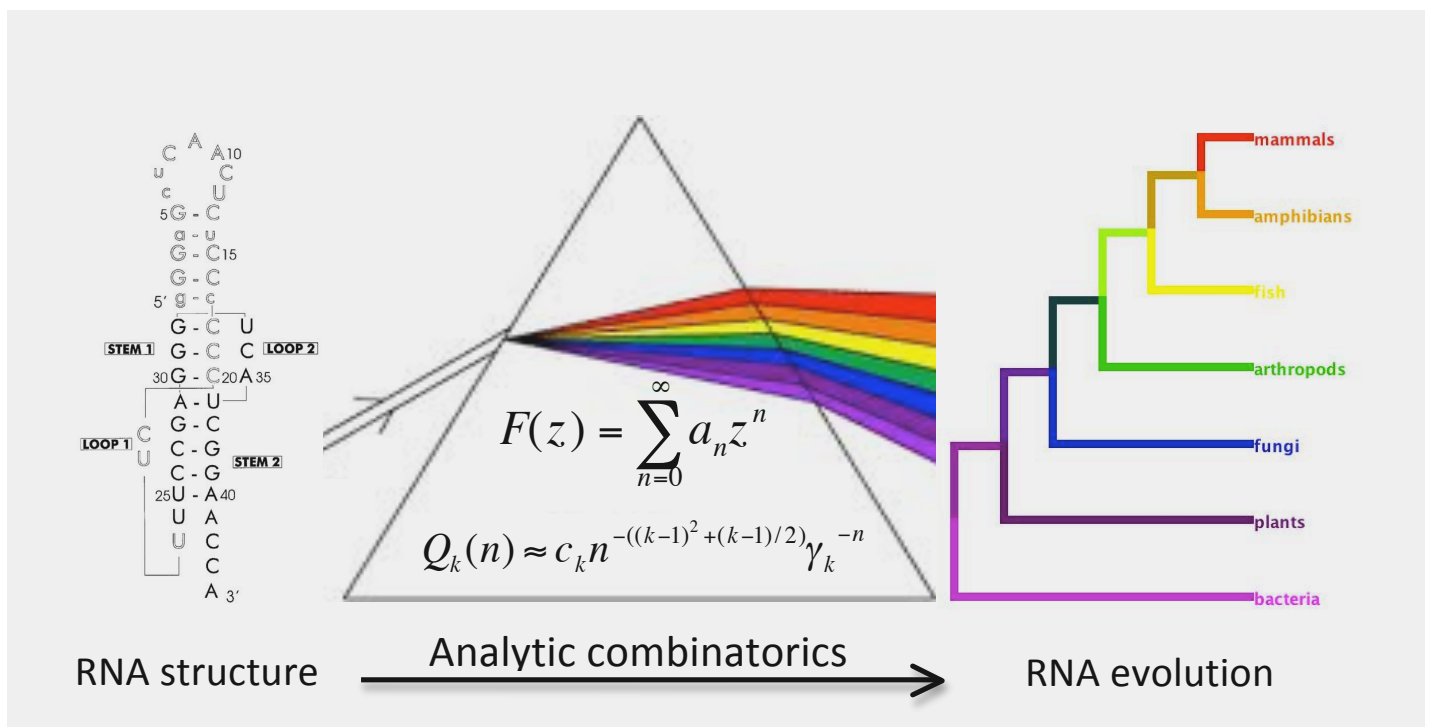


Χρήση γεννητριών συναρτήσεων και ασυμπτωτικής αρίθμησης για τη μελέτη της εξέλιξης δομών RNA



Διπλωματική εργασία του
Νικόλαου Κωνσταντινίδη

Υπό την επίβλεψη του Καθηγητή Μιχάλη Κολουτζάκη

Περιεχόμενα

Περιεχόμενα	3
Κεφάλαιο 0	
Εισαγωγή	4
0.1 Γεννήτριες συναρτήσεις	4
0.2 Ψευδοκόμποι RNA	5
0.3 Χάρτες αλληλουχίας – δομής	6
Κεφάλαιο 1	
Μέθοδοι συμβολικής απαρίθμησης και ασυμπτωτικών εκτιμήσεων	7
1.1 Βασικοί ορισμοί	7
1.2 Μέθοδοι συμβολικής απαρίθμησης	8
1.3 Μέθοδοι ασυμπτωτικών εκτιμήσεων	10
Κεφάλαιο 2	
Γεννήτριες Συναρτήσεις	13
2.1 Ορισμοί	13
2.2 Γεννήτριες συναρτήσεις απλούστερων δομών	15
2.3 Αρθρωτές k-noncrossing δομές	24
Κεφάλαιο 3	
Ασυμπτωτική απαρίθμηση	33
3.1 D-finiteness και P-recursiveness	33
3.2 Η εξίσωση $F_k(z)$	34
3.3 Ασυμπτωτική αρίθμηση άλλων δομών	38
Κεφάλαιο 4	
Στοιχεία σχετικά με την εξέλιξη των RNA ψευδοκόμπων	43
4.1 Ευρωστία και εξελιξιμότητα RNA δομών	43
4.2 Τα ουδέτερα δίκτυα των RNA ψευδοκόμπων	45
4.3 Συζήτηση	47
Βιβλιογραφία	48

Κεφάλαιο 0

Εισαγωγή

Σκοπός της πτυχιακής αυτής εργασίας είναι η μελέτη των χαρτών αλληλουχίας-δομής των RNA ψευδοκόμπων, με στόχο την εξαγωγή συμπερασμάτων σχετικά με το εξελικτικό δυναμικό των δομών αυτών. Για να επιτευχθεί αυτό, αφού παρουσιαστούν οι μέθοδοι συμβολικής απαρίθμησης για την εξαγωγή γεννητριών συναρτήσεων (Κεφάλαιο 1), υπολογίζουμε τις γεννήτριες συναρτήσεις διαφόρων τύπων ψευδοκόμπων RNA (Κεφάλαιο 2), κάνουμε τις κατάλληλες ασυμπτωτικές εκτιμήσεις για τις συναρτήσεις αυτές (Κεφάλαιο 3) και, τέλος, συζητάμε τις εκτιμήσεις αυτές στα πλαίσια των χαρτών αλληλουχίας-δομής.

Πριν προχωρήσουμε στο κυρίως μέρος της πτυχιακής, ακολουθούν λίγα λόγια για τις γεννήτριες συναρτήσεις, τους ψευδοκόμπους RNA και τους χάρτες αλληλουχίας-δομής.

0.1 Γεννήτριες συναρτήσεις

Έστω ότι έχουμε ένα πρόβλημα στο οποίο η απάντηση είναι μια αλληλουχία αριθμών $\alpha_0, \alpha_1, \alpha_2, \dots$. Αν δεν μπορούμε να βρούμε μια απλή συνάρτηση, η οποία να υπολογίζει το $\alpha_n = f(n)$, τότε μία πιθανή λύση μπορεί να είναι η εύρεση μίας γεννήτριας συνάρτησης. Η γεννήτρια συνάρτηση είναι μία συνάρτηση της μορφής

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

οι συντελεστές της οποίας, $\alpha_0, \alpha_1, \alpha_2, \dots$ είναι η αλληλουχία των αριθμών που ψάχνουμε. Παραδείγματος χάριν, η γεννήτρια συνάρτηση που δίνει την ακολουθία Fibonacci είναι η παρακάτω:

$$f(z) = \frac{z}{1 - z - z^2}$$

Δηλαδή, ο n -οστός αριθμός της ακολουθίας Fibonacci F_n είναι ο συντελεστής του x_n στην επέκταση (expansion) της παραπάνω συνάρτησης ως δυναμοσειρά. Η εύρεση, βέβαια, μιας τέτοιας συνάρτησης δεν είναι εξίσου ικανοποιητική με μια συνάρτηση της μορφής $F_n = f(n)$, και κυρίως είναι δύσκολο να υπολογίσουμε το F_n όταν το n είναι πολύ μεγάλο. Σε πολλές περιπτώσεις, όμως, που δεν έχουμε την πολυτέλεια εύρεσης ακριβούς συνάρτησης, η γεννήτρια συνάρτηση μπορεί να είναι η μοναδική δυνατή προσέγγιση. Επιπλέον, υπάρχουν πολλοί τρόποι με τους οποίους μπορεί κανείς να υπολογίσει πώς συμπεριφέρεται το α_n όταν το n είναι πολύ μεγάλο, κάποιοι από τους οποίους περιγράφονται στο Κεφάλαιο 3.

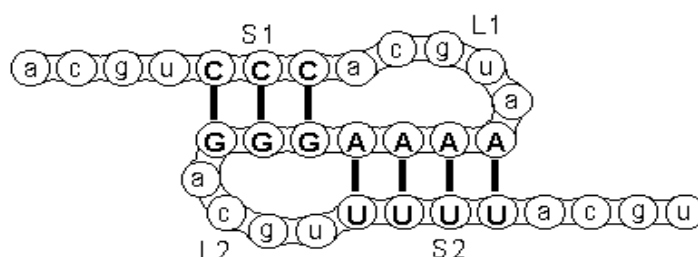
Επίσης, σε κάποιες περιπτώσεις, γνωρίζοντας τη γεννήτρια συνάρτηση, μπορεί κάποιος να βρει μια ακριβή ή αναδρομική συνάρτηση για την ακολουθία, να βρει μέσους όρους και στατιστικά στοιχεία της ακολουθίας, να αποδείξει μονοτροπικότητα (unimodality) ή κυρτότητα (convexity) της ακολουθίας κοκ. Οι γεννήτριες συναρτήσεις μπορούν, επίσης να χρησιμοποιηθούν για να αποδείξουν ισότητες.

0.2 Ψευδοκόμποι RNA

Το RNA είναι μια μονοκλωνη ακολουθία των βάσεων ουρακίλη, γουανίνη, αδερίνη και κυτοσίνη. Οι διαφορετικές τρισδιάστατες δομές που μπορεί να λάβει ένα μόριο RNA αντικατοπτρίζουν την ποικιλία λειτουργιών αυτού. Οι λειτουργίες ποικίλλουν από την μεταφορά της γενετικής πληροφορίας απο το DNA στις πρωτεΐνες έως την ενζυμική κατάλυση αντιδράσεων. Ως εκ τούτου, η πρόβλεψη των δομών που μπορεί να πάρει ένα μόριο RNA είναι εξαιρετικά δύσκολη.

Αν και έχουν βρεθεί αλγόριθμοι για την πρόβλεψη διαφόρων RNA δομών, οι υπάρχοντες αλγόριθμοι δεν μπορούν να αντιμετωπίσουν τη δομή των ψευδοκόμπων RNA (Lyngso and Pedersen 2000). Ένας ψευδοκόμπος είναι μια δομή RNA που αποτελείται τουλάχιστον από δύο ελικοειδή δίκλιωνα τμήματα τα οποία συνδέονται με μονόκλωνες περιοχές ή θηλιές. Με τη σειρά τους οι θηλιές, μπορούν να αλληλεπιδρούν με κοντινές μονόκλωνες περιοχές σχηματίζοντας δεσμούς υδρογόνου και συμμετέχοντας έτσι στην τελική τριτοταγή δομή του μορίου (Εικόνα 0.1). Έτσι, αυτή η σχετικά απλή δομή μπορεί να λάβει πολύπλοκες και σταθερές μορφές. Λόγω της ποικιλίας του μήκους των θηλιών και των μίσχων, οι ψευδοκόμποι ποικίλλουν δομικά, αποκτώντας διαφορετικούς βιολογικούς ρόλους. Τέτοιοι ρόλοι είναι οι εξής: σχηματισμός του καταλυτικού πυρήνα ριβοενζύμων (Rastogi, Beattie et al. 1996; Ke, Zhou et al. 2004), αυτοματιζόμενων ιντρονίων (Adams, Stahley et al. 2004) και της τελομεράσης (Theimer, Blois et al. 2005). Επίσης, οι ψευδοκόμποι συμμετέχουν στη γονιδιακή ρύθμιση επάγοντας τη ριβοσωμική αλλαγή πλαισίου ανάγνωσης (frameshifting) σε πολλούς ιούς (Shen and Tinoco 1995; Michiels, Versleijen et al. 2001; Egli, Minasov et al. 2002; Nixon, Rangan et al. 2002).

Όπως γίνεται εύκολα αντιληπτό, οι ψευδοκόμποι αποτελούν μια ιδιαίτερα σημαντική ομάδα δομών RNA, η πρόβλεψη της δομής των οποίων αποτελεί υποσχόμενο πεδίο έρευνας.



Εικόνα 0.1: Δομή RNA ψευδοκόμπου. Οι δίκλιωνες περιοχές συμβολίζονται με S1 και S2, ενώ οι θηλιές με L1 και L2 (βλέπε και Κεφάλαιο 2).

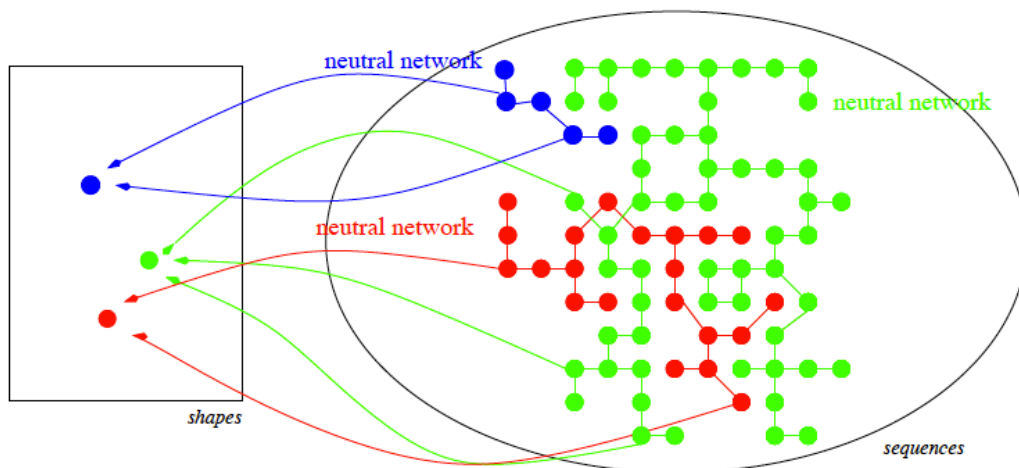
(Πηγή: <http://www.ekevanbatenburg.nl/PKBASE/PKBABOUT.HTML>)

0.3 Χάρτες αλληλουχίας - δομής

Οι χάρτες που συνδέουν τις αλληλουχίες με τις δομές είναι ιδιαίτερα σημαντικοί για τη μελέτη της εξέλιξης του RNA. Οι χάρτες αυτοί αποτελούνται από δύο σύνολα, (α) το σύνολο των αλληλουχιών και (β) το σύνολο των δομών, καθώς και από γραμμές που ενώνουν σημεία του ενός με σημεία του άλλου συνόλου (Εικόνα 0.2). Είναι προφανές ότι πολλά μέλη του συνόλου των αλληλουχιών συνδέονται με ένα μέλος του συνόλου των δομών.

Τα ουδέτερα σύνολα αποτελούνται από όλες τις αλληλουχίες που έχουν την ίδια δομή. Ένα ουδέτερο σύνολο ονομάζεται ουδέτερο δίκτυο αν όλες οι αλληλουχίες που περιλαμβάνονται σε αυτό μπορούν να συνδεθούν μεταξύ τους με αλληπαλλήλες μοναδικές μεταλλαγές που δεν οδηγούν σε δομές εκτός του συνόλου. Όσον αφορά τους RNA φαινοτύπους, τα ουδέτερα σύνολα είναι σχεδόν πάντα καλά συνδεδεμένα, οπότε χρησιμοποιούμε αυτούς τους όρους εναλλακτικά (Jorg, Martin et al. 2008) (Εικόνα 0.2).

Το μέγεθος και η δομή αυτών των συνόλων έχει σημαντική επίπτωση στην ευρωστία ενός βιολογικού συστήματος σε μεταλλαγές καθώς και στην εξελικτικότητα του, την ικανότητα δηλαδή να παράγει φαινοτυπική ποικιλία. Σε μελέτες μοριακών φαινοτύπων έχει φανεί ότι όσο μεγαλύτερο είναι αυτό το σύνολο, τόσο μεγαλύτερη η ανοχή και η εξελικτικότητα των φαινοτύπων (Wagner 2005). Όσο μεγάλο και να είναι ένα ουδέτερο σύνολο, είναι σημαντικό να τονιστεί ότι περιλαμβάνει μόνο ένα μικρό υποσύνολο του γενοτυπικού χώρου (του συνόλου δηλαδή των διαφορετικών αλληλουχιών).



Εικόνα 0.2: Χάρτης αλληλουχίας - δομής. Με διαφορετικά χρώματα είναι σημασμένα τα διαφορετικά ουδέτερα δίκτυα. (Πηγή: <http://www.tbi.univie.ac.at/~ulim/probA/ulli.pdf>)

Κεφάλαιο 1

Μέθοδοι Συμβολικής Απαρίθμησης και Ασυμπτωτικών Εκτιμήσεων

Βασικός στόχος του πρώτου κεφαλαίου είναι η δημιουργία της κατάλληλης μαθηματικής βάσης για την εξαγωγή των γεννητριών συναρτήσεων και των ασυμπτωτικών τους εκτιμήσεων που θα γίνει στα Κεφάλαια 2 και 3. Πολλά από τα στοιχεία προέρχονται από το βιβλίο *Analytic combinatorics*. (Flajolet and Sedgewick 2009)

1.1 Βασικοί ορισμοί

Οι γεννήτριες συναρτήσεις “μετράνε” διακριτά αντικείμενα και ο στόχος είναι η απαρίθμηση αυτών των αντικειμένων βάσει κάποιων χαρακτηριστικών. Πριν προχωρήσουμε στις μεθόδους απαρίθμησης, καλό είναι να αναφέρουμε κάποιους βασικούς ορισμούς:

Ορισμός 1.1: Συνδυαστική τάξη είναι ένα πεπερασμένο ή αριθμήσιμο σύνολο στο οποίο ορίζεται μια συνάρτηση μεγέθους, που ικανοποιεί τις παρακάτω συνθήκες:

- το μέγεθος ενός στοιχείου είναι ένας μη αρνητικός ακέραιος
- ο αριθμός των στοιχείων δεδομένου μεγέθους είναι πεπερασμένος.

Ορισμός 1.2: Η μετρούσα ακολουθία (counting sequence) μιας συνδυαστικής τάξης είναι μια ακολουθία ακεραίων $(A_n)_{n \geq 0}$, όπου A_n είναι ο πληθικός αριθμός, δηλαδή ο αριθμός των στοιχείων στην τάξη A που έχουν μέγεθος n .

Ορισμός 1.3: Δύο συνδυαστικές τάξεις A και B ονομάζονται συνδυαστικά ισόμορφες, $A \cong B$, αν και μόνο αν οι μετρούσες ακολουθίες τους είναι ίδιες. Αυτή η συνθήκη είναι ίδια με την ύπαρξη ενός μονομορφισμού από την A στη B που διατηρεί το μέγεθος.

Ορισμός 1.4: Η συνήθης γεννήτρια συνάρτηση μιας ακολουθίας (A_n) είναι η τυπική δυναμοσειρά:

$$A(z) = \sum_{n=0}^{\infty} A_n z^n.$$

Η συνήθης γεννήτρια συνάρτηση μιας συνδυαστικής τάξης A είναι η γεννήτρια συνάρτηση των πληθικών αριθμών A_n . Αντίστοιχα, η συνήθης γεννήτρια συνάρτηση έχει τη συνδυαστική μορφή:

$$A(z) = \sum_{a \in A} z^{|a|}.$$

1.2 Μέθοδοι συμβολικής απαρίθμησης

Υπάρχουν δύο βασικοί τρόποι που χρησιμοποιούνται για την εξαγωγή γεννητριών συναρτήσεων. Ένας είναι η ανάλυση των δομών που θέλουμε να αριθμήσουμε σε μικρότερες δομές ίδιου ή απλούστερου τύπου και, κατόπιν, η εξαγωγή αναδρομικών σχέσεων που ικανοποιούνται από το $\{A_n\}$. Εναλλακτικά, οι τάξεις των συνδυαστικών δομών δημιουργούνται κυρίως από απλούστερες τάξεις χρησιμοποιώντας μια συλλογή βασικών συνδυαστικών κατασκευών.

Ορισμός 1.5: Έστω Φ μια κατασκευή που σχετίζει μια οποιαδήποτε συλλογή τάξεων $B^{(1)}, B^{(2)}, \dots, B^{(m)}$ με μια νέα τάξη

$$A = \Phi(B^{(1)}, B^{(2)}, \dots, B^{(m)}).$$

Η κατασκευή Φ λέγεται αποδεκτή αν και μόνο αν η μετρούσα ακολουθία (A_n) εξαρτάται μόνο από τις αντίστοιχες ακολουθίες $(B_n^{(1)}), (B_n^{(2)}), \dots, (B_n^{(m)})$. Για τις αποδεκτές κατασκευές, υπάρχει ένας τελεστής Ψ που δρα πάνω στις αντίστοιχες συνήθεις γεννήτριες συναρτήσεις:

$$A(z) = \Psi[B^{(1)}(z), \dots, B^{(m)}(z)].$$

Καρτεσιανό γινόμενο: Το καρτεσιανό γινόμενο δύο τάξεων B και C δημιουργεί διατεταγμένα ζεύγη,

$$A = B \times C \text{ αν } A = \{\alpha = (b, c) \mid b \in B, c \in C\},$$

με το μέγεθος του ζεύγους να ορίζεται ως εξής:

$$|a|_A = |b|_B + |c|_C.$$

Μπορούμε εύκολα να δούμε ότι οι μετρούσες ακολουθίες που αντιστοιχούν στα A, B, C συνδέονται με την εξίσωση:

$$A_n = \sum_{k=0}^n B_k C_{n-k},$$

που σημαίνει ότι η κατασκευή είναι αποδεκτή και μεταφράζεται σε γινόμενο των συνήθων γεννητριών συναρτήσεων

$$A(z) = B(z) \cdot C(z).$$

Ένωση συνόλων: Έστω $A = B \cup C$ με $B \cap C = \emptyset$, με το μέγεθος ορισμένο ως εξής: $|\omega|_A = |\omega|_B$, αν $\omega \in B$, και $|\omega|_A = |\omega|_C$, αν $\omega \in C$. Ισχύει, λοιπόν,

$$A_n = B_n + C_n,$$

που σε επίπεδο γεννητριών συναρτήσεων σημαίνει:

$$A(z) = B(z) + C(z)$$

που σημαίνει ότι η κατασκευή είναι αποδεκτή και μεταφράζεται σε αθροισμα των συνήθων γεννητριών συναρτήσεων.

