



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΧΗΜΙΚΩΝ ΜΗΧΑΝΙΚΩΝ  
ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ**

# Εφαρμογή μεθόδων Συστημικής Βιολογίας στη μελέτη των μηχανισμών ρύθμισης οργανισμών βιοτεχνολογικού ενδιαφέροντος

# Διδακτορική Διατριβή

## Ελευθέριος Πιλάλης

Ιούλιος 2011

Η έγκριση της διδακτορικής διατριβής από την Ανωτάτη Σχολή Χημικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα (Ν. 5343/1932, Άρθρο 202).

## **Ευχαριστίες**

Θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στον Καθηγητή Φραγκίσκο Κολίση, επιβλέποντα της διατριβής αυτής, για την πολύτιμη υποστήριξη και καθοδήγηση που κατέστησαν δυνατή την πραγματοποίησή της. Θα ήθελα επίσης να ευχαριστήσω θερμά τον Δρ. Αριστοτέλη Χατζηιωάννου, Ερευνητή Γ' στο Ινστιτούτο Βιολογικών Ερευνών και Βιοτεχνολογίας (IBEB) του Εθνικού Ιδρύματος Ερευνών (ΕΙΕ) για την επιστημονική καθοδήγηση και τη συμπαράσταση καθόλη τη διάρκεια εκπόνησης της διατριβής. Εκφράζω επίσης τις θερμές μου ευχαριστίες στους συνεργάτες μου Δρ. Όλγα Παπαδόδημα, Δρ. Παναγιώτη Μούλο (IBEB), Ιωάννη Κανάρη και Χαράλαμπο Δουύκα (υποψήφιοι διδάκτορες Πανεπιστημίου Αιγαίου) για τη συμπαράσταση, το ιδιαίτερα φιλικό κλίμα συνεργασίας και τις εποικοδομητικές συζητήσεις.

Παράλληλα, ευχαριστώ θερμά για τη συνεισφορά τους, τον Καθηγητή Δημήτριο Κυριακίδη, τον Καθηγητή Χρήστο Παναγιωτίδη και τον Δρ. Αστέριο Γρηγορούδη από το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.

Τέλος, θα ήθελα ν'αναφερθώ στην πολύτιμη και αμέριστη συμπαράσταση της συζύγου μου, Δέσποινας Σακλαμάκη και να αφιερώσω τη διατριβή σε αυτή και στις δύο κόρες μας.



## **Περιεχόμενα**

Ευχαριστίες .....	3
Εισαγωγή .....	9
Κεφάλαιο 1 – Συστημική ανάλυση του κεντρικού μεταβολισμού της Ελαιοκράμβης με τη χρήση υπολογιστικού μοντέλου μεγάλης κλίμακας.....	15
1.1    Εισαγωγή.....	15
1.2 Μέθοδοι.....	19
1.2.1 Ανάπτυξη του μοντέλου .....	19
1.2.2 Υπολογισμός των μεταβολικών ροών του εμβρύου της ελαιοκράμβης.....	24
1.2.3 Συστημική ανάλυση της κατανομής ροών σταθερής κατάστασης.....	27
1.3    Αποτελέσματα.....	30
1.3.1 Μοντελοποίηση του κεντρικού μεταβολισμού της ελαιοκράμβης .....	30
1.3.2 Ανάλυση Διαφοροποίησης Ροών (FVA).....	50
1.3.3 Συστημική ανάλυση του χώρου ροών .....	58
1.3.4 <i>In silico</i> διερεύνηση του ρόλου του μεταγραφικού παράγοντα WRINKLED1 (WRL1) .....	60
1.4 Συμπεράσματα .....	67
Κεφάλαιο 2 – Βιοπληροφορική πρόβλεψη στόχων του μεταγραφικού παράγοντα AtoC στο γονιδίωμα του βακτηρίου <i>Escherichia coli</i> .....	69
2.1 Εισαγωγή.....	69
2.2 Μέθοδοι.....	74
2.2.1 Γενική περιγραφή της υπολογιστικής διαδικασίας ανίχνευσης και εκτίμησης μοτίβων .....	74
2.2.2 Πληροφορικά και μαθηματικά εργαλεία .....	75
2.2.3 Επαναληπτική διαδικασία de novo ανίχνευσης συντηρημένης περιοχής DNA και κατασκευή μοτίβων .....	77
2.2.4 Λειτουργική ανάλυση μοτίβων.....	79

2.2.5 Πειραματική επιβεβαίωση επιλεγμένων στόχων των μοτίβων για <i>in vivo</i> σύνδεση με τον ρυθμιστή απόκρισης AtoC.....	80
<b>2.3 Αποτελέσματα .....</b>	<b>81</b>
2.3.1 Εφαρμογή της υπολογιστικής διαδικασίας πρόβλεψης στόχων του μεταγραφικού παράγοντα AtoC .....	81
2.3.3 Αξιολόγηση ενδιάμεσων μοτίβων .....	85
2.3.4 Συσχέτιση μοτίβων με γονιδιακές λειτουργίες και μεταβολικά μονοπάτια	87
2.3.5 Πειραματική επιβεβαίωση επιλεγμένων στόχων των μοτίβων για <i>in vivo</i> σύνδεση με τον ρυθμιστή απόκρισης AtoC.....	89
2.3.6 Εφαρμογή της υπολογιστικής διαδικασίας στον μεταγραφικό παράγοντα LexA .....	94
<b>2.4 Συμπεράσματα .....</b>	<b>98</b>
2.4.1 Αναγνώριση νέων στόχων του AtoC σε ολόκληρο το γονιδίωμα του <i>E. coli</i> . .....	98
2.4.2 Ανάδειξη cis-ρυθμιστικών στοιχείων με πολύ χαμηλή συντήρηση πρωτοταγούς δομής μέσω της στατιστικής ανάλυσης γονιδιακών οντολογιών... .....	99
2.4.3 Προοπτικές ανάπτυξης της υπολογιστικής διαδικασίας.....	99
<b>Βιβλιογραφία .....</b>	<b>101</b>
<b>Παράρτημα Α.....</b>	<b>107</b>
1. Κώδικας Κεφαλαίου 1 .....	107
1. Κώδικας κεφαλαίου 2 .....	109
<b>Παράρτημα Β .....</b>	<b>143</b>
Πειραματική επιβεβαίωση στόχων του παράγοντα AtoC.....	143
Συνθήκες καλλιέργειας.....	143
Αναλύσεις ανοσοκατακρήμνισης χρωματίνης (ChIP assays) .....	143
Electrophoretic Mobility Shift Assays .....	144
Αποτελέσματα πειραματικής επιβεβαίωσης.....	144
<b>Παράρτημα Γ .....</b>	<b>147</b>

Δημοσιεύσεις διατριβής σε διεθνή περιοδικά .....	147
Άλλες δημοσιεύσεις του Ελευθέριου Πιλάλη .....	147
Ανακοινώσεις σε διεθνή συνέδρια .....	147



## Εισαγωγή

Αντικείμενο της παρούσης διδακτορικής διατριβής είναι η μελέτη των μηχανισμών ρύθμισης οργανισμών βιοτεχνολογικού ενδιαφέροντος με μεθοδολογίες Βιολογίας Συστημάτων. Ο όρος Βιολογία Συστημάτων (Kitano 2001) περιγράφει ένα γενικό θεωρητικό πλαίσιο ολιστικής προσέγγισης προβλημάτων βιολογικών συστημάτων, εν αντιθέσει με τις αναγωγιστικές (reductionist) προσεγγίσεις που εξετάζουν τις ιδιότητες μεμονωμένων θεμελιωδών παραγόντων ενός συστήματος. Ο όρος σύστημα αναφέρεται ως ένας αριθμός αλληλοεξαρτώμενων και αλληλεπιδρώντων επιμέρους στοιχείων που ορίζουν ένα πολύπλοκο και δομημένο λειτουργικό σύνολο. Οι μεθοδολογίες της Βιολογίας Συστημάτων αποσκοπούν στην ανάδειξη και μελέτη των αναδυόμενων ιδιοτήτων (emergent properties) των βιολογικών συστημάτων, οι οποίες δεν εξηγούνται άμεσα από τις επι μέρους ιδιότητες των θεμελιωδών συστατικών αλλά προκύπτουν μέσω της πολυπλοκότητας των αλληλεπιδράσεων τους. Προκειμένου οι βιολογικοί μηχανισμοί να γίνουν κατανοητοί με την έννοια του συστήματος, είναι απαραίτητη η ολιστική εξέταση των βιολογικών δομών σε διάφορα επίπεδα οργάνωσης (γονιδιώματος, κυττάρου, ιστού κ.α.) καθώς και των δυναμικών ιδιοτήτων τους και όχι η μελέτη των χαρακτηριστικών μεμονωμένων μερών ενός κυττάρου ή ενός οργανισμού (Kitano, 2002). Κατά συνέπεια, η ανάπτυξη της Βιολογίας Συστημάτων ωφελήθηκε καθοριστικά από την πρόοδο των πειραματικών μεθόδων υψηλής τροφοδοσίας/απόδοσης (high throughput) δεδομένων και μετρήσεων, η οποία κατέστησε δυνατή την παροχή βιολογικών πληροφοριών σε μεγάλη κλίμακα. Επιπρόσθετα, η υψηλή διαθεσιμότητα πληροφοριών ανέδειξε το πρόβλημα της πολυπλοκότητας των βιολογικών μηχανισμών ως πολυπαραγοντικών συστημάτων, της οποίας η αντιμετώπιση είναι δύσκολη χωρίς τη συνδρομή της πληροφορικής και μαθηματικών εργαλείων. Κατά συνέπεια, η Βιολογία Συστημάτων είναι άρρηκτα συνδεδεμένη με την βιοπληροφορική, τα μαθηματικά και άλλα συναφή πεδία που προσφέρουν εργαλεία διαχείρισης της πολυπλοκότητας και αναγωγής της σε συστημικές ιδιότητες.

Μια από τις πιο θεμελιώδεις έννοιες της Βιολογίας Συστημάτων είναι η έννοια του δικτύου αλληλεπιδράσεων, δηλαδή των ολοκληρωμένων δικτύων μορίων, γονιδίων ή βιοχημικών αντιδράσεων, η δομή των οποίων καθορίζει τους κυτταρικούς φαινότυπους. Τα δίκτυα αλληλεπιδράσεων δεν αποτελούν απλώς ένα άθροισμα

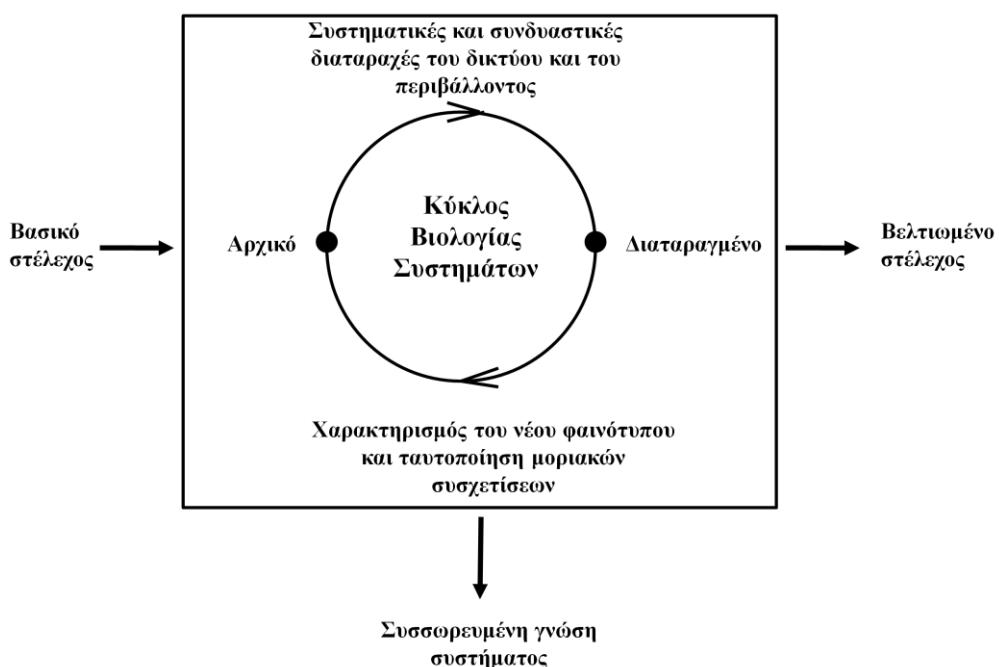
παρατεθημένων βιομορίων, αλλά δομημένα συστήματα μεγάλης κλίμακας, όπου όλα τα επιμέρους στοιχεία έχουν έναν ρόλο που μπορεί να αποδοθεί ποσοτικοποιημένα. Κατά συνέπεια, στα επιμέρους πεδία της Βιολογίας Συστημάτων, όπως είναι η Γονιδιωματική, η Πρωτεινωμική και η Μεταβολωμική, δεν εξετάζεται ο ρόλος μεμονωμένων γονιδίων, πρωτεΐνων, ή μεταβολιτών αντίστοιχα, αλλά η συνδυαστική δράση όλων μαζί, ή τουλάχιστον ενός μεγάλου μέρους τους. Οι αναδυόμενες, λόγω της συνδυαστικής δράσης, ιδιότητες καθορίζουν τη λειτουργία του δικτύου, με αποτέλεσμα να είναι αδύνατη η ερμηνεία του βιολογικού μηχανισμού από μεμονωμένα επιμέρους συστατικά του. Στόχος, επίσης, της Βιολογίας Συστημάτων είναι η ολοκληρωμένη συσχέτιση (integration) διαφορετικών επιπέδων της κυτταρικής φυσιολογίας. Η ολοκλήρωση του γονιδιακού με το μεταβολικό επίπεδο αποτελεί ένα χαρακτηριστικό παράδειγμα, το οποίο εκφράζεται μέσω της κατασκευής υπολογιστικών κυτταρικών μοντέλων (*in silico* cellular models) που λαμβάνουν συνδυαστικά υπόψη το γονιδίωμα με τον μεταβολισμό του κυττάρου και αποτελούν εργαλεία προσομοίωσης της ρύθμισης της κυτταρικής φυσιολογίας σε συστημικό επίπεδο.

Οι εφαρμογές Βιολογίας Συστημάτων στη Βιοτεχνολογία είναι ποικίλες και συνήθως εστιάζουν στη μελέτη συστημικών ιδιοτήτων του κυτταρικού μεταβολισμού και της κυτταρικής ρύθμισης, με σκοπό την εφαρμογή στρατηγικών μεταβολικής μηχανικής για την παραγωγή σε μεγάλη κλίμακα αλλά και νεοφανών προϊόντων υψηλής προστιθέμενης αξίας (Lee et al. 2007, Park et al. 2007). Στα πλαίσια των προσεγγίσεων αυτών παρουσιάζεται μια σειρά ανοιχτών θεμάτων και προκλήσεων, όπως η ταυτοποίηση γενετικών στόχων για την τροποποίηση και βελτίωση κυτταρικών στελεχών, η ανάδειξη δικτύων αλληλεπίδρασης βιομορίων, η αξιοποίηση μεταβολικών πληροφοριών και ποσοτικών μετρήσεων σε μεγάλη κλίμακα και η ολοκλήρωση (integration) γενετικής πληροφορίας και μεταβολισμού (Stephanopoulos et al. 2004; Feist et al. 2007; Henry et al. 2006).

Στην παρούσα διατριβή επιχειρείται η προσέγγιση των μηχανισμών ρύθμισης δυο οργανισμών, ενός φυτικού και ενός μικροβιακού, με έντονο βιοτεχνολογικό ενδιαφέρον.

Στο Κεφάλαιο 1, παρουσιάζεται η ανάπτυξη υπολογιστικού μοντέλου μεγάλης κλίμακας του κεντρικού μεταβολισμού του εμβρύου της ελαιοκράμβης (*Brassica*

*napus*) με σκοπό την εξαγωγή συστηματικών ιδιοτήτων του φυτικού μεταβολισμού. Η ανάπτυξη μοντέλων για την εφαρμογή προσομοιώσεων βρίσκεται στην καρδιά της Βιολογίας Συστημάτων, καθώς μια από τις βασικές αρχές της δέχεται ότι η διαλεύκανση των θεμελιωδών μηχανισμών που διέπουν ένα βιολογικό σύστημα μπορεί να επιταχυνθεί με την συστηματική εισαγωγή διαταραχών και συνεπακόλουθη μελέτη των επιπτώσεών τους (Σχήμα 1). Τα μεταβολικά μοντέλα γονιδιωματικής κλίμακας αποτελούν, επίσης, ένα σημαντικό στάδιο ολοκλήρωσης της κυτταρικής φυσιολογίας, καθώς συνδέουν το γονιδίωμα με τον μεταβολισμό. Η αναγωγή των βιοχημικών φαινοτύπων σε δίκτυο αλληλεπίδρασης μεταβολιτών επιτρέπει την ανάδειξη συστηματικών ιδιοτήτων, όπως την απόκριση του μεταβολισμού σε εξωτερικές (σε σχέση με το περιβάλλον) ή εσωτερικές (σε σχέση με το γονιδίωμα, την μεταφορά σήματος) αλλαγές.



**Σχήμα 1.** Ο κύκλος της Βιολογίας Συστημάτων στη Βιοτεχνολογία. Η συστηματική εισαγωγή και μελέτη διαταραχών συμβάλλει στη συσσώρευση γνώσης των μοριακών αλληλεπιδράσεων και στη βελτίωση των στελεχών.

