Research Articles

On Finding All Suboptimal Foldings of an RNA Molecule

MICHAEL ZUKER

An algorithm and a computer program have been prepared for determining RNA secondary structures within any prescribed increment of the computed global minimum free energy. The mathematical problem of determining how well defined a minimum energy folding is can now be solved. All predicted base pairs that can participate in suboptimal structures may be displayed and analyzed graphically. Representative suboptimal foldings are generated by selecting these base pairs one at a time and computing the best foldings that contain them. A distance criterion that ensures that no two structures are "too close" is used to avoid multiple generation of similar structures. Thermodynamic parameters, including free-energy increments for single-base stacking at the ends of helices and for terminal mismatched pairs in interior and hairpin loops, are incorporated into the underlying folding model of the above algorithm.

THE RNA SECONDARY STRUCTURE MODEL HAS BEEN IN existence since Fresco *et al.* (*1*) first showed that single-stranded RNA folds back onto itself in structures stabilized by hydrogen bonds between complementary bases. This model is not concerned with three-dimensional aspects of structure, but focuses solely on which hydrogen bonds form. This approach is appropriate, because while detailed three-dimensional structure data exists only for transfer RNA (*2*), three-dimensional modeling is premature for general RNA molecules.

This folding model is an example of what mathematicians call a discrete model. There are no continuously varying parameters such as bond lengths, angles, or interatomic distances. Instead, either a hydrogen bond exists between two complementary bases or it does not. One of the principal advantages of dealing with such a structural model is that mathematical tools exist to compute an optimal folding based on free-energy minimization. The pitfalls of becoming trapped in local energy minima that are encountered in models with a large number of continuous parameters can be avoided.

The model, however, has the mathematical property that there can be numerous foldings within 5 or 10 percent of the computed minimum free energy. Moreover, these foldings can be topologically very different from one another. For example, an alternative folding to the computed minimum free-energy folding of the 5.8*S* RNA from *Crypthecodinium cohnii* has an energy within 5 percent of the global minimum and yet shares not a single base pair with the optimal folding (*3*). The uncertainties inherent in the model and in the thermodynamic data on which folding is based can be mitigated if a means of predicting suboptimal foldings is available.

Two types of RNA folding algorithms have the ability to find a minimum energy secondary structure. The "combinatorial" method, first introduced by Pipas and McMahon (*4*), forms structures by combining all potential helices in all possible ways. By their nature, combinatorial algorithms predict alternative foldings. The program developed by Ninio and co-workers (*5–7*) is based on a time-saving tree search method, but it does not escape from combinatorial reality. The number of possible foldings, and hence the computation time, grow exponentially with the size of the sequence (*8*), and it is not surprising that this and similar programs are limited to folding about 150 to 200 bases.

Minimum energy foldings can also be computed with recursive, or dynamic programming, algorithms. They were first used in the RNA folding problem by Nussinov *et al.* (*9*) to maximize base pairing. This method was subsequently extended to energy minimization (*10, 11*). These programs work in two stages. The first part, called the fill algorithm, computes and stores minimum folding energies for all fragments of the sequence. The process begins with all pentanucleotides and builds up to larger fragments in a recursive fashion. The second algorithm, called the traceback, computes a minimum energy structure by searching systematically through the matrix of stored energies. The main advantages over combinatorial algorithms are speed and the ability to fold relatively large molecules. By examining possible base pairs in the context of what neighboring base pairs might be, the algorithm escapes the tyranny of an exponentially growing number of structures. If the treatment of multibranched loops is sufficiently simple (*12*), a recursive folding algorithm can execute in time proportional to the cube of the sequence length. My own algorithm (*11*) can fold about 2000 bases on a VAX 11/750.

The main weakness of many recursive-folding algorithms (*9–11*) is that by design they yield only a single solution. The entire folding process can be repeated with slightly perturbed energy rules, but this is a prohibitively expensive way to generate alternative foldings. Williams and Tinoco (*13*) have extended a dynamic programming algorithm similar to others (*11, 14*) so that multiple foldings are predicted. However, its selection of computed foldings is arbitrary in that it depends on the idiosyncrasies of the algorithm, which must choose from a myriad of possibilities. These foldings do not provide a total picture on how much variation is possible and on how robust the predictions are.

The previously described recursive-folding algorithm (*11*) computes and stores the minimum folding energy for each subsequence of the given RNA sequence. Also, for each subsequence, we calculated the minimum folding energy for the fragment with the ends constrained to form a base pair with each other if possible. For a fragment stretching from ribonucleotides i to j, this number is denoted by $V(i,j)$ and is needed for proper function of the algorithm.

