Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 7706–7710, October 1989
Biochemistry

Improved predictions of secondary structures for RNA

JOHN A. JAEGER*, DOUGLAS H. TURNER*†, AND MICHAEL ZUKER‡

*Department of Chemistry, University of Rochester, Rochester, NY 14627; and ‡Division of Biological Sciences, National Research Council of Canada, Ottawa, ON K1A OR6, Canada

Communicated by I. Tinoco, Jr., July 3, 1989

ABSTRACT **The accuracy of computer predictions of RNA secondary structure from sequence data and free energy parameters has been increased to roughly 70%. Performance is judged by comparison with structures known from phylogenetic analysis. The algorithm also generates suboptimal structures. On average, the best structure within 10% of the lowest free energy contains roughly 90% of phylogenetically known helixes. The algorithm does not include tertiary interactions or pseudoknots and employs a crude model for single-stranded regions. The only favorable interactions are base pairing and stacking of terminal unpaired nucleotides at the ends of helixes. The excellent performance is consistent with these interactions being the primary interactions determining RNA secondary structure.**

RNA is important for functions such as catalysis, RNA splicing, regulation of transcription and translation, and transport of proteins across membranes (1). Many RNA sequences are known. Determination of secondary structures, however, is difficult. Thermodynamics has been applied to predict RNA secondary structure from sequence (2, 3), but with modest success. Predictions of suboptimal structures make the method more useful (4). We report combining several recent advances to improve predictions of RNA secondary structures.

Three advances are incorporated in this work. (*i*) New methods for synthesizing RNA make it possible to obtain model systems with a large variety of sequence (5, 6). This technique has led to measurements of improved parameters and the realization that non-base-paired nucleotides contribute sequence-dependent interactions that stabilize secondary structure (7, 8). (*ii*) A computer algorithm has been developed that allows incorporation of non-base-paired interactions in the prediction of optimal and suboptimal secondary structures (9). (*iii*) Several RNA secondary structures have been determined by phylogeny (10–15). Comparison of predicted and known structures allows optimization of parameters that have not been measured. Resultant predictions appear sufficiently reliable to aid planning and interpretation of experiments on RNA.

MATERIALS AND METHODS

Thermodynamic Parameters. When possible, free energy increments at 37°C, ΔG°_{37}, were taken from experiments in 1 M NaCl. For fully base-paired regions, experiments on dGCATGC indicate that 1 M NaCl mimics solutions containing 1–100 mM Mg^{2+} in the presence of 0.15–1 M NaCl (16).

Relatively few experiments are available for loop structures, and, therefore, little is known about interactions determining loop stability. This situation forced approximations. (*i*) Jacobson–Stockmayer theory (17) was used to extrapolate the length dependence of ΔG°_{37} for bulge, hairpin, and internal loops: $\Delta G^\circ(n) = \Delta G^\circ(n_{max}) + 1.75\,RT \ln(n/n_{max})$. For this equation n is the number of unpaired nucleotides in the loop, n_{max} is the maximum-length loop for which experimental data is available, R is the gas constant (1.987 cal·mol^{-1}·K^{-1}; 1 cal = 4.184 J), and T is the temperature in K (310.15 K for 37°C). (*ii*) When experimental data were available for loops of length n and $n + 2$, but not $n + 1$, ΔG° for $n + 1$ was obtained by interpolation. (*iii*) The sequence dependence of hairpin and internal-loop stability was approximated by assuming that for most such loops the sequence dependence arises from stacking interactions of the first mismatch on the adjacent base pair. These interactions were approximated by ΔG°_{37} values measured for terminal unpaired nucleotides and mismatches on oligonucleotide duplexes (8). This approximation is suggested by the correlation seen between these ΔG° values and the three-dimensional structure of yeast phenylalanine tRNA (8). Details of various approximations and parameters are described below. Although the treatment of loops is crude, the program can incorporate future refinements.

Watson–Crick and GU Pairs. Free energy increments for hydrogen-bonded AU, GC, and GU pairs are based on the nearest-neighbor model (5, 7, 8, 18). Measured free energy increments for terminal hydrogen-bonded GU pairs are an average of 0.2 kcal/mol more favorable than internal GU mismatches (19). This information has not been included in the model.

