STUDIES IN FOUNDATIONS AND COMBINATORICS
ADVANCES IN MATHEMATICS SUPPLEMENTARY STUDIES, VOL. 1

Secondary Structure of Single-Stranded Nucleic Acids[†]

MICHAEL S. WATERMAN

Los Alamos Scientific Laboratory
Los Alamos, New Mexico

DEDICATED TO JOHN R. KINNEY

The primary structure of a single-stranded nucleic acid, such as a tRNA, is the sequence of nucleotides or bases making up the molecule. Secondary structure of such a molecule is a class of graphs in the plane which preserves the bonds in the primary structure but allows helical regions. Prediction of the most stable secondary structure is an important problem at the most basic biological level. This paper gives the first graph theoretic definition of secondary structure and derives some associated properties. A classification of secondary structures is given and used as a basis for new and efficient algorithms to find the most stable secondary structure.

1. INTRODUCTION

The sequence of nucleotides or bases of a tRNA is known as its primary structure. When primary structure of a single-stranded tRNA is known, the question arises of which bases form pairs and allow the sequence to form helical regions in two dimensions. This latter structure is known as secondary structure and was proposed quite early [9, 10]. In fact, a cloverleaf model was proposed when the first tRNA primary structure was known [14]. Secondary structure has received much attention [1] and has, as at least part of its function, a role in the interactions of tRNAs with proteins [19], in stabilizing mRNA, in packing RNA into virus particles, and in recognition of specific sites by components of the translating system [23]. Prediction of the minimum free energy (i.e., most stable) secondary structure, then, is an important problem at the most basic biological level.

To approximate the Helmholtz free energy of a proposed secondary structure, much biochemical information must be obtained. Work by Uhlenbeck *et al.* [21], Gralla and Crothers [11, 12], DeLisi and Crothers [3], and DeLisi [4] has contributed to this task. Once this information is obtained, secondary structure is carefully defined, and rules for evaluation of free energy are established, then work on prediction of secondary structure can start.

[†] Research performed under the auspices of the U.S. Department of Energy.

ISBN 0-12-599101-0

Much effort has been given to the prediction of secondary structure. One of the most important methods, due to Tinoco *et al.* [20], considers the base pairing matrix for a tRNA. This method has been modified and extended but appears to be the basis of the algorithms that have since been proposed or used [3, 7, 8, 13, 17, 19]. Below a connection is made between a Tinoco-like method and the adjacency matrix for the graph corresponding to the secondary structure. This connection allows a clear understanding of the method and its severe limitations.

This paper gives a careful and general definition of secondary structure and derives some properties that follow from the definition. A classification of secondary structure by complexity is given, and this classification is used as a basis for efficient algorithms to find the optimal secondary structure. The algorithms are new, and careful proofs are given that show they solve the general problem of optimal structure of single-stranded nucleic acids. Attention is also given to computational efficiency.

2. The Graph Theory of Secondary Structure

In this section, graph theoretic properties of secondary structures are studied. There will be no consideration of the specific pairing rules or properties of a specific single-stranded nucleic acid. Instead, the object is to give a precise definition of secondary structure and of the components of secondary structure for an arbitrary sequence of length n. The total number of secondary structures for a sequence of length n is considered. Also, the set of possible secondary structures will be decomposed into a disjoint union indexed by $k = 0, 1, \ldots$. The kth set in this union will be called the set of kth order secondary structures and will be given a simple definition based on the complexity of the secondary structure. This classification of secondary structures will allow a proof (Theorem 7.5) that the algorithms stated later in this paper do find the best secondary structure given a specific single stranded nucleic acid with pairing rules.

Figure 2.1 gives some examples of secondary structure. In the following discussion, a sequence of n points ($1-2-\cdots-n$) will be assumed given in

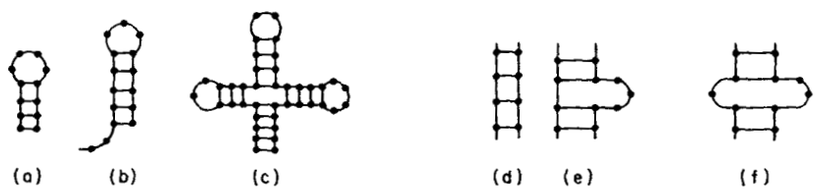
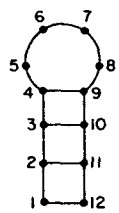


FIG. 2.1. Secondary structure.

which each point i, $1 < i < n$, is joined or bonded to $i - 1$ and $i + 1$. A point i unpaired will mean i is not joined to any points other than $i - 1$ and $i + 1$. Figure 2.1a is the configuration of the sequence $(1\text{–}2\text{–}3\cdots\text{–}12)$ with, in addition to the bonding between adjacent members of the sequence, the bonds (1, 12), (2, 11), (3, 10), (4, 9). Thus, when labeled, Fig. 2.1a becomes:



This structure is called a hairpin. The region (1–2–3–4) bonded with (9–10–11–12) is called a ladder (also called helical region in the literature) and appears in Fig. 2.1d. The unpaired region (5–6–7–8) forms what is called a hairpin loop in the literature. (In this paper hairpin loop and loop have identical meanings.) The unpaired region in the initial part of the sequence in Fig. 2.1b will be called a tail. The secondary structure in Fig. 2.1c is called a cloverleaf. The unpaired region in Fig. 2.1e is known as a bulge, while the unpaired members of Fig. 2.1f are known as an interior loop.

The problem is to construct a usable mathematical model of all possible secondary structures and to use this model to study secondary structure. It is important to realize that these molecules have a polarity (the 3′ end to the 5′ end), and that only bases of opposite polarity bond. The most useful approach to the model seems to be graph theoretic since secondary structures are clearly graphs in the plane. Also, a and b are said to be *adjacent* if (a, b) is a line of the graph. This leads to the definition of the *adjacency matrix* $A = (a_{ij})$ of a labeled graph G with P points. The element $a_{ij} = 1$ if i and j are adjacent, and $a_{ij} = 0$ otherwise. (Set the elements $a_{ii} = 0$.)

A study of the properties of these structures and the adjacency matrices has led to the following definition of secondary structure.

DEFINITION 2.1. A *secondary structure* is a graph on the set of n labeled points $\{1, 2, \ldots, n\}$ such that the adjacency matrix $A = (a_{ij})$ has the following three properties:

(i) $a_{i,i+1} = 1$ for $1 \leqslant i \leqslant n - 1$.
(ii) For each fixed i, $1 \leqslant i \leqslant n$, there is at most one $a_{ij} = 1$ where $j \neq i \pm 1$.
(iii) If $a_{ij} = a_{kl} = 1$, where $i < k < j$, then $i \leqslant l \leqslant j$.

If $a_{ij} = 1$, i and j are said to be *bonded*.

Part (i) of Definition 2.1 requires adjacent points to be bonded. Part (ii) states that each point can be bonded to at most one other point (besides the adjacent points). Finally, part (iii) is to assure that if i and j are bonded, then all bonding of points $i < k < j$ is with points l between i and j. Part (iii) is an essential part of the definition as it keeps the structure from "folding" and becoming a three-dimensional or tertiary structure.

It is interesting to note that Kleitman [16] and Hsieh [15] have studied irreducible diagrams which suggest a different representation of secondary structure. This representation, given in Fig. 2.2, is perhaps useful for combinatorial reasons but is not used here since the definitions and results here are motivated by the models of secondary structure in the biological literature.

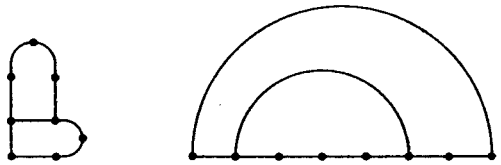


FIG. 2.2. (a) Usual diagram of secondary structure. (b) "Loop" diagram of secondary structure.

Before proceeding to the enumeration of secondary structures, it is necessary to clearly indicate what structures are to be enumerated. Two secondary structures for $(1-2-\cdots-n)$ will be considered distinct if their adjacency matrices are not equal. Consider $n = 6$, where the seventeen secondary structures are enumerated in Fig. 2.3.

An examination of Fig. 2.3 shows that the enumeration depends heavily on the labeling of the sequence. In fact, there are eleven such structures if the sequence is unlabeled.

It would be an interesting problem to enumerate the unlabeled secondary structures, although of no direct use in this paper. Also, Paul Stein of Los Alamos Scientific Laboratory has suggested the problem of determining whether or not a given unlabeled graph has a labeling such that the graph is a secondary structure.

THEOREM 2.1. *Let $S(n)$ be the number of secondary structures for n points. Then $S(1) = S(2) = 1$, and for $n > 2$, $S(n)$ satisfies*

$$S(n+1) = S(n) + \sum_{k=0}^{n-2} S(k)S(n-k-1),$$

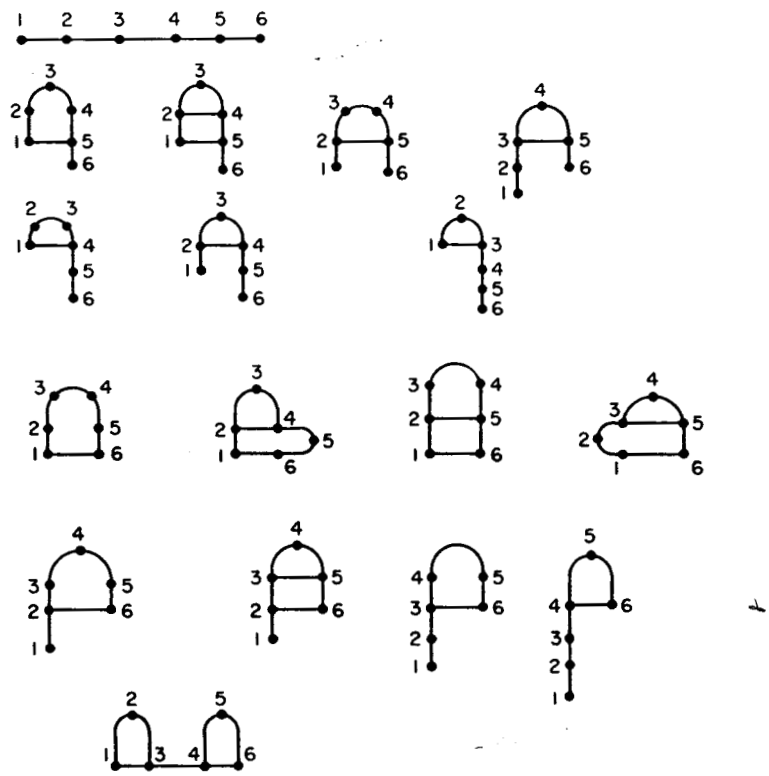


FIG. 2.3. The seventeen secondary structures for (1–2–3–4–5–6).

where $S(0) \equiv 1$. *Also,* $S(n) \geq 2^{n-2}$ *for* $n \geq 2$. (*The table below gives the values of* $S(n)$ *for* $n = 1, \ldots, 10$.)

n	0	1	2	3	4	5	6	7	8	9	10
$S(n)$	1	1	1	2	4	8	17	37	82	185	423

Proof. It is easy to see from Definition 2.1 that for $n = 1$ and $n = 2$ the only secondary structures are

$$\underset{(n=1)}{1} \qquad \underset{(n=2)}{1 \quad 2}$$

and therefore $S(1) = S(2) = 1$.

Suppose $S(k)$ is known for $1 \leq k \leq n$. Now consider $(1\text{–}2\text{–}\cdots\text{–}n+1)$. Either $n+1$ is not paired or $n+1$ is paired with j for $1 \leq j \leq n-1$. If $n+1$ is not paired, then $(1\text{–}2\text{–}\cdots\text{–}n)$ can form any possible secondary structure.

If $n+1$ is paired with j, then $(1\text{–}2\text{–}\cdots\text{–}(j-1))$ and $(j+1\text{–}\cdots\text{–}n)$ can each form any possible secondary structure. This technique forms secondary structures for $(1\text{–}2\text{–}\cdots\text{–}n+1)$ as a check of Definition 2.1 shows. That this technique enumerates all secondary structures for the sequence follows from (iii) of that definition.

From the preceeding paragraph, it follows that

$$S(n+1) = S(n) + S(n-1) + S(1)S(n-2) + \cdots + S(n-2)S(1).$$

Defining $S(0) = 1$, the equation takes the form $(n \geqslant 2)$

$$\begin{aligned} S(n+1) &= S(n) + \sum_{k=0}^{n-2} S(k)S(n-k-1) \\ &= S(n) + S(n-1) + \sum_{k=1}^{n-2} S(k)S(n-k-1) \\ &= S(n) + S(n-1) + \sum_{k=0}^{n-3} S(k+1)S(n-k-2). \end{aligned}$$

Since $S(k+1) \geqslant S(k)$,

$$S(n+1) \geqslant S(n) + S(n-1) + \sum_{k=0}^{n-3} S(k)S(n-k-2) = 2S(n).$$

Now $S(2) = 1$ so that the above inequality implies

$$S(n) \geqslant 2^{n-2}.$$

This completes the proof.

To show that 2^{n-2} is an unsatisfactory bound, assume $\lambda_2\alpha^n \leqslant S(n) < \lambda_1\alpha^n$, where $\lambda_2 > 0$. Then

$$\lambda_1\alpha^{n+1} \geqslant \lambda_2\left[\alpha^n + \lambda_2 \sum_{k=0}^{n-2} \alpha^{n-1}\right] = \lambda_2[\alpha^n + \lambda_2(n-2)\alpha^{n-1}],$$

and

$$(\lambda_1/\lambda_2)\alpha^2 \geqslant \alpha + \lambda_2(n-2),$$

which is a contradiction. Thus, the rate of growth of $S(n)$ is not geometric. However, it will be shown that $S(n)$ is bounded by a geometric growth rate. If the generating function $\phi(x)$ is defined by

$$\phi(x) = \sum_{n=0}^{\infty} S(n)x^n,$$

the recursion formula in Theorem 2.1 can be multiplied by x^{n+1} and summed to obtain

$$x^2\phi^2(x) + (x - 1 - x^2)\phi(x) + 1 = 0.$$

This equation can be solved to yield the next corollary.

COROLLARY 2.1. *If* $\phi(x) = \sum_{n=0}^{\infty} S(n)x^n$, *then*

$$\phi(x) = \frac{x^2 - x + 1 - [1 + x(x^3 - 2x^2 - x - 2)]^{1/2}}{2x^2}.$$

Next it is shown that $S(n)$ is bounded by a geometric series.

COROLLARY 2.2. *For* $n \geqslant 2$ *there is a fixed* $M > 0$ *such that*

$$2^{n-2} \leqslant S(n) \leqslant M4^n.$$

Proof. Now $S(n+1) \leqslant \sum_{k=0}^{n} S(k)S(n+1-k)$ so that if $g(0) = g(1) = 1$ and

$$g(n) = \sum_{k=0}^{n-1} g(k)g(n-k),$$

then $S(n) \leqslant g(n)$; and DeBruijn and Erdos [2] show

$$g(n) = \frac{1}{2^{n-1}}\binom{2n}{n} = O(4^n).$$

Now some additional definitions will be given so that several of the features in Fig. 2.1 can be identified from the adjacency matrix.