Ας δουμε, τώρα, τις βασικές κατασκευές που αποτελούν τον πυρήνα της γλώσσας για τις συνδυαστικές κατασκευές. Στα επόμενα κεφάλαια, δε θα χρησιμοποιηθούν όλες αυτές οι κατασκευές, αλλά θα περιγραφούν περιληπτικά εδώ στα πλαίσια μιας πλήρους παρουσίασης.

Αρχικά, ας υποθέσουμε ότι δίνεται μια τάξη E που καλείται ουδέτερη τάξη και αποτελείται από ένα μοναδικό στοιχείο $\{\varepsilon\}$ μεγέθους 0. Κάθε τέτοιο στοιχείο ονομάζεται ουδέτερο στοιχείο. Υποθέτουμε, επίσης, ότι δίνεται μια ατομική τάξη Z που αποτελείται από ένα μοναδικό στοιχείο μεγέθους 1. Κάθε τέτοιο στοιχείο ονομάζεται άτομο. Είναι προφανές ότι οι γεννήτριες συναρτήσεις των ανωτέρω τάξεων είναι:

$$\begin{aligned} E(z) &= 1, \\ Z(z) &= z. \end{aligned}$$

Ακολουθία: Αν B είναι μια τάξη τότε η τάξη-ακολουθία $SEQ(B)$ ορίζεται ως το άπειρο άθροισμα

$$SEQ(B) = \{\varepsilon\} + B + (B \times B) + (B \times B \times B) + \dots$$

όπου ε είναι το ουδέτερο στοιχείο (μεγέθους 0). Από την ένωση συνόλων, μπορούμε εύκολα να δούμε ότι:

$$SEQ(B)(z) = 1 + B(z) + B(z)^2 + B(z)^3 + \dots = \frac{1}{1 - B(z)}.$$

Κύκλος: Οι ακολουθίες των οποίων τα στοιχεία μετακινούνται κυκλικά ορίζουν κύκλους, που συμβολίζονται ως $CYC(B)$. Έχουμε, ουσιαστικά,

$$CYC(B) := SEQ(B)/S,$$

όπου S είναι η σχέση ισοδυναμίας μεταξύ ακολουθιών των οποίων τα στοιχεία διαφέρουν μόνο κατά μία κυκλική μετακίνηση. Η απόδειξη της σχέσης που συνδέει τις γεννήτριες συναρτήσεις είναι εκτός στόχων αυτής της εργασίας:

$$CYC(B)(z) = \sum_{k=1}^{\infty} \frac{\phi(k)}{k} \log \frac{1}{1 - B(z^k)}.$$

Multiset: Τα multisets είναι πεπερασμένα σύνολα (η σειρά των στοιχείων του συνόλου δε μετράει), όπου επιτρέπονται αυθαίρετες επαναλήψεις στοιχείων. Ισχύει

$$MSET(B) := SEQ(B)/R,$$

όπου R είναι η σχέση ισοδυναμίας μεταξύ ακολουθιών των οποίων τα στοιχεία διαφέρουν μόνο κατά κάποια μετάθεση στοιχείων. Αποδεικνύεται εύκολα ότι:

$$MSET(B)(z) = \prod_{n \geq 1} (1 - z^n)^{-B_n}.$$

Powerset: Η τάξη αυτή ορίζεται ως η τάξη $PSET(B) \subset MSET(B)$ που αποτελείται από τα multisets που δεν έχουν επαναλήψεις. Αποδεικνύεται, επίσης, εύκολα ότι:

$$PSET(B)(z) = \prod_{n \geq 1} (1 + z^n)^{B_n}.$$

1.3 Μέθοδοι ασυμπτωτικών εκτιμήσεων

Έστω $f(z) = \sum_n a_n z^n$ μια γεννήτρια συνάρτηση. Δεδομένου ότι οι ακριβείς εξισώσεις για τους συντελεστές μπορεί να είναι πολύπλοκες ή και αδύνατο να εξαχθούν, μπορούμε να μελετήσουμε τη συνάρτηση με το να αποκτήσουμε πληροφορίες για τα a_n όταν το n είναι μεγάλο. Στη συνέχεια, μας ενδιαφέρει να έχουμε μια εκτίμηση για τους συντελεστές συναρτήσεως ενός εκθετικού και ενός πολυωνμικού παράγοντα, δηλαδή $a_n \approx P(n) \cdot \gamma^n$, όπου το γ είναι μία σταθερά και το $P(n)$ ένα πολυώνυμο.

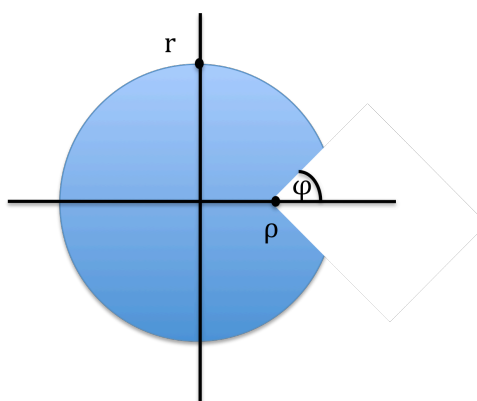
Το κύριο αποτέλεσμα που θα μας χρειαστεί στο Κεφάλαιο 3 είναι το Θεώρημα 1.9. Πριν από αυτό, όμως, θα δώσουμε κάποια άλλα αποτελέσματα χωρίς απόδειξη, που θα μας οδηγήσουν στο τελικό Θεώρημα.

Η εύρεση των γ και $P(n)$ βασίζεται κυρίως σε singular expansions και θεωρήματα μεταφοράς (transfer theorems). Τα θεωρήματα μεταφοράς πραγματοποιούν τη μετάφραση των error terms από συναρτήσεις σε συντελεστές.

Η συμπεριφορά της $f(z)$ κοντά στο επικρατές ανώμαλο σημείο ρ (dominant singularity) είναι ο καθοριστικός παράγοντας της ασυμπτωτικής συμπεριφοράς των συντελεστών της.

Ορισμός 1.6: Δεδομένων δύο αριθμών ϕ, r , όπου $r > |\rho|$ και $0 < \phi < \pi/2$, το ανοιχτό πεδίο $\Delta_\rho(\phi, r)$ (Εικόνα 1.1) ορίζεται ως:

$$\Delta_\rho(\phi, r) = \{z \mid |z| < r, z \neq \rho, |\text{Arg}(z - \rho)| > \phi\}.$$



Εικόνα 1.1: Το ανοιχτό πεδίο $\Delta_\rho(\phi, r)$ σημαίνεται με μπλε χρώμα.

Ένα πεδίο είναι Δ_ρ -πεδίο στο ρ αν είναι της μορφής $\Delta_\rho(\varphi, r)$ για κάποια r και φ . Μια συνάρτηση είναι Δ_ρ -αναλυτική αν είναι αναλυτική σε κάποιο Δ_ρ -πεδίο.

Έστω $[z^n]f(z)$ ο συντελεστής του z^n στο ανάπτυγμα της δυναμοσειράς του $f(z)$ στο 0. Δεδομένου ότι οι συντελεστές Taylor έχουν την ιδιότητα

$$\forall \gamma \in \mathbb{C} \setminus \{0\}; \quad [z^n]f(z) = \gamma^n [z^n]f\left(\frac{z}{\gamma}\right),$$

μπορούμε, χωρίς βλάβη της γενικότητας, να περιορίσουμε την ανάλυσή μας στην περίπτωση που το $z=1$ είναι το μοναδικό επικρατές ανώμαλο σημείο. Χρησιμοποιούμε το $U(a,r) = \{z \in \mathbb{C} \mid |z-a| < r\}$ για να ορίσουμε την ανοιχτή γειτονιά του a στο \mathbb{C} . Επίσης, χρησιμοποιούμε τους παρακάτω συμβολισμούς:

$$(f(z) = O(g(z)) \text{ όταν } z \rightarrow \rho) \Leftrightarrow (f(z)/g(z) \text{ είναι φραγμένο όταν } z \rightarrow \rho),$$

$$(f(z) = o(g(z)) \text{ όταν } z \rightarrow \rho) \Leftrightarrow (f(z)/g(z) \rightarrow 0 \text{ όταν } z \rightarrow \rho),$$

$$(f(z) = \Theta(g(z)) \text{ όταν } z \rightarrow \rho) \Leftrightarrow (f(z)/g(z) \rightarrow c \text{ όταν } z \rightarrow \rho),$$

$$(f(z) \approx g(z) \text{ όταν } z \rightarrow \rho) \Leftrightarrow (f(z)/g(z) \rightarrow 1 \text{ όταν } z \rightarrow \rho),$$

όπου c είναι μία σταθερά. Αν γράψουμε $f(z) = O(g(z))$ ή $f(z) = o(g(z))$ ή $f(z) = \Theta(g(z))$ ή $f(z) \approx g(z)$, εννοείται ότι το z τείνει σε ένα (μοναδικό) ανώμαλο σημείο. Αντίστοιχα, για την περίπτωση των ακολουθιών ο συμβολισμός έχει ως εξής:

$$(a_n = O(b_n) \text{ όταν } n \rightarrow \infty) \Leftrightarrow (a_n/b_n \text{ είναι φραγμένο όταν } n \rightarrow \infty),$$

$$(a_n = o(b_n) \text{ όταν } n \rightarrow \infty) \Leftrightarrow (a_n/b_n \rightarrow 0 \text{ όταν } n \rightarrow \infty),$$

$$(a_n = \Theta(b_n) \text{ όταν } n \rightarrow \infty) \Leftrightarrow (a_n/b_n \rightarrow c \text{ όταν } n \rightarrow \infty),$$

$$(a_n \approx b_n \text{ όταν } n \rightarrow \infty) \Leftrightarrow (a_n/b_n \rightarrow 1 \text{ όταν } n \rightarrow \infty).$$

Θεώρημα 1.7 (Flajolet and Sedgewick 2009): (α) Έστω $f(z) = (1-z)^{-a}$, $a \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$, τότε

$$[z^n]f(z) \approx \frac{n^{a-1}}{\Gamma(a)} \left[1 + \frac{a(a-1)}{2n} + \frac{a(a-1)(a-2)(3a-1)}{24n^2} + \frac{a^2(a-1)^2(a-2)(a-3)}{48n^3} + O\left(\frac{1}{n^4}\right) \right].$$

(β) Έστω $f(z) = (1-z)^r \log\left(\frac{1}{1-z}\right)$, $r \in \mathbb{Z}_{\geq 0}$, τότε

$$[z^n]f(z) \approx (-1)^r \frac{r!}{n(n-1)\dots(n-r)}.$$

Θεώρημα 1.8 (Flajolet and Sedgewick 2009): Έστω $f(z)$ Δ_1 -αναλυτική συνάρτηση στο μοναδικό ανώμαλο της σημείο $z=1$. Έστω $g(z)$ γραμμικός συνδυασμός συναρτήσεων στο σύνολο B , όπου

$$B = \{(1-z)^a \log^\beta\left(\frac{1}{1-z}\right) \mid a, \beta \in R\},$$

δηλαδή, έχουμε στην τομή της γειτονιάς του 1 με το Δ_1 -πεδίο

$$f(z) = o(g(z)) \text{ για } z \rightarrow 1.$$

Τότε έχουμε

$$[z^n]f(z) = o([z^n]g(z)),$$

όπου $o \in \{O, o, \Theta, \approx\}$.

Τα Θεωρήματα 1.7 και 1.8 είναι πολύ σημαντικά για την ανάλυση ανώμαλων σημείων των δομών των RNA ψευδοκόμπων.

Αν το $f(z)$ έχει πολλά επικρατή ανώμαλα σημεία, το $[z^n]f(z)$ ορίζεται ασυμπτωτικά από το άθροισμα όλων των επικρατών ανώμαλων σημείων.

Μία συνέπεια του Θεωρήματος 1.8 είναι το supercritical paradigm, το οποίο περιγράφεται στο Θεώρημα 1.9. Αναφέρεται στη σύνθεση δύο συναρτήσεων, όπου η “εσωτερική” συνάρτηση είναι ομαλή στο ανώμαλο σημείο της “εξωτερικής” συνάρτησης. Σε αυτή την περίπτωση, ο τυπος του ανώμαλου σημείου είναι της “εξωτερικής” συνάρτησης. Αυτό που συμβαίνει είναι ότι η εσωτερική συνάρτηση “μεταφέρει” απλά το ανώμαλο σημείο της “εξωτερικής” συνάρτησης.

Το Θεώρημα 1.9 είναι προσαρμοσμένο για τα αποτελέσματα του Κεφαλαίου 3, και η απόδειξη του παραλείπεται (Flajolet and Sedgewick 2009).

Θεώρημα 1.9: Έστω $\psi(z,s)$ μια αλγεβρική, αναλυτική συνάρτηση σε ένα πεδίο $D = \{(z,s) \mid |z| \leq r, |s| < \varepsilon\}$, τέτοιο ώστε $\psi(0,s) = 0$. Επίσης, ας υποθέσουμε ότι το $\gamma(s)$ είναι το μοναδικό επικρατές ανώμαλο σημείο της $F_k(\psi(z,s))$ και η μοναδική αναλυτική λύση του $\psi(\gamma(s),s) = \rho_k^2$, $|\gamma(s)| \leq r$, $\partial_z \psi(\gamma(s),s) \neq 0$ για $|s| < \varepsilon$. Τότε η $F_k(\psi(z,s))$ έχει μοναδικό ασυμπτωτικό ανάπτυγμα και

$$[z^n]F_k(\psi(z,s)) \approx A(s)n^{-((k-1)^2 + (k-1)/2)} \left(\frac{1}{\gamma(s)}\right)^n,$$

όπου $A(s)$ συνεχής.

Κεφάλαιο 2

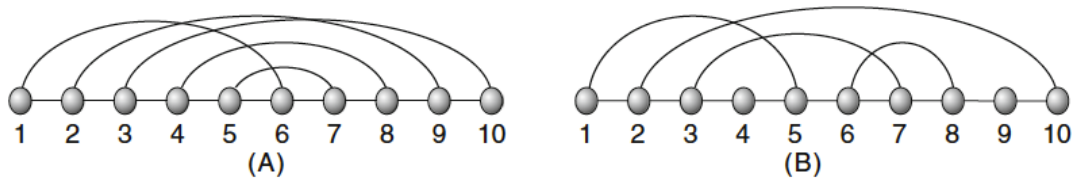
Γεννήτριες συναρτήσεις

Στο κεφάλαιο αυτό θα καταλήξουμε στη γεννήτρια συνάρτηση για k -noncrossing canonical RNA ακολουθίες με ελάχιστο μήκος τόξου 4, οι οποίες θεωρούνται ως η πιο σχετική τάξη στα πλαίσια των χαρτών αλληλουχίας - δομής.

Για να καταλήξουμε εκεί θα χρειαστούμε κάποιους ορισμούς καθώς και τις γεννήτριες συναρτήσεις απλούστερων δομών.

2.1 Ορισμοί

Ορισμός 2.1: Διάγραμμα (Εικόνα 2.1) είναι ένα γράφημα πάνω από το σύνολο σημείων $[n]=\{1,2,\dots,n\}$. Το διάγραμμα αναπαριστάται ως μια ευθεία σημείων $1,2,\dots,n$ και τόξων (i,j) , όπου $i < j$, πάνω από τα σημεία. Ως **μήκος τόξου** ορίζεται το $s = j - i$ και το αντίστοιχο τόξο ονομάζεται s -τόξο.



Εικόνα 2.1: Διαγράμματα μήκους n με 5 και 4 τόξα, αντίστοιχα. (Reidys 2011)

Stack μεγέθους σ , $S_{i,j}^\sigma$, είναι η μέγιστη ακολουθία “παράλληλων” τόξων,
 $((i, j), (i + 1, j - 1), \dots, (i + (\sigma - 1), j - (\sigma - 1)))$.

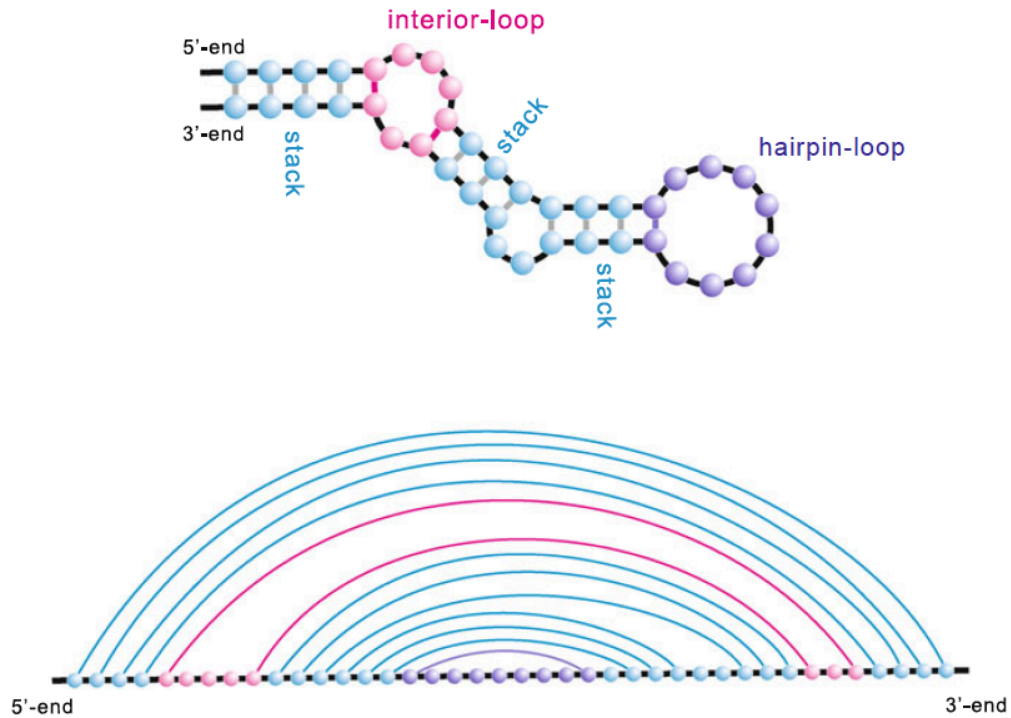
Ένα stack μεγέθους σ ονομάζεται σ -stack.

Stem μήκους s είναι η ακολουθία

$$(S_{i_1, j_1}^{\sigma_1}, S_{i_2, j_2}^{\sigma_2}, \dots, S_{i_s, j_s}^{\sigma_s}),$$

όπου το $S_{i_m, j_m}^{\sigma_m}$ είναι φωλιασμένο από το $S_{i_{m-1}, j_{m-1}}^{\sigma_{m-1}}$, έτσι ώστε κάθε τόξο φωλιασμένο από το $S_{i_{m-1}, j_{m-1}}^{\sigma_{m-1}}$ είτε να περιέχεται είτε να είναι φωλιασμένο εντός του $S_{i_{m-1}, j_{m-1}}^{\sigma_{m-1}}$, για $2 \leq m \leq s$.

Ως **hairpin** ορίζεται το ζεύγος $((i, j), [i+1, j-1])$, όπου το (i, j) είναι ένα τόξο και $[i+1, j-1]$ ένα διάστημα, δηλαδή μια ακολουθία διαδοχικών, απομονωμένων κορυφών $(i, i+1, \dots, j-1, j)$. (Εικόνα 2.2)

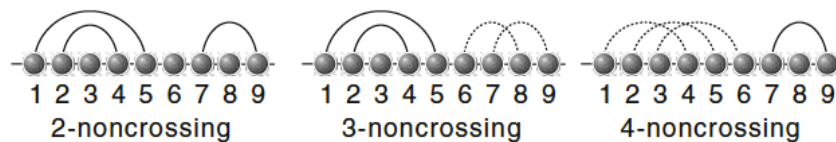


Εικόνα 2.2: Στην εικόνα παρατηρούμε τη δομή και το διάγραμμα ενός μορίου RNA, που σχηματίζει stem, το οποίο αποτελείται από τρία stacks. Επίσης, παρατηρούμε και τις θηλιές, μία εσωτερική και μία hairpin. (Reidys 2011)

Ορισμός 2.2: Το **k-crossing** είναι ένα σύνολο διακριτών τόξων $(i_1, j_1), (i_2, j_2), \dots, (i_k, j_k)$, τέτοιο ώστε:

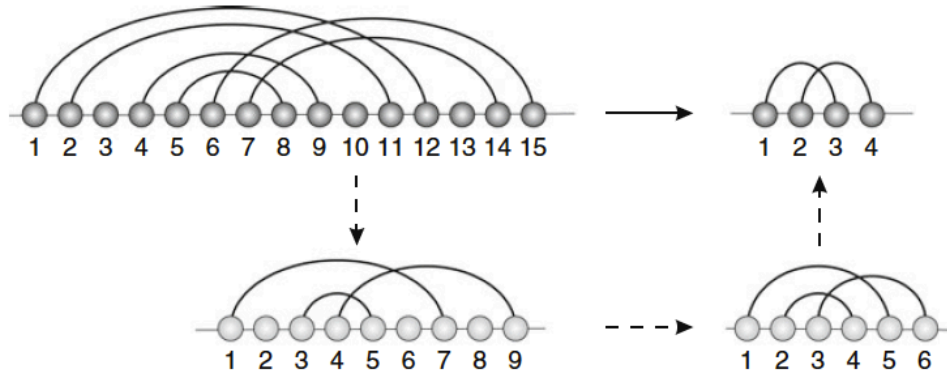
$$i_1 < i_2 < \dots < i_k < j_1 < j_2 < \dots < j_k.$$

Ένα διάγραμμα χωρίς k-crossings ονομάζεται **k-noncrossing διάγραμμα** και ένα k-noncrossing διάγραμμα χωρίς απομονωμένα σημεία ονομάζεται **k-noncrossing matching** (Εικόνα 2.3).



