# Internal loops in RNA secondary structure prediction

Rune B. Lyngsø* Michael Zuker† Christian N. S. Pedersen‡

**Abstract**

We present an analysis of currently used free energy functions for internal loop stability in RNA secondary structure. This analysis enables us to present an $O(|s|^3)$ algorithm for evaluating internal loops thus improving the overall complexity of RNA secondary structure prediction from $O(|s|^4)$ to $O(|s|^3)$. Using an implementation of this algorithm we examine how reasonable a commonly used heuristic of limiting the size of internal loops evaluated has been.

## 1 Introduction

Structure prediction remains one of the most compelling yet elusive areas of computational biology. Not yielding to overwhelming numbers and resources this area still poses a lot of interesting questions for future research. One of the areas where most progress has been achieved is the prediction of the secondary structure of RNA molecules.

In RNA, the major forces governing structure formation seems to be well understood as the nearest neighbor model proposed by Tinoco et. al. in [11, 10] combined with experimentally determined free energy parameters yields fair predictions of structure stabilities. Furthermore, this model allows for efficient dynamic programming algorithms (e.g. [17, 7]) for rigorously computing the structures of minimum free energy in the model and thus for prediction of the secondary structure of the optimal folding.

The most time consuming part of these algorithms is the evaluation of internal loops which in general requires time $O(|s|^4)$ where $|s|$ is the length of the RNA sequence for which we want to find the structures of minimum free energy. This problem has previously been addressed, e.g. in [12] and [1]. These solutions unfortunately do not allow for the complexity of the functions currently used for estimating the stability of internal loops, and therefore a heuristic limiting the size of internal loops evaluated to some cutoff size (e.g. 30 as suggested in [4]) is usually invoked.

In this paper we will analyze the structure of currently used internal loop stability functions to obtain an $O(|s|^3)$ algorithm for rigorously evaluating all internal loops regardless of size. Furthermore, we will present a study using this algorithm of how reasonable the hitherto used cutoff size of 30 is. The paper is structured as follows. In section 2 we present the basic RNA secondary structure prediction algorithm. In section 3 we analyze the internal loop stability function and, based on this analysis present our $O(|s|^3)$ algorithm for evaluating internal loops. In section 4 we compare secondary structure predictions performed with and without a cutoff size of 30.

## 2 Basic dynamic programming algorithm

A secondary structure of a sequence $s$ is a set $S$ of base pairs $i \cdot j$ with $1 \leq i < j \leq |s|$ such that $\forall i \cdot j, i' \cdot j' \in S : i = i' \Leftrightarrow j = j'$. Thus any base can take part in at most one base pair. We will further assume that the structure does not contain pseudo-knots. A pseudo-knot is two 'overlapping' base pairs, that is base pairs $i \cdot j$ and $i' \cdot j'$ with $i < i' < j < j'$.

One can view a pseudo-knot free secondary structure $S$ as a collection of *loops* together with some *external* unpaired bases (see figure 1). Let $i < k < j$ with $i \cdot j \in S$. Then $k$ is said to be *accessible* from $i \cdot j$ if for all

*Dept. of Computer Science, University of Aarhus, Århus, Denmark; email: rlyngsoe@daimi.au.dk. Work done while visiting the Institute for Biomedical Computing at Washington University.

†Institute for Biomedical Computing, Washington University, St. Louis, USA; email: zuker@ibc.wustl.edu.

‡Basic Research in Computer Science, University of Aarhus, Århus, Denmark; email: cstorm@brics.dk. Supported by the ESPRIT Long Term Research Programme of the EU under project number 20244 (ALCOM-IT). Work done while visiting the University of California at Davis and DIMACS at Rutgers University.

$i' \cdot j' \in S$ it is not the case that $i < i' < k < j' < j$. The base pair $i \cdot j$ is said to be the *exterior* base pair of (or *closing*) the loop consisting of $i \cdot j$ and all bases accessible from it. If $i'$ and $j'$ are accessible from $i \cdot j$ and $i' \cdot j' \in S$ – observe that for a structure without pseudo-knots either both or none of $i'$ and $j'$ will be accessible from $i \cdot j$ if $i' \cdot j' \in S$ – then $i' \cdot j'$ is called an *interior* base pair of the loop and is said to be accessible from $i \cdot j$. If there are no interior base pairs the loop is called a *hairpin* loop. With one interior base pair it is called a *stacked pair* if $i' = i + 1$ and $j' = j - 1$ and otherwise an *internal* loop (*bulges* are a special kind of internal loops with either $i' = i + 1$ or $j' = j - 1$). Loops with more than one interior base pair are called *multibranched* loops. Unpaired bases and base pairs not accessible from any base pair are called external.

