COMBINATORICS OF RNA SECONDARY STRUCTURES *

IVO L. HOFACKER†, PETER SCHUSTER‡ AND PETER F. STADLER*

Abstract. Secondary structures of polynucleotides can be view as a certain class of planar vertex-labeled graphs. We construct recursion formulae enumerating various sub-classes of these graphs as well as certain structural elements (sub-graphs). First order asymptotics are derived and their dependence on the logic of base pairing is computed and discussed.

Key words. Planar Graphs, Generating Functions, Asymptotic Enumeration, Secondary Structure

AMS subject classifications. 05A15, 05A16, 05C30, 92C40

1. Introduction. Presumably the most important problem and the greatest challenge in present day theoretical biophysics deals with deciphering the code that transforms sequences of biopolymers into spatial molecular structures. A sequence is properly visualized as a string of symbols which together with the environment encodes the molecular architecture of the biopolymer. In case of one particular class of biopolymers, the ribonucleic acid (RNA) molecules, decoding of information stored in the sequence can be properly decomposed into two steps. Transformation of the string into a planar graph, and folding of the string into a three-dimensional structure under conservation of the neighborhood relation determined by the graph. We are concerned here only with the first step, the transformation of the sequence into the graph (Fig. 1), which is much simpler than other known sequence-to-structure relations in biophysics. We are not concerned here with the physical rules that govern this transformation. Instead we are interested in the combinatorics of RNA secondary structures which in essence is an exercise in combining structural elements into valid structures under certain additional constraints.

Previous results on combinatorial aspects of secondary structures of RNA molecules are due to Waterman and coworkers [1-5]. Particularly important for the work reported here are a recursion for the number of different secondary structures formed by strings of constant length [1] as well as analytical expression for its asymptotic values [2]. Secondary structures are labeled planar graphs and as such they are closely related to the *linked diagrams* of Touchard [6-10].

In section 2 we introduce the basic definitions of secondary structures and recall their various representations. Section 3 presents the recursion formulas for the exact enumeration of various types of constrained secondary structures as well as their structural elements. Constrained secondary structures are of primary importance in

*Received by the editors ???? ??? 1992; accepted by the editors ????, ??, ????. This work was supported financially by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Projects No. S 5305-PHY and P 8526-MOB.

†Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria. (ivo@tbi.univie.ac.at).

‡Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria; The Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe, NM 87501, U.S.A.; Institut für Molekulare Biotechnologie e.V., PF 100813, D-07708 Jena, Germany (pks@imb-jena.de).

*Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Vienna, Austria; The Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe, NM 87501, U.S.A. (studla@tbi.univie.ac.at).

biophysics since not every conceivable element of a secondary structure will be found in reality. For example, hairpin loops containing one or two nucleotides are so strongly disfavored by the energetics that they do not occur in RNA secondary structures. In section 4 first order asymptotics to these recursions are devised. Although the class of graphs formed by secondary structures is interesting in its own rights, secondary structures in biology make sense only when they are related to sequences. Implications resulting from the condition that secondary structures have to be built on sequences are discussed in section 5. The results reported here are particularly interesting in relation to the data which were obtained from RNA secondary structure statistics performed by folding large ensembles of sequences into minimum free energy structures [11,12]. The asymptotic values show the influence of the logic of base pairing as expressed in terms of *stickiness*. Stickiness accounts for the possible base pairings supported by the nucleotide alphabet but ignores the energetic effect of different strenghts of the base pairs. Numerically computed data based on empirical energetic parameters include both effects, and the comparison allows to separate the influence of the pairing logic from the energetics. A detailed comparison will be given in a forthcoming paper [13].

2. Secondary Structures and Structural Elements.

2.1. Definitions.

DEFINITION 2.1. [1] A secondary structure is a vertex-labeled graph on n vertices with an adjacency matrix A fulfilling

(i) $a_{i,i+1} = 1$ for $1 \leq i \leq n-1$;

(ii) For each i there is at most a single $k \neq i-1, i+1$ such that $a_{ik} = 1$;

(iii) If $a_{ij} = a_{kl} = 1$ and $i < k < j$ then $i < l < j$.

We will call an edge (i,k), $|i-k| \neq 1$ a bond or base pair. A vertex i connected only to $i-1$ and $i+1$ will be called unpaired. A vertex i is said to be *interior* to the base pair (k,l) if $k < i < l$. If, in addition, there is no base pair (p,q) such that $k < p < i < q$ we will say that i is *immediately interior* to the base pair (k,l).

DEFINITION 2.2. A secondary structure consists of the following structure elements

(i) A *stack* consists of subsequent base pairs $(p-k, q+k)$, $(p-k+1, q+k-1)$, …, (p,q) such that neither $(p-k-1, q+k+l)$ nor $(p+1, q-1)$ is a base pair. $k+1$ is the *length* of the stack, $(p-k, q+k)$ is the terminal base pair of the stack.

(ii) A *loop* consists of all unpaired vertices which are immediately interior to some base pair (p,q).

(iii) An *external vertex* is an unpaired vertex which does not belong to a loop. A collection of adjacent external vertices is called an external element. If it contains the vertex 1 or n it is a free end, otherwise it is called joint.

LEMMA 2.3. *Any secondary structure $\mathcal{S}$ can be uniquely decomposed into stacks, loops, and external elements.*

Proof. Each vertex which is contained in a base pair belongs to a unique stack. Since an unpaired vertex is either external or immediately interior to a unique base pair the decomposition is unique: Each loop is characterized uniquely by its "closing" base pair. □

DEFINITION 2.4. A stack $[(p,q), \ldots, (p+k, q-k)]$ is called *terminal* if $p-1=0$ or $q+1 = n+1$ or if the two vertices $p-1$ and $q+1$ are not interior to any base pair. The sub-structure enclosed by the terminal base pair (p,q) of a terminal stack will be called a *component* of $\mathcal{S}$. We will say that a structure on n vertices has a terminal base pair if $(1,n)$ is a base pair.

LEMMA 2.5. *A secondary structure may be uniquely decomposed into components and external vertices. Each loop is contained in a component.*
The proof is trivial. Note that by definition the open structure has 0 components.

DEFINITION 2.6. The degree of a loop is given by 1 plus the number of terminal base pairs of stacks which are interior to the closing bond of the loop. A loop of degree 1 is called hairpin (loop), a loop of a degree larger than 2 is called multiloop. A loop of degree 2 is called bulge if the closing pair of the loop and the unique base pair immediately interior to it are adjacent; otherwise a loop of degree 2 is termed interior loop.

DEFINITION 2.7. Let $\mathcal{S}$ be an arbitrary secondary structure. For all $\mathcal{S}$ let us denote by $\Omega(\mathcal{S})$ the unique secondary structure which is obtained from $\mathcal{S}$ the following procedure:

(i) For each hairpin, open its stack and add the corresponding bases to the hairpin loop.

(ii) If a bulge or interior loop follows, then add its digits also to the hairpin and continue by opening its stack.

(iii) If a multiloop or a joint follows, then add the now unpaired digits to the multiloop and stop.

Waterman [1] used the above procedure to define the order $\omega(\mathcal{S})$ of a secondary structure as the smallest number of repetitions of Ω necessary to obtain the open structure. Of course, the open structure has order $\omega = 0$ and any structure without a multiloop has order $\omega = 1$.

2.2. Representation of Secondary Structures. A secondary structure $\mathcal{S}$ can be translated into a rooted ordered tree (linear tree) Υ by introducing an additional root and representing a base pair (p, q) by a vertex x such that the sons $y_1, \ldots y_k$ of x correspond to the base pairs $(p_1, q_1) \ldots (p_k, q_k)$ immediately interior to (p, q) [11,12]. For each unpaired vertex z a half-vertex is added to the vertex representing the closing pair of the loop containing z. (For external digits this is the root.) The tree-representation of a secondary structure is shown in fig. 1B.

A string representation $\mathbf{S}$ can by obtained by the following rules:

(i) If vertex i is unpaired then $\mathbf{S}_i =$ ".".

(ii) If (p, q) is a base pair and $p < q$ then $\mathbf{S}_p =$ "(" and $\mathbf{S}_q =$ ")".

These rules yield a sequence of matching brackets and dots (*cf.* fig. 1C). A related representation is derived in [14].

Waterman's definition of secondary structures implies that each branch of the corresponding tree representation Υ has at least one terminal half-vertex, or equivalently, each matching pair of brackets contains at least one dot. In biological applications the number of unpaired positions is at least 3, implying at least 3 dots within each pair of matching brackets. From the combinatorial point of view it makes perfect sense to consider the general problem with a minimum number $m \geq 0$ of unpaired vertices in each hairpin loop. In fact, for $m = 0$ one recovers three well known Motzkin families [2,15].