Terminal Unpaired Nucleotides and Mismatches. Parameters for terminal unpaired nucleotides (dangling ends) and terminal mismatches are those collected by Turner *et al.* (8). AU, GC, and GU pairs at the ends of helixes are allowed not to be hydrogen bonded because in some cases stacking without pairing gives a more favorable ΔG°. Increments for terminal nonhydrogen-bonded AU, GC, and GU pairs are approximated by the corresponding 3′-dangling end made more favorable by 0.2 kcal/mol to approximate the effect of the opposing 5′-dangling end (20). Only one measurement is available for ΔG°_{37} associated with a dangling end adjacent to a terminal GU mismatch (19). For a 3′-dangling adenosine phosphate in the sequence $\begin{smallmatrix}UA\\G\end{smallmatrix}$, the free energy increment was roughly the average of increments measured for 3′-dangling ends adjacent to the corresponding AU and GC pairs. Thus dangling-end increments for $\begin{smallmatrix}UN\\G\end{smallmatrix}$ sequences are approximated with the average measured for $\begin{smallmatrix}UN\\A\end{smallmatrix}$ and $\begin{smallmatrix}CN\\G\end{smallmatrix}$ sequences. The sequence $\begin{smallmatrix}GN\\U\end{smallmatrix}$ was approximated as $\begin{smallmatrix}AN\\U\end{smallmatrix}$. The 5′-dangling ends, $\begin{smallmatrix}NU\\G\end{smallmatrix}$ and $\begin{smallmatrix}NG\\U\end{smallmatrix}$, are given ΔG°_{37} values of −0.2 kcal/mol, corresponding to the average measured for other 5′-dangling ends (8). Terminal mismatches adjacent to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ΔG°_{37}, free energy increments at 37°C; nt, nucleotide; *S*-num, number of times a nucleotide is unpaired; *P*-num, number of different base pairs formed by a base in structures.

†To whom reprint requests should be addressed.

GU pairs are treated in a similar way. Thus $\begin{matrix}UN\\GY\end{matrix}$ is approximated as the average of $\begin{matrix}UN\\AY\end{matrix}$ and $\begin{matrix}CN\\GY\end{matrix}$. Analogously, $\begin{matrix}GN\\UY\end{matrix}$ is approximated as $\begin{matrix}AN\\UY\end{matrix}$.

Bulge Loops. Because little is known about sequence dependence of bulge-loop stability, these motifs are approximated with a nearest-neighbor model. This oversimplified model has three assumptions based on enthalpy increments measured for bulges of 1, 2, and 3 nucleotides (nt) (21). (*i*) Stability of a bulge is independent of sequence in the loop. (*ii*) Base pairs adjacent to a single nucleotide bulge stack so that the ΔG°_{37} for this nearest-neighbor interaction is retained. (*iii*) Bulges with >1 nt prevent stacking of the adjacent base pairs so this potential nearest-neighbor ΔG°_{37} is omitted. When ΔG°_{37} values were available for oligonucleotide sequences with and without dangling ends, they were simply averaged to fit the nearest-neighbor model. Parameters for bulges are given in Table 1 (21–23).

Hairpin Loops. The smallest allowed hairpin has three unpaired nucleotides (3). For loops >3 nucleotides, the first nucleotide on both the 3′ and 5′ sides of the loop is assumed to form a mismatch with the same stability as the corresponding mismatch on an oligonucleotide duplex. Free energies of loop formation were calculated from published melting temperatures (T_m) (24–27). In these calculations, ΔH° and ΔS° for the stem and for the first unpaired nucleotide on the 5′ side of the loop (a 3′-dangling end) are approximated with parameters measured for base pairs and 3′-dangling ends in oligonucleotide duplexes (8). The first unpaired nucleotide on the 3′ side of the loop is assumed to have $\Delta H^\circ = 0$ kcal/mol and $\Delta S^\circ = 0.6$ cal·mol^{-1}·K^{-1} corresponding to a 5′-dangling end on an oligonucleotide duplex. Loop closure is assumed to have $\Delta H^\circ = 0$ kcal/mol. With these assumptions, ΔS° and ΔG°_{37} for loop closure are calculated from the following equation, where the subscript de represents dangling end:

$$\Delta S^\circ_{loop} = [\Delta H^\circ_{stem} + \Delta H^\circ_{3'de} - T_m(\Delta S^\circ_{stem} + \Delta S^\circ_{3'de} + 0.6)]/T_m \quad [1]$$

$$\Delta G^\circ_{37,\,loop} = 310.15\,\Delta S^\circ_{loop} \quad [2]$$

Values for ΔG°_{37} of loop closure are listed in Table 1.