DEFINITION 2.2. Suppose A is the adjacency matrix for a secondary structure on $(1\text{–}2\text{–}\cdots\text{–}n)$.

(i) The point j is said to be *paired* if there is some point $i \neq j \pm 1$ such that $a_{ij} = 1$.

(ii) The region $(i+1\text{–}i+2\text{–}\cdots\text{–}(j-1))$ is a *loop* if $i+1, i+2, \ldots, j-1$ are all unpaired and $a_{ij} = 1$. The pair (i, j) is said to be the *foundation of the loop.*

(iii) The sequence $(i+1\text{–}i+2\text{–}\cdots\text{–}(j-1))$ is a *bulge* if $i+1, i+2, \ldots, j-1$ are all unpaired, i and j are both paired, and $a_{ij} \neq 1$.

(iv) An *interior loop* is two bulges $(i+1\text{–}i+2\text{–}\cdots\text{–}(j-1))$ and $(k+1\text{–}k+2\text{–}\cdots\text{–}(l-1))$ such that $a_{il} = 1$ and $a_{jk} = 1$. (Here $i < j < k < l$.)

(v) A *join* is a bulge $(i\text{–}i+1\text{–}\cdots\text{–}j)$ such that $a_{kl} = 1$ for $k < i$ implies $l \leqslant i$, and $a_{kl} = 1$ for $k > j$ implies $l \geqslant j$.

(vi) A *tail* is a sequence $(1\text{–}2\text{–}\cdots\text{–}j)$ where $1, 2, \ldots, j$ are unpaired and $j+1$ is paired.

(vii) A *ladder* is two sequences $(i+1\text{–}i+2\text{–}\cdots\text{–}i+j)$ and $(k+1\text{–}k+2\text{–}\cdots\text{–}k+j)$ such that $i+j+1<k$, $a_{i+l,k+j-l+1}=1$ for $1\leqslant l\leqslant j$ and $a_{i,k+j+1}=a_{i+j+1,k}=0$. If $i+j+3=k+1$, this last requirement is dropped.

(viii) A *hairpin* is the longest sequence $(i+1\text{–}i+2\text{–}\cdots\text{–}(j-1))$ containing exactly one loop such that $a_{i+1,j-1}=1$ and $a_{i,j}=0$. The paired points $i+1$ and $j-1$ will be called the *foundation* of the hairpin.

The above Definition 2.2 shows that the definition of secondary structure given here is rich enough to include the elementary structures in Fig. 2.1. These structures can be easily identified from the graph or from the adjacency matrix. For example, a ladder corresponds to a sequence of 1's on some negative diagonal. It is this observation that Tinoco *et al.* [20] utilize in their work on prediction of RNA secondary structure. It has been the basis of all previous algorithms to predict secondary structure (see, e.g., [19].) All these algorithms rely on the examination of all possible secondary structures. The combinatorial results shown later in this section indicate that the number of cases in such an examination is extremely large.

However, the question of whether the definition of secondary structure is too broad remains. The next theorem shows that any secondary structure is made up of loops, ladders, bulges, and tails. No secondary structure can be drawn with a point that is not a member of a loop, ladder, bulge, or tail.

THEOREM 2.2. *Any secondary structure can be uniquely decomposed into loops, ladders, bulges, and tails.*

Proof. If $a_{ij}=1$, where $i\neq j\pm 1$, then i and j are members of sequences (possibly of length 1) which are a ladder. Thus, assume i is an unpaired point. Then let $(i-j)\text{–}\cdots\text{–}i\text{–}\cdots\text{–}i+k$ be the longest sequence of unpaired points that i is a member of. If $i-j=1$ or $i+k=n$, then i belongs to a tail. Otherwise $i-j-1$ and $i+k+1$ are paired. If $a_{i-j-1,i+k+1}=1$, then i belongs to a loop. If $a_{i-j-1,i+k+1}=0$, then i belongs to a bulge.

The next theorem shows that any nontrivial secondary structure contains a loop.

THEOREM 2.3. *If a secondary structure has at least one pair, it has at least one loop.*

Proof. Clearly, there exists at least one ladder. Consider the ladders in the secondary structure: $L_1, L_2, \ldots, L_k$. By the definition of ladder, the

union of the two sequences making up the ladder does not make a new sequence. Since there are a finite number of ladders and every paired point belongs to a ladder, there exists a ladder such that the nonempty sequence of points between the two sequences making up the ladder has the property that they are all unpaired. By definition this sequence is a loop.

The next result shows the secondary structure in more detail. This result is necessary to make a classification of secondary structures by complexity.

THEOREM 2.4. *Every secondary structure can be uniquely decomposed into* (i) *hairpins and* (ii) *ladders, bulges, and tails which are not members of a hairpin.*

Proof. Each loop is contained in a hairpin. Since by definition a hairpin has exactly one loop, there are as many (nonintersecting) hairpins as loops. The structures remaining must, by Theorem 2.2, be ladders, bulges, and tails.

Next, secondary structures are classified by a certain complexity criterion. A simple lemma is necessary to make certain this definition can be accomplished.

LEMMA 2.1. *If A is the adjacency matrix for some secondary structure and, if $A' = (a'_{ij})$ is formed from A by setting $a'_{ij} = a'_{ji} = 0$ for any set of choices of i and j ($i \neq j \pm 1$), then A' is the adjacency matrix for another secondary structure.*

Proof. It is easy to check Definition 2.1 for $A' = (a'_{ij})$.

DEFINITION 2.3. Let A be the adjacency matrix for a secondary structure. A sequence $A^{(i)}$ of adjacency matrices of secondary structure is formed as follows:

(i) $A^{(0)} = A$.

(ii) Form $A^{(i+1)}$ from $A^{(i)}$ by setting $a_{kl}^{(i+1)} = a_{lk}^{(i+1)} = 0$ whenever $a_{kl}^{(i)} = a_{lk}^{(i)} = 1$, k and l are members of some hairpin, and $k \neq l \pm 1$.

The secondary structure for A is said to be *kth order* if $A^{(k)}$ is the first matrix in the sequence $\{A^{(k)}\}_{k=0}^{\infty}$ such that the secondary structure for $A^{(k)}$ has no hairpins. (Of course, this means $A^{(k)}$ is a matrix such that $a_{ij}^{(k)} = 0$ if $i \neq j \pm 1$.)

It is clear from Lemma 2.1 that the algorithm of Definition 2.3 is well defined. The next theorem states some additional properties of order.

THEOREM 2.5. *Any secondary structure is a kth order secondary structure for some unique $k \geq 0$. If the sequence is of length n, then $k \leq [n/3]$.*

Proof. Since at least three points are required to make up a hairpin, there are a finite number of hairpins for any secondary structure. If there are no such hairpins, the secondary structure is 0th order and there is nothing to show. If there is exactly one hairpin, there can be no paired points not belonging to the hairpin; and therefore the secondary structure is first order. Otherwise there are at least two hairpins.

Assume $A^{(i)}$ has at least two hairpins. At least one pair of these hairpins has only unpaired points between them. Thus $A^{(i+1)}$ can have at least one less hairpin than $A^{(i)}$, and the algorithm must terminate in a finite number of steps.

In fact, each hairpin has at least one point in its loop and at least one pair of points on or above the foundation. Therefore, there are at most $\lceil n/3 \rceil$ loops and the secondary structure cannot have order greater than $\lceil n/3 \rceil$.

It is very convenient to determine the order of a secondary structure from a graph, and this is illustrated in Fig. 2.4. An arrow will denote passage from the secondary structure for $A^{(i)}$ to that for $A^{(i+1)}$ in Fig. 2.4.

It is clear that order is related to complexity and that, for sufficiently large n, secondary structures of arbitrarily large order exist. A proof of this can be based on Fig. 2.4f. If a secondary structure of order k can be constructed, add it to each side of a hairpin. The new structure is of order $k + 1$.

THEOREM 2.6. *For any $k > 0$ there is an $n = n(k)$ such that some secondary structure for (1–2–$\cdots$–n) is of order k.*

Next, the secondary structures with exactly one loop are enumerated. These structures are the basis of the lower bound in Theorem 2.1.

THEOREM 2.7. *There are exactly $2^{n-2} - 1$ secondary structures of length $n (n \geq 2)$ which have exactly one loop.*

Proof. Let $L(n)$ be the number of secondary structures of length n which have *no more* than one loop. By inspection $L(1) = L(2) = 1$.

The proof follows that of Theorem 2.1. Assume $L(1), L(2), \ldots, L(n)$ are known and consider (1–2–$\cdots$–$n + 1$). Either $n + 1$ is not paired or it is paired with j, where $1 \leq j \leq n - 1$. In the first case, (1–2–$\cdots$–$n + 1$) can form $L(n)$ secondary structures of interest. In the second case, ($j + 1$–$\ldots$–$(n - 1)$) can form any possible secondary structure of interest. (1–2–$\cdots$–$(j - 1)$) cannot form a loop as this would make two loops.

Thus it is clear that

$$\begin{aligned} L(n+1) &= L(n) + L(n-1) + \cdots + L(1) \\ &= L(n) + L(n) = 2L(n) \qquad \text{for} \quad n \geq 2. \end{aligned}$$

Since $L(2) = 1$,

$$L(n) = 2^{n-2} \qquad \text{for} \quad n \geqslant 2.$$

The function $L(n)$ counts all secondary structures with one loop plus the unique case of no pairing. The result now follows.

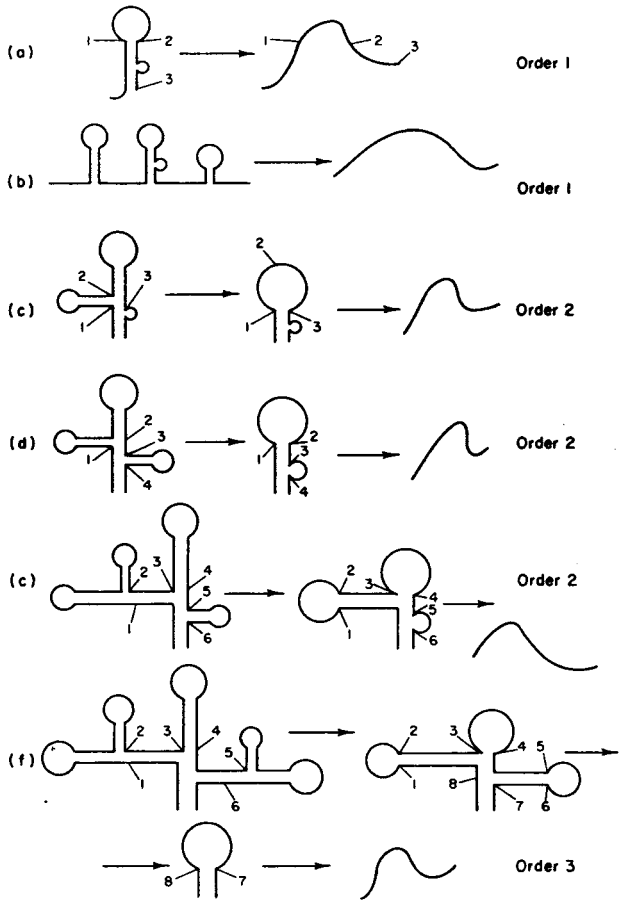


FIG. 2.4. Example of order.

Although Theorem 2.6 is a satisfactory answer to the question of how many (order 1) secondary structures have exactly one loop, the question of the number of order 1 secondary structures remains. This problem is dealt with in the next theorem, and the solution follows from an application of the discrete renewal theorem.

THEOREM 2.8. *Let $S_n^{(1)}$ be the number of secondary structures of order 1 for a sequence of length n. There is a $\lambda \in (2, 3)$ ($\lambda = 2.2055\ldots$) such that λ is a solution of $x^3 - 2x^2 - 1 = 0$ and*

$$\lim_{n\to\infty} \lambda^{-n} S_n^{(1)} = \mu^{-1},$$

where $\mu = \lambda^{-1} + 3\lambda^{-3} + 2^{-4}[2\lambda(\lambda - 2)^{-2} - 2\lambda^{-1} - 8\lambda^{-2} - 24\lambda^{-3}]$.

Proof. Let s_n be the number of zero or first order secondary structures for a sequence of length n. (Clearly, $s_n = S_n^{(1)} - 1$.) Also, let $h_1 = 1$, $h_2 = 0$, and, for $k \geqslant 3$, let h_k be the number of secondary structures with no more than one loop for a sequence of $k - 2$ points.

It easily follows from the definition of first order that if i and j are paired there can be no more than one loop in the sequence $(i + 1\text{–}\cdots\text{–}(j - 1))$. By an argument similar to that of Theorem 2.1, the system of equations below follows:

$$\begin{aligned}
s_0 &= 1,\\
s_1 &= 1 = s_0h_1,\\
s_2 &= 1 = s_0h_2 + s_1h_1,\\
s_3 &= s_2 + s_0 = s_0h_3 + s_1h_2 + s_2h_1,\\
s_4 &= s_3 + s_01 + s_1 = s_0h_4 + s_1h_3 + s_2h_2 + s_3h_1\\
s_5 &= s_4 + s_0h_5 + s_1h_4 + s_2 = s_0h_5 + s_1h_4 + s_2h_3 + s_3h_2 + s_4h_1,\\
&\vdots\\
s_n &= s_{n-1} + s_0h_n + s_1h_{n-1} + \cdots + s_{n-3}h_3,\\
&= s_0h_n + s_1h_{n-1} + \cdots + s_{n-3}h_3 + s_{n-2}h_2 + s_{n-1}h_1.
\end{aligned}$$

Define $b_0 = 1$, and, for $\lambda > 2$,

$$v_n = s_n\lambda^{-n}, \qquad f_n = h_n\lambda^{-n}.$$

Then,

$$\begin{aligned}
v_0 &= b_0,\\
v_n &= v_0f_n + v_1f_{n-1} + \cdots + v_{n-1}f_1, \qquad n \geqslant 1.
\end{aligned}$$

Now,

$$\begin{aligned}
f = \sum_{n=1}^{\infty} f_n &= \lambda^{-1} + \lambda^{-3} + \sum_{n=4}^{\infty} 2^{n-4}\lambda^{-n}\\
&= \lambda^{-1} + \lambda^{-3} + 2^{-4}(2/\lambda)^4 \sum_{m=0}^{\infty} (2/\lambda)^m\\
&= \lambda^{-1} + \lambda^{-3} + \lambda^{-3}(\lambda - 2)^{-1}.
\end{aligned}$$

If $\lambda = 3$, $f = 3^{-1} + 3^{-3} + 3^{-3}(3-2)^{-1} < 1$. But $\lim_{\lambda \downarrow 2} f = +\infty$. Thus there is a $\lambda \in (2, 3)$ such that $f = 1$. The equation $f = 1$ can be rewritten as the equation in the theorem statement by factoring and deleting the factor $x - 1$.