For some applications it is useful to work with simplified representations [16,17]. A tree T is obtained by denoting a stack by single vertex. In terms of the representation Υ this means that each vertex of degree 2 not carrying a half-vertex (except for the root) is merged with its son and then the half-vertices are removed (*cf.* fig.1D). The number of vertices in T is then just the number of stacks in $\mathcal{S}$, the number of components of $\mathcal{S}$ coincides with the number of sons of the root in T. An alternative "coarse grained" representation of a secondary structures is the homeomorphically

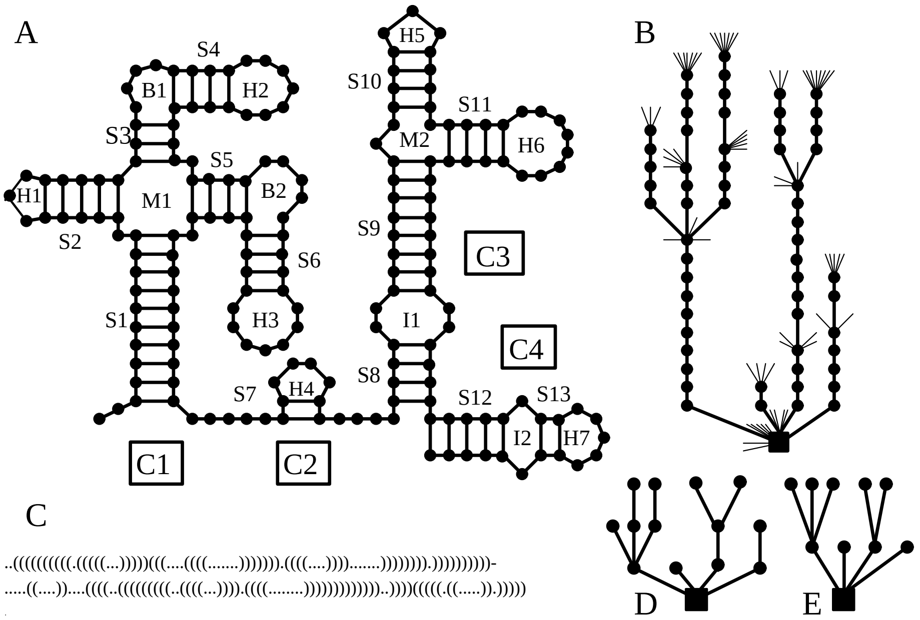

FIG. 1. *Representations of secondary structures. The notation* **A** *is common in biology. Structure elements are indicated as follows: H hairpin loops, I interior loops, B bulges, M multiloops; S stacks. The structure consists of four components, indicated as C1 through C4.*

B *is the corresponding tree notation, and* **C** *is the linear encoding of this tree. For details see section 2.2.*

D *is a coarse grained representation obtained from* **B** *by contracting each stack to a single vertex and omitting the half-vertices representing the unpaired positions.* **E** *is the homeomorphically irreducible tree obtained from* **D**.

irreducible tree $\mathcal{H}$ corresponding to Υ which is obtained by removing all vertices of degree 2 (except for the root) and all half-vertices. Again the number of components of $\mathcal{S}$ equals the number of sons of the root. Waterman's degree ω coincides with the height of $\mathcal{H}$ (*cf.* fig.1E).

2.3. The Basic Recursion. A secondary structure on $n+1$ digits may be obtained from a structure on n digits either by adding a free end at the right hand end or by inserting a base pair $1 \equiv (k+2)$. In the second case the substructure enclosed by this pair is an arbitrary structure on k digits, and the remaining part of length $n-k-1$ is also an arbitrary valid secondary structure. Therefore, we obtain the following recursion formula for the number S_n of secondary structures:

$$
\begin{gathered}
S_{n+1} = S_n + \sum_{k=m}^{n-1} S_k S_{n-k-1}, \qquad n \geq m+1 \\
S_0 = S_1 = \ldots = S_{m+1} = 1
\end{gathered} \tag{1}
$$

Equ.(1) has first been derived by Waterman [1]; m denotes the minimum number of unpaired digits in a hairpin loop. Note that our definition of S_n differs from Waterman's for $n < m$: he used $S_n = 0$.

The above recursion can be used to develop an algorithm for generating random secondary structures with a uniform distribution

$$\text{Prob}\{\mathcal{S}\} = 1/S_n \tag{2}$$

in the *shape space* of all secondary structures over a given chain length.

3. Recursions.

3.1. Structures with Certain Properties. Let $J_n(b)$ denote the number of structures on n vertices with exactly b components. The derivation of the recursion relations parallels the argument leading to equ.(1):

$$\begin{aligned} J_{n+1}(b) &= J_n(b) + \sum_{k=m}^{n-1} S_k J_{n-k-1}(b-1), \qquad b>0,\ n\geq m+1 \\ J_n(b) &= 0,\ b>0, n\leq m+1, \qquad J_n(0)=1, n\geq 0 \end{aligned} \tag{3}$$

because adding an unpaired digits to a structure on n digits does not change the number of components, while introducing an additional bracket makes the bracketed part of length k a single component and does not affect the remainder of the sequence.

Let $H_n(b)$ denote the number of structures with exactly b base pairs (bonds) on n vertices. The recursion

$$\begin{aligned} H_{n+1}(b) &= H_n(b) + \sum_{k=m}^{n-1}\sum_{\ell=0}^{b-1} H_k(\ell) H_{n-k-1}(b-\ell-1), \qquad b>0,\ n\geq m+1 \\ H_n(b) &= 0,\ b>0, n\leq m+1, \qquad H_n(0)=1, n\geq 0 \end{aligned} \tag{4}$$

is also immediate. One just has to observe that an additional sum over the number of unpaired digits in the newly bracketed part of the structure has to be introduced. This recursion has also been considered in [18]. Recently Schmitt and Waterman [5] obtained the closed expression $H_n(b) = \frac{1}{b}\binom{n-b}{b+1}\binom{n-b-1}{b-1}$ for the special case $m=1$.

Analogously we obtain for the number $E_n(b)$ of structures with b external digits

$$\begin{aligned} E_{n+1}(b) &= E_n(b-1) + \sum_{k=m}^{n-1} S_k E_{n-k-1}(b) \qquad b\geq 0,\ n\geq m+1 \\ E_n(n) &= 1, \quad E_n(b)=0 \quad b\neq n,\ n\leq m+1, \quad E_n(-1)=0 \end{aligned} \tag{5}$$

It is a bit more tricky to obtain a recursion for number $N_n(b)$ of sequences with a given number of stacks. To this end we introduce an auxiliary variable $Z_n(b)$ denoting the number of secondary structures with exactly b stacks *given* that its $3'$ and $5'$ ends are paired. We obtain then

$$\begin{aligned} N_{n+1}(b) &= N_n(b) + \sum_{k=m}^{n-1}\sum_{\ell=0}^{b} Z_{k+2}(\ell) N_{n-k-1}(b-\ell), \quad b>0,\ n\geq m+1 \\ N_n(0) &= 1, \quad N_n(b)=0, \quad b>0,\ n\leq m+1 \end{aligned} \tag{6}$$

For the auxiliary variably we find

$$Z_n(b) = Z_{n-2}(b) + N_{n-2}(b-1) - Z_{n-2}(b-1), \qquad Z_0(b)=Z_1(b)=0 \tag{7}$$

by enclosing structures on $n-2$ digits by a base pair.

Let $A_n(b)$ denote the number of structures with exactly b hairpins. Since the number of hairpins is unchanged by enclosing a substructure which already contains a base pair in an additional base pair we get

$$\begin{aligned} A_{n+1}(b) &= A_n + \sum_{k=m}^{n-1}\left[\sum_{\ell=1}^{b} A_k(\ell)A_{n-k-1}(b-\ell) + A_{n-k-1}(b-1)\right] \\ & \qquad n \geq m+1 \\ A_n(b) &= \delta_{0,b} \qquad n \leq m+1 \end{aligned} \tag{8}$$

where $\delta_{0,b}$ is Kronecker's δ, i.e. $\delta_{0,0}=1$ and $\delta_{0,b}=0$, $b \neq 0$.

3.2. Structure Elements. The total number U_n of unpaired bases in the set of all structures can be obtained as follows: By adding an unpaired base to each structure on n digits we have the U_n unpaired digits present in them plus the S_n newly added ones. By introducing the base pair $1 \equiv k+2$ we have S_k times all the unpaired digits in the reminder of the sequence plus all the unpaired digits in the newly bracket part of length k times the the number of structures which can be formed from the reminder of the structure. Summing over k we find

$$\begin{aligned} U_{n+1} &= (U_n + S_n) + \sum_{k=m}^{n-1}[S_k U_{n-k-1} + S_{n-k-1}U_k], \qquad n \geq m+1 \\ U_n &= n, \quad n \leq m+1 \end{aligned} \tag{9}$$

Denote the total number of base pairs by P_n. It is clear that $2P_n + U_n = nS_n$. For sake of completeness we state the recursion for P_n:

$$\begin{aligned} P_{n+1} &= P_n + \sum_{k=m}^{n-1}\{S_k P_{n-k-1} + S_{n-k-1}(P_k + S_k)\} \\ P_n &= 0, \quad n \leq m+1 \end{aligned} \tag{10}$$

By an analogous reasoning we find for the total number I_n of components in the set of all secondary structures on n vertices.

$$\begin{aligned} I_{n+1} &= I_n + \sum_{k=m}^{n-1} S_k[I_{n-k-1} + S_{n-k-1}] \\ I_n &= 0 \quad n \leq m+1 \end{aligned} \tag{11}$$

The number N_{n+1} of stacks in the set of structures on $n+1$ digits consists of all stacks on n digits plus all stacks in the tail times the number of structures with the newly introduced base pair plus all stacks within the newly formed base pair times the number of structures in the tail. The newly formed base pair introduces an additional stack for all $S_k - S_{k-2}$ structures in its interior which do not have a terminal base pair. (For the S_{k-2} structures with terminal base pair a stack is elongated.) Therefore

$$\begin{aligned} N_{n+1} &= N_n + \sum_{k=m}^{n-1}\{S_k N_{n-k-1} + S_{n-k-1}(N_k + S_k)\} \\ &\quad - \sum_{k=m+2}^{n-1} S_{k-2}S_{n-k-1} \qquad n \geq m+1 \\ N_n &= 0, \; n \leq m+1 \end{aligned} \tag{12}$$

Let $Q_n(b)$ denote the number of loops with b unpaired digits in the set of all secondary structures. For $n+1$ vertices we retain all loops from the set of loops on n digits by adding a vertex to the $3'$ end; additional we find all loops in the tail-substructure for each possible structure interior to the new base pair. The third contributions consists of all loops interior to the new base pair times all possible structures in the tail. A loop with b unpaired vertices remains unchanged and additionally each structure with exactly b external vertices within the new base pair gives rise to an additional loop with b unpaired digit.