Εικόνα 2.3: k-noncrossing διαγράμματα.

Ορισμός 2.3: Μία **k-noncrossing, σ -canonical δομή** είναι ένα k-noncrossing διάγραμμα με ελάχιστο μήκος τόξου, $\lambda \geq 2$, και ελάχιστο μήκος stack, σ . **Πυρήνας** είναι μια k-noncrossing δομή με ελάχιστο μήκος τόξου $\lambda \geq 2$, στην οποία κάθε stack έχει μήκος 1. Το **V_k σχήμα** είναι ένα k-noncrossing matching με stacks μεγέθους ακριβώς 1. Ουσιαστικά, το V_k σχήμα είναι ένας πυρήνας χωρίς απομονωμένες κορυφές (Εικόνα 2.4).



Εικόνα 2.4: Παραγωγή ενός V_k σχήματος (πάνω δεξιά) από την αρχική δομή (πάνω αριστερά). Κάτω αριστερά, είναι ο πυρήνας της δομής αυτής. (Reidys 2011)

2.2 Γεννήτριες συναρτήσεις απλούστερων δομών

Η πρώτη γεννήτρια συνάρτηση που χρειάζεται να υπολογιστεί είναι αυτή των k -noncrossing matchings. Έστω, λοιπόν, $f_k(n)$ ο αριθμός των k -noncrossing matchings χωρίς μεμονωμένες κορυφές. Προφανώς, αν $n=2a+1$ ($a \in \mathbb{N}$), τότε $f_k(n)=0$.

Για την:

$$F_k(z) = \sum f_k(2n) \cdot z^n$$

δεν έχουμε ακριβή τύπο για τη γεννήτρια συνάρτηση. Υπάρχουν, όμως, συνήθειες διαφορικές εξισώσεις, τις οποίες ικανοποιεί και με τις οποίες μπορούμε να αποκτήσουμε πληροφορίες για αυτή (θα τις συζητήσουμε στο επόμενο κεφάλαιο).

Τις γεννήτριες συναρτήσεις των περισσότερο πολύπλοκων δομών θα τις βρούμε συναρτήσει της F_k .

Έστω $g_k(s,m)$ ο πληθικός αριθμός των k -noncrossing matchings μήκους $2s$ και με m 1-τόξα (τόξα μήκους 1) και $G_k(x,y)$ η γεννήτρια συνάρτησή τους.

$$G_k(x,y) = \sum_{s \geq 0} \sum_{m=0}^s g_k(s,m) x^s y^m$$

Λήμμα 2.4 (Reidys and Wang 2010): Έστω $k, s, m \in \mathbb{N}$, $k \geq 2$, $0 \leq m \leq s$. Τότε το $g_k(s,m)$ έχει τις παρακάτω ιδιότητες:

$$g_k(s,m) = 0 \text{ για } m > s,$$

$$\sum_{m=0}^s g_k(s,m) = f_k(2s)$$

και ισχύει η αναδρομική σχέση:

$$(m+1) \cdot g_k(s+1,m+1) = (m+1) \cdot g_k(s,m+1) + (2s+1-m) \cdot g_k(s,m).$$

Τέλος, η γεννήτρια συνάρτηση $G_k(x,y)$ δίνεται από τον παρακάτω τύπο:

$$G_k(x,y) = \frac{1}{x+1-yx} F_k\left(\frac{x}{(x+1-yx)^2}\right)$$

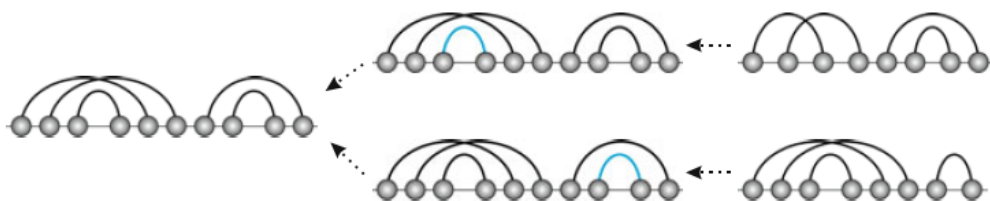
Απόδειξη:

Ο πρώτος τύπος ισχύει καθώς δεν μπορεί να υπάρχουν k -noncrossing matchings με περισσότερα από s 1-τόξα καθώς αυτό θα σήμαινε ότι μία κορυφή συμμετέχει σε περισσότερα από ένα τόξα.

Ο δεύτερος τύπος ισχύει εξ ορισμού, καθώς τα k -noncrossing matchings μήκους $2s$ μπορεί να έχουν από 0 έως s 1-τόξα. Ο τύπος αυτός είναι ισοδύναμος με τον:

$$G_k(x,1) = F_k(x).$$

Έστω, ένα k -noncrossing matching $\delta \in G_k(s+1, m+1)$ και ας σημάνουμε το ένα 1-τόξο. Έχουμε $(m+1)g_k(s+1, m+1)$ διαφορετικά σημασμένα k -noncrossing matchings. Αντίστοιχα, για να επιτύχουμε ένα τέτοιο σημασμένο matching, μπορούμε απλά να εισάγουμε ένα σημασμένο 1-τόξο σε ένα k -noncrossing matching $\delta' \in G_k(s, m+1)$. Σ' αυτή την περίπτωση, μπορούμε να τοποθετήσουμε το τόξο μέσα σε ένα προϋπάρχον 1-τόξο, ώστε να διατηρήσουμε τον αριθμό των 1-τόξων. Μπορούμε επίσης να τοποθετήσουμε ένα σημασμένο 1-τόξο σε ένα k -noncrossing matching $\delta'' \in G_k(s, m)$. Σε αυτή την περίπτωση πρέπει να τοποθετήσουμε το τόξο ανάμεσα σε δύο κορυφές που δε σχηματίζουν 1-τόξο (Εικόνα 2.5).



Εικόνα 2.5: Σημαίνοντας τα 1-τόξα, μπορούμε να ακολουθήσουμε το σχηματισμό συγκεκριμένων συνδυασμών τόξων. (Reidys 2011)

Έτσι έχουμε: $(m+1)g_k(s, m+1) + (2s+1-m)g_k(s, m)$ διαφορετικά σημασμένα matchings και:

$$(m+1)g_k(s+1, m+1) = (m+1)g_k(s, m+1) + (2s+1-m)g_k(s, m).$$

Η παραπάνω αναδρομική σχέση είναι ισοδύναμη με τη μερική διαφορική εξίσωση:

$$\frac{\partial G_k(x, y)}{\partial y} = x \frac{\partial G_k(x, y)}{\partial y} + 2x^2 \frac{\partial G_k(x, y)}{\partial x} + xG_k(x, y) - xy \frac{\partial G_k(x, y)}{\partial y}.$$

Έστω, τώρα, ότι α) η:

$$G_k^*(x, y) = \frac{1}{x+1-yx} F_k\left(\frac{x}{(x+1-yx)^2}\right)$$

είναι μια λύση της μερικής διαφορικής εξίσωσης, β) οι συντελεστές $g_k^*(s, m) = [x^s y^m] G_k^*(x, y)$, ικανοποιούν την $g_k^*(s, m) = 0$ για $m > s$ και γ) $G_k^*(x, 1) = F_k(x)$.

Πράγματι, για το α) έχουμε:

$$\begin{aligned}\frac{\partial G_k^*(x,y)}{\partial y} &= xuF_k(xu) + 2xuF_k'(xu), \\ \frac{\partial G_k^*(x,y)}{\partial x} &= (y-1)uF_k(xu) + \frac{(1+yx-x)u}{x}F_k'(xu), \\ u &= (x+1-yx)^{-2}, \\ F_k'(xu) &= \sum_{s \geq 0} sf_k(2s)(xu)^s\end{aligned}$$

Συνεπώς,

$$(1+xy-x)\frac{\partial G_k^*(x,y)}{\partial y} = 2x^2\frac{\partial G_k^*(x,y)}{\partial x} + xG_k^*(x,y)$$

που συμπίπτει με τη διαφορική εξίσωση. Για να δείξουμε το β), αρκεί να παρατηρήσουμε ότι η $G_k^*(x,y)$ είναι δυναμοσειρά, καθώς είναι αναλύτική στο $(0,0)$. Επίσης, παρατηρούμε ότι το y εμφανίζεται μόνο ως γινόμενο xy , οπότε το β) ισχύει. Τέλος, το γ) είναι προφανές.

Έστω ότι:

$$G_k^*(x,y) = G_k(x,y).$$

Από κατασκευής:

$$\begin{aligned}g_k^*(s,m) &= 0, \text{ αν } m > s \\ \sum_{m=0}^s g_k^*(s,m) &= f_k(2s)\end{aligned}$$

$$(m+1)g_k^*(s+1,m+1) = (m+1)g_k^*(s,m+1) + (2s+1-m)g_k^*(s,m).$$

Χρησιμοποιώντας αυτές τις ιδιότητες και με αναγωγή στο s ,

$$\forall s, m \geq 0, \text{ ισχύει } g_k^*(s,m) = g_k(s,m).$$

Αποδείχτηκε.

Στη συνέχεια, θα μελετήσουμε τη γεννήτρια συνάρτηση $I_k(s,m)$ των k -noncrossing διαγραμμμάτων, μήκους $2s$ και με m 1-τόξα.

Θεώρημα 2.5 (Reidys and Wang 2010): Έστω k, s, m φυσικοί αριθμοί όπου $k \geq 2$. Τότε ισχύει η παρακάτω εξίσωση:

$$I_k(z,u) = \frac{1+z}{1+2z-zu} F_k\left(\frac{z(1+z)}{(1+2z-zu)^2}\right).$$

Απόδειξη:

Έστω η αντιστοιχία μεταξύ των k -noncrossing matchings με m 1-τόξα και των V_k σχημάτων τους:

$$g: G_k(s,m) \rightarrow \bigcup_{b=0}^{s-m} [I_k(s-b,m) \times \{(a_j)_{1 \leq j \leq s-b} \mid \sum_{j=1}^{s-b} a_j = b, a_j \geq 0\}], \text{ όπου } s \geq 1.$$

Εδώ, για κάθε $\delta \in G_k(s,m)$, έχουμε το $g(\delta) = (c(\delta), (a_j)_{1 \leq j \leq s-b})$, όπου το $c(\delta)$ είναι ο πυρήνας του δ και το a_j μετρά τα διαγραμμένα τόξα. Είναι εύκολο να δειχτεί ότι ο g είναι καλά ορισμένος, δεδομένου ότι όλα τα 1-τόξα του $c(\delta)$ είναι απλά τα 1-τόξα του δ . Επίσης, παρατηρούμε ότι η απεικόνιση c είναι 1 προς ένα και επί (αμφιριπτική). Δεδομένου ότι η c είναι επιριπτική (επί), έχουμε:

$$G_k(x,y) = \sum_{s,m} g_k(s,m) x^s y^m = \sum_m \sum_{\gamma \in I_k(m)} G_\gamma(x,y),$$

όπου $I_k(m)$ είναι το σύνολο των V_k σχημάτων που έχουν m 1-τόξα και $G_\gamma(x,y)$ είναι η γεννήτρια συνάρτηση όλων των k -noncrossing matchings που έχουν m 1-τόξα που προβάλλουν στο σχήμα γ . Έστω ότι το γ έχει s τόξα. Έστω και οι συνδυαστικές τάξεις των τόξων R και των 1-τόξων R^* με γεννήτριες συναρτήσεις $R(x)=x$ και $R^*(x,y)=yx$. Τότε:

- κάθε k -noncrossing matching που έχει V_k σχήμα γ ανακτάται “φουσκώνοντας” τα τόξα του γ σε stacks και η συνδυαστική τάξη των stacks δίνεται από το $R \times SEQ(R)$,
- το “φούσκωμα” δεν επηρεάζει τον αριθμό των 1-τόξων.

Έτσι έχουμε:

$$G_\gamma(x,y) = \left(\frac{x}{1-x}\right)^s y^m. \text{ (βλ. Κεφάλαιο 1.2)}$$

Για κάθε $\gamma, \gamma_1 \in I_k(m)$, έχοντας s τόξα, έχουμε $G_\gamma(x,y) = G_{\gamma_1}(x,y)$, οπότε:

$$G_k(x,y) = \sum_m \sum_{\gamma \in I_k(m)} G_\gamma(x,y) = \sum_{s \geq 0} \sum_{m=0} i_k(s,m) \left(\frac{x}{1-x}\right)^s y^m.$$

Σύμφωνα με το Λήμμα 2.4, έχουμε:

$$G_k(x,y) = \frac{1}{x+1-yx} F_k\left(\frac{x}{(x+1-yx)^2}\right)$$

Θέτοντας $z = \frac{x}{1-x}$ και $u = y$ και αντικαθιστώντας στην εξίσωση, καταλήγουμε στο επιθυμητό:

$$I_k(z,u) = \frac{1+z}{1+2z-zu} F_k\left(\frac{z(1+z)}{(1+2z-zu)^2}\right).$$

Αποδείχτηκε.

Στη συνέχεια, θα υπολογίσουμε κάποιες γεννήτριες συναρτήσεις χρησιμοποιώντας συμβολική απαρίθμηση και “φουσκώνοντας” V_k σχήματα. Τα V_k σχήματα παράγονται αν από μια δοσμένη k -noncrossing σ -canonical RNA δομή αφαιρέσουμε αρχικά τις μεμονωμένες κορυφές και στη συνέχεια περιορίσουμε όλα τα stacks σε απλά τόξα. Αρχικά, θα υπολογίσουμε τη γεννήτρια συνάρτηση των πυρήνων.

Θεώρημα 2.6 (Reidys 2011): Έστω $k \in \mathbb{N}$, $k \geq 2$, z μεταβλητή και $r(z) = \frac{1}{1+z^2}$.

Τότε

$$C_k(z) = \frac{1}{r(z)z^2 - z + 1} F_k\left(\left(\frac{\sqrt{r(z)}z}{r(z)z^2 - z + 1}\right)^2\right).$$

Απόδειξη:

Ως C_k θα συμβολίζουμε το σύνολο των k -noncrossing πυρήνων, ως I_k το σύνολο των k -noncrossing σχημάτων και ως $I_k(m)$ το σύνολο των k -noncrossing σχημάτων που έχουν m 1-τόξα. Έχουμε, λοιπόν, την επιρριπτική απεικόνιση:

$$\varphi: C_k \rightarrow I_k$$

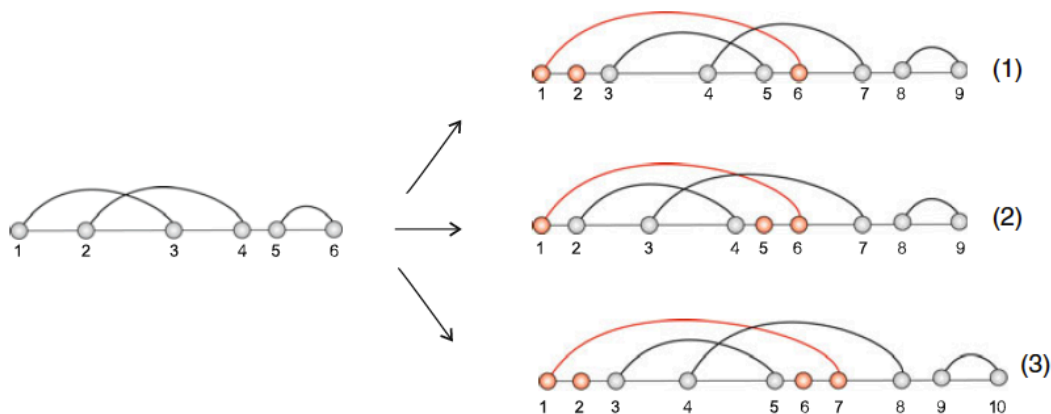
που επάγει τη διαμέριση του $C_k = \bigcup_{\gamma} \varphi^{-1}(\gamma)$, όπου $C_k(\gamma)$ είναι το σύνολο των k -noncrossing πυρήνων που έχουν σχήμα γ . Τότε

$$C_k(z) = \sum_{m \geq 0} \sum_{\gamma \in I_k(m)} C_{\gamma}(z),$$

όπου $C_{\gamma}(z)$ είναι η γεννήτρια συνάρτηση της συνδυαστικής τάξης $C_k(\gamma)$. Στη συνέχεια, υπολογίζουμε τη $C_{\gamma}(z)$ συμβολικά “φουσκώνοντας” V_k σχήματα. Έστω C_{γ} η συνδυαστική τάξη των πυρήνων που προέρχονται από το “φούσκωμα” του V_k σχήματος γ . Για να δημιουργήσουμε αυτή την τάξη, θεωρούμε τις τάξεις M (stems τόξων), L (απομονωμένες κορυφές), R (τόξα), R' (επαγόμενα τόξα) και Z (κορυφές), όπου $Z(z)=z$ και $R(z)=z^2$.

Η ιδέα είναι να “φουσκώσουμε” ένα V_k σχήμα $\gamma \in I_k(s)$ σε δύο βήματα:

Βήμα 1: Αρχικά, το $\gamma \in I_k(m)$ που έχει s τόξα, με $s \geq \max\{1, m\}$ “φουσκώνεται” σε πυρήνα, “φουσκώνοντας” κάθε τόξο του V_k σχήματος σε ένα stem τόξων (Εικόνα 2.6).



Εικόνα 2.6: Βήμα 1 (Reidys 2011)

Τα εξωτερικά τόξα που καλούνται επαγόμενα πρέπει να διαχωρίζονται με απομονωμένες κορυφές, όπως φαίνεται και στο σχήμα. Βάζουμε λοιπόν ομάδες απομονωμένων κορυφών, είτε αριστερά (1), είτε δεξιά (2), είτε κι από τις δύο πλευρές (3). Έτσι, παράγονται τα παρακάτω:

- Αρχικά, τα απομονωμένα τμήματα, δηλαδή ομάδες απομονωμένων κορυφών. Απλώς, έχουμε $L=SEQ(Z)$, όπου

$$L(z) = \frac{1}{1-z}$$

- Μετά, τα επαγόμενα τόξα, δηλαδή ζεύγη αποτελούμενα από τόξα R και τουλάχιστον μια ομάδα απομονωμένων κορυφών στη μία ή και τις δύο πλευρές. Μιας και η κορυφή μπορεί να συνδυαστεί ελεύθερα με αυτά τα διαστήματα, έχουμε $R' = R \times ((Z \times L) + (Z \times L) + (Z \times L)^2)$, με γεννήτρια συνάρτηση:

$$R'(z) = z^2 \left(\frac{z}{1-z} + \frac{z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right).$$

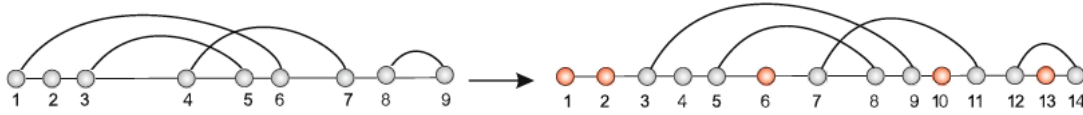
- Και, τέλος, τα stems, δηλαδή ζεύγη αποτελούμενα από ένα ελάχιστο (minimal) τόξο R και μία αυθαίρετα μακριά αλληλουχία επαγόμενων τόξων $M = R \times SEQ(R')$, με γεννήτρια συνάρτηση

$$M(z) = z^2 \frac{1}{1-R'(z)}.$$

Ο κατασκευασμένος πυρήνας έχει s stems τόξων και $(2s+1)$ (πιθανώς άδεια) διαστήματα απομονωμένων κορυφών.