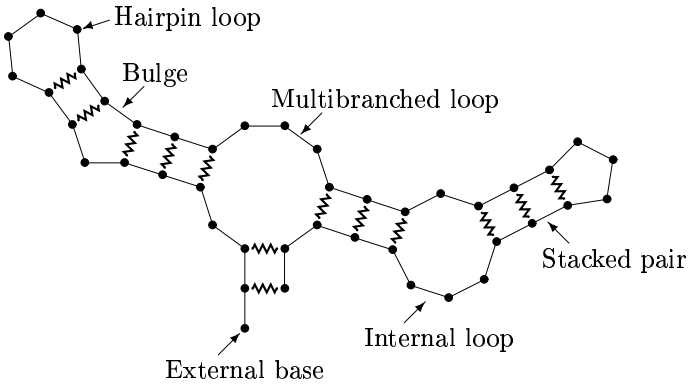



Figure 1: An example RNA structure. Bases are depicted by circles, the RNA backbone by straight lines and base pairings by zigzagged lines.

RNA secondary structure prediction is the problem of determining the most stable structure for a given sequence. We measure stability in terms of the free energy of the structure. Thus we want to find a structure of minimal free energy which we will also call an optimal structure. The energy of a secondary structure is assumed to be the sum of the energies of the loops of the structure and furthermore the loops are assumed to be independent, that is the energy of a loop only depends on that loop and not on the rest of the structure[10].

Based on these assumptions one can specify a recursion to calculate the energy of the optimal structure for a sequence $s$. We will use four arrays[1] – $W$, $V$, $VBI$ and $VM$ – to hold the minimal free energy of certain structures of subsequences of $s$. The entries of these arrays are interdependent and can be calculated recursively using pre-specified free energy functions – $eS$, $eH$, $eL$ and $eM$ – for the contributions from the various types of loops as follows.

- The energy of an optimal structure of the subsequence from 1 through $i$:

$$W(i) = \min\{W(i-1), \min_{1<j\leq i}\{W(j-1) + V(j,i)\}\}.$$

- The energy of an optimal structure of the subsequence from $i$ through $j$ closed by $i \cdot j$:

$$V(i,j) = \min\{eH(i,j), eS(i,j) + V(i+1,j-1), VBI(i,j), VM(i,j)\}$$

where $eH(i,j)$ is the energy of a hairpin loop closed by $i \cdot j$ and $eS(i,j)$ is the energy of stacking base pair $i \cdot j$ with $i+1 \cdot j-1$.

- The energy of an optimal structure of the subsequence from $i$ through $j$ where $i \cdot j$ closes a bulge or an internal loop:

$$VBI(i,j) = \min_{\substack{i<i'<j'<j \\ i'-i+j-j'>2}}\{eL(i,j,i',j') + V(i',j')\}$$

where $eL(i,j,i',j')$ is the energy of a bulge or internal loop with exterior base pair $i \cdot j$ and interior base pair $i' \cdot j'$.

- The energy of an optimal structure of the subsequence from $i$ through $j$ where $i \cdot j$ closes a multibranched loop:

$$VM(i,j) = \min_{\substack{i<i_1<j_1< \\ \cdots \\ <i_k<j_k<j}}\{eM(i,j,i_1,j_1,\ldots,i_k,j_k) + \sum_{l=1}^{k} V(i_l,j_l)\}$$

where $k > 1$ and $eM(i,j,i_1,j_1,\ldots,i_k,j_k)$ is the energy of a multibranched loop with exterior base pair $i \cdot j$ and interior base pairs $i_1 \cdot j_1, \ldots, i_k \cdot j_k$.

When all entries of these arrays have been filled out $W(|s|)$ contains the free energy for optimal structures and an optimal structure can be determined by backtracking the calculations that led to this free energy.

To make the problem of determining the optimal secondary structure tractable the following simplifying assumption is often made. The energy of multibranched loops can be decomposed into linear contributions from the number of unpaired bases in the loop, the number of branches in the loop and a constant[16][2], that is

$$eM(i,j,i_1,j_1,\ldots,i_k,j_k) = a + bk + c(i_1 - i - 1 + j - j_k - 1 + \sum_{l=1}^{k-1}(i_{l+1} - j_l - 1)). \quad (1)$$

We introduce an extra array

[1] Actually two arrays – $V$ and $W$ – suffices but we will use four arrays to simplify the description. Below we will introduce a fifth array $WM$ that will also be needed in an efficient implementation.