Το έμβρυο της ελαιοκράμβης είναι ένα κυτταρικό εργοστάσιο παραγωγής ελαίου με υψηλή προστιθέμενη αξία ως βιοκαύσμου αλλά και ως διατροφικού προϊόντος. Η συμβολή στη μελέτη των ιδιοτήτων του φυτικού κεντρικού μεταβολισμού, με έμφαση

στη βιοσύνθεση λιπαρών οξέων, παρουσιάζει εξαιρετικό ενδιαφέρον για την βελτίωση ή και αύξηση της παραγωγής ελαίων υψηλής προστιθέμενης αξίας. Στην εργασία αυτή, αναπτύχθηκε βιοχημικό μοντέλο μεγάλης κλίμακας της ελαιοκράμβης στο οποίο ελήφθη υπόψη για πρώτη φορά η διαμερισματοποίηση του φυτικού κυττάρου και χρησιμοποιήθηκε για την εφαρμογή συστηματικής ανάλυσης στην μελέτη των δομικών ιδιοτήτων του μεταβολισμού. Στην ανάλυση αυτή, εφαρμόστηκε μαθηματική μέθοδος μείωσης της πολυπλοκότητας έτσι ώστε οι ιδιότητες του μεταβολικού δικτύου να αναχθούν σε μικρό αριθμό σημαντικών ρυθμιστικών αντιδράσεων. Συνεπώς, επετεύχθη η ποσοτικοποίηση του ρυθμιστικού ρόλου των ενζυμικών δραστικοτήτων και αναδείχθηκαν οι πιο σημαντικοί εν δυνάμει ενζυμικοί στόχοι για την βελτίωση του μεταβολισμού. Οι υπολογιστικές προσομοιώσεις της ανάπτυξης του εμβρύου που πραγματοποιήθηκαν, ήταν σύμφωνες με πειραματικές παρατηρήσεις από τη βιβλιογραφία και απέδειξαν τη λειτουργικότητα του μοντέλου αλλά και τη χρησιμότητά του για τη μελέτη του ρυθμιστικού ρόλου μεταγραφικών παραγόντων. Τέλος, οι πιθανοί ενζυμικοί στόχοι για την ρύθμιση της βιοσύνθεσης των λιπιδίων εξετάστηκαν μέσω της εφαρμογής υπολογιστικής προσομοίωσης γενετικών τροποποιήσεων (διαγραφών γονιδίων).

Στο Κεφάλαιο 2, παρουσιάζεται η ανάπτυξη μιας υπολογιστικής διαδικασίας πρόβλεψης στόχων μεταγραφικών παραγόντων, σε επίπεδο ολόκληρου γονιδιώματος. Οι μεταγραφικοί παράγοντες είναι πρωτεΐνες που ελέγχουν την μεταγραφή των γονιδίων σε RNA, με την συνδυαστική πρόσδεσή τους στις μεσογονιδιακές επαγωγικές περιοχές. Συνεπώς, αποτελούν το κύριο μέσο ρύθμισης της έκφρασης των γονιδίων και μετάδοσης σήματος στο κύτταρο. Μια κατηγορία μεταγραφικών παραγόντων, οι ρυθμιστές απόκρισης, ελέγχουν την μεταγραφική απόκριση των βακτηριακών κυττάρων στα εξωτερικά ερεθίσματα του περιβάλλοντος. Οι ρυθμιστές απόκρισης, συνδεδεμένοι με αισθητήριες διαμεμβρανικές κινάσες, σχηματίζουν τα «συστήματα δυο συστατικών» (Two-component Systems) έτσι ώστε να δέχονται τα εξωτερικά ερεθίσματα από το περιβάλλον και να τα μετατρέπουν σε σήματα γονιδιακής ρύθμισης, επιτρέποντας στο βακτήριο να προσαρμόσει ανάλογα τον μεταβολισμό του. Η μελέτη των συστημάτων δυο συστατικών παρουσιάζει ιδιαίτερο ενδιαφέρον στη Βιοτεχνολογία, καθώς αυτά αποτελούν τη μοριακή βάση μιας σειράς βασικών βακτηριακών ιδιοτήτων. Συγκεκριμένα, τα συστήματα δυο συστατικών ρυθμίζουν την ανθεκτικότητα των βακτηρίων σε χημικές ενώσεις, την αντίσταση σε

αντιβιοτικά, τη βιοσύνθεση μεταβολικών προϊόντων υψηλής προστιθέμενης αξίας και τον καταβολισμό τοξικών μορίων. Η ανίχνευση των γονιδιακών στόχων των «συστημάτων δυο συστατικών», αναμένεται να συμβάλει καθοριστικά σε βιοτεχνολογικές εφαρμογές, όπως η βιοσύνθεση επιθυμητών μεταβολικών προϊόντων, η μικροβιακή απορρύπανση από τοξικές χημικές ενώσεις, η αντιμετώπιση της βακτηριακής παθογένειας στη διατροφή και η σύνθεση νέων αποτελεσματικότερων αντιβιοτικών, κ.α.

Εν τούτοις, η ανίχνευση των θέσεων πρόσδεσης των ρυθμιστών απόκρισης στο DNA και η σύνδεσή τους με τα αντίστοιχα μεταβολικά μονοπάτια παραμένει ένα ανοιχτό θέμα, καθώς προσπίπτει στην έλλειψη αξιόλογης συντήρησης αλληλουχιών των επαγωγικών περιοχών του βακτηριακού γονιδιώματος. Η υπολογιστική διαδικασία που αναπτύχθηκε στα πλαίσια της διατριβής στοχεύει στη αυτόματη ανίχνευση πολλαπλών θέσεων πρόσδεσης και συνεπακόλουθης στατιστικής ανάλυσης, με σκοπό τη σύνδεση των θέσεων πρόσδεσης με ανάλογα βιολογικά μονοπάτια. Η διαδικασία αυτή εφαρμόστηκε στη μελέτη του «συστήματος δυο συστατικών» AtoSC, του βακτηρίου *Escherichia coli*. Οι κύριες λειτουργίες του συστήματος AtoSC είναι ο καταβολισμός των λιπαρών οξέων μικρής αλυσίδας και η βιοσύνθεση των πολυαμινών. Το σύστημα AtoSC εμπλέκεται, επίσης, σε μια σειρά κυτταρικών λειτουργιών, όπως η βακτηριακή κινητικότητα και η χημειόταξη, η απόκριση στην υψηλή οσμωτική πίεση και η ευαισθησία στα αμινογλυκοσικά αντιβιοτικά. Σε αντίθεση με τον καταβολισμό των λιπαρών οξέων, για τον οποίο η μοριακή βάση έχει διαλευκανθεί, παρεμένει άγνωστος ένας αριθμός γονιδιακών στόχων του συστήματος AtoSC και πιο συγκεκριμένα του δεύτερου συστατικού του, της μεταγραφικού παράγοντα AtoC. Η προσπάθεια αναγνώρισης δευτερευουσών θέσεων πρόσδεσης του AtoC στο βακτηριακό χρωμόσωμα είχε αποβεί άκαρπη λόγω της μη συντήρησης της μιας γνωστής αλληλουχίας πρόσδεσης σε άλλο σημείο του γονιδιώματος. Το πρόβλημα αυτό αντιμετωπίστηκε με την εφαρμογή μιας υπολογιστικής διαδικασίας ανίχνευσης μοτίβων DNA σε επίπεδο ολόκληρου γονιδιώματος, η οποία βασίζεται στην στατιστική ανάλυση όρων γονιδιακής οντολογίας (Gene Ontology Terms). Η στατιστική ανάλυση όρων γονιδιακής οντολογίας, που είναι μια μορφή δομημένης και iεραρχημένης αναπαράστασης της γονιδιακής λειτουργίας, επέτρεψε τη συνδυαστική εκτίμηση ομάδων πιθανών θέσεων πρόσδεσης ως προς τα μεταβολικά μονοπάτια τα οποία ελέγχουν. Η υπολογιστική

διαδικασία ανέδειξε πιθανούς γονιδιακούς στόχους του AtoC, εκ των οποίων 21 επιβεβαιώθηκαν και πειραματικά. Συνεπώς, η υπολογιστική διαδικασία που παρουσιάζεται, εκτός από την ανάδειξη της μοριακής βάσης γνωστών λειτουργιών του AtoSC, αποτελεί συμβολή στη διαλεύκανση των πολύπλοκων δικτύων των βακτηριακών σηματοδοτικών μηχανισμών ρύθμισης, μέσω της στατιστικής σύνδεσης βιολογικών μονοπατιών με ελάχιστα διατηρημένες ρυθμιστικές αλληλουχίες DNA.

# Κεφάλαιο 1 – Συστημική ανάλυση του κεντρικού μεταβολισμού της Ελαιοκράμβης με τη χρήση υπολογιστικού μοντέλου μεγάλης κλίμακας

## 1.1 Εισαγωγή

Η ανάλυση του κυτταρικού μεταβολισμού σε επίπεδο συστημάτων έχει αποδειχθεί ιδιαίτερα σημαντική στο σχεδιασμό στρατηγικών γενετικής μηχανικής που στοχεύουν στην ανακατανομή των μεταβολικών ροών προς επιθυμητά τελικά προϊόντα (Stephanopoulos et al. 2004). Η εκμετάλλευση βάσεων δεδομένων σε επίπεδο ολόκληρου του γονιδιώματος σε συνδυασμό με τις κατάλληλες μαθηματικές μεθόδους, οδήγησαν στην ανάπτυξη Μεταβολικών Μοντέλων Γονιδιωματικής Κλίμακας (MMGK, Genome-Scale Metabolic Reconstructions) (Oberhardt et al. 2009). Τα Μεταβολικά Μοντέλα Γονιδιωματικής Κλίμακας (MMGK) προκύπτουν από την επισημείωση πλήρως αλληλουχημένων γονιδιωμάτων και καλύπτουν το χάσμα μεταξύ της γενετικής πληροφορίας και των μεταβολικών φαινοτύπων. Ο όρος επισημείωση (annotation) αναφέρεται στον συστηματικό χαρακτηρισμό των ιδιοτήτων (λειτουργικών, δομικών κ.α.) των γονιδιωμάτων και των προϊόντων τους. Οι βασικές επισημειώσεις ενός γονιδίου αφορούν τη θέση του στο γονιδίωμα καθώς και τη λειτουργία της πρωτεΐνης που κωδικοποιεί, όπως για παράδειγμα τα σηματοδοτικά ή μεταβολικά μονοπάτια στα οποία συμμετέχει. Τα MMGK βασίζονται στις μεταβολικές επισημειώσεις των γονιδιωμάτων και προσφέρουν το πλαίσιο αξιοποίησης μαζικών πειραματικών δεδομένων σχετικών με τον μεταβολισμό των κυττάρων, αλλά και δίνουν την δυνατότητα υπολογιστικής (*in silico*) προσομοίωσης του μεταβολισμού σε επίπεδο ολόκληρου κυττάρου.

Το ενδιαφέρον για την ανάπτυξη MMGK είναι όμως συνεχώς αυξανόμενο, τροφοδοτούμενο από τις νέες τεχνολογίες τόσο στον τομέα της υπολογιστικής επισημείωσης όσο και στον τομέα των βιοχημικών και μοριακών τεχνικών μεγάλης κλίμακας που υπεισέρχονται στη ανάκτηση λειτουργικής βιολογικής πληροφορίας σε διάφορα επίπεδα (π.χ. μεταβολισμό, μεταγραφή, ενζυμική δραστικότητα κ.α.). Τα MMGK καθιστούν δυνατή την ανάλυση των συστημικών ιδιοτήτων του κυτταρικού μεταβολισμού υπό δεδομένες συνθήκες, καθώς και την υπολογιστική πρόβλεψη της ανταπόκρισης του μεταβολισμού στη μεταβολή των συνθηκών αυτών. Τα δίκτυα αλληλεπίδρασης ενέχουν την ιδιότητα να διατηρούν τα λειτουργικά χαρακτηριστικά

τους ενάντια σε διάφορες διαταραχές. Η ιδιότητα αυτή, που ονομάζεται *ευρωστία*, (robustness) είναι θεμελιώδης και καθορίζει την συμπεριφορά ενός μεταβολικού συστήματος σε εξωτερικές ή εσωτερικές μεταβολές. Οι εξωτερικοί παράγοντες εκφράζονται μέσω των ροών εισόδου και εξόδου οι οποίες συνδέουν τον μεταβολισμό με το εξωτερικό περιβάλλον (για παράδειγμα, οι συγκεντρώσεις μεταβολιτών στο θρεπτικό μέσο). Οι εσωτερικοί παράγοντες είναι τα ίδια τα στοιχεία του συστήματος ή ρυθμιστικά στοιχεία αυτών. Ένα κατεξοχήν πρόβλημα που επιλύεται με τις μεθόδους μοντελοποίησης σε γονιδιακή κλίμακα είναι η διαγραφή ή εισαγωγή γονιδίων. Η διαγραφή (και αντίστροφα η εισαγωγή) ενός γονιδίου αντιστοιχεί στην εξάλειψη μιας ή πολλαπλών χημικών αντιδράσεων ή στην μεταβολή τους σε περίπτωση που το γονίδιο αποτελεί ρυθμιστικό παράγοντα. Στα πλαίσια της Βιολογίας Συστημάτων, τα προβλήματα που πρέπει να επιλυθούν αφορούν την τροποποίηση όχι ενός αλλά συνδυασμού πολλών παραγόντων οι οποίοι τυπικά υπεισέρχονται σε πολλαπλά σημεία του μεταβολισμού. Η πολυπλοκότητα της αλληλεπίδρασης όλων αυτών των παραγόντων είναι δυνατόν να προσεγγιστεί μόνο με όρους Βιολογίας Συστημάτων, καθιστώντας την ανάγκη χρήσης ΜΜΓΚ επιτακτική.

Εν τούτοις, η ανάπτυξη ενός ΜΜΓΚ δεν απαιτεί μόνο την αυτόματα επισημειωμένη αλληλουχία του γονιδιώματος, αλλά και την εντατική, μη αυτοματοποιημένη ενσωμάτωση διαθέσιμης πληροφορίας για τον μεταβολισμό και τη φυσιολογία του οργανισμού υπό μελέτη. Η βαθύτερη γνώση του βιολογικού συστήματος, που προέρχεται από πειραματικές διαδικασίες, είναι απαραίτητη για την ανάπτυξη ενός λειτουργικού μοντέλου που να μπορεί αποδόσει ρεαλιστικές προσομοιώσεις. Τέτοια μοντέλα κατασκευάζονται «από κάτω προς τα πάνω» (*bottom-up approach*) και συνδυάζουν γονιδιωματική και βιβλιογραφική πληροφορία. Αντίθετα, η αυτόματη εξαγωγή μεταβολικών δικτύων από προσεγγίσεις «από πάνω προς τα κάτω» (*top-down approach*) δεν οδηγεί σε λειτουργικά μοντέλα (Thiele and Palsson, 2010). Σύμφωνα, επίσης, με τους Thiele and Palsson, 2010, η πλήρης αυτόματη κατασκευή μεταβολικών μοντέλων δεν είναι ακόμα εφικτή λόγω μιας σειράς προβλημάτων που επιλύονται μόνο με χειροκίνητη εκτίμηση, όπως η κατεύθυνση των αντιδράσεων, το ενδοκυτταρικό pH και η χρησιμοποίηση συγκεκριμένων υποστρωμάτων και συμπαραγόντων από τα ένζυμα. Για τον λόγο αυτό, η κατασκευή λειτουργικών ΜΜΓΚ υψηλής ποιότητας είναι χρονοβόρα και

μπορεί να κυμανθεί από 6 μήνες για καλά χαρακτηρισμένα βακτήρια με μικρά γονιδιώματα, μέχρι 2 χρόνια για τον ανθρώπινο μεταβολισμό (Thiele and Palsson, 2010).

Συμπερασματικά, ο αριθμός των διαθέσιμων ΜΜΓΚ είναι αρκετά περιορισμένος, ακόμα και για οργανισμούς με πολύ μικρά γονιδιώματα, όπως είναι οι προκαρυωτικοί. Για τον λόγο αυτό, οι περισσότερες προσπάθειες κατασκευής υπολογιστικών μεταβολικών μοντέλων μεγάλης κλίμακας αφορούν προκαρυωτικούς οργανισμούς, αφού οι ευκαρυωτικοί πολυκύτταροι οργανισμοί χαρακτηρίζονται από αυξημένη πολυπλοκότητα. Όσον αφορά τις προσπάθειες μοντελοποίησης πολυκύτταρων οργανισμών, αυτές έχουν επικεντρωθεί περισσότερο στον ανθρώπινο μεταβολισμό (Duarte et al. 2007; Ma et al. 2007). Επισήμως, έχει παρουσιαστεί ένα έλλειμμα στην διαθεσιμότητα μοντέλων ανώτερων φυτών. Συγκεκριμένα, ενώ περισσότερα από 30 μεταβολικά μοντέλα γονιδιωματικής κλίμακας έχουν δημοσιευθεί μέχρι σήμερα, ελάχιστα αφορούν φυτά (Boyle and Morgan 2009; Grafahrend-Belau et al. 2009; Oberhardt et al. 2009). Η ανάγκη ανάπτυξης φυτικών μοντέλων είναι ιδιαίτερα επιτακτική, δεδομένου του τεράστιου βιοτεχνολογικού και βιομηχανικού ενδιαφέροντος που παρουσιάζουν πολλά από αυτά.

Στα πλαίσια της παρούσης διδακτορικής διατριβής αναπτύχθηκε ένα ΜΜΓΚ μεγάλης κλίμακας του φυτού *Brassica napus* (ελαιοκράμβη) για να καλυφθεί η έλλειψη μοντέλου της οικογένειας *Brassicaceae*, η οποία περιλαμβάνει και το φυτό *Arabidopsis thaliana* που χρησιμοποιείται διεθνώς ως μοντέλο για τη μελέτη όλου του φυτικού βασιλείου. Συνεπώς, επιτυγχάνεται ένας διπλός στόχος, από την μια πλευρά η ανάπτυξη υπολογιστικού μοντέλου εκπροσώπου της οικογένειας *Brassicaceae* που παρουσιάζει ιδιαίτερο ενδιαφέρον για τη μελέτη μιας ευρείας οικογένειας φυτών και από την άλλη πλευρά η μοντελοποίηση της ελαιοκράμβης, σημαντικού ενεργιακού φυτού στη βιομηχανία παραγωγής ελαίων, αφού η βιομάζα των σπόρων της αποτελείται μέχρι και 60% από λιπίδια (Broun et al. 1999; Thelen and Ohlrogge 2002). Η ανάπτυξη του μεταβολικού μοντέλου της ελαιοκράμβης βασίστηκε στο γονιδίωμα του φυτού *Arabidopsis thaliana*, λόγω της ήδη διαθέσιμης πλήρους αλληλούχισής του, καθώς και τη διαθεσιμότητα πλήθους γονιδιακών επισημειώσεων. Αντίθετα, το γονιδίωμα της ελαιοκράμβης δεν είναι ακόμα διαθέσιμο. Το *Arabidopsis thaliana* είναι όμως ένας τυπικός εκπρόσωπος της οικογένειας των *Brassicaceae* και συνεπώς η στενή φυλογενετική συγγένεια ανάμεσα

στα δύο φυτά, ιδίως σε ό,τι αφορά τον κεντρικό μεταβολισμό, δικαιολογεί απόλυτα τη χρησιμοποίηση του γονιδιώματος του *Arabidopsis thaliana* για τον κεντρικό μεταβολισμό της ελαιοκράμβης.

