On Reidys and Stadler's metrics for RNA contact structures*

Francesc Rosselló

Department of Mathematics and Computer Science,
Research Institute of Health Science (IUNICS),
University of the Balearic Islands,
07122 Palma de Mallorca (Spain)
E-mail: cesc.rossello@uib.es

Abstract

We compute explicitly several abstract metrics for RNA contact structures defined by Reidys and Stadler.

Keywords: RNA contact structure, RNA secondary structure, metric, distance, symmetric group.

1 Introduction

As it is well known, an RNA molecule can be viewed as a chain of (ribo)nucleotides with a definite orientation. Each of these nucleotides is characterized by (and in practice identified with) the base attached to it, which can be adenine (A), cytosine (C), guanine (G), or uracil (U). Thus, an RNA molecule with N nucleotides can be mathematically described as a word of length N over the alphabet $\{A, C, G, U\}$, called the *primary structure* of the molecule.

In the cell and *in vitro* each RNA molecule folds into a three-dimensional structure, which determines its biochemical function. This structure is held together by weak interactions called *hydrogen bonds* between pairs of non-consecutive bases: actually, a hydrogen bond can only form between bases that are several positions apart in the chain, but for simplicity we shall not take this restriction into account here. Most of these bonds form between *Watson-Crick complementary bases*, i.e., between A and U and between C and G, but a significant amount of bonds also form between other pairs of bases [1]. The *contact structure* of an RNA molecule [2] is a simplified model of this three-dimensional structure, consisting of an undirected graph with nodes its bases and arcs its base pairs or *contacts*; the *length* of a contact structure is the number of its nodes. A restriction, called the *unique bonds condition*, is added to the definition of RNA contact structure [3]: a base can only pair with at most one base. It is usual to impose a third restriction on *RNA secondary structures*, by forbidding the existence of (*pseudo*)*knots*, i.e., of contacts that "cross" each other: if $i < k < j < l$, then there cannot simultaneously

*This work has been partially supported by the Spanish DGES, grant BFM2000-1113-C02-01.

exist a contact between bases at the ith and jth positions in the primary structure and a contact between bases at the kth and lth positions. This restriction has its origin in the first dynamic programming methods to predict RNA secondary structures [3, 4]. Since real RNA structures can contain knots, which are moreover important structural elements in many RNA molecules [5], we shall not take either this restriction into account here.

An important problem in molecular biology is the comparison of these RNA contact structures, because it is assumed that a preserved three-dimensional structure corresponds to a preserved function. The measurement of the similarity of RNA contact structures of a fixed length has an interest in itself. For instance, it is used in the prediction of RNA secondary structures to reduce the output of alternate structures when suboptimal solutions, and not only optimal, are considered [6, §IX], and it can also be used to compare the output of different contact structure prediction algorithms applied to the same RNA molecule, to assess their performance.

In a seminal paper on the algebraic representation of biomolecular structures [7], C. Reidys and P. F. Stadler introduced three abstract metrics on the set of RNA contact structures of a fixed length based on their algebraic models and independent of any notion of graph edition, and they discussed their biophysical relevance. They ended that paper by asking, among other questions, whether there exists any relation between the metrics for RNA contact structures they had defined. In this paper we answer this question by explicitly computing these metrics. In a subsequent paper [8] we generalize one of these metrics to contact structures without unique bonds, as for instance protein contact structures.

2 Main results

From now on, let $[n]$ denote the set $\{1, \ldots, n\}$, for every positive integer n.

Definition 1 *An* RNA contact structure *of length n is an undirected graph without multiple edges or self-loops $\Gamma = ([n], Q)$, for some $n \geq 1$, whose arcs $\{j, k\} \in Q$, called* contacts*, satisfy the following two conditions:*

i) For every $j \in [n]$, $\{j, j+1\} \notin Q$.

ii) For every $j \in [n]$, if $\{j, k\}, \{j, l\} \in Q$, then $k = l$.

Condition (i) translates the impossibility of a contact between two consecutive bases, while condition (ii) translates the unique bonds condition. Our RNA contact structures have also been called in the literature *contact structures with unique bonds* [7, 9] and *1-diagrams* [10].

We shall denote from now on a contact $\{j, k\}$ by $j{\cdot}k$ or $k{\cdot}j$, without distinction. A node is said to be *isolated* in an RNA contact structure when it is not involved in any contact.