$$\begin{aligned} Q_{n+1}(b) &= Q_n(b) + \sum_{k=m}^{n-1} \{Q_{n-k-1}(b)S_k + S_{n-k-1}[Q_k(b) + E_k(b)]\} \\ & \qquad\qquad n \geq m+1,\, b > 0 \\ Q_n(b) &= 0,\, n \leq m+1 \end{aligned} \tag{13}$$

For loops without unpaired digits the recursion is slightly different since structures without external digits within the new base pair do not provide a loop if they consist of a single component, i.e. if they end in base pair. There are S_{k-2} such structures on k vertices.

$$\begin{aligned} Q_{n+1}(0) &= Q_n(0) + \sum_{k=m}^{n-1} \{Q_{n-k-1}(0)S_k + S_{n-k-1}[Q_k(0) + E_k(0)]\} \\ & \qquad - \sum_{k=m+2}^{n-1} S_{n-k-1}S_{k-2} \quad n \geq m+1 \\ Q_n(0) &= 0,\, n \leq m+1 \end{aligned} \tag{14}$$

Let $W_n(b)$ denote the number of stacks with exactly b base pairs in the set of secondary structures. For each structure in the bracketed part there are $W_{n-k-1}(b)$ stacks of suitable size while for each structure in the tail there only $W_k(b) - W_k^-(b) + W_k^+(b)$ such stacks; $W_k^-(b)$ denotes all stacks of correct length which are elongated by the new bracket and $W_k^+(b)$ denotes all stacks which are too short by one base pair and are elongated by the new pair. Clearly, $W_k^-(b)$ is just the number of structures on k digits with a terminal stack of length b, while $W_k^+(b)$ is the number of structures with a terminal stack of length $b-1$. We have

$$W_k^-(b) = \begin{cases} S_{k-2b} - S_{k-2b-2} & k > m+2b+1 \\ 1 & k = m+2b, m+2b+1 \\ 0 & k < m+2b \end{cases} \tag{15}$$

and $W_k^+(b) = W_k^-(b-1)$ for these auxiliary variables.

$$
\begin{aligned}
W_{n+1}(b) = & \\
&= W_n(b) + \sum_{k=m}^{n-1} \left\{ W_{n-k-1}(b) S_k + S_{n-k-1}[W_k(b) - W_k^-(b) + W_k^+(b)] \right\} \\
&= W_n(b) + \sum_{k=m}^{n-1} \left[W_{n-k-1}(b) S_k - S_{n-k-1} W_k(b) \right] + \\
&\qquad + \sum_{k=m+2b+2}^{n-1} S_{k-2b-2} S_{n-k-1} - 2 \sum_{k=m+2b}^{n-1} S_{k-2b} S_{n-k-1} + \\
&\qquad + \sum_{k=m+2b-2}^{n-1} S_{k-2b+2} S_{n-k-1} \\
W_n(b) = 0 & \quad \text{for } n \leq m+1
\end{aligned}
\tag{16}
$$

Let $L_n(d)$ denote the number of loops of degree d in the set of all secondary structures. By Y_n and B_n, resp., we will denote the number of interior loops and bulges. Let us start with bulges and interior loops: Let X_n^* denote the number of structures that yield a bulge if included into an extra pair of brackets, and let X_n^{**} denote the number of structures that yield an interior loop if included into an extra bracket, i.e. the number of structures having a free end on both sides. Clearly $X_n^{**} = J_{n-2}(1)$, as structures with zero components would yield a hairpin while structures with more components would yield a multiloop. In order to calculate X_n^* we observe that a bulge is formed by an new bracket if the structure enclosed has only a single component and ends neither in a base pair nor in free ends on both sides. As there are S_{n-2} structures resulting in a stack elongation if $n \geq m+2$ (and none otherwise) we have

$$
X_n^* = J_n(1) - J_{n-2}(1) - S_{n-2} \quad n \geq m+2 \tag{17}
$$

The recursions for loops of degree 2 are now straight forward:

$$
\begin{aligned}
B_{n+1} &= B_n + \sum_{k=m}^{n-1} \left\{ S_k B_{n-k-1} + S_{n-k-1}[B_k + X_k^*] \right\} \\
Y_{n+1} &= Y_n + \sum_{k=m}^{n-1} \left\{ S_k Y_{n-k-1} + S_{n-k-1}[Y_k + J_{k-2}(1)] \right\} \\
L_{n+1}(2) &= L_n(2) + \sum_{k=m}^{n-1} \left\{ S_k L_{n-k-1}(2) + S_{n-k-1}[L_k(2) + J_k(1)] \right\} \\
&\qquad - \sum_{k=m+2}^{n-1} S_{n-k-1} S_{k-2} \\
B_n = Y_n &= L_n(2) = 0 \qquad n \leq m+1
\end{aligned}
\tag{18}
$$

Hairpins are generated either by stack-elongation of a structure with a single hairpin or by enclosing the open structure into the additional bracket. Thus

$$
\begin{aligned}
L_{n+1}(1) &= L_n(1) + \sum_{k=m}^{n-1} \left\{ S_k L_{n-k-1}(1) + S_{n-k-1}[L_k(1) + 1] \right\} \quad n \geq m+1 \\
L_n(1) &= 0 \quad n \leq m+1
\end{aligned}
\tag{19}
$$

For multiloops finally we obtain the recursion

$$
\begin{aligned}
L_{n+1}(d) &= L_n(d) + \sum_{k=m}^{n-1}\left\{S_k L_{n-k-1}(d) + S_{n-k-1}[L_k(d)+J_k(d-1)]\right\} \\
&\qquad \text{for} \quad d\geq 2,\ n\geq m+1 \\
L_n(d) &= 0 \qquad \text{for } n\leq m+1
\end{aligned} \tag{20}
$$

Summing over all loop degrees d we recover the recursion for the total number of stacks, since for each stack there is exactly one loop.

The total number of external digits, E_n, can be obtained directly as sum of the numbers $E_n(b)$. For sake of completeness we mention that it fulfills the recursion

$$
\begin{aligned}
E_{n+1} &= E_n + S_n + \sum_{k=m}^{n-1} S_k E_{n-k-1} \quad n\geq m+1 \\
E_n &= n \quad n\leq m+1
\end{aligned} \tag{21}
$$

3.3. Secondary Structures of a Given Order. Let $D_n(c,\omega)$ be the number of secondary structures with c components and order ω. Furthermore let $D_n^*(\omega)$ be the number of structures which yield a structure of order ω when enclosed by an additional base pair. It is clear that the following recursion holds

$$
\begin{aligned}
D_{n+1}(c,\omega) &= D_n(c,\omega) + \sum_{k=m}^{n-1}\Bigg\{D_k^*(\omega)\sum_{\ell=0}^{\omega-1} D_{n-k-1}(c-1,\ell) \\
&\qquad + D_{n-k-1}(c-1,\omega)\sum_{\ell=0}^{\omega-1} D_k^*(\ell) + D_k^*(\omega) D_{n-k-1}(c-1,\omega)\Bigg\} \\
D_n(0,0) &= 1,\ D_n(0,d) = D_n(c,0) = 0 \qquad n\leq m+1
\end{aligned} \tag{22}
$$

since a structure with a base pair $1\equiv k+2$ has order d and c components iff either the bracketed part has order ω and the tail has a order at most ω and $c-1$ components or the bracketed part has a degree smaller than ω and the tail has $c-1$ components and order ω. It remain to calculate $D_n^*(\omega)$. By inspection we find for $n>m$

$$
\begin{aligned}
D_n^*(0) &= 0 \\
D_n^*(1) &= 1 + D_n(1,1) \\
D_n^*(\omega) &= D_n(1,\omega) + \sum_{\ell=2}^{\infty} D_k(\ell,\omega-1), \quad \omega\geq 2
\end{aligned} \tag{23}
$$

while for $n\leq m$ we have $D_n^*(\omega)=0$. There is no structure of order 0 with a bracket in it; order one is obtained by either bracketing the open structure or by bracketing a structure with a single component and order 1. If the bracketed part has only a single components its order is preserved by adding a terminal bracket. If it consists of more than one components, the addition of the multiloop increases the order by one.

Summing over the number of components we obtain the number of Structures with given order $\tilde{D}_n(\omega)$. Let us further introduce the number of structure of order at most one, $D'_n(1)$. It is easy to derive the following system of recursions from the above ones:

$$
\begin{aligned}
\tilde{D}_{n+1}(\omega) &= \tilde{D}_n(\omega) + \sum_{k=m}^{n-1}\left\{ D^*_k(\omega)\sum_{\ell=0}^{\omega-1}\tilde{D}_{n-k-1}(\ell) + \tilde{D}_{n-k-1}(d)\sum_{\ell=0}^{\omega} D^*_k(\ell)\right\}\\
D^*_k(\omega) &= \tilde{D}_k(\omega-1) + D_k(1,\omega) - D_k(1,\omega-1) \qquad\qquad n\geq m+2\\
D_{n+1}(1,\omega) &= D_n(1,\omega) + \sum_{k=m}^{n-1} D^*_k(\omega)\\
&\tilde{D}_n(0)=1,\ \tilde{D}_n(\omega)=0 \text{ for } \omega\geq 1, \qquad n\leq m+1
\end{aligned} \tag{24}
$$

For the number of structures with a degree at most one we find

$$
\begin{aligned}
D'_{n+1} &= D'_n + \sum_{k=m}^{n-1} D^*_k(1) D'_{n-k-1}\\
D^*_{n+1}(1) &= \sum_{k=m}^{n} D^*_k(1)
\end{aligned} \tag{25}
$$

3.4. Secondary Structures with Minimum Stack Length. Let $\Psi_n(l)$ be the number of structures with minimal stack length l, and let $\Psi^*_n(l)$ be the number of structures on n digits which have only stacks of length at least l if an additional terminal base pair is attached. Furthermore let $\Psi^{**}_n(l)$ be the number of structures on n digits with all stacks of length at least l for which $1\equiv n$ is not a base pair.