Table 1. Free energy parameters for loops

Size, nt	Internal	Bulge	Hairpin
1		+3.9	
2	+4.1	+3.1	
3	+4.5	+3.5	+4.5
4	+4.9	(+4.2)	+5.5
5	(+5.3)	+4.8	+4.9
6	+5.7	(+5.0)	(+5.1)
7	(+5.9)	(+5.2)	+5.2
8	(+6.0)	(+5.3)	(+5.5)
9	(+6.1)	(+5.4)	+5.8
10	(+6.3)	(+5.5)	(+5.9)

Free energy parameters in kcal/mol for RNA loops at 37°C in 1 M NaCl. For larger loop sizes, n, use $\Delta G^\circ(n) = \Delta G^\circ(n_{max}) + 1.75\,RT \ln(n/n_{max})$, where n_{max} is 6, 5, and 9 for internal, bulge, and hairpin loops, respectively. Parameters not derived from experimental measurements are listed in parentheses. Parameters for hairpin loops >3 nt and for internal loops assume additional stability is conferred by stacking of terminal mismatches at helix ends. Asymmetric internal loops with branches of N_1 and N_2 nt must be penalized additionally by the minimum of 3.0 or $N \times f(M)$ kcal/mol, where $N = |N_1 - N_2|$, M is the minimum of 4, N_1, or N_2, and $f(1) = 0.4$, $f(2) = 0.3$, $f(3) = 0.2$, $f(4) = 0.1$. The parameter for bulge loops of one only assumes additional stability is conferred by stacking of the adjacent base pairs as approximated by nearest-neighbor parameters (7, 8). No stacking across bulges of 2 or more nucleotides is assumed.

Recently, Tuerk *et al.* (28) reported that UUCG forms an unusually stable hairpin loop. J. Haney and O. C. Uhlenbeck (personal communication) find that GAAA behaves similarly. They have also noticed that these and six other sequences (GCAA, GAGA, GUGA, GGAA, UACG, and GCGA) account for >60% of tetraloops in a set of 16S and 23S rRNAs (29). Appropriate ΔG°_{37} values for these sequences are not known. Nevertheless, they are clearly more stable than other tetraloops. To test the effect of such enhanced stability on structure prediction, hairpins of 4 nt containing the eight most prevalent tetraloop sequences were each made more stable by 2 kcal/mol; this increment corresponds to a single strong hydrogen bond or stacking interaction (8, 30, 31).

Internal Loops. The first mismatch on each side of an internal loop is assumed to provide a favorable ΔG°_{37} equal to that provided by a terminal mismatch on an oligonucleotide duplex. All internal loops, including those containing 2 nt (single mismatches) have two such free energy increments. This assumption is consistent with the observation that the favorable free energy of a terminal mismatch in oligonucleotide duplexes comes primarily from 3′-terminal stacking (19). With these assumptions, ΔG°_{37} values for internal loops were calculated from data on $A_4GC_nU_4$ (32), A_7CU_7 (33), and $CGCA_nGCG$ and $CGCA_nCGC$ plus $GCGA_mGCG$ (N. Sugimoto, R. Kierzek, C. E. Longfellow, and D.H.T., unpublished experiments). For example, from published plots of T_m^{-1} vs. oligomer concentration, the measured difference in stabilities between A_7CU_7 and A_7U_7 is +4.3 kcal/mol (33). This difference is assumed to include the loss of two $\begin{matrix}AU\\U\end{matrix}$ stacking interactions ($\Delta G^\circ_{37} = -0.6$ kcal/mol for each) from the central $\begin{matrix}AU\\UA\end{matrix}$ nearest neighbor in A_7U_7 and the gain of two $\begin{matrix}AC\\U\end{matrix}$ ($\Delta G^\circ_{37} = -0.5$ kcal/mol for each) and two $\begin{matrix}pU\\A\end{matrix}$ ($G^\circ_{37} = -0.2$ kcal/mol for each) interactions. Thus the calculated ΔG°_{37} for the C_2 internal loop is as follows: $4.3 + 2(-0.6) - 2(-0.5) - 2(-0.2) = 4.5$ kcal/mol. Table 1 lists average ΔG°_{37} values for internal loops.