Βήμα 2: Εισάγουμε απομονωμένες κορυφές στις εναπομείνουσες $2s-1+2$ θέσεις. Αυτό το δεύτερο “φούσκωμα” (Εικόνα 2.7) εκφράζεται ως εξής: $J = L^{2s+1-m} \times (Z \times L)^m$, όπου:

$$J(z) = \left(\frac{1}{1-z} \right)^{2s+1-m} \left(\frac{z}{1-z} \right)^m.$$



Εικόνα 2.7: Βήμα 2 (Reidys 2011)

Συνδυάζοντας τα βήματα 1 και 2, καταλήγουμε στην:

$$C_\gamma = M^s \times L^{2s+1-m} \times (Z \times L)^m,$$

και υπολογίζουμε:

$$\begin{aligned} C_\gamma(z) &= \left(\frac{z^2}{1-z^2 \left(\frac{2z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} \right)^s \cdot \left(\frac{1}{1-z} \right)^{2s+1-m} \cdot \left(\frac{z}{1-z} \right)^m \\ &= (1-z)^{-1} \left(\frac{z^2}{1-2z+z^2-2z^3+z^4} \right)^s z^m. \end{aligned}$$

Δεδομένου ότι για κάθε γ , $\gamma_1 \in I_k(s,m)$ έχουμε $C_\gamma(z) = C_{\gamma_1}(z)$

$$C_k(z) = \sum_{m \geq 0} \sum_{\gamma \in I_k(m)} C_\gamma(z) = \sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) C_\gamma(z).$$

Συνεπώς, έχουμε για τη γεννήτρια συνάρτηση των k -noncrossing πυρήνων

$$(2.1) \quad \begin{aligned} C_k(z) &= \sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) C_\gamma(z) \\ &= (1-z)^{-1} \sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) \left(\frac{z^2}{1-2z+z^2-2z^3+z^4} \right)^s z^m. \end{aligned}$$

Σύμφωνα με το Θεώρημα 2.5:

$$\sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) x^s y^m = \frac{1+x}{1+2x-xy} \sum_{s \geq 0} f_k(2s) \left(\frac{x(1+x)}{(1+2x-xy)^2} \right)^s.$$

Αντικαθιστώντας $x = \frac{z^2}{1-2z+z^2-2z^3+z^4}$ και $y = z$ στην προηγούμενη εξίσωση, έχουμε:

$$\sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) \left(\frac{z^2}{1-2z+z^2-2z^3+z^4} \right)^s z^m = \frac{(1-z)(1+z^2)}{1-z+2z^2-z^3} F_k \left(\frac{z^2(1+z^2)}{(1-z+2z^2-z^3)^2} \right).$$

Αντικαθιστώντας στην εξίσωση 2.1, υπολογίζουμε:

$$\begin{aligned} C_k(z) &= \frac{1}{1-z} \cdot \frac{(1-z)(1+z^2)}{1-z+2z^2-z^3} F_k \left(\frac{z^2(1+z^2)}{(1-z+2z^2-z^3)^2} \right) \\ &= \frac{1}{1-z + \frac{1}{1+z^2} z^2} F_k \left(\left(\frac{\sqrt{\frac{1}{1+z^2}} z}{\frac{1}{1+z^2} z^2 - z + 1} \right)^2 \right). \end{aligned}$$

Αποδείχτηκε.

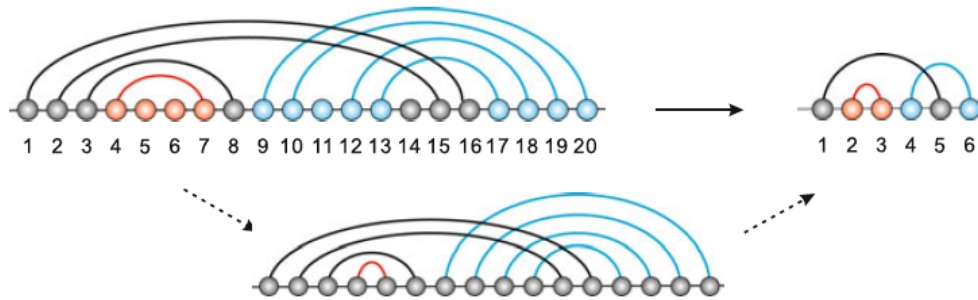
Στη συνέχεια, θα υπολογίσουμε τη γεννήτρια συνάρτηση των k -noncrossing σ -canonical δομών. Αξίζει να σημειωθεί ότι κανένα από αυτά τα αποτελέσματα δεν ισχύει στην περίπτωση των k -noncrossing, 2-canonical δομών με ελάχιστο μήκος τόξου 4. Η ανάλυση αυτών των δομών θα γίνει στην επόμενη ενότητα.

Θεώρημα 2.7 (Reidys 2011): Έστω $k, \sigma \in \mathbb{N}$, $k \geq 2$, $\sigma \geq 1$ και $u_\sigma = \frac{(z^2)^{\sigma+1}}{z^{2\sigma} - z^2 + 1}$. Τότε

$$T_{k,\sigma}(z) = \frac{1}{u_\sigma(z)z^2 - z + 1} F_k \left(\left(\frac{\sqrt{u_\sigma(z)}z}{(u_\sigma(z)z^2 - z + 1)} \right)^2 \right).$$

Απόδειξη:

Έστω $T_{k,\sigma}$ το σύνολο των k -noncrossing σ -canonical δομών και I_k το σύνολο όλων των k -noncrossing V_k σχημάτων και $I_k(m)$ αυτών που έχουν m 1-τόξα (Εικόνα 2.8).



Εικόνα 2.8: Μια 3-noncrossing, 2-canonical δομή RNA αντιστοιχίζεται στο V_k σχήμα της. Πρώτα αφαιρούνται οι μεμονωμένες κορυφές και μετά το μπλε stem μετατρέπεται σε μοναδικό τόξο και το κόκκινο hairpin μετατρέπεται σε 1-τόξο. (Reidys 2011)

Έχουμε, λοιπόν, την επιρριπτική απεικόνιση:

$$\varphi: T_{k,\sigma} \rightarrow I_k$$

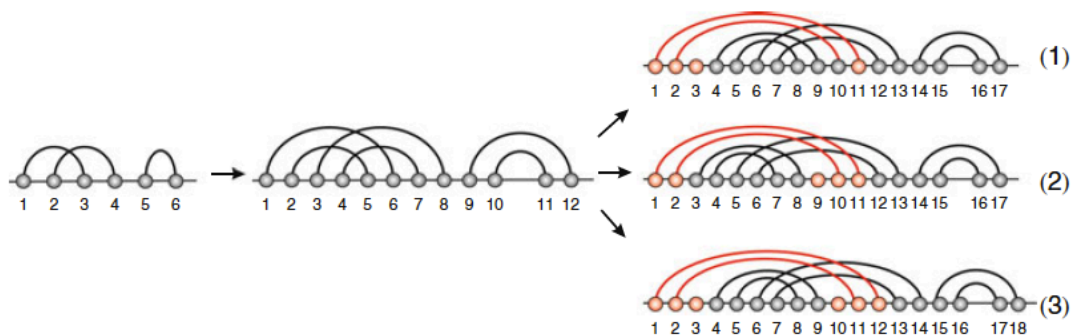
Πράγματι, για κάθε V_k σχήμα γ στο $I_k(m)$, μπορούμε να κατασκευάσουμε μία k -noncrossing σ -canonical δομή με m hairpins, προσθέτοντας τουλάχιστον $\sigma-1$ τόξα σε κάθε stack και εισάγοντας τουλάχιστον μία απομονωμένη κορυφή σε κάθε 1-τόξο. Η απεικόνιση αυτή επάγει τη διαμέριση του $T_{k,\sigma} = \bigcup_{\gamma} \varphi^{-1}(\gamma)$. Τότε

$$(2.2) \quad T_{k,\sigma}(z) = \sum_{m \geq 0} \sum_{\gamma \in I_k(m)} T_{\gamma}(z).$$

Στη συνέχεια, υπολογίζουμε τη γεννήτρια συνάρτηση $T_{\gamma}(z)$. Θα την κατασκευάσουμε από απλούστερες συνδυαστικές τάξεις θεωρώντας τις τάξεις M (stems), K^{σ} (stacks), N^{σ} (επαγόμενα stacks), L (απομονωμένες κορυφές), R (τόξα) και Z (κορυφές), όπου $Z(z)=z$ και $R(z)=z^2$.

Η ιδέα είναι να “φουσκώσουμε” ένα σχήμα $\gamma \in I_k(m)$, που έχει s τόξα, με $s \geq \max\{1,m\}$ σε δύο βήματα:

Βήμα 1: “Φουσκώνουμε” κάθε τόξο του σχήματος σε ένα stack μεγέθους τουλάχιστον σ και στη συνέχεια προσθέτουμε κι άλλα stacks. Τα τελευταία καλούνται επαγόμενα stacks και πρέπει να διαχωρίζονται με απομονωμένες κορυφές, αντίστοιχα με το Θεώρημα 2.6, όπως φαίνεται και στην Εικόνα 2.9.



Εικόνα 2.9: Βήμα 1: Ένα V_k σχήμα (αριστερά) “φουσκώνεται” σε 3-noncrossing, 2-canonical δομή. Αρχικά, κάθε τόξο “φουσκώνεται” σε stack μεγέθους τουλάχιστον 2 (μέση) και μετά σε μια νέα 3-noncrossing, 2-canonical δομή (δεξιά) προσθέτοντας ένα stack μεγέθους 2. Υπάρχουν τρεις τρόποι να προστεθούν οι απομονωμένες κορυφές. (Reidys 2011)

Πρέπει να σημειωθεί ότι κατά το πρώτο βήμα δεν εισάγονται άλλα διαστήματα απομονωμένων κορυφών, εκτός από τα απαραίτητα για να διαχωρίσουν τα ένθετα stacks. Έτσι, παράγονται τα παρακάτω:

- Αρχικά, απομονωμένα τμήματα, δηλαδή ομάδες απομονωμένων κορυφών. Έχουμε $L=SEQ(Z)$, όπου

$$L(z) = \frac{1}{1-z}$$

- Μετά, stacks, δηλαδή ζεύγη αποτελούμενα από (α) την ελάχιστη ακολουθία τόξων R^σ και (β) μια αυθαίρετη επέκταση αποτελούμενη από τόξα αυθαίρετου πεπερασμένου μήκους. $K^\sigma = R^\sigma \times SEQ(R)$, με γεννήτρια συνάρτηση

$$K^\sigma(z) = z^{2\sigma} \cdot \frac{1}{1-z^2}$$

- Στη συνέχεια, επαγόμενα stacks, δηλαδή stacks μαζί με τουλάχιστον ένα, μη άδειο διάστημα απομονωμένων κορυφών στη μία ή και τις δύο πλευρές.

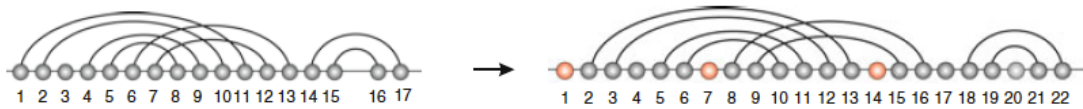
$N^\sigma = K^\sigma \times ((Z \times L) + (Z \times L) + (Z \times L)^2)$, με γεννήτρια συνάρτηση:

$$N^\sigma(z) = \frac{z^{2\sigma}}{1-z^2} \left(2 \frac{z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)$$

- Τέλος, stems, δηλαδή ζεύγη αποτελούμενα από stacks και μία αυθαίρετα μακριά αλληλουχία επαγόμενων stacks $M^\sigma = K^\sigma \times SEQ(N^\sigma)$, με γεννήτρια συνάρτηση

$$M^\sigma(z) = \frac{K^\sigma(z)}{1-N^\sigma(z)} = \frac{\frac{z^{2\sigma}}{1-z^2}}{1 - \frac{z^{2\sigma}}{1-z^2} \left(2 \frac{z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)}$$

Βήμα 2: Εισάγουμε απομονωμένες κορυφές στις εναπομείνουσες $2s+1$ θέσεις. Για κάθε 1-τόξο πρέπει να εισαχθεί τουλάχιστον μία απομονωμένη κορυφή (Εικόνα 2.10).



Εικόνα 2.10: Βήμα 2: Η δομή του Βήματος 1 (αριστερά) “φουσκώνει” σε μια νέα 3-noncrossing, 2-canonical δομή (δεξιά) προσθέτοντας απομονωμένες κορυφές. (Reidys 2011)

Αυτό το δεύτερο “φούσκωμα” εκφράζεται ως εξής: $J = L^{2s+1-m} \times (Z \times L)^m$, όπου:

$$J(z) = \left(\frac{1}{1-z} \right)^{2s+1-m} \left(\frac{z}{1-z} \right)^m.$$

Συνδυάζοντας τα βήματα 1 και 2, καταλήγουμε στην:

$$T_\gamma = (M^\sigma)^s \times L^{2s+1-m} \times (Z \times L)^m,$$

και αντίστοιχα:

$$T_\gamma(z) = \left(\frac{\frac{z^{2\sigma}}{1-z^2}}{1 - \frac{z^{2\sigma}}{1-z^2} \left(2 \frac{z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} \right)^s \left(\frac{1}{1-z} \right)^{2s+1-m} \left(\frac{z}{1-z} \right)^m$$

$$= (1-z)^{-1} \left(\frac{z^{2\sigma}}{(1-z^2)(1-z)^2 - (2z-z^2)z^{2\sigma}} \right)^s z^m$$

Δεδομένου ότι για κάθε $\gamma, \gamma_1 \in I_k(s,m)$ έχουμε $T_\gamma(z) = T_{\gamma_1}(z)$, βρίσκουμε ότι

$$T_k(z) = \sum_{m \geq 0} \sum_{\gamma \in I_k(m)} T_\gamma(z) = \sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) T_\gamma(z).$$

Θέτοντας

$$\eta_\sigma(z) = \frac{z^{2\sigma}}{(1-z^2)(1-z)^2 - (2z-z^2)z^{2\sigma}},$$

έχουμε βάσει της εξίσωσης 2.2 και του Θεωρήματος 2.5 την παρακάτω κατάσταση:

$$T_{k,\sigma}(z) = \sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) T_\gamma(z)$$

και σύμφωνα με το Θεώρημα 2.5

$$\sum_{s \geq 0} \sum_{m=0}^s i_k(s,m) x^s y^m = \frac{1+x}{1+2x-xy} \sum_{s \geq 0} f_k(2s) \left(\frac{x(1+x)}{(1+2x-xy)^2} \right)^s.$$

Αντικαθιστώντας $x = \eta_\sigma(z)$ και $y = z$ και $w_\sigma(z) = z^{2\sigma} - z^2 + 1$ στην προηγούμενη εξίσωση, έχουμε:

$$T_{k,\sigma}(z) = \frac{w_\sigma(z)}{(1-z)w_\sigma(z) + z^{2\sigma}} F_k \left(\frac{z^2(z^2)^{\sigma-1} w_\sigma(z)}{((1-z)w_\sigma(z) + z^{2\sigma})^2} \right)$$

$$= \frac{1}{(1-z) + u_\sigma(z)z^2} F_k \left(\frac{z^2 u_\sigma(z)}{((1-z) + u_\sigma(z)z^2)^2} \right)$$

Αποδείχτηκε.

2.3 Αρθρωτές k-noncrossing δομές

Στα πλαίσια των χαρτών αλληλουχίας - δομής, οι 2-canonical (από δω και πέρα θα αναφέρονται ως canonical) δομές με ελάχιστο μήκος τόξου ≥ 4 θεωρούνται η πιο σχετική τάξη. Στην προηγούμενη ενότητα, χρησιμοποιήσαμε τα 1-τόξα για τον υπολογισμό των γεννητριών συναρτήσεων. Θα δούμε ότι κλειδί για τον υπολογισμό των γεννητριών συναρτήσεων των k-noncrossing canonical δομών είναι τα 2-τόξα των V_k σχημάτων. Αυτά, αν και περισσότερο πολύπλοκα από τα 1-τόξα αφού μπορούν να διασταυρώνονται, είναι εύκολο να ταξινομηθούν. Αντίστοιχα με το Λήμμα 2.4 μπορούμε να εξάγουμε αναδρομικές σχέσεις (δεν παρουσιάζονται στα πλαίσια αυτής της εργασίας), που τελικά οδηγούν στην

εύρεση της επιθυμητής γεννήτριας συνάρτησης των k -noncrossing canonical δομών με ελάχιστο μήκος τόξου 4.

Έστω, λοιπόν, Q_k το σύνολο αυτών των δομών (που ονομάζονται και αρθρωτές k -noncrossing δομές) και:

$$Q_k(z) = \sum_{n=0}^{\infty} Q_k(n)z^n$$

η γεννήτρια συνάρτηση.

Πρόταση 2.8: Η γεννήτρια συνάρτηση των αρθρωτών noncrossing δομών δίνεται από τον τύπο:

$$Q_2(z) = \frac{1 - z^2 + z^4}{1 - z - z^2 + z^3 + 2z^4 + z^6} \cdot F_2\left(\frac{z^4 - z^6 + z^8}{(1 - z - z^2 + z^3 + 2z^4 + z^6)^2}\right).$$

Απόδειξη:

Χρησιμοποιώντας την ορολογία του Θεωρήματος 2.7, μπορούμε να εκφράσουμε το $Q_2(z)$ μέσω V_k σχημάτων γ , που έχουν s τόξα, m από τα οποία είναι 1-τόξα. Αυτό οδηγεί στις παρακάτω συνδυαστικές τάξεις:

$$Q_\gamma = M^s \times L^{2s+1-m} \times (Z^3 \times L)^m$$

$$M = K^\sigma \times \text{SEQ}(N^\sigma)$$

$$N^\sigma = K^\sigma \times ((Z \times L) + (Z \times L) + (Z \times L)^2)$$

$$K^\sigma = R \times \text{SEQ}(R)$$

$$L = \text{SEQ}(Z)$$

Αρχικά, εισάγουμε ένα τμήμα αποτελούμενο από τουλάχιστον τρεις απομονωμένες κορυφές μέσα σε οποιοδήποτε 1-τόξο $J^{[3]} = L^{2s+1-m} \times (Z^3 \times L)^m$, δηλαδή,

$$J^{[3]}(z) = \left(\frac{1}{1-z}\right)^{2s+1-m} \left(\frac{z^3}{1-z}\right)^m.$$

Μιας και έχουμε μόνο φωλιασμένα τόξα στα V_2 σχήματα (δε διασταυρώνονται), κάθε μη-1-τόξο μπορεί, μετά την παραπάνω εισαγωγή απομονωμένων κορυφών, να "φουσκωθεί" αυθαίρετα. Έτσι,

$$\begin{aligned} Q_\gamma(z) &= \left(\frac{\frac{z^{2\sigma}}{1-z^2}}{1 - \frac{z^{2\sigma}}{1-z^2} \left(\frac{2z}{1-z} + \left(\frac{z}{1-z}\right)^2\right)}\right)^s \cdot \left(\frac{1}{1-z}\right)^{2s+1-m} \left(\frac{z^3}{1-z}\right)^m \\ &= (1-z)^{-1} \left(\frac{z^{2\sigma}}{(1-z)^2(1-z^2) - z^{2\sigma}(2z-z^2)}\right)^s (z^3)^m. \end{aligned}$$

Δεδομένου ότι για κάθε $\gamma, \gamma' \in I_2(s, m)$ έχουμε $Q_\gamma(z) = Q_{\gamma'}(z)$, βρίσκουμε ότι

$$Q_2(z) = \sum_{m, s \geq 0} \sum_{\gamma \in I_2(s, m)} Q_\gamma(z) = \sum_{s \geq 0} \sum_{m=0}^s i_2(s, m) Q_\gamma(z).$$

Οπότε:

$$Q_2(z) = (1-z)^{-1} \sum_{s \geq 0} \sum_{m=0}^s i_2(s,m) \eta_\sigma(z)^\sigma (z^3)^m$$

Αντικαθιστώντας $z = \eta_\sigma(z)$ και $u = z^3$ στην εξίσωση του Θεωρήματος 2.5, έχουμε:

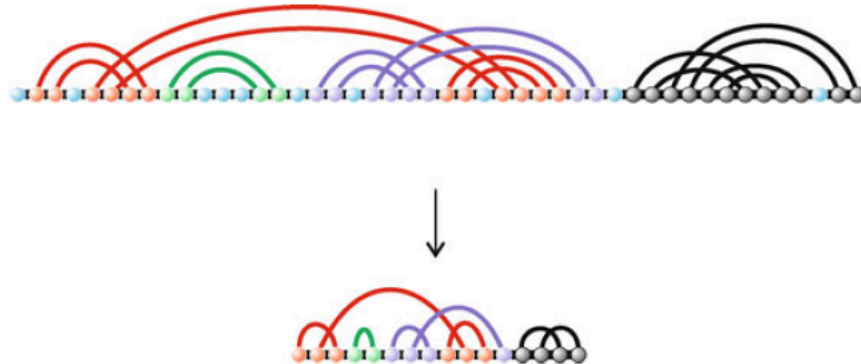
$$Q_2(z) = (1-z)^{-1} \frac{1 + \eta_\sigma(z)}{1 + 2\eta_\sigma(z) - \eta_\sigma(z)z^3} F_2\left(\frac{\eta_\sigma(z)(1 + \eta_\sigma(z))}{(1 + 2\eta_\sigma(z) - \eta_\sigma(z)z^3)^2}\right)$$

$$= \frac{(1-z)(1-z^2+z^{2\sigma})}{1-2z+2z^3-z^4+2z^{2\sigma}-2z^{2\sigma+1}+z^{2\sigma+2}-z^{2\sigma+3}} F_2\left(\frac{(1-z)^2 z^{2\sigma} (1-z^2+z^{2\sigma})}{(1-2z+2z^3-z^4+2z^{2\sigma}-2z^{2\sigma+1}+z^{2\sigma+2}-z^{2\sigma+3})^2}\right).$$

Αν θέσουμε $\sigma=2$, αποδείχτηκε.