These three numbers fulfill for $l>1$ the coupled recursions

$$
\begin{aligned}
\Psi_{n+1}(l) &= \Psi_n(l) + \sum_{k=m+2l-2}^{n-1} \Psi^*_k(l)\Psi_{n-k-1}(l)\\
\Psi^*_n(l) &= \sum_{p=l-1}^{(n-m)/2} \Psi^{**}_{n-2p}(l)\\
\Psi^{**}_n(l) &= \Psi_n(l) - \Psi^*_{n-2}(l)\\
&\Psi_n(l) = \Psi^{**}_{n+1}(l) = 1 \quad n < m+2l,\\
&\Psi^*_n(l) = 0 \quad m+2l-2
\end{aligned} \tag{26}
$$

The first recursion is obvious. A structure which has only stacks of length at least l after addition of the terminal base pair must have a terminal stack of length $p\geq l-1$. The remaining part of the structure must have stacks of length at least l without a terminal base pair. Of course there is no such structure if $n-2p<m$. For the numbers $\Psi^{**}_n(l)$ we obtain the explicit recursion:

$$
\begin{aligned}
\Psi^{**}_{n+1}(l) &= \Psi_n(l) + \sum_{k=m+2l-2}^{n-2} \Psi^*_k(l)\Psi_{n-k-1}(l)\\
&\Psi^{**}_n = 1 \quad n < m+2l
\end{aligned} \tag{27}
$$

because structures without a terminal base pair and stacks of length at least l are obtained by adding a new base pair to structures which including this base pair have stacks of sufficient length (first factor in the sum) provided the structures in the remaining part of the structure have also sufficient stack length. Of course there may not by a terminal base pair by construction. Comparing the sum in (27) and in the recursion for $\Psi_n(l)$ yields the final result. We have of course $\Psi_n(1) = S_n$ for all n and $\Psi_n(l+1) < \Psi_n(l)$ for all l and sufficiently large n.

Remark. It is possible of course to obtain recursions of the above type for the number of structure elements or the number of structures with particular properties also for $l > 1$. If Ξ_n is the counting series of interest one has to replace $S_k \Xi_{n-k-1}$ by $\Psi^*_k \Xi_{n-k-1}$ and $\Xi_k S_{n-k-1}$ by $\Xi^*_k \Psi_{n-k-1}$, where Ξ^* counts the objects of interest subject to the restriction that the secondary structure has a terminal stack of length at least l.

4. Asymptotics.

Notation 4.1. The symbols $\sim$ and $\mathcal{O}$ have their usual meaning:

(i) $f(x) = \mathcal{O}(x)$ means $f(x)/x$ is bounded.

(ii) $f(n) \sim g(n)$ means $f(n)/g(n) \longrightarrow 1$ as $n \longrightarrow \infty$.

The symbol o, however, does not have its standard meaning in this paper (*cf.* theorem 4.5). If not explicitly stated, asymptotic formulas assume $n \longrightarrow \infty$.

4.1. Asymptotics from Generating Functions.

We will use the following simplified version of Darboux' theorem (*cf.* [19, p.205])

THEOREM 4.2. *Let $y(x) = \sum_{n=0}^{\infty} y_n x^n$ be of the form*

$$y(x) = \beta(x) + \sum_k g_k(x)(1 - \frac{x}{\alpha})^{\omega_k} \tag{28}$$

where β, g_k are analytic on a circle larger than the circle of convergence of $y(x)$, ω_k real but not a non-negative integer. Suppose y has only a single singularity at $x = \alpha$ Denote by ω the smallest exponent ω and by $g(x)$ the corresponding analytic factor. Then

$$y_n \sim \frac{g(\alpha)}{\Gamma(-\omega)}\, n^{-1-\omega} \left(\frac{1}{\alpha}\right)^n \tag{29}$$

THEOREM 4.3. [20, Theorem 5]. *Assume that $y_n \geq 0$, and $y(x) = \sum_{n=0}^{\infty} y_n x^n$ satisfies $F(x,y) \equiv 0$. Suppose there are real numbers $\alpha > 0$, $\beta > y_0$ such that*

(i) *for some $\delta > 0$, $F(x,y)$ is analytic whenever $|x| < \alpha + \delta$ and $|y| < \beta + \delta$;*

(ii) *$F(\alpha,\beta) = 0$, $F_y(\alpha,\beta) = 0$;*

(iii) *$F_x(\alpha,\beta) \neq 0$, $F_{yy}(\alpha,\beta) \neq 0$;*

(iv) *(α,β) is the only solution in the interior of the complex rectangle $|x| < \alpha$ and $|y| < \beta$.*

Then

$$y_n \sim \sqrt{\frac{\alpha F_x(\alpha,\beta)}{2\pi F_{yy}(\alpha,\beta)}}\, n^{-3/2} \left(\frac{1}{\alpha}\right)^n \tag{30}$$

Remark 4.4. By comparison of theorem 4.2 and 4.3 we find immediately

$$g(\alpha) = -\sqrt{\frac{2\alpha F_x(\alpha,\beta)}{F_{yy}(\alpha,\beta)}} \qquad \text{with} \quad \beta = \beta(\alpha) \tag{31}$$

THEOREM 4.5. *Let $\Phi(x,y)$ be analytic for $|x|<\alpha+\delta$ and $|y|<\beta(\alpha)+\delta$, $\delta>0$. Suppose y is of the form*

$$y(x)=\beta(x)+(1-\frac{x}{\alpha})^{1/2}g(x). \tag{32}$$

Let $z(x)=\sum_{n=0}^{\infty}z_nx^n$ be a generating function of the form $z=\Phi(x,y)$. Then

$$\lim_{n\to\infty}\frac{z_n}{y_n}=\Phi_y(\alpha,\beta(\alpha)) \tag{33}$$

Proof. We will use the short hand o for any function $o(x)$, which is analytic for $|x|<\alpha+\delta$, $\delta>0$, and fulfils $o(\alpha)=0$.

$$\begin{aligned}\Phi(x,y)&=\sum_{k=0}^{\infty}a_k(x)y^k=\\&=\sum_{k=0}^{\infty}a_k(x)\left\{\beta(x)^k+\left[k\beta(x)^{k-1}g(x)+o\right](1-x/\alpha)^{1/2}\right\}=\\&=\Phi(x,\beta(x))+\left[\Phi_y(x,\beta(x))g(x)+o\right]\left(1-\frac{x}{\alpha}\right)^{1/2}\end{aligned} \tag{34}$$

Darboux' theorem shows that

$$z_n\sim\frac{g(\alpha)\Phi_y(\alpha,\beta)}{\Gamma(-\frac{1}{2})}\cdot n^{-3/2}\left(\frac{1}{\alpha}\right)^n \tag{35}$$

THEOREM 4.6. *Let $\Phi(x,y)$ be analytic as in the previous Theorem. Suppose y is of the form*

$$y(x)=\beta(x)+(1-\frac{x}{\alpha})^{1/2}g(x) \tag{36}$$

Let $z(x)=\sum_{n=0}^{\infty}z_nx^n$ be a generating function of the form

$$z(x)=\frac{1}{\alpha\beta-xy}\Phi(x,y), \tag{37}$$

where $\beta=\beta(\alpha)$. Then

$$\frac{z_k}{y_k}\sim\frac{2\Phi(\alpha,\beta)}{\alpha g^2(\alpha)}\cdot n \tag{38}$$

Proof. Consider first

$$\begin{aligned}\frac{1}{\alpha\beta-xy}&=\frac{1}{\alpha\beta-x\beta(x)-xg(x)(1-x/\alpha)^{1/2}}=\\&=\frac{\alpha\beta-x\beta(x)+xg(x)(1-x/\alpha)^{1/2}}{[\alpha\beta-x\beta(x)]^2-x^2g^2(x)(1-x/\alpha)}=\\&=\frac{o+xg(x)(1-x/\alpha)^{1/2}}{\mathcal{O}(1-x/\alpha)^2-x^2g^2(x)(1-x/\alpha)}=\\&=\eta(x)-\frac{1}{xg(x)}[1+o](1-x/\alpha)^{-1/2}\end{aligned}$$

where $\eta(x)$ is analytic on circle larger than the circle of convergence of $y(x)$. Multiplying this expression by a Taylor expansion of $\Phi(x,y)$ yields

$$\begin{aligned} \frac{1}{\alpha\beta - xy}\Phi(x,y) &= \eta\Phi - \sum_{k=0}^{\infty} a_k(x)\beta(x)^k \frac{1}{xg(x)}(1-x/\alpha)^{-1/2}[1-o] \\ &= \eta\Phi - \frac{\Phi(x,y)}{xg(x)}[1-o](1-\frac{x}{\alpha})^{-1/2} \end{aligned} \tag{39}$$

Applying Darboux' theorem and using that $\Gamma(\frac{1}{2}) = -\frac{1}{2}\Gamma(-\frac{1}{2})$ completes the proof. □

COROLLARY 4.7. *Let y as in the previous theorem and let u, v be of the same form as z above. Suppose there is an analytic function $\Phi(x,y)$ such that $u = \Phi(x,y)v$. Then*

$$\lim_{n\to\infty} \frac{u_n}{v_n} = \Phi(\alpha,\beta) \tag{40}$$

Proof. Assuming that both u and v are of the form (37) we find from equation (38) that $u_n/v_n = \Phi^u(\alpha,\beta)/\Phi^v(\alpha,\beta)$. The conditions of theorem 4.6 ensure that this quotient exists and $\Psi = \Phi^u/\Phi^v$. □