Papanicolaou *et al.* (34) suggested asymmetric internal loops may be less stable than symmetric internal loops. Preliminary experiments (A. Peritz, R. Kierzek, and D.H.T., unpublished experiments) indicate penalties for asymmetric internal loops are about half those suggested (34). These penalties are described in the legend to Table 1 and must be added to the ΔG° values in Table 1.

Multibranched Loops. Unpaired nucleotides adjacent to helixes in multibranched loops are assumed to act like dangling ends, and the favorable free energy increments are set equal to those determined from studies on oligonucleotide duplexes (8). If a nucleotide can stack on either of two helixes, the stacking with the most favorable ΔG° is chosen.

For speed of computation, a linear approximation was used for the length dependence of ΔG° for a multibranched loop. The functional form is as follows: $\Delta G^\circ_{37} = a + bn + ch$, where a, b, and c are adjustable parameters, n is the number of unpaired nucleotides in the loop, and h is the number of helixes in the loop. Best results were obtained with $a = 4.6$, $b = 0.4$, and $c = 0.1$.

Computational Methods. The algorithm (9) is an extension of the dynamic programming algorithm of Zuker and Stiegler (35), with addition of sequence dependence for single-stranded regions. Pseudoknots (36) are not allowed. The program finds the lowest free energy structure and a representative set of suboptimal structures within any desired range of higher free energies. The program is written in FORTRAN and runs in a VAX/VMS environment. Directions for its use are described elsewhere (37).

Optimization of Unmeasured Parameters. Many factors affecting nucleic acid structure are poorly understood. When parameters were not available or when there was a choice among approximations, different combinations were tested by predicting secondary structures previously deduced from phylogenetic evidence. The test set of secondary structures contained tRNAs (38), 5S rRNAs (10), four domains of *Escherichia coli* 16S rRNA (39), and the self-splicing group I LSU intron from *Tetrahymena thermophila* (12, 13). In all cases, data suggest the phylogenetic structure forms in the absence of proteins. Predicted structures were compared with known secondary structure helixes as described in the legend for Table 2.

RESULTS

Performance on Optimized Structures. On average, the lowest free energy or optimal structure has 73% of the phylogenetically deduced helixes when the minimal parameter set is used, and 77% when special tetraloop sequences are made more favorable by 2 kcal/mol (Table 2). Short-range helixes containing at least 1 bp between nucleotides separated by 30 or fewer nucleotides are predicted with roughly the same percentages, 70 and 72%, without and with special tetraloops, respectively. Similar percentages are also found when base pairs are compared directly. For example, with special tetraloop sequences included in the predictions for 16S rRNA, 63 and 67%, respectively, of phylogenetically deduced helixes and base pairs are found in the optimal structure. For the LSU intron, the values are 88 and 89%, respectively. The predictions are considerably better than those from the program of Zuker and Stiegler (35) with parameters collected by Salser (43). Improvement is even greater than indicated in Table 2 because isolated base pairs were previously forbidden but are now allowed.

Similar performance for optimal structures is obtained when stacking increments in internal and hairpin loops are replaced by penalties of 0.9 kcal/mol for each closing AU pair. Thus, the detailed sequence dependence for stacking is not required for the performance listed in Table 2.

The folding algorithm can be constrained by forcing certain nucleotides to be single stranded based on chemical modification data. To see whether this had any effect, spinach 5S and *E. coli* 16S rRNAs were folded with strongly modified nucleotides constrained (39, 44). For the 5S and 16S rRNAs, the predictions improved from 4 to 5 and 41 to 42 helixes

Table 2. Comparison of structures deduced from computer prediction and phylogenetic data

			Phylogenetic helixes predicted*, %					
			Without tetraloops			With tetraloops		
				Suboptimal			Suboptimal	
RNA	Helix, no.	Salser†	Optimal	Within 5%	Within 10%	Optimal	Within 5%	Within 10%
				Optimized				
tRNA	553	69	90	93	96	90	94	96
5S rRNA	335	45	60	77	84	66	79	88
16S rRNA								
Domain 1	24	8	54	83	88	63	100	100
Domain 2	15	53	67	93	100	67	93	100
Domain 3	22	14	50	77	86	59	91	96
Domain 4	4	0	75	75	75	75	75	75
Total	65	20	55	83	89	63	94	97
LSU	17	59	88	100	100	88	100	100
Average		48	73	88	92	77	92	95
				Nonoptimized 16S-like RNA				
Rat mitochondria	35	37	40	54	66	40	54	66
H. volcanii	62	50	80	95	98	79	97	98
C. r. chloroplast	65	15	43	80	88	40	83	91
				Introns				
Yeast OX5$^{\alpha}$ (I)	11	55	82	91	100	82	91	100
ND1 (I)	23	35	22	61	65	39	61	70
T4D (I)	12	42	83	92	92	83	92	100
Yeast A1 (II)	32	47	66	81	91	69	81	91
Yeast A5 (II)	30	50	77	97	100	80	97	100
Yeast B2 (II)	31	48	55	84	84	68	84	94
Average		42	61	82	87	64	82	90