Η κατάσταση είναι αρκετά διαφορετική στην περίπτωση του $k>2$. Για να καταλάβουμε τις αρθρωτές, k -noncrossing δομές πρέπει να διαχωρίσουμε τα 2-τόξα, δηλαδή τα τόξα της μορφής $(i, i+2)$, σε τάξεις. Κάθε τάξη απαιτεί ξεχωριστή διαδικασία “φουσκώματος” όπως θα δούμε και στο τελικό θεώρημα. Ποιες είναι όμως αυτές οι τάξεις; (Εικόνα 2.11)

- **C₁**: η τάξη των 1-τόξων
- **C₂**: η τάξη των ζευγών τόξων που αποτελούνται από αλληλοδιασταυρούμενα τόξα
- **C₃**: η τάξη των ζευγών τόξων (α, β) όπου το α είναι το μοναδικό 2-τόξο που διασταυρώνεται με το β και το β έχει μήκος τουλάχιστον 3
- **C₄**: η τάξη των τριάδων τόξων $(\alpha_1, \beta, \alpha_2)$ όπου τα α_1 και α_2 είναι 2-τόξα που διασταυρώνονται με το β .



Εικόνα 2.11: Μια αρθρωτή, 2-noncrossing δομή και το V_k σχήμα της. Στο V_k σχήμα φαίνονται οι τέσσερις τάξεις με πράσινο (**C₁**), μαύρο (**C₂**), μπλε (**C₃**) και κόκκινο (**C₄**). (Reidys 2011)

Πριν καταλήξουμε στο τελευταίο Θεώρημα, το οποίο και είναι ο τελικός στόχος αυτού του Κεφαλαίου, χρειάζεται να αναφερθούν κάποια Λήμματα και Προτάσεις, απαραίτητα για το τελικό θεώρημα. Η απόδειξή τους είναι εκτός στόχου αυτής της εργασίας και μπορεί να βρεθεί στο *Combinatorial Computational Biology of RNA* (Reidys 2011).

Αρχικά, ας υπολογίσουμε τη γεννήτρια συνάρτηση των V_k σχημάτων,

$$W_k(x, y, w) = \sum_{s \geq 0} \sum_{u_1=0}^s \sum_{u_2=0}^{\lfloor \frac{s-u_1}{2} \rfloor} i_k(s, u_1, u_2) x^s y^{u_1} w^{u_2}$$

λαμβάνοντας υπόψιν μας τις τάξεις C_1 και C_2 . Έστω $I_k(s, u_1, u_2)$ το σύνολο των V_k σχημάτων με s -τόξα, u_1 1-τόξα και u_2 αλληλοδιασταυρούμενα 2-τόξα και $i_k(s, u_1, u_2)$ ο πληθικός τους αριθμός.

Λήμμα 2.9: Για $k > 2$, για τους συντελεστές $i_k(s, u_1, u_2)$ ισχύει:

$$i_k(s, u_1, u_2) = 0 \text{ για } u_1 + 2u_2 > s$$

$$\sum_{u_2=0}^{\lfloor \frac{s-u_1}{2} \rfloor} i_k(s, u_1, u_2) = i_k(s, u_1),$$

όπου $i_k(s, u_1)$ είναι ο αριθμός των V_k σχημάτων που έχουν s τόξα και u_1 1-τόξα. Επίσης ισχύει η αναδρομική σχέση:

$$(u_2 + 1)i_k(s + 1, u_1, u_2 + 1) = (u_1 + 1)i_k(s, u_1 + 1, u_2) + (u_1 + 1)i_k(s - 1, u_1 + 1, u_2).$$

Έστω $I_k(s, u_1, u_2, u_3, u_4)$ το σύνολο των V_k σχημάτων με s -τόξα, u_i στοιχεία της C_i τάξης και $i_k(s, u_1, u_2, u_3, u_4)$ ο πληθικός τους αριθμός.

Πρόταση 2.10: Για $k > 2$, έχουμε

$$W_k(x, y, w) = \frac{1 + x}{(1 - w)x^3 + (1 - w)x^2 + (2 - y)x + 1} F_k\left(\frac{x(1 + x)}{((1 - w)x^3 + (1 - w)x^2 + (2 - y)x + 1)^2}\right).$$

Λήμμα 2.11: Για $k > 2$, οι συντελεστές $i_k(s, \mathbf{u})$, ικανοποιούν τα παρακάτω:

$$I_k(s, u_1, u_2, u_3, u_4) = 0 \text{ για } u_1 + 2u_2 + 2u_3 + 3u_4 > s$$

$$\sum_{u_3, u_4 \geq 0} i_k(s, u_1, u_2, u_3, u_4) = i_k(s, u_1, u_2).$$

Επίσης, ισχύουν οι αναδρομικές σχέσεις:

$$\begin{aligned} & (u_3 + 1)i_k(s + 1, u_1, u_2, u_3 + 1, u_4) = \\ & 2u_1 i_k(s - 1, u_1, u_2, u_3, u_4) + 4(u_2 + 1)i_k(s - 1, u_1, u_2 + 1, u_3, u_4) + 4(u_2 + 1)i_k(s - 1, u_1, u_2 + 1, u_3 - 1, u_4) \\ & + 4(u_2 + 1)i_k(s - 2, u_1, u_2 + 1, u_3 - 1, u_4) + 2u_3 i_k(s - 1, u_1, u_2, u_3, u_4) + 6(u_3 + 1)i_k(s - 1, u_1, u_2, u_3 + 1, u_4) \\ & + 2(u_3 + 1)i_k(s - 2, u_1, u_2, u_3 + 1, u_4) + 2u_3 i_k(s - 2, u_1, u_2, u_3, u_4) + 4(u_4 + 1)i_k(s - 1, u_1, u_2, u_3 - 1, u_4 + 1) \\ & + 4u_4 i_k(s - 1, u_1, u_2, u_3, u_4) + 4(u_4 + 1)i_k(s - 1, u_1, u_2, u_3, u_4 + 1) + 4u_4 i_k(s - 2, u_1, u_2, u_3, u_4) \\ & + 2(u_4 + 1)i_k(s - 2, u_1, u_2, u_3, u_4 + 1) + 2(2(s - 1) - 2u_1 - 4u_2 - 4u_3 - 6u_4)i_k(s - 1, u_1, u_2, u_3, u_4) \\ & + (2(s - 2) - 4u_2 - 4u_3 - 6u_4)i_k(s - 2, u_1, u_2, u_3, u_4) + 2(u_3 + 1)i_k(s, u_1, u_2, u_3 + 1, u_4) \\ & + 4(u_4 + 1)i_k(s, u_1, u_2, u_3 - 1, u_4 + 1) + (2s - 2u_1 - 4u_2 - 4u_3 - 6u_4)i_k(s, u_1, u_2, u_3, u_4) \end{aligned}$$

και

$$2(u_4 + 1)i_k(s + 1, u_1, u_2, u_3, u_4 + 1) = (u_3 + 1)i_k(s, u_1, u_2, u_3 + 1, u_4) + 2(u_2 + 1)i_k(s, u_1, u_2 + 1, u_3, u_4).$$

Πρόταση 2.12: Για $k > 2$, η γεννήτρια συνάρτηση των ανωτέρω V_k σχημάτων δίνεται από τον παρακάτω τύπο:

$$I_k(x, y, z, w, t) = \frac{1 + x}{1 - (y - 2)x + (2w - z - 1)x^2 + (2w - z - 1)x^3} F_k\left(\frac{x(1 + (2w - 1)x + (t - 1)x^2)}{(1 - (y - 2)x + (2w - z - 1)x^2 + (2w - z - 1)x^3)^2}\right).$$

Θεώρημα 2.13: Έστω $k > 2$, τότε

$$Q_k(z) = \frac{1 - z^2 + z^4}{1 - z - z^2 + z^3 + 2z^4 + z^6 - z^8 + z^{10} - z^{12}} F_k\left(\frac{z^4(1 - z^2 - z^4 + 2z^6 - z^8)}{(1 - z - z^2 + z^3 + 2z^4 + z^6 - z^8 + z^{10} - z^{12})^2}\right)$$

Απόδειξη:

Έστω Q_k το σύνολο των αρθρωτών, k -noncrossing δομών, I_k το σύνολο όλων των k -noncrossing V_k σχημάτων και $I_k(s, \mathbf{u})$ το σύνολο των k -noncrossing V_k σχημάτων που έχουν s -τόξα και u_i στοιχεία που ανήκουν στην τάξη C_i , όπου $1 \leq i \leq 4$ (Εικόνα). Τότε έχουμε την επιρριπτική απεικόνιση,

$$\varphi: Q_k \rightarrow I_k,$$

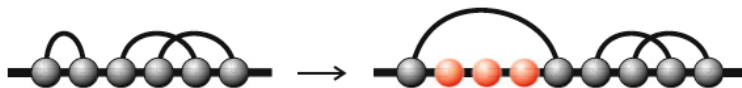
που επάγει τη διαμέριση $Q_k = \bigcup_{\gamma} \varphi^{-1}(\gamma)$. Αυτή η διαμέριση μας επιτρέπει να οργανώσουμε το $Q_k(z)$ με βάση τα V_k σχήματα, γ , ως εξής:

$$Q_k(z) = \sum_{s, \mathbf{u}} \sum_{\gamma \in I_k(s, \mathbf{u})} Q_{\gamma}(z).$$

Στη συνέχεια υπολογίζουμε τη γεννήτρια συνάρτηση $Q_{\gamma}(z)$ με αντίστοιχο τρόπο όπως στο Θεώρημα 2.7. Σημείο κλειδί αποτελεί το γεγονός ότι οι διαδικασίες “φουσκώματος” είναι ειδικές για κάθε τάξη C_i . Θα “φουσκώσουμε” όλα τα “κρίσιμα” τόξα, αυτά δηλαδή που απαιτούν την εισαγωγή απομονωμένων κορυφών για να ικανοποιήσουν τη συνθήκη ελάχιστου μήκους τόξου. Θα αναφερόμαστε σε κάθε stem διαφορετικό από το 2-stack ως \dagger -stem. Αντίστοιχα, η συνδυαστική τάξη των \dagger -stems δίνεται από το $(M - R^2)$, χρησιμοποιώντας την ονοματολογία του Θεωρήματος 2.7.

1) C_1 - τάξη: Εδώ εισάγουμε απομονωμένες κορυφές (Εικόνα 2.12), και έχουμε κατευθείαν:

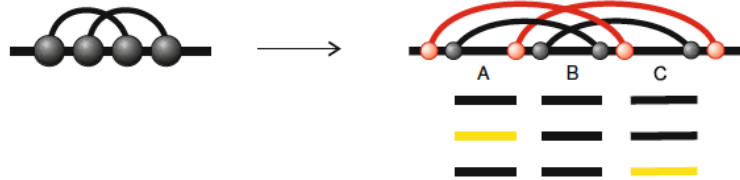
$$C_1(z) = \frac{z^3}{1 - z}.$$



Εικόνα 2.12: Τάξη C_1 : εισαγωγή τουλάχιστον τριών κορυφών (Reidys 2011)

2) C_2 – τάξη: Κάθε τέτοιο στοιχείο είναι ένα ζεύγος $((i, i+2), (i+1, i+3))$ και πρέπει να ξεχωρίσουμε τα ακόλουθα σενάρια:

A) Και τα δύο τόξα “φουσκώνονται” σε stacks μήκους 2. Αφαιρώντας τις περιπτώσεις όπου δεν προστίθεται καμία απομονωμένη κορυφή και τα δύο σενάρια κατά τα οποία δεν υπάρχει εισαγωγή στο διάστημα $[i+1, i+2]$ αλλά μόνο στο $[i, i+1]$ ή στο $[i+2, i+3]$, (Εικόνα 2.13),



Εικόνα 2.13: Τάξη C_2 : “Φούσκωμα” και των δύο τόξων σε 2-stacks. Θέτουμε $A = [i, i+1]$, $B = [i+1, i+2]$ και $C = [i+2, i+3]$ και σημαίνουμε τα σενάρια ενθέσεων που αφαιρούνται ως εξής: η ένθεση απομονωμένων κορυφών σημαίνεται με κίτρινη γραμμή ενώ η μη ένθεση με μαύρη. (Reidys 2011)

καταλήγουμε στο:

$$C_2^{(a)} = R^4 \times [(SEQ(Z))^3 - E - 2(Z \times SEQ(Z))]$$

Η συνδυαστική αυτή τάξη έχει γεννήτρια συνάρτηση

$$C_2^{(a)}(z) = z^8 \left(\left(\frac{1}{1-z} \right)^3 - 1 - \frac{2z}{1-z} \right).$$

B) Ένα τόξο, το $(i+1, i+3)$ ή το $(i, i+2)$, “φουσκώνεται” σε 2-stack ενώ το άλλο σε ένα αυθαίρετο \dagger -stem, Εικόνα 2.14.



Εικόνα 2.14: “Φούσκωμα” ενός τόξου σε 2-stack (Reidys 2011)

Αφαιρώντας τις περιπτώσεις, όπου καμία κορυφή δεν εισάγεται στα $[i+1, i+2]$ και $[i+2, i+3]$ ή στα $[i, i+1]$ και $[i+2, i+3]$, έχουμε:

$$C_2^{(b)} = 2R^2 \times (M - R^2) \times ((SEQ(Z))^2 - E) \times SEQ(Z),$$

με την παρακάτω γεννήτρια συνάρτηση:

$$C_2^{(b)}(z) = 2z^4 \left(\frac{\frac{z^4}{1-z^2}}{1 - \frac{z^4}{1-z^2} \left(\frac{2z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} - z^4 \right) \left(\left(\frac{1}{1-z} \right)^2 - 1 \right) \left(\frac{1}{1-z} \right).$$

Γ) Και τα δύο τόξα “φουσκώνονται” σε ένα αυθαίρετο \dagger -stem, Εικόνα 2.15.



Εικόνα 2.15: Τάξη C_2 : “Φούσκωμα” και των δύο τόξων σε ένα αυθαίρετο \dagger -stem. (Reidys 2011)

Σε αυτή την περίπτωση η εισαγωγή απομονωμένων κορυφών γίνεται αυθαίρετα, οπότε

$$C_2^{(c)} = (M - R^2)^2 \times (\text{SEQ}(Z))^3$$

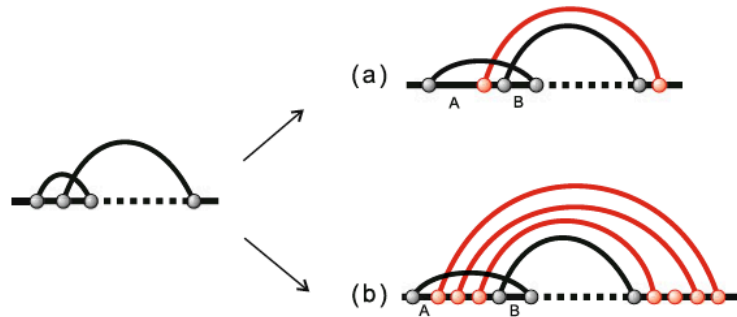
και, αντίστοιχα:

$$C_2^{(c)}(z) = 2z^4 \left(\frac{\frac{z^4}{1-z^2}}{1 - \frac{z^4}{1-z^2} \left(\frac{2z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} - z^4 \right)^2 \left(\frac{1}{1-z} \right)^3.$$

Μιας και τα παραπάνω σενάρια είναι αλληλοαποκλειόμενα, η γεννήτρια συνάρτηση της C_2 - τάξης δίνεται από την

$$C_2(z) = C_2^{(a)}(z) + 2C_2^{(b)}(z) + C_2^{(c)}(z)$$

- 3) C_3 - τάξη:** Αυτή η τάξη αποτελείται από ζεύγη τόξων (α, β) όπου το α είναι το μοναδικό 2-τόξο που διασταυρώνεται με το β και το β έχει μήκος τουλάχιστον 3. Χώρις βλάβη της γενικότητας, μπορούμε να περιορίσουμε την ανάλυση μας στην περίπτωση $((i, i+2), (i+1, j))$, όπου $j > i+3$.



Εικόνα 2.16: Τάξη C_3 : Μόνο ένα τόξο “φουσκώνεται” και το “φούσκωμά” του καθορίζει δύο περιπτώσεις. (Reidys 2011)

Ξεχωρίζουμε και πάλι τις παρακάτω περιπτώσεις:

- A) Το τόξο $(i+1, j)$ “φουσκώνεται” σε 2-stack. Στη συνέχεια, πρέπει εισάγουμε τουλάχιστον μία απομονωμένη κορυφή είτε στο $[i, i+1]$ είτε στο $[i+1, i+2]$, Εικόνα 2.16. Έτσι, έχουμε:

$$C_3^{(a)} = R^2 \times (\text{SEQ}(Z)^2 - E),$$

με γεννήτρια συνάρτηση:

$$C_3^{(a)}(z) = z^4 \left(\left(\frac{1}{1-z} \right)^2 - 1 \right).$$

- B) Το τόξο $(i+1, j)$ “φουσκώνεται” σε ένα αυθαίρετο \dagger -stem, Εικόνα 2.16. Τότε:

$$C_3^{(b)} = (M - R^2) \times \text{SEQ}(Z)^2,$$

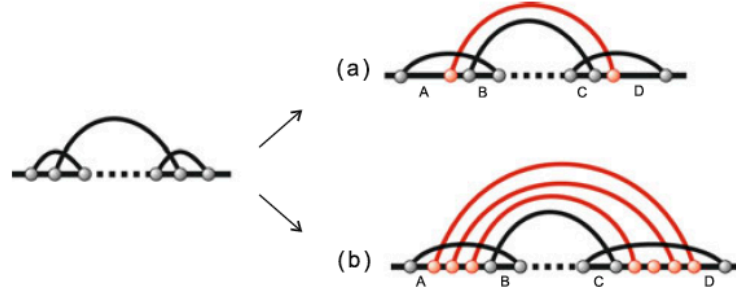
με γεννήτρια συνάρτηση:

$$C_3^{(b)}(z) = \left(\frac{\frac{z^4}{1-z^2}}{1 - \frac{z^4}{1-z^2} \left(\frac{2z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} - z^4 \right) \left(\frac{1}{1-z} \right)^2.$$

Συνεπώς:

$$C_3(z) = C_3^{(a)}(z) + C_3^{(b)}(z)$$

4) **C₄ - τάξη:** Αυτή η τάξη αποτελείται από τριάδες τόξων $(\alpha_1, \beta, \alpha_2)$ όπου τα α_1 και α_2 είναι 2-τόξα που διασταυρώνονται με το β .



Εικόνα 2.17: Τάξη C₄: όπως και στο “φούσκωμα” του C₃, μόνο το μη 2-τόξο “φουσκώνεται”, ξεχωρίζοντας δύο περιπτώσεις (Reidys 2011)

Διακρίνουμε τις παρακάτω περιπτώσεις:

A) Το τόξο β “φουσκώνεται” σε 2-stack, Εικόνα 2.17. Αντίστοιχα, με την τάξη C₃, έχουμε

$$C_4^{(a)} = R^2 \times (\text{SEQ}(Z)^2 - E) \times (\text{SEQ}(Z)^2 - E)$$

με γεννήτρια συνάρτηση:

$$C_4^{(a)}(z) = z^4 \left(\left(\frac{1}{1-z} \right)^2 - 1 \right)^2.$$

B) Το τόξο β “φουσκώνεται” σε ένα αυθαίρετο \dagger -stem, Εικόνα 2.17.

$$C_4^{(b)} = (M - R^2) \times \text{SEQ}(Z)^4,$$

με γεννήτρια συνάρτηση:

$$C_4^{(b)}(z) = \left(\frac{\frac{z^4}{1-z^2}}{1 - \frac{z^4}{1-z^2} \left(\frac{2z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} - z^4 \right) \left(\frac{1}{1-z} \right)^4.$$

Αντίστοιχα, καταλήγουμε στο:

$$C_4(z) = C_4^{(a)}(z) + C_4^{(b)}(z).$$

Το “φούσκωμα” οποιουδήποτε τόξου του γ που δεν υπολογίστηκε στα προηγούμενα βήματα ακολουθεί τη λογική του Θεωρήματος 2.7. Παρατηρούμε ότι δεν υπολογίστηκαν $s - 2u_2 - u_3 - u_4$ τόξα του V_k σχήματος γ . Επίσης, $2s + 1 - u_1 - 3u_2 - 2u_3 - 4u_4$ διαστήματα δεν υπολογίστηκαν για την εισαγωγή απομονωμένων κορυφών. Το “φούσκωμα” των παραπάνω σύμφωνα με τα λεχθέντα στο Θεώρημα 2.7 δίνει τη συνδυαστική τάξη:

$$S = M^{s-2u_2-u_3-u_4} \times (\text{SEQ}(Z))^{2s+1-u_1-3u_2-2u_3-4u_4}$$

που έχει την παρακάτω γεννήτρια συνάρτηση:

$$S(z) = \left(\frac{\frac{z^4}{1-z^2}}{1 - \frac{z^4}{1-z^2} \left(\frac{2z}{1-z} + \left(\frac{z}{1-z} \right)^2 \right)} \right)^{s-2u_2-u_3-u_4} \left(\frac{1}{1-z} \right)^{2s+1-u_1-3u_2-2u_3-4u_4}.$$

Δεδομένου ότι όλες οι διαδικασίες “φουσκώματος” μπορούν να συνδυαστούν ελεύθερα έχουμε:

$$Q_\gamma = C_1^{u_1} \times C_2^{u_2} \times C_3^{u_3} \times C_4^{u_4} \times S$$

οπότε:

$$Q_\gamma(z) = C_1(z)^{u_1} \cdot C_2(z)^{u_2} \cdot C_3(z)^{u_3} \cdot C_4(z)^{u_4} \cdot S(z) = \frac{1}{1-z} \zeta_0(z)^s \zeta_1(z)^{u_1} \zeta_2(z)^{u_2} \zeta_3(z)^{u_3} \zeta_4(z)^{u_4},$$

όπου:

$$\begin{aligned} \zeta_0(z) &= \frac{z^4}{1-2z+2z^3-z^4-2z^5+z^6}, \\ \zeta_1(z) &= z^3, \\ \zeta_2(z) &= z(1-4z^3+2z^4+8z^5-6z^6-7z^7+8z^8+2z^9-4z^{10}+z^{11}), \\ \zeta_3(z) &= z(2-2z+z^3+2z^4-z^5), \\ \zeta_4(z) &= z^2(5-4z-3z^2+6z^3+2z^4-4z^5+z^6). \end{aligned}$$

Παρατηρώντας ότι $\gamma_1, \gamma_2 \in I_k(s, \mathbf{u})$ έχουμε $Q_{\gamma_1}(z) = Q_{\gamma_2}(z)$, βρίσκουμε, σύμφωνα με

την $Q_k(z) = \sum_{s, \mathbf{u}} \sum_{\gamma \in I_k(s, \mathbf{u})} Q_\gamma(z)$, ότι

$$Q_k(z) = \sum_{s, \mathbf{u} \geq 0} i_k(s, \mathbf{u}) Q_\gamma(z)$$

Σύμφωνα με την Πρόταση 2.12:

$$\begin{aligned} \sum_{s, \mathbf{u} \geq 0} i_k(s, \mathbf{u}) x^n y^{u_1} z^{u_2} w^{u_3} t^{u_4} &= \\ &= \frac{1+x}{1-(y-2)x+(2w-z-1)x^2+(2w-z-1)x^3} F_k \left(\frac{x(1+(2w-1)x+(t-1)x^2)}{(1-(y-2)x+(2w-z-1)x^2+(2w-z-1)x^3)^2} \right). \end{aligned}$$

Θέτοντας $x = \zeta_0(z), y = \zeta_1(z), r = \zeta_2(z), w = \zeta_3(z), t = \zeta_4(z)$, έχουμε:

$$Q_k(z) = \frac{1-z^2+z^4}{1-z-z^2+z^3+2z^4+z^6-z^8+z^{10}-z^{12}} F_k \left(\frac{z^4(1-z^2-z^4+2z^6-z^8)}{(1-z-z^2+z^3+2z^4+z^6-z^8+z^{10}-z^{12})^2} \right).$$

Αποδείχτηκε.