Given two contact structures of the same length $\Gamma_1 = ([n], Q_1), \Gamma_2 = ([n], Q_2)$, their *union* is the graph $\Gamma_1 \cup \Gamma_2 = ([n], Q_1 \cup Q_2)$; notice that it need not be a contact structure.

Let $\mathcal{C}_n$ stand for the set of all RNA contact structures of length n, S_n the symmetric group of permutations of $[n]$, and $\mathrm{Sub}(S_n)$ the set of subgroups of S_n.

Definition 2 *For every $\Gamma = ([n], Q) \in \mathcal{C}_n$, say with $Q = \{i_1{\cdot}j_1, \ldots, i_k{\cdot}j_k\}$, let*

$$\pi(\Gamma) = \prod_{t=1}^{k} (i_t, j_t) \in S_n,$$

where (i, j) denotes the transposition in S_n defined by $i \leftrightarrow j$.

Reidys and Stadler proved in [7] that the mapping $\pi : \mathcal{C}_n \to S_n$ is injective and that $\pi(\Gamma)$ is an involution for every $\Gamma \in \mathcal{C}_n$. This representation of RNA contact structures as involutions is then used by these authors to define the following metric, called the *involution metric*.

Proposition 1 *The mapping $d_{inv} : \mathcal{C}_n \times \mathcal{C}_n \to \mathbb{R}$ sending every $(\Gamma_1, \Gamma_2) \in \mathcal{C}_n^2$ to the least number $d_{inv}(\Gamma_1, \Gamma_2)$ of transpositions necessary to represent the permutation $\pi(\Gamma_1)\pi(\Gamma_2)$, is a metric.*

Our first goal is to compute explicitly this metric. To do this, we need to introduce a further notion. Given two RNA contact structures $\Gamma_1 = ([n], Q_1)$ and $\Gamma_2 = ([n], Q_2)$, consider the subgroup $D(\Gamma_1, \Gamma_2) = \langle \pi(\Gamma_1), \pi(\Gamma_2) \rangle \in \mathrm{Sub}(S_n)$ generated by the involutions associated to them. This subgroup, formed of permutations of $[n]$, acts on this set. Recall that when a group G acts on a set X, the *orbits* of X induced by this action are those subsets of X of the form $\{g(x) \mid g \in G\}$ for every $x \in X$ (where $g(x)$ denotes the image of x under the action of $g \in G$). In our case, the orbits of $[n]$ induced by the action of $D(\Gamma_1, \Gamma_2)$ are exactly those subsets $\{i_1, i_2, \ldots, i_m\} \subseteq [n]$, $m \geq 1$, such that

$$i_1 \cdot i_2, i_2 \cdot i_3, \ldots, i_{m-1} \cdot i_m \in Q_1 \cup Q_2$$

and maximal with this property, i.e., such that the only other contact in $Q_1 \cup Q_2$ involving some element of this subset can possibly be $i_1 \cdot i_m$. The unique bonds condition (or, in group-theoretical terms, the fact that the transpositions defining each $\pi(\Gamma_i)$ are pairwise disjoint) implies that if $\{i_1, i_2, \ldots, i_m\}$ is such an orbit, then $i_1 \cdot i_2, i_3 \cdot i_4, \ldots$ belong to one of the sets Q_1 or Q_2 and $i_2 \cdot i_3, i_4 \cdot i_5, \ldots$ to the other one.

Such an orbit is *closed* if $m = 2$ and $i_1 \cdot i_2 \in Q_1 \cap Q_2$, or $m \geq 3$ and $i_1 \cdot i_m \in Q_1 \cup Q_2$, and an orbit is *open* in all other cases. We shall call the cardinal of an orbit its *length*. The length of a closed orbit is always even: if $i_1 \cdot i_2 \in Q_1$ in a closed orbit $\{i_1, i_2, \ldots, i_m\}$ with $m \geq 3$, then $i_1 \cdot i_m \in Q_2$ and hence $i_{m-1} \cdot i_m \in Q_1$.

We shall say that a contact $i \cdot j \in Q_1 \cup Q_2$ is *involved* in an orbit when its vertices i, j belong to this orbit. Notice that every contact is involved in one, and only one, orbit, and that every pair of contacts in $Q_1 \cup Q_2$ sharing a node are involved in the same orbit, and a contact belongs to $Q_1 \Delta Q_2$ if and only if it is involved in an open orbit or in a closed orbit of length $m \geq 4$.

Notice finally that the orbits induced on $[n]$ by the action of $D(\Gamma_1, \Gamma_2)$ are exactly the connected components of the graph $\Gamma_1 \cup \Gamma_2$, and that that the closed orbits of length greater than 2 are exactly the cycles in this union of graphs.