Next apply the discrete renewal theorem as given in Feller [5, pp. 330–331]. The only additional computation to obtain the result is

$$\mu = \sum_{n=1}^{\infty} nf_n = \lambda^{-1} + 3\lambda^{-3} + \sum_{n=4}^{\infty} n2^{n-4}\lambda^{-n}$$
$$= \lambda^{-1} + 3\lambda^{-3} + 2^{-4}[2\lambda(\lambda-2)^{-2} - 2\lambda^{-1} - 8\lambda^{-2} - 24\lambda^{-3}].$$

Of course, there is only one secondary structure of order 0 so

$$\lim_{n\to\infty} \lambda^{-n} S_n^{(1)} = \lim_{n\to\infty} \lambda^{-n} s_n = \mu^{-1}.$$

The proof is now completed.

Since there are exactly $2^{n-2} - 1$ secondary structures with exactly one loop ($n \geq 2$), it follows that, if $S_n^{(1)*}$ is the number of first order secondary structures with more than one loop,

$$\lim_{n\to\infty} \lambda^{-n} S_n^{(1)*} = \mu^{-1}.$$

3. First Order Single Loop Secondary Structures

In this section results are derived which allow the determination of the minimum free energy first order single loop secondary structure for a single-stranded nucleic acid. The results are quite general and allow input of more precise information as it becomes available. Of course, restrictions on the possible secondary structure are imposed, and the set of possible secondary structures is not nearly as general as those described in Section 2. The algorithms which are presented in this section are motivated in form and proof by previous work [22] on sequence homology, but the results of [22] cannot be directly adapted to this problem. In fact, some quite distinct properties arise as a result of the assumptions and restrictions imposed here. Symmetries exist in the sequence homology problem that are absent here.

Some fundamental definitions are now made.

Definition 3.1. Let $\mathscr{A}$ be a finite set called the *alphabet*. Then the following functions are assumed to be given.

(i) A *pairing function* is a symmetric function $p(\cdot, \cdot)$ defined on $\mathscr{A} \times \mathscr{A}$ such that $p(a, b) \in \{0, 1\}$ for $(a, b) \in \mathscr{A} \times \mathscr{A}$.

(ii) A real valued symmetric function $\alpha(\cdot, \cdot)$ defined on $\mathscr{A} \times \mathscr{A}$ is called the free energy associated with the pair (a, b). Using $\alpha(\cdot, \cdot)$, the *ladder function* can be defined on $\mathscr{A}^k \times \mathscr{A}^k$ for $k \geqslant 1$ by

$$\alpha(a_1 \cdots a_k, b_1 \cdots b_k) = \sum_{i=1}^{k} \alpha(a_i, b_i).$$

(iii) A real valued function $\beta(\cdot)$ defined on $\mathscr{A}^k$ for $k \geqslant 1$ is called the *bulge function.* Its value $\beta(a_1 \cdots a_k)$ is said to be the free energy associated with a bulge $a_1 \cdots a_k$.

(iv) The real valued symmetric function $\gamma(\cdot, \cdot)$, defined on $\mathscr{A}^k \times \mathscr{A}^l$ for $\min\{k, l\} \geqslant 1$, is called the *interior loop function.* Its value $\gamma(a_1 \cdots a_k, b_1 \cdots b_l)$ is said to be the free energy associated with an interior loop $a_1 \cdots a_k$ and $b_1 \cdots b_l$.

(v) The real valued function $\zeta(\cdot; \cdot)$ defined on $\mathscr{A}^2 \times \mathscr{A}^k$ for $k \geqslant 1$ is called the *loop function.* The value $\zeta(cd; a_1 \cdots a_k)$ is called the free energy associated with a loop $a_1 \cdots a_k$ with foundation cd.

(vi) The real valued function $\xi(\cdot)$ defined on $\mathscr{A}^k$ for $k \geqslant 1$ is called the *join function.* Its value $\xi(a_1 \cdots a_k)$ is called the free energy associated with a join.

(vii) The real valued function $\tau(\cdot)$ defined on $\mathscr{A}^k$ for $k \geqslant 1$ is called the *tail function.* Its value $\tau(a_1 \cdots a_k)$ is called the free energy associated with a tail. In this work $\tau(a_1 \cdots a_k) \equiv 0$ for all elements of $\mathscr{A}^k$ for all k.

To establish an efficient algorithm, several assumptions are made about these five functions. While these assumptions might seem rather arbitrary, they appear natural from energy considerations; and all are satisfied by current estimates of the functions [6, 11, 12, 19, 20]. They are necessary for the proof of Theorem 3.1 below. For ease of expression below when, say, $\beta(k)$ is written, it is to be interpreted as the bulge function β evaluated at an arbitrary sequence of k letters. An inequality like

$$\beta(k + l) \leqslant \beta(k) + \beta(l)$$

means

$$\beta(a_1 \cdots a_k, b_1 \cdots b_l) \leqslant \beta(a_1 \cdots a_k) + \beta(b_1 \cdots b_l).$$

Also, since ζ is a function of $\mathscr{A}^2 \times \mathscr{A}^k$, ζ is written as $\zeta(k)$.

DEFINITION 3.2. The functions $(p, \alpha, \beta, \gamma, \zeta)$ with an alphabet are said to be *regular* if

(i) for all arguments, β, γ, ζ are positive,

$$\alpha(a, b) \text{ is negative} \quad \text{if} \quad p(a, b) = 1,$$
$$\alpha(a, b) = +\infty \quad \text{if} \quad p(a, b) = 0,$$

and

(ii) the following inequalities hold:

$$\begin{aligned}
\beta(k+l) &\leqslant \beta(k)+\beta(l),\\
\gamma(k,l) &\leqslant \beta(k)+\beta(l),\\
\max\{\gamma(k+q,l),\gamma(q+k,l),\gamma(k,l+q),\gamma(k,q+l)\} &\leqslant \beta(q)+\gamma(k,l),\\
\max\{\zeta(k+l),\zeta(l+k)\} &\leqslant \beta(k)+\zeta(l),\\
\zeta(l+q+n) &\leqslant \gamma(l,n)+\zeta(q),\\
\gamma(k+q,l+n) &\leqslant \gamma(k,l)+\gamma(q,n).
\end{aligned}$$

Let **a** be a finite sequence of elements of $\mathscr{A}$ with the functions $p, \alpha, \beta, \gamma, \zeta, \xi$ defined as in Definition 3.1. By Theorem 2.2, any secondary structure $S(\mathbf{a})$ for (**a**) can be decomposed into loops (LO), bulges (B'), ladders (LA), and tails (T). If the interior loops are denoted by IL and joins by J, then let $B = (B' \sim IL) \sim J$.

The next definition concerns the free energy of a class of secondary structures. Free energy here refers to the difference in Helmholtz free energy between a given secondary structure $S(\mathbf{a})$ and the completely unpaired structure $S_0(\mathbf{a})$ (a tail). $S_0(\mathbf{a})$ is the random coil state and is a member of $\mathscr{T}$. While the traditional notation for free energy is ΔF or ΔA [24], the delta has been dropped as its mathematical connotations might be misleading. The assumption of additivity for free energy is equivalent to independence of energy contributions of the various structural components. This results from the assumption of statistical mechanics that the partition function is a product for independent energy contributions [24].

DEFINITION 3.3. Assume **a** is a sequence as given above and that a first order secondary structure $S(\mathbf{a})$ has at least one pair. Then the *first order free energy associated with* $S(\mathbf{a})$ is defined by

$$F_1(S(a)) = \sum_{v \in J} \xi(v) + \sum_{w \in LO} \zeta(w) + \sum_{x \in B} \beta(x) + \sum_{y \in IL} \gamma(y) + \sum_{z \in LA} \alpha(z).$$

If $S(\mathbf{a})$ has no pair, $F_1(S(\mathbf{a})) = 0$ by definition. (This is consistent with the omission of T in the above equation.) The *first order free energy associated with* **a** is defined by

$$F_1(\mathbf{a}) = \min_{S(\mathbf{a}) \in \mathscr{S}} F_1(S(\mathbf{a})),$$

where

$$\begin{aligned}
\mathscr{S} = \{&S(a): S(a) \text{ is zero or first order,}\\
&w \in LO \text{ implies } w \text{ is a sequence of at least } m \text{ elements,}\\
&\text{and } z \in LA \text{ implies } \alpha(z) \leqslant \delta < 0\}.
\end{aligned}$$

If $\mathscr{S}^* = \mathscr{S} \cap \{S(\mathbf{a}): S(a)$ has no more than one loop$\}$, then define

$$F_1^*(\mathbf{a}) = \min_{S(\mathbf{a}) \in \mathscr{S}^*} F(S(\mathbf{a})).$$

Since any loop must have at least one element, the requirement of at least m elements is no loss of generality and can be used to eliminate computation. The requirement of $\alpha(z) \leqslant \delta < 0$ is to assure that ladders have enough free energy to be stable. Previously, at least three (consecutive) pairs in a ladder had been required; the present requirement is more realistic but can include the previous requirement.

Definition 3.3 is a mathematical statement of the problem of "best" secondary structure. It will require some slight alterations in succeeding sections but is entirely adequate for first order secondary structures.

A few more comments are in order before the next theorem. The algorithm given by the theorem essentially considers, for $a_1 \cdots a_i$ and $a_n \cdots a_{n-j+1}$, the free energies associated with pairs, bulges, and interior loops between the two sequences. The algorithm chooses the "structure" which has the smallest free energy. As the "structure" formed has no loop and is not a secondary structure, the word *alignment* is used. This usage is familiar from the literature on homology of biological sequences [22]. Another convention is that, if an alignment ends in a_i and a_{n-j+1} paired, it is said that that alignment ends in a pair. Of course, the final step of the theorem obtains the value of $F_1^*(\mathbf{a})$.

THEOREM 3.1. *Let $\mathbf{a} = a_1 \cdots a_n$ be a sequence from $\mathscr{A}^n$ and assume the functions $(p, \alpha, \beta, \gamma, \zeta)$ are regular. Let $\delta < 0$ and m, a positive integer, be used in the definition of F_1^*. Define*

$$b_j = a_{n-j+1}, \qquad 1 \leqslant j \leqslant n$$

and $\rho(0, j) = \rho(i, 0) = 0$ for $0 \leqslant i, j \leqslant n$. Then inductively define $\rho(i, j)$, for $0 < ij$; $i + j \leqslant n - m$, to be the minimum of zero and:

(i)

$$\rho(i - k, j - k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant \min\{i, j\}$ and $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \delta$;

(ii)

$$\rho(i - k, j) + \beta(a_{i-k+1} \cdots a_i),$$

where $1 \leqslant k \leqslant i - 1$ and some alignment for $\rho(i - k, j)$ ends in a pair;

(iii)

$$\rho(i, j - k) + \beta(b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant j - 1$ and some alignment for $\rho(j, j - k)$ ends in a pair; and

(iv)

$$\rho(i-k, j-l) + \gamma(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j),$$

where $1 \leqslant k \leqslant i-1, 1 \leqslant l \leqslant j-1$, *and some alignment for* $\rho(i-k, j-l)$ *ends in a pair.*

Then

$$F_1^*(\mathbf{a}) = \min\{\rho(i,j) + \zeta(a_i b_j; a_{i+1} \cdots a_{n-j}) : i + j \leqslant n - m \text{ and some alignment for } \rho(i,j) \text{ ends in a pair}\},$$

where $F_1^*(\mathbf{a}) = 0$ *if the above set is empty.*

Proof. The proof that $\rho(i, j)$ is the minimum free energy associated with $a_1 \cdots a_i$ and $b_1 \cdots b_j = a_n \cdots a_{n-j+1}$ is postponed at present, and the statement is assumed true.

If a best first order secondary structure with no more than one loop has at least one pair, then it has exactly one loop. Let $a_i b_j$ be the foundation of that loop. Then the free energy e for $a_1 \cdots a_i$ and $b_1 \cdots b_j$ in this structure satisfies $\rho(i, j) \leqslant e < 0$. If $\rho(i, j)$ has an alignment with i and j paired, then $\rho(i, j) < e$ is a contradiction and $\rho(i, j) = e$. Otherwise the assumption of the paragraph above requires $\rho(i, j) < e$. There are two situations for $\rho(i, j)$. The first situation is

$$\rho(i,j) = \rho(i-k, j-l) + \gamma(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j),$$

which implies

$$\begin{aligned}\rho(i-k, j-l) &+ \zeta(a_{i-k} b_{j-l}; a_{i-k+1} \cdots a_{n-j+l}) \\ &\leqslant \rho(i-k, j-l) + \gamma(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j) + \zeta(a_i b_j; a_{i+1} \cdots a_{n-j}) \\ &< e + \zeta(a_i b_j; a_i \cdots a_{n-j}) = f_1^*(\mathbf{a}).\end{aligned}$$

The other situation is

$$\rho(i,j) = \rho(i-k, j) + \beta(a_{i-k+1} \cdots a_i)$$

so that

$$\begin{aligned}\rho(i-k, j) &+ \zeta(a_{i-k} b_j; a_{i-k+1} \cdots a_{n-j}) \\ &\leqslant \rho(i-k, j) + \beta(a_{i-k+1} \cdots a_i) + \zeta(a_i b_j; a_{i+1} \cdots a_{n-j}) \\ &< e + \zeta(a_i b_j; a_{i+1} \cdots a_{n-j}) = f_1^*(\mathbf{a}).\end{aligned}$$

Each of these situations (the second covers two cases, one of which is omitted) results in a contradiction since a member of $\mathcal{S}^*$ has been exhibited with free energy smaller than F_1^*. Therefore, subject to the assumption on $\rho(i, j)$ made above, the theorem holds.

To show $\rho(i, j)$ has the required property, assume that the property holds for $\rho(k, l)$, where $0 \leqslant k \leqslant i$, $0 \leqslant l \leqslant j$, $k + l < i + j$. Consider an optimal (minimum free energy) alignment for $a_1 \cdots a_i$ and $b_1 \cdots b_j$. Either (i) a_i and b_j are paired, (ii) exactly one of a_i and b_j is paired, or (iii) neither a_i or b_j is paired. Each case will be handled below. It must be shown that, in each case, the optimal alignment is one that is obtained by the theorem.

In case (i), the pair $a_i b_j$ belongs to a ladder z satisfying $\alpha(z) \leqslant \delta$. The free energy associated with the optimal alignment for $a_1 \cdots a_i$ and $b_1 \cdots b_j$ has, by the equation of Definition 3.3, the form

$$r(i - k, j - k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where an alignment for $r(i - k, j - k)$ does not end with a_{i-k} and b_{j-k} paired. If $\rho(i - k, j - k) < r(i - k, j - k)$, an easy contradiction results so that, in case (i), $\rho(i, j)$ has the required property.

For case (ii), suppose, without loss of generality, that a_i is paired. Then the minimum free energy associated with $a_1 \cdots a_i$ and $b_1 \cdots b_j$ is of the form

$$r(i, j - k) + \beta(b_{j-k+1} \cdots b_j)$$

since b_j must belong to a bulge. In the alignment for $r(i, j - k)$, a_i and b_{j-k} are paired. If $\rho(i, j - k) = r(i, j - k)$, then the result holds. Therefore, assume $\rho(i, j - k) < r(i, j - k)$. If a_i and b_{j-k} are paired in some alignment for $\rho(i, j - k)$, then an easy contradiction results. There are three more possibilities to be considered.