[2] It is known that the stability of a multibranched loop also depends on the stacking effects of the base pairs in the loop and their neighboring unpaired bases. These effects can also be handled efficiently but for simplicity we have omitted the details here.

- The energy of an optimal structure of the subsequence from $i$ through $j$ that constitutes part of a multibranched loop structure, that is where unpaired bases and external base pairs are penalized according to 1:

$$\begin{aligned} WM(i,j) = \min\{&V(i,j)+b,\\ &WM(i,j-1)+c,\\ &WM(i+1,j)+c,\\ &\min_{i<k\leq j}\{WM(i,k-1)+WM(k,j)\}\} \end{aligned}$$

which enables us to restate the calculation of the energy of the optimal multibranched loop as

$$VM(i,j) = \min_{i+1<k\leq j-1}\{WM(i+1,k-1)+WM(k,j-1)+a\}.$$

Based on these recurrence relations we can by dynamic programming calculate the energy of the optimal structure in time $O(|s|^3)$ – assuming that the free energy functions can be evaluated in constant time – except for the calculation of the entries of $VBI$ which requires $O(|s|^4)$ in total.

## 3 Internal loop evaluation

Examining the recursion for internal loops one observes that two base pairs, $i \cdot j$ and $i' \cdot j'$, may be compared as candidates for the interior base pair for numerous exterior base pairs. If $V(i,j) \ll V(i',j')$ it is evident that we would not have to consider $i' \cdot j'$ as a candidate interior base pair for any entry of $VBI$ where $i \cdot j$ would also be a candidate interior base pair.

Though it would often in practice be the case that we could a priori discard many candidate interior base pairs by the above observation, we can not in general guarantee this to be the case. To get an improvement in the worst case performance of the evaluation of internal loops we thus have to examine properties of the energy functions for internal loop stability that will allow us to group base pairs and entries of $VBI$ such that we only have to make one comparison between $i \cdot j$ and $i' \cdot j'$ to determine which one would yield the more stable structure for the entire group of entries. In this section we will exploit such properties of currently used energy functions leading to an algorithm for evaluating internal loops requiring worst case time $O(|s|^3)$.

Currently used energy rules for internal loop stability (cf. [14]) split the contributions into three parts:

- An entropic term that depends on the size of the loop.
- Stacking energies for the mismatched base pairs adjacent to the closing (exterior *and* interior) base pairs.
- An asymmetry penalty for asymmetric loops.

With this separation we can rewrite the internal loop energy function as

$$\begin{aligned} eL(i,j,i',j') = &\text{size}(i'-i+j-j'-2)+\\ &\text{stacking}(i\cdot j)+\text{stacking}(i'\cdot j')+\\ &\text{asymmetry}(i'-i-1,j-j'-1) \end{aligned} \quad (2)$$

thus turning the recursion for internal loops into

$$\begin{aligned} &VBI(i,j) = \\ &\min_{\substack{i<i'<j'<j\\ i'-i+j-j'>2}} \{V(i',j')+\\ &\quad \text{stacking}(i'\cdot j')+\text{stacking}(i\cdot j)+\\ &\quad \text{size}(i'-i+j-j'-2)+\\ &\quad \text{asymmetry}(i'-i-1,j-j'-1)\}. \end{aligned} \quad (3)$$

To eliminate the size dependence from the comparison between candidate interior base pairs we group base pairs by the distance between the bases. So let $j'-i' = j''-i''$ and assume that $V(i',j')+\text{stacking}(i'\cdot j') < V(i'',j'')+\text{stacking}(i''\cdot j'')$. Now consider the internal loops with $i \cdot j$ (where $i < i', i''$ and $j > j', j''$) as exterior base pair and $i' \cdot j'$ respectively $i'' \cdot j''$ as interior base pairs. As the contribution to the stability from size and the contribution from mismatched base pair stacking with $i \cdot j$ are the same for these internal loops, the structure with interior base pair $i' \cdot j'$ would always be the most stable of the two if we could disregard any penalty due to asymmetry. This is used in [12] to obtain an $O(|s|^3)$ algorithm for RNA folding when internal loop stability only depends on size.