The author is a fellow of the Canadian Institute for Advanced Research and is head of the Biomolecular Modeling Section in the Protein Laboratory of the Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6.

The first step toward this multiple folding algorithm came with attempts to extend the algorithm to fold circular RNA such as viroids (*15*). In a circular RNA, the choice of an origin is arbitrary. The key observation is that, in a circular molecule composed of ribonucleotides $r_1, r_2, \ldots, r_n$, a base pair linking r_i and r_j divides the secondary structure into two parts. There is a folding of the "included fragment" from r_i to r_j, and another folding of the "excluded fragment" from r_j through the origin to r_i. In a linear molecule, this symmetry is lost since the "excluded fragment" is broken into two linear segments, r_1 to r_i and r_j to r_n. The additivity assumption characteristic of recursive algorithms implies that the total folding energy is the sum of the energies of the two foldings. Steger *et al.* (*16*) extend the algorithm of Zuker and Stiegler (*11*) by computing additional numbers $V(j,i)$, analogous to $V(i,j)$, but referring to the "excluded fragments" instead. These numbers can also be computed recursively. They observe that $V(i,j) + V(j,i)$ is the minimum free energy of a structure containing the base pair r_i-r_j, and that the minimum value of $V(i,j) + V(j,i)$ over all possible base pairs is the minimum folding energy, E_{min}, for the circular RNA molecule. A similar extension to folding circular RNA was made subsequently (*17*).

The above extension provides all that is necessary for the realization of a multiple folding algorithm, at least for circular RNA. The time-consuming fill algorithm is executed normally, although the circular algorithm requires twice as much time and computer storage as the regular algorithm for a sequence of the same size. Instead of merely identifying a base pair r_i-r_j that gives E_{min} and computing an optimal folding, the strategy is to identify all base pairs for which $V(i,j) + V(j,i)$ is "close" to E_{min}. If P is a number between 0 and 100, then a "P-optimal" base pair is a base pair r_i-r_j for which $V(i,j) + V(j,i) \geq (1 - P/100) \times E_{min}$. Thus a P-optimal base pair is contained in at least one folding within P percent of the minimum free energy. Such a folding is defined as a P-optimal folding. The collection of all P-optimal base pairs is the mathematical union of all P-optimal foldings. This information must then be displayed and interpreted. Also, the actual number of foldings within 5 or 10 percent of the optimal energy might be very large. We can first plot each P-optimal base pair r_i-r_j as a point at the ith row and jth column of a triangular half-matrix and thus produce a picture of the superposition of all P-optimal foldings. Such plots are called energy dot plots (*18*). An examination of which regions are empty and which are full of dots and the comparison of energy dot plots as P increases reveals how well determined the various motifs of the RNA structure are. An optimal or suboptimal folding can be generated by choosing an optimal or suboptimal base pair and computing the best folding containing that base pair. This procedure does not generate all possible foldings, but local motifs that can be part of P-optimal foldings are found.

The procedure for circular RNA generalizes to linear RNA. The linear molecule is handled as if it were circular, provided that the first and last bases, now regarded as adjacent, be allowed to pair with each other if necessary. In addition, loops containing the origin must be treated as special cases. For example, a hairpin loop containing the origin becomes two single-stranded regions at the 5′ and 3′ ends of the molecule. This artificial circularization would bias the results in a dynamic simulation of folding, but causes no problems with this algorithm in which foldings are computed independently of folding pathways.

The choice of P-optimal foldings, rather than foldings within a fixed energy of the global minimum, is deliberate. If the folding rules and energy parameters were well determined, it would suffice to look within 3 kcal/mole from the optimal energy to find all structures that occur 99.5 percent of the time. This guideline is the result of the Boltzmann energy distribution at 300 K. A deviation of 5 or 10 percent from an optimal folding of −100 kcal/mole would correspond to rare events with probabilities 2×10^{-4} and 6×10^{-8}, respectively. These large energy increments are chosen not for thermodynamic reasons, but because of the large uncertainties in the energy data; thus the biochemically correct folding should be within a 5 or 10 percent energy increment.

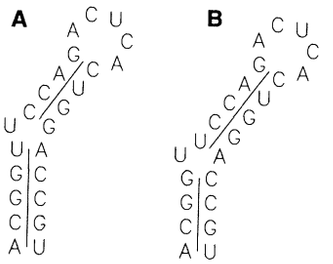

Fig. 1. Two foldings of the same oligonucleotide fragment that are a distance of 1 apart. The base pair U^5-A^{20} of (**A**) does not occur in (**B**), but its base numbers (*5, 20*) are both within 1 of the base numbers of G^4-C^{21} in (B). Similarly, U^6-A^{20} of (B) is equally close to C^7-G^{19} of (A). All other base pairs are common.