First, suppose an alignment for $\rho(i, j - k)$ is of the form

$$\rho(i - l, j - k) + \beta(a_{i-l+1} \cdots a_i) = \rho(i, j - k).$$

Then,

$$\begin{aligned} &\rho(i - l, j - k) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j) \\ &\quad \leqslant \rho(i - l, j - k) + \beta(a_{i-l+1} \cdots a_i) + \beta(b_{j-k+1} \cdots b_j) \\ &\quad < r(i, j - k) + \beta(b_{j-k+1} \cdots b_j), \end{aligned}$$

and a contradiction has been obtained.

Next, assume an alignment for $\rho(i, j - k)$ has the form

$$\rho(i, j - k - l) + \beta(b_{j-k-l+1} \cdots b_{j-k}) = \rho(i, j - k).$$

Then,

$$\begin{aligned} &\rho(i, j - k - l) + \beta(b_{j-k-l+1} \cdots b_j) \\ &\quad \leqslant \rho(i, j - k - l) + \beta(b_{j-k-l+1} \cdots b_{j-k}) + \beta(b_{j-k+1} \cdots b_j) \\ &\quad < r(i, j - k) + \beta(b_{j-k+1} \cdots b_j) \end{aligned}$$

and, again, a contradiction is shown.

To conclude case (ii), suppose an alignment for $\rho(i, j-k)$ is of the form

$$\rho(i-l, j-k-q) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k-q+1} \cdots b_{j-k}) = \rho(i, j-k).$$

Then,

$$\begin{aligned}
&\rho(i-l, j-k-q) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k-q+1} \cdots b_j) \\
&\quad \leqslant \rho(i-l, j-k-q) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k-q+1} \cdots b_{j-k}) \\
&\qquad + (b_{j-k+1} \cdots b_j) \\
&\quad < r(i, j-k) + \beta(b_{j-k+1} \cdots b_j)
\end{aligned}$$

and, with this contradiction, case (ii) is concluded.

In case (iii), neither a_i nor b_j is paired. Then the minimum free energy associated with $a_1 \cdots a_i$ and $b_1 \cdots b_j$ has the form

$$r(i-l, j-k) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j).$$

If $\rho(i-l, j-k) = r(i-l, j-k)$, the proof is complete. Therefore, assume $\rho(i-l, j-k) < r(i-l, j-k)$. If a_{i-l} and b_{j-k} are paired in some alignment for $\rho(i-l, j-k)$, an immediate contradiction results. There are essentially two distinct situations to consider.

First, suppose an alignment for $\rho(i-l, j-k)$ has the form

$$\rho(i-l-q, j-k) + \beta(a_{i-l-q+1} \cdots a_{i-l}) = \rho(i-l, j-k).$$

Then,

$$\begin{aligned}
&\rho(i-l-q, j-k) + \gamma(a_{i-l-q+1} \cdots a_i, b_{j-k+1} \cdots b_j) \\
&\quad \leqslant \rho(i-l-q, j-k) + \beta(a_{i-l-q+1} \cdots a_{i-l}) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j) \\
&\quad < r(i-l, j-k) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j),
\end{aligned}$$

which is a contradiction.

The last (nonredundant) situation to consider is that an alignment for $\rho(i-l, j-k)$ has the form

$$\rho(i-l-q, j-k-p) + \gamma(a_{i-l-q+1} \cdots a_{i-l}, b_{j-k-p+1} \cdots b_{j-k}).$$

Then,

$$\begin{aligned}
&\rho(i-l-q, j-k-p) + \gamma(a_{i-l-q+1} \cdots a_i, b_{j-k-p+1} \cdots b_j) \\
&\quad \leqslant \rho(i-l-q, j-k-p) + \gamma(a_{i-l-q+1} \cdots a_{i-l}, b_{j-k-p+1} \cdots b_{j-k}) \\
&\qquad + \gamma(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j) \\
&\quad < r(i-l, j-k) + \gamma(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j).
\end{aligned}$$

With this last contradiction, case (iii) and therefore the proof of the theorem is complete.

Hopefully, it is evident that the algorithm of Theorem 3.1 can be implemented on a computer and that a matrix formulation is the most convenient. The only points (i, j), except for ladders, to search over are those whose optimal alignments ended in a ladder. It is possible to make other observations to reduce the search even more, and this is handled in the next corollaries and theorem. For the theorem it is necessary to make further assumptions about β and γ. These assumptions are quite reasonable and satisfied by current estimates of β and γ [19]. The set I denotes the set of integers.

COROLLARY 3.1. *For Theorem* 3.1, *let* $\lambda = \min_{\mathscr{A}^2} \alpha(a, b) < 0$. *Then*

(i) *if* $\delta/\lambda \in I$, *then* $\rho(i, j) = 0$ *for* $\min\{i, j\} < \delta/\lambda$,
(ii) *if* $\delta/\lambda \notin I$, *then* $\rho(i, j) = 0$ *for* $\min\{i, j\} \leqslant \delta/\lambda$.

Proof. An examination of Theorem 3.1 shows that $\rho(i, j) \neq 0$ must first be achieved by part (i) of that theorem; but part (i) does not have a contribution unless $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \delta$. The first occasion that this can occur is for the smallest integer k such that $k\lambda \leqslant \delta$ or $k \leqslant \delta/\lambda < 0$. The corollary follows.

COROLLARY 3.2. *In Theorem* 3.1, (i) *can be replaced by*

(i)′

$$\rho(i - k, j - k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant \min\{i, j\}$, $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j)$, *and no alignments for* $\rho(i - k, j - k)$, $\rho(i - k + 1, j - k + 1), \ldots, \rho(i - 1, j - 1)$ *end in a pair, and* (ii) *can be replaced by*

(ii)′

$$\rho(i - k, j - k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant \min\{i, j\}$, $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant 0$, *no alignments for* $\rho(i - k + 1, j - k + 1), \ldots, \rho(i - 1, j - 1)$ *end in a pair, and some alignment for* $\rho(i - k, j - k)$ *ends in a pair.*

Proof. The proof follows easily from the additivity of α.

THEOREM 3.2. *Assume the situation of Theorem* 3.1 *and, in addition, that*

$$\beta(k) \leqslant \min\{\beta(k + l), \beta(l + k)\},$$
$$\beta(k) \leqslant \min\{\gamma(k + l), \gamma(l, k)\},$$

and

$$\gamma(k, l) \leqslant \min\{\gamma(k + i, l + j), \gamma(i + k, j + l)\},$$

where i, j, k, and l are arbitrary nonnegative integers. Suppose that $\rho(i - k, j - l)$ and $\rho(i, j)$ each possess an alignment that ends in a pair. Then, if $\rho(i, j) \leqslant \rho(i - k, j - l)$, $\rho(i - k, j - l)$ need not be used in steps (ii), (iii), *or* (iv) *to determine $\rho(i + p, j + q)$ (where $0 < p + q$). Also, $\rho(i - k, j - l)$ need not be used to determine $F_1^*(\mathbf{a})$.*

Proof. Suppose $q = 0$ and $p \neq 0$. Then three cases need to be considered:

(i) $l = 0, k \neq 0$,
(ii) $l \neq 0, k \neq 0$, and
(iii) $l \neq 0, k = 0$.

For case (i),

$$\begin{aligned}\rho(i, j) + \beta(a_{i+1} \cdots a_{i+p}) &\leqslant \rho(i - k, j) + \beta(a_{i+1} \cdots a_{i+p}) \\ &\leqslant \rho(i - k, j) + \beta(a_{i-k+1} \cdots a_{i+p}).\end{aligned}$$

For case (ii),

$$\begin{aligned}\rho(i, j) + \beta(a_{i+1} \cdots a_{i+p}) &\leqslant \rho(i - k, j - l) + \beta(a_{i+1} \cdots a_{i+p}) \\ &\leqslant \rho(i - k, j - l) + \beta(a_{i-k+1} \cdots a_{i+p}) \\ &\leqslant \rho(i - k, j - l) + \gamma(a_{i-k+1} \cdots a_{i+p}, b_{j-l+1} \cdots b_j).\end{aligned}$$

In case (iii),

$$\begin{aligned}\rho(i, j) + \beta(a_{i+1} \cdots a_{i+p}) &\leqslant \rho(i, j - l) + \beta(a_{i+1} \cdots a_{i+p}) \\ &\leqslant \rho(i, j - l) + \gamma(a_{i+1} \cdots a_{i+p}, b_{j-l+1} \cdots b_j).\end{aligned}$$

In each of the three cases above, the quantity for $\rho(i, j)$ was smaller than the corresponding quantity for $\rho(i - k, j - l)$. The only (nonredundant) situation remaining is $p \neq 0, q \neq 0$. Then,

$$\begin{aligned}&\rho(i - j) + \gamma(a_{i+1} \cdots a_{i+p}, b_{j+1} \cdots b_{j+q}) \\ &\quad\leqslant \rho(i - k, j - l) + \gamma(a_{i+1} \cdots a_{i+p}, b_{j+1} \cdots b_{j+q}) \\ &\quad\leqslant \rho(i - k, j - l) + \gamma(a_{i-k+1} \cdots a_{i+p}, b_{j-l+1} \cdots b_{j+q}),\end{aligned}$$

and the proof is concluded.

EXAMPLE. For a simple example, let $A = \{a, u, g, c, n\}$ with $p(g, c) = p(a, u) = 1$ and $p(x, y) = 0$ elsewhere. (The base n is to be thought of as a

neutral element.) Define

$$\alpha(g, c) = \alpha(a, u) = -2,$$
$$\beta(k) = k,$$
$$\gamma(k, l) = k + l,$$

and

$$\zeta(k) = k.$$

The functions $(p, q, \beta, \gamma, \zeta)$ are easily seen to be regular. With $\delta = -2 < 0$ and $m = 2$, Theorems 3.1, 3.2, and Corollary 3.1 all hold.

Now consider

$$\mathbf{a} = \textit{ggguaunnnauagggnnncccauannnuauccc}.$$

The elements of this sequence of length will be numbered from 2 to 34. The structure for $F_1^*(\mathbf{a}) = -7$ has four ladder regions and appears in Fig. 4.1.

Base number	14 pairs with base number 22	:	*gc*
Base number	15 pairs with base number 21	:	*gc*
Base number	16 pairs with base number 20	:	*gc*
Base number	11 pairs with base number 24	:	*au*
Base number	12 pairs with base number 23	:	*ua*
Base number	5 pairs with base number 30	:	*ua*
Base number	6 pairs with base number 29	:	*au*

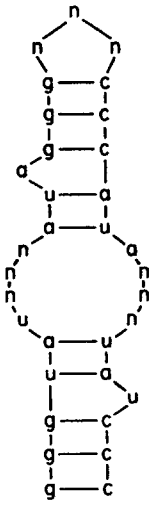


FIG. 4.1. Best single loop secondary structure for **a**.

Base number	2 pairs with base number 34	:	*gc*
Base number	3 pairs with base number 33	:	*gc*
Base number	4 pairs with base number 32	:	*gc*

4. The Hairpin Matrix

The purpose of this section is to define a matrix which will allow efficient calculation of the best first order secondary structure $F_1(\mathbf{a})$ and the best second order secondary structure. In addition, the concepts of the last section and of this section motivate the solution of the problem of finding the best Nth order secondary structure.

Motivation for the matrix defined below was the computation of the best single loop secondary structures for $a_i a_{i+1} \cdots a_j$ for $1 \leqslant i < j \leqslant n$. There are, of course, $n(n-1)/2$ such sequences. The use for such a matrix is to use the free energies

$$F_1{}^*(a_i a_{i+1} \cdots a_j)$$

in much the same way $\beta(a_i \cdots a_j)$ was used in Section 3. Certain problems immediately arise. A major difficulty is that $F_1{}^*(a_i \cdots a_j)$ gives tails weight zero when a tail of such a structure could become a bulge or a join when the entire sequence $a_1 a_2 \cdots a_n$ is considered. This difficulty is overcome by the next definition.

Definition 4.1. For a given sequence $\mathbf{a} = a_1 \cdots a_n$, regular functions $(p, \alpha, \beta, \gamma, \zeta)$ and parameters $\delta < 0$ and m, define the *hairpin matrix* H to be the symmetric $n \times n$ matrix with $h_{ii} = 0$ for $1 \leqslant i \leqslant n$ and, for $1 \leqslant i < j \leqslant n$,

$$h_{ij} = \min_{S \in \mathscr{S}_{ij}} F_1(S),$$

where $\mathscr{S}_{ij} = \{S(a_i \cdots a_j) : S(a_i \cdots a_j)$ is first order with exactly one loop, $w \in LO$ implies w has at least m elements, $z \in LA$ implies $\alpha(z) \leqslant \delta < 0$, and $p(a_i, a_j) = 1\}$. If $\mathscr{S}_{ij} = \varnothing$, let $h_{ij} = K$, where K is large enough so that the structure corresponding to h_{ij} is not used.

Since this definition is fundamental in the next two sections, it must be carefully examined. The value h_{ij} is the minimum free energy of single loop secondary structures for $a_i \cdots a_j$ with the restriction that a_i and a_j are paired. This avoids the problem of tails with weight zero and also allows the omission of calculation of

$$F_1{}^*(a_i \cdots a_j)$$

for many values of i and j.

The calculation of h_{ij} cannot be handled directly by Theorem 3.1. The next theorem modifies Theorem 3.1 to the present problem for the sequence $a_1 \cdots a_n$. Of course, $a_i \cdots a_j$ can be similarly handled. Notice that $h_{1,n} > 0$ is allowed.

THEOREM 4.1. *Let $a_1 \cdots a_n$ be a sequence from $\mathscr{A}^n$ with regular functions $(p, \alpha, \beta, \gamma, \zeta)$, $\delta < 0$, and m a positive integer. Define $b_j = a_{n-j+1}$, $1 \leqslant j \leqslant n$. Let*

$$v = \min\{k \geqslant 1 : \alpha(a_1 \cdots a_k, b_1 \cdots b_k) \leqslant \delta, \text{ and } 2k \leqslant n - m\}.$$

If v does not exist, then $h_{1n} = K$. Otherwise, let

$$\rho(v, v) = \alpha(a_1 \cdots a_v, b_1 \cdots b_v),$$

where the alignment for $\rho(v, v)$ is said to end in a pair, and inductively define $\rho(i, j)$, $v \leqslant i$, $v \leqslant j$, $2v < i + j \leqslant n - m$, to be the minimum of zero and:

(i)

$$\rho(i - k, j - k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant \min\{i - v, j - v\}$ and either (A) $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \delta$ and no alignments for $\rho(i - k, j - k), \cdots, \rho(i - 1, j - 1)$ end in a pair, or (B) $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant 0$ and no alignments for $\rho(i-k+1, j-k+1), \ldots, \rho(i - 1, j - 1)$ end in a pair, and some alignment for $\rho(i - k, j - k)$ ends in a pair;

(ii)

$$\rho(i - k, j) + \beta(a_{i-k+1} \cdots a_i),$$

where $1 \leqslant k \leqslant i - v$ and some alignment for $\rho(i - k, j)$ ends in a pair;

(iii)

$$\rho(i, j - k) + \beta(b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant j - v$ and some alignment for $\rho(i, j - k)$ ends in a pair; and

(iv)

$$\rho(i - k, j - l) + \gamma(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j)$$

where $1 \leqslant k \leqslant i - v$, $1 \leqslant l \leqslant j - v$ and some alignment for $\rho(i - k, j - l)$ ends in a pair.