4.2. The Number of Secondary Structures. The series S_n has been extensively studied in [1]. Consider the series Ψ_n of secondary structures with a prescribed minimum stack length l and minimum size m for hairpin loops. Denote by

$$\psi(x) = \sum_{n=0}^{\infty} \Psi_n x^n, \qquad \phi(x) = \sum_{n=0}^{\infty} \Psi_n^* x^n, \qquad \theta(x) = \sum_{n=0}^{\infty} \Psi_n^{**} x^n \tag{41}$$

the generating functions. We introduce furthermore the notation

$$t_m(x) = \sum_{k=0}^{m-1} x^k \qquad \tau_m(x) = \sum_{k=1}^{m-1} kx^k = x\frac{d}{dt}t_m(x) \tag{42}$$

THEOREM 4.8. *The generating function ψ, ϕ and θ fulfill the coupled functional equations*

$$\begin{aligned} \psi &= 1 + x\psi + x^2\phi\psi \\ \phi &= \frac{x^{2(l-1)}}{1-x^2}\big(\theta - t_m(x)\big) \\ \theta &= \psi - x^2\phi \end{aligned} \tag{43}$$

Proof. The first and third line are obvious. The second line is obtained from

$$
\begin{aligned}
\phi &= \sum_{n=0}^{\infty} x^n \sum_{p=l-1}^{(n-m)/2} \Psi^{**}_{n-2p} = \\
&= \sum_{n=0}^{\infty}\sum_{p=0}^{n/2} x^2 p \Psi^{**}_{n-2p} x^{n-2p} - \sum_{p=0}^{l-2} x^{2p} \sum_{n=0}^{\infty} \Psi^{**}_{n-2p} x^{n-2p} - \\
&\quad - \sum_{p>\frac{n-m}{2}}^{n/2} x^{2p} \sum_{n=0}^{\infty} \Psi^{**}_{n-2p} x^{n-2p} + \sum_{p>\frac{n-m}{2}}^{l-2} x^{2p} \sum_{n=0}^{\infty} \Psi^{**}_{n-2p} x^{n-2p} \\
&= \frac{1}{1-x^2}\theta - \sum_{p=0}^{l-2} x^{2p}\theta - t_m(x) + \sum_{p=0}^{l-2} x^{2p} t_m(x) = \\
&= \frac{x^{2(l-1)}}{1-x^2}\big(\theta - t_m(x)\big)
\end{aligned} \tag{44}
$$

□

COROLLARY 4.9. *The generating function ψ fulfills the functional equation*

$$F(x,\psi) = x^{2l}\psi^2 - \big[(1-x)(1-x^2+x^{2l}) + x^{2l}t_m(x)\big]\psi + (1-x^2+x^{2l}) = 0 \tag{45}$$

COROLLARY 4.10. *For $l = 1$ we recover the generating function $s(x) = \sum_{n=0}^{\infty} S_k x^k$ for the number of secondary structures. It fulfills the functional equation*

$$F(x,y) = 1 + \big(2x - \sum_{k=0}^{m+1} x^k\big)y + x^2 \cdot y^2 = 0 \tag{46}$$

COROLLARY 4.11.

$$\Psi_n \sim \frac{-g(\alpha)}{2\sqrt{\pi}} n^{-3/2} \left(\frac{1}{\alpha}\right)^n \tag{47}$$

where $g(\alpha)$ is given in equ.(31) with

$$\beta = \frac{1}{\alpha^l}\sqrt{1-\alpha^2+\alpha^{2l}}, \tag{48}$$

and α is the smallest positive solution of

$$2x^l\sqrt{1-x^2+x^{2l}} - \big[(1-x)(1-x^2+x^{2l}) + x^{2l}t_m(x)\big] = 0 \tag{49}$$

Proof. From (43) some simple algebra yield the functional equation (45). From $F(\alpha,\beta) - \beta F_y(\alpha,\beta) = 0$ one obtains immediately (48) and (49) for the zeros necessary for the application of theorem 4.3. The latter proves equ.(47). □

Corollary 4.12. For $l = 1$ the above equations simplify to $\beta = 1/\alpha$ and

$$\sum_{k=0}^{m+1} \alpha^k - 4\alpha = 0 \tag{50}$$

Numerical values are given in table 1.

Throughout the remainder of this paper we will assume $l = 1$ if l is not mentioned explicitly, while α and β will denote the solutions of equations (49) and (48) respectively.

TABLE 1
Coefficients for the asymptotics of Ψ_n.

l	m= 0	1	2	3	5	∞
			α			
1	0.3333	0.3820	0.4142	0.4369	0.4658	0.5000
2	0.4836	0.5081	0.5266	0.5409	0.5610	0.5958
3	0.5672	0.5828	0.5952	0.6053	0.6204	0.6537
4	0.6227	0.6336	0.6428	0.6504	0.6623	0.6938
5	0.6629	0.6712	0.6783	0.6843	0.6941	0.7237
10	0.7704	0.7737	0.7766	0.7793	0.7840	0.8066
20	0.8713	0.8518	0.8530	0.8540	0.8559	0.8713
100	0.9520	0.9521	0.9522	0.9523	0.9525	0.9571
∞	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
			$-g(\alpha)/(2\sqrt{\pi})$			
1	1.4658	1.1044	0.8766	0.7131	0.4880	0.0000
2	2.7155	2.1614	1.7742	1.4848	1.0769	0.0000
3	3.9640	3.2711	2.7558	2.3561	1.7741	0.0000
4	5.2305	4.4238	3.7990	3.3003	2.5537	0.0000
5	6.5194	5.6142	4.8923	4.3033	3.4009	0.0000
10	13.309	12.026	10.921	9.962	8.3820	0.0000
20	28.365	26.557	24.913	23.414	20.787	0.0000
100	189.31	185.30	181.41	177.63	170.40	0.0000

4.3. Average Number of Structure Elements. Denote by Ξ_n the number of structural elements. From the biological point of view it is very interesting to know the average number of structural elements in a single structure, i.e. the asymptotic behaviour of Ξ_n/S_n. It is clear that the counting series for the total number of structure elements, including the total number of base pairs and unpaired digits is bounded from above by nS_n.

LEMMA 4.13.

$$t_m(\alpha) = \frac{3\alpha-1}{\alpha^2} \qquad \tau_m(\alpha) = \frac{3\alpha-1}{\alpha(1-\alpha)} - m\frac{(1-2\alpha)^2}{\alpha^2(1-\alpha)}$$
$$g^2(\alpha) = \frac{(1-2\alpha)(2+m-2m\alpha)}{(1-\alpha)\alpha^3} \tag{51}$$

THEOREM 4.14. *The number of components, I_n, fulfils*

$$\lim_{n\to\infty} \frac{I_n}{S_n} = 2\beta(1-\alpha) - 1 = 2/\alpha - 3 \tag{52}$$

Proof. Let $i(x) = \sum_{k=0}^{\infty} I_k x^k$ be the generating function for the number of components. The recursion can be brought to the form

$$I_{n+1} = I_n + \sum_{k=0}^{n-1} S_k I_{n-k-1} + \sum_{k=0}^{n-1} S_k S_{n-k-1} - \sum_{k=0}^{m-1} [I_{n-k-1} + S_{n-k-1}] \tag{53}$$

Multiplying by x^{n+1} and summing over n yields

$$i(x) = xi(x) + x^2 s(x) i(x) + x^2 s^2(x) - x^2 t_m(x)[s(x) + i(x)] \tag{54}$$

We find by using twice the functional equation for $s(x)$

$$\begin{aligned} i(x) &= \frac{x^2 s^2(x) - s(x)x^2 t_m(x)}{1 - x - x^2 s(x) + x^2 t_m(x)} \\ &= s(x) \cdot x^2 s(x)\big[s(x) - t_m(x)\big] \\ &= s^2(x)(1-x) - s(x) \end{aligned} \tag{55}$$

Application of theorem 4.5 immediately yields the desired result. □

Remark 4.15. In the first form of equ. (52) this result holds for arbitrary minimal stack length l as well.

THEOREM 4.16. *The number of external digits, E_n, fulfils*

$$\lim_{n\to\infty} \frac{E_n}{S_n} = 2\alpha\beta = 2 \tag{56}$$

Proof. The functional equation for the generating function reads $e(x) = x \cdot s^2(x)$. Theorem 4.5 completes the proof. □

THEOREM 4.17. *The number of unpaired digits, U_n, fulfils*

$$\frac{U_n}{S_n} \sim \frac{2\alpha + m(1-2\alpha)}{2 + m(1-2\alpha)} \cdot n \tag{57}$$

Proof. Let $u(x) = \sum_{n=0}^{\infty} U_n x^n$ be the generating function of the number of unpaired digits. From recursion (9) we find immediately the functional equation

$$u = xu + xs + 2x^2 us - x^2 u t_m(x) - x^2 s\tau_m(x) \tag{58}$$

Using the functional equation for s, some algebra yields

$$u(x) = \frac{1}{1 - x^2 s^2} \cdot s^2 x(1 - x\tau_m) \tag{59}$$

Application of theorem 4.6 completes the proof. □

Remark 4.18. Let $p(x)$ be the generating function for the number of base pairs. Since $U_n + 2P_n = nS_n$ we have $u(x) + 2p(x) = xs'(x)$.