Η μοντελοποίηση εστιάστηκε στο έμβρυο της ελαιοκράμβης, με σκοπό την συστημική ανάλυση της παραγωγής βιομάζας (πρωτεΐνων, λιπιδίων και αμύλου) και ιδιαίτερα τη βιοσύνθεση λιπαρών οξέων. Το μοντέλο περιλαμβάνει 262 μεταβολίτες και 313 βιοχημικές αντιδράσεις, κατανεμημένες σε 3 κυτταρικά διαμερίσματα, το κυτταρόπλασμα, το μιτοχόνδριο και τον χλωροπλάστη. Τα δεδομένα για την ανάπτυξη του μοντέλου αντλήθηκαν από εξειδικευμένες βάσεις βιοχημικών δεδομένων και από τη βιβλιογραφία. Η λειτουργικότητα του μοντέλου επιβεβαιώθηκε με τη μέθοδο της Flux Balance Analysis (FBA) και της γραμμικής βελτιστοποίησης. Στη συνέχεια εφαρμόστηκε μια συστημική ρυθμιστική ανάλυση του βιοχημικού δικτύου, η οποία μείωσε τη διαστασιμότητα του προβλήματος ανάγοντας τους παρατηρήσιμους βιοχημικούς φαινότυπους (την κατανομή των βιοχημικών ροών) σε ένα μικρό αριθμό ρυθμιστικών αντιδράσεων. Κατά αυτόν τον τρόπο εξήγησαν κρίσιμα συμπεράσματα αναφορικά με τον τρόπο που η ίδια η δομή του μεταβολικού δικτύου υπαγορεύει κάποιες πλευρές τις κυτταρικής φυσιολογίας, με έμφαση στη βιοσύνθεση των λιπαρών οξέων.

## 1.2 Μέθοδοι

### 1.2.1 Ανάπτυξη του μοντέλου

#### Συλλογή γονιδιακών και ενζυμικών επισημειώσεων

Η κατασκευή του ΜΜΓΚ ξεκινάει με τη συλλογή των γονιδίων και των αντίστοιχων ενζυμικών αντιδράσεων. Για το μοντέλο της ελαιοκράμβης οι πληροφορίες αυτές εξήχθησαν από τη βάση δεδομένων Aracyc 6.0 (Mueller et al. 2003; Zhang et al. 2005). Η βάση αυτή περιέχει το πλήρες γονιδίωμα του *A. thaliana* επισημειωμένο ως προς τις ενζυμικές δραστικότητες των γονιδίων. Ωστόσο, η βάση Aracyc δεν περιέχει επισημειώσεις ως προς την αντιστοίχιση των αντιδράσεων σε κυτταρικά διαμερίσματα. Η αντιστοίχιση αυτή επετεύχθη μέσω της αξιοποίησης της βάσης ενζυμικών δεδομένων BRENDA (Chang et al. 2009) καθώς και την κατάλληλη βιβλιογραφία που αφορά το εμβρυικό στάδιο ανάπτυξης του σπόρου της ελαιοκράμβης 14 ημέρες μετά την άνθηση (Schwender et al. 2004; Schwender et al. 2006). Εξήχθησαν συνολικά 1286 αντιδράσεις από τις οποίες ορίστηκαν οι συστημικές, δηλαδή αυτές που συνιστούν ένα επιλύσιμο σύστημα, όπως περιγράφεται λεπτομερώς παρακάτω. Το τελικό μοντέλο συμπεριέλαβε 262 μεταβολίτες και 313 βιοχημικές αντιδράσεις που αντιπροσωπεύουν τα μονοπάτια του κεντρικού μεταβολισμού στο κυτταρόπλασμα, τον χλωροπλάστη και το μιτοχόνδριο. Το σύνολο των αντιδράσεων μετατράπηκε στο format SBML (Systems Biology Markup Language) με το οποίο έγινε εφικτή η περαιτέρω αξιοποίησή του στα ειδικά προγράμματα.

#### Εξαγωγή συστημικών αντιδράσεων μέσω Ανάλυσης Ιδιαζουσών Τιμών του στοιχειομετρικού πίνακα

Το σύνολο των βιοχημικών αντιδράσεων που αντιστοιχούν στα επισημειωμένα γονίδια δεν συνιστούν ένα λειτουργικό βιοχημικό μοντέλο, καθώς το πιθανότερο είναι να περιέχουν ελλιπή περιγραφή των μεταβολικών μονοπατιών. Μη πλήρως περιγεγραμμένα μεταβολικά μονοπάτια περιέχουν μεταβολίτες-αδιέξοδα (dead-end metabolites), των οποίων η παραγωγή ή κατανάλωση δεν προβλέπεται από το σύστημα, καθώς και κομμάτια του μεταβολισμού ασύνδετα με τον κεντρικό μεταβολισμό. Σύμφωνα με τους Famili I. and Palsson BO., 2003, προκειμένου να οριστεί ένα μεταβολικό μοντέλο, είναι αναγκαίο να οριστούν οι συστημικές αντιδράσεις, δηλαδή οι αντιδράσεις που συναποτελούν ένα επιλύσιμο αλγεβρικό

σύστημα. Αυτό επιτυγχάνεται με τη μέθοδο της Ανάλυσης Ιδιαζουσών Τιμών (Singular Value Decomposition, SVD).

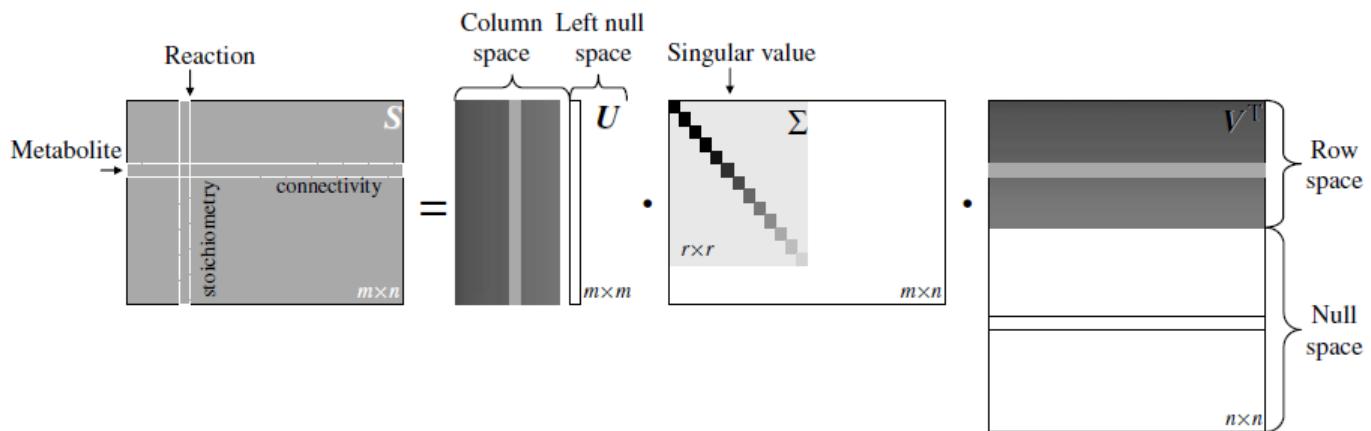
Η SVD αποτελεί μια αλγεβρική μέθοδο μείωσης της διαστασιμότητας ενός πίνακα. Χρησιμοποιείται ευρέως σε προβλήματα μείωσης διαστασιμότητας δεδομένων, όπως για παράδειγμα στην επεξεργασία σήματος, στη συμπίεση εικόνας κ.α. Στην περίπτωση ενός μεταβολικού μοντέλου, το οποίο αναπαριστάται με ένα στοιχειομετρικό πίνακα  $S$ , η μέθοδος SVD μπορεί να εφαρμοστεί για την εξαγωγή των θεμελιωδών ιδιοτήτων του πίνακα  $S$  (Famili I. and Palsson BO., 2003, Barrett et al., 2006). Ο στοιχειομετρικός πίνακας, αντίθετα με τους πίνακες πειραματικών δεδομένων, είναι ένας τέλειος πίνακας που περιέχει μόνο ακέραιους. Συνεπώς η μέθοδος SVD μπορεί να χρησιμοποιηθεί για την ανάλυση των βασικών ιδιοτήτων του μεταβολικού συστήματος μέσω της εξαγωγής των θεμελιωδών υπο-χώρων (sub-spaces) του στοιχειομετρικού πίνακα.

Έστω ένας ορθογώνιος πίνακας  $S$  με διαστάσεις  $m \times n$  (Σχήμα 1.1). Τότε υπάρχουν δύο ορθομοναδιαίοι πίνακες  $U$  και  $V$  οι οποίοι ικανοποιούν την εξίσωση:

$$S = U \Sigma V^T$$

Το αποτέλεσμα της εφαρμογής SVD στον στοιχειομετρικό πίνακα δίνεται στο Σχήμα 1.1. Οι στήλες των πινάκων  $U$  και  $V$  αποτελούν το αριστερό και δεξί μοναδιαίο διάνυσμα, τα οποία είναι τα modes. Η διαγώνιος του πίνακα  $\Sigma$  περιέχει τις ιδιοτιμές (singular values), οι οποίες δίνουν τη σχετική συμμετοχή των δυο διανυσμάτων στην ανακατασκευή του πίνακα  $S$ . Συνεπώς, ο πίνακας  $S$  μπορεί να ανακατασκευαστεί από τους  $U$ ,  $\Sigma$  και  $V$  ως:

$$S = \sum_{i=0}^n \sigma_i (u_i v_i^T)$$



**Σχήμα 1.1.** Η εφαρμογή της μεθόδου SVD στον στοιχειομετρικό πίνακα  $S$ . Ο μηδενικός χώρος (null space) περιέχει τις ροές σταθερής κατάστασης (από Palsson B., 2006).

Ο μηδενικός χώρος (null space) του πίνακα  $S$  ορίζεται ως εξής:

$$S \cdot v_{ss} = 0$$

Όπου  $v_{ss}$  οι κατανομές ροών σταθερής κατάστασης (steady-state). Ο μηδενικός χώρος αντιπροσωπεύει γραμμικά ανεξάρτητα διανύσματα βιοχημικών ροών (αντιδράσεων), οι γραμμικοί συνδυασμοί των οποίων δίνουν οποιαδήποτε κατανομή ροών σταθερής κατάστασης. Πρακτικά, ο μηδενικός χώρος δίνει την πληροφορία για το αν μια μεταβολική αντίδραση συμμετέχει ή σε κάποιο από τα γραμμικά ανεξάρτητα διανύσματα. Στην περίπτωση που δεν συμμετέχει, θεωρείται παρεμποδισμένη (blocked reaction), επειδή περιέχει κάποιον μεταβολίτη-αδιέξοδο. Από μια παρεμποδισμένη αντίδραση δεν μπορεί να περάσει μεταβολική ροή. Ο μηδενικός χώρος λοιπόν, επιτρέπει την επιθεώρηση του μοντέλου, με σκοπό την ανίχνευση ελλιπούς βιοχημικής περιγραφής και την αποκατάστασή της με την προσθήκη επιπρόσθετων αντιδράσεων. Στο στάδιο αυτό, είναι απαραίτητη η μη αυτόματη επιμέλεια (manual curation) του μοντέλου, με βάση την ευρύτερη γνώση της βιοχημείας του συστήματος, ώστε να προστεθούν οι απαραίτητες μη επισημειωμένες από το γονιδίωμα μεταβολικές αντιδράσεις (Thiele and Palsson, 2010).

### Ορισμός αντικειμενικής συνάρτησης βελτιστοποίησης

Ο υπολογισμός των μεταβολικών ροών σταθερής κατάστασης (steady-state) του εμβρύου της ελαιοκράμβης πραγματοποιήθηκε με τη μέθοδο της Ανάλυσης

Εξισορρόπησης Ροών (Flux Balance Analysis, FBA) (Varma and Palsson 1994). Η FBA είναι μια περιοριστική μέθοδος (constraint-based) η οποία προβλέπει τις μεταβολικές ροές σταθερής κατάστασης εφαρμόζοντας περιορισμούς εξισορρόπησης μάζας σε στοιχειομετρικά μοντέλα. Αυτό επιτυγχάνεται μέσω της γραμμικής βελτιστοποίησης (ή, αντίθετα, ελαχιστοποίησης) μια αντικειμενικής συνάρτησης που θεωρείται ότι έχει μια συνολική επίδραση στην κυτταρική φυσιολογία. Συνήθως χρησιμοποιούνται ενεργειακές διαδικασίες του κυττάρου όπως η παραγωγή ATP ή ο ρυθμός παραγωγής βιομάζας. Στην παρούσα μελέτη, ως αντικειμενική συνάρτηση, χρησιμοποιήθηκε ο ρυθμός παραγωγής βιομάζας από το έμβρυο της ελαιοκράμβης στο στάδιο της αύξησης της συγκέντρωσης ελαίου, 15 ημέρες μετά την άνθηση. Η σύνθεση της βιομάζας κατά το στάδιο αυτό, όπως επίσης και η σύνθεση του θρεπτικού μέσου εξήχθησαν από τη βιβλιογραφία (Schwender et al. 2004; Schwender et al. 2006) και παρουσιάζονται στον Πίνακα 1.1. Η παραγώμενη στο στάδιο αυτό βιομάζα αποτελείται από έλαιο (τριγλυκερίδια), άμυλο και πρωτεΐνες (κυρίως cruciferin, napin and oleosin). Στον Πίνακα 1.2 παρουσιάζεται η μοριακή σύνθεση των τριγλυκεριδίων που αποτελούν το έλαιο του σπόρου. Τα νουκλεϊκά οξέα δεν συμπεριελήφθησαν θεωρούμενα ως αμελητέα, καθώς τα έμβρυα στο στάδιο ανάπτυξης παράγουν κυρίως βιομόρια που αντιπροσωπεύουν ενεργειακά αποθέματα, όπως είναι το έλαιο και το άμυλο και δομικά βιομόρια, όπως είναι οι πρωτεΐνες. Ως ενεργειακά υποστρώματα στο θρεπτικό μέσο συμπεριελήφθησαν η σακχαρόζη, τα αμινοξέα αλανίνη και γλουταμίνη, η αμμωνία, nitrate, phosphate, sulfate, οι παράγοντες NAD, ADP, GDP and UDP. Οι περιορισμοί των αντιδράσεων παρουσιάζονται στον Πίνακα 1.3. Η φωτοσύνθεση θεωρήθηκε ενεργή στο στάδιο αυτό, καθώς έχει αποδειχθεί ότι αποτελεί πηγή πλαστιδικού NADPH για τα έμβρυα της ελαιοκράμβης (Schwender et al. 2003). Αντίθετα η φωτοαναπνοή περιορίστηκε ως μη παρατηρήσιμη (Schwender and Ohlrogge 2002).

Συστατικό Βιομάζας	g.g DW <sup>-1</sup> βιομάζας	Μοριακό Βάρος	mmoles.g DW <sup>-1</sup> βιομάζας
Λιπίδια (τριγλυκερίδια)	0.54	826	0.654
L-Glu	0.013	129.13	0.1
L-Gln	0.041	128.15	0.322
L-Pro	0.014	97.13	0.144
L-Arg	0.024	156.2	0.156
L-Asp	0.015	115.11	0.128
L-Asn	0.014	114.12	0.125
L-Thr	0.013	101.12	0.131
L-Ile	0.014	113.18	0.126
L-Met	0.006	131.21	0.049
L-Lys	0.013	128.19	0.104
L-Ala	0.013	71.1	0.181
L-Val	0.017	99.15	0.174
L-Leu	0.025	113.18	0.224
L-Ser	0.017	87.09	0.192
L-Gly	0.015	57.07	0.26
L-Cys	0.004	103.16	0.043
L-Phe	0.016	147.19	0.111
L-Tyr	0.012	163.19	0.072
L-Trp	0.005	186.23	0.026
L-His	0.007	137.16	0.049
άμυλο	0.162	181	0.895
ATP	-	-	49.9

**Πίνακας 1.1.** Συστατικά βιομάζας του εμβρύου της ελαιοκράμβης

Συστατικό ελαίου	mol/mol τριγλυκεριδίων
Λιπαρά οξέα C16	0.074
Λιπαρά οξέα C18	0.53
Λιπαρά οξέα C20	0.1
Λιπαρά οξέα C22	0.296
γλυκερόλη	1

**Πίνακας 1.2.** Μοριακή σύνθεση των λιπιδίων της ελαιοκράμβης

Αντίδραση	Περιορισμός (nmol.h <sup>-1</sup> .gDW <sup>-1</sup> )
Πρόσληψη σακχαρόζης	≤ 58.3
Πρόσληψη αλανίνης	≤ 7.36
Πρόσληψη γλουταμίνης	≤ 4.25
Πρόσληψη νιτρικού οξέος	≤ 1.2
Εκπομπή CO2	≥ 142
Φωτοαναπνοή (4.1.1)	0
Αμφίδρομες αντιδράσεις	≥ -2000, ≤ 2000
Μονόδρομες αντιδράσεις	≥0, ≤ 2000