Now, next proposition explicitly computes Reidys and Stadler's involution metric. In it, and henceforth, $A \Delta B$ denotes the symmetric difference $(A \cup B) - (A \cap B)$ of the sets A and B, and $|A|$ stands for the cardinal of the finite set A.

Proposition 2 *For every $\Gamma_1 = ([n], Q_1), \Gamma_2 = ([n], Q_2) \in \mathcal{C}_n$,*

$$d_{inv}(\Gamma_1, \Gamma_2) = |Q_1 \Delta Q_2| - 2\Omega(\Gamma_1, \Gamma_2),$$

where $\Omega(\Gamma_1, \Gamma_2)$ is the number of cycles in the graph $\Gamma_1 \cup \Gamma_2$.

Proof. As we have already mentioned, the cycles in $\Gamma_1 \cup \Gamma_2$ are exactly the closed orbits induced on $[n]$ by the action of $D(\Gamma_1, \Gamma_2)$ and thus, to prove the statement, we shall use a group-theoretical argument. To simplify the language, we shall refer to the orbits induced by the action on $[n]$ of $D(\Gamma_1, \Gamma_2)$ simply by orbits. To begin with, notice that if two transpositions appearing in the product $\pi(\Gamma_1)\pi(\Gamma_2)$ are not disjoint, then the indexes involved in them belong to the same orbit. Moreover, two disjoint transpositions always commute. This allows us to reorganize the transpositions in the product $\pi(\Gamma_1)\pi(\Gamma_2)$, assembling them into subproducts corresponding to orbits. More specifically, if for every orbit O and for every $i = 1, 2$ we let

$$\pi(O, \Gamma_i) = \prod_{\substack{k \cdot l \in Q_i \\ k,l \in O}} (k, l),$$

then

$$\pi(\Gamma_1)\pi(\Gamma_2) = \prod_{O \in \{\text{orbits}\}} \pi(O, \Gamma_1)\pi(O, \Gamma_2).$$

Since the orbits are pairwise disjoint, this finally shows that the least number of transpositions which $\pi(\Gamma_1)\pi(\Gamma_2)$ decomposes into is equal to the sum of the least numbers of transpositions which $\pi(O, \Gamma_1)\pi(O, \Gamma_2)$ decompose into, for every orbit O. It remains to compute this last number for each type of orbit O.

If O is an open orbit of length $m = 1$, then $\pi(O, \Gamma_1)\pi(O, \Gamma_2) = \mathrm{Id}$, and it corresponds to a node that is isolated both in Γ_1 and in Γ_2.

Let now $O = \{i_1, \ldots, i_m\}$ be an open orbit of length $m \geq 2$. Consider first the case when $i_1 \cdot i_2, i_3 \cdot i_4, \ldots, i_{m-1} \cdot i_m \in Q_1$ and $i_2 \cdot i_3, i_4 \cdot i_5, \ldots \in Q_2$; in particular, m is even. Then

$$\begin{aligned} \pi(O, \Gamma_1)\pi(O, \Gamma_2) &= (i_1, i_2)(i_3, i_4) \cdots (i_{m-1}, i_m)(i_2, i_3) \cdots (i_{m-2}, i_{m-1}) \\ &= (i_2, i_4, \ldots, i_m, i_{m-1}, i_{m-3}, \ldots, i_3, i_1), \end{aligned}$$

a cycle of length m that decomposes into the product of $m-1$ transpositions (and it is the least number of transpositions required to represent it), which is exactly the number of contacts of $Q_1 \cup Q_2$ involved in this orbit.

A similar argument shows that in all other cases for an open orbit O, the permutation $\pi(O, \Gamma_1)\pi(O, \Gamma_2)$ is equal to a cycle of length the number of elements of the orbit, and thus the least number of transpositions this product decomposes into is equal to the number of contacts of $Q_1 \cup Q_2$ involved in this orbit O, all of them belonging to $Q_1 \Delta Q_2$.

If O is a closed orbit of length $m = 2$, say $O = \{i_1, i_2\}$, then $\pi(O, \Gamma_1)\pi(O, \Gamma_2) = (i_1, i_2)(i_1, i_2) = \mathrm{Id}$. Notice that closed orbits of length 2 correspond to contacts in $Q_1 \cap Q_2$.