Implementation. The multiple folding algorithm is programmed in Fortran 77 and runs in a VAX/VMS environment; energy rules used are those set by Freier, Turner, and colleagues (*19, 20*). They differ from the rules summarized by Salser (*21*) in a number of ways: (i) They are computed for folding at 37°C rather than 25°C. (ii) The new rules add single-base stacking energies for dangling bases adjacent to helices. Mismatched pairs adjacent to the closing base pair (or pairs) of interior and hairpin loops are also taken into account. (iii) Ninio's correction for lopsided interior loops is used (*6*). Multiple branched loops are assigned a fixed penalty of 4.7 kcal/mole plus 0.4 kcal/mole per single-stranded base and 0.1 kcal/mole per closing base pair. These are adjustable constants. Single-base stacking is computed in these loops where applicable.

There are two modes of operation. In the first, energy dot plots appear on the screen to any desired percent from the minimum energy. Base pairs can be selected with the use of the cross-hair feature of Tektronix-type terminals, and optimal or suboptimal foldings can be computed containing the chosen base pair. In the second mode of operation, foldings are generated automatically and sorted by energy. The automatic procedure may generate a large number of foldings within 5 or 10 percent of the minimum free energy, many of which are similar. For this reason, a distance function was developed as a way of measuring topological differences between two structures. The distance between two foldings is the smallest whole number d such that for every base pair r_i-r_j of one, there is a base pair r_h-r_k of the other satisfying $|i - h| \leq d$ and $|j - k| \leq d$. This dimensionless quantity is zero if and only if the two structures are identical (*12*). The automatic procedure to generate P-optimal foldings can be adjusted so that the distances between all pairs of computed structures are greater than a preassigned d. This procedure decreases the number of computed structures. Setting $d = 0$ rules out one of the two nearly identical foldings in Fig. 1.

Folding of a viroid. The 359-base potato spindle tuber viroid (PSTV) was folded with the circular version of the new program. The viroid is known to fold into a long rodlike structure (*15*), as predicted by the Zuker-Stiegler algorithm (*11*). The energy dot plot of this structure is a jagged diagonal of points (base pairs) comprising the helices of the rodlike folding, which are interrupted by single-stranded regions (Fig. 2A). We call this collection of points the rod. An examination of alternative structures shows that the optimal rodlike folding is well determined (Fig. 2, B and C). Within 10 percent of the minimum free energy, there will be deviations, but these would be minor perturbations of the basic rodlike structure. The rod gradually thickens as the degree of suboptimality is increased (Fig. 2, B and C). Points near the rod correspond to base pairs that migrate lengthwise along the structure. Points close to the diagonal (upper left to lower right) correspond to small hairpin

structures that have been "pinched out" from the rodlike folding (Fig. 3, A and B).

The computed folding of PSTV is well determined in a mathematical sense. Significant deviations from the optimal folding are not observed within 5 or 10 percent of the minimum free energy. A significantly different structure does not emerge until 20-optimal foldings are computed, when a highly branched folding totally different from the rod is found. These results for PSTV, in which small changes in the energy parameters are unlikely to perturb significantly the predicted folding, are not typical.

Folding of a class I intron. The folding analysis of the self-splicing intervening sequence (IVS) from the 26*S* ribosomal RNA of *Tetrahymena thermophila* (*22, 23*) is more complicated than that for PSTV. It is similar to the analysis by Jacobson *et al.* (*18*) using a prototype version of the multiple folding program with the old energy rules. This work assessed bacteriophage folding predictions in regions of structural stability as indicated by electron microscopy. At first glance, the optimal folding drawn in Fig. 4 seems quite different from the computed folding in Cech *et al.* (*24*), but this difference is mostly due to the absence of some long-range base pairs in the former. These computer-generated foldings differ because the energy rules have been updated. The folding in Fig. 4 is closer to the more recent model (*25, 26*) based on phylogeny and biochemical evidence, containing 89 out of its 121 base pairs (74 percent).