Then

$$h_{1,n} = \min\{\rho(i, j) + \zeta(a_i b_j; a_{i+1} \cdots a_{n-j}) : v \leqslant i, v \leqslant j, \\ i + j \leqslant n - m, \text{ and some alignment for } \rho(i, j) \text{ ends in a pair}\}.$$

Proof. If $\{k \leqslant 1 : \alpha(a_1 \cdots a_k, b_1 \cdots b_k) \leqslant \delta \text{ and } 2k \leqslant n - m\} = \varnothing$, then $\mathscr{S}_{1,n} = \varnothing$ and $h_{1,n} = K$.

Otherwise, v is well defined. The main thing to be proved is that $\rho(i, j)$ is the minimum free energy of alignments of $a_1 \cdots a_i$ and $b_1 \cdots b_j$ subject to $a_1 \cdots a_v$ and $b_1 \cdots b_v$ paired. The only step that can be taken, where the previous $\rho(l, m)$ does not end in a pair, is part (A) of step (i), where $\rho(i - k, j - l)$ cannot have an alignment which ends in a pair. But, $\rho(i - k, j - l)$ *must* have been obtained in steps (ii), (iii), or (iv), where its previous $\rho(l, m)$ did end in a pair. Since the induction began with $\rho(v, v)$, which ends in a pair, all alignments trace back to $\rho(v, v)$ and therefore have $a_1 \cdots a_v$ and $b_1 \cdots b_v$ paired.

That $h_{1,n}$ is the minimum value follows as in Theorem 3.1.

A few cases can be eliminated by the next corollary.

COROLLARY 4.1. *For Theorem* 4.1, *let* $\lambda = \min_{\mathscr{A}^2} \alpha(a, b) < 0$. *Then* $h_{i,j} = 0$ *if*

(i) $i - j + 1 < m + 2(\delta/\lambda)$ *when* $\delta/\lambda \in I$, *or*
(ii) $i - j + 1 < m + 2(\delta/\lambda) + 2$ *when* $\delta/\lambda \notin I$.

Proof. Obviously, the sequence $a_i \cdots a_j$ must be long enough to have a loop of length m and to have pairing of weight $\leqslant \delta$. To have the pairing in the shortest possible time, k pairs are needed, where k is the smallest integer k such that $0 < \delta/\lambda \leqslant k$ (see Corollary 3.1). The result follows as two elements are needed to form each pair.

EXAMPLE. The simple example of Section 3, **a** = *ggguaunnnauaggg-nnncccauannnuauccc*, has the 33 × 33 hairpin matrix shown in Table 4.1. Here there is no harm in letting $K = 0$.

5. FIRST ORDER SECONDARY STRUCTURES

In this section a simple algorithm is given to find the best first order secondary structure from the hairpin matrix H. To review first order secondary structures, a simple theorem is stated. This theorem follows easily from Definition 2.3 and was used in the proof of Theorem 2.8.

THEOREM 5.1. *A secondary structure is first order if and only if* (i) *there is at least one pair and* (ii) *if i is paired with j then there is no more than one loop in* $a_i \cdots a_j$.

TABLE 4.1
Hairpin matrix for a ($K = 0$)

	g	g	g	u	a	u	n	n	n	a	u	a	g	g	g	n	n	n	c	c	c	a	u	a	n	n	n	u	a	u	c	c	c
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	-4	-7
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-5	-4
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	-2	0	0	0	0
a	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	-2	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-6	0	0	0	0	0	0	-3	0	0	0
u	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-6	0	0	0	0	-1	0	0	0	0
a	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0	-3
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0
g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	-3	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0
u	0	0	0	0	-1	0	0	0	0	-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0
a	0	0	0	-3	0	0	0	0	0	0	-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a	0	0	0	-2	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
u	0	0	0	0	-2	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	-4	-5	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	-7	-4	0	0	0	0	0	0	0	0	0	0	-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

It was not necessary to consider joins in first order structures with exactly one loop; but if first order structures have more than one loop, they must have joins. A problem is created by two hairpins separated by no elements of $\mathscr{A}$. This is said to be a join of length 0 and is assigned a free energy of $0 \leqslant \xi(0)$. In general, all joins have been assigned free energy 0 in the literature; but as the addition is not too difficult, it is included here. These considerations will be important in second and higher order secondary structures.

THEOREM 5.2. *For* $\mathbf{a} = a_1 \cdots a_n$, *a sequence from* $\mathscr{A}^n$, *with regular functions* $(p, \alpha, \beta, \gamma, \zeta)$, $\delta > 0$, *and m a positive integer, let* $H = (h_{ij})$ *be the hairpin matrix. Assume* $\xi \geqslant 0$, $\xi(k + l) \leqslant \xi(k) + \xi(l)$, $F_1(0) = 0$, *and inductively define* $F_1(i)$ *to be the minimum of zero and the following quantities for* $1 \leqslant j \leqslant i < n$:

(i)

$$F_1(i - j) + h_{i-j+1,i},$$

if $h_{i-j+1,i} \neq 0$ *and* $F_1(i - j)$ *ends in a (nonempty) join;*

(ii)

$$F_1(i - j) + \xi(0) + h_{i-j+1,i},$$

if $h_{i-j+1,i} \neq 0$ *and* $F_1(i - j)$ *ends in a hairpin;*

(iii)

$$F_1(i - j) + \xi(a_{i-j+1} \cdots a_i),$$

if $h_{i-j+1,i} = 0$ *and* $F_1(i - j)$ *ends in a hairpin.*

Then $F_1(\mathbf{a})$ *is the minimum of the following quantities for* $1 \leqslant j \leqslant n$:

(i)′

$$F_1(n - j) + h_{n-j+1,n},$$

if $h_{n-j+1,n} \neq 0$ *and* $F_1(i - j)$ *ends in a join;*

(ii)′

$$F_1(n - j) + \xi(0) + h_{n-j+1,n},$$

if $h_{n-j+1,n} \neq 0$ *and* $F_1(n - j)$ *ends in a hairpin;*

(iii)′

$$F_1(n - j),$$

if $h_{n-j+1,n} = 0$ *and* $F_1(n - j)$ *ends in a hairpin.*

Proof. If $S(\mathbf{a})$ is a first order secondary structure whose l hairpins have foundations

$$i_{j_1}i_{k_1}; \quad i_{j_2}i_{k_2}; \quad \ldots; \quad i_{j_l}i_{k_l};$$

then

$$F(S(a)) = \sum_{m=1}^{l-1} \left(h_{i_{j_m} i_{k_m}} + \xi(a_{i_{k_m}+1} \cdots a_{i_{j_{m+1}}-1}) \right) + h_{i_{j_1} i_{k_1}}.$$

The proof is by induction. Assume, for $1 \leqslant j \leqslant i-1 < n-1$, that $f(j)$ is the minimum free energy of $a_1 \cdots a_j$. Then the best such structure for $a_1 \cdots a_i$ either is a tail or ends in a join or a hairpin. The number of ways $a_1 \cdots a_i$ can end in a hairpin are those hairpins whose numbers $h_{i-j+1,i} \neq 0$ $(1 \leqslant j \leqslant i-1)$. This handles (i) and (ii) of the theorem. The only other possible cases are where $a_1 \cdots a_i$ ends in a join $a_{i-j+1} \cdots a_i$. If the optimal structure for $a_1 \cdots a_{i-j}$ also ended in a join, say $a_{k+1} \cdots a_{i-j}$, then

$$\begin{aligned} F(i) &= F(k) + \xi(a_{k+1} \cdots a_{i-j}) + \xi(a_{i-j+1} \cdots a_i) \\ &\geqslant F(k) + \xi(a_{k+1} \cdots a_i). \end{aligned}$$

Therefore, all necessary situations are covered by (iii).

To complete the proof, simply note that for $i = n$ ending in a "join" is actually ending in a tail and (iii) becomes (iii)'.

As the join function complicates Theorem 5.2, the theorem is restated with $\xi \equiv 0$. This restatement as a corollary is the practical version of the theorem for first order secondary structures.

COROLLARY 5.1. *Assume* $\xi \equiv 0$ *in Theorem* 5.2. *Then*

$$F_1(a_1 \cdots a_i) = F_1(i) = \min\{F_1(i-j) + h_{i-j+1,i} : 1 \leqslant j \leqslant i\}.$$

EXAMPLE. The example **a** = *ggguaunnnauagggnnncccauannnuauccc* given at the end of Section 3 has $F_1(\mathbf{a}) = -9$ and the structure is shown in Fig. 5.1.

1st hairpin from 23 to 31; hairpin free energy = −3.0:
Base number 23 pairs with base number 31 : *au*
Base number 24 pairs with base number 30 : *ua*
Base number 25 pairs with base number 29 : *au*

2nd hairpin from 14 to 22; hairpin free energy = −3.0:
Base number 14 pairs with base number 22 : *gc*
Base number 15 pairs with base number 21 : *gc*
Base number 16 pairs with base number 20 : *gc*

3rd hairpin from 5 to 13; hairpin free energy = −3.0:
Base number 5 pairs with base number 13 : *ua*
Base number 6 pairs with base number 12 : *au*
Base number 7 pairs with base number 11 : *ua*

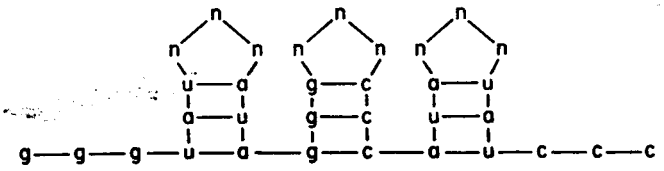


FIG. 5.1. Best first order secondary structure for **a**.

6. SECOND ORDER SECONDARY STRUCTURES

Next, second order secondary structures are considered. The definitions and algorithms of this section are a crucial step in solving the problem of the best Nth order secondary structure. The hairpin matrix of Section 4 and the algorithm of Section 5 for the best first order secondary structure are the tools for the solution obtained in this section.

The first task is to study some graph theory of second order secondary structures. Assume the adjacency matrix of a secondary structure is A. If

$$A = A^{(0)} \to A^{(1)} \to A^{(2)} = (a_{ij}^{(2)}),$$

where $A^{(2)}$ is the first $A^{(i)}$ such that $a_{ij}^{(2)} = 0$ if $i \neq j \pm 1$, then A is second order by Definition 2.3. Consider the matrix $A^{(1)}$ in this sequence. By definition, $A^{(1)}$ is the matrix for some first order secondary structure. Therefore, this first order secondary structure has exactly one loop or more than one loop. The problem of best second order structures will first be solved for the one loop case and then extended to the multiple loop case.

Next, a definition is given to classify the parts of a second order secondary structure in such a way to make the computation of free energy possible. Theorems 2.2 and 2.4 are not adequate for this purpose. A lemma is necessary to justify the definition.

LEMMA 6.1. *Let A be a second order secondary structure and suppose $A = A^{(0)} \to A^{(1)} \to A^{(2)}$. Assume $a_i a_{i+1} \cdots a_j$ is a sequence of unpaired elements in $A^{(1)}$ and that a_{i-1} and a_{j+1} are either paired or not in the sequence. Then $a_i a_{i+1} \cdots a_j$ is zero or first order in $A^{(2)}$.*

Proof. If $a_i a_{i+1} \cdots a_j$ were (at least) second order in $A^{(0)}$, then $a_i a_{i+1} \cdots a_j$ could not be unpaired in $A^{(1)}$.

DEFINITION 6.1. Let A be a second order secondary structure and suppose $A = A^{(0)} \to A^{(1)} \to A^{(2)}$ as in Definition 2.3. For $A^{(1)}$, let LO be the set of loops, LA_0 the set of ladders, IL the set of interior loops, B' the set of bulges, J the set of joins, T the set of tails, and let $B = B' \sim IL \sim J$.

(i) Assume $a_1 \cdots a_k \in LO$. If $a_1 \cdots a_k$ is a loop in $A^{(0)}$, $a_1 \cdots a_k$ is said to be a *zero order loop in* $A^{(0)}$. Otherwise, $a_1 \cdots a_k$ is said to be a *first order loop in* $A^{(0)}$. These relations are written as $a_1 \cdots a_k \in LO_0$ and $a_1 \cdots a_k \in LO_1$, respectively.

(ii) If $a_1 \cdots a_k \in B$, then, if $a_1 \cdots a_k$ is a member of B in $A^{(0)}$, $a_1 \cdots a_k$ is a *zero order bulge in* $A^{(0)}$ or $a_1 \cdots a_k \in B_0$. Otherwise, $a_1 \cdots a_k$ is called a *first order bulge in* $A^{(0)}$, which is written as $a_1 \cdots a_k \in B_1$.

(iii) Assume $(a_1 \cdots a_k, b_1 \cdots b_l) \in IL$. If this pair of sequences is an interior loop in $A^{(0)}$, then the pair is said to be a *zero order interior loop in* $A^{(0)}$. Otherwise, the pair is said to be a *first order interior loop in* $A^{(0)}$. These relations are written as $(a_1 \cdots a_k, b_1 \cdots b_l) \in IL_0$ or $(a_1 \cdots a_k, b_1 \cdots b_l) \in IL_1$, respectively.

(iv) Assume $a_1 \cdots a_k \in J$. If this sequence is a join in $A^{(0)}$, then it is said to be a *zero order join in* $A^{(0)}$. Otherwise, $a_1 \cdots a_k$ is said to be a *first order join in* $A^{(0)}$. These relations are written as $a_1 \cdots a_k \in J_0$ or $a_1 \cdots a_k \in J_1$, respectively.

(v) If $a_1 \cdots a_k \in T$, then if $a_1 \cdots a_k$ is also a tail in $A^{(0)}$, $a_1 \cdots a_k$ is said to be a *zero order tail in* $A^{(0)}$. Otherwise, $a_1 \cdots a_k$ is said to be a *first order tail in* A^0. These relations are written as $a_1 \cdots a_k \in T_0$ and $a_1 \cdots a_k \in T_1$, respectively.

(vi) LA_0 will be called the set of *zero order ladders in* $A^{(0)}$.

(vii) A sequence in $A^{(0)}$ is called a *cloverleaf* if the sequence is a hairpin in $A^{(1)}$ and the loop sequence in $A^{(1)}$ has exactly three loops in $A^{(0)}$.

The following theorem shows that Definition 6.1 adequately describes second order secondary structures.

THEOREM 6.1. *Any second order secondary structure can be uniquely decomposed into the sets* $LO_0, LO_1, B_0, B_1, IL_0, IL_1, J_0, J_1, T_0, T_1,$ *and* LA_0.

Proof. Theorem 2.2 asserts that $A^{(1)}$ can be uniquely decomposed into loops, ladders, bulges, joins, and tails. These ladders are LA_0. Lemma 6.1 and Definition 6.1 allow decomposition of the remaining sets, say Q, into Q_0 and Q_1.