Finally, assume that O is a closed orbit of length $m \geq 3$, say $O = \{i_1, \ldots, i_m\}$ with $i_1 \cdot i_2, i_3 \cdot i_4, \ldots, i_{m-1} \cdot i_m \in Q_1$ and $i_2 \cdot i_3, \ldots, i_{m-2} \cdot i_{m-1}, i_m \cdot i_1 \in Q_2$; remember that m is in this case even. Then

$$\begin{aligned} \pi(O, \Gamma_1)\pi(O, \Gamma_2) &= (i_1, i_2)(i_3, i_4) \cdots (i_{m-1}, i_m)(i_2, i_3) \cdots (i_{m-2}, i_{m-1})(i_m, i_1) \\ &= (i_2, i_4, \ldots, i_m)(i_{m-1}, i_{m-3}, \ldots, i_3, i_1), \end{aligned}$$

the product of two disjoint cycles of length $m/2$. Since each cycle requires $m/2 - 1$ transpositions, the least number of transpositions the permutation $\pi(O, \Gamma_1)\pi(O, \Gamma_2)$ decomposes into is equal to $m - 2$, the number of contacts of $Q_1 \cup Q_2$ involved in this orbit O (all of them belonging again to $Q_1 \Delta Q_2$) minus 2.

To sum up,

$$\begin{aligned} d_{inv}(\Gamma_1,\Gamma_2) &= |\{\text{contacts involved in linear orbits}\}| \\ &\quad +|\{\text{contacts involved in closed orbits of length ¿ 2}\}| - 2\Omega(\Gamma_1,\Gamma_2) \\ &= |Q_1 \Delta Q_2| - 2\Omega(\Gamma_1,\Gamma_2), \end{aligned}$$

as we claimed. ∎

The number and structure of the orbits induced by the action of $\langle \pi(\Gamma_1), \pi(\Gamma_2)\rangle$ on $[n]$ are related to the probability of transition from the neutral network of Γ_1 (the set of sequences that fold into it) to that of Γ_2: see [7, §3] and the references cited therein.

Definition 3 *For every $\Gamma = ([n], Q) \in \mathcal{C}_n$, say with $Q = \{i_1 \cdot j_1, \ldots, i_k \cdot j_k\}$, let*

$$T(\Gamma) = \{(i_1, j_1), \ldots, (i_k, j_k)\}$$

be the set of the transpositions corresponding to the contacts in Q and let $G(\Gamma) = \langle T(\Gamma)\rangle$ be the subgroup of S_n generated by this set of transpositions.

Reidys and Stadler also proved in [7] that the mapping $G : \mathcal{C}_n \to \mathrm{Sub}(S_n)$ is injective, and then they used this representation of RNA contact structures as permutation subgroups to define the following *subgroup metric*.

Proposition 3 *The mapping $d_{sgr} : \mathcal{C}_n \times \mathcal{C}_n \to \mathbb{R}$ defined by*

$$d_{sgr}(\Gamma_1,\Gamma_2) = \ln\left(\frac{|G(\Gamma_1)\cdot G(\Gamma_2)|}{|G(\Gamma_1)\cap G(\Gamma_2)|}\right)$$

is a metric.

Next proposition shows that this metric simply measures, up to a constant factor, the cardinal of the symmetric difference of the sets of contacts.

Proposition 4 *For every $\Gamma_1 = ([n], Q_1), \Gamma_2 = ([n], Q_2) \in \mathcal{C}_n$,*

$$d_{sgr}(\Gamma_1,\Gamma_2) = (\ln 2)|Q_1 \Delta Q_2|.$$

Proof. Since the transpositions generating a group $G(\Gamma)$, with $\Gamma \in \mathcal{C}_n$, are pairwise disjoint, there is a bijection between $G(\Gamma)$ and the powerset $\mathcal{P}(T(\Gamma))$: each element of $G(\Gamma)$ is the product of a subset of $T(\Gamma)$ in a unique way. Hence, $|G(\Gamma_1)| = 2^{|Q_1|}$ and $|G(\Gamma_2)| = 2^{|Q_2|}$.

On the other hand, by the uniqueness of the decomposition of a permutation into a product of disjoint cycles, a permutation belongs to $G(\Gamma_1) \cap G(\Gamma_2)$ if and only if it is a product of transpositions belonging to both $G(\Gamma_1)$ and $G(\Gamma_2)$. Therefore,

$$G(\Gamma_1)\cap G(\Gamma_2) = \langle T(\Gamma_1)\cap T(\Gamma_2)\rangle = \langle (i,j) \mid i\cdot j \in Q_1 \cap Q_2\rangle,$$

and then, arguing as in the previous paragraph, we see that $|G(\Gamma_1)\cap G(\Gamma_2)| = 2^{|Q_1\cap Q_2|}$.