Sequence comparison with homologous introns from five closely related *Tetrahymena* species (*27*) reveals that hairpin structures homologous to regions A, B, and C (Fig. 4) can form in all of the compared species, even though sequence heterogeneity is greatest in these regions. In addition, there are two pairs of sequence elements and a double-stranded region that are conserved in a number of nuclear and mitochondrial introns (*24*). Using the nomenclature of Burke *et al.* (*26*), the first pair of elements is *P* (U^{106} to G^{111}) and Q (C^{207} to A^{213}). These form an imperfect helix, P4, with A^{209} bulging out. Elements R (G^{263} to A^{267}) and S (U^{306} to C^{310}) form a helix, denoted by P7. The conserved double-stranded region occurs between G^{95} to A^{102} and U^{270} to C^{277} and is called P3. Both P4 and P7 are found in the optimal folding, whereas P3 is not (Fig. 4). All three of these double-stranded regions are in the model of Burke *et al.* (*26*). However, the presence of P3, P4, and P7 in a secondary structure creates a pseudo-knot (*28, 29*), and these are excluded by all energy-minimizing secondary structure prediction programs in use today (*30*). Including pseudo-knots in structure prediction algorithms would require a revised model, a new set of energy rules, and a much more complicated and slower algorithm. By finding two out of three conserved regions, the new program did as well as it could. It was also successful in finding the three regions A, B, and C in an optimal folding.

The program detects two optimal foldings: the one shown in Fig. 4, and an almost identical folding in which the U^{129}-A^{191} base pair is

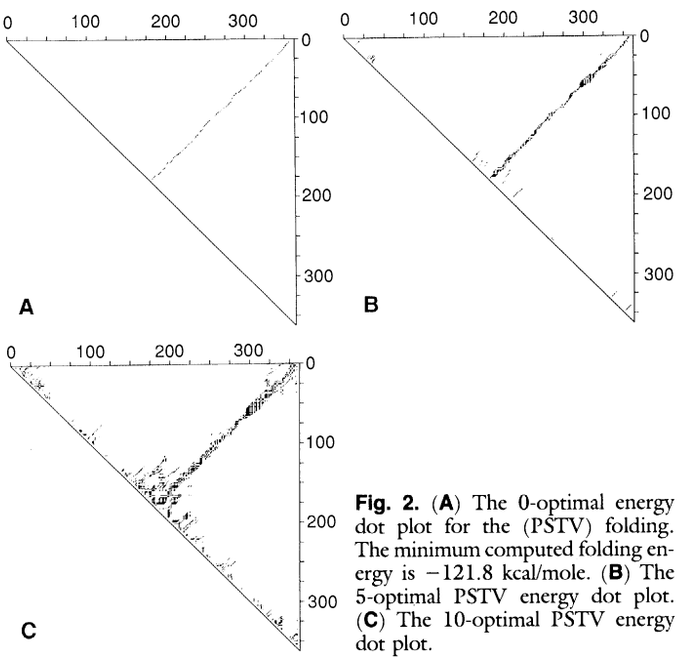

Fig. 2. (**A**) The 0-optimal energy dot plot for the (PSTV) folding. The minimum computed folding energy is −121.8 kcal/mole. (**B**) The 5-optimal PSTV energy dot plot. (**C**) The 10-optimal PSTV energy dot plot.

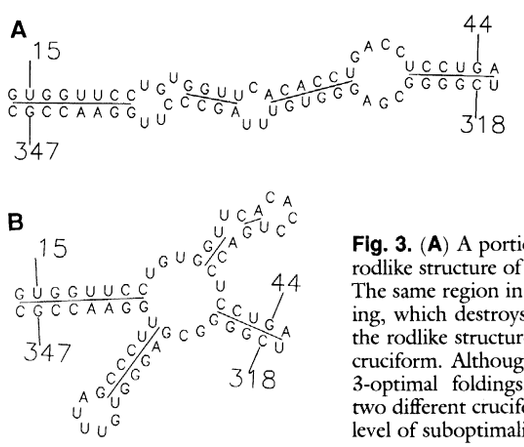

Fig. 3. (**A**) A portion of the optimal rodlike structure of PSTV RNA. (**B**) The same region in a 3-optimal folding, which destroys 12 base pairs of the rodlike structure to create a local cruciform. Although there are many 3-optimal foldings, there are only two different cruciform motifs at this level of suboptimality.