Secondary structure was predicted with the program of Zuker (9). A helix is defined as a region containing 3 or more bp with no bulge or interior loops containing 3 or more nucleotides. A helix is considered correct when it contains all but 1 or 2 bp deduced phylogenetically. Pseudoknots were not considered because the program cannot predict them. For tRNA and 5S rRNA, structure predictions were made on 141 and 67 randomly chosen sequences and compared with phylogenetic models (10, 38). For tRNA, modified nucleotides unable to base pair were not allowed to pair. *E. coli* 16S rRNA was divided into four major domains (26–557, 561–913, 913–1397, and 1397–1542) and compared with the "naked" 16S rRNA structure (39). The self-splicing group I LSU intron in *Tetrahymena thermophila* (LSU) is composed of nt −10 to −1 and 2 to 424. Calculated structure (37) was compared to consensus structure (12). 16S-like rRNAs (11) were each divided into four domains [rat mitochondrial: 20–279, 279–509, 526–829, and 829–953; *Halobacterium* (*H.*) *volcanii*: 21–495, 501–857, 865–1342, and 1342–1474; and *Chlamydomonas reinhardii* (*C. r.*) chloroplast: 27–509, 515–857, 866–1326, and 1329–1476]. Self-splicing group I introns were folded with large regions of undetermined structure replaced by 3 Ns [T4 *td*, −8–1021 omitting 95–850 (40); *Podospora anserina* ND1, −7–1818 omitting 19–369 and 407–1396 (14); and *Saccharomyces cerevisiae* OX5$^{\alpha}$: −10–1375 omitting 80–943 (41)]. Group II introns (15, 42) were split into two domains (*S. cerevisiae* A1: 1–10 with 353–2450, replacing 505–2348 with 3 Ns, and 5–358; *S. cerevisiae* A5, 1–11 with 411–887, and 6–416; and *S. cerevisiae* B1, 1–11 with 380–768, and 6–385).

*Parameters from this work.

†Parameters from Ref. 43.

correct, respectively. The prediction for domain 3 of 16S rRNA, however, included five fewer correct helixes when modification data were included. One reason this occurs is that base-paired nt 951 is modified. Thus, modification data improve predictions in most, but not all, cases.

Suboptimal structures within 5 and 10% of the free energy of the optimal structure were searched to find the predicted structure closest to the phylogenetic structure. Because the program provides only a representative set of structures (9, 37), this search included different combinations of various local motifs from the representative structures. The percentages of phylogenetic helixes present in these structures are also listed in Table 2. On average, 88 and 92% of known helixes are found within 5% of the minimum free energy, without and with tetraloops; this increases to 92 and 95% within 10% of the minimum free energy. Often, a significantly better structure was found within a few kcal/mol of the optimal structure (Fig. 1).

Performance on Nonoptimized Structures. Several RNAs not included in the optimization set were considered (Table 2). With inclusion of tetraloops, on average, 64 and 90% of known helixes are predicted in the optimal and best suboptimal structures within 10%, respectively. In contrast to the RNAs used for optimization of parameters, the only evidence these RNAs fold into the phylogentic structure in the absence of proteins is the self-splicing activity of the chosen introns. Furthermore, the details of these structures are supported by less evidence than those used for optimization of parameters.

Graphical Presentation. Consideration of suboptimal structures can be aided by graphical methods. For example, regions of ambiguous pairing can be identified by plotting the number of different base pairs formed by a base in optimal and suboptimal structures, *P*-num, versus sequence position (9) (Fig. 2). Base pairing of nucleotides with a large *P*-num is predicted ambiguously. For example, this may indicate insufficient knowledge for prediction or may indicate fluxional regions. A related plot shows the number of times a nucleotide is unpaired, *S*-num, in optimal and suboptimal structures (Fig. 2). A large *S*-num indicates a region is probably single-stranded.