Now, it is well known that

$$|G(\Gamma_1)\cdot G(\Gamma_2)| = \frac{|G(\Gamma_1)|\cdot|G(\Gamma_2)|}{|G(\Gamma_1)\cap G(\Gamma_2)|},$$

and hence

$$d_{sgr}(\Gamma_1,\Gamma_2) = \ln\left(\frac{|G(\Gamma_1)|\cdot|G(\Gamma_2)|}{|G(\Gamma_1)\cap G(\Gamma_2)|^2}\right) = \ln 2^{|Q_1|+|Q_2|-2|Q_1\cap Q_2|} = \ln 2^{|Q_1\Delta Q_2|},$$

as we claimed. ∎

Notice in particular that, should Reidys and Stadler had defined their subgroup metric as $\log_2(|G(\Gamma_1)\cdot G(\Gamma_2)|/|G(\Gamma_1)\cap G(\Gamma_2)|)$, it would coincide with $|Q_1\Delta Q_2|$.

The third metric on $\mathcal{C}_n$ proposed by Reidys and Stadler is actually a general way of defining metrics, rather than a single one, and it uses Magarshak and coworkers' algebraic representation of RNA contact structures [11, 12], recently extended in [13] to cope with contacts other than Watson-Crick complementary base pairs. These authors represent an RNA contact structure $\Gamma=([n],Q)$ as an $n\times n$ complex symmetric matrix $S_\Gamma=(s_{i,j})_{i,j=1,\ldots,n}$ where

$$s_{i,j}=\begin{cases}-1 & \text{if } i\neq j \text{ and } i\cdot j\in Q\\ 1 & \text{if } i=j \text{ and } i\cdot l\notin Q \text{ for every } l\\ 0 & \text{otherwise}\end{cases}$$

Since $S_\Gamma^{-1}=S_\Gamma$ for every $\Gamma\in\mathcal{C}_n$, one can define for any $\Gamma_1,\Gamma_2\in\mathcal{C}_n$ the *transfer matrix* $T_{\Gamma_1,\Gamma_2}=S_{\Gamma_2}\circ S_{\Gamma_1}$. Then, Reidys and Stadler propose to measure the difference between two RNA contact structures by defining a metric through

$$(\Gamma_1,\Gamma_2)\mapsto\|T_{\Gamma_1,\Gamma_2}\|,$$

where $\|\cdot\|$ stands for some *length function* on the group $GL(n,\mathbb{C})$ of $n\times n$ invertible complex matrices [7, Def. 9, Lem. 6] (actually, Reidys and Stadler propose to use a matrix norm $\|\cdot\|$, but it is probably a misprint, as it would not yield a metric). A simple and well-known length function on $GL(n,\mathbb{C})$ is

$$\|A\|=\operatorname{rank}(A-\mathrm{Id}),$$

which allows to define a *matrix metric* on $\mathcal{C}_n$

$$d_{mag}(\Gamma_1,\Gamma_2)=\operatorname{rank}(T_{\Gamma_1,\Gamma_2}-\mathrm{Id}).$$

This metric turns out to be equal to the involution metric d_{inv} defined above.

Proposition 5 *For every $\Gamma_1,\Gamma_2\in\mathcal{C}_n$, $d_{mag}(\Gamma_1,\Gamma_2)=d_{inv}(\Gamma_1,\Gamma_2)$.*

The proof of this proposition is similar to (and simpler than) the proof of [13, Thm. 17], which establishes essentially this equality for the generalized algebraic representation of RNA contact structures in the sense of Magarshak introduced in that paper, and therefore we omit it.

3 Conclusion

In this paper we have computed explicitly Reidys and Stadler's involution d_{inv}, subgroup d_{sgr} and matrix d_{mag} metrics for RNA contact structures of the same length. As it has turned out, for every $\Gamma_1=([n],Q_1),\Gamma_2=([n],Q_2)\in\mathcal{C}_n$,

$$d_{sgr}(\Gamma_1,\Gamma_2)=(\ln 2)|Q_1\Delta Q_2|,\quad d_{inv}(\Gamma_1,\Gamma_2)=d_{mag}(\Gamma_1,\Gamma_2)=|Q_1\Delta Q_2|-2\Omega(\Gamma_1,\Gamma_2)$$

where $\Omega(\Gamma_1,\Gamma_2)$ is the number of cycles in the graph $\Gamma_1\cup\Gamma_2$. In particular,

$$d_{inv}(\Gamma_1,\Gamma_2)=\frac{1}{\ln 2}d_{sgr}(\Gamma_1,\Gamma_2)-2\Omega(\Gamma_1,\Gamma_2).$$