The free energy functions of Definition 3.1 are not adequate to directly define the free energy of a second order secondary structure. The main difficulty is associated with IL_1 and B_1, where it is not entirely clear from the literature how to assign the free energy. The approach taken here is motivated by DeLisi [4].

DEFINITION 6.2. Let $\mathcal{A}$ be an alphabet, ξ a join function, and $(p, \alpha, \beta, \gamma, \zeta)$ regular functions.

(i) A function r is called a replacement function if, for each $(a, b) \in \mathscr{A}^2$ such that $p(a, b) = 1$, $r(a, b)$ is a finite sequence of elements of $\mathscr{A}$.

(ii) It will be useful to let $\xi_0 = \xi$, $\alpha_0 = \alpha$, $\beta_0 = \beta$, $\gamma_0 = \gamma$, and $\zeta_0 = \zeta$.

(iii) Let $a_1 \cdots a_k$ be a sequence in B_1, J_1, LO_1, T_1, or one of the arguments in IL_1. Then, as a sequence in $A^{(0)}$, $a_1 \cdots a_k$ is composed of a tail $a_1 \cdots a_{t_1}$, a hairpin, a join $a_{i_1} \cdots a_{j_1}, \ldots,$ a hairpin, a join $a_{i_n} \cdots a_{j_n}$, a hairpin, and a tail $a_{t_2} \cdots a_k$. Notice that there must be at least one hairpin in this sequence, unless, perhaps, $a_1 \cdots a_k$ is an argument of IL_1. Now define

$$R(a_1 \cdots a_k) = a_1 \cdots a_{t_1} r(a_{t_1+1}, a_{i_1-1}) \cdots a_{i_n} \cdots a_{j_n} r(a_{j_n+1}, a_{t_2-1}) a_{t_2} \cdots a_k,$$

$$\zeta_1(a, b; a_1 \cdots a_k) = \zeta_0(a, b; R(a_1 \cdots a_k)),$$

$$\beta_1(a_1 \cdots a_k) = \beta_0(R(a_1 \cdots a_k)),$$

$$\gamma_1(a_1 \cdots a_k, b_1 \cdots b_l) = \gamma_0(R(a_1 \cdots a_k), R(b_1 \cdots b_l)),$$

where, if $a_1 \cdots a_k$ is unpaired in $A^{(0)}$, $R(a_1 \cdots a_k) = a_1 \cdots a_k$.

Next, the above concepts are used to give a definition of the free energy of a second order secondary structure.

DEFINITION 6.3. Assume **a** is a sequence in $\mathscr{A}^n$, where ξ is a join function, $(p, \alpha, \beta, \gamma, \zeta)$ are regular functions, and r is a replacement function. Let $S(\mathbf{a})$ be a secondary structure of no more than second order. Then the *second order free energy associated with* $S(\mathbf{a})$ is defined by

$$F_2(S(a)) = \sum_{w_0 \in LO_0} \zeta_0(w_0) + \sum_{w_1 \in LO_1} \zeta_1(w_1) + \sum_{t_1 \in T_1} F_1(t_1) + \sum_{v_0 \in J_0} \xi_0(v_0) + \sum_{v_1 \in J_1} \xi_1(v_1) + \sum_{x_0 \in B_0} \beta_0(x_0) + \sum_{x_1 \in B_1} \beta_1(x_1) + \sum_{x_0 \in IL_0} \gamma_0(y_0) + \sum_{y_1 \in IL_1} \gamma_1(y_1) + \sum_{z_0 \in LA_0} \alpha_0(z_0).$$

Of course, $F_2(S(\mathbf{a})) = 0$ if $S(\mathbf{a})$ has no pair. The *second order free energy for* **a** is given by

$$F_2(\mathbf{a}) = \min_{S(\mathbf{a}) \in \mathscr{S}} F_2(S(\mathbf{a})),$$

where $\mathscr{S} = \{S(\mathbf{a}): S(\mathbf{a})$ is of no more than second order, w a loop in $S(\mathbf{a})$ implies w has no more than m elements, and z a ladder in $S(\mathbf{a})$ implies $\alpha(z) \leqslant \delta\}$. If $\mathscr{S}^* = \mathscr{S} \cap \{S(\mathbf{a}): A^{(1)}$ has no more than one loop$\}$, then let

$$F_2^*(\mathbf{a}) = \min_{S(\mathbf{a}) \in \mathscr{S}^*} F_2(S(\mathbf{a})).$$

Remark. It is easy to see from Definitions 3.2 and 3.3 that $F_2(\mathbf{a})$ and $F_2{}^*(\mathbf{a})$ can be found from the functions $r, \xi, p, \alpha, \beta, \gamma, \zeta$, and a procedure for calculating $F_1(S(\mathbf{b}))$ for any sequence b.

The next theorem generalizes the algorithm for $F_1{}^*(\mathbf{a})$ to $F_2{}^*(\mathbf{a})$.

THEOREM 6.2. *Let $a_1 \cdots a_n$ be a sequence from $\mathscr{A}^n$ and assume the functions $(p, \alpha, \beta, \gamma, \zeta)$ are regular, ξ is a join function, r is a replacement function, and δ and m satisfy the conditions of the definition of $F_2{}^*$. Define*

$$b_j = a_{n-j+1}, \qquad 1 \leqslant j \leqslant n,$$

and $\rho(0, j) = \rho(i, 0) = 0$ for $0 \leqslant i, j \leqslant n$. Then, for $0 < ij$ and $i + j \leqslant n - m$, inductively define $\rho(i, j)$ to be the minimum of zero and:

(i)

$$\rho(i-k, j-k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant \min\{i, j\}$ and $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \delta$;

(ii)

$$\rho(i-k, j) + \min_{v=0,1} \beta_v(a_{i-k+1} \cdots a_i);$$

where $1 \leqslant k \leqslant i-1$ and some alignment for $\rho(i-k, j)$ ends in a pair;

(iii)

$$\rho(i, j-k) + \min_{v=0,1} \beta_v(b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant j-1$ and some alignment for $\rho(i, j-k)$ ends in a pair, and

(iv)

$$\rho(i-k, j-l) + \min_{v=0,1} \gamma_v(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j),$$

where $1 \leqslant k \leqslant i-1$, $1 \leqslant l \leqslant j-1$, and some alignment for $\rho(i-k, j-l)$ ends in a pair.

Then

$$F_2{}^*(\mathbf{a}) = \min\{\rho(i, j) + \min_{v=0,1} \zeta_v(a_i b_j; a_{i+1} \cdots a_{n-j})$$

$$: i + j \leqslant n - m \text{ and some alignment for } \rho(i, j) \text{ ends in a pair}\},$$

where $F_2{}^*(a) = 0$ if the above set is empty.

Proof. The proof proceeds exactly as in Theorem 3.1 as soon as the inequalities of $(p, \alpha, \beta, \gamma, \zeta)$ regular are shown to hold with subscripts of 0

and 1 inserted in a consistent manner. For example, to show the inequality $\beta(k+l) \leqslant \beta(k) + \beta(l)$, it may be necessary to show that

$$\beta_1(k+l) \leqslant \beta_1(k) + \beta_0(l).$$

The difficulty is in inserting the proper choice of 0 or 1 as subscripts. This is done below in a case-by-case fashion for the steps of Theorem 3.1.

The first situation is to assume $\rho(i,j)$ is the minimum free energy for alignments of $a_1 \cdots a_i$ and $b_1 \cdots b_j$. If $A = A^{(0)}$ is the adjacency matrix for the best second order structure with the structure for $A^{(1)}$ having exactly one loop, let $a_i b_j$ be the foundation of that loop. The free energy for $a_1 \cdots a_i$ and $b_1 \cdots b_j$ in this structure satisfies $\rho(i,j) \leqslant e < 0$. As before, assume $\rho(i,j) < e$. Suppose first that

$$F_2^*(a) = e + \zeta_{v_3}(a_i b_j; a_{i+1} \cdots a_{n-j})$$

and

$$\rho(i,j) = \rho(i-k, j-l) + \gamma_{v_1}(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j).$$

Then

$$\begin{aligned}\rho(i-k, j-l) + \zeta_{v_2}(a_{i-k} b_{j-l}; a_{i-k+1} \cdots a_{n-j+1}) \\ \leqslant \rho(i,j) + \zeta_{v_3}(a_i b_j; a_{i+1} \cdots a_{n-j}) < F_2^*(\mathbf{a}),\end{aligned}$$

where the structure for ζ_{v_2} is formed by "addition" of the structures for ζ_{v_1} and ζ_{v_3}. Therefore, $v_2 = 1$ if and only if $\max\{v_1, v_3\} = 1$. The first inequality holds since $\zeta_0(l+q+n) \leqslant \gamma_0(l,n) + \zeta_0(q)$ holds for the replacement bulges and loops.

To complete the first situation, assume $F_2^*(\mathbf{a})$ is as above and suppose

$$\rho(i,j) = \rho(i-k, j) + \beta_{v_1}(a_{i-k+1} \cdots a_i).$$

Then,

$$\begin{aligned}\rho(i-k, j) + \zeta_{v_2}(a_{i-k} b_j; a_{i-k+1} \cdots a_{n-j}) \\ \leqslant \rho(i,j) + \zeta_{v_3}(a_i b_j; a_{i+1} \cdots a_{n-j}) < F_2^*(a).\end{aligned}$$

Again, $v_2 = \max\{v_1, v_3\}$.

To show $\rho(i,j)$ had the optimal property desired, three cases were considered in Theorem 3.1. Case (i) proceeds without change. For Case (ii), assume the minimum free energy for $a_1 \cdots a_i$ and $b_1 \cdots b_j$ has the form

$$r^*_{(i-j)} = r(i, j-k) + \beta_{v_3}(b_{j-k+1} \cdots b_j),$$

and $\rho(i, j-k) < r(i, j-k)$. First, suppose an alignment for $\rho(i, j-k)$ is of the form

$$\rho(i, j-k) = \rho(i-l, j-k) + \beta_{v_1}(a_{i-l+1} \cdots a_i).$$

Then ($v_2 = \max\{v_1, v_3\}$)

$$\rho(i-l, j-l) + \gamma_{v_2}(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \rho(i, j-k) + \beta_{v_3}(b_{j-k+1} \cdots b_j) < r^*_{ij}.$$

Next, suppose $\rho(i, j-k)$ is of the form

$$\rho(i, j-k-l) + \beta_{v_1}(b_{j-k-l+1} \cdots b_{j-k}).$$

Then, with $v_2 = \max\{v_1, v_3\}$,

$$\rho(i, j-k-l) + \beta_{v_2}(b_{j-k-l+1} \cdots b_j) \leqslant \rho(i, j-k) + \beta_{v_1}(b_{j-k+1} \cdots b_j) < r^*(i, j).$$

The last situation for case (ii) is

$$\rho(i, j-k) = \rho(i-l, j-k-q) + \gamma_{v_1}(a_{i-l+1} \cdots a_i, b_{j-k-q+1} \cdots b_{j-k}).$$

Then with v_2 as usual,

$$\rho(i-l, j-k-q) + \gamma_{v_2}(a_{i-l+1} \cdots a_i, b_{j-k-q+1} \cdots b_j) \leqslant \rho(i, j-k) + \beta_{v_3}(b_{j-k+1} \cdots b_j) < r^*(i, j).$$

Now, proceed to case (iii), where

$$r^*(i, j) = r(i-l, j-k) + \gamma_{v_3}(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

and assume $\rho(i-l, j-k) < r(i-l, j-k)$. First, assume an alignment for $\rho(i-l, j-k)$ such that

$$\rho(i-l, j-k) = \rho(i-l-q, j-k) + \beta_{v_1}(a_{i-l-q+1} \cdots a_{i-l}).$$

Then

$$\rho(i-l-q, j-k) + \gamma_{v_2}(a_{i-l-q+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \rho(i-l, j-k) + \gamma_{v_3}(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j) < r^*(i, j).$$

The last situation is for

$$\rho(i-l, j-k) = \rho(i-l-q, j-k-p) + \gamma_{v_1}(a_{i-l-q+1} \cdots a_{i-l}, b_{j-k-p+1} \cdots b_{j-k}).$$

Then

$$\rho(i-l-q, j-k-p) + \gamma_{v_2}(a_{i-l-q+1} \cdots a_i, b_{j-k-p+1} \cdots b_j) \leqslant \rho(i-l, j-k) + \gamma_{v_3}(a_{i-l+1} \cdots a_i, b_{j-k+1} \cdots b_j) < r^*(i, j).$$

This completes the proof.

It is possible to obtain results analogous to Corollaries 3.1 and 3.2 and to Theorem 3.2. Those results were to reduce computation in $F_1^*(\mathbf{a})$. But for $F_2^*(\mathbf{a})$ much more serious difficulties are encountered, and a discussion to

make computation feasible will be given later in this section. Presently, the discussion will be aimed at algorithms for $F_2(\mathbf{a})$.

The first task is to generalize the hairpin matrix H.

DEFINITION 6.4. Let the sequence $\mathbf{a} = a_1 \cdots a_n$ be given along with regular functions $(p, \alpha, \beta, \gamma, \zeta)$, a join function ξ, and a replacement function r. Define *the second order hairpin matrix* $H^{(2)}$ to be the symmetric $n \times n$ matrix with $h_{ii}^{(2)} = 0$ for $1 \leqslant i \leqslant n$ and, for $1 \leqslant i < j \leqslant n$,

$$h_{ij}^{(2)} = \min_{S \in \mathscr{S}_{ij}} F_2(S),$$

where

$\mathscr{S}_{ij} = \{S(a_i \cdots a_j) : S = S(a_i \cdots a_j)$ is no more than second order, w a loop implies w has at least m elements, z a ladder implies $\alpha(z) \leqslant \delta < 0$. If S is first order, S has exactly one loop and $p(a_i, a_j) = 1$. If S is second order, the structure corresponding to $A^{(1)}$ has exactly one loop and $p(a_i, a_j) = 1\}$.

If $\mathscr{S}_{ij} = \varnothing$, $h_{ij}^{(2)} = K$.

Theorem 4.1 handled the algorithm for computing H. The modification of Theorem 4.1 is easily made for $H^{(2)}$.

THEOREM 6.3. *Theorem* 4.1 *holds for* $H^{(2)}$ *if* $\tau \in \{\beta, \gamma, \zeta\}$ *is replaced by* $\min\{\tau_0, \tau_1\}$ *and* $h_{1,n}$ *is replaced by* $h_{1,n}^{(2)}$.

It is also easy to generalize Theorem 5.2 and Corollary 5.1. Only the generalization of the corollary is given here.

THEOREM 6.4. *Assume* $\xi \equiv 0$ *in Theorem* 6.2. *Then*

$$F_2(a_1 \cdots a_i) = F_2(i) = \min\{F_2(i-j) + h_{i-j+1,i}^{(2)} : 1 \leqslant j \leqslant i\}.$$

Theorem 6.4 completes the general discussion of the best second order structure. The next section will handle the problem of best Nth order secondary structures.

