for the pivot to using a logarithmic penalty for unpaired nucleotides. This summary includes the lower and upper limits on each parameter, as well as the step size used. The parameters used are *lower, lower + step, lower + 2step, ..., upper*.

	Lower	Upper	Step
Initiation ( <i>A</i> )	3	20	0.1
Branch Cost ( <i>B</i> )	-3	3	0.1
Unpaired Cost ( <i>C</i> )	-3	3	0.1
Logarithm Coefficient ( <i>D</i> )	-3	3	0.1
Pivot to Logarithmic ( <i>E</i> )	1 <i>nt</i>	8 <i>nt</i>	1 <i>nt</i>

Supplementary Table 3: A summary of the space of parameters optimized over using the IBF algorithm for the Aalberts & Nandagopal model. This summary includes the lower and upper limits on each parameter, as well as the step size used. The parameters used are *lower, lower + step, lower + 2step, ..., upper*.

	Lower	Upper	Step
<i>a</i> -length	1.2 Å	8.7 Å	0.05 Å
<i>b</i> -length	10 Å	17.5 Å	0.05 Å
Offset ( <i>C</i> )	-5 <i>kcal/mol</i>	5 <i>kcal/mol</i>	0.1 <i>kcal/mol</i>

Supplementary Table 4: Paired *t*-test against the existing linear model on the complete validation data set. Contains the *p*-value of a paired *t*-test of each set of model parameters against the results of the existing linear model. Results using parameters determined using linear regression (LR), iterative brute force (IBF), and the existing parameters are included. RNA families used for training are denoted by parentheses.

	(16S rRNAs)	(23S rRNAs)	(5S rRNAs)	RNase P RNAs	Grp. I In- trons	Grp. II In- trons	(SRP RNAs)	(tRNAs)	(Telomerase tmRNAs RNAs)	
IBF Linear	0.802	0.617	0.000	0.000	0.263	0.109	0.000	0.000	0.012	0.122
LR Linear	0.122	0.042	0.000	0.021	0.310	0.060	0.830	0.000	0.618	0.000
Existing Logarithmic	0.528	0.952	0.000	0.000	0.552	0.115	0.000	0.000	0.868	0.000
IBF Logarithmic	0.606	0.430	0.190	0.000	0.072	0.889	0.420	0.000	0.682	0.000
LR Logarithmic	0.003	0.001	0.337	0.000	0.091	0.601	0.733	0.271	0.665	0.000
Number of RNAs	88	30	1283	454	98	11	928	6430	37	462

Supplementary Table 5: Paired *t*-test against the existing linear model on the small validation data set. Contains the *p*-value of a paired *t*-test of each set of model parameters against the results of the existing linear model. Results using parameters determined using linear regression (LR), iterative brute force (IBF), and the existing parameters are included. RNA families used for training are denoted by parentheses.

	16S rRNAs	23S rRNAs	(5S rRNAs)	RNase P RNAs	Grp. I Introns	(SRP RNAs)	(tRNAs)	tmRNAs
Existing Aalberts & Nandagopal	0.485	0.191	0.935	0.003	0.433	0.000	0.000	0.271
IBF Aalberts & Nandagopal	0.388	0.198	0.028	0.009	0.212	0.726	0.003	0.793
LR Aalberts & Nandagopal	0.685	0.105	0.574	0.000	0.031	0.000	0.000	0.085
Number of RNAs	29	5	1283	111	21	767	6430	10