Κεφάλαιο 3

Ασυμπτωτική απαρίθμηση

Αυτό είναι το πιο σημαντικό Κεφάλαιο, καθώς έχοντας τις γεννήτριες συναρτήσεις του Κεφαλαίου 2, καθώς και τις μεθόδους του Κεφαλαίου 1, μπορούμε να προχωρήσουμε στις ασυμπτωτικές εκτιμήσεις για τις συναρτήσεις αυτές. Στα πλαίσια του Κεφαλαίου αυτού θα παρουσιαστούν για λόγους πληρότητας και πολλά Θεωρήματα, των οποίων η απόδειξη είναι εκτός στόχων αυτής της εργασίας.

Αρχικά, αφού δώσουμε απαραίτητους ορισμούς και ιδιότητες, θα συζητήσουμε την ασυμπτωτική εκτίμηση της $F_k(z)$, δίνοντας και κάποια στοιχεία που είναι απαραίτητα για τον υπολογισμό των ασυμπτωτικών εκτιμήσεων και των υπόλοιπων γεννητριών συναρτήσεων που εξήχθησαν στο Κεφάλαιο 2.

3.1 D-finiteness και P-recursiveness

Ορισμός 3.1: Μια δυναμοσειρά ονομάζεται **D-finite** (differentially finite – διαφορικά πεπερασμένη) αν ικανοποιεί μια γραμμική ομογενή διαφορική εξίσωση με πολυωνυμικούς συντελεστές. Η σημασία του D-finiteness στην απαρίθμηση προέρχεται από το γεγονός ότι μια γεννήτρια συνάρτηση είναι D-finite αν και μόνο αν οι συντελεστές της είναι P-recursive.

P-recursive (polynomially recursive) ονομάζεται μία ακολουθία αν υπάρχουν πολυώνυμα $p_0(n), \dots, p_m(n) \in C[n]$, με $p_m(n) \neq 0$, τέτοια ώστε για κάθε $n \in \mathbb{N}$:

$$p_m(n)f(n+m) + p_{m-1}(n)f(n+m-1) + \dots + p_0(n)f(n) = 0.$$

Μία δυναμοσειρά ονομάζεται **αλγεβρική** αν υπάρχουν πολυώνυμα $q_0(x), \dots, q_m(x) \in C[x]$, με $q_m(x) \neq 0$, τέτοια ώστε

$$q_m F^m(x) + q_{m-1} F^{m-1}(x) + \dots + q_1 F(x) + q_0(x) = 0.$$

Θεώρημα 3.2 (Stanley 2012): Για τις P-recursive ακολουθίες, τις D-finite και τις αλγεβρικές δυναμοσειρές ισχύουν τα παρακάτω:

- α) Αν οι f, g είναι P-recursive τότε και η $f \cdot g$ είναι P-recursive.
- β) Αν οι F, G είναι D-finite και $\alpha, \beta \in C$, τότε οι $\alpha F + \beta G$ και FG είναι D-finite.
- γ) Αν η F είναι D-finite και η G είναι αλγεβρική με $G(0)=0$, τότε η $F(G(x))$ είναι D-finite.

Απόδειξη:

Θα αποδείξουμε το γ) το οποίο θα μας είναι χρήσιμο και στη συνέχεια. Εφ'όσον $G(0)=0$, η $F(G(x))$ είναι καλά ορισμένη. Έστω $K = F(G(x))$. Τότε η $K^{(i)}$ είναι ένας γραμμικός συνδυασμός των $F(G(x)), F'(G(x)), \dots$, πάνω από το $C[G, G', \dots]$, δηλαδή το δακτύλιο των πολυωνύμων στα G, G', \dots με μιγαδικούς συντελεστές.

Έστω ότι $G^{(i)} \in C, i \geq 0$ οπότε $C[G, G', \dots] \subset C(x, G)$, όπου το $C(x, G)$ υποδηλώνει το σώμα που παράγεται από το x και το G . Δεδομένου ότι το G είναι αλγεβρικό, ικανοποιεί την

$$q_d(x)G^d(x) + q_{d-1}(x)G^{d-1}(x) + \dots + q_1(x)G(x) + q_0(x) = 0,$$

όπου $q_0(x), \dots, q_d(x) \in C[x], q_d(x) \neq 0$ και το d είναι ελάχιστο, δηλαδή το $(G^i(x))_{i=0}^{d-1}$ είναι γραμμικά ανεξάρτητο πάνω από το $C[x]$.

Θέτουμε $P(x, G) = q_d(x)G^d(x) + q_{d-1}(x)G^{d-1}(x) + \dots + q_1(x)G(x) + q_0(x)$.

Παραγωγίζουμε κι έχουμε

$$0 = \frac{d}{dx} P(x, G) = \frac{\partial P(x, y)}{\partial x} \Big|_{y=G} + G' \frac{\partial P(x, y)}{\partial y} \Big|_{y=G}.$$

Ο βαθμός του $\frac{\partial P(x, y)}{\partial y} \Big|_{y=G}$ στο G είναι μικρότερος από $d-1$ και $q_d(x) \neq 0$, οπότε

$\frac{\partial P(x, y)}{\partial y} \Big|_{y=G} \neq 0$. Άρα:

$$G' = - \frac{\frac{\partial P(x, y)}{\partial x} \Big|_{y=G}}{\frac{\partial P(x, y)}{\partial y} \Big|_{y=G}} \in C(x, G).$$

Επαναλαμβάνοντας την παραπάνω διαδικασία, βρίσκουμε ότι $G^{(i)} \in C(x, G), i \geq 0$, οπότε $C[G, G', \dots] \subset C(x, G)$, οπότε ισχύει η υπόθεση.

Έστω V ο διανυσματικός χώρος στο $C(x, G)$ που παράγεται από τα $F(G(x)), F'(G(x)), \dots$. Δεδομένου ότι η F είναι D -finite, ισχύει ότι $\dim_{C(x)} \langle F, F', \dots \rangle < \infty$, υπονοώντας ότι και το $\dim_{C(G)} \langle F(G), F'(G), \dots \rangle$ είναι πεπερασμένο. Έτσι, δεδομένου ότι το $C(G)$ είναι υπόσωμα του $C(x, G)$, έχουμε

$$\dim_{C(x, G)} \langle F, F', \dots \rangle < \infty$$

κι επομένως $\dim_{C(x, G)} V < \infty$ και $\dim_{C(x)} C(x, G) < \infty$. Αυτό συνεπάγεται ότι

$$\dim_{C(x)} V = \dim_{C(x, G)} V \cdot \dim_{C(x)} C(x, G) < \infty$$

κι επειδή κάθε $K^{(i)} \in V$, συμπεραίνουμε ότι η $F(G(x))$ είναι D -finite.

3.2 Η εξίσωση για την $F_k(z)$

Πόρισμα 3.3: Έστω $I_r(2x) = \sum_{j \geq 0} \frac{x^{2j+r}}{j!(r+j)!}$ η υπερβολική συνάρτηση πρώτου

τύπου του Bessel τάξης r . Η γεννήτρια συνάρτηση των k -noncrossing matchings δίνεται από τον παρακάτω τύπο:

$$\sum_{n \geq 0} f_k(2n) \frac{x^{2n}}{(2n)!} = \det [I_{i-j}(2x) - I_{i+j}(2x)]_{i, j=1}^{k-1}.$$

Η κύρια σημασία της παραπάνω πρότασης είναι ότι συνεπάγεται ότι η $H_k(z) = \sum_{n \geq 0} f_k(2n) \frac{z^{2n}}{(2n)!}$ είναι D-finite, το οποίο θα μας φανεί χρήσιμο στη συνέχεια.

Τα Θεωρήματα 3.4 και 3.5 θα δοθούν χωρίς απόδειξη καθώς οι αποδείξεις ξεφεύγουν από το σκοπό αυτής της εργασίας.

Θεώρημα 3.4 (Reidys 2011): Για τυχαίο $k \in \mathbb{N}, k \geq 2, \arg(z) \neq \pm \frac{\pi}{2}$, ισχύει ότι

$$H_k(z) = \left[\prod_{i=1}^{k-1} \Gamma(i+1 - \frac{1}{2}) \prod_{r=1}^{k-2} r! \right] \left(\frac{e^{2\pi}}{\pi} \right)^{k-1} z^{-(k-1)^2 - \frac{k-1}{2}} (1 + O(|z|^{-1})),$$

όπου $\Gamma(z)$ είναι η συνάρτηση γάμμα.

Θεώρημα 3.5 (Reidys 2011): Για τυχαίο $k \in \mathbb{N}, k \geq 2$, έχουμε

$$f_k(2n) \approx c_k n^{-((k-1)^2 + (k-1)/2)} (2(k-1))^{2n}, \text{ όπου } c_k > 0.$$

Πόρισμα 3.6: Η γεννήτρια συνάρτηση των k-noncrossing αντιστοιχιών $2n$ κορυφών, $F_k(z) = \sum_{n \geq 0} f_k(2n) z^n$ είναι D-finite.

Απόδειξη:

Από την Πρόταση 3.3 γνωρίζουμε την εκθετική γεννήτρια συνάρτηση του $f_k(2n)$

$$\sum_{n \geq 0} f_k(2n) \frac{x^{2n}}{(2n)!} = \det[I_{i-j}(2x) - I_{i+j}(2x)] \Big|_{i,j=1}^{k-1}$$

όπου $I_m(x)$ είναι η εξίσωση Bessel πρώτης τάξης. Η εξίσωση Bessel πρώτης τάξης ικανοποιεί την $I_n(x) = i^{-n} J_n(ix)$ και η $J_n(x)$ είναι η λύση της διαφορικής εξίσωσης Bessel

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - n^2)y = 0.$$

Για κάθε $n \in \mathbb{N}$, η $J_n(x)$ είναι D-finite. Έστω $G(x) = ix$. Προφανώς, $G(x) \in C_{alg}[x]$ και $G(0) = 0$, οπότε οι $J_n(ix)$ και $I_n(x)$ είναι D-finite, βάσει του (γ) του Θεωρήματος 3.2. Αναλογα, αποδεικνύεται ότι και το $I_n(2x)$ είναι D-finite για

σταθερό $n \in \mathbb{N}$. Από την $\sum_{n \geq 0} f_k(2n) \frac{x^{2n}}{(2n)!} = \det[I_{i-j}(2x) - I_{i+j}(2x)] \Big|_{i,j=1}^{k-1}$ και το (β) του

Θεωρήματος 3.2, συμπεραίνουμε ότι η $H_k(x) = \sum_{n \geq 0} \frac{f_k(2n)}{(2n)!} x^{2n}$ είναι D-finite. Με

άλλα λόγια η ακολουθία $f(n) = \frac{f_k(2n)}{(2n)!}$ είναι P-recursive και, επιπλέον, η $g(n) = (2n)!$ είναι P-recursive, καθώς $(2n+1)(2n+2)g(n) - g(n+1) = 0$. Συνεπώς, η

$f_k(2n) = f(n)g(n)$ είναι P-recursive, που αποδεικνύει ότι η $F_k(z) = \sum_{n \geq 0} f_k(2n)z^n$ είναι D-finite.

Αφού η $F_k(z)$ είναι D-finite, υπάρχει ένα $e \in \mathbb{N}$, τέτοιο ώστε η $F_k(z)$ να ικανοποιεί μια συνήθη διαφορική εξίσωση της μορφής:

$$q_{0,k}(z) \frac{d^e}{dz^e} F_k(z) + q_{1,k}(z) \frac{d^{e-1}}{dz^{e-1}} F_k(z) + \dots + q_{e,k}(z) F_k(z) = 0,$$

όπου τα $q_{j,k}(z)$ είναι πολυώνυμα. Γνώση της παραπάνω συνήθους διαφορικής εξίσωσης είναι σημαντική για δύο λόγους: α) κάθε επικρατούσα ανωμαλία (dominant singularity) μιας λύσης περιέχεται στις ρίζες του $q_{0,k}(z)$. Με άλλα λόγια, η ΣΔΕ ελέγχει τις επικρατούσες ανωμαλίες που είναι σημαντικές για την ασυμπτωτική αρίθμηση. β) Κάτω από συγκεκριμένες συνθήκες κανονικότητας, η singular expansion της $F_k(z)$ έπεται από τη ΣΔΕ, βλ. Πόρισμα 3.8.

Στη συνέχεια, παραθέτουμε χωρίς απόδειξη (Reidys 2011) τις ΣΔΕ της $F_k(z)$ για $2 \leq k \leq 9$, καθώς και το singular expansion της $F_k(z)$.

Πρόταση 3.7: Για $2 \leq k \leq 9$, η $F_k(z)$ ικανοποιεί τις ΣΔΕ του Πίνακα 2.1 (Reidys 2011) και, συγκεκριμένα, έχουμε:

$$\begin{aligned} q_{0,2} &= (4z - 1)z, \\ q_{0,3} &= (16z - 1)z^2, \\ q_{0,4} &= (144z^2 - 40z + 1)z^3, \\ q_{0,5} &= (1024z^2 - 80z + 1)z^4, \\ q_{0,6} &= (14400z^3 - 4144z^2 + 140z - 1)z^5, \\ q_{0,7} &= (147456z^3 - 12544z^2 + 224z - 1)z^6, \\ q_{0,8} &= (2822400z^4 - 826624z^3 + 31584z^2 - 336z + 1)z^7, \\ q_{0,9} &= (37748736z^4 - 3358720z^3 + 69888z^2 - 480z + 1)z^8. \end{aligned}$$

Οι παραπάνω εξισώσεις καθώς και το Θεώρημα 3.5 δείχνουν ότι για $2 \leq k \leq 9$ η μοναδική επικρατούσα ανωμαλία της $F_k(z)$ δίνεται από το ρ_k^2 , όπου $\rho_k = \frac{1}{2(k-1)}$.

k	
2	$(4x - 1)xf''(x) + (10x - 2)f'(x) + 2f(x) = 0$
3	$(16x^3 - x^2)f^{(3)}(x) + (96x^2 - 8x)f''(x) + (108x - 12)f'(x) + 12f(x) = 0$
4	$(144x^5 - 40x^4 + x^3)f^{(4)}(x) + (1584x^4 - 556x^3 + 20x^2)f^{(3)}(x)$ $+ (4428x^3 - 1968x^2 + 112x)f''(x) + (3024x^2 - 1728x + 168)f'(x)$ $+ (216x - 168)f(x) = 0$
5	$(1024x^6 - 80x^5 + x^4)f^{(5)}(x) + (20,480x^5 - 2256x^4 + 40x^3)f^{(4)}(x)$ $+ (121,600x^4 - 19,380x^3 + 532x^2)f^{(3)}(x) + (241,920x^3 - 56,692x^2 + 2728x)$ $f''(x) + (130,560x^2 - 46,048x + 4400)f'(x) + (7680x - 4400)f(x) = 0$
6	$(14,400x^8 - 4144x^7 + 140x^6 - x^5)f^{(6)}(x)$ $+ (367,200x^7 - 148,368x^6 + 7126x^5 - 70x^4)f^{(5)}(x)$ $+ (3,078,000x^6 - 1,728,900x^5 + 123,850x^4 - 1792x^3)f^{(4)}(x)$ $+ (10,179,000x^5 - 7,880,640x^4 + 880,152x^3 - 20,704x^2)f^{(3)}(x)$ $+ (12,555,000x^4 - 13,367,880x^3 + 2,399,184x^2 - 106,016x)f''(x)$ $+ (4,374,000x^3 - 6,475,680x^2 + 1,922,736x - 187,200)f'(x)$ $+ (162,000x^2 - 350,640x + 187,200)f(x) = 0$
7	$(147,456x^9 - 12,544x^8 + 224x^7 - x^6)f^{(7)}(x)$ $+ (6,193,152x^8 - 757,760x^7 + 18,816x^6 - 112x^5)f^{(6)}(x)$ $+ (89,800,704x^7 - 16,035,456x^6 + 582,280x^5 - 4872x^4)f^{(5)}(x)$ $+ (561,254,400x^6 - 146,691,840x^5 + 8,254,664x^4 - 104,480x^3)f^{(4)}(x)$ $+ (1,535,708,160x^5 - 585,419,280x^4 + 54,069,792x^3 - 1,151,984x^2)f^{(3)}(x)$ $+ (1,651,829,760x^4 - 916,833,600x^3 + 144,777,216x^2 - 6,094,528x)f''(x)$ $+ (516,741,120x^3 - 421,901,280x^2 + 117,590,208x - 11,797,632)f'(x)$ $+ (17,418,240x^2 - 22,034,880x + 11,797,632)f(x) = 0$
8	$(2,822,400x^{11} - 826,624x^{10} + 31,584x^9 - 336x^8 + x^7)f^{(8)}(x)$ $+ (129,830,400x^{10} - 55,968,384x^9 + 3,026,208x^8 - 43,512x^7 + 168x^6)f^{(7)}(x)$ $+ (2,202,883,200x^9 - 1,363,532,352x^8 + 107,691,912x^7 - 2,188,752x^6$ $+ 11,424x^5)$ $f^{(6)}(x) + (17455132800x^8 - 15,140,260,128x^7 + 1,789,953,376x^6$ $- 54349,728x^5 + 405,200x^4)f^{(5)}(x)$ $+ (67,586,778,000x^7 - 80,551,356,480x^6 + 14,421,855,200x^5$ $- 698,609,104x^4 + 8,035,104x^3)f^{(4)}(x)$ $+ (122,393,376,000x^6 - 197,784,236,160x^5 + 53,661,386,080x^4$ $- 4437573,920x^3 + 88,180,864x^2)f^{(3)}(x)$ $- 4437573,920x^3 + 88,180,864x^2)f^{(3)}(x)$ $+ (90,239,184,000x^5 - 196,676,000,640x^4 + 80,758,975,680x^3$ $- 11,973,419,104x^2 + 488,846,272x)f''(x)$ $+ (19,559,232,000x^4 - 57,892,907,520x^3 + 35,467,753,520x^2$ $- 9,969,500,032x + 1,033,305,728)f'(x)$ $+ (444,528,000x^3 - 1,852,865,280x^2 + 186,993,760x - 1,033,305,728)f(x) = 0$
9	$(37,748,736x^{12} - 3,358,720x^{11} + 69,888x^{10} - 480x^9 + x^8)f^{(9)}(x)$ $+ (2,717,908,992x^{11} - 351,387,648x^{10} + 10,065,408x^9 - 90,912x^8$ $+ 240x^7)f^{(8)}(x)$ $+ (72,873,934,848x^{10} - 1,378,440,8064x^9$ $+ 563,449,728x^8 - 6,950,616x^7 + 24,024x^6)f^{(7)}(x)$ $+ (940,566,380,544x^9 - 258,478,202,880x^8 + 15,638,941,312x^7$ $- 2,368,505,160x^6 + 1,304,336x^5)f^{(6)}(x)$ $+ (6,273,464,795,136x^8 - 2,467,959,432,192x^7 + 227,994,061,392x^6$ $- 18,674,432,128x^5 + 41,782,224x^4)f^{(5)}(x)$ $+ (21,523,928,186,880x^7 - 119,317,461,350,40x^6 + 17,131,29,509,184x^5$ $- 75,115,763,872x^4 + 802,970,368x^3)f^{(4)}(x)$ $+ (35,583,374,131,200x^6 - 27,454,499,6659,20x^5 + 614,7724,228,704x^4$ $- 475,182,777,504x^3 + 8,956,331,968x^2)f^{(3)}(x)$ $+ (24,400,027,975,680x^5 - 26,056,335,882,240x^4 + 9,086,553,292,608x^3$ $- 1,308,864,283,488x^2 + 52,313,960,192x)f''(x)$ $+ (4,976,321,495,040x^4 - 740,2528,051,200x^3 + 4,051,342,551,744x^2$ $- 1,122,348,764,928x + 120,086,385,408)f'(x)$ $+ (107,017,666,560x^3 - 230,051,819,520x^2 + 208,033,076,736x - 120,$ $086,385,408)f(x) = 0$

Πίνακας 2.1: Οι διαφορικές εξισώσεις για την $\mathbf{F}_k(z)$, από το Maple package gfun. (Reidys 2011)

Πόρισμα 3.8: Για $2 \leq k \leq 9$, η singular expansion της $F_k(z)$ για $z \rightarrow \rho_k^2$ δίνεται από τον τύπο:

$$F_k(z) = P_k(z - \rho_k^2) + c'_k (z - \rho_k^2)^{((k-1)^2 + \frac{k-1}{2})^{-1}} \log(z - \rho_k^2)(1 + o(1)), \text{ αν το } k \text{ είναι περιττό,}$$

$$F_k(z) = P_k(z - \rho_k^2) + c'_k (z - \rho_k^2)^{((k-1)^2 + \frac{k-1}{2})^{-1}} (1 + o(1)), \text{ αν το } k \text{ είναι άρτιο.}$$

Επιπλέον, οι όροι $P_k(z)$ είναι πολυώνυμα βαθμού όχι μεγαλύτερου από $(k-1)^2 + (k-1)/2 - 1$, το c'_k είναι μία σταθερά και $\rho_k = \frac{1}{2(k-1)}$.