THEOREM 4.19. *The number of stacks or loops, N_n, fulfils*

$$\frac{N_n}{S_n} \sim \frac{(1-\alpha)^2(1+\alpha)}{2 + m - 2m\alpha} \cdot n \tag{60}$$

Proof. Let $\nu(x) = \sum_{n=0}^{\infty} N_n x^n$ be the generating function of the number of stacks. Observe that

$$\sum_{k=m+p}^{n-1} S_{k-p} S_{n-k-1} = \sum_{k=m}^{n-p-1} S_k S_{n-p-k-1} \tag{61}$$

and therefore gives rise to a term $x^{p+2}[(s^2 - st_m(x)] = x^p[(1-x)s-1]$ in the functional equation for the generating function. Thus recursion (6) translates to

$$\nu = x\nu + 2x^2 s\nu - x^2 \nu t_m(x) + (1-x^2)[s\cdot(1-x) - 1] \tag{62}$$

or, after some simple algebra,

$$\nu = \frac{1}{1 - x^2 s^2} s(1-x^2)[s(1-x) - 1] \tag{63}$$

The proof is completed by theorem 4.6. □

4.4. The Number of Structures with Certain Properties.

THEOREM 4.20. *The number of structures with b base pairs is*

$$H_n(b) \sim \frac{1}{(b+1)!b!} n^{2b} \tag{64}$$

Proof. From recursion (4) one finds the functional equation

$$\begin{aligned} h_b &= xh_b + x^2 \sum_{k=0}^{b-1} h_{b-k-1} h_k - x^2 t_m(x) h_{b-1} \qquad b > 0 \\ &= xh_b + x^2 \sum_{k=1}^{b} h_k h_{b-k-1} + x^{m+2} h_{b-1} \end{aligned} \tag{65}$$

and $h_0(x) = 1/(1-x)$. With the ansatz

$$h_b(x) = \eta_b(x) \cdot \frac{1}{1-x} \left(\frac{x}{1-x} \right)^{2b} \tag{66}$$

one checks finds that $\eta_b(x)$ are polynomials fulfilling

$$\eta_b(x) = \sum_{k=1}^{b} \eta_k(x) \eta b - k - 1(x) + x^m \eta_{b-1} \tag{67}$$

Theorem 4.2 assures now that

$$H_n(b) \sim \frac{\eta_b(1)}{\Gamma(2b+1)} \cdot n^{2b}. \tag{68}$$

Since $\eta_0(1) = 1$, equ.(67) becomes the well known recursion for the Catalan numbers

$$\eta_b(1) = C_b = \frac{1}{b+1} \binom{2b}{b}. \tag{69}$$

□

THEOREM 4.21. *The number of structures with exactly b stacks is*

$$N_n(b) \sim \frac{C_b}{2^b (3b)!} \cdot n^{3b} \tag{70}$$

Proof. Let $\nu_b(x) = \sum_{n=0}^{\infty} N_n(b) x^n$ be the generating function for the number of structures with exactly b stacks and denote by $\zeta_b(x)$ the generating function for the auxiliary variable $Z_n(b)$. It is straight forward to derive the functional equations

$$\begin{aligned} \zeta_b &= \frac{x^2}{(1-x)(1+x)} [\nu_{n-1} - \eta_{b-1}] \\ \nu_b &= \frac{x^2}{(1-x)} \sum_{l=1}^{b} \zeta_l - \nu_{b-l} \end{aligned} \tag{71}$$

One easily checks that these generating functions are of the form

$$\begin{aligned} \nu_b(x) &= \mu_b(x)\frac{1}{(x+1)^b}\frac{1}{(x-1)^{3b+1}} \\ \zeta_b(x) &= \xi_b(x)\frac{1}{(x+1)^b}\frac{1}{(x-1)^{3b+1}} \end{aligned} \tag{72}$$

where $\mu_b(x)$ and $\xi_b(x)$ are polynominals. Theorem 4.2 thus yields

$$N_n(b) \sim \frac{1}{2^b}\frac{\mu_b(1)}{\Gamma(3b+1)} \cdot n^{3b} \tag{73}$$

where $\mu_b(1)$ and $\xi_b(1)$ fulfill the recursions

$$\xi_b(1) = \mu_{b-1}(1) \qquad \mu_n(1) = \sum_{l=1}^{b} \xi_l(1)\mu_{b-l}(1) = \sum_{l=0}^{b-1} \mu_l(1)\mu_{b-l-1}(1) \tag{74}$$

Again, the coefficients $\mu_b(1)$ coincide with the Catalan numbers. □

THEOREM 4.22. *The number of structures with b hairpins is*

$$A_n(b) \sim \frac{4}{2^{(3+m)b}b!(b-1)!}n^{2(b-1)}2^n \tag{75}$$

Proof. Let $a_b(x)$ denote the generating function $\sum A_n(b)x^n$. From recursion (8) we obtain with some simple algebra

$$a_b = xa_b + x^2\sum_{i=1}^{b} a_i a_{b-i} + x^2 t_m a_{b-1} \qquad b>0 \tag{76}$$

and $a_0(x) = 1/(1-x)$. Collecting all terms containing $a_b(x)$ yields

$$a_b(1-2x) = x^{m+2}a_{b-1} + x^2\sum_{i=1}^{b-1} a_i a_{b-i} \tag{77}$$

With the ansatz

$$a_b(x) = \left(\frac{x^{m+2}}{1-x}\right)^b \frac{1}{(1-2x)^{2b-1}}\eta_b(x) \tag{78}$$

we find the following recursion for the polynomials $\eta_b(x)$:

$$\eta_b(x) = (1-2x)(1-x)\eta_{b-1} + x^2\sum_{i=1}^{b-1}\eta_i(x)\eta_{b-1} \qquad \eta_1(x) = 1 \tag{79}$$

Theorem 4.2 now implies that the relevant singularity occurs at $x = 1/2$ leaving us with the recursion

$$\eta_b(\tfrac{1}{2}) = \frac{1}{4}\sum_{i=1}^{b-1}\eta_i(\tfrac{1}{2})\eta_{b-i}(\tfrac{1}{2}) \tag{80}$$

It is easy to check that this is solved by

$$\eta_b(\tfrac{1}{2}) = \frac{1}{2^{2(b-1)}} C_{b-1} \tag{81}$$

From theorem 4.2 we find now that

$$A_n(b) \sim \frac{C_{b-1}}{2^{2(b-1)} 2^{b(m+1)} \Gamma(2b+1)} n^{2(b-1)} 2^n \tag{82}$$

Some simple algebra completes the proof. □

THEOREM 4.23. *The number of structures with b components, $J_n(b)$, fulfils*

$$\lim_{n\to\infty} J_n(b)/S_n = \frac{\alpha^2}{(1-\alpha)^3} b \left(\frac{1-2\alpha}{1-\alpha}\right)^{b-1} \tag{83}$$

Proof. Let $j_b(x) = \sum_{n=0}^{\infty} J_n(b) x^n$ be the generating function for the number of secondary structures with exactly b components. Its is straight forward to derive

$$j_b(x) = \frac{x^2}{1-x} \left(s - t_m(x)\right)^b \cdot j_0(x) \qquad b \geq 1 \tag{84}$$

and from $J_n(0) = 1$ we obtain $j_0(x) = 1/(1-x)$. From theorem 4.5 we find that

$$\lim_{n\to\infty} J_n(b)/S_n = \frac{1}{1-\alpha} \left(\frac{\alpha^2}{1-\alpha}\right)^b \cdot b \left(\beta - t_m(\alpha)\right)^{b-1} \tag{85}$$

□

THEOREM 4.24. *The number of structures with b external digits, $E_n(b)$, fulfils*

$$\lim_{n\to\infty} E_n(b)/S_n = \frac{1}{4}(b+1)\left(\frac{1}{2}\right)^b \tag{86}$$

Proof. Let $e_b(x)$ be the generating function of the number of secondary structures with exactly b external digits. Recursion (21) yields the functional equation

$$e_b - \delta_{0b} = x e_{b-1} + x^2 s e_b - x^2 e_b t_m(x) \tag{87}$$

Substituting the functional equation for s and some algebra finally yields $e_0 = s/(1+xs)$ and $e_b = \frac{xs}{1+xs} e_{b-1}$. Therefore,

$$e_b = \left(\frac{xs}{1+xs}\right)^b \cdot \frac{s}{1+xs} \tag{88}$$

Application of theorem 4.5 and observing $\alpha\beta = 1$ yields the desired expression. □

THEOREM 4.25. *For any finite order ω there is a positive constant ϵ such that*

$$\lim_{n\to\infty} \frac{\tilde{D}_n(\omega-1) e^{\epsilon n}}{\tilde{D}_n(\omega)} = 0 \tag{89}$$

Proof. We will need the generating functions

$$\Delta_\omega = \sum_{n=0}^{\infty} \tilde{D}_n(\omega) x^n \quad \Delta^*_\omega = \sum_{n=0}^{\infty} D^*_n(\omega) x^n \quad \Delta'_\omega = \sum_{n=0}^{\infty} D_n(1,\omega) x^n \tag{90}$$

TABLE 2

Secondary structures with order ω.
The base of the exponential part of the asymptotic is given.