What can be said about $\Omega(\Gamma_1, \Gamma_2)$? To begin with, notice that, although it is the difference of two metrics, it is not even a pseudometric: if $\Gamma_0 = ([n], \emptyset)$, then $\Omega(\Gamma_0, \Gamma) = 0$ for every $\Gamma \in \mathcal{C}_n$, while there are pairs $\Gamma_1, \Gamma_2 \in \mathcal{C}_n$ such that $\Omega(\Gamma_1, \Gamma_2) > 0$, which shows that Ω does not satisfy the triangle inequality.

Notice also that

$$0 \leq \Omega(\Gamma_1, \Gamma_2) \leq \min\{|Q_1|, |Q_2|\}/2;$$

the upper bound is reached when all closed orbits have length 4 and every contact in one of the RNA contact structures is involved in one of these closed orbits. Jointly with R. Alberich and J.Miró, we are currently working on the relation between d_{inv}, d_{sgr} and other algebraically defined abstract metrics on $\mathcal{C}_n$, as for instance the metric d_4 introduced in [8], and in particular on the problem of estimating the probability that, given two RNA contact structures Γ_1, Γ_2 of the same length, $\Omega(\Gamma_1, \Gamma_2) = 0$. Both numerical simulations with randomly generated RNA secondary structures of large length and heuristic arguments show that this probability is high: for instance, among all 35 RNA contact structures of length 6, only 2 form a closed orbit with $\Gamma = \{[6], \{1{\cdot}3, 4{\cdot}6\}\})$. We hope to report on it soon.

Acknowledgments. We acknowledge with thanks R. Alberich, J. Miró and G. Valiente for several discussions on the topic of this paper. We also thank the anonymous referees for their comments and suggestions, which have led to a substantial improvement of this paper.

References

[1] E. Westhof, V. Fritsch, RNA folding: Beyond Watson-Crick pairs, *Structure with Folding & Design* **8** (2000) R55–R65.

[2] Chan, H. S., Dill, K. A., 1989. Compact polymers. *Macromolecules* 22, 4559–4573.

[3] Waterman, M. S., Smith, T. F., 1978. RNA secondary structure: a complete mathematical analysis. *Math. Biosci.* 42, 257–266.

[4] Zuker, M., Sankoff, D., 1984. RNA secondary structures and their prediction. *Bull. Math. Biol.* 46, 591–621.

[5] Westhof, E., Jaeger, L., 1992. RNA pseudoknots. *Current Opinion Struct. Biol.* 2, 327–333.

[6] Zuker, M., 1989. The use of dynamic programming algorithms in RNA secondary structure prediction. In *Mathematical methods for DNA sequences* (M. Waterman, ed.), CRC Press, pp. 159–184.

[7] C. Reidys, P. F. Stadler, Bio-molecular shapes and algebraic structures, *Computers & Chemistry* **20** (1996), 85–94.

[8] M. Llabrés, F. Rosselló, A new family of metrics for biopolymer contact structures, to appear in *Comput. Biol. and Chem.* (doi:10.1016/j.compbiolchem.2003.10.002).

[9] P. Schuster, P. F. Stadler, Discrete models of biopolymers, to appear in *Handbook of Computational Chemistry* (M.J.C. Crabbe, M. Drew and A. Konopka, eds.), Marcel Dekker (in press); see also Univ. Wien TBI Preprint No. pks-99-012 (1999).

[10] C. Haslinger, P. F. Stadler, RNA structures with pseudo-knots: Graph-theoretical, combinatorial, and statistical properties, *Bulletin of Mathematical Biology* **61** (1999), 437–467.

[11] A. Kister, Y. Magarshak, J. Malinsky, The theoretical analysis of the process of RNA molecule self-assembly, *BioSystems* **30** (1993), 31–48.

[12] Y. Magarshak, C. J. Benham, An algebraic representation of RNA contact structures, *J. of Biomolecular Structures & Dynamics* **10** (1992) 465–488.

[13] J. Casasnovas, J. Miró, F. Rosselló, On the algebraic representation of RNA contact structures with G.U pairs, *Journal Math. Biol.* 47 (2003), 1–22.