Now the computational problems associated with Theorem 6.2 are taken up. Results corresponding to Corollary 3.2 and Theorem 3.2 no longer hold for second order structures. Even more important is the computation of β_1 and γ_1. Attention is now restricted to a smaller class of first order bulges and interior loops.

DEFINITION 6.5. Let $\mathscr{S}^{**} = \mathscr{S}^* \cap \{S(a) : b_1 \cdots b_k$ a bulge, join, or tail in $A^{(1)}$ implies $b_1 \cdots b_k$ has no more than one loop in $A^{(2)}\}$. Then define F_2^{**} by

$$F_2^{**}(a) = \min_{S(a) \in \mathscr{S}^{**}} F_2(S(a)).$$

An $n \times n$ symmetric matrix H^β, called the *hairpin bulge matrix*, is defined by

$$h_{ij}^\beta = \min\{h_{kl} + \beta_1(a_i \cdots a_j) : i \leqslant k \leqslant l \leqslant j\}.$$

Similarly, an $(n \times n)^2$ symmetric matrix H^γ called the *hairpin interior loop matrix* is defined by

$$h^\gamma(i, j; k, l) = \min\{h_{i'j'} + h_{k'l'} + \gamma_1(a_i \cdots a_j, a_k \cdots a_l): i \leqslant i' \leqslant j' \leqslant j \text{ and } k \leqslant k' \leqslant l' \leqslant j\}.$$

Computation of H^β is now considered. The assumptions on β are r are consistent with the literature.

THEOREM 6.5. *Let the assumptions of Theorem* 6.2 *hold. Assume*

$$\beta(c_1 \cdots c_k) = \beta(k) \quad \textit{independent of} \quad c_1 \cdots c_k,$$

$$\beta \quad \text{is a strictly increasing function,}$$

and

$$r(a, b) = r \quad \textit{for all} \quad (a, b) \quad \textit{such that} \quad p(a, b) = 1.$$

Then, if $\min_{i \leqslant k \leqslant l \leqslant j} h_{kl} = h_{i_0 j_0}$, (i) *and* (ii) *below hold*:

(i) If

$$h_{i_0 j_0} + \beta(r + j + i_0 - i - j_0) - \min_{\substack{i \leqslant k \leqslant l \leqslant j \\ (k,l) \neq (i_0, j_0)}} h_{kl} - \beta(r) \leqslant 0,$$

then

$$h_{ij}^\beta = h_{i_0 j_0} + \beta(r + j + i_0 - i - j_0).$$

(ii) *If*

$$h_{ij}^\beta = h_{i'j'} + \beta(r + j + i' - i - j'),$$

then

$$j_0 - i_0 \leqslant j' - i'.$$

Proof. (i) The hairpin weight for $i_0 \cdots j_0$ is

$$h_{i_0 j_0} + \beta(r + j + i_0 - i - j_0),$$

and the best possible weight for any remaining structure is

$$\min_{(k,l) \neq (i_0, j_0)} h_{kl} + \beta(r) = h' + \beta(r).$$

The conclusion of (i) holds if

$$h_{i_0j_0} + \beta(r + j + i_0 - i - j_0) \leqslant h' + \beta(r).$$

(ii) Assume the conclusion is false. Then

$$j' - i' < j_0 - i_0,$$

or

$$i_0 - j_0 < i' - j',$$

and

$$\beta(r + j - i + i_0 - j_0) < \beta(r + j - i + i' - j').$$

Then

$$h_{i_0j_0} + \beta(r + j - i + i_0 - j_0) < h_{i'j'} + \beta(r + j - i + i' - j'),$$

which is a contradiction.

Next it is seen that if β is linear for arguments greater than or equal to r then H^β can easily be computed.

THEOREM 6.6. *Let the assumptions of Theorem* 6.5 *hold and suppose*

$$\beta(r + k) = \beta(r) + s_\beta k \qquad \text{where} \quad s_\beta > 0.$$

Then

$$h^\beta_{i,\,j+1} = \min\{h^\beta_{ij} + s_\beta;\, h_{k,\,j+1} + \beta(r + k - i) : h_{k,\,j+1} < 0,\, i \leqslant k \leqslant j\}.$$

Proof. Let

$$\begin{aligned} h^\beta_{i,j} &= h_{i_0j_0} + \beta(r + j - i + i_0 - j_0) \\ &= h_{i_0j_0} + \beta(r) + s_\beta(j - i + i_0 - j_0). \end{aligned}$$

Now, if

$$\begin{aligned} h^\beta_{i,\,j+1} &= h_{i'j'} + \beta(r + 1 + j - i + i' - j') \\ &= h_{i'j'} + \beta(r) + s_\beta(j - i + i_0 - j_0) + s_\beta, \end{aligned}$$

where $i \leqslant i' \leqslant j' \leqslant j$, then $i' = i_0$ and $j' = j_0$. The other possibilities are covered by $j' = j + 1$ and are included in the above minimization.

It is now clear that H^β can easily and efficiently be computed. However, H^γ is a very large computational job. By making another linearity assumption, H^γ can be computed from the algorithm for H^β.

THEOREM 6.7. *Let the assumptions of Theorem* 6.2 *hold and assume*

$\gamma(c_1 \cdots c_k, d_1 \cdots d_l) = \gamma(k + l)$ *independent of* $c_1 \cdots c_k$ *and* $d_1 \cdots d_l$;
γ *is a strictly increasing function in each argument*;
$r(a, b) = r$ *for* (a, b) *such that* $p(a, b) = 1$; *and*
$\gamma(r + k) = \gamma(r) + s_\gamma k$ *where* $s_\gamma > 0$.

Then, if $h_{ij}^{\beta *} = \min\{h_{i'j'} + s_\gamma(j - i + i' - j') : h_{i'j'} < 0\}$,

$$h^\gamma(i, j; k, l) = \min\{\gamma(r) + s_\gamma r + h_{ij}^{\beta *} + h_{kl}^{\beta *}, \gamma(r) + s_\gamma(l - k + 1) + h_{ij}^{\beta *}, \gamma(r) + s_\gamma(i - j + 1) + h_{kl}^{\beta *}, \gamma(j - i + 1 + l - k + 1)\}.$$

Proof. Assume $h^\gamma(i, j; k, l)$ has a hairpin in $a_i \cdots a_j$ or $a_k \cdots a_l$. Then, in the case both have hairpins,

$$\begin{aligned} h^\gamma(i, j; k, l) &= h_{i_0 j_0} + h_{k_0 l_0} + \gamma_1(a_i \cdots a_j, a_k \cdots a_l) \\ &= h_{i_0 j_0} + h_{k_0 l_0} + \gamma(2r + j - i + i_0 - j_0 + l - k + k_0 - l_0) \\ &= h_{i_0 j_0} + h_{k_0 l_0} + \gamma(r) + s_\gamma(r + j - i + i_0 - j_0 + l - k + k_0 - l_0) \\ &= \gamma(r) + s_\gamma r + (h_{i_0 j_0} + s_\gamma(j - i + i_0 - j_0)) \\ &\quad + (h_{k_0 l_0} + s_\gamma(l - k + k_0 - l_0)). \end{aligned}$$

In case exactly one has a hairpin, say $a_i \cdots a_j$,

$$\begin{aligned} h^\gamma(i, j; k, l) &= h_{i_0 j_0} + \gamma_1(a_i \cdots a_j, a_k \cdots a_l) \\ &= h_{i_0 j_0} + \gamma((r + j - i + i_0 - j_0) + (l - k + 1)) \\ &= \gamma(r) + (h_{i_0 j_0} + s_\gamma(j - i + i_0 - j_0)) + (0 + s_\gamma(l - k + 1)). \end{aligned}$$

In case neither have hairpins,

$$h^\gamma(i, j; k, l) = \gamma_1(a_i \cdots a_j, a_k \cdots a_l) = \gamma(j - i + 1 + l - k + 1).$$

Now H^β can be used in steps (ii) and (iii) of Theorem 6.2 and H^γ in step (iv) to compute $F_2^{**}(\mathbf{a})$.

EXAMPLE. The example $\mathbf{a} = ggguaunnnauagggnnncccauannnuauccc$ given at the end of Section 3 and 5 has $F_2{}^*(\mathbf{a}) = -12$ if $R(a_1 \cdots a_k) = 1$. (That is, any hairpin is replaced by a sequence of length 1.) The structure for $F_2{}^*(\mathbf{a})$ is a cloverleaf and is shown in Fig. 6.1.

Ladder regions for the second order hairpin:

Base number	2 pairs with base number 34	:	*gc*
Base number	3 pairs with base number 33	:	*gc*
Base number	4 pairs with base number 32	:	*gc*

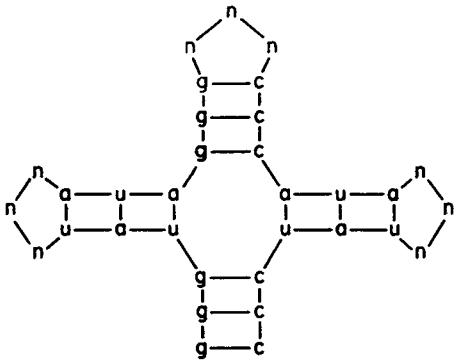


FIG. 6.1. Best second order secondary structure for **a**.

The loop has a hairpin from 5 to 13,

Base number 5 pairs with base number 13 : *ua*
Base number 6 pairs with base number 12 : *au*
Base number 7 pairs with base number 11 : *ua*

another hairpin from 23 to 31,

Base number 23 pairs with base number 31 : *au*
Base number 24 pairs with base number 30 : *ua*
Base number 25 pairs with base number 29 : *au*

and a final hairpin from 14 to 22.

Base number 14 pairs with base number 22 : *gc*.
Base number 15 pairs with base number 21 : *gc*.
Base number 16 pairs with base number 20 : *gc*.

7. *N*th Order Secondary Structures

Finally, *N*th order secondary structures are considered. This section is a direct generalization of Section 6. The structures of this section are not easy to visualize, but hopefully the work on second order structures provides a natural motivation. Also, at the conclusion of this section, a result is given to characterize the solution of the best secondary structure (of any order) for a given sequence. Due to the previous work of Section 6, the results of this section will be briefly stated and proofs omitted. It is assumed throughout that $N \geqslant 2$.

First, the components of an *N*th order structure are classified.

DEFINITION 7.1. Let A be the adjacency matrix for an Nth order secondary structure and suppose $A = A^{(0)} \to A^{(1)} \to \cdots \to A^{(N-1)} \to A^{(N)}$ as in Definition 2.3. $A^{(N-1)}$ is first order and let LO be the set of loops, LA_0 the set of ladders, IL the set of interior loops, B' the set of bulges, J the set of joins, and T the set of tails, and let $B = B' \sim IL$.

(i) Assume $a_1 \cdots a_k \in LO$. If $a_1 \cdots a_k$ is a loop in $A^{(0)}$, $a_1 \cdots a_k$ is a *zero order loop in* $A^{(0)}$. Otherwise, $a_1 \cdots a_k$ is said to be an $(N-1)$*st order loop in* $A^{(0)}$. These relations are written as $a_1 \cdots a_k \in LO_0$ or $a_1 \cdots a_k \in LO_{N-1}$, respectively.

(ii) If $a_1 \cdots a_k \in B$, then, if $a_1 \cdots a_k$ is a bulge in $A^{(0)}$, $a_1 \cdots a_k$ is a *zero order bulge in* $A^{(0)}$ or $a_1 \cdots a_k \in B_0$. Otherwise, $a_1 \cdots a_k$ is called an $(N-1)$*st order bulge in* $A^{(0)}$ or $a_1 \cdots a_k \in B_{N-1}$.

(iii) Assume $(a_1 \cdots a_k, b_1 \cdots b_l) \in IL$. If the pair is an interior loop in $A^{(0)}$, it is said to be a *zero order interior loop in* $A^{(0)}$. Otherwise, the pair is said to be an $(N-1)$*st order interior loop in* $A^{(0)}$. The relations are written as $(a_1 \cdots a_k, b_1 \cdots b_l) \in IL_0$ or $(a_1 \cdots a_k, b_1 \cdots b_l) \in IL_{N-1}$, respectively.

(iv) Assume $a_1 \cdots a_k \in J$. If this sequence is a join in $A^{(0)}$, it is said to be a *zero order join in* $A^{(0)}$. Otherwise, it is said to be an $(N-1)$*st order join in* $A^{(0)}$. The relations are written as $a_1 \cdots a_k \in a_k \in J_0$ or $a_1 \cdots a_k \in J_{N-1}$, respectively.

(v) If $a_1 \cdots a_k \in T$ and $a_1 \cdots a_k$ is also a tail in $A^{(0)}$, then $a_1 \cdots a_k$ is said to be a *zero order tail in* $A^{(0)}$. Otherwise, $a_1 \cdots a_k$ is said to be an $(N-1)$*st order tail in* $A^{(0)}$. The relations are written as $a_1 \cdots a_k \in T_0$ or $a_1 \cdots a_k \in T_{N-1}$, respectively.

(iv) LA_0 will be called the set of *zero order ladders in* $A^{(0)}$.

THEOREM 7.1. *Any Nth order secondary structure can be uniquely decomposed into the sets LO_0, LO_{N-1}, B_0, B_{N-1}, IL_0, IL_{N-1}, J_0, J_{N-1}, T_0, T_{N-1}, and LA_0.*

It is clear that for $N > 2$, J_{N-1}, for example, could be further decomposed. However, this will not be necessary for the algorithm given below and is therefore omitted.

The functions ζ_1, β_1, and γ_1 are extended in the following manner. Let $a_1 \cdots a_k$ be a sequence in B_{N-1}, say. Then define

$$\beta_{N-1}(a_1 \cdots a_k) = \beta_{N-2}(R(a_1 \cdots a_k)),$$

where R is defined on $a_1 \cdots a_k$ considered as a sequence in $A^{(N-2)}$. The remaining functions ζ_{N-1} and γ_{N-1} are similarly defined.

DEFINITION 7.2. Let $\mathbf{a} \in \mathscr{A}^N$ where ξ is a join function, p, α, β, γ, ζ are regular functions, and r is a replacement function. Let $S(\mathbf{a})$ be a secondary

structure of no more than Nth order. Then the *Nth order free energy for $S(\mathbf{a})$* is defined by

$$F_N(S(\mathbf{a})) = \sum_{w_0 \in LO_0} \zeta_0(w_0) + \sum_{w_N \in LO_{N-1}} \zeta_N(w_N) + \sum_{t \in T_{N-1}} F_{N-1}(t)$$
$$+ \sum_{v_0 \in J_0} \xi_0(v_0) + \sum_{v_N \in J_{N-1}} \xi_N(v_N) + \sum_{x_0 \in B_0} \beta_0(x_0) + \sum_{x_N \in B_{N-1}} \beta_N(x_N)$$
$$+ \sum_{y_0 \in IL_0} \gamma_0(y_0) + \sum_{y_N \in IL_{N-1}} \gamma_N(y_N) + \sum_{z_0 \in LA_0} \alpha_0(z_0).$$

If $S(\mathbf{a})$ has order less than N, let $S(\mathbf{a}) \in T_{N-1}$. The *$N$th order free energy for* $\mathbf{a}$ is given by

$$F_N(\mathbf{a}) = \min_{S \in \mathscr{S}} F_N(S),$$

where $\mathscr{S} = \{S(\mathbf{a}): S(\mathbf{a})$ is no more than Nth order, w a loop in $S(\mathbf{a})$ implies w has no more than m elements, and z a ladder in $S(\mathbf{a})$ implies $\alpha(z) \leqslant \delta\}$. If $\mathscr{S}^* = \mathscr{S} \cap \{S(\mathbf{a}): A^{(N-1)}$ has no more than one loop$\}$, then define

$$F_N{}^*(\mathbf{a}) = \min_{S \in \mathscr{S}^*} F_N(S).$$

The proof of the next theorem follows that of Theorem 6.2.