The correlation of low *P*-num regions with accepted structures for *E. coli* 16S rRNA and LSU intron was checked. When *P*-num values from structures within 5% of the minimum free energy (including special tetraloops) are divided by the total length of the RNA, *n*, 38 regions were found with *P*-num/*n* <0.01. Of these, 50 and 18%, respectively, correspond to conserved, and correct but not conserved helixes. A further 14% of regions with *P*-num/*n* <0.01 are correctly identified as single-stranded from *S*-num plots. Thus, only 18% of low *P*-num/*n* regions have foldings that differ from phylogenetically determined foldings. This result is a higher correlation than for the entire RNA (see 16S and LSU in Table 2).

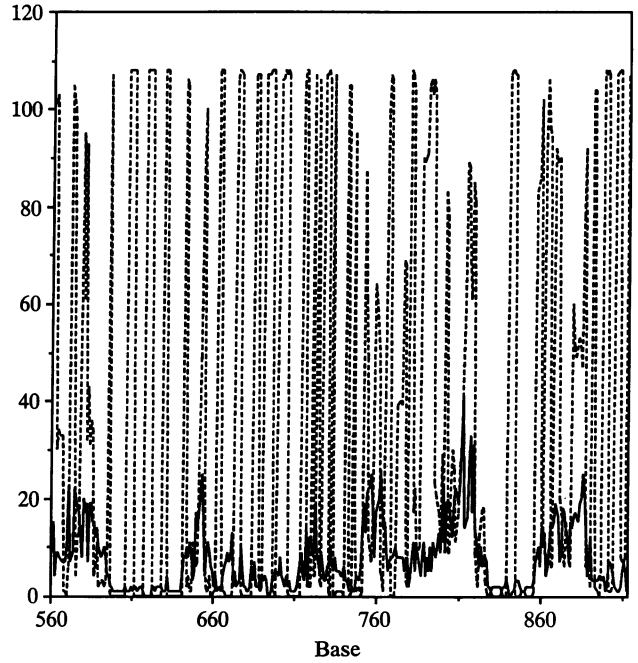

FIG. 2. *P*-num (—) and *S*-num (---) plots at 5% for domain 2 of *E. coli* 16S rRNA. *P*-num plots the number of different base pairs formed by a base in structures within 5% of the free energy minimum vs. nucleotide number of that base. *S*-num plots the number of times a nucleotide is not base paired in those structures. *P*-num and *S*-num data were smoothed by taking a 5-point running average of the data; *S*-num data were scaled by a factor of 0.256.