Having already mentioned the asymmetry penalty we obviously do not intend to disregard the asymmetry penalty. We will however exploit properties of currently used functions for the asymmetry penalty to obtain an efficient algorithm for evaluating internal loops. In [8] an asymmetry penalty function of the form

$$\text{asymmetry}(n_1,n_2) = \min\{E_{max}, n\cdot f(m)\} \quad (4)$$

where $n = |n_1 - n_2|$, $m = \min\{n_1, n_2, c\}$ and $E_{max}$ is the maximum penalty was proposed. The $c$ is a constant that in [8] was set to 5; for currently used parameters based on thermodynamic studies the value of $c$ is 1, cf. [9]. The key observation is that if we fix the lopsidedness (denoted by $n$ in equation 4) the asymmetry penalty will not change with size[3], that is

$$\text{asymmetry}(n_1,n_2) = \text{asymmetry}(n_1+1,n_2+1) \quad (5)$$

for $n_1, n_2 \geq c$. Now let $j'-i' = j''-i''$, $i+c < i', i''$ and $j-c > j', j''$. Assume that an internal loop with $i \cdot j$ and $i' \cdot j'$ as closing base pairs yields a more stable

[3] Actually we only need that the size dependence is identical for all choices of lopsidedness.

structure than an internal loop with $i \cdot j$ and $i'' \cdot j''$ as closing base pairs (that is $V(i',j') + eL(i,j,i',j') < V(i'',j'') + eL(i,j,i'',j'')$). Then for all $l \geq 0$ we know that $i' \cdot j'$ is a better choice for interior base pair than $i'' \cdot j''$ for an internal loop with exterior base pair $i-l \cdot j+l$ as

$$\begin{aligned}
&V(i',j') + eL(i-l, j+l, i', j') \\
&\quad = V(i',j') + \text{size}(i'-i+j-j'+2l-2) + \\
&\quad\quad \text{stacking}(i-l \cdot j+l) + \text{stacking}(i' \cdot j') + \\
&\quad\quad \text{asymmetry}(i'-i+l-1, j+l-j'-1) \\
&\quad = V(i',j') + \text{size}(i'-i+j-j'-2) + \\
&\quad\quad \text{stacking}(i \cdot j) + \text{stacking}(i' \cdot j') + \\
&\quad\quad \text{asymmetry}(i'-i-1, j-j'-1) + \\
&\quad\quad \text{size}(i'-i+j-j'+2l-2) + \\
&\quad\quad \text{stacking}(i-l \cdot j+l) - \text{stacking}(i \cdot j) - \\
&\quad\quad \text{size}(i'-i+j-j'-2) \\
&\quad < V(i'',j'') + \text{size}(i''-i+j-j''-2) + \\
&\quad\quad \text{stacking}(i \cdot j) + \text{stacking}(i'' \cdot j'') + \\
&\quad\quad \text{asymmetry}(i''-i-1, j-j''-1) + \\
&\quad\quad \text{size}(i''-i+j-j''+2l-2) + \\
&\quad\quad \text{stacking}(i-l \cdot j+l) - \text{stacking}(i \cdot j) - \\
&\quad\quad \text{size}(i''-i+j-j''-2) \\
&\quad = V(i'',j'') + \text{size}(i''-i+j-j''+2l-2) + \\
&\quad\quad \text{stacking}(i-l \cdot j+l) + \text{stacking}(i'' \cdot j'') + \\
&\quad\quad \text{asymmetry}(i''-i+l-1, j+l-j''-1) \\
&\quad = V(i'',j'') + eL(i-l, j+l, i'', j'').
\end{aligned} \tag{6}$$

Assume we know the optimal interior base pair, $i' \cdot j'$, for internal loops of size $l$ (and shortest stretch of unpaired bases at least $c$) with exterior base pair $i \cdot j$. We can then find the optimal interior base pair for internal loops of size $l+2$ (and shortest stretch of unpaired bases at least $c$) with exterior base pair $i-1 \cdot j+1$ in constant time. This is done by comparing the structure with interior base pair $i' \cdot j'$ with the two structures with an interior base pair that gives a shortest stretch of unpaired bases of exactly $c$ bases. This method is used in algorithm 1 to obtain an $O(|s|^3)$ solution to the evaluation of internal loops with shortest stretch of unpaired bases at least $c$. As there are only $O(c|s|^3)$ internal loops with shortest stretch of unpaired bases shorter than $c$ we can evaluate these individually and still get an $O(|s|^3)$ – or more precisely $O(c|s|^3)$ – algorithm for the evaluation of internal loops. It is an easy observation that algorithm 1 requires time $O(|s|)$ for each choice of $i, j$, thus yielding an $O(|s|^3)$ algorithm when iterating over all choices of $i, j$ with $i < j$.