**Πίνακας 1.3.** Οι περιορισμοί που εφαρμόστηκαν στο μοντέλο της ελαιοκράμβης

### 1.2.2 Υπολογισμός των μεταβολικών ροών του εμβρύου της ελαιοκράμβης

Οι μεταβολικές καταστάσεις του κυττάρου καθώς και η εξέλιξή τους στον χρόνο περιγράφονται πρωταρχικά από τιμές συγκεντρώσεων μεταβολιτών και ρυθμών μεταβολής των συγκεντρώσεων αυτών (reaction rates). Οι ποσότητες αυτές αλληλοσχετίζονται υπό τον νόμο της διατήρησης της μάζας, ο οποίος εκφράζεται με τους στοιχειομετρικούς συντελεστές των βιοχημικών αντιδράσεων. Οι ταχύτητες των βιοχημικών αντιδράσεων εξαρτώνται από τις συγκεντρώσεις των μεταβολιτών σε συνάρτηση με τις αντίστοιχες ενζυμικές κινητικές, οι οποίες μεταβάλλονται ως συνέπεια της μεταγραφικής και μεταβολικής ρύθμισης (Smallbone et al. 2007). Ένα ΜΜΓΚ τυπικά αποτελείται από εκατοντάδες βιοχημικές αντιδράσεις καθιστώντας την επίλυσή τους ως συστήματα κινητικών εξισώσεων δύσκολη, όχι μόνο από μαθηματική σκοπιά αλλά και λόγω της δυσκολίας μέτρησης ή διαθεσιμότητας των τιμών των παραμέτρων του συστήματος. Κατά συνέπεια στην γονιδιωματική κλίμακα απαιτείται η χρήση διαφορετικών μεθοδολογιών από την κινητική μοντελοποίηση. Η επικρατέστερη σήμερα μεθοδολογία μοντελοποίησης του μεταβολισμού σε γονιδιωματική κλίμακα είναι τα Περιοριστικά Μοντέλα (PM, constraint-based models) (Barrett et al. 2006). Τα ΠΜ παρακάμπτουν τις προαναφερθείσες δυσκολίες εστιάζοντας στους μέσους ρυθμούς αντιδράσεων που διατηρούνται στο κύτταρο σε κάποια δεδομένη μεταβολική κατάσταση. Η κεντρική παραδοχή της προσέγγισης αυτής είναι ότι στην παρατηρούμενη χρονική κλίμακα οι μεταβολικές ροές βρίσκονται σε σταθερή κατάσταση (steady-state) εώς ότου οι εξωτερικές συνθήκες μεταβληθούν. Συνεπώς σε μια δεδομένη κατάσταση οι μεταβολικές ροές περιορίζονται από τους στοιχειομετρικούς συντελεστές και η σχέση μεταξύ αυτών

ανάγεται σε γραμμική. Με αυτόν τον τρόπο καθίσταται δυνατόν να εφαρμοσθούν και επιπρόσθετοι περιορισμοί ως προς τα κατώτερα και ανώτερα όρια των ρυθμών μεταβολής, αυξάνοντας την αποτελεσματικότητα του μοντέλου στην ρεαλιστική αναπαραγωγή του κυτταρικού μεταβολισμού. Το σύνολο των περιορισμών εκφράζεται με τις κατανομές των ροών που επιτυγχάνονται όταν το σύστημα επιλυθεί με μεθόδους γραμμικής άλγεβρας. Σε κάθε δεδομένη μεταβολική κατάσταση αντιστοιχεί ένα σύνολο δυνατών κατανομών ροών. Προκειμένου να περιοριστεί ο αριθμός των δυνατών κατανομών εφαρμόζεται η μεθοδολογία της Ανάλυσης Εξισορρόπησης Ροών (Flux Balance Analysis, FBA)(Oberhardt et al. 2009; Varma and Palsson 1994) η οποία χρησιμοποιεί Γραμμική Βελτιστοποίηση για να περιορίσει τις κατανομές των ροών σε αυτές που μεγιστοποιούν μια ορισμένη Αντικειμενική Συνάρτηση (ΑΣ) η οποία διέπει τον κυτταρικό μεταβολισμό (π.χ. αύξηση μακρομορίων βιομάζας σε ένα κύτταρο υπό ανάπτυξη).

### **Ανάλυση Εξισορρόπησης Ροών (Flux Balance Analysis, FBA)**

Το πρώτο βήμα στη μέθοδο FBA είναι η αναπαράσταση του ΜΜΓΚ μέσω του στοιχειομετρικού του πίνακα  $S$ . Ο πίνακας  $S$  έχει διαστάσεις  $m \times n$  και περιέχει τους στοιχειομετρικούς συντελεστές των βιοχημικών αντιδράσεων του μεταβολικού δικτύου. Οι γραμμές του πίνακα  $S$  αντιστοιχούν στους μεταβολίτες, ενώ οι στήλες αντιστοιχούν στις αντιδράσεις. Οι θετικοί στοιχειομετρικοί συντελεστές περιγράφουν την παραγωγή των αντίστοιχων μεταβολιτών, ενώ οι αρνητικοί περιγράφουν την κατανάλωσή τους.

Ο πίνακας  $S$  περιγράφει σχέσεις διατήρησης μάζας. Η ισορροπία μάζας εκφράζεται από την ροή που περνάει μέσα από κάθε αντίδραση και από τη στοιχειομετρία της αντίδρασης. Ειδικότερα:

$$\frac{dx}{dt} = S * v$$

όπου  $v$  είναι το διάνυσμα των ροών που αντιστοιχούν στις αντιδράσεις ή στήλες του πίνακα  $S$  και  $dx$  είναι η μεταβολή της ποσότητας ενός μεταβολίτη προς το χρονικό διάστημα  $dt$ .

Στην σταθερή κατάσταση, ο λόγος  $dx/dt$  γίνεται 0 για όλες τις ροές, αφού οι συγκεντρώσεις των μεταβολιτών δεν μεταβάλλονται. Αυτό ισχύει λόγω της

παραδοχής ότι οι αντιδράσεις επιτελούνται πολύ γρήγορα σε σχέση με το ρυθμό μεταβολής των κυτταρικών φαινοτύπων όπως η κυτταρική αύξηση (συγκέντρωση βιομάζας) (Lee et al. 2006). Συνεπώς η προηγούμενη εξίσωση γίνεται:

$$S * v = 0$$

Η ισορροπία των ροών που εκφράζεται από την παραπάνω σχέση αποτελεί τον κύριο περιορισμό στη μέθοδο FBA. Επιπρόσθετοι περιορισμοί, όπως η αναστρεψιμότητα της αντίδρασης, η ενέργεια ομοιόστασης και άλλοι στοχευμένοι περιορισμοί μπορούν να προστεθούν για να μειωθεί η διαστασιμότητα του προβλήματος και ο χώρος των λύσεων. Το πρόβλημα ανάγεται λοιπόν σε ένα σύστημα γραμμικών εξισώσεων διατήρησης μάζας. Το σύστημα εξισώσεων είναι σε μεγάλο βαθμό υπο-ορισμένο (underdetermined), καθώς οι εξισώσεις είναι πολύ λιγότερες από τους αγνώστους. Συνεπώς, για την επίλυσή του χρησιμοποιείται η τεχνική της γραμμικής βελτιστοποίησης, που επιτυγχάνεται με τον ορισμό μιας αντικειμενικής συνάρτησης. Η αντικειμενική συνάρτηση είναι της μορφής:

$$Z = c * v$$

όπου  $c$  είναι το διάνυσμα των συντελεστών για κάθε μεταβολική ροή του διανύσματος  $v$ . Η βελτιστοποίηση αποσκοπεί στην εύρεση μιας λύσης  $v$ , δηλαδή ενός συνδυασμού μεταβολικών ροών που βελτιστοποιεί τη συνάρτηση  $Z$ , μέσα στα όρια των περιορισμών. Η συνάρτηση  $Z$  είναι γραμμική και επιλύεται εύκολα με γραμμικό προγραμματισμό. Η αντικειμενική συνάρτηση μπορεί να πάρει πολλές μορφές και μπορεί να εκφράζει είτε μια φυσιολογική κυτταρική λειτουργία, είτε μια επιθυμητή ή ύπο διερεύνηση ιδιότητα. Τυπικές αντικειμενικές συναρτήσεις είναι η αύξηση της βιομάζας και η παραγωγή ATP. Στην παρούσια εργασία για τη γραμμική βελτιστοποίηση πραγματοποιήθηκε το Cobra Toolbox 1.3.5 για περιβάλλον Matlab (glpkme solver) (Becker et al. 2007).

### **Ανάλυση Διαφοροποίησης Ροών (Flux variability analysis, FVA)**

Επειδή ο στοιχειομετρικός πίνακας είναι σε μεγάλο βαθμό υπο-ορισμένος, η μέθοδος FBA δεν δίνει μοναδική λύση ως προς τις τιμές των μεταβολικών ροών. Αντίθετα, καταλήγουμε σε πολλαπλές βέλτιστες λύσεις που ουσιαστικά αντιπροσωπεύουν ένα εύρος πιθανών τιμών. Η μέθοδος FVA εξετάζει τα ανώτατα και κατώτατα όρια τιμών κάθε αντίδρασης ώστε να επιτυγχάνεται ένα

προκαθορισμένο ποσοστό ελάχιστης τιμής της αντικειμενικής συνάρτησης. Το πρόβλημα επιλύεται με την εφαρμογή πολλαπλών βελτιστοποιήσεων FBA, όπου ορίζεται μια ελάχιστη τιμή της αντικειμενικής συνάρτησης και διερευνάται για κάθε ροή η ανώτατη και η ελάχιστη τιμή της που αντιστοιχεί στα επιθυμητά όρια της αντικειμενικής συνάρτησης (Mahadevan and Schilling, 2003).

### **1.2.3 Συστημική ανάλυση της κατανομής ροών σταθερής κατάστασης**

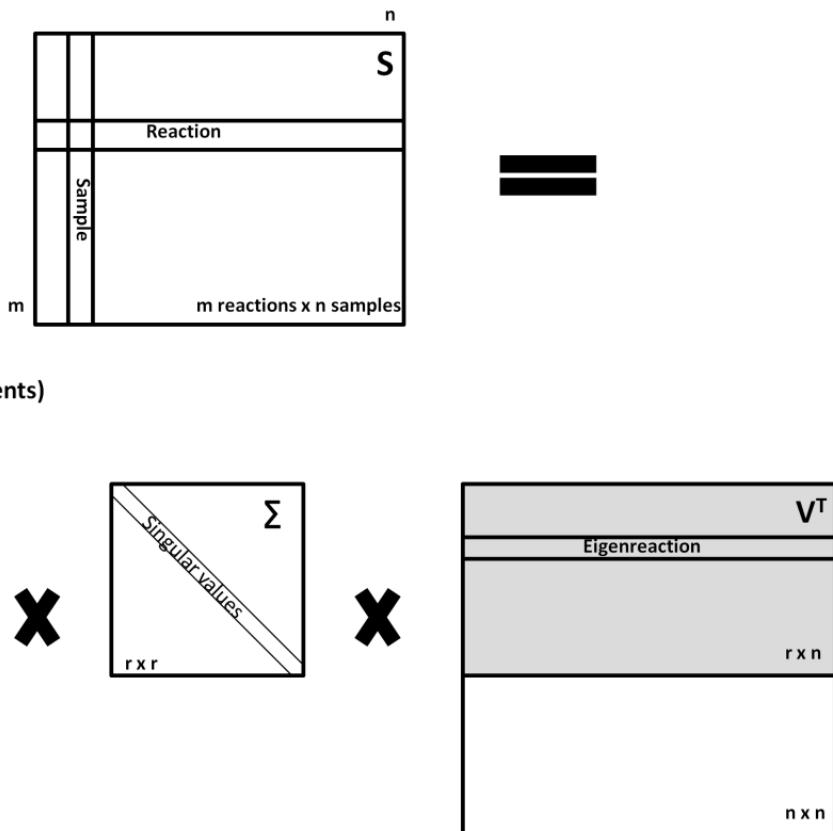
Τα μεταβολικά μοντέλα μεγάλης κλίμακας αποτελούν υπο-ορισμένα συστήματα (underdetermined), που περιλαμβάνουν πολλούς βαθμούς ελευθερίας και συνεπώς δεν επιδέχονται μοναδική επίλυση. Αυτό συμβαίνει επειδή πολύ λίγοι ρυθμοί μεταβολής αντιδράσεων είναι γνωστοί και οι οποίοι συνήθως αφορούν τις ανταλλαγές με το θρεπτικό μέσο, δηλαδή τις εξωτερικές ροές, αφού οι εσωτερικές ροές είναι δύσκολα μετρήσιμες. Συνεπώς, η μέθοδος της γραμμικής βελτιστοποίησης που χρησιμοποιήθηκε για την FBA δεν παρέχει μοναδική λύση για τις τιμές των μεταβολικών ροών. Ως εκ τούτου, η ανάλυση του συνολικού χώρου των λύσεων (χώρος ροών, flux space) μπορεί να παράσχει πολλές περισσότερες πληροφορίες για τις δυναμικές ιδιότητες του μεταβολικού δικτύου και ιδίως για το πώς αυτό ρυθμίζεται. Μια αποτελεσματική προσέγγιση για την εξερεύνηση του χώρου ροών αποτελεί η ομοιόμορφη τυχαία δειγματοληψία (uniform random sampling) του χώρου των λύσεων ακολουθούμενη από Singular Value Decomposition (SVD) του πίνακα των δειγμάτων (sample matrix) (Barrett et al. 2006). Η μέθοδος αυτή, που περιγράφηκε στην προηγούμενη ενότητα, ως μέθοδος μείωσης της διαστασιμότητας ενός πίνακα, είναι κατάλληλη και για την ανάλυση του πίνακα δειγμάτων με διαφορετική φυσική σημασία. Στην περίπτωση της SVD στον πίνακα δειγμάτων, ανακτούμε ορθοκανονικά ιδιοδιανύσματα (eigenvectors) βιοχημικών αντιδράσεων, ανάγοντας το δίκτυο σε μικρότερες ομάδες αντιδράσεων, η μεταβλητότητα των οποίων ευθύνεται για όλες τις συνολικότερες κατανομές ροών που παρατηρούνται. Οι ομάδες αντιδράσεων αυτές (ιδιοδιανύσματα) ονομάζονται Πρωτεύουσες συνιστώσες (Principal Components) του δικτύου καθώς η μεταβλητότητά τους καθορίζει τη δυναμική συμπεριφορά του μεταβολικού δικτύου να ανακατευθύνει τις μεταβολικές ροές. Αντίστοιχα, κάθε Πρωτεύουσα Συνιστώσα χαρακτηρίζεται από τους βαθμούς συνεισφορών των αντιδράσεων που την αποτελούν. Οι αντιδράσεις με υψηλή τιμή συνεισφοράς ελέγχουν ρυθμιστικά το συνολικό μεταβολικό δίκτυο και αποτελούν πιθανούς ρυθμιστικούς στόχους. Συνεπώς, η μεταβλητότητα των ροών του μοντέλου

εκφράζεται ως ο γραμμικός συνδυασμός μιας σειράς ανεξάρτητων μεταβλητών, δηλαδή των πρωτευουσών συνιστωσών, οι οποίες συσχετίζονται με τις αρχικές μεταβλητές μέσω ορθοκανονικής μετατροπής (Jolliffe 2002). Εφόσον ένα πολυμεταβλητό σύστημα, όπως το σύνολο των πιθανών ροών, αναπαριστάται ως ένα σύνολο συντεταγμένων σε χώρο πολλών διαστάσεων (ένας άξονας ανά μεταβλητή), η ανάλυση πρωτευουσών συνιστωσών επιτρέπει τη μείωση της διαστασιμότητας και την ανάδειξη των σημείων του χώρου με τη μεγαλύτερη πληροφορία, δηλαδή των πιο σημαντικών ροών. Η μεταβλητότητα των ροών αυτών ελέγχει την μεταβλητότητα όλου του συστήματος, καθώς η συνολική διαφοροποίηση ανάγεται στη διαφοροποίηση των πρωτευουσών συνιστωσών. Συμπερασματικά, μπορεί να θεωρηθεί ότι οι πρωτεύουσες συνιστώσες εκφράζουν διακριτούς και ανεξάρτητους φαινότυπους, οι οποίοι καθορίζονται από τους βαθμούς συνεισφορών των αντιδράσεων.

Η παραπάνω μέθοδος εφαρμόστηκε με το πρόγραμμα COBRA Toolbox 1.3.5 (Becker et al. 2007). Αρχικά πραγματοποιήθηκε δειγματοληψία που περιέλαβε συνολικά 20M βήματα και 50K δείγματα (samples). Στη συνέχεια, στον πίνακα δειγμάτων που προέκυψε από τη δειγματοληψία, εφαρμόστηκε η μέθοδος SVD που οδήγησε στην ανάκτηση των ιδιοδιανυσμάτων (Σχήμα 1.2).

### **Ανάλυση Ιδιαζουσών Τιμών του πίνακα δειγμάτων**

Ο πίνακας δειγμάτων είναι διαστάσεων  $m$  αντιδράσεων  $\times n$  δειγμάτων (samples), δηλαδή των διαφορετικών τιμών που παίρνει μια βιοχημική ροή κατά την εκτέλεση της τυχαίας δειγματοληψίας. Οι στήλες του πίνακα  $U$  αποτελούν τα ιδιοδιανύσματα των οποίων η μεταβλητότητα εκφράζει όλες τις καταστάσεις του μεταβολικού δικτύου (Σχήμα 1.2). Τα ιδιοδιανύσματα είναι ορθοκανονικά και άρα γραμμικά ανεξάρτητα μεταξύ τους. Κάθε ιδιοδιάνυσμα αποτελείται από τις τιμές συμμετοχής (loadings) των αντιδράσεων. Οι τιμές της διαγωνίου του πίνακα  $\Sigma$  λέγονται ιδιάζουσες ή ιδιοτιμές και εκφράζουν την σχετική συμμετοχή του κάθε ιδιοδιανύσματος στη μεταβλητότητα του συστήματος.



**Σχήμα 1.2.** Ανάλυση Ιδιαζουσών Τιμών του πίνακα δειγμάτων  $S$  που προκύπτει από ομοιόμορφη τυχαία δειγματοληψία (uniform random sampling) του χώρου των λύσεων.