3.3 Ασυμπτωτική αρίθμηση άλλων δομών

Τώρα που έχουμε τα απαραίτητα εργαλεία, καθώς και τις γεννήτριες συναρτήσεις του Κεφαλαίου 2, μπορούμε να προχωρήσουμε στην ασυμπτωτική αρίθμηση κάποιων από τις δομές που είδαμε στο Κεφάλαιο 2. Θα καταλήξουμε, όπως και στο προηγούμενο Κεφάλαιο, στην ασυμπτωτική αρίθμηση των k -noncrossing canonical RNA ακολουθιών με ελάχιστο μήκος τόξου 4.

Όπως είδαμε στο Θεώρημα 2.5, η γεννήτρια συνάρτηση $I_k(s, m)$ των k -noncrossing διαγραμμμάτων, μήκους $2s$ και με m 1-τόξα, δίνεται από τον τύπο:

$$I_k(z, u) = \frac{1+z}{1+2z-zu} F_k\left(\frac{z(1+z)}{(1+2z-zu)^2}\right).$$

Θεώρημα 3.9 (Reidys and Wang 2010): Για $2 \leq k \leq 9$, ο αριθμός των V_k σχημάτων μήκους $2s$ δίνεται ασυμπτωτικά από τον τύπο:

$$i_k(s) \approx c_k s^{-((k-1)^2 + (k-1)/2)} (\mu_k^{-1})^s,$$

όπου μ_k είναι η μοναδική θετική πραγματική λύση του $\frac{z}{1+z} = \rho_k^2$ και c_k είναι κάποια θετική σταθερά.

Απόδειξη:

Το Θεώρημα 3.9 είναι άμεση συνέπεια του Θεωρήματος 1.9. Η γεννήτρια συνάρτηση $F_k(z) = \sum_{n \geq 0} f_k(2n)z^n$ είναι D-finite (Πόρισμα 3.6) και η εσωτερική συνάρτηση $\theta(z) = \frac{z}{1+z}$ είναι αλγεβρική, ικανοποιεί το $\theta(0) = 0$ και είναι αναλυτική για $|z| < 1$. Χρησιμοποιώντας το γεγονός ότι όλες οι ανωμαλίες της $F_k(z)$ περιλαμβάνονται στις ρίζες του $q_{0,k}(z)$ (Πρόταση 3.7), μπορούμε να επιβεβαιώσουμε ότι η $F_k(\theta(z))$ έχει τη μοναδική επικρατούσα πραγματική ανωμαλία $\mu_k < 1$, που ικανοποιεί τη $\theta(\mu_k) = \rho_k^2$ για $2 \leq k \leq 9$. Εφ'όσον $\theta'(\mu_k) \neq 0$, το Θεώρημα 1.9 εγγυάται ότι:

$$i_k(s) \approx c_k s^{-((k-1)^2 + (k-1)/2)} (\mu_k^{-1})^s.$$

Αποδείχτηκε.

Από το Θεώρημα 2.6, γνωρίζουμε ότι η γεννήτρια συνάρτηση των πυρηνών δίνεται από τον τύπο:

$$C_k(z) = \frac{1}{r(z)z^2 - z + 1} F_k\left(\left(\frac{\sqrt{r(z)}z}{r(z)z^2 - z + 1}\right)^2\right), \text{ με } r(z) = \frac{1}{1+z^2}.$$

Θεώρημα 3.10: Για $k \in \mathbb{N}, k \geq 2$, έχουμε:

$$C_k(n) \approx c_k n^{-((k-1)^2 + (k-1)/2)} \left(\frac{1}{\kappa_k}\right)^n, k = 3, 4, \dots, 9,$$

όπου κ_k είναι η μοναδική θετική πραγματική επικρατούσα ανωμαλία της $C_k(z)$ και η ελάχιστη θετική πραγματική λύση της εξίσωσης $\left(\frac{\sqrt{r(x)}x}{r(x)x^2 - x + 1}\right)^2 = \rho_k^2$, για $k = 3, 4, \dots, 9$.

Απόδειξη:

Το Θεώρημα του Prigsheim (Titchmarsh 1939) εγγυάται ότι η $C_k(z)$ έχει μια πραγματική θετική επικρατούσα ανωμαλία κ_k . Επιβεβαιώνουμε ότι υπάρχει μοναδική λύση του $w(x) = \left(\frac{\sqrt{r(x)}x}{r(x)x^2 - x + 1}\right)^2 = \rho_k^2$, για $k = 3, 4, \dots, 9$. Αυτή η λύση ισούται με το κ_k , τη, συνεπώς μοναδική, επικρατούσα πραγματική ανωμαλία της $C_k(z)$. Επίσης, δεδομένου ότι το κ_k είναι αυστηρά μικρότερο από την ανωμαλία του $w(x)$ και $w'(\kappa_k) \neq 0$, μπορούμε να εφαρμόσουμε το Θεώρημα 1.9 στη σύνθετη συνάρτηση $F_k(w(x))$. Δεδομένου, ακόμη, ότι το κ_k είναι μικρότερο από τις ανωμαλίες του παράγοντα $\frac{1}{r(x)x^2 - x + 1}$ και του ότι το $w(x)$ είναι αλγεβρικό, χρησιμοποιώντας τις εξισώσεις της Πρότασης 3.7, μπορούμε να επιβεβαιώσουμε ότι το Θεώρημα 1.9 ισχύει και για $k = 3, 4, \dots, 9$. Έτσι, έχουμε:

$$C_k(n) \approx c_k n^{-((k-1)^2 + (k-1)/2)} \left(\frac{1}{\kappa_k}\right)^n, \text{ για κάποιο } c_k > 0.$$

Αποδείχτηκε.

Στο Θεώρημα 2.7 δείξαμε ότι η γεννήτρια συνάρτηση των k -noncrossing σ -canonical δομών ισούται με:

$$T_{k,\sigma}(z) = \frac{1}{u_\sigma(z)z^2 - z + 1} F_k\left(\left(\frac{\sqrt{u_\sigma(z)}z}{(u_\sigma(z)z^2 - z + 1)}\right)^2\right), \text{ με } u_\sigma(z) = \frac{(z^2)^{\sigma+1}}{z^{2\sigma} - z^2 + 1}.$$

Θεώρημα 3.11: Έστω $k, \sigma \in \mathbb{N}, k \geq 2$, $u_\sigma(x) = \frac{(x^2)^{\sigma+1}}{x^{2\sigma} - x^2 + 1}$ και ρ_k^2 η επικρατούσα θετική πραγματική ανωμαλία της $F_k(z)$. Τότε:

$$T_{k,\sigma}(n) \approx c_{k,\sigma} n^{-((k-1)^2 + (k-1)/2)} \left(\frac{1}{\gamma_{k,\sigma}}\right)^n, 2 \leq k \leq 9, 1 \leq \sigma \leq 9,$$

όπου $\gamma_{k,\sigma}$ είναι η επικρατούσα πραγματική ανωμαλία του $T_{k,\sigma}(x)$ και η ελάχιστη θετική πραγματική λύση της εξίσωσης:

$$q_{\sigma}(x) = \left(\frac{\sqrt{u_{\sigma}(x)}x}{u_{\sigma}(x)x^2 - x + 1} \right)^2 = \rho_k^2, \text{ βλ. Πίνακα 2.2}$$

k	2	3	4	5	6	7	8	9
$\gamma_{k,1}$	2.6180	4.7913	6.8541	8.8875	10.9083	12.9226	14.9330	16.9410
$\gamma_{k,2}$	1.9680	2.5881	3.0382	3.4138	3.7438	4.0420	4.3162	4.5715
$\gamma_{k,3}$	1.7160	2.0477	2.2704	2.4466	2.5955	2.7259	2.8427	2.9490
$\gamma_{k,4}$	1.5782	1.7984	1.9410	2.0511	2.1423	2.2209	2.2904	2.3529
$\gamma_{k,5}$	1.4899	1.6528	1.7561	1.8347	1.8991	1.9540	2.0022	2.0454
$\gamma_{k,6}$	1.4278	1.5563	1.6368	1.6973	1.7466	1.7883	1.8248	1.8573
$\gamma_{k,7}$	1.3815	1.4872	1.5528	1.6019	1.6415	1.6750	1.7041	1.7300
$\gamma_{k,8}$	1.3454	1.4351	1.4903	1.5314	1.5645	1.5923	1.6165	1.6378
$\gamma_{k,9}$	1.3164	1.3941	1.4417	1.4770	1.5054	1.5291	1.5497	1.5679

Πίνακας 2.2

Απόδειξη:

Επιβεβαιώνουμε ότι όλες οι επικρατούσες ανωμαλίες του $T_{k,\sigma}(x)$ είναι και ανωμαλίες του $F_k(\frac{\sqrt{u_{\sigma}(x)}x}{u_{\sigma}(x)x^2 - x + 1})$ και ότι το $\gamma_{k,\sigma}$ είναι η μοναδική επικρατής ανωμαλία και για τις δύο συναρτήσεις. Ανάλογα με το Θεώρημα 3.10, συμπεραίνουμε μέσω και του Θεωρήματος 1.9, ότι, πράγματι:

$$T_{k,\sigma}(n) \approx c_{k,\sigma} n^{-((k-1)^2 + (k-1)/2)} \left(\frac{1}{\gamma_{k,\sigma}} \right)^n.$$

Αποδείχτηκε.

Όσον αφορά τις αρθρωτές noncrossing δομές, των οποίων η γεννήτρια συνάρτηση δίνεται στην Πρόταση 2.8 και είναι η παρακάτω:

$$Q_2(z) = \frac{1 - z^2 + z^4}{1 - z - z^2 + z^3 + 2z^4 + z^6} \cdot F_2\left(\frac{z^4 - z^6 + z^8}{(1 - z - z^2 + z^3 + 2z^4 + z^6)^2}\right)$$

έχουμε τον εξής τύπο για την ασυμπτωτική της αριθμηση.

Πρόταση 3.12: Ισχύει:

$$Q_2(n) \approx c_2 n^{-3/2} \gamma_2^{-n},$$

όπου γ_2 είναι η ελάχιστη θετική πραγματική λύση του $\theta(z) = 1/4$ και

$$\theta(z) = \frac{z^4 - z^6 + z^8}{(1 - z - z^2 + z^3 + 2z^4 + z^6)^2}.$$

Επίσης, έχουμε $\gamma_2 \approx 1.8489$ και $c_2 \approx 1.4848$.

Απόδειξη:

Η απόδειξη είναι απλή εφαρμογή του Θεωρήματος 1.9.

Τέλος, η γεννήτρια συνάρτηση για τις k -noncrossing canonical RNA δομές με ελάχιστο μήκος τόξου 4 είναι (Θεώρημα 2.13):

$$Q_k(z) = \frac{1 - z^2 + z^4}{1 - z - z^2 + z^3 + 2z^4 + z^6 - z^8 + z^{10} - z^{12}} F_k\left(\frac{z^4(1 - z^2 - z^4 + 2z^6 - z^8)}{(1 - z - z^2 + z^3 + 2z^4 + z^6 - z^8 + z^{10} - z^{12})^2}\right)$$

ενώ η ασυμπτωτική αρίθμηση δίνεται από το παρακάτω θεώρημα:

Θεώρημα 3.13: Για $3 \leq k \leq 9$

$$Q_k(n) \approx c_k n^{-((k-1)^2 + (k-1)/2)} \gamma_k^{-n}, c_k > 0$$

όπου γ_k είναι η ελάχιστη θετική πραγματική λύση του $\theta(z) = \rho_k^2$, βλ. Πίνακα 2.3.

k	3	4	5	6	7	8	9
$\theta(n)$	n^{-5}	$n^{-\frac{21}{2}}$	n^{-18}	$n^{-\frac{55}{2}}$	n^{-39}	$n^{-\frac{105}{2}}$	n^{-68}
γ_k^{-1}	2.5410	3.0132	3.3974	3.7319	4.0327	4.3087	4.5654

Πίνακας 2.3: Οι εκθετικοί παράγοντες γ_k και τα $\theta(n)$ για τις αρθρωτές, k -noncrossing δομές.

Απόδειξη:

Από το Θεώρημα 3.6, η $Q_k(z)$ είναι D-finite. Το Θεώρημα του Pringsheim (Titchmarsh 1939) εγγυάται ότι η $Q_k(z)$ έχει μια επικρατούσα θετική πραγματική ανωμαλία γ_k . Επιβεβαιώνουμε ότι για $3 \leq k \leq 9$, το γ_k είναι η μοναδική λύση της εξίσωσης $\theta(z) = \rho_k^2$. Σύμφωνα με τον Πίνακα 2.4,

k	$\vartheta(z) = \rho_k^2$	$ z $	k	$\vartheta(z) = \rho_k^2$	$ z $
3	0.3935	0.3935	7	0.2480	0.2480
	$0.1979 \pm 0.4983i$	0.5361		$0.0468 \pm 0.2928i$	0.2965
	$0.1979 \pm 0.4986i$	0.7309		-0.3274	0.3274
	Other solutions	≥ 0.8762		Other solutions	≥ 0.8684
4	0.3319	0.3319	8	0.2680	0.2680
	$0.1116 \pm 0.4181i$	0.4327		$0.0393 \pm 0.2705i$	0.3691
	-0.4984	0.4984		- - 0.3003	-0.3003
	Other solutions	≥ 0.8734		Other solutions	≥ 0.8684
5	0.2943	0.2943	9	0.2190	0.2190
	$0.0763 \pm 0.3611i$	0.3691		$0.0339 \pm 0.2526i$	0.2548
	-0.4144	0.4144		-0.2789	0.2789
	Other solutions	≥ 0.8693		Other solutions	≥ 0.8685
6	0.2680	0.2680			
	$0.0580 \pm 0.3218i$	0.3269			
	-0.3633	0.3633			
	Other solutions	≥ 0.8685			

Πίνακας 2.4

η λύση αυτή είναι αυστηρά μικρότερη από την ανωμαλία του $\theta(z)$ και ο Πίνακας 2.5 δείχνει ότι $\theta'(z) \neq 0$.

k	3	4	5	6	7	8	9
$\vartheta'(\gamma_k)$	0.739849	0.402041	0.253552	0.176398	0.131015	0.101895	0.081989

Πίνακας 2.5

Συνεπώς, η σύνθετη συνάρτηση:

$$F_k\left(\frac{z^4(1-z^2-z^4+2z^6-z^8)}{q(z)^2}\right)$$

έγκειται στο Θεώρημα 1.9 κι έτσι έχουμε:

$$Q_k(n) \approx c_k n^{-((k-1)^2+(k-1)/2)} \gamma_k^{-n}, c_k > 0.$$

Αποδείχτηκε.

Κεφάλαιο 4

Στοιχεία σχετικά με την εξέλιξη των RNA ψευδοκόμπων

Τα αποτελέσματα που παρουσιάστηκαν στα προηγούμενα κεφάλαια, εκτός από το θεωρητικό ενδιαφέρον που παρουσιάζουν σχετικά με τη χρήση των γεννητριών συναρτήσεων για τη μελέτη βιολογικών μορίων, μπορούν να μας δώσουν πληροφορίες σχετικά με την εξέλιξη των RNA ψευδοκόμπων. Οι RNA ψευδοκόμποι φαίνεται να έχουν ιδιαίτερο εξελικτικό ενδιαφέρον, καθώς αποτελούν κατάλοιπο του RNA κόσμου που προηγήθηκε του DNA κόσμου, όπως τον γνωρίζουμε τώρα. Τα μόρια που φέρουν τέτοιες δομές (RNaseP, τελομεράση, ιντρόνια τάξης I) είναι εξαιρετικά συντηρημένα, και οι RNA ψευδοκόμποι είναι απαραίτητοι για την καταλυτική τους δράση.

Πριν δούμε τις συνέπειες των αποτελεσμάτων των Κεφαλαίων 2 και 3 στην εξελικτικότητα των δομών αυτών, ας δούμε πρώτα τι είναι η ευρωστία (robustness) και εξελικτικότητα (evolvability) των RNA δομών και από τι καθορίζεται. Τέλος, θα κλείσουμε συζητώντας τη σημασία αυτών των παρατηρήσεων.

4.1 Ευρωστία και εξελικτικότητα RNA δομών

Οι πιο διεξοδικές αναλύσεις σχετικά με την ευρωστία βιολογικών μορίων (ανοχή σε μεταλλαγές) αφορούν τη δομή του RNA. Οι αναλύσεις αυτές υποδεικνύουν όχι μόνο την ανοχή των RNA δομών σε μεταλλαγές (ευρωστία) αλλά παρουσιάζουν και την ύπαρξη ενός μεγάλου ουδέτερου δικτύου μέσα στο οποίο η ευρωστία των RNA δομών εξελίσσεται. Επίσης, δείχνουν πώς οι ουδέτερες μεταλλαγές μπορούν να είναι σημαντικές για την εξελικτικότητα των RNA δομών – την εμφάνιση δηλαδή εξελικτικών νεωτερισμών (Wagner 2005).

Η σωστή δευτεροταγής δομή είναι πολύ σημαντική για τη λειτουργία πολλών RNA μορίων. Υπάρχουν πολλά τέτοια παραδείγματα. Είναι χαρακτηριστικό, επίσης, ότι πολλά από αυτά τα μόρια σχηματίζουν RNA ψευδοκόμπους. RNA μόρια, λοιπόν, των οποίων η δευτεροταγής δομή συνδέεται άμεσα με τη λειτουργία είναι: η εσωτερική θέση έναρξης της μετάφρασης (Internal Ribosome Entry Site - IRES) των picornaviruses, η 3' μη μεταφραζόμενη περιοχή των flavivirus, τα rRNAs των ριβοσωμάτων, η RNaseP, η 3' μη μεταφραζόμενη περιοχή πολλών mRNAs, το RNA τμήμα της τελομεράσης κοκ (Dayton, Konings et al. 1992; Jackson and Kaminski 1995; Powell, Zhang et al. 1995; Mandl, Holzmann et al. 1998).

Για τη διατήρηση της δευτεροταγούς τους δομής, τα RNA μόρια χρησιμοποιούν δύο τρόπους: πρώτον, τη συντήρηση της πρωτοταγούς τους δομής, ή την ανοχή σε μεταλλαγές που διατηρούν τη δευτεροταγή δομή σταθερή. Παραδείγματος χάριν, αν ένα είδος έχει στη δομή ενός RNA μορίου του δύο βάσεις G-C που

σχηματίζουν δεσμούς υδρογόνου μεταξύ τους, τότε αν το ίδιο μόριο σε άλλο είδος έχει A-T σ' αυτή τη θέση, τα οποία επίσης μπορούν να σχηματίζουν δεσμούς υδρογόνου, η δευτεροταγής δομή μπορεί να διατηρηθεί अपαράλλακτη. Τέλος, αξίζει να αναφερθεί ότι τόσο λειτουργικές όσο και εξελικτικές μελέτες υποδεικνύουν ότι η δευτεροταγής δομή είναι απαραίτητη αν και όχι επαρκής για τη λειτουργία του RNA (James, Olsen et al. 1988; James, Olsen et al. 1989; Muse 1995; Parsch, Braverman et al. 2000). Συνεπώς, ο σχηματισμός της δομής του RNA και η ευρωστία της δομής αυτής απαιτούν περαιτέρω ανάλυση.