	α_ω		
ω	$m=0$	$m=1$	$m=3$
0	1	1	1
1	0.41421256	0.4533977	0.4863890
2	0.37597060	0.4221456	0.4680050
3	0.35978154	0.4076474	0.4577424

Recursion (24) yields the following system of coupled functional equations for the above generating functions

$$\begin{aligned} \Delta_\omega &= x\Delta_\omega + x^2\Delta^*_\omega \sum_{i=0}^{\omega-1}\Delta_i + x^2\Delta_\omega\sum_{i=0}^{\omega}\Delta^*_i \\ \Delta^*_\omega &= \Delta_{\omega-1} + \Delta'_\omega - \Delta'_{\omega-1} \qquad \omega\geq 2 \\ \Delta'_\omega &= x\Delta'_\omega + x^2\Delta^*_\omega\frac{1}{1-x} \end{aligned} \tag{91}$$

For $\omega=0$ we have $\Delta_0 = 1/(1-x)$ and for $\omega=1$ we find explicitly

$$\begin{aligned} \Delta^*_1(x) &= \frac{1-x}{1-2x}x^m \\ \Delta_1(x) &= \frac{x^{m+2}}{1-x}\cdot\frac{1}{1-2x-x^{m+2}} \end{aligned} \tag{92}$$

Eliminating Δ'_ω we find for $\omega\geq 2$

$$\begin{aligned} \Delta^*_\omega &= \frac{(1-x)^2}{1-2x}\Delta_{\omega-1} - \frac{x^2}{1-2x}\Delta^*_{\omega-1} \\ \Delta_\omega &= \frac{x^2\Delta^*_\omega\sum_{i=0}^{\omega-1}\Delta_i}{1-x-x^2\sum_{i=0}^{\omega}\Delta^*_i} \end{aligned} \tag{93}$$

Unfortunately these expressions become to clumsy to be of much practical use.
Denote $f_\omega(x) = 1-x-x^2\sum_{i=0}^{\omega}\Delta^*_i$ and let λ be the unique solution of $1-2x-x^{m+2}$ in the interval $[0,1/2[$. Obviously $f_\omega(x)$ is strictly monotone decreasing and has at least one zero in $(0,\alpha^*)$, where α^* denotes the position of the singularity with the smallest x value among the function $\Delta_i(x)$, $i<\omega$. Therefore, $\Delta_\omega(x)$ has a singularity $\alpha_\omega<\alpha^*$. By induction, therefore, $\alpha_\omega<\alpha_{\omega-1}$ for all ω, since explicitly we have $\alpha_1=\lambda$ and the first singularity in Δ^*_ω occurs at $x=\alpha_{\omega-1}$. By theorem 4.2 we have $\Delta_n(\omega)\sim c_1 n^{c_2}\alpha^n_\omega$. The inequality $1/\alpha_\omega > 1/\alpha_{\omega-1}$ completes the proof. □

Numerical results for the constants obtained for calculating $\Delta_\omega(x)$ explicitly by using `Mathematica` and numerically solving for the smallest zero of the denominator in (93,2) are tabulated in table 2. The case $m=1$, $\omega=1$ has been calculated by Waterman [1,3].

4.5. The Distribution of Structure Elements.

THEOREM 4.26. *The number of loops with b unpaired digits, $Q_n(b)$, fulfils*

$$\lim_{n\to\infty}\frac{Q_n(b)}{N_n} = \frac{\alpha^2}{(1-\alpha^2)(1-2\alpha)}\left[\frac{1}{2\alpha\cdot 2^b} - \Theta(m-b)\alpha^b - (1-2\alpha)\delta_{b0}\right] \tag{94}$$

Proof. Let $q_b(x) = \sum_{k=0}^{\infty} Q_n(b)x^n$ denote the generating function for the number of loops with b unpaired digits. From recursion (13) we find immediately

$$\begin{aligned} q_b &= xq_b + 2x^2sq_b + x^2se_b - x^2q_bt_m - \Theta(m-b)x^b \\ q_0 &= xq_0 + 2x^2sq_0 + x^2se_0 - x^2q_bt_m - \Theta(m) - x^2[s(1-x)-1] \end{aligned} \tag{95}$$

where $\Theta(n)$ denote the Heaviside function, $\Theta(n) = 1$ for $n > 0$ and $\Theta(n) = 0$ for $n \leq 0$. Some simple algebra confirms

$$\begin{aligned} q_b &= \frac{1}{1-x^2s^2}x^2s^2[e_b - \Theta(m-b)x^b] \qquad b > 0 \\ q_0 &= \frac{1}{1-x^2s^2}x^2s[se_b - s(1-x) + 1 - \Theta(m-0)] \end{aligned} \tag{96}$$

Inserting e_b from equ. (88) and using theorem 4.6 now proves the assertion. □

THEOREM 4.27. *The asymptotic distribution of stack length is exponential:*

$$\lim_{n\to\infty} \frac{W_n(b)}{N_n} = \frac{1-\alpha^2}{\alpha^2}\alpha^{2b} \tag{97}$$

Proof. Let $w_b(x) = \sum_{k=0}^{\infty} W_n(b)x^n$ denote the generating function for the number of stacks of length b. From recursion (16) we find

$$w_b = xw_b + 2x^2sw_b - x^2w_bt_m + (x^{2b+2} - 2x^{2b} + x^{2b-2})[(1-x)s-1] \tag{98}$$

Some simple algebra assure that

$$w_b = \frac{1}{1-x^2s^2}x^{2b-2}s(1-x^2)^2[(1-x)s-1] = x^{2b-2}(1-x^2)\cdot\nu(x) \tag{99}$$

Corollary 4.7 completes the proof. □

4.6. Loop Types.

THEOREM 4.28. *The distribution of loop degrees fulfils*

$$\begin{aligned} \lim_{n\to\infty} \frac{L_n(d)}{N_n} &= \frac{\alpha^2}{(1-\alpha^2)(1-2\alpha)}\times \\ &\times \left[\frac{1}{1-2\alpha}\left(\frac{1-2\alpha}{1-\alpha}\right)^d - \begin{cases} \frac{3\alpha-1}{\alpha^2} & d=1 \\ (1-2\alpha) & d=2 \\ 0 & d>2 \end{cases} \right] \end{aligned} \tag{100}$$

Proof. Let $\ell_d(x) = \sum_{n=0}^{\infty} L_n(d)x^n$ be the generating function for the number of loops with degree d. For hairpins one finds from recursion (18)

$$\ell_1 = x\ell_1 + 2x^2\ell_1 s - x^2\ell_1 t_m(x) + \frac{x^{m+2}}{1-x}s \tag{101}$$

Similar functional equations can be obtained for loops of higher degree from recursions (19) and (20). They can be brought to the form

$$\begin{aligned} \ell_1 &= \frac{1}{1-x^2s^2}\frac{x^{m+2}}{1-x}s^2 \\ \ell_2 &= \frac{1}{1-x^2s^2}\left[x^2s^2[j_1(x) - (1-x)] + x^2s\right] \\ \ell_d &= \frac{1}{1-x^2s^2}x^2s^2j_{d-1}(x) \end{aligned} \tag{102}$$

Using the explicit expressions for j_d and theorem 4.6, some tedious algebra finally yield equ.(100). □

Remark 4.29. The average loop degree $\bar{d}$ can be most easily calculated from the following balance equation which holds for all secondary structures

$$\sum_{\text{loops } \lambda} \deg(\lambda) = 2\#[\text{stacks}] - \#[\text{components}] \tag{103}$$

From equ.(52) and equ.(60) we find immediately that the average loop degree fulfills

$$\lim_{n\to\infty} \bar{d}_n = 2 \tag{104}$$

THEOREM 4.30. *The ratio of bulges and true interior loops fulfills*

$$\lim_{n\to\infty} \frac{B_n}{Y_n} = \frac{2}{\alpha}(1-\alpha) \tag{105}$$

Proof. Denote by $b(x)$ and $y(x)$ the generating function for the number of bulges and interior loops respectively. By construction they fulfil $b(x) + y(x) = \ell_2(x)$. It is thus sufficient to calculate $y(x)$ from recursion (18). We find

$$y(x) = \frac{1}{1 - x^2 s^2} s^2 x^4 j_1(x) \tag{106}$$

and thus

$$b(x) = \ell_2(x) - y(x) = \frac{1}{1 - x^2 s^2} x^2 s \left[s(1 - x^2) j_1 - (1 - x)s + 1 \right] \tag{107}$$

Corollary 4.7 and some algebra complete the proof. □

5. Secondary Structures on a Sequence.

DEFINITION 5.1. Let $\mathcal{A}$ be some finite alphabet of size κ, let Π be a symmetric Boolean $\kappa \times \kappa$-matrix and let $\Sigma = [\sigma_1 \dots \sigma_N]$ be a string of length N over $\mathcal{A}$. A secondary structure is *compatible* with the sequence Σ if for all base pairs (p, q) holds $\Pi_{\sigma_p,\sigma_q} = 1$.

Following [4,18] the number of secondary structures $\mathcal{S}$ compatible with some string can be enumerated as follows: Denote by $S_{p,q}$ the number of structures compatible with the substring $[\sigma_p \dots \sigma_q]$. Then

$$S_{l,n+1} = S_{l,n} + \sum_{k=l}^{n-m} S_{l,k-1} S_{k+1,n} \Pi_{\sigma_k,\sigma_{n+1}} \tag{108}$$

For a random sequence, the expected number $\bar{S}_n$ of compatible structures is then [21]

$$\bar{S}_{n+1} = \bar{S}_n + p \sum_{k=1}^{n-m} \bar{S}_{k-1} \bar{S}_{n-k} = \bar{S}_n + p \sum_{k=m}^{n-1} \bar{S}_k \bar{S}_{n-k-1} \tag{109}$$

where

$$p = \frac{1}{\kappa^2} \sum_{i,j=1}^{\kappa} \Pi_{ij} \tag{110}$$

is called the *stickiness* [22].

Remark 5.2. A secondary structure compatible with a given sequence with maximal number of base pairs can be deterined by a dynamic programming algorithm [23]. This observation was the starting point for the construction of reliable energy-directed folding algorithms (for a review see, e.g., [21]).