To see that we do in fact evaluate all internal loops

**Algorithm 1** Evaluation of internal loops of size $2l+a$ with exterior base pair $i-l \cdot j+l+a$ and shortest stretch of unpaired bases at least $c$.

**Require:** $i, j$ with $i < j$.
  **for** $a = 0$ to $1$ **do**
    $E = \infty$
    **for** $l = c+1$ to $\min\{i-1, |s|-j-a\}$ **do**
      $E = \min\{E,$
        $V(i-l+c+1, j-l+c+1) +$
          $\text{asymmetry}(c, 2l+a-c-2) +$
          $\text{stacking}(i-l+c+1, j-l+c+1),$
        $V(i+a+l-c-1, j+a+l-c-1) +$
          $\text{asymmetry}(2l+a-c-2, c) +$
          $\text{stacking}(i+a+l-c-1, j+a+l-c-1)\}$
      $VBI(i-l, j+a+l) =$
        $\min\{VBI(i-l, j+a+l),$
        $E + \text{size}(2l+a-2) + \text{stacking}(i-l, j+a+l)\}$
    **end for**
  **end for**

with shortest stretch of unpaired bases at least $c$ consider such a loop with exterior base pair $i' \cdot j'$ and interior base pair $i'' \cdot j''$. We will now specify for what choice of $i, j$ algorithm 1 will evaluate this loop.

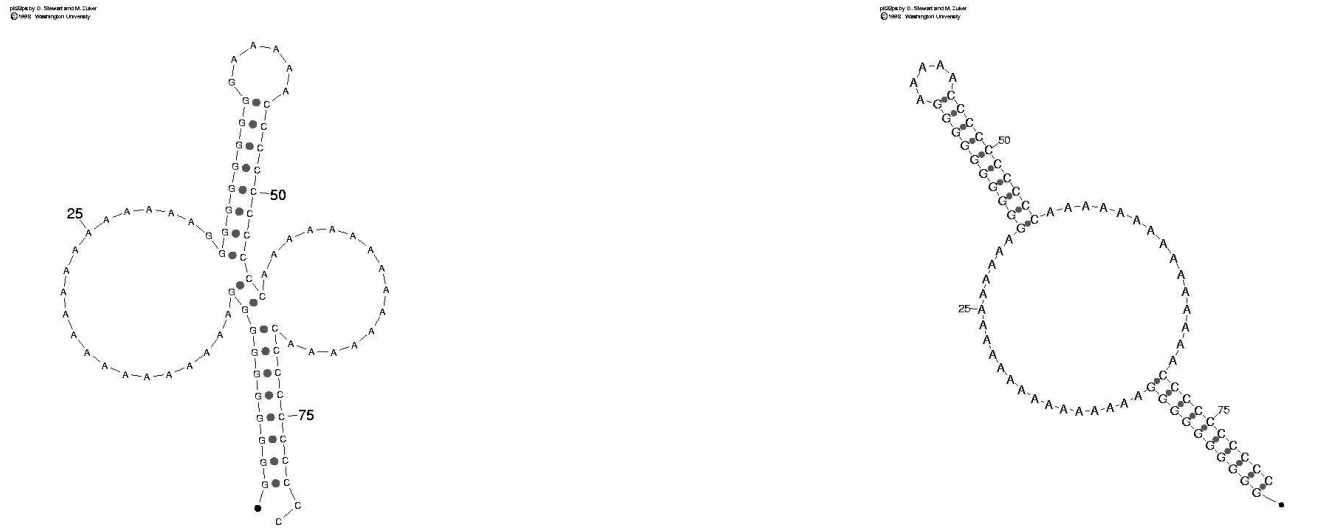
Let $u = \lfloor (i''-i'+j'-j'')/2 \rfloor$ and $v$ be the parity of the size of the loop (0 if the size of the loop is even and 1 if the size of loop is odd). Let $i = i'+u$, $j = j'-u-v$ and $a = v$. One can check that for $l = |i''-i| + c + 1$ ($-a$ if $i'' > i$) the base pair $i'' \cdot j''$ is one of the two base pairs considered for a new value of $E$. For $l = u$ we get that $i-l \cdot j+a+l = i' \cdot j'$ and thus – through $E$ – that $V(i'',j'') + eL(i',j',i'',j'')$ is considered for the value of $VBI(i',j')$.