Για να καταλάβουμε πώς αλληλεπιδρούν η ευρωστία και η εξελικτικότητα των RNA δομών και πώς επηρεάζονται από τις μεταλλαγές χρειάζεται να συζητήσουμε διεξοδικότερα τα ουδέτερα δίκτυα. Όπως είδαμε και στο εισαγωγικό κεφάλαιο, ουδέτερα δίκτυα είναι τα σύνολα των πρωτοταγών δομών που καταλήγουν στην ίδια δευτεροταγή δομή. Συνεπώς, αν έχουμε μια αλληλουχία μπορούμε να εισάγουμε σημειακές μεταλλαγές περνώντας από αλληλουχία σε αλληλουχία, χωρίς, επι της ουσίας, να αλλάζουμε δομή. Ακραίο παράδειγμα εισαγωγής σημειακών μεταλλαγών που δεν αλλάζουν τη λειτουργική δευτεροταγή δομή αποτελούν τα ιντρόνια τάξης I. Σ' αυτά ενά πολύ μικρό ποσοστό βάσεων παραμένει συντηρημένο (μόλις επτά). Παρολ' αυτά, η δευτεροταγής τους δομή που είναι υπεύθυνη και για την καταλυτική τους δράση (αυτομάτισμα) είναι πανομοιότυπη (Lisacek, Diaz et al. 1994).

Επομένως, όταν κατά την εξέλιξη, σχηματίζεται μια αλληλουχία, η οποία έχει μια συγκεκριμένη λειτουργική δομή, μπορούμε να φανταστούμε ότι καταλαμβάνει ένα συγκεκριμένο σημείο μέσα στο χώρο των αλληλουχιών (sequence space). Αν, τώρα, η δομή αυτή είναι απαραίτητη για τη λειτουργία της και η λειτουργία, με τη σειρά της, επιλέγεται θετικά από τη φυσική επιλογή, είναι προφανές ότι οι μόνες μεταλλαγές που θα περνούν στην επόμενη γενιά είναι αυτές που δεν επηρεάζουν τη δευτεροταγή δομή. Μπορούμε, λοιπόν, στη συνέχεια, να φανταστούμε, ότι οι διαφορετικές αλληλουχίες που εμφανίζονται θα καταλαμβάνουν διαφορετικά σημεία στο χώρο των αλληλουχιών, τα οποία όμως, θα παραμένουν στο ουδέτερο δίκτυο της αρχικής αλληλουχίας. Έχει δείξει, ότι αλληλουχίες με την ίδια δομή σχηματίζουν νησίδες μέσα στο χώρο των αλληλουχιών (Bornberg-Bauer 1997; van Nimwegen, Crutchfield et al. 1999). Ανάμεσα στις νησίδες αυτές, υπάρχουν συχνά γέφυρες που ανήκουν στο ίδιο ουδέτερο δίκτυο. Μετά από πολλές ουδέτερες μεταλλαγές, η εικόνα της εξέλιξης της αρχικής αλληλουχίας μέσα στο χώρο των αλληλουχιών, θα αποτελείται ουσιαστικά από πυκνές περιοχές (νησίδες), ενωμένες με λεπτές γέφυρες.

Γνωρίζοντας, τώρα, πώς λειτουργούν οι ουδέτερες μεταλλαγές και τα ουδέτερα δίκτυα στην εξέλιξη μιας αλληλουχίας, ας δούμε τι επιπτώσεις έχουν οι μεταλλαγές αυτές στην ευρωστία και την εξελικτικότητα μιας RNA αλληλουχίας και πώς σχετίζονται οι δύο αυτοί όροι μεταξύ τους. Είναι ξεκάθαρο, πλέον, ότι οι ουδέτερες μεταλλαγές συνεισφέρουν στην ευρωστία μας αλληλουχίας, όσο περισσότερες είναι οι πιθανές ουδέτερες μεταλλαγές τόσο πιο εύρωστη είναι μια αλληλουχία. Όμως, έχει φανεί ότι οι ουδέτερες μεταλλαγές συνεισφέρουν με έμμεσο τρόπο και στην εξελικτικότητα της αλληλουχίας. Ας επανέλθουμε στο προηγούμενο παράδειγμα της αλληλουχίας μέσα στο χώρο των αλληλουχιών. Πριν την εισαγωγή ουδέτερων μεταλλάγων, η αλληλουχία περιβαλλόταν από

έναν αριθμό αλληλουχιών, κάποιες από τις οποίες ήταν ουδέτερες και κάποιες όχι. Δεδομένου, όμως, του ότι η αλληλουχία αυτή καταλάμβανε ένα μόνο σημείο, ο αριθμός των γειτονικών αλληλουχιών ήταν περιορισμένος. Μετά την εισαγωγή των ουδέτερων μεταλλαγών, οι αλληλουχίες (που διατηρούν την αρχική δευτεροταγή δομή και λειτουργία) γειτνιάζουν με πολλές διαφορετικές αλληλουχίες. Αν υπάρχει, λοιπόν, μια αλληλουχία στο χώρο των αλληλουχιών, που μπορεί να εισάγει ένα νεωτερισμό (είτε να βελτιώσει τη λειτουργία της υπάρχουσας δομής είτε να εισάγει μια νέα λειτουργία), είναι προφανές ότι οι ουδέτερες μεταλλαγές τη φέρνουν πιο κοντά. Επίσης, ως συνέπεια, όσο πιο εύρωστη είναι μια αλληλουχία, τόσο περισσότερες αλληλουχίες (διαφορετικού ουδέτερου δικτύου) βρίσκονται κοντά στο ουδέτερο δίκτυο της αρχικής.

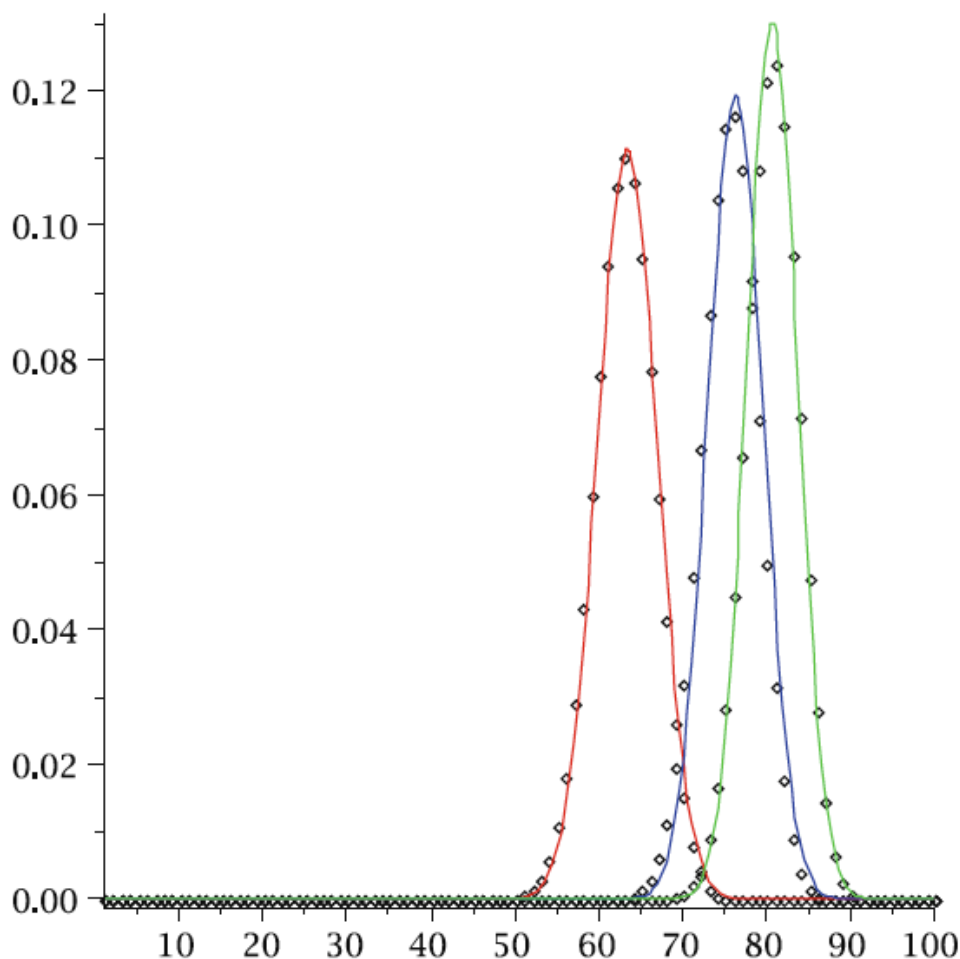
4.2 Τα ουδέτερα δίκτυα των RNA ψευδοκόμπων

Από τα αποτελέσματα του Κεφαλαίου 3, μπορούμε να δείξουμε ότι οι RNA ψευδοκόμποι διαθέτουν εκθετικό αριθμό διακριτών δομών ψευδοκόμπων που έχουν εκθετικά μεγάλα ουδέτερα δίκτυα.

Η παρουσία ουδέτερων δικτύων φαίνεται από τον εκθετικό δείκτη της ασυμπτωτικής εκτίμησης των k -noncrossing canonical RNA ψευδοκόμπων με ελάχιστο μήκος τόξου 4. Αρχικά, να αναφέρουμε ότι οι δομές ψευδοκόμπων που παρουσιάζουν βιολογικό ενδιαφέρον είναι οι 3-noncrossing canonical RNA δομές, καθώς αυτές έχουν βρεθεί στη φύση, ως τώρα. Ο εκθετικός δείκτης αυτών γ_3^{-1} ισούται με 2,541 (Θεώρημα 3.13), ο οποίος είναι πολύ μικρότερος από τον εκθετικό δείκτη των RNA αλληλουχιών που είναι 4 (ο αριθμός των διαφορετικών νουκλεοτιδίων). Αυτό υποδεικνύει την παρουσία ουδέτερων δικτύων, καθώς ο λόγος αλληλουχιών προς 3-noncrossing canonical RNA δομές ψευδοκόμπων αυξάνει εκθετικά.

Για να δείξουμε την ύπαρξη εκθετικά πολλών ψευδοκόμπων με μεγάλα ουδέτερα δίκτυα, πρέπει να αναλύσουμε το μέγεθος ενός ουδέτερου δικτύου. Δεσμοί υδρογόνου για το σχηματισμό δευτεροταγούς δομής μπορούν να παρουσιαστούν μόνο μεταξύ των G-C, A-T (δεσμοί Watson-Crick) και G-U νουκλεοτιδίων. Άρα, από τα $4^2 = 16$ πιθανά δινουκλεοτίδια, μόνο 6 από αυτά μπορούν να σχηματίσουν δεσμούς υδρογόνου. Συνεπώς, τα ουδέτερα δίκτυα είναι εκθετικά μικρότερα συγκριτικά με το χώρο των αλληλουχιών. Συγκεκριμένα, αν ένας χάρτης αλληλουχίας – δομής υποδεικνύει δομές στις οποίες υπάρχουν $O(n)$ ζεύγη βάσεων, τα ουδέτερα δίκτυα είναι εκθετικά μικρότερα από το χώρο των αλληλουχιών. Έτσι, μπορούμε να συμπεράνουμε ότι υπάρχουν εκθετικά περισσότερες τέτοιες δομές.

Είναι συνεπώς ενδιαφέρον να υπολογίσουμε την κατανομή του αριθμού των ζευγών βάσεων στις k -noncrossing δομές. Η ιδέα είναι να μπει μια «συνδυαστική» ετικέτα σε κάθε τόξο (ζεύγος βάσεων) στη γεννήτρια συνάρτηση, περνώντας σε μια γεννήτρια συνάρτηση δύο μεταβλητών. Τα αποτελέσματα αυτά αποτελούν συνέπεια αυτών που παρουσιάστηκαν στα Κεφάλαια 2 και 3, αλλά είναι εκτός στόχων αυτής της εργασίας (Reidys 2011). Η οριακή κατανομή των ζευγών βάσεων στις k -noncrossing δομές λαμβάνει κατανομή Gauss (Εικόνα 4.1).



Εικόνα 4.1: Η οριακή κατανομή (γραμμή) και τα ακριβή δεδομένα απαρίθμησης για canonical 2- (κόκκινη), 3- (μπλε) και 4- (πράσινη) noncrossing δομές μήκους 200 νουκλεοτιδίων. (Reidys 2011)

Όπως βλέπουμε και στον Πίνακα 4.1,

	$k = 2$		$k = 3$		$k = 4$		$k = 5$	
τ	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$
1	0.276393	0.0447214	0.390891	0.0415653	0.425464	0.0314706	0.443020	0.0251601
2	0.317240	0.0643144	0.381701	0.0559928	0.403574	0.0470546	0.416068	0.0413361
3	0.336417	0.0791378	0.383555	0.0670987	0.400288	0.0559818	0.410087	0.0517052
4	0.348222	0.0916871	0.386408	0.0767872	0.400412	0.0667094	0.408701	0.0603242
5	0.356484	0.1028563	0.389134	0.0855937	0.401402	0.0748305	0.408741	0.0680229

Πίνακας 4.1: Μέση τιμή και διακύμανση του αριθμού των ζευγών βάσεων των k-noncrossing, τ -canonical δομών. (Reidys 2011)

ο αριθμός των ζευγών βάσεων των 3-noncrossing canonical RNA δομών μήκους n έχει μέση τιμή $0,38n$. Αυτό σημαίνει ότι οι χάρτες αλληλουχίας – δομής παρουσιάζουν μια πληθώρα δομών με μεγάλα ουδέτερα δίκτυα. Όπως θα συζητήσουμε και στο τελευταίο κομμάτι, αυτοί οι χάρτες είναι ιδανικοί για να διευκολύνουν την εξελικτική βελτιστοποίηση μέσω τυχαίων μεταλλαγών.

4.3 Συζήτηση

Το κυριότερο, λοιπόν, αποτέλεσμα που προκύπτει από αυτή την εργασία είναι η ύπαρξη εκθετικά αυξανόμενων ουδέτερων δικτύων των 3-noncrossing canonical RNA ψευδοκόμπων, όπως προέκυψε από την ασυμπτωτική εκτίμηση της γεννήτριας συνάρτησης αυτών. Ως εκ τούτου, οι δομές αυτές παρουσιάζουν μια ιδιαίτερη ευρωστία, γεγονός που είναι απαραίτητο αν αναλογιστούμε τη σημασία της καταλύτικης τους δράσης και λογικό δεδομένου ότι παραμένουν εξαιρετικά συντηρημένες, σε ορισμένες περιπτώσεις ανάμεσα σε όλους τους ευκαρυωτικούς οργανισμούς. Επίσης, όπως συζητήσαμε στην προηγούμενη ενότητα, οι δομές αυτές είναι ιδανικές για την εξελικτική βελτιστοποίηση τους, δηλαδή τη γρήγορη κτήση της κατάλληλότερης δομής για την εκάστοτε λειτουργία, κάτι το οποίο συνέβη, πιθανώς, όταν οι δομές αυτές απέκτησαν τη συγκεκριμένη δράση.

Ένα άλλο σημαντικό χαρακτηριστικό που έχει παραβλεφθεί στη βιβλιογραφία είναι το γεγονός ότι τα περισσότερα μόρια RNA που έχουν αποδεδειγμένα λειτουργική δευτεροταγή δομή, και η οποία είναι πολύ συντηρημένη, έχουν δομή ψευδοκόμπου (και, συγκεκριμένα, 3-noncrossing δομή). Το γεγονός αυτό δεν μπορεί να είναι τυχαίο. Αυτό σημαίνει, είτε ότι η δομή αυτή εξασφαλίζει κάποιο λειτουργικό πλεονέκτημα το οποίο δεν μπορεί να παραχθεί από άλλη δομή, είτε ότι έχει τόση ευρωστία ώστε ευνοήθηκε εξελικτικά και διατηρήθηκε από τον RNA κόσμο μέχρι σήμερα.

Έτσι, κοιτώντας στη βιβλιογραφία, βλέπουμε ότι ανάμεσα στα RNA μόρια που λαμβάνουν μορφή ψευδοκόμπου είναι: το RNA συστατικό της τελομεράσης (η δομή του οποίου είναι συντηρημένη από τους μύκητες και τα πρωτόζωα μέχρι τον άνθρωπο) (Theimer, Blois et al. 2005), η RNA αλυσίδα της RNaseP (συντηρημένη δομή του ψευδοκόμπου σε όλου τους οργανισμούς – βακτήρια, αρχαία, ευκαρυωτικοί οργανισμοί) (Hartmann and Hartmann 2003), τα ιντρόνια τάξης I (τα οποία εντοπίζονται συντηρημένα σε βακτήρια, πρώτιστα και βακτηριοφάγους, ενώ δεν υπάρχουν σε ανώτερους οργανισμούς καθώς η λειτουργία τους αντικαταστάθηκε πιθανώς από το ματισμόσωμα) (Adams, Stahley et al. 2004) καθώς και οι ιικοί RNA ψευδοκόμπου (τα οποία ανάλογα με τη λειτουργία τους διαφέρουν από ιό σε ιό) (Brierley, Pennell et al. 2007).

Είναι σίγουρο ότι χρειάζονται πειράματα τα οποία θα αξιολογήσουν τη σημασία των δομών ψευδοκόμπων τόσο για αυτή καθαυτή τη λειτουργία τους, όσο και για την ικανότητα τους να διατηρούν σταθερή τη δομή τους παρά τη συσσώρευση μεταλλαγών. Παρ'όλα αυτά, όπως φαίνεται από τις έρευνες που αναφέρονται σ'αυτή την εργασία, τα μαθηματικά μπορούν να προσφέρουν μια κλειδαρότρυπα, μέσα από την οποία μπορούμε να κοιτάξουμε στο παρελθόν και να εξάγουμε χρήσιμα συμπεράσματα σχετικά με την εξελικτική δυναμική, την ευρωστία και την εξελικτικότητα ζωντανών συστημάτων.

Βιβλιογραφία

Adams, P. L., M. R. Stahley, et al. (2004). "Crystal structure of a self-splicing group I intron with both exons." *Nature* **430**(6995): 45-50.

Bornberg-Bauer, E. (1997). "How are model protein structures distributed in sequence space?" *Biophys J* **73**(5): 2393-2403.

Brierley, I., S. Pennell, et al. (2007). "Viral RNA pseudoknots: versatile motifs in gene expression and replication." *Nat Rev Microbiol* **5**(8): 598-610.

Dayton, E. T., D. A. Konings, et al. (1992). "Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein." *J Virol* **66**(2): 1139-1151.

Egli, M., G. Minasov, et al. (2002). "Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution." *Proc Natl Acad Sci U S A* **99**(7): 4302-4307.

Flajolet, P. and R. Sedgewick (2009). *Analytic combinatorics*. Cambridge ; New York, Cambridge University Press.

Hartmann, E. and R. K. Hartmann (2003). "The enigma of ribonuclease P evolution." *Trends Genet* **19**(10): 561-569.

Jackson, R. J. and A. Kaminski (1995). "Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond." *RNA* **1**(10): 985-1000.

James, B. D., G. J. Olsen, et al. (1988). "The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme." *Cell* **52**(1): 19-26.

James, B. D., G. J. Olsen, et al. (1989). "Phylogenetic comparative analysis of RNA secondary structure." *Methods Enzymol* **180**: 227-239.

Jorg, T., O. C. Martin, et al. (2008). "Neutral network sizes of biological RNA molecules can be computed and are not atypically small." *BMC Bioinformatics* **9**: 464.

- Ke, A., K. Zhou, et al. (2004). "A conformational switch controls hepatitis delta virus ribozyme catalysis." Nature **429**(6988): 201-205.
- Lisacek, F., Y. Diaz, et al. (1994). "Automatic identification of group I intron cores in genomic DNA sequences." J Mol Biol **235**(4): 1206-1217.
- Lyngso, R. B. and C. N. Pedersen (2000). "RNA pseudoknot prediction in energy-based models." J Comput Biol **7**(3-4): 409-427.
- Mandl, C. W., H. Holzmann, et al. (1998). "Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus." J Virol **72**(3): 2132-2140.
- Michiels, P. J., A. A. Versleijen, et al. (2001). "Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting." J Mol Biol **310**(5): 1109-1123.
- Muse, S. V. (1995). "Evolutionary analyses of DNA sequences subject to constraints of secondary structure." Genetics **139**(3): 1429-1439.
- Nixon, P. L., A. Rangan, et al. (2002). "Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot." J Mol Biol **322**(3): 621-633.
- Parsch, J., J. M. Braverman, et al. (2000). "Comparative sequence analysis and patterns of covariation in RNA secondary structures." Genetics **154**(2): 909-921.
- Powell, D. M., M. J. Zhang, et al. (1995). "Sequence specificity in the higher-order interaction of the Rev protein of HIV-1 with its target sequence, the RRE." J Acquir Immune Defic Syndr Hum Retrovirol **10**(3): 317-323.
- Rastogi, T., T. L. Beattie, et al. (1996). "A long-range pseudoknot is required for activity of the Neurospora VS ribozyme." EMBO J **15**(11): 2820-2825.
- Reidys, C. M. (2011). Combinatorial computational biology of RNA : pseudoknots and neutral networks. New York, Springer Science+Business Media.
- Reidys, C. M. and R. R. Wang (2010). "Shapes of RNA pseudoknot structures." J Comput Biol **17**(11): 1575-1590.

Shen, L. X. and I. Tinoco, Jr. (1995). "The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus." J Mol Biol **247**(5): 963-978.

Stanley, R. P. (2012). Enumerative combinatorics. Volume 1. Cambridge, NY, Cambridge University Press.

Theimer, C. A., C. A. Blois, et al. (2005). "Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function." Mol Cell **17**(5): 671-682.

Titchmarsh, E. C. (1939). The theory of functions. London, Oxford university press.

van Nimwegen, E., J. P. Crutchfield, et al. (1999). "Neutral evolution of mutational robustness." Proc Natl Acad Sci U S A **96**(17): 9716-9720.

Wagner, A. (2005). Robustness and evolvability in living systems. Princeton, N.J., Princeton University Press.