Analogously one can derive recursions of the number of structure elements compatible with a string Σ. All recursions in this paper are sums of linear terms of the form A_n and quadratic terms of the type

$$\sum_{k=m}^{n-1} B_k C_{n-k-1} = \sum_{k=1}^{n-m} C_{k-1} B_{n-k} \tag{111}$$

The corresponding recursions for structures compatible with a string can then be found by the rule

$$\begin{aligned} A_n &\longrightarrow A_{l,n} \\ \sum_{k=1}^{n-m} C_{k-1} B_{n-k} &\longrightarrow \sum_{k=l}^{n-m} C_{l,k-1} B_{k+1,n} \Pi_{\sigma_k,\sigma_{n+1}} \end{aligned} \tag{112}$$

For expected numbers for random sequences the above rules simplify to

$$\begin{aligned} A_n &\longrightarrow A_n \\ \sum_{k=m}^{n-1} B_k C_{n-k-1} &\longrightarrow p \sum_{k=m}^{n-1} B_k C_{n-k-1} \end{aligned} \tag{113}$$

As a single example we calculate the expected proportion of unpaired digits in a secondary structure compatible with a random sequence with stickiness p. Application of the above rules to equ.(9) immediately yields

$$\begin{aligned} U_{n+1} &= (U_n + S_n) + p \sum_{k=m}^{n-1} [S_k U_{n-k-1} + S_{n-k-1} U_k], \qquad n \geq m+1 \\ U_n &= n, \quad n \leq m+1 \end{aligned} \tag{114}$$

From equ.(109) and equ.(114) we obtain the generating functions

$$1 = s[1 - x - px^2 s + px^2 t_m] \tag{115}$$

$$u = xu + xs + p[2x^2 us - x^2 u t_m - x^2 s \tau_m] \tag{116}$$

For α and β we find from the functional equation for s

$$\begin{aligned} &\alpha\beta = 1/\sqrt{p} \\ &\frac{1}{\sqrt{p}} - (2 + \frac{1}{\sqrt{p}})\alpha + \sqrt{p}\alpha^2 t_m(\alpha) = 0 \end{aligned} \tag{117}$$

Theorem 4.3 allows to calculate the asymptotic for $\bar{S}_n(p)$ from s in equation (115). Furthermore we have the following generalization of lemma lemma 4.13 for arbitrary $p \neq 1$:

TABLE 3
Asymptotics of some structure elements
as a function of stickiness

p	1	0.5 **GC**	0.375 **GCAU**	0.25 **GCXK**
α	0.4369	0.5092	0.5391	0.5809
U_n/nS_n	0.5265	0.5897	0.6147	0.6487
P_n/nS_n	0.2368	0.2051	0.1926	0.1756
N_n/nS_n	0.1915	0.1786	0.1717	0.1608
I_n/S_n	1.5776	1.7266	1.7918	1.8855
$L_n(1)/N_n$	0.2769	0.3062	0.3183	0.3352
$L_n(2)/N_n$	0.5082	0.4692	0.4537	0.4325
B_n/Y_n	2.5776	1.9280	1.7096	1.4428
StackLength	1.2363	1.1487	1.1220	1.0924
LoopSize	2.7493	3.3018	3.5801	4.0342
E_n/S_n	2	2.828	3.266	4

LEMMA 5.3.

$$\begin{aligned} t_m(\alpha) &= \frac{(1+2\sqrt{p})\alpha-1}{p\alpha^2} \\ \tau_m(\alpha) &= \frac{\alpha-1+2\alpha\sqrt{p}}{\alpha p(1-\alpha)} - m\frac{\left(\alpha-1+\alpha\sqrt{p}\right)^2}{\alpha^2 p(1-\alpha)} \\ g^2(\alpha) &= \frac{(1-\alpha-\sqrt{p}\alpha)(2+m(1-\alpha-\sqrt{p}\alpha))}{\sqrt{p}^3(1-\alpha)\alpha^3} \end{aligned} \tag{118}$$

Equation (116) thus simplifies to

$$u = \frac{s^2 x\,(1-p\,\tau_m x)}{1-ps^2x^2} \tag{119}$$

and theorem 4.6 shows that

$$\lim_{n\to\infty}\frac{U_n}{nS_n} = \frac{1}{\alpha g^2(\alpha)p}\left[\frac{1}{\sqrt{p}\alpha} - \sqrt{p}\tau_m(\alpha)\right] = \frac{2\alpha+m(1-\alpha-\sqrt{p}\alpha)}{2+m(1-\alpha-\sqrt{p}\alpha)} \tag{120}$$

Number of Loops and stacks:

$$\begin{aligned} \nu &= \frac{s\,(1-s(1-x))\,(px^2-1)}{1-ps^2x^2} \\ \lim_{n\to\infty}\frac{N_n}{S_n} &= \frac{(1-\alpha)(1-\alpha^2 p)}{2+m(1-\alpha-\alpha\sqrt{p})} \end{aligned} \tag{121}$$

Number of components:

$$\begin{aligned} i &= s^2(1-x)-s \\ \lim_{n\to\infty}\frac{I_n}{S_n} &= 2\beta(1-\alpha)-1 \end{aligned} \tag{122}$$

Loops with degree 2, i.e., interior loops and bulges:

$$
\begin{aligned}
l_2 &= \frac{psx^2\left[(1-x)^2 - s(1-x)^3 + psx^2(s-t_m)\right]}{(1-x)^2(1-ps^2x^2)} \\
\lim_{n\to\infty}\frac{L_n(2)}{N_n} &= \frac{(2-\alpha)\alpha^3 p}{(1-\alpha)^2(1-\alpha^2 p)} \\
\lim_{n\to\infty}\frac{B_n}{Y_n} &= 2/\alpha - 2
\end{aligned}
\tag{123}
$$

Hairpins:

$$
\begin{aligned}
l_1 &= \frac{ps^2x^2\,(1-(1-x)t_m)}{(1-x)\,(1-ps^2x^2)} \\
\lim_{n\to\infty}\frac{L_n(1)}{N_n} &= \frac{1-\alpha-\alpha\sqrt{p}}{1-\alpha-\alpha^2\,p+\alpha^3\,p}
\end{aligned}
\tag{124}
$$

Numerical values for the stickiness of the most important alphabets are given in table 2.

Acknowledgements.
Stimulating discussions with Drs. Walter Fontana and Danielle Konings are greatfully acknowledged.

REFERENCES

[1] M. S. Waterman, *Secondary Structure of Single-Stranded Nucleic Acids*, Adv. Math. Suppl. Studies 1 (1978), pp. 167–212.

[2] P. R. Stein and M. S. Waterman. *On Some New Sequences Generalizing the Catalan and Motzkin Numbers*, J. Discr. Math., 26 (1978), pp. 261–272.

[3] M. S. Waterman, *Combinatorics of RNA Hairpins and Cloverleaves*, Studies Appl. Math., 60 (1978), pp. 91–96.

[4] M. S. Waterman and T. F. Smith, *RNA Secondary Structure. A Complete Mathematical Analysis.* Math. Biosc., 42 (1978), pp. 257–266.

[5] W. R. Schmitt and M. S. Waterman, *Plane Trees and RNA Secondary Structure*, Preprint 1992.

[6] J. Touchard, *Sur une Problème de Configurations et sur les Fractions Continues*, Canad. J. Math., 4 (1952), pp. 2–25.

[7] D. Kleitman, *Proportions of Irreducible Diagrams*, Studies in Appl. Math., 49 (1970), pp. 297–299.

[8] W. N. Hsieh, *Proportions of Irreducible Diagrams*, Studies in Appl. Math., 52 (1973), pp. 277–283.

[9] P. R. Stein, *On a Class of Linked Diagrams, I. Enumeration*, J. Comb. Theory A, 24 (1978), pp. 357–366.

[10] P. R. Stein and C. J. Everett, *On a Class of Linked Diagrams, II. Asymptotics*, Discr. Math., 22 (1978), pp. 309–318

[11] W. Fontana, T. Griesmacher, W. Schnabl, P. F. Stadler, and P. Schuster, *Statistics of Landscapes Based on Free Energies Replication and Degradation Rate Constants of RNA Secondary Structures*, Mh. Chem., 122, (1991), pp. 795–819.

[12] W. Fontana, D. A. M. Konings, P. F. Stadler, and P. Schuster, *Statistics of RNA Secondary Structures*, Biopolymers, 33, (1993), pp. 1389–1404.

[13] M. Tacker, P. F. Stadler, E. G. Bornberg-Bauer, I. L. Hofacker, and P. Schuster, *Robust Properties of RNA Secondary Structure Folding Algorithms*, Preprint, Santa Fe Institute (1994).

[14] P. Hogeweg and B. Hesper, *Energy Directed Folding of RNA Sequences*, Nucl. Acid Res., 12 (1984), pp. 67–74.

[15] R. Donaghey and L. W. Shapiro, *Motzkin Numbers*, J. Comb. Theor. A, 23 (1977), pp. 291–301.
[16] B. A. Shapiro, *An Algorithm for Comparing Multiple RNA Secondary Structures*, CABIOS, 4 (1988), pp. 387–397.
[17] B. A. Shapiro and K. Zhang, *Comparing Multiple RNA Secondary Structures Using Tree Comparison*, CABIOS, 6 (1990), pp. 309–318.
[18] J. A. Howell, T. F. Smith, and M. S. Waterman, *Computation of Generating Functions for Biological Molecules*, SIAM J. Appl. Math., 39 (1980), pp. 119–133.
[19] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Publ., vol. XXIII, rev. ed., Amer. Math. Soc., New York, 1959.
[20] E. A. Bender, *Asymptotic Methods in Enumeration*, SIAM Review, 16 (1974), pp. 485–515.
[21] M. Zuker and S. Sankoff, *RNA Secondary Structures and their Prediction*, Bull. Math. Biol., 46 (1984), pp. 591–621.
[22] A. M. Lesk, *A Combinatorial Study of the Effects af Admitting Non-Watson-Crick Base Pairings and of Base Composition on the Helix-Forming Potential of Polynucleotides of Random Sequences*, J. Theor. Biol., 44 (1974), pp. 7–17.
[23] R. Nussinov, G. Pieczenik, J. R. Griggs, and D. J. Kleitman, *Algorithms for Loop Matchings*, SIAM J. Appl. Math., 35 (1978), pp. 68–82.