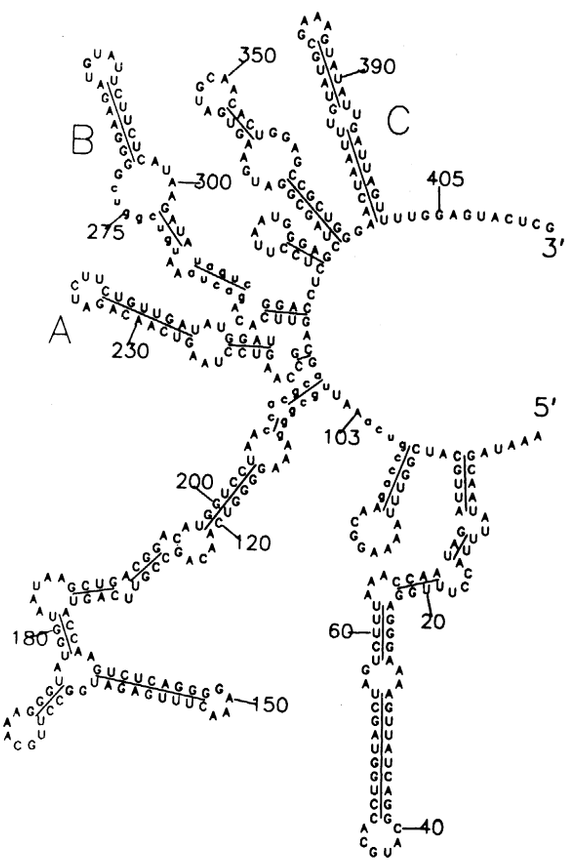

Fig. 4. A minimum-energy folding (−106.1 kcal/mole) of the 413-base self-excising *Tetrahymena* IVS. The folding required 4.25 hours of CPU time on a VAX 11/750 computer. Regions A (G^{226} to U^{246}), B (G^{279} to C^{297}), and C (A^{368} to U^{401}) are local hairpin structures for which there is phylogenetic evidence. Three pairs of (almost) complementary segments whose base pairing is conserved in a number of related nuclear and mitochondrial introns are shown in lower case.

replaced by U^{130}-A^{191} and in which U^{129} is single-stranded. The distance criterion introduced earlier was designed to eliminate the prediction of two such close structures. Thirty separate runs were made with the automatic feature to select foldings (Table 1). It is remarkable that so many trivially different 10-optimal foldings are found. When the distance between these foldings is forced to be greater than 2, the number falls dramatically. The 96 10-optimal foldings with $d = 10$ were examined in some detail. All of the structural motifs in the model of Burke *et al.* (*26*) occur in this collection, as do the structural elements contained in the model of Cech *et al.* (*24*). The P3 region is found without the two base pairs after the U-U mismatch. This entire motif appears only when $d \leq 2$. The reason is that those two base pairs are energetically unfavorable even when the rest of the motif forms. The entire P3 motif occurs in an 8.2-optimal structure (Fig. 5).

In the 5- and 10-optimal energy dot plots for the IVS (Fig. 6, B and C), the added lines create three triangular regions above the diagonal, corresponding to base pairs within the segments from 1 to 105, 106 to 213, and 214 to 413. In the 5-optimal plot, there are very few dots outside these triangular regions, which means that, within 5 percent of the minimum energy, base pairing between the three segments is unlikely. Alternative structures most likely occur from alternative foldings within these segments. The third and largest triangle is the most cluttered, implying that the greatest variability is in the last segment. In the 10-optimal dot plot, the number of possible long-range base pairs is considerable. However, the rectangles above and to the right of the middle triangle contain relatively few dots, which means that the segment from nucleotides 106 to 213 is likely to base pair only with itself in 10-optimal foldings. The growth of dots in the middle triangle from 0- to 5- to 10-optimal suggests a blurring of the branched motif formed by bases 106 to 213 (Fig. 4). The conclusion is that this branched motif is well determined and is likely to occur in 10-optimal structures. Nevertheless, it can partially disappear even within 5 percent of the minimum energy. The 5-optimal dot plot (Fig. 6B) contains three consecutive helices that intrude on the rectangle to the right of the middle triangular region. Selecting a base pair (such as G^{109}-U^{321}) in the region results in a 4.6-optimal folding that eliminates 13 base pairs in the stem region of the branched motif (Fig. 7). Similar analyses of the energy dot plots for the IVS show that motif A (Fig. 4) is also well determined, as is the hairpin on A^{30} to U^{55}. In contrast, bases 75 to 105 of the IVS can participate in many alternative structures within 10 percent of the minimum energy.