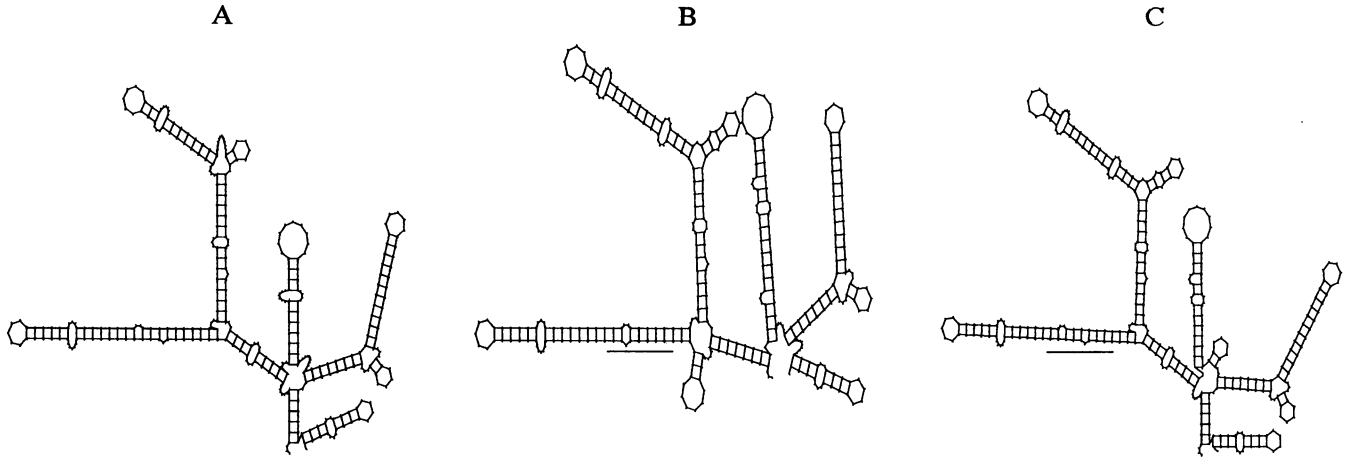

FIG. 1. *E. coli* 16S rRNA domain 2 (561–913) as deduced from phylogenetic data (*A*) (39), optimal structure prediction (*B*) (67% correct), and best suboptimal prediction within 5% of the optimal free energy (*C*) (93% correct). Structure predictions used the data set that includes a bonus of −2.0 kcal/mol for eight of the most common tetraloops. Energies for the optimal and suboptimal structures are −108.5 and −103.5 kcal/mol, respectively. Often, slipping of several base pairs as shown for helix 590–650 (underlined) improves the comparison with phylogeny. Tick marks represent unpaired nucleotides. A bulge loop is formed when there are unpaired nucleotides on only one strand of a double helix; an internal loop forms when there are unpaired nucleotides opposite each other that interrupt a double helix. A multibranched loop is formed by the intersection of more than 2 double helixes. Hairpin loops form when the strand bends on itself locally to form base pairs. For reference, the first, second, and last hairpins in each structure are closed by nucleotides 617:623, 690:697, and 897:902, respectively.

DISCUSSION

Predicted optimal structures contain on average ≈70% of known helixes for the sequences tested. The best suboptimal structures within 10% of the optimal free energy contain, on average, ≈90% of known helixes. The results are encouraging considering how little is known about sequence dependence of loop stability. For example, the structures predicted least reliably contain several multibranched loops, the motif for which no experimental data are available. Clearly, predictions are improving as information accumulates. The results suggest base pairing and possibly stacking of terminal unpaired nucleotides are the primary interactions determining secondary structure. Evidently, tertiary interactions, including base pairing in pseudoknots (36), are less important.

Although tertiary interactions cannot be incorporated in the algorithm, they can be used to refine predictions. For example, information on tertiary interactions and nonnearest-neighbor effects could be used to reorder suboptimal foldings. Because 90% of known helixes are found within 10% of the minimum free energy, such a two-stage approach may lead to further improvement in predictions.

Interactions determining nucleic acid structure are insufficiently understood to allow perfect predictions of structure. Prediction of suboptimal structures, however, provides a guide for designing experiments to determine structures empirically and should also be helpful for interpreting experiments indicating a single sequence has multiple conformations. Several studies provide hints of such conformational flexibility (45, 46). Another application may be choice of sequences for structural studies with x-ray and NMR methods. Sequences with clearly defined optimal structures will be preferred over those with fluxional character (47). Performance of predictions with various parameter sets also indicates parameters requiring further study. For example, the sequence dependence of loops must be measured to take full advantage of the program.

Although a large number of suboptimal structures are often found, graphical methods facilitate focusing on important properties. *P*-num plots indicate regions where base pairing is predicted ambiguously. *S*-num plots indicate regions that are probably single stranded. Both such regions are attractive targets for site-directed mutagenesis and other methods. Ambiguous pairing regions can be studied to refine secondary structure. Single-stranded regions can be searched for functionally important nucleotides, for tertiary interactions that determine three-dimensional structure, and for targets suitable for primers or antisense oligonucleotides. Computer-aided rational design of such experiments should speed discovery of important relationships between sequence, structure, and function.

This work was supported by National Institutes of Health Grant GM22939. Michael Zuker is a Fellow of the Canadian Institute for Advanced Research.

1. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. M. (1987) *Molecular Biology of the Gene* (Benjamin Cummings, Inc., Menlo Park, CA).
2. Tinoco, I., Jr., Uhlenbeck, O. C. & Levine, M. D. (1971) *Nature (London)* **230,** 362–367.
3. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246,** 40–41.
4. Williams, A., Jr. & Tinoco, I., Jr. (1986) *Nucleic Acids Res.* **14,** 299–315.
5. Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H. & Freier, S. M. (1986) *Biochemistry* **25,** 7840–7846.
6. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 8783–8798.
7. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
8. Turner, D. H., Sugimoto, N. & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17,** 167–192.
9. Zuker, M. (1989) *Science* **244,** 48–52.
10. Wolters, J. & Erdmann, V. A. (1988) *Nucleic Acids Res.* **16,** r1–r70.
11. Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. (1985) *Prog. Nucleic Acid Res. Mol. Biol.* **32,** 155–216.
12. Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W. & Tabak, H. F. (1987) *Nucleic Acids Res.* **15,** 7217–7221.
13. Michel, F. & Dujon, B. (1983) *EMBO J.* **2,** 33–38.
14. Michel, F. & Cummings, D. J. (1985) *Curr. Genet.* **10,** 69–79.
15. Michel, F., Umesono, K. & Ozeki, H. (1989) *Gene*, in press.
16. Williams, A. P., Longfellow, C. E., Freier, S. M., Kierzek, R. & Turner, D. H. (1989) *Biochemistry* **28,** 4283–4291.
17. Jacobson, H. & Stockmayer, W. H. (1950) *J. Chem. Phys.* **18,** 1600–1606.
18. Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol.* **86,** 843–853.
19. Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Biochemistry* **25,** 3209–3213.
20. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T. & Turner, D. H. (1985) *Biochemistry* **24,** 4533–4539.
21. Longfellow, C. E., Kierzek, R. & Turner, D. H. (1989) *Biochemistry*, in press.
22. Fink, T. R. & Crothers, D. M. (1972) *J. Mol. Biol.* **66,** 1–12.
23. Yuan, R. C., Steitz, J. A., Moore, P. B. & Crothers, D. M. (1979) *Nucleic Acids Res.* **7,** 2399–2418.
24. Groebe, D. R. & Uhlenbeck, O. C. (1988) *Nucleic Acids Res.* **16,** 11725–11735.
25. Gralla, J. & Crothers, D. M. (1973) *J. Mol. Biol.* **73,** 497–511.
26. Riesner, D., Maass, G., Thiebe, R., Philippsen, P. & Zachau, H. G. (1973) *Eur. J. Biochem.* **36,** 76–88.
27. Coutts, S. M., Gangloff, J. & Dirheimer, G. (1974) *Biochemistry* **13,** 3938–3948.
28. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N. & Gold, L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1364–1368.
29. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
30. Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H. & Turner, D. H. (1986) *Biochemistry* **25,** 3214–3219.
31. Turner, D. H., Sugimoto, N., Kierzek, R. & Dreiker, S. (1987) *J. Am. Chem. Soc.* **109,** 3783–3785.
32. Gralla, J. & Crothers, D. M. (1973) *J. Mol. Biol.* **78,** 301–319.
33. Uhlenbeck, O. C., Martin, F. H. & Doty, P. (1971) *J. Mol. Biol.* **57,** 217–229.
34. Papanicolaou, C., Gouy, M. & Ninio, J. (1984) *Nucleic Acids Res.* **12,** 31–44.
35. Zuker, M. & Stiegler, P. (1981) *Nucleic Acids Res.* **9,** 133–148.
36. Pleij, C. W. A., Rietveld, K. & Bosch, L. (1985) *Nucleic Acids Res.* **13,** 1717–1731.
37. Jaeger, J. A., Turner, D. H. & Zuker, M. (1989) *Methods Enzymol.* **183,** in press.
38. Sprinzl, M., Hartmann, T., Meissner, F., Moll, J. & Vorderwülbecke, T. (1987) *Nucleic Acids Res.* **15,** r53–r188.
39. Moazed, D., Stern, S. & Noller, H. F. (1986) *J. Mol. Biol.* **187,** 399–416.
40. Shub, D. A., Gott, J. M., Xu, M.-Q., Lang, B. F., Michel, F., Tomaschewski, J. & Belfort, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1151–1155.
41. Waring, R. B. & Davies, R. W. (1984) *Gene* **28,** 277–291.
42. Michel, F., Jacquier, A. & Dujon, B. (1982) *Biochimie* **64,** 867–881.
43. Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42,** 985–1002.
44. Romby, P., Westhof, E., Toukifimpa, R., Mache, R., Ebel, J., Ehresmann, C. & Ehresmann, B. (1988) *Biochemistry* **27,** 4721–4730.
45. Sugimoto, N., Kierzek, R. & Turner, D. H. (1988) *Biochemistry* **27,** 6384–6392.
46. Moore, P. B. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 721–728.
47. Gewirth, D. T. & Moore, P. B. (1988) *Nucleic Acids Res.* **16,** 10717–10732.