BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis
Volume 6(4)

A new Motzkin class for joint RNA secondary structures

Athanasios Alexiou*, Panayiotis Vlamos

Department of Informatics, Ionian University, Plateia Tsirigoti 7, 49100 Corfu, Greece; Athanasios T Alexiou – Email: alexiou@ionio.gr; Phone: +30 210 9533296; Fax: +30 210 9533296; *Corresponding author

Received March 16, 2011; Accepted May 07, 2011; Published May 07, 2011

Abstract:
In general RNA prediction problem includes genetic mapping, physical mapping and structure prediction. The ultimate goal of structure prediction is to obtain the three dimensional structure of bimolecules through computation. The key concept for solving the above mentioned problem is the appropriate representation of the biological structures. Even though, the problems that concern representations of certain biological structures like secondary structures either are characterized as NP-complete or with high complexity, few approximation algorithms and techniques had been constructed, mainly with polynomial complexity, concerning the prediction of RNA secondary structures. In this paper, a new class of Motzkin paths is introduced, the so-called semi-elevated inverse Motzkin peakless paths for the representation of two interacting RNA molecules. The basic combinatorial interpretations on single RNA secondary structures are extended via these new Motzkin paths on two RNA molecules and can be applied to the prediction methods of joint structures formed by interacting RNAs.

Keywords:
RNA-RNA interaction problem, RNA secondary Structures, semi-elevated Motzkin paths

Background:
The secondary structure of an RNA molecule is the collection of base pairs that occur in its 3D structure. When the 5'-end of one nucleotide fits to the 3'-end of another, a p-bond is formed, while the sequence of p-bonds defines the backbone of the molecules. On the other hand certain base pairs like C-G, A-U and G-U form h-bonds, which cause folding of the molecular backbone into a configuration of minimal energy **[1]**. In some cases unusual non-canonical base pairs, like G/U, G/A and C/A, replace the canonical Watson-Crick base pairs, which maintained a stable helical structure. While these non-canonical pairings allow possible hydrogen-bonding interactions and can be treated as neutral evidence for a helical structure, there seems to be evidence against pairing **[2]**. Additionally, an RNA molecule can be viewed as an ordered sequence of n bases, while secondary structures can be generally defined as a set of pairs $i\,.\,j$, $1 \leq i \leq j \leq n$, , indexed starting at 1 from the so-called 5'-end and with each index in, at most, one pair. The above folds, cooperatively, form pseudoknot free secondary structures, where no base pairs overlap, that is there are no pair of bases $i\,.\,j$ and $i'.\,j'$ with $i < i' < j < j'$. In literature **[3]** except for hairpin and interior loops, definitions for bans, multiloops, external loops, pseudo knot loops, interior-pseudo knotted loops and multi-pseudo knotted loops, can also be found. A secondary structure of size n is closed **[1]** if there is an h-bond connecting bases 1 and n and for given integers n > 2, l > 0, there are $S^{(l)}(n - 2)$ secondary structures of size n and rank l. A bijection between the set of all closed secondary structures $Z^{(l)}$ (n) and the set of all plane trees with exactly n leaves $T^{(1)}(n)$ has also been proved. A more extended definition of closed secondary structures has been given **[3]**, through the closed regions of a secondary structure. Representing a secondary structure as an arc diagram, in which base indices are shown as vertices on a straight line, ordered from the 5`-end and arcs (always above the straight line) indicate base pairs, a region [i; j] will be referred to as: weakly closed if it contains at leat one base pair and for all base pairs $i'.\,j'$ of R, $i' \in [i; j]$ if and only if $j' \in [i; j]$ and closed if either $i = 1, j = n$ or if it is weakly closed and for all l with $i < 1 < j$ the regions [i; 1] and [1; j] are not weakly closed. Finally, it should be mentioned that secondary structures are in a simple bijection with Motzkin paths without peaks **[4]**.

Furthermore there are cases, such as antisense RNAs which exhibit their functions by establishing stable joint structures with target mRNA, introducing the well known RNA-RNA Interaction Prediction Problem (RIP). Although it is proved **[5]** that RIP is an NP-complete problem in its general form, there are a number of algorithms in the literature for RNA interactions. We can mention the latest studies concerning methods with polynomial complexity: minimum free energy models for joint RNA structures **[6, 7]**, multiple sequence alignments **[8]**, grammar based approach **[9]**, base pair counting, stacked pair energy models and loop energy models **[5]**. Even though in some cases the running time and the space complexity is of better performance **[7]** the proposed methods for the RIP, concern mainly joint secondary structures that do not contain pseudoknots, crossing interactions or zigzags. In this paper we introduce a new combinatorial technique for the representation of two interacting RNA molecules. The extension of the well known Motzkin paths to a more complex form, will probably overcome the limitations of RNAs structure comparison, giving the opportunity to apply new operations between joint sequences in order to identify, analyze and compare them.