This qualitative image analysis can be made more precise by introducing a new kind of plot. If r_i is the ith ribonucleotide in a sequence, then P-Num(i) can be defined as the total number of different base pairs in which r_i can participate in all P-optimal foldings. Thus P-Num(i) is the number of points in the ith row and column of the P-optimal energy dot plot. In plots of 5-Num and 10-Num for the IVS (Fig. 8), the 10-Num plot forms a trough in the region of the branched motif (U^{106} to A^{213}), indicating a relatively well-defined structure. The average value of 10-Num is 15.9 for this segment. However, the P-Q base pairing at the base of this region is not well defined, with average 10-Num values of 44.5 and 29.0 for P and Q, respectively. At the 10-percent-level of suboptimality, P and Q can take part in numerous alternative foldings. Thus a study of the 10-Num plot leads to the more conservative prediction that only the middle portion of the branched motif (bases 115 to 204) is well defined. The average value of 10-Num for the segment from 75 to 105 is high (41.1), confirming the earlier observation based on a visual inspection of the 10-optimal dot plot. The best defined regions are the A^{30} to U^{55} hairpin and the A motif (bases 226 to 246), with 10-Num averages of 4.8 and 6.2, respectively. At the 5 percent level, more precise statements can be made. There are 20 bases that are always single-stranded and 42 base pairs that always

Table 1. The number of foldings computed for the *Tetrahymena* IVS at different percentages from the minimum folding energy P for various minimum pairwise-distance criteria d.

P	$d = 0$	$d = 2$	$d = 5$	$d = 10$	$d = 20$	$d = 50$
0	2	1	1	1	1	1
1	17	4	2	2	2	1
2	40	10	5	4	4	2
5	325	91	36	21	14	4
10	3140	677	230	96	39	9

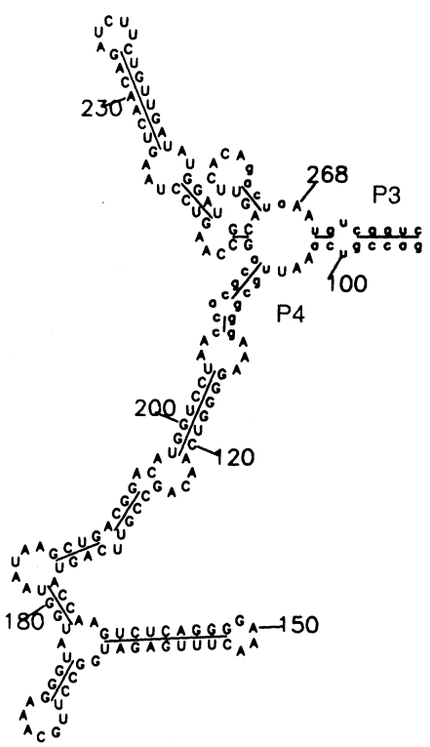

Fig. 5. Part of a suboptimal folding of the IVS containing the entire P3 structural motif not found in the optimal structure (Fig. 4). The P4 region is retained in this folding while P7 is lost. The bases of these structural features are shown in lower case.

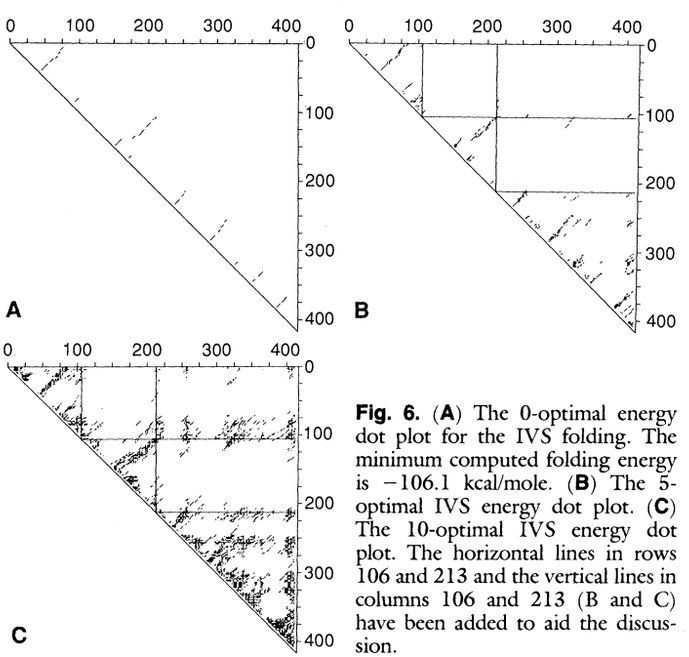

Fig. 6. (**A**) The 0-optimal energy dot plot for the IVS folding. The minimum computed folding energy is −106.1 kcal/mole. (**B**) The 5-optimal IVS energy dot plot. (**C**) The 10-optimal IVS energy dot plot. The horizontal lines in rows 106 and 213 and the vertical lines in columns 106 and 213 (B and C) have been added to aid the discussion.

occur in 5-optimal structures. In particular, the hairpin G^{31} to C^{54} and the A motif without the bottom two base pairs must always form in 5-optimal structures.