## 4 Results

To make the problem of determining the optimal secondary structure for an RNA sequence more tractable it has hitherto been common practice to limit the size of internal loops. The `mfold` server has a built-in limit of 30 and in [4] a limit of 30 is also hinted at. With the ability to make a rigorous search for the optimal structure we decided to see whether this limit has been reasonable.

### 4.1 A constructed 'mean' sequence

The easiest way to find a loop of size larger than 30 is of course to construct it yourself. We constructed a sequence of length 80 consisting only of C's, G's and A's (but no U's) designed to fold into two stems of 10 base pairing C's and G's separated by an internal loop of 35 unpaired A's and with a hairpin loop consisting of




(a) Maximum loop size 30; Energy: −29.6 kcal/mol

(b) No maximum loop size; Energy: −42.9 kcal/mol

Figure 2: Foldings of the sequence GGGGGGGGGGAAAAAAAAAAAAAAAAAAAGGGGGGGGGGAAAAACCCCCCCCCCCAAAAAAAAAAAAAAACCCCCCCCCC

5 A's. The result of folding this sequence at 37 °C with and without a size limit of 30 respectively is shown in figure 2

One can observe that the prediction with a cutoff size of 30 does in fact pair most of the C's with G's – but instead of having the A's in one big internal loop they are folded out as two bulges. A further observation is that there can indeed be a major increase in stability by choosing one large internal loop instead of two smaller bulges.

Though this example may be cute, the interesting question of course is whether RNA sequences for which the optimal structure contains a large internal loop occur naturally. The reason that a cutoff size of 30 has been deemed reasonable is of course that no internal loops even close to this size are observed in a standard structure prediction at 37 °C. But when the temperature is increased, base pairs become less stable which may cause short stems of stacking base pairs to break up. We thus decided to look at a couple of sequences for which structure prediction at higher temperatures would be interesting.

### 4.2 Q$\beta$

Jacobson [5] reported on some experiments on determining structural features in Q$\beta$ denatured to various extents. It is believed that denaturing effects relates to temperature effects and we thus chose to fold this sequence at nine different temperatures in the range from 45 °C to 100 °C to see whether we would find any of the structural features reported by Jacobson.

None of these predicted foldings showed any signs of the features Jacobson reported – at higher temperatures the structure simply came apart as small structural fragments usually covering less than 100 nucleotides. Furthermore we did not observe any internal loops larger than size 25. An example prediction is shown in figure 3.

### 4.3 *Thermococcus celer*

*Thermococcus celer* is an organism that lives in solfataric marine water holes of Vulcano, Italy, at temperatures around 90 °C; it's optimal growth temperature is reported to be around 88 °C [13]. Furthermore, the structure of the 23S subunit exhibits an internal loop of size 33 closed by base pairs $1139 \cdot 1268$ and $1155 \cdot 1249$, cf. [3, 2].

Folding this sequence at 88 °C we did (almost) get the inner stem of this internal loop but the outer stem came apart as two single strands (cf. figure 4(b)). When lowering the temperature to 75 °C we did get both stems but the internal loop was split into two loops of size 2 and 27 respectively by a short stem consisting of the base pairs $1141 \cdot 1266$ and $1142 \cdot 1265$ (cf. figure 4(c)).

We then tried to search the range of temperatures between 75 °C and 88 °C and at 82 °C we did in fact correctly predict the internal loop of size 33 (cf. fig-

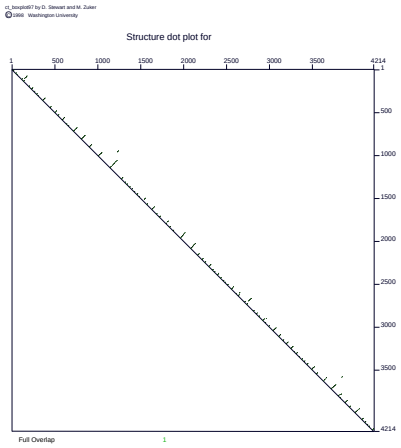



Figure 3: Dot-plot of the prediction of the Q$\beta$ structure at 65 °C. The absence of long range base pairings (dots far away from the diagonal) is apparent.

ure 4(d)). At this temperature we on the other hand missed the structure inside the inner stem, a structure that is quite well predicted at 75 °C; no temperature thus seemed decisively best for predicting this structural fragment. Generally, as with the Q$\beta$ predictions, these predictions missed long-range base pairings and predicted structures consisting of fragments covering less than 300 bases.