### Ρυθμιστικός Βαθμός Αντιδράσεων

Ο Ρυθμιστικός Βαθμός μιας αντίδρασης είναι μια τιμή που κυμαίνεται από 0 εώς 1 και εκφράζει τη συνολική συμμετοχή μιας αντίδρασης στη μεταβλητότητα του μεταβολικού δικτύου λαμβάνοντας υπόψη τη συμμετοχή της σε όλες τις πρωτεύουσες συνιστώσες. Η τιμή αυτή υπολογίζεται ως η  $L^2$ -norm του διανύσματος των συνεισφορών  $x$  (loadings) της αντίδρασης:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$|x| = \sqrt{\sum_{k=1}^n |x^k|^2}$$

## 1.3 Αποτελέσματα

### 1.3.1 Μοντελοποίηση του κεντρικού μεταβολισμού της ελαιοκράμβης

Το τελικό μοντέλο αποτελείται από 313 αντιδράσεις και 262 μεταβολίτες (Πίνακας 1.4). Η λειτουργικότητα του μοντέλου και η ικανότητά του προσομοίωσης μεταβολικών καταστάσεων επιβεβαιώθηκε μέσω της μεθόδου FBA. Συγκεκριμένα, υπολογίστηκαν οι μεταβολικές καταστάσεις ανάπτυξης υπό αερόβιες, αναερόβιες και αυτότροφες συνθήκες. Το σημαντικότερο αρχικό βήμα επιβεβαίωσης ενός υπολογιστικού μεταβολικού μοντέλου είναι η δυνατότητά του να προβλέπει έναν ρεαλιστικό ρυθμό παραγωγής βιομάζας (growth rate) (Thiele and Palsson 2010). Η γενική παραδοχή είναι ότι, αν το μοντέλο περιέχει κάποιο σημαντικό λάθος στην περιγραφή του μεταβολισμού, αυτό θα έχει σίγουρα επίπτωση στον ρυθμό παραγωγής βιομάζας, ο οποίος θα αποκλίνει υπερβολικά από πειραματικές παρατηρήσεις. Στην προκειμένη περίπτωση, πραγματοποιήθηκε προσομοίωση των συνθηκών καλλιέργειας των (Schwender et al. 2006) (βλέπε Μεθοδολογία) που περιλαμβάνουν θρεπτικό μέσο με σακχαρόζη, αλανίνη και γλουταμίνη. Επειδή τα έμβρυα της ελαιοκράμβης είναι πράσινα και μικρή ποσότητα φωτός φτάνει σε αυτά, η προσομοίωση έγινε δύο φορες, με ύπαρξη και χωρίς ύπαρξη φωτοσύνθεσης αντίστοιχα. Η φωτοσύνθεση εκφράζεται ως μια μη περιορισμένη ή περιορισμένη στο 0 βιοχημική αντίδραση, που περιλαμβάνει τα φωτοσυστήματα I και II, καθώς και την ATP συνθετάση του χλωροπλάστη. Ο ρυθμός ανάπτυξης βιομάζας με με περιορισμένη φωτοσύνθεση υπολογίστηκε σε  $0.018 \text{ h}^{-1}$ . Η τιμή αυτή είναι απόλυτα ρεαλιστική και βρίσκεται πολύ κοντά στην πειραματική τιμή που είναι  $0.014 \text{ h}^{-1}$ . Εν τούτοις, η προσομοίωση χωρίς φωτοσύνθεση που είναι πιο κοντά στις πραγματικές συνθήκες ανάπτυξης των εμβρύων, καθώς η φωτοσύνθεση συντελείται περιορισμένα, έδωσε ακριβώς την πειραματική τιμή  $0.014 \text{ h}^{-1}$ . Το αποτέλεσμα αυτό δικαιώνει την ορθότητα του μοντέλου ως προς την βιοχημική περιγραφή του μεταβολισμού της ελαιοκράμβης.

**Πίνακας 1.4.** Οι αντιδράσεις του μεταβολικού μοντέλου της ελαιοκράμβης

Ένζυμο	Όνομα αντίδρασης	(=> : αμφίδρομη, -> μονόδρομη)
-	<i>ATP_Maintenance</i>	$ATP_c + H_2O_c \rightarrow ADP_c + H_c + phosphate_c$
(exchange)	<i>EX_UDP_glucose</i>	$UDP_D\_glucose_c \rightarrow UDP_D\_glucose_e$
(exchange)	<i>EX_acetate</i>	$acetate_c \rightarrow acetate_e$
(exchange)	<i>EXAlanine</i>	$L\_glutamine_c + H_c \Rightarrow L\_glutamine_e + H_e$
(exchange)	<i>EX_alpha_D_glucose</i>	$alpha\_D\_glucose_c + H_c \Rightarrow alpha\_D\_glucose_e + H_e$
(exchange)	<i>EX_ammonia_c</i>	$ammonia_c \Rightarrow ammonia_e$
(exchange)	<i>EX_biomassIg</i>	$biomass1g \rightarrow biomass1g_e$
(exchange)	<i>EX_CO2</i>	$CO2_c \Rightarrow CO2_e$
(exchange)	<i>EX_ethanol</i>	$ethanol_c \rightarrow ethanol_e$
(exchange)	<i>EX_formate</i>	$formate_p \rightarrow formate_e$
(exchange)	<i>EX_glutamine</i>	$L\_alanine_c + H_c \Rightarrow L\_alanine_e + H_e$
(exchange)	<i>EX_lactate</i>	$lactate_c \rightarrow lactate_e$
(exchange)	<i>EX_nitricOxide</i>	$nitricOxide_c \rightarrow nitricOxide_e$
(exchange)	<i>EX_NO3</i>	$NO3_p \Rightarrow NO3_e$
(exchange)	<i>EX_O2</i>	$O2_c \Rightarrow O2_e$
(exchange)	<i>EX_phosphate</i>	$phosphate_c \Rightarrow phosphate_e$
(exchange)	<i>EX_sucrose</i>	$sucrose_c + H_c \Rightarrow sucrose_e + H_e$
(exchange)	<i>EX_sulfate</i>	$sulfate_p \Rightarrow sulfate_e$
Starch synthesis	<i>J_starch</i>	$0.89489 ADP_D\_glucose_p \rightarrow 0.89489 ADP_p + starch$
fatty acid synthase (c16)	<i>JFAc16</i>	$acetyl_CoA_p + 7 malonyl_CoA_p + 14 H_p + 14 NADPH_p \Rightarrow FAc16 + 7 CO2_p +$

---

fatty acid synthase (c18)	<i>JFAc18</i>	9 coenzyme_A_p + 14 NADP_p FAC16 + malonyl_CoA_p + 2 H_p + 2 NADPH_p => FAc18 + CO2_p + coenzyme_A_p + 2 NADP_p
Pyruvate Dehydrogenase Complex	<i>Jpyr_dehydrog_complex_p</i>	pyruvate_p + coenzyme_A_p + NAD_p => acetyl_CoA_p + CO2_p + H_p + NADH_p
histidinol dehydrogenase	<i>RXN_I_I_1_23_p</i>	histidinol_p + NAD_p => H_p + NADH_p + histidinal_p
histidinol dehydrogenase	<i>RXN_I_I_1_23_p2</i>	H2O_c + histidinal_p + NAD_p => L_histidine_p + H_p + NADH_p
shikimate dehydrogenase	<i>RXN_I_I_1_25_p</i>	H_p + NADPH_p + _3_dehydro_shikimate_p => shikimate_p + NADP_p
glycerate dehydrogenase	<i>RXN_I_I_1_29_p</i>	H_p + NADH_p + hydroxypyruvate_p => NAD_p + glycerate_p
homoserine dehydrogenase	<i>RXN_I_I_1_3_p</i>	L_aspartate_semaldehyde_p + NADPH_p + H_p => NADP_p + homoserine_p
malate dehydrogenase	<i>RXN_I_I_1_38_p</i>	malate_p + NAD_p => CO2_p + H_p + NADH_p + pyruvate_p
phosphogluconate dehydrogenase	<i>RXN_I_I_1_44_p</i>	NADP_p + _6_phospho_D_gluconate_p => D_ribulose_5_phosphate_p + CO2_p + H_p + NADPH_p
glucose-6-phosphate dehydrogenase	<i>RXN_I_I_1_49_p</i>	NADP_p + alpha_D_glucose_6_phosphate_p => H_p + NADPH_p + D_glucono_delta_lactone_6_phosphate_p
3-isopropylmalate dehydrogenase	<i>RXN_I_I_1_85_p</i>	_3_isopropylmalate_p + NAD_p => NADH_p + _2_isopropyl_3_oxosuccinate_p + H_p
ketol-acid reductoisomerase	<i>RXN_I_I_1_86_p</i>	H_p + NADPH_p + _2_acetolactate_p + H_p => NADP_p + _2_3_dihydroxy_isovalerate_p
ketol-acid reductoisomerase	<i>RXN_I_I_1_86_p2</i>	_2_aceto_2_hydroxy_butyrate_p + H_p + NADPH_p => NADP_p + _2_3_dihydroxy_3_methylvalerate_p
phosphoglycerate dehydrogenase	<i>RXN_I_I_1_95_p</i>	_3_phosphoglycerate_p + NAD_p => H_p + NADH_p + _3_phospho_hydroxypyruvate_p
glycolate dehydrogenase	<i>RXN_I_I_99_14_p</i>	glycolate_p + NAD_p => H_p + NADH_p + glyoxylate_p
catalase	<i>RXN_I_II_1_6_p</i>	2 H2O2_p => O2_p + 2 H2O_c
aspartate-semaldehyde dehydrogenase	<i>RXN_I_2_I_11_p</i>	H_p + NADPH_p + L_aspartyl_4_phosphate_p => L_aspartate_semaldehyde_p + NADP_p + phosphate_p

---

glyceraldehyde-3-phosphate dehydrogenase	<i>RXN_1_2_1_12_p</i>	D_glyceraldehyde_3_phosphate_p + phosphate_p + NAD_p => H_p + NADH_p + _1_3_diphosphateglycerate_p
glycolaldehyde dehydrogenase	<i>RXN_1_2_1_21_p</i>	H2O_c + NAD_p + glycolaldehyde_p => H_p + NADH_p + glycolate_p
glutamate-5-semialdehyde dehydrogenase	<i>RXN_1_2_1_41_p</i>	NADPH_p + L_glutamate_5_phosphate_p + H_p => NADP_p + L_glutamate_gamma_semialdehyde_p + phosphate_p
formate dehydrogenase	<i>RXN_1_2_1_43_p</i>	NADP_p + formate_p => H_p + NADPH_p + CO2_p
dihydrodipicolinate reductase	<i>RXN_1_3_1_26_p</i>	L_2_3_dihydrodipicolinate_p + NADPH_p + H_p => tetrahydrodipicolinate_p + NADP_p
arogenate dehydrogenase	<i>RXN_1_3_1_43_p</i>	NADP_p + L_arogenate_p => H_p + NADPH_p + L_tyrosine_p + CO2_p
L-aspartate oxidase	<i>RXN_1_4_3_16_p</i>	O2_p + L_aspartate_p => H2O2_p + iminoaspartate_p + H_p
glutamate synthase (ferredoxin)	<i>RXN_1_4_7_1_p</i>	2 redferredoxin_p + L_glutamine_p + alpha_ketoglutarate_p => 2 oxferredoxin_p + 2 L_glutamate_p
pyrroline-5-carboxylate reductase	<i>RXN_1_5_1_2_p</i>	pyrroline_5_carboxylate_p + NADPH_p + H_p => NADP_p + L_proline_p
methylenetetrahydrofolate reductase	<i>RXN_1_5_1_20_p</i>	H_p + NADH_p + _5_10_methylene_THF_p => _5_methyl_THF + NAD_p
dihydrofolate reductase	<i>RXN_1_5_1_3_p</i>	H_p + NADPH_p + _7_8_dihydrofolate_p => NADP_p + tetrahydrofolate_p
methylenetetrahydrofolate reductase (ferredoxin)	<i>RXN_1_5_7_1_p</i>	2 redferredoxin_p + _5_10_methylene_THF_p + 2 H_p => 2 oxferredoxin_p + _5_methyl_THF
nitrate reductase	<i>RXN_1_7_1_1_p</i>	NO3_p + NADH_p + H_p -> NAD_p + nitrite_p + H2O_c
ferredoxin-nitrite reductase	<i>RXN_1_7_7_1_p</i>	7 H_p + 6 redferredoxin_p + nitrite_p -> 2 H2O_c + 6 oxferredoxin_p + ammonia_p
sulfite oxidase	<i>RXN_1_8_3_1_p</i>	sulfite_p + O2_p + H2O_c => H2O2_p + sulfate_p
adenylyl-sulfate reductase (glutathione)	<i>RXN_1_8_4_9_p</i>	adenosine_5_Psulfate_p + 2 glutathione_p => 2 H_p + glutathione_disulfide_p + AMP_p + sulfite_p
sulfite reductase	<i>RXN_1_8_7_1_p</i>	sulfite_p + 3 redferredoxin_p => 3 oxferredoxin_p + 3 H2O_c + hydrogen_sulfide_p
methionine synthase	<i>RXN_2_1_1_13_p</i>	_5_methyl_THF_p + L_homocysteine_p => 2 H_p + tetrahydrofolate_p + L_methionine_p
methionine synthase	<i>RXN_2_1_1_14_p</i>	L_homocysteine_p => L_methionine_p

glycine hydroxymethyltransferase	<i>RXN_2_1_2_1_p</i>	$2 \text{H\_p} + \text{tetrahydrofolate\_p} + \text{L\_serine\_p} \Rightarrow \text{H}_2\text{O\_c} + \text{5\_10\_methylene\_THF\_p} + \text{glycine\_p}$
3-methyl-2-oxobutanoate hydroxymethyltransferase	<i>RXN_2_1_2_11_p</i>	$\text{2\_dehydropantoate\_p} + \text{tetrahydrofolate\_p} + 2 \text{H\_p} \Rightarrow \text{H}_2\text{O\_c} + \text{2\_keto\_isovalerate\_p} + \text{5\_10\_methylene\_THF\_p}$
transketolase	<i>RXN_2_2_1_1_p</i>	$\text{D\_xylulose\_5\_phosphate\_p} + \text{D\_erythrose\_4\_phosphate\_p} \Rightarrow \text{D\_glyceraldehyde\_3\_phosphate\_p} + \text{D\_fructose\_6\_phosphate\_p}$
transketolase	<i>RXN_2_2_1_1_p</i>	$\text{D\_ribose\_5\_phosphate\_p} + \text{D\_xylulose\_5\_phosphate\_p} \Rightarrow \text{D\_sedoheptulose\_7\_phosphate\_p} + \text{D\_glyceraldehyde\_3\_phosphate\_p}$
transaldolase	<i>RXN_2_2_1_2_p</i>	$\text{D\_sedoheptulose\_7\_phosphate\_p} + \text{D\_glyceraldehyde\_3\_phosphate\_p} \Rightarrow \text{D\_fructose\_6\_phosphate\_p} + \text{D\_erythrose\_4\_phosphate\_p}$
acetolactate synthase	<i>RXN_2_2_1_6_p</i>	$2 \text{pyruvate\_p} \Rightarrow \text{2\_acetolactate\_p} + \text{CO}_2\text{\_p}$
acetolactate synthase	<i>RXN_2_2_1_6_p2</i>	$\text{pyruvate\_p} + \text{2\_oxobutanoate\_p} \Rightarrow \text{2\_aceto\_2\_hydroxy\_butyrate\_p} + \text{CO}_2\text{\_p}$
serine O-acetyltransferase	<i>RXN_2_3_1_30_p</i>	$\text{L\_serine\_p} + \text{acetyl\_CoA\_p} \Rightarrow \text{O\_acetyl\_L\_serine\_p} + \text{coenzyme\_A\_p}$
gamma-glutamyltransferase	<i>RXN_2_3_2_2_p</i>	$\text{H}_2\text{O\_c} + \text{glutathione\_p} \rightarrow \text{L\_glutamate\_p} + \text{L\_cysteinylglycine\_p}$
2-isopropylmalate synthase	<i>RXN_2_3_3_13_p</i>	$\text{H}_2\text{O\_c} + \text{2\_keto\_isovalerate\_p} + \text{acetyl\_CoA\_p} \Rightarrow \text{coenzyme\_A\_p} + \text{2\_isopropylmalate\_p}$
ATP citrate synthase	<i>RXN_2_3_3_8_p</i>	$\text{citrate\_p} + \text{ATP\_p} + \text{coenzyme\_A\_p} \rightarrow \text{oxaloacetate\_p} + \text{ADP\_p} + \text{acetyl\_CoA\_p} + \text{phosphate\_p}$
imidazoleglycerol phosphate synthase	<i>RXN_2_4_2____p</i>	$\text{L\_glutamine\_p} + \text{phosphoribulosylformimino\_AICAR\_P\_p} \Rightarrow \text{D\_erythro\_imidazole\_glycerol\_phosphate\_p} + \text{L\_glutamate\_p}$
ATP phosphoribosyltransferase	<i>RXN_2_4_2_17_p</i>	$\text{ATP\_p} + \text{5\_phosphoribosyl\_1\_pyrophosphate\_p} \Rightarrow \text{phosphoribosyl\_ATP\_p} + \text{diphosphate\_p}$
anthranilate phosphoribosyltransferase	<i>RXN_2_4_2_18_p</i>	$\text{5\_phosphoribosyl\_1\_pyrophosphate\_p} + \text{anthranilate\_p} \Rightarrow \text{N\_5\_phosphoribosyl\_anthranilate\_p} + \text{diphosphate\_p}$
nicotinate-nucleotide diphosphorylase	<i>RXN_2_4_2_19_p</i>	$\text{quinolinate\_p} + \text{5\_phosphoribosyl\_1\_pyrophosphate\_p} \Rightarrow \text{nicotinic\_acid\_mononucleotide\_p} + \text{CO}_2\text{\_p} + \text{diphosphate\_p}$