Suboptimal foldings versus dot plot analysis. The automatic-traceback procedure is not intended to generate all nearly optimal foldings. Even with the distance constraint, there can be too many structures to examine within 10 or even 5 percent of the minimum energy. For example, within 5 percent of the minimum energy, the three segments 1 to 105, 106 to 213, and 214 to 413 of the IVS fold more or less independently of one another (Fig. 6B). Selecting a suboptimal base pair in one of the segments produces a suboptimal structure in that segment, whereas the folding in the rest of the sequence is optimal. If ten suboptimal base pairs are chosen in each triangular region, a total of 30 structures would be generated by the existing program. However, it is possible to combine each suboptimal folding in each segment with every other suboptimal folding found in the other two segments. This procedure yields $10 \times 10 \times 10 = 1000$ suboptimal structures. In general, this sort of combinatorial argument can be used to increase the output of the program by many orders of magnitude. Thus the algorithm was designed to find the best structures containing single given base pairs, instead of proceeding to compute structures containing two or more prescribed base pairs.

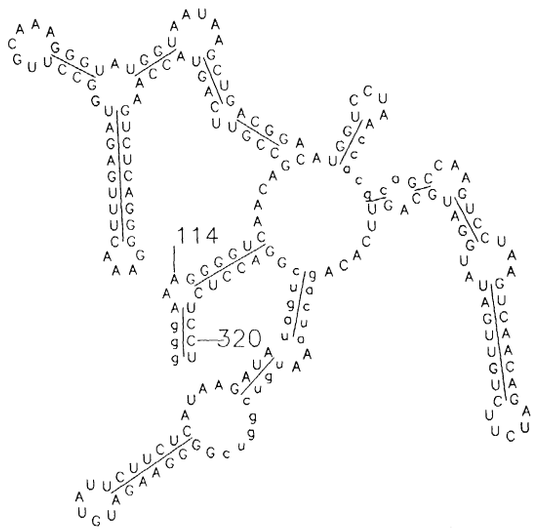

Fig. 7. Part of a 5-optimal folding in which the branched structural motif from Fig. 4 (bases U^{106} to A^{213}) is partly lost. This folding also contains long-range base pairs not found in the optimal structure.

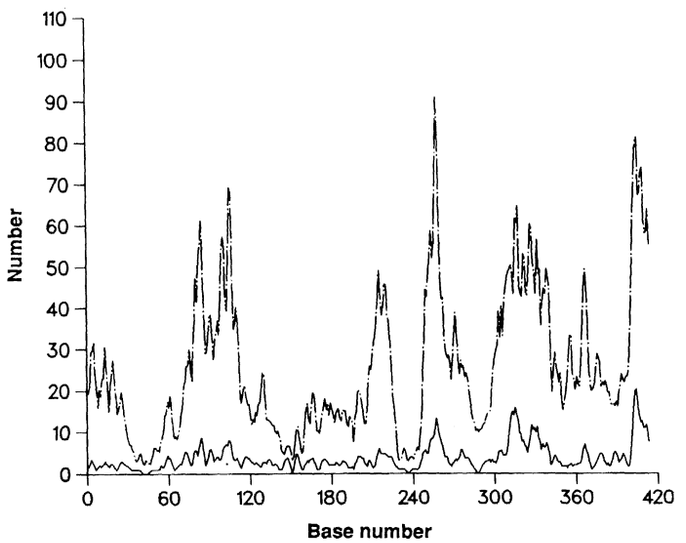

Fig. 8. The 5- and 10-Num plots for the IVS shown in solid and dashed lines, respectively. The total number of base pairs in which the ith base can take part in 5- or 10-optimal foldings (ordinate) is plotted against i (abscissa). Plotted ordinates are the averages over three consecutive bases.

The selection of a single suboptimal base pair usually results in the discovery of a novel local folding motif including that base pair. The rest of the folding often contains base pairs that have been found in previous foldings. Occasionally, selecting a base pair produces a folding that is different from previous structures not only near the selected base pair, but farther away as well. At the very least, the procedure of selecting P-optimal base pairs not too close to base pairs that have already occurred in a folding should yield all possible local motifs that can take part in P-optimal foldings. At the 10-optimal level, this procedure predicts 96 percent of phylogenetically determined helices in 141 transfer RNA sequences and 88 percent of the corresponding helices for 67 5*S* RNA sequences (*31*).