It should however be mentioned that a prediction at 82 °C with a cutoff size of 30 completely misses the outer stem and thus makes a prediction of this fragment identical to the prediction at 88 °C. Thus we get a decisively better prediction at this temperature when examining internal loops of all sizes than when using a cutoff size of 30.

## 5 Conclusion

Our studies of structure predictions at high temperatures did not show an abundance of internal loops larger than the hitherto used cutoff size. There is thus no reason to suspect that predictions using this cutoff size are generally erroneous. We were however able to predict one internal loop that exceeds this size limit. Furthermore we predicted a number of internal loops with size larger than 20. This indicates that the cutoff size of 30 is probably a little bit to small for safe predictions at high temperatures. Especially if also suboptimal foldings (cf. [15]) are sought for, or if calculating the partition functions[4] as in [6], the cutoff size – if used at all – should be set somewhat higher.

Another observation is that the energy parameters estimated for higher temperatures by extrapolation of parameters experimentally determined at lower temperatures do not seem to allow for a prediction of the long range base pairings. One reason for this might be that structures at higher temperatures tend to have more unpaired bases in multibranched loops. The effect of the number of unpaired bases on the stability of multibranched loops should theoretically be logarithmic but are modeled by a linear function for reasons of computational efficiency. This might be acceptable for multibranched loops with only a few unpaired bases but becomes prohibitive as the number of unpaired bases grows.

[4] The method we have described in this paper can also be applied to the calculations of internal loops contributions to partition functions, thus allowing for a rigorous $O(|s|^3)$ computation of these too.

## Acknowledgments

The authors would like to thank Darrin Stewart for his help with generating the figures for this article. This work was supported, in part, by NIGMS grant GM54250 to MZ.

## References

[1] David Eppstein, Zvi Galil, and Raffaele Giancarlo. Speeding up dynamic programming. In *Proc. 29th Symp. Foundations of Computer Science*, pages 488–496. Assoc. Comput. Mach., October 1988.

[2] R. R. Gutell. http://pundit.icmb.utexas.edu/. RNA secondary structures.

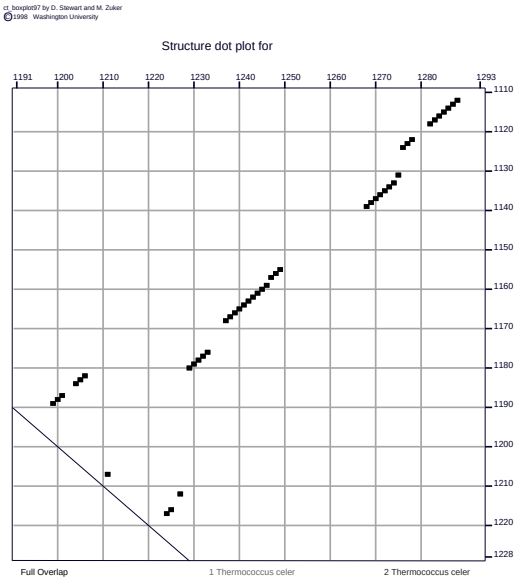
[3] R. R. Gutell, M. W. Gray, and M. N. Schnare. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. *Nucleic Acids Res.*, 21:3055–3074, 1993. Database issue.

[4] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie*, (125):167–188, 1994.

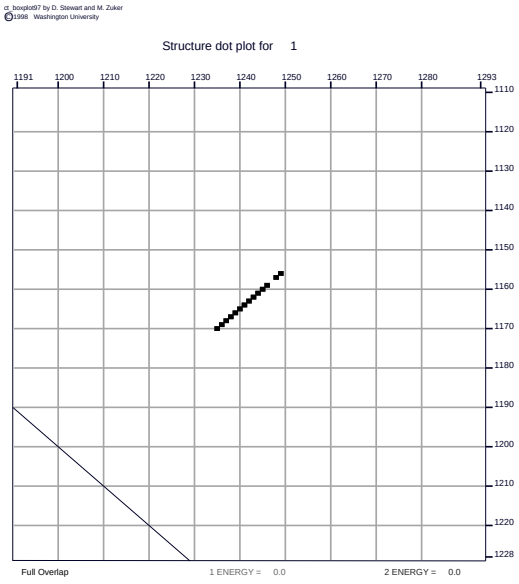
[5] Ann B. Jacobson. Secondary structure of coliphage Q$\beta$ RNA. *Journal of Molecular Biology*, 221:557–570, 1991.