Methodology:
A Motzkin path on n steps is a lattice path in the Cartesian plane from (0,0) to (n,0) using only steps of the type (1,1)-up, (1,-1)-down and (1,0)-level, that never run below the x-axis. Of particular interest is the set of elevated paths that never touch the x-axis except initially and perhaps finally. If the above steps are written as: u-up, d-down and u-level, an elevated path lies between an initial extra u and a final extra d.

A new class of Motzkin Paths:

In the case in question, we assume that the 5'-end of an RNA molecule corresponds to the initial element of an elevated Motzkin path, while the 3'-end treats like a semi-elevated step, able to transform the final *d* to *h*, in order to create a new structure **(Figure 1a)**. Therefore, we introduce the term semi-elevated generalized Motzkin peakless path for RNA secondary structures. Additionally, the use of these paths, may offer a solution to the opening of RNA secondary structures in specific bonds. A method for the opening of RNA secondary structures was firstly introduced by using movable specific bond indices **[10]**, which failed in certain structures. Furthermore, an inverse generalized Motzkin peakless path on *n* steps is a lattice path in the Cartesian plane from (n,1) to (0,0), using only steps of the type (-1,1)-inverse up ($_iu$), (-1,-1)-inverse down ($_id$) and (1,0)-inverse level($_ih$), that never run below the *x*-axis. The map $z : Mn \rightarrow Rn$ can now be constructed. Given two independent generalized Motzkin paths K and L of length n and m respectively, it is possible for their paths *z*(*K* o *L*) to be combined using the following procedure: (i)The last step of the corresponding Motzkin path P is followed by a semi-elevated tail; (ii) In order to obtain the joint secondary structure on the 3'-end, the elevated-tail step changes to a semi-elevated and becomes a horizontal step; (iii) For each up step of the form *hd…d* the corresponding $_iu…_iu_ih$ will be designed; (iv) For each down step of the form *u…uh* the corresponding $_ih_id…_id$ will be designed; (v) Every step *h…h* will be designed in the same length towards the inverse direction $_ih…_ih$

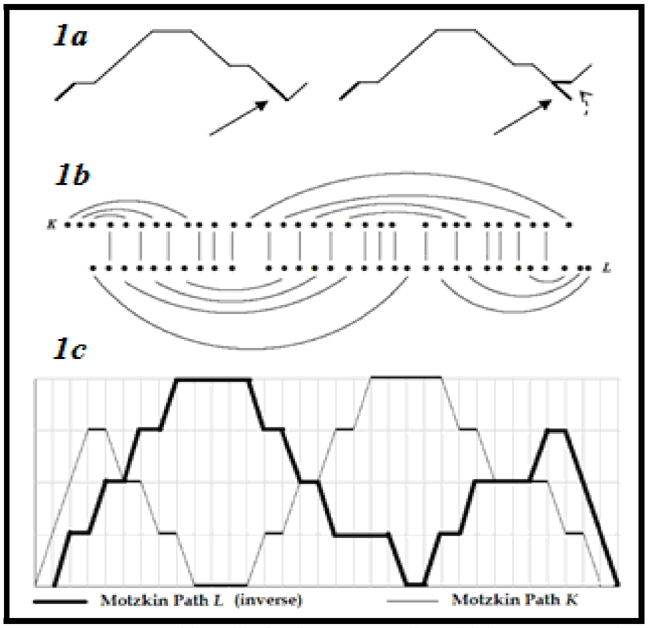

Figure 1: New Motzkin paths for the representation of joint RNA secondary structure. **(1a):** The semi elevated Motzkin peakless path, transforming the final d to h, in order to create a new structure. **(1b):** An example of joint RNA structure. **(1c):** The corresponding Motzkin paths for the joint structure, using the definition of semi elevated Motzkin peakless path

Representation of joint RNA secondary structures:

We apply the above procedure on the joint RNA structure of two secondary sequences K and L, according to the method of the longest common subsequence of multiple binary strings (mLCS) **[5]**. Let's assume the following Motzkin paths **(Figure 1b)**:

K: uuuhdhdhdhhhuhuhuhuhhhhdhdhhhdhd (5' to 3' direction)
L: uhuhuhuhhhhdhdhdhhhdhuhuhhhuhddd (3' to 5' direction)

The steps of the ***L*** Motzkin path are transformed with the following procedure: each *d*-step is converted to $_iu$, each *u*-step is converted to $_id$ and each *h*-step is converted to $_ih$. Hence, the Motzkin path of ***L***-RNA secondary structure becomes an inverse Motzkin path and the composition of these two paths defines the Motzkin path of the RNA's interaction **(Figure 1c)**.