transferase	<i>RXN_2_5_1_p</i>	O_phospho_L_homoserine_p + L_cysteine_p => cystathionine_p + phosphate_p
dihydropteroate synthase	<i>RXN_2_5_1_15_p</i>	_2_amino_4_hydroxy_6_hydroxymethyl_7_8_dihydropteridine_diphosphate_p + p_aminobenzoate_p => _7_8_dihydropteroate_p + diphosphate_p
3-phosphoshikimate 1-carboxyvinyltransferase	<i>RXN_2_5_1_19_p</i>	shikimate_3_phosphate_p + phosphoenolpyruvate_p => _5_enolpyruvyl_shikimate_3_phosphate_p + phosphate_p
cysteine synthase	<i>RXN_2_5_1_47_p</i>	O_acetyl_L_serine_p + hydrogen_sulfide_p => acetate_p + L_cysteine_p
3-deoxy-7-phosphoheptulonate synthase	<i>RXN_2_5_1_54_p</i>	H2O_c + D_erythrose_4_phosphate_p + phosphoenolpyruvate_p => phosphate_p + _3_deoxy_D_arabino_heptulosonate_7_phosphate_p
quinolinate synthase	<i>RXN_2_5_1_72_p</i>	iminoaspartate_p + dihydroxyacetone_phosphate_p => quinolinate_p + 2 H2O_c + phosphate_p
Transaminase	<i>RXN_2_6_1_p</i>	tetrahydروpicolinate_p + H2O_c + L_glutamate_p => L_L_diaminopimelate_p + alpha_ketoglutarate_p
aspartate transaminase	<i>RXN_2_6_1_1_p</i>	oxaloacetate_p + L_glutamate_p => alpha_ketoglutarate_p + L_aspartate_p
alanine transaminase	<i>RXN_2_6_1_2_p</i>	LAlanine_p + alpha_ketoglutarate_p => pyruvate_p + L_glutamate_p
glycine transaminase	<i>RXN_2_6_1_4_p</i>	glyoxylate_p + L_glutamate_p => glycine_p + alpha_ketoglutarate_p
branched-chain-amino-acid transaminase	<i>RXN_2_6_1_42_p</i>	L_valine_p + alpha_ketoglutarate_p => _2_keto_isovalerate_p + L_glutamate_p
branched-chain-amino-acid transaminase	<i>RXN_2_6_1_42_p</i>	L_isoleucine_p + alpha_ketoglutarate_p => L_glutamate_p + _2_keto_3_methyl_valerate_p
branched-chain-amino-acid transaminase	<i>RXN_2_6_1_42_p2</i>	L_leucine_p + alpha_ketoglutarate_p => _2_ketoisopropane_p + L_glutamate_p
serine-glyoxylate transaminase	<i>RXN_2_6_1_45_p</i>	L_serine_p + glyoxylate_p => glycine_p + hydroxypyruvate_p
phosphoserine transaminase	<i>RXN_2_6_1_52_p</i>	L_glutamate_p + _3_phospho_hydroxypyruvate_p => _3_phospho_serine_p + alpha_ketoglutarate_p
glutamate-prephenate aminotransferase	<i>RXN_2_6_1_79_p</i>	L_glutamate_p + prephenate_p => L_rogenate_p + alpha_ketoglutarate_p
histidinol-phosphate transaminase	<i>RXN_2_6_1_9_p</i>	imidazole_acetol_phosphate_p + L_glutamate_p => L_histidinol_phosphate_p + alpha_ketoglutarate_p
hexokinase	<i>RXN_2_7_1_1_p</i>	ATP_p + alpha_D_glucose_p -> 2 H_p + alpha_D_glucose_6_phosphate_p + ADP_p

6-phosphofructokinase	<i>RXN_2_7_1_11_p</i>	ATP_p + D_fructose_6_phosphate_p -> ADP_p + fructose_1_6_bisphosphate_p
phosphoribulokinase	<i>RXN_2_7_1_19_p</i>	ATP_p + D_ribulose_5_phosphate_p -> D_ribulose_1_5_bisphosphate_p + ADP_p
glycerate kinase	<i>RXN_2_7_1_31_p</i>	ATP_p + glycerate_p -> ADP_p + _3_phosphoglycerate_p
homoserine kinase	<i>RXN_2_7_1_39_p</i>	ATP_p + homoserine_p -> O_phospho_L_homoserine_p + ADP_p
pyruvate kinase	<i>RXN_2_7_1_40_p</i>	ADP_p + phosphoenolpyruvate_p -> ATP_p + pyruvate_p
shikimate kinase	<i>RXN_2_7_1_71_p</i>	shikimate_p + ATP_p -> shikimate_3_phosphate_p + ADP_p
diphosphate-fructose-6-phosphate 1-phosphotransferase	<i>RXN_2_7_1_90_p</i>	D_fructose_6_phosphate_p + diphosphate_p => phosphate_p + fructose_1_6_bisphosphate_p
glutamate 5-kinase	<i>RXN_2_7_2_11_p</i>	ATP_p + L_glutamate_p -> L_glutamate_5_phosphate_p + ADP_p
phosphoglycerate kinase	<i>RXN_2_7_2_3_p</i>	ATP_p + _3_phosphoglycerate_p -> _1_3_diphosphoglycerate_p + ADP_p
aspartate kinase	<i>RXN_2_7_2_4_p</i>	ATP_p + L_aspartate_p -> ADP_p + L_aspartyl_4_phosphate_p
adenylate kinase	<i>RXN_2_7_4_3_p</i>	AMP_p + ATP_p => 2 ADP_p
nucleoside-diphosphate kinase	<i>RXN_2_7_4_6_p</i>	ATP_p + GDP_p => H_p + GTP_p + ADP_p
ribose-phosphate diphosphokinase	<i>RXN_2_7_6_1_p</i>	ATP_p + D_ribose_5_phosphate_p -> AMP_p + _5_phosphoribosyl_1_pyrophosphate_p + 2 H_p
2-amino-4-hydroxy-6-hydroxymethylidihydropteridine diphosphokinase	<i>RXN_2_7_6_3_p</i>	ATP_p + _6_hydroxymethyl_dihydropterin_p -> AMP_p + _2_amino_4_hydroxy_6_hydroxymethyl_7_8_dihydropteridine_diphosphate_p
nicotinate-nucleotide adenylyltransferase	<i>RXN_2_7_7_18_p</i>	ATP_p + nicotinic_acid_mononucleotide_p -> nicotinic_acid_adenine_dinucleotide_p + diphosphate_p
glucose-1-phosphate adenylyltransferase	<i>RXN_2_7_7_27_p</i>	ATP_p + alpha_D_glucose_1_phosphate_p -> ADP_D_glucose_p + diphosphate_p
sulfate adenylyltransferase	<i>RXN_2_7_7_4_p</i>	ATP_p + sulfate_p -> diphosphate_p + adenosine_5_Psulfate_p
-	<i>RXN_238I</i>	indole_3_glycerol_phosphate_p => D_glyceraldehyde_3_phosphate_p + indole_p
6-phosphogluconolactonase	<i>RXN_3_1_1_31_p</i>	H2O_c + D_glucono_delta_lactone_6_phosphate_p => H_p + _6_phospho_D_gluconate_p
histidinol-phosphatase	<i>RXN_3_1_3_15_p</i>	H2O_c + L_histidinol_phosphate_p => phosphate_p + histidinol_p
phosphoglycolate phosphatase	<i>RXN_3_1_3_18_p</i>	H2O_c + _2_phosphoglycolate_p => glycolate_p + phosphate_p

phosphoserine phosphatase	<i>RXN_3_I_3_3_p</i>	$H_2O_c + \text{L\_phospho\_serine\_p} \Rightarrow \text{L\_serine\_p} + \text{phosphate\_p}$
NMN nucleosidase	<i>RXN_3_2_2_14_p</i>	$H_2O_c + \text{nicotinamide\_mononucleotide\_p} \Rightarrow \text{H\_p} + \text{nicotinamide\_p} + \text{D\_ribose\_5\_phosphate\_p}$
alanyl aminopeptidase	<i>RXN_3_4_11_2_p</i>	$H_2O_c + \text{L\_cysteinylglycine\_p} \rightarrow \text{glycine\_p} + \text{L\_cysteine\_p}$
glutaminase	<i>RXN_3_5_1_2_p</i>	$H_2O_c + \text{L\_glutamine\_p} \Rightarrow \text{ammonia\_p} + \text{L\_glutamate\_p}$
GTP cyclohydrolase I	<i>RXN_3_5_4_16_p</i>	$H_2O_c + \text{GTP\_p} \rightarrow \text{formate\_p} + \text{7\_8\_dihydronopterin\_3\_triphosphate\_p}$
phosphoribosyl-AMP cyclohydrolase	<i>RXN_3_5_4_19_p</i>	$H_2O_c + \text{phosphoribosyl\_AMP\_p} \Rightarrow \text{phosphoribosylformiminoAICAR\_phosphate\_p}$
hydrolase	<i>RXN_3_6_1_p</i>	$H_2O_c + \text{7\_8\_dihydronopterin\_3\_triphosphate\_p} \Rightarrow \text{dihydronopterin\_phosphate\_p} + \text{diphosphate\_p}$
hydrolase	<i>RXN_3_6_1_p2</i>	$H_2O_c + \text{dihydronopterin\_phosphate\_p} \Rightarrow \text{dihydro\_neo\_pterin\_p} + \text{phosphate\_p}$
inorganic diphosphatase	<i>RXN_3_6_1_1_p</i>	$H_2O_c + \text{diphosphate\_p} \rightarrow 2 \text{ phosphate\_p}$
phosphoribosyl-ATP diphosphatase	<i>RXN_3_6_1_31_p</i>	$H_2O_c + \text{phosphoribosyl\_ATP\_p} \Rightarrow \text{diphosphate\_p} + \text{phosphoribosyl\_AMP\_p}$
carboxylyase (photorespiration)	<i>RXN_4_I_1_p</i>	$D\_ribulose\_1\_5\_bisphosphate\_p + O_2\_p \rightarrow \text{3\_phosphoglycerate\_p} + \text{2\_phosphoglycolate\_p} + H\_p$
Diaminopimelate decarboxylase	<i>RXN_4_I_1_20_p</i>	$\text{meso\_diaminopimelate\_p} \Rightarrow \text{L\_lysine\_p} + CO_2\_p$
Ribulose-bisphosphate carboxylase	<i>RXN_4_I_1_39_p</i>	$D\_ribulose\_1\_5\_bisphosphate\_p + CO_2\_p + H_2O_c \Rightarrow 2 \text{ 3\_phosphoglycerate\_p} + 2 \text{ H\_p}$
indole-3-glycerol-phosphate synthase	<i>RXN_4_I_1_48_p</i>	$\text{1\_o\_carboxyphenylamino\_1\_deoxyribulose\_5\_phosphate\_p} \Rightarrow \text{indole\_3\_glycerol\_phosphate\_p} + H_2O_c + CO_2\_p$
Fructose-bisphosphate aldolase	<i>RXN_4_I_2_13_p</i>	$\text{fructose\_1\_6\_bisphosphate\_p} \Rightarrow D\_glyceraldehyde\_3\_phosphate\_p + \text{dihydroxyacetone\_phosphate\_p}$
Dihydronopterin aldolase	<i>RXN_4_I_2_25_p</i>	$\text{dihydro\_neo\_pterin\_p} \Rightarrow \text{glycolaldehyde\_p} + \text{6\_hydroxymethyl\_dihydropterin\_p}$
Threonine aldolase	<i>RXN_4_I_2_5_p</i>	$L\_threonine\_p \Rightarrow \text{acetaldehyde\_p} + \text{glycine\_p}$
anthranilate synthase	<i>RXN_4_I_3_27_p</i>	$\text{chorismate\_p} + \text{L\_glutamine\_p} \Rightarrow \text{anthranilate\_p} + \text{pyruvate\_p} + \text{L\_glutamate\_p}$
aminodeoxychorismate lyase	<i>RXN_4_I_3_38_p</i>	$\text{4\_amino\_4\_deoxychorismate\_p} \Rightarrow p\_aminobenzoate\_p + \text{pyruvate\_p}$
carbonate dehydratase	<i>RXN_4_2_I_1_p</i>	$H_2O_c + CO_2\_p \Rightarrow HCO_3\_p + H\_p$

3-dehydroquinate dehydratase	<i>RXN_4_2_1_10_p</i>	$_3\text{dehydroquinate}_p \Rightarrow H2O_c + _3\text{dehydro_shikimate}_p$
phosphopyruvate hydratase	<i>RXN_4_2_1_11_p</i>	$_2\text{phosphoglycerate}_p \Rightarrow H2O_c + \text{phosphoenolpyruvate}_p$
imidazoleglycerol-phosphate dehydratase	<i>RXN_4_2_1_19_p</i>	$D\text{erythro_imidazole_glycerol_phosphate}_p \Rightarrow H2O_c +$ $\text{imidazole_acetol_phosphate}_p$
tryptophan synthase	<i>RXN_4_2_1_20_p</i>	$L\text{serine}_p + \text{indole}_p \Rightarrow H2O_c + L\text{tryptophan}_p$
3-isopropylmalate dehydratase	<i>RXN_4_2_1_33_p</i>	$_2\text{isopropylmalate}_p \Rightarrow _3\text{isopropylmalate}_p$
dihydrodipicolinate synthase	<i>RXN_4_2_1_52_p</i>	$L\text{aspartate_semaldehyde}_p + \text{pyruvate}_p \Rightarrow 2 H2O_c +$ $L\text{2-3_dihydrodipicolinate}_p$
dihydroxy-acid dehydratase	<i>RXN_4_2_1_9_p</i>	$_2\text{3_dihydroxy_3_methylvalerate}_p \Rightarrow H2O_c + _2\text{keto_3_methyl_valerate}_p$
dihydroxy-acid dehydratase	<i>RXN_4_2_1_9_p</i>	$_2\text{3_dihydroxy_isovalerate}_p \Rightarrow H2O_c + _2\text{keto_isovalerate}_p$
arogenate dehydratase	<i>RXN_4_2_1_9I_p</i>	$L\text{arogenate}_p \Rightarrow H2O_c + L\text{phenylalanine}_p + CO2_p$
threonine synthase	<i>RXN_4_2_3_1_p</i>	$O\text{phospho_L_homoserine}_p + H2O_c \Rightarrow L\text{threonine}_p + \text{phosphate}_p$
3-dehydroquinate synthase	<i>RXN_4_2_3_4_p</i>	$_3\text{deoxy_D_arabino_heptulosonate}_p + \text{phosphate}_p \Rightarrow _3\text{dehydroquinate}_p +$ $\text{phosphate}_p$
chorismate synthase	<i>RXN_4_2_3_5_p</i>	$_5\text{enolpyruvyl_shikimate}_3\text{phosphate}_p \Rightarrow \text{chorismate}_p + \text{phosphate}_p$
threonine ammonia-lyase	<i>RXN_4_3_1_19_p</i>	$L\text{threonine}_p \Rightarrow _2\text{oxobutanoate}_p + \text{ammonia}_p$
cystathionine beta-lyase	<i>RXN_4_4_1_8_p</i>	$H2O_c + \text{cystathionine}_p \Rightarrow L\text{homocysteine}_p + \text{pyruvate}_p + \text{ammonia}_p$
Diaminopimelate epimerase	<i>RXN_5_1_1_7_p</i>	$L\text{L_diaminopimelate}_p \Rightarrow \text{meso_diaminopimelate}_p$
Ribulose-phosphate 3-epimerase	<i>RXN_5_1_3_1_p</i>	$D\text{ribulose}_5\text{phosphate}_p \Rightarrow D\text{xylulose}_5\text{phosphate}_p$
Triose-phosphate isomerase	<i>RXN_5_3_1_1_p</i>	$D\text{glyceraldehyde}_3\text{phosphate}_p \Rightarrow \text{dihydroxyacetone}_p + \text{phosphoribosylformiminoAICAR}_p \Rightarrow$
1-(5-phosphoribosyl)-5-[(5-	<i>RXN_5_3_1_16_p</i>	$\text{phosphoribosylformiminoAICAR}_p \Rightarrow$ $\text{phosphoribulosylformimino_AICAR}_P_p$
phosphoribosylamino)methylideneamino]imidazole-4-		
carboxamide isomerase		
phosphoribosylanthranilate isomerase	<i>RXN_5_3_1_24_p</i>	$N\text{5_phosphoribosyl_anthraniolate}_p \Rightarrow$ $_1\text{o_carboxyphenylamino}_1\text{deoxyribulose}_5\text{phosphate}_p$
Ribose-5-phosphate isomerase	<i>RXN_5_3_1_6_p</i>	$D\text{ribose}_5\text{phosphate}_p \Rightarrow D\text{ribulose}_5\text{phosphate}_p$