[6] John S. McCaskill. The equilibrium partiton function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

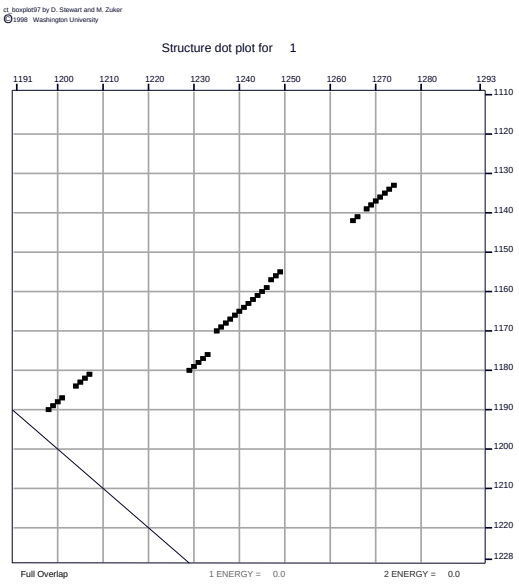
[7] Ruth Nussinov and Ann B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. USA*, 77(11):6309–6313, 1980.



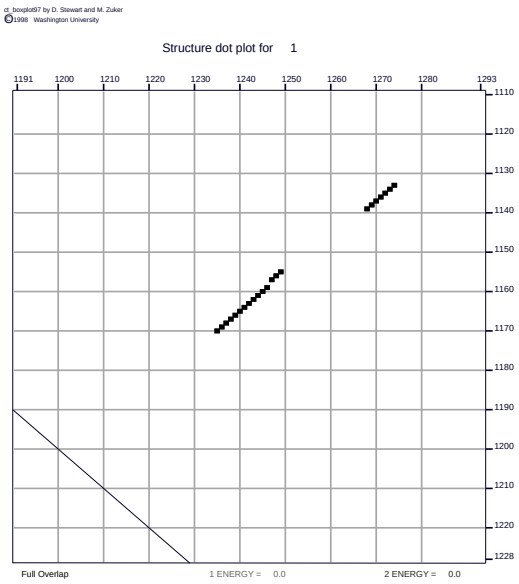

(a) Fragment of the structure between bases 1112 and 1288.




(b) Prediction of the same fragment at 88 °C.




(c) Prediction of the same fragment at 75 °C.




(d) Prediction of the same fragment at 82 °C.

Figure 4: Known and Predicted structures for *thermococcus celer*.

[8] Chaterine Papanicolaou, Manolo Gouy, and Jacques Ninio. An energy model that predicts the correct folding of both the tRNA and the 5S RNA molecules. *Nucleic Acids Research*, 12:31–44, 1984.

[9] Adam E. Peritz, Ryszard Kierzek, Naoki Sugimoto, and Douglas H. Turner. Thermodynamic study of internal loops in oligoribonucleotides: symmetric loops are more stable than asymmetric loops. *Biochemistry*, 30:6428–6436, 1991.

[10] Ignacio Tinoco, Jr., Philip N. Borer, Barbara Dengler, Mark D. Levine, Olke C. Uhlenbeck, Donald M. Crothers, and Jay Gralla. Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology*, 246:40–41, 1973.

[11] Ignacio Tinoco, Jr., Olke C. Uhlenbeck, and Mark D. Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230:362–367, 1971.

[12] Michael S. Waterman and T. F. Smith. Rapid dynamic programming methods for RNA secondary structure. *Advances in Applied Mathematics*, 7:455–464, 1986.

[13] W. Zillig, I. Holz, D. Janekovic, W. Schäfer, and W. D. Reiter. The archaebacterium *Thermococcus celer* represents, a novel genus within the thermophilic branch of archaebacteria. *Systematic and Applied Microbiology*, 4:88–94, 1983.

[14] Michael Zuker. http://www.ibc.wustl.edu/~zuker/rna/energy/node2.html. RNA web-page.

[15] Michael Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244:48–52, 1989.

[16] Michael Zuker and David Sankoff. RNA secondary structures and their prediction. *Bull. Mathematical Biology*, 46:591–621, 1984.

[17] Michael Zuker and Patrick Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.