Discussion:

In the general case of RNA-RNA Interaction Prediction Problem (RIP), it is assumed that there exist two independent RNA sequences *K* and *L* of length *n* and *m* respectively. In the joint secondary structure of K and L each nucleotide is paired with at most one nucleotide in the same or the other strand, while these two strands interact in opposite directions. If we assume that the K strand is indexed from 1 to n in 5' to 3' direction and L is indexed from 1 to m in 3' to 5' direction, then we refer to the i_{th} nucleotide in K and L by i_K and i_S respectively and to any base pair between nucleotides *i* and *j* with the notion *i* . *j*. We present an efficient algorithm for the generation of invertible walking and the formulation of joint RNA secondary structures.

Algorithm - Motzkin Paths for Joint RNA Secondary Structures:

Input: Two independent RNA sequences K_n and L_m
Output: The combined Motzkin paths of the joint RNA secondary structures
Design the generalized Motzkin peakless path of *K*, from 1 to n in 5' to 3' direction; Transform the last *d*-step (elevated) with an *h*-step; Design the generalized semi-elevated Motzkin peakless path of *L*, from 1 to m in 3' to 5' direction
End

It is necessary to recall and adjust some basic definitions of the 'lattice walk'. The assumption that the path is peakless means that there are no $_iu_id$ or $_id_iu$; A $h_id…_id$ is called maximal if it is not followed by another down step; a horizontal step is said to be single if it is not adjacent to a down step; a down step is called tail if it is the last down step of a maximal ($h_id…_id$); For an inverse up step *a* (at level *k*) its corresponding inverse down step is *b* (the leftmost down step at level *k-1*, right to *a*) and for an inverse down step its corresponding inverse up step will be the rightmost up step at level *k+1* to the left. Additionally the creation of a semi-elevated inverse Motzkin peakless path is totally reversible, using reflection techniques and with 1-1 correspondence to a Motzkin path and the number of Dyck paths of these semi-elevated inverse Motzkin peakless paths is equal to that of the corresponding Motzkin paths. It can be easily observed that several interesting relationships are valid in the case of joint Motzkin paths such as the bijection between unit squares and the auxiliary paths **[11]**. Nevertheless, the application of any merging algorithm to these joint Motzkin paths or to their corresponding labelled trees **[4]** is currently uncertain while the process of the RNA-RNA is not completely understood. While our improvements concern an alternative representation of the RIP, we can implement new operations on RNA interactions in future studies, similar to the node fusion and edge fusion on trees representations **[12]**. These new operations will improve our understanding on multilevel RNA structure comparison and will extend the limitations of the classical mathematical representations on individual RNA structures.

Conclusion:

In this study, a new class of generalized Motzkin paths is introduced, the semi-elevated inverse generalized Motzkin peakless paths. As an application of this Cartesian walk the RNA Interaction Problem was studied, concerning an optimal representation of joint RNA secondary structures. The proposed technique makes it possible to extend close secondary structures without peaks towards the 3' direction ignoring any steps of the form *ud*. While in most of the cases the algorithms for the prediction of RNAs interaction have a polynomial complexity, our model specifies a new field for the design of prediction models and computer tools. Future study and implementation of new operations among interacting RNA structures, will clarify the principles of RNA Interaction Problem.

References:

[1] Doslic T & Veljan D. *Mathematical Communications* 2007 **12**: 163
[2] Chen JL & Blasco M. *Cell* 2000 **100**: 503 [PMID: 10721988]
[3] Rastegari B & Condon A. *Springer WABI.* 2005 **3692**: 341
[4] Deutsch E & Shapiro L. *Discrete Mathematics* 2002 **256**: 655
[5] Alkan C *et al. Journal of Computational Biology* 2006 **13**: 267 [PMID: 16597239]
[6] Chitsaz H *et al. Bioinformatics* 2009 **25**: 365
[7] Salari R. *et al. Algorithms for Molecular Biology* 2010 **5**: 5 [PMID: 20047661]
[8] Andrew X Li *et al. Bioinformatics* 2011 **27**: 546
[9] Kato Y *et al. Pattern Recognition* 2009 **42**: 531
[10] Willenbring R. *Discrete Applied Mathematics* 2009 **157**: 1607
[11] Pergola E. *et al. Advances in Applied Mathematics* 2002 **28**: 580
[12] Allali J & Sagot M. *IEEE/ACM transactions on computational biology and bioinformatics* 2005 **2**: 3 [PMID: 17044160]

Edited by P Kangueane
Citation: Alexiou & Vlamos. Bioinformation 6(4): 162-163 (2011)