THEOREM 7.2. *Let $\mathbf{a} \in \mathscr{A}^n$ and assume p, α, β, γ, ζ are regular, ξ is a join function, r is a replacement function, and δ, m are as in the definition of $F_N(\mathbf{a})$. Define*

$$b_j = a_{n-j+1}, \qquad 1 \leqslant j \leqslant n,$$
$$\rho(0, j) = \rho(i, 0) = 0, \qquad 0 \leqslant i, j \leqslant n.$$

Then inductively define $\rho(i, j)$ for $0 < ij$, $i + j \leqslant n - m$ to be the minimum of zero and:

(i)

$$\rho(i - k, j - k) + \alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant \min\{i, j\}$ and $\alpha(a_{i-k+1} \cdots a_i, b_{j-k+1} \cdots b_j) \leqslant \delta$;

(ii)

$$\rho(i - k, j) + \min_{v=0,N} \beta_v(a_{i-k+1} \cdots a_i),$$

where $1 \leqslant k \leqslant i - 1$ and some alignment for $\rho(i - k, j)$ ends in a pair;

(iii)

$$\rho(i, j - k) + \min_{v=0,N} \beta_v(b_{j-k+1} \cdots b_j),$$

where $1 \leqslant k \leqslant j-1$ *and some alignment for* $\rho(i, j-k)$ *ends in a pair*; and

(iv)

$$\rho(i-k, j-l) + \min_{v=0,N} \gamma_v(a_{i-k+1} \cdots a_i, b_{j-l+1} \cdots b_j),$$

where $1 \leqslant k \leqslant i-1, 1 \leqslant l \leqslant j-1$, *and some alignment for* $\rho(i-k, j-l)$ *ends in a pair.*

Then

$$F_N^*(\mathbf{a}) = \min\{\rho(i,j) + \min_{v=0,N} \zeta_v(a_i b_j; a_{i+1} \cdots a_{n-j}):$$
$$i + j \leqslant n - m \text{ and some alignment for } \rho(i,j) \text{ ends in a pair}\},$$

where $F_N^*(\mathbf{a}) = 0$ *if the above set is empty.*

DEFINITION 7.3. Define the *Nth order hairpin matrix* $H^{(N)}$ to be the symmetric $n \times n$ matrix with

$$h_{ij}^{(N)} = \min_{S \in \mathscr{S}_{ij}} F_N(S), \qquad 1 \leqslant i \leqslant j' \leqslant n,$$

where $\mathscr{S}_{ij} = \{S(a_i \cdots a_j) = S : S$ is no more than Nth order, if w is a loop then w has at least m elements, if z is a ladder then $\alpha(z) \leqslant \delta < 0$. If S is first order, S has exactly one loop and $p(a_i, a_j) = 1$. If S is order $O > 1$, the structure corresponding to $\mathscr{A}^{(O-1)}$ has exactly one loop and $p(a_i, a_j) = 1$. If $\mathscr{S}_{ij} = \varnothing, h_{ij}^{(N)} = 0.\}$

THEOREM 7.3. *Theorem* 4.1 *holds for* $H^{(N)}$ *if* $\tau \in \{\beta, \gamma, \zeta\}$ *is replaced by* $\min\{\tau_0, \tau_N\}$ *and* h_{1n} *is replaced by* $h_{1n}^{(N)}$.

THEOREM 7.4. *Assume* $\xi = 0$ *in Theorem* 7.2. *Then*

$$F_N(a_1 \cdots a_i) = F_N(i) = \min\{F_N(i-j) + h_{i-j+1,i}^{(N)} : 1 \leqslant j \leqslant i\}.$$

Finally, the order of the best structure for $\mathbf{a}$ is characterized in terms of the algorithm.

THEOREM 7.5. *The order* N *of* $S(\mathbf{a})$ *such that* $F_N(S(\mathbf{a})) = \min_M F_M(\mathbf{a})$ *satisfies*

$$N = \min\{M : H^{(M)} = H^{(M+1)}\}.$$

8. CONCLUSION

A general algorithm for the evaluation of free energy can be obtained from the above work. (i) First, set $f = 0$. (ii) Then evaluate all hairpins in

the structure and add the free energy to f. (iii) Replace the evaluated hairpins using the replacement functions. (iv) If the new structure is unpaired, f is the free energy of the original structure. Otherwise go to (ii).

If replacement functions are found to be the wrong approach, then (iii) can be altered. For example, it might be necessary to have a loop function of order 1, 2, ... The approach used in this paper, then, is easy to generalize to fit more accurate models. It does seem entirely adequate for current information.

A computer program is being developed, and a preliminary version was used for the naive example given above. As an example of its power, a portion of the first order loop in the Min Jou flower model [18] is examined using the same (combinatorial) energy assignments given in the earlier example. The sequence of length 68 is

$$\mathbf{a} = ucaaacgacgcuaacgacucccuuagcccaaagg$$

$$\cdot\, uagaaaauccucuggaacguaacggaauuguuau,$$

which composes a first order structure of three loops in the flower model. There are 21 bases in the joins for this structure, and 16 base pairs (including two *gu* pairs and two ladders with only one base pair). The algorithm (with parameters as specified earlier) predicts a best single loop hairpin of $F_1^*(a) = -12$ with 12 bases in the tails, and 17 base pairs with *no gu* pairs. The best first order structure of $F_1(a) = -12$ has 12 bases in tails and joins and 17 base pairs. Of course, more precise energy functions would make the comparison more realistic, but the two structures given above seem to improve on that portion of the flower model. The three structures are displayed in Fig. 8.1.

A study is being planned in which good estimates of the energy functions will be used to study a set of tRNAs. Of course, that work will be useful in a final evaluation of this paper, and the results will be compared with other studies. We conjecture that, except for minor modifications, the algorithms proposed are the most efficient possible for determination of secondary structure on computers.

While the ladder function α given in this paper is strictly additive, it is important to note that it can be modified to handle general α. In particular, nearest neighbor effects [24] can be treated by adding a (negative) free energy for adjacent bonds. This allows the use of the best estimates of the energy functions.

Finally, the connection between the graph theory of secondary structure of single stranded nucleic acids and that for secondary structure in single stranded proteins is currently being considered.

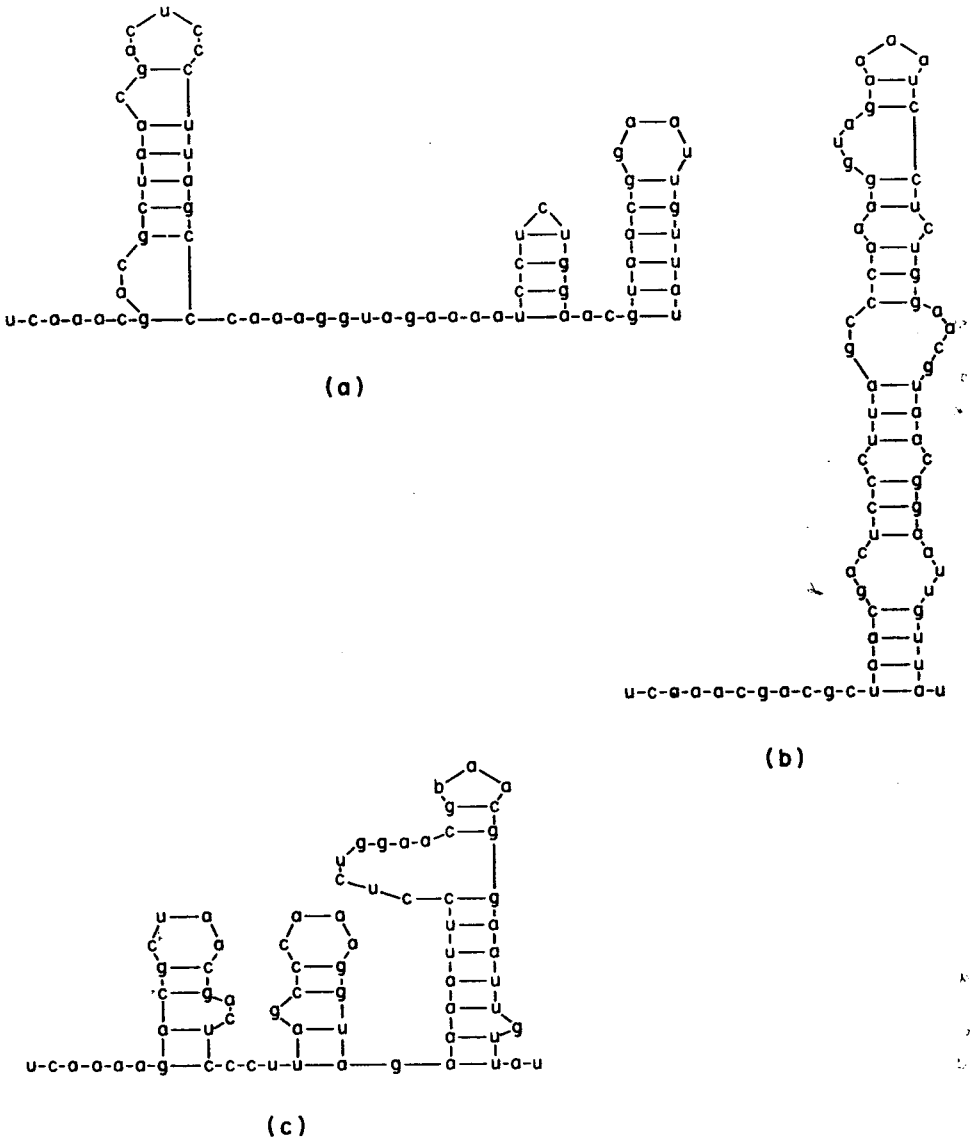


FIG. 8.1. Secondary structures for a portion of the coat protein gene. (a) Structure appearing in the flower model, (b) best hairpin, and (c) best first order structure.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Charles DeLisi of the National Cancer Institute who initially proposed the problem of secondary structure to me in such clear and straightforward terms and who provided much insight, encouragement, and advice during the course of this investigation. Thanks are also due William Beyer and Paul Stein of Los Alamos Scientific Laboratory and Temple F. Smith of Northern Michigan University for helpful discussions.

REFERENCES

1. V. A. BLOOMFIELD, D. M. CROTHERS, AND I. TINOCO, "Physical Chemistry of Nucleic Acids," Harper, New York, 1974.
2. N. G. DEBRUIJN AND P. ERDOS, Some linear and some quadratic recursion formulas II, *Indag. Math.* **55** (1952), 152–163.
3. C. DELISI AND D. M. CROTHERS, Prediction of RNA secondary structure, *Proc. Nat. Acad. Sci. U.S.A.* **68** (1971), 2682–2685.
4. C. DELISI, Conformational changes in transfer RNA. I. Equilibrium theory, *Biopolymers* **12** (1973), 1713–1728.
5. W. FELLER, "An Introduction to Probability Theory and Its Applications. Volume I" (3rd ed.), Wiley, New York 1967.
6. T. R. FINK AND D. M. CROTHERS, Free energy of imperfect nucleic acid helices. I. The bulge defect, *J. Mol. Biol.* **66** (1972), 1–12.
7. W. M. FITCH, Considerations regarding the regulation of gene transcription and messenger translation, *J. Mol. Evol.* **1** (1972), 185–207.
8. W. M. FITCH, The large extent of putative secondary nucleic acid structure in random nucleotide sequences or amino acid derived messenger-RNA, *J. Mol. Evol.* **3** (1974), 279–291.
9. J. R. FRESCO, B. M. ALBERTS, AND P. DOTY, Some molecular details of the secondary structure of ribonucleic acid, *Nature* **188** (1960), 98–101.
10. J. R. FRESCO, L. C. KLOTZ, AND E. G. RICHARDS, A new spectroscopic approach to the determination of helical secondary structure in ribonucleic acids, *Cold Spring Harbor Symp. Quant. Biol.* **28** (1963), 83.
11. J. GRALLA AND D. CROTHERS, Free energy of imperfect nucleic acid helices, II. Small hairpin loops, *J. Mol. Biol.* **73** (1973), 497–517.
12. J. GRALLA AND D. CROTHERS, Free energy of imperfect nucleic acid helices, III. Small internal loops resulting, *J. Mol. Biol.* **78** (1973), 301–319.
13. J. GRALLA AND C. DELISI, M-RNA is expected to form stable secondary structures, *Nature* **248** (1974), 330–332.
14. R. W. HOLLEY, J. APGAR, G. A. EVERETT, J. T. MADISON, M. MARQUISEE, S. H. MERRILL, J. R. PENSWICH, AND A. ZAMER, Structure of a ribonucleic acid, *Science* **147** (1965), 1462–1465.
15. W. N. HSIEH, Proportions of irreducible diagrams, *Studies in Applied Math.* **52** (1973), 277–283.
16. D. J. KLEITMAN, Proportions of irreducible diagrams, *Studies in Applied Math.* **49** (1970), 297–299.
17. A. J. MARK AND J. A. PETRUSKA, A method for predictory RNA structure from amino-acid sequence data, *J. Mol. Biol.* **72** (1972), 609–617.
18. W. MIN JOU, G. HAEGEMAN, M. YSEBAERT, AND W. FIERS, Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein, *Nature* **237** (1972), 82–88.

19. J. M. PIPAS AND J. E. MCMAHON, Method for predicting RNA secondary structure, *Proc. Nat. Acad. Sci. U.S.A.* **72** (1975), 2017–2021.
20. I. TINOCO, O. C. UHLENBECK, AND M. D. LEVINE, Estimation of secondary structure in ribonucleic acids, *Nature* **230** (1971), 362–367.
21. O. C. UHLENBECK, P. N. BORER, B. DENGLER, AND I. TINOCO, Stability of RNA hairpin loops: A_6–C_m–U_6, *J. Mol. Biol.* **73** (1973), 483–496.
22. M. S. WATERMAN, T. F. SMITH, AND W. A. BEYER, Some biological sequence metrics, *Advances in Math.* **20** (1976), 367–387.
23. A. M. LESK, A combinatorial study of the effects of admitting non-Watson–Crick base pairing and of base composition on the helix-forming potential of polynucleotides of random sequences, *J. Theor. Biol.* **44** (1974), 7–17.
24. D. POLAND AND H. A. SCHERAGA, "Theory of Helix-Coil Transitions in Biopolymers," Academic Press, New York, 1970.

AMS (MOS) subject classifications: 05A15, 05C99, 68A10, 92A05.