Glucose-6-phosphate isomerase	<i>RXN_5_3_1_9_p</i>	$\text{alpha\_D\_glucose\_6\_phosphate\_p} \Rightarrow \text{D\_fructose\_6\_phosphate\_p}$
phosphoglycerate mutase	<i>RXN_5_4_2_1_p</i>	$\text{3\_phosphoglycerate\_p} \Rightarrow \text{2\_phosphoglycerate\_p}$
phosphoglucomutase	<i>RXN_5_4_2_2_p</i>	$\text{alpha\_D\_glucose\_1\_phosphate\_p} \Rightarrow \text{alpha\_D\_glucose\_6\_phosphate\_p}$
Chorismate mutase	<i>RXN_5_4_99_5_p</i>	$\text{chorismate\_p} \Rightarrow \text{prephenate\_p}$
Acetate-CoA ligase	<i>RXN_6_2_1_1_p</i>	$\text{ATP\_p} + \text{acetate\_p} + \text{coenzyme\_A\_p} \rightarrow \text{AMP\_p} + \text{acetyl\_CoA\_p} + \text{diphosphate\_p}$
Glutamate-ammonia ligase	<i>RXN_6_3_1_2_p</i>	$\text{ATP\_p} + \text{ammonia\_p} + \text{L\_glutamate\_p} \rightarrow \text{ADP\_p} + \text{L\_glutamine\_p} + \text{phosphate\_p} + \text{H\_p}$
NAD+ synthase	<i>RXN_6_3_1_5_p</i>	$\text{ATP\_p} + \text{nicotinic\_acid\_adenine\_dinucleotide\_p} + \text{ammonia\_p} \rightarrow \text{AMP\_p} + \text{diphosphate\_p} + \text{NAD\_p}$
dihydrofolate synthase	<i>RXN_6_3_2_12_p</i>	$\text{ATP\_p} + \text{7\_8\_dihydropteroate\_p} + \text{L\_glutamate\_p} \rightarrow \text{7\_8\_dihydrofolate\_p} + \text{ADP\_p} + \text{phosphate\_p}$
Glutamate-cysteine ligase	<i>RXN_6_3_2_2_p</i>	$\text{ATP\_p} + \text{L\_glutamate\_p} + \text{L\_cysteine\_p} \rightarrow 2 \text{ H\_p} + \text{ADP\_p} + \text{phosphate\_p} + \text{L\_g\_glutamylcysteine\_p}$
Glutathione synthase	<i>RXN_6_3_2_3_p</i>	$\text{ATP\_p} + \text{L\_g\_glutamylcysteine\_p} + \text{glycine\_p} \rightarrow 2 \text{ H\_p} + \text{ADP\_p} + \text{phosphate\_p} + \text{glutathione\_p}$
NAD+ synthase (glutamine-hydrolysing)	<i>RXN_6_3_5_1_p</i>	$\text{ATP\_p} + \text{nicotinic\_acid\_adenine\_dinucleotide\_p} + \text{L\_glutamine\_p} + \text{H}_2\text{O\_c} \rightarrow \text{AMP\_p} + \text{L\_glutamate\_p} + \text{diphosphate\_p} + \text{NAD\_p} + 2 \text{ H\_p}$
aminodeoxychorismate synthase	<i>RXN_6_3_5_8_p</i>	$\text{chorismate\_p} + \text{L\_glutamine\_p} \Rightarrow \text{4\_amino\_4\_deoxychorismate\_p} + \text{L\_glutamate\_p}$
Acetyl-CoA carboxylase	<i>RXN_6_4_1_2_p</i>	$\text{ATP\_p} + \text{HCO}_3\text{p} + \text{acetyl\_CoA\_p} + \text{H\_p} \rightarrow \text{malonyl\_CoA\_p} + \text{ADP\_p} + \text{phosphate\_p}$
-	<i>RXN_7800</i>	$\text{2\_isopropyl\_3\_oxosuccinate\_p} \Rightarrow \text{2\_ketoisocaproate\_p} + \text{CO}_2\text{p}$
-	<i>RXN_GCVMULTI_p</i>	$2 \text{ H\_p} + \text{tetrahydrofolate\_p} + \text{glycine\_p} + \text{NAD\_p} \Rightarrow \text{ammonia\_p} + \text{NADH\_p} + \text{CO}_2\text{p} + \text{5\_10\_methylene\_THF\_p}$
NAD kinase	<i>RXN_nadkinase_p</i>	$\text{NAD\_p} + \text{ATP\_p} \rightarrow \text{ADP\_p} + \text{NADP\_p}$
Photosystems I&II, ATP synthase	<i>RXN_PSII_PSI_ATPs_p</i>	$3 \text{ H}_2\text{O} + 3 \text{ phosphate\_p} + 3 \text{ ADP\_p} + \text{light\_p} + 2 \text{ H}_2\text{O\_c} + 2 \text{ NADP\_p} \rightarrow 3 \text{ ATP\_p} + \text{O}_2\text{p} + 2 \text{ NADPH\_p} + 2 \text{ H\_p}$

-	<i>RXN_SPONT_PRO_p</i>	L_glutamate_gamma_semaldehyde_p => H2O_c + pyrroline_5_carboxylate_p
Protein synthesis	<i>J_protein</i>	49.9049 ATP_c + 0.14393 L_proline_c + 0.17373 L_valine_c + 0.071818 L_tyrosine_c + 0.025926 L_tryptophan_c + 0.13082 L_threonine_c + 0.19191 L_serine_c + 0.11086 L_phenylalanine_c + 0.048872 L_methionine_c + 0.12784 L_aspartate_c + 0.18089 LAlanine_c + 0.2238 L_leucine_c + 0.1037 L_lysine_c + 0.12576 L_isoleucine_c + 0.048872 L_histidine_c + 0.25986 glycine_c + 0.09983 L_glutamate_c + 0.32184 L_glutamine_c + 0.042912 L_cysteine_c + 0.12486 L_asparagine_c + 0.15585 L_arginine_c -> 49.9049 phosphate_c + 49.9049 ADP_c + protein
Triacylglycerol synthesis	<i>J_tg</i>	0.074 FAc16 + 0.53 FAc18 + 0.101 FAc20 + 0.296 FAc22 + sn_glycerol_3_phosphate_c + ATP_c -> tg + ADP_c + phosphate_c
Biomass synthesis	<i>J_biomass</i>	0.65394 tg + 0.15585 L_arginine_c + 0.12486 L_asparagine_c + 0.042912 L_cysteine_c + 0.32184 L_glutamine_c + 0.09983 L_glutamate_c + 0.25986 glycine_c + 0.048872 L_histidine_c + 0.12576 L_isoleucine_c + 0.1037 L_lysine_c + 0.2238 L_leucine_c + 0.18089 LAlanine_c + 0.12784 L_aspartate_c + 0.048872 L_methionine_c + 0.11086 L_phenylalanine_c + 0.19191 L_serine_c + 0.13082 L_threonine_c + 0.025926 L_tryptophan_c + 0.071818 L_tyrosine_c + 0.17373 L_valine_c + 0.14393 L_proline_c + 49.9049 ATP_c + 0.89489 ADP_D_glucose_p -> biomass1g + 50.8 ADP_c + 49.9049 phosphate_c
Fatty acid synthase (C20)	<i>JFAc20</i>	FAc18 + malonyl_CoA_c + 2 H_c + 2 NADPH_c => FAc20 + CO2_c + coenzyme_A_c + 2 NADP_c
Fatty acid synthase (C22)	<i>JFAc22</i>	FAc20 + malonyl_CoA_c + 2 H_c + 2 NADPH_c => FAc22 + CO2_c + coenzyme_A_c + 2 NADP_c
Pyruvate Dehydrogenase Complex	<i>Jpyr_dehydrg_complex_c</i>	pyruvate_c + coenzyme_A_c + NAD_c -> acetyl_CoA_c + CO2_c + H_c + NADH_c
alcohol dehydrogenase	<i>RXN_I_I_I_I_c</i>	H_c + NADH_c + acetaldehyde_c => ethanol_c + NAD_c
L-lactate dehydrogenase	<i>RXN_I_I_I_27_c</i>	H_c + NADH_c + pyruvate_c -> NAD_c + lactate_c
malate dehydrogenase	<i>RXN_I_I_I_38_c</i>	NAD_c + malate_c => H_c + NADH_c + CO2_c + pyruvate_c

nitric-oxide synthase	<i>RXN_1_14_I3_39_c</i>	H_c + NADPH_c + L_arginine_c + O2_c -> citrulline_c + nitricOxide_c + NADP_c
acetaldehyde dehydrogenase	<i>RXN_1_2_I10_c</i>	acetaldehyde_c + coenzyme_A_c + NAD_c => H_c + NADH_c + acetyl_CoA_c
glyceraldehyde-3-phosphate dehydrogenase	<i>RXN_1_2_I12_c</i>	D_glyceraldehyde_3_phosphate_c + phosphate_c + NAD_c => H_c + NADH_c + _1_3_diphosphoglycerate_c
aldehyde dehydrogenase	<i>RXN_1_2_I3_c</i>	acetaldehyde_c + H2O_c + NAD_c => H_c + NADH_c + acetate_c
N-acetyl-gamma-glutamyl-phosphate reductase	<i>RXN_1_2_I38_c</i>	H_c + NADPH_c + N_acetylglutamyl_phosphate_c => phosphate_c + N_acetyl_L_glutamate_5_semialdehyde_c + NADP_c
glyceraldehyde-3-phosphate dehydrogenase	<i>RXN_1_2_I9_c</i>	NADP_c + H2O_c + D_glyceraldehyde_3_phosphate_c => H_c + NADPH_c + _3_phosphoglycerate_c
glutamate synthase	<i>RXN_1_4_I14_c</i>	H_c + NADH_c + L_glutamine_c + alpha_ketoglutarate_c => 2 L_glutamate_c + NAD_c
glutamate dehydrogenase	<i>RXN_1_4_I3_c</i>	ammonia_c + alpha_ketoglutarate_c + H_c + NADH_c => NAD_c + H2O_c + L_glutamate_c
ornithine carbamoyltransferase	<i>RXN_2_I3_3_c</i>	L_ornithine_c + carbamoyl_phosphate_c => citrulline_c + phosphate_c + H_c
amino-acid N-acetyltransferase	<i>RXN_2_3_I1_c</i>	acetyl_CoA_c + L_glutamate_c -> N_acetyl_L_glutamate_c + coenzyme_A_c
ATP citrate synthase	<i>RXN_2_3_3_8_c</i>	citrate_c + ATP_c + coenzyme_A_c -> oxaloacetate_c + ADP_c + acetyl_CoA_c + phosphate_c
sucrose synthase	<i>RXN_2_4_I13_c</i>	UDP_D glucose_c + D_fuctose_c => UDP_c + sucrose_c
sucrose-phosphate synthase	<i>RXN_2_4_I14_c</i>	UDP_D glucose_c + D_fuctose_6_phosphate_c -> H_c + UDP_c + sucrose_6_phosphate_c
aspartate transaminase	<i>RXN_2_6_I1_c</i>	oxaloacetate_c + L_glutamate_c => alpha_ketoglutarate_c + L_aspartate_c
acetylornithine transaminase	<i>RXN_2_6_I11_c</i>	L_glutamate_c + N_acetyl_L_glutamate_5_semialdehyde_c => alpha_ketoglutarate_c + N_acetyl_L_ornithine_c
ornithine aminotransferase	<i>RXN_2_6_I13_c</i>	L_ornithine_c + alpha_ketoglutarate_c => L_glutamate_gamma_semialdehyde_c + L_glutamate_c
alanine transaminase	<i>RXN_2_6_I2_c</i>	LAlanine_c + alpha_ketoglutarate_c => pyruvate_c + L_glutamate_c

hexokinase	<i>RXN_2_7_1_1_c</i>	ATP_c + alpha_D_glucose_c -> 2 H_c + alpha_D_glucose_6_phosphate_c + ADP_c
6-phosphofructokinase	<i>RXN_2_7_1_11_c</i>	ATP_c + D_fructose_6_phosphate_c -> ADP_c + fructose_1_6_bisphosphate_c
fructokinase	<i>RXN_2_7_1_4_c</i>	ATP_c + D_fructose_c -> 2 H_c + ADP_c + D_fructose_6_phosphate_c
pyruvate kinase	<i>RXN_2_7_1_40_c</i>	ADP_c + phosphoenolpyruvate_c -> ATP_c + pyruvate_c
diphosphate-fructose-6-phosphate 1-phosphotransferase	<i>RXN_2_7_1_90_c</i>	D_fructose_6_phosphate_c + diphosphate_c => phosphate_c + fructose_1_6_bisphosphate_c
phosphoglycerate kinase	<i>RXN_2_7_2_3_c</i>	ATP_c + _3_phosphoglycerate_c -> _1_3_diphosphoglycerate_c + ADP_c
acetylglutamate kinase	<i>RXN_2_7_2_8_c</i>	N_acetyl_L glutamate_c + ATP_c -> ADP_c + N_acetylglutamyl_phosphate_c
adenylate kinase	<i>RXN_2_7_4_3_c</i>	ATP_c + AMP_c => 2 ADP_c
nucleoside-diphosphate kinase	<i>RXN_2_7_4_6_c</i>	ATP_c + UDP_c => H_c + ADP_c + UTP_c
glucose-1-phosphate adenylyltransferase	<i>RXN_2_7_7_27_c</i>	ATP_c + alpha_D_glucose_1_phosphate_c -> ADP_D_glucose_c + diphosphate_c
UTP-glucose-1-phosphate uridylyltransferase	<i>RXN_2_7_7_9_c</i>	H_c + UTP_c + alpha_D_glucose_1_phosphate_c => diphosphate_c + UDP_D_glucose_c
sucrose-phosphate phosphatase	<i>RXN_3_1_3_24_c</i>	H2O_c + sucrose_6_phosphate_c -> phosphate_c + sucrose_c
beta-fructofuranosidase	<i>RXN_3_2_1_26_c</i>	H2O_c + sucrose_c -> alpha_D_glucose_c + D_fructose_c
acetylornithine deacetylase	<i>RXN_3_5_1_16_c</i>	N_acetyl_L_ornithine_c + H2O_c => acetate_c + L_ornithine_c
urease	<i>RXN_3_5_1_5_c</i>	H2O_c + urea_c -> CO2_c + 2 ammonia_c
N-carbamoylputrescine amidase	<i>RXN_3_5_1_53_c</i>	N_carbamoylputrescine_c + H2O_c => putrescine_c + ammonia_c + CO2_c
arginase	<i>RXN_3_5_3_1_c</i>	H2O_c + L_arginine_c => L_ornithine_c + urea_c
agmatine deiminase	<i>RXN_3_5_3_12_c</i>	agmatine_c + H2O_c => ammonia_c + N_carbamoylputrescine_c
inorganic diphosphatase	<i>RXN_3_6_1_1_c</i>	H2O_c + diphosphate_c => 2 phosphate_c
Pyruvate decarboxylase	<i>RXN_4_1_1_1_c</i>	pyruvate_c -> acetaldehyde_c + CO2_c
Glutamate decarboxylase	<i>RXN_4_1_1_15_c</i>	L glutamate_c => CO2_c + _4_aminobutyrate_c
Arginine decarboxylase	<i>RXN_4_1_1_19_c</i>	L_arginine_c => CO2_c + agmatine_c
phosphoenolpyruvate carboxykinase (ATP)	<i>RXN_4_1_1_49_c</i>	oxaloacetate_c + ATP_c => ADP_c + CO2_c + phosphoenolpyruvate_c
Fructose-bisphosphate aldolase	<i>RXN_4_1_2_13_c</i>	fructose_1_6_bisphosphate_c => D_glyceraldehyde_3_phosphate_c +

		dihydroxyacetone_phosphate_c
carbonate dehydratase	<i>RXN_4_2_1_1_c</i>	$H_2O_c + CO2_c \Rightarrow HCO3_c + H_c$
phosphopyruvate hydratase	<i>RXN_4_2_1_11_c</i>	$_2\_phosphoglycerate_c \Rightarrow H_2O_c + phosphoenolpyruvate_c$
argininosuccinate lyase	<i>RXN_4_3_2_1_c</i>	$L\_arginino\_succinate_c \Rightarrow fumarate_c + L\_arginine_c$
Triose-phosphate isomerase	<i>RXN_5_3_1_1_c</i>	$D\_glyceraldehyde_3\_phosphate_c \Rightarrow dihydroxyacetone\_phosphate_c$
Glucose-6-phosphate isomerase	<i>RXN_5_3_1_9_c</i>	$alpha\_D\_glucose\_6\_phosphate_c \Rightarrow D\_fructose\_6\_phosphate_c$
phosphoglycerate mutase	<i>RXN_5_4_2_1_c</i>	$_3\_phosphoglycerate_c \Rightarrow _2\_phosphoglycerate_c$
phosphoglucomutase	<i>RXN_5_4_2_2_c</i>	$alpha\_D\_glucose\_1\_phosphate_c \Rightarrow alpha\_D\_glucose\_6\_phosphate_c$
-	<i>RXN_5257</i>	$sn\_glycerol\_3\_phosphate_c + NAD_c \Rightarrow dihydroxyacetone\_phosphate_c + H_c + NADH_c$
Carbamoyl-phosphate synthase	<i>RXN_6_3_4_16_c</i>	$2 ATP_c + H_2O_c + CO2_c + ammonia_c \rightarrow 2 ADP_c + phosphate_c + carbamoyl\_phosphate_c + 4 H_c$
Argininosuccinate synthase	<i>RXN_6_3_4_5_c</i>	$ATP_c + citrulline_c + L\_aspartate_c \rightarrow AMP_c + L\_arginino\_succinate_c + diphosphate_c + 2 H_c$
Asparagine synthase	<i>RXN_6_3_5_4_c</i>	$ATP_c + H_2O_c + L\_glutamine_c + L\_aspartate_c \rightarrow AMP_c + L\_asparagine_c + L\_glutamate_c + diphosphate_c$
Carbamoyl-phosphate synthase	<i>RXN_6_3_5_5_c</i>	$2 ATP_c + H_2O_c + L\_glutamine_c + HCO3_c \rightarrow 2 ADP_c + L\_glutamate_c + phosphate_c + carbamoyl\_phosphate_c$
Acetyl-CoA carboxylase	<i>RXN_6_4_1_2_c</i>	$ATP_c + HCO3_c + acetyl\_CoA_c + H_c \rightarrow malonyl\_CoA_c + ADP_c + phosphate_c$
NAD kinase	<i>RXN_nadkinase_c</i>	$NAD_c + ATP_c \rightarrow ADP_c + NADP_c$
-	<i>RXN_PROLINE_MULTI</i>	$ATP_c + H_c + NADPH_c + L\_glutamate_c \rightarrow NADP_c + L\_glutamate\_gamma\_semialdehyde_c + ADP_c + phosphate_c$
ATP synthase	<i>ATPsynthase_m</i>	$ADP_m + phosphate_m + 3 H_c \rightarrow ATP_m + H_2O_c + 2 H_m$
malate dehydrogenase	<i>RXN_I_I_1_37_m</i>	$malate_m + NAD_m \Rightarrow oxaloacetate_m + H_m + NADH_m$
malate dehydrogenase	<i>RXN_I_I_1_38_m</i>	$NAD_m + malate_m \Rightarrow H_m + NADH_m + CO2_m + pyruvate_m$
isocitrate dehydrogenase	<i>RXN_I_I_1_42_m</i>	$NAD_m + isocitrate_m \Rightarrow H_m + NADH_m + CO2_m + alpha\_ketoglutarate_m$

ubiquinol-cytochrome-c reductase	<i>RXN_1_10_2_2_m</i>	$1.5 \text{ H}_\text{m} + 2 \text{ ox\_cytc}_\text{m} + \text{UQH2}_\text{m} \rightarrow 1.5 \text{ H}_\text{c} + 2 \text{ red\_cytc}_\text{m} + \text{UQ}_\text{m}$
succinate-semialdehyde dehydrogenase	<i>RXN_1_2_1_24_m</i>	$\text{H}_2\text{O}_\text{c} + \text{succinate\_semialdehyde}_\text{m} + \text{NAD}_\text{m} \rightarrow \text{H}_\text{m} + \text{NADH}_\text{m} + \text{succinate}_\text{m}$
succinate dehydrogenase (ubiquinone)	<i>RXN_1_3_5_1_m</i>	$\text{succinate}_\text{m} + \text{UQ}_\text{m} \Rightarrow \text{fumarate}_\text{m} + \text{UQH2}_\text{m}$
glutamate dehydrogenase	<i>RXN_1_4_1_2_m</i>	$\text{H}_2\text{O}_\text{c} + \text{L\_glutamate}_\text{m} + \text{NAD}_\text{m} \Rightarrow \text{H}_\text{m} + \text{NADH}_\text{m} + \text{ammonia}_\text{m} + \text{alpha\_ketoglutarate}_\text{m}$
NADH dehydrogenase (ubiquinone)	<i>RXN_1_6_5_3_m</i>	$\text{H}_\text{m} + \text{NADH}_\text{m} + \text{UQ}_\text{m} \rightarrow \text{UQH2}_\text{m} + \text{NAD}_\text{m}$
cytochrome-c oxidase	<i>RXN_1_9_3_1_m</i>	$4 \text{ red\_cytc}_\text{m} + \text{O}_2\text{}_\text{m} + 6 \text{ H}_\text{m} \rightarrow 4 \text{ ox\_cytc}_\text{m} + 6 \text{ H}_\text{c} + 2 \text{ H}_2\text{O}_\text{c}$
citrate synthase	<i>RXN_2_3_3_1_m</i>	$\text{oxaloacetate}_\text{m} + \text{H}_2\text{O}_\text{c} + \text{acetyl\_CoA}_\text{m} \rightarrow \text{citrate}_\text{m} + \text{coenzyme A}_\text{m}$
malate synthase	<i>RXN_2_3_3_9_m</i>	$\text{H}_2\text{O}_\text{c} + \text{glyoxylate}_\text{m} + \text{acetyl\_CoA}_\text{m} \rightarrow \text{coenzyme A}_\text{m} + \text{malate}_\text{m}$
aspartate transaminase	<i>RXN_2_6_1_1_m</i>	$\text{oxaloacetate}_\text{m} + \text{L\_glutamate}_\text{m} \Rightarrow \text{alpha\_ketoglutarate}_\text{m} + \text{L\_aspartate}_\text{m}$
4-aminobutyrate transaminase	<i>RXN_2_6_1_19_m</i>	$\text{pyruvate}_\text{m} + \text{4\_aminobutyrate}_\text{m} \rightarrow \text{L\_alanine}_\text{m} + \text{succinate\_semialdehyde}_\text{m}$
4-aminobutyrate transaminase	<i>RXN_2_6_1_19_m2</i>	$\text{alpha\_ketoglutarate}_\text{m} + \text{4\_aminobutyrate}_\text{m} \Rightarrow \text{succinate\_semialdehyde}_\text{m} + \text{L\_glutamate}_\text{m}$
2-oxoglutarate dehydrogenase	<i>RXN_2OXOGLUTARATEDEH</i>	$\text{coenzyme A}_\text{m} + \text{alpha\_ketoglutarate}_\text{m} + \text{NAD}_\text{m} \Rightarrow \text{succinyl\_CoA}_\text{m} + \text{H}_\text{m} + \text{NADH}_\text{m} + \text{CO}_2\text{}_\text{m}$
isocitrate lyase	<i>RXN_4_1_3_1_m</i>	$\text{isocitrate}_\text{m} \Rightarrow \text{glyoxylate}_\text{m} + \text{succinate}_\text{m}$
fumarate hydratase	<i>RXN_4_2_1_2_m</i>	$\text{fumarate}_\text{m} + \text{H}_2\text{O}_\text{c} \Rightarrow \text{malate}_\text{m}$
aconitate hydratase	<i>RXN_4_2_1_3_m</i>	$\text{citrate}_\text{m} \Rightarrow \text{cis\_aconitate}_\text{m} + \text{H}_2\text{O}_\text{c}$
aconitate hydratase	<i>RXN_4_2_1_3_m2</i>	$\text{cis\_aconitate}_\text{m} + \text{H}_2\text{O}_\text{c} \Rightarrow \text{isocitrate}_\text{m}$
Succinate-CoA ligase	<i>RXN_6_2_1_5_m</i>	$\text{succinyl\_CoA}_\text{m} + \text{ADP}_\text{m} + \text{phosphate}_\text{m} \Rightarrow \text{ATP}_\text{m} + \text{coenzyme A}_\text{m} + \text{succinate}_\text{m}$
transport	<i>transp_acet_p</i>	$\text{acetate}_\text{c} \Rightarrow \text{acetate}_\text{p}$
transport	<i>transp_acetaldehyde_p</i>	$\text{acetaldehyde}_\text{c} \Rightarrow \text{acetaldehyde}_\text{p}$
transport	<i>transp_ADPGluc_p</i>	$\text{ADP\_D\_glucose}_\text{c} \Rightarrow \text{ADP\_D\_glucose}_\text{p}$
transport	<i>transp_ala_p</i>	$\text{L\_alanine}_\text{p} \Rightarrow \text{L\_alanine}_\text{c}$
transport	<i>transp_atpAdp_p</i>	$\text{ATP}_\text{p} + \text{ADP}_\text{c} \Rightarrow \text{ATP}_\text{c} + \text{ADP}_\text{p}$

transport	<i>transp_cit_p</i>	$\text{citrate}_c \Rightarrow \text{citrate}_p$
transport	<i>transp_co2_p</i>	$\text{CO2}_c \Rightarrow \text{CO2}_p$
transport	<i>transp_cys_p</i>	$\text{L\_cysteine}_p \Rightarrow \text{L\_cysteine}_c$
transport	<i>transp_dhap_p</i>	$\text{phosphate}_c + \text{dihydroxyacetone\_phosphate}_p \Rightarrow \text{phosphate}_p + \text{dihydroxyacetone\_phosphate}_c$
transport	<i>transp_glc6p_p</i>	$\alpha\text{D\_glucose}_6\text{phosphate}_c + \text{phosphate}_p \Rightarrow \alpha\text{D\_glucose}_6\text{phosphate}_p + \text{phosphate}_c$
transport	<i>transp_gln_p</i>	$\text{L\_glutamine}_c \Rightarrow \text{L\_glutamine}_p$
transport	<i>transp_gluc_p</i>	$\alpha\text{D\_glucose}_c \Rightarrow \alpha\text{D\_glucose}_p$
transport	<i>transp_gly_p</i>	$\text{glycine}_c \Rightarrow \text{glycine}_p$
transport	<i>transp_his_p</i>	$\text{L\_histidine}_p \Rightarrow \text{L\_histidine}_c$
transport	<i>transp_ileu_p</i>	$\text{L\_isoleucine}_p \Rightarrow \text{L\_isoleucine}_c$
transport	<i>transp_leu_p</i>	$\text{L\_leucine}_p \Rightarrow \text{L\_leucine}_c$
transport	<i>transp_lys_p</i>	$\text{L\_lysine}_p \Rightarrow \text{L\_lysine}_c$
transport	<i>transp_mal_p</i>	$\text{malate}_c \Rightarrow \text{malate}_p$
transport	<i>transp_malAsp_p</i>	$\text{malate}_c + \text{L\_aspartate}_p \Rightarrow \text{malate}_p + \text{L\_aspartate}_c$
transport	<i>transp_malGlu_p</i>	$\text{malate}_c + \text{L\_glutamate}_p \Rightarrow \text{malate}_p + \text{L\_glutamate}_c$
transport	<i>transp_malKetog_p</i>	$\text{malate}_c + \alpha\text{ketoglutarate}_p \Rightarrow \text{malate}_p + \alpha\text{ketoglutarate}_c$
transport	<i>transp_mallCit_p</i>	$\text{malate}_c + \text{citrate}_p \Rightarrow \text{malate}_p + \text{citrate}_c$
transport	<i>transp_met_p</i>	$\text{L\_methionine}_p \Rightarrow \text{L\_methionine}_c$
transport	<i>transp_nh3_p</i>	$\text{ammonia}_c \Rightarrow \text{ammonia}_p$
transport	<i>transp_o2_p</i>	$\text{O2}_c \Rightarrow \text{O2}_p$
transport	<i>transp_oaaMal_p</i>	$\text{oxaloacetate}_c + \text{malate}_p \Rightarrow \text{oxaloacetate}_p + \text{malate}_c$
transport	<i>transp_P_m</i>	$\text{phosphate}_c \Rightarrow \text{phosphate}_m$
transport	<i>transp_P_p</i>	$\text{phosphate}_c \Rightarrow \text{phosphate}_p$
transport	<i>transp_pep_p</i>	$\text{phosphate}_c + \text{phosphoenolpyruvate}_p \Rightarrow \text{phosphate}_p + \text{phosphoenolpyruvate}_c$

transport	<i>transp_phe_p</i>	$L_{phenylalanine\_p} \Rightarrow L_{phenylalanine\_c}$
transport	<i>transp_pro_p</i>	$L_{proline\_p} \Rightarrow L_{proline\_c}$
transport	<i>transp_pyr_p</i>	$pyruvate\_c \rightarrow pyruvate\_p$
transport	<i>transp_ser_p</i>	$L_{serine\_c} \Rightarrow L_{serine\_p}$
transport	<i>transp_thr_p</i>	$L_{threonine\_p} \Rightarrow L_{threonine\_c}$
transport	<i>transp_trp_p</i>	$L_{tryptophan\_p} \Rightarrow L_{tryptophan\_c}$
transport	<i>transp_tyr_p</i>	$L_{tyrosine\_p} \Rightarrow L_{tyrosine\_c}$
transport	<i>transp_val_p</i>	$L_{valine\_p} \Rightarrow L_{valine\_c}$
transport	<i>transp_ala_m</i>	$L_{alanine\_c} \Rightarrow L_{alanine\_m}$
transport	<i>transp_atpAdp_m</i>	$ATP\_m + ADP\_c \Rightarrow ATP\_c + ADP\_m$
transport	<i>transp_co2_m</i>	$CO2\_c \Rightarrow CO2\_m$
transport	<i>transp_gaba_m</i>	$L_{glutamate\_c} + _4_{aminobutyrate\_m} \Rightarrow L_{glutamate\_m} + _4_{aminobutyrate\_c}$
transport	<i>transp_glu_m</i>	$L_{glutamate\_c} + L_{aspartate\_m} \Rightarrow L_{glutamate\_m} + L_{aspartate\_c}$
transport	<i>transp_ketoglCit_m</i>	$citrate\_c + alpha_{ketoglutarate\_m} \Rightarrow citrate\_m + alpha_{ketoglutarate\_c}$
transport	<i>transp_ketoglSucc_m</i>	$succinate\_c + alpha_{ketoglutarate\_m} \Rightarrow succinate\_m + alpha_{ketoglutarate\_c}$
transport	<i>transp_malKetog_m</i>	$malate\_c + alpha_{ketoglutarate\_m} \Rightarrow malate\_m + alpha_{ketoglutarate\_c}$
transport	<i>transp_mallCit_m</i>	$malate\_c + citrate\_m \Rightarrow malate\_m + citrate\_c$
transport	<i>transp_MalP_m</i>	$phosphate\_c + malate\_m \Rightarrow phosphate\_m + malate\_c$
transport	<i>transp_nh3_m</i>	$ammonia\_c \Rightarrow ammonia\_m$
transport	<i>transp_o2_m</i>	$O2\_c \Rightarrow O2\_m$
transport	<i>transp_oaaAsp_m</i>	$oxaloacetate\_c + L_{aspartate\_m} \Rightarrow oxaloacetate\_m + L_{aspartate\_c}$
transport	<i>transp_oaaCit_m</i>	$oxaloacetate\_c + citrate\_m \Rightarrow oxaloacetate\_m + citrate\_c$
transport	<i>transp_oaaKetogl_m</i>	$oxaloacetate\_c + alpha_{ketoglutarate\_m} \Rightarrow oxaloacetate\_m + alpha_{ketoglutarate\_c}$
transport	<i>transp_oaaMal_m</i>	$oxaloacetate\_c + malate\_m \Rightarrow oxaloacetate\_m + malate\_c$
transport	<i>transp_oaaSucc_m</i>	$oxaloacetate\_c + succinate\_m \Rightarrow oxaloacetate\_m + succinate\_c$
transport	<i>transp_pyr_m</i>	$pyruvate\_c \rightarrow pyruvate\_m$

transport	<i>transp_succFum_m</i>	succinate_c + fumarate_m => succinate_m + fumarate_c
transport	<i>transp_succMal_m</i>	malate_c + succinate_m => malate_m + succinate_c
transport	<i>transp_succP_m</i>	phosphate_c + succinate_m => phosphate_m + succinate_c
transport	<i>transp_glyoxylate_p2m</i>	glyoxylate_p => glyoxylate_m

**Πίνακας 1.4.** Οι αντιδράσεις του μεταβολικού μοντέλου της ελαιοκράμβης.

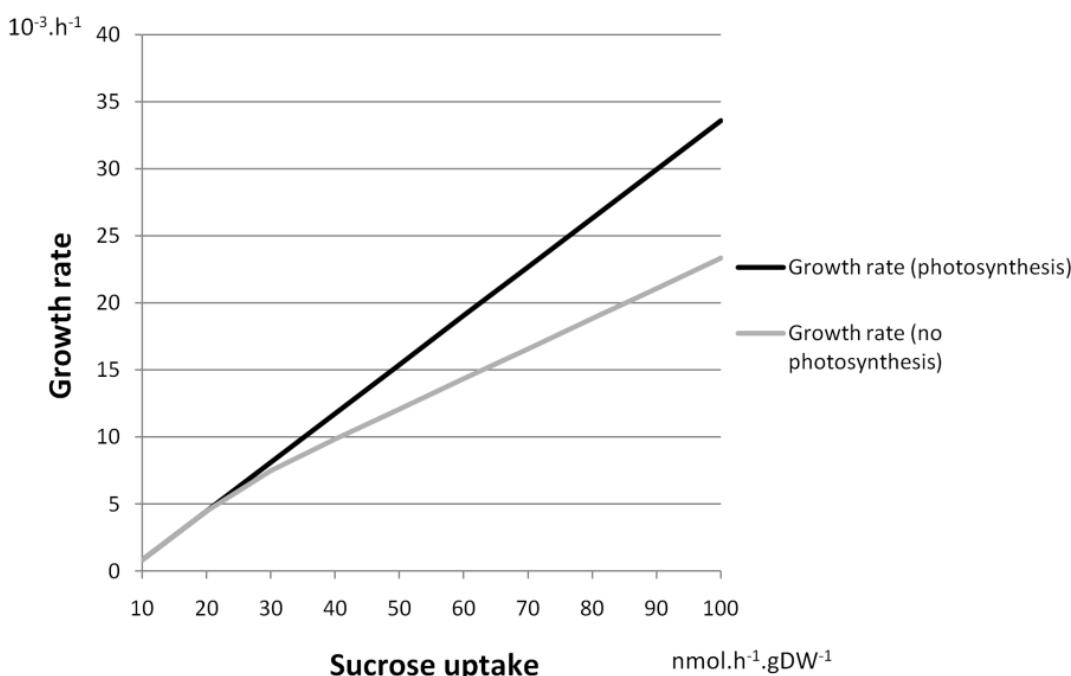
c: κυτταρόπλασμα

p: χλωροπλάστης

m: μιτοχόνδριο

e: εξωτερικό

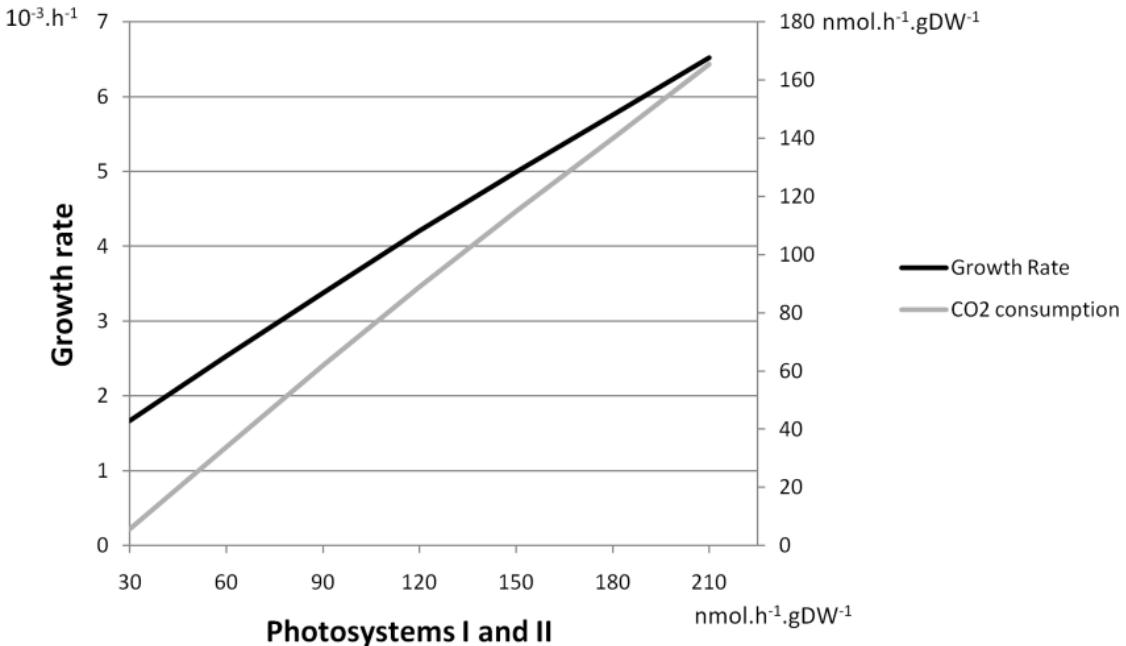
Στο Σχήμα 1.3 συνοψίζονται οι τιμές του ρυθμού ανάπτυξης για διαφορετικές τιμές κατανάλωσης γλυκόζης. Παρατηρείται η αναμενόμενη αύξηση του ρυθμού παραγωγής βιομάζας παρουσία φωτοσύνθεσης ως προς τις αντίστοιχες τιμές κατανάλωσης σακχαρόζης.



**Σχήμα 1.3.** Γραφική παράσταση του ρυθμού παραγωγής βιομάζας προς την κατανάλωση σακχαρόζης, απουσία και παρουσία φωτοσύνθεσης.