The analysis of the energy dot plot and the derived P-Num function is an effective way of viewing and appreciating the entire range of solutions within a given percentage of the minimum folding energy. This approach makes it possible to assign a confidence to a secondary structure, or to decide that it is very unlikely that one segment base pairs with another. The program already allows for the incorporation of nuclease data indicating single- or double-stranded regions, so that only base pairs compatible with such data would be viewed in the dot plot. An automated procedure to compare energy dot plots of two or more homologous sequences in the search for a common folding that would combine energy minimization and phylogeny remains to be developed.

REFERENCES AND NOTES

1. J. R. Fresco, B. M. Alberts, P. Doty, *Nature* **188**, 98 (1960).
2. S. H. Kim *et al.*, *Science* **185**, 435 (1974).
3. M. Zuker, *Lect. Math. Life Sci.* **17**, 87 (1986).
4. J. M. Pipas and J. E. McMahon, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2017 (1975).
5. J.-P. Dumas and J. Ninio, *Nucleic Acids Res.* **10**, 197 (1982).
6. C. Papanicolaou, M. Gouy, J. Ninio, *ibid.* **12**, 31 (1984).
7. M. Gouy, P. Marliere, C. Papanicolaou, J. Ninio, *Biochemie* **67**, 523 (1985).
8. P. R. Stein and M. S. Waterman, *Discrete Math.* **26**, 261 (1978).
9. R. Nussinov, G. Pieczenik, J. R. Griggs, D. J. Kleitman, *SIAM (Soci. Ind. Appl. Math.) J. Appl. Math.* **35**, 68 (1978).
10. R. Nussinov and A. B. Jacobson, *Proc. Natl. Acad. Sci. U.S.A.* 77, 6309 (1980).
11. M. Zuker and P. Stiegler, *Nucleic Acids Res.* **9**, 133 (1981).
12. M. Zuker, in *Mathematical Methods for DNA Sequences*, M.S. Waterman, Ed. (CRC Press, Boca Raton, FL, 1989), pp. 159–184.
13. A. L. Williams, Jr., and I. Tinoco, Jr., *Nucleic Acids Res.* **14**, 299 (1986).
14. D. Sankoff, J. B. Kruskal, S. Mainville, R. J. Cedergren, in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, D. Sankoff and J. B. Kruskal, Eds. (Addison-Wesley, Reading, MA, 1983), p. 93.
15. D. Riesner *et al.*, *J. Biomol. Struct. Dyn.* **1**, 669 (1983).
16. G. Steger *et al.*, *ibid.* **2**, 543 (1984).
17. M. Zuker and D. Sankoff, *Bull. Math. Biol.* **46**, 591 (1984).
18. A. B. Jacobson, M. Zuker, A. Hirashima, in *Molecular Biology of RNA: New Perspectives*, M. Inouye and B. S. Dudock, Eds. (Academic Press, New York, 1987), p. 331.
19. S. M. Freier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9373 (1986).
20. D. H. Turner *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 123 (1987).
21. W. Salser, *ibid.* **42**, 985 (1977).
22. T. R. Cech, A. J. Zaug, P. J. Grabowski, *Cell* **27**, 487 (1981).
23. K. Kruger *et al.*, *ibid.* **31**, 147 (1982).
24. T. R. Cech *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3903 (1983).
25. R. B. Waring, P. Towner, S. J. Minter, R. W. Davies, *Nature* **321**, 133 (1986).
26. J. M. Burke *et al.*, *Nucleic Acids Res.* **15**, 7217 (1987).
27. H. Nielson and J. Engberg, *ibid.* **13**, 7445 (1985).
28. C. W. A. Pleij, R. Krijn, L. Bosch, *ibid.*, p. 1717.
29. J. D. Puglisi, J. R. Wyatt, I. Tinoco, Jr., *Nature* **331**, 283 (1988).
30. M. Gouy, in *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*, M. J. Bishop and C. J. Rawlings, Eds. (IRL Press, Oxford, 1987), p. 259.
31. J. A. Jaeger, D. H. Turner, M. Zuker, unpublished results.

32. I thank A. Jacobson for encouragement and discussions during the development of the algorithm and for thinking of the term "energy dot plot"; D. Turner and J. Jaeger for help in testing the new program and adapting the latest energy rules; and R. Somorjai for suggestions during the preparation of the manuscript. E. Nelson performed much of the programming work. The program is available from the author upon written request. This is NRCC publication No. 30147.

4 November 1988; accepted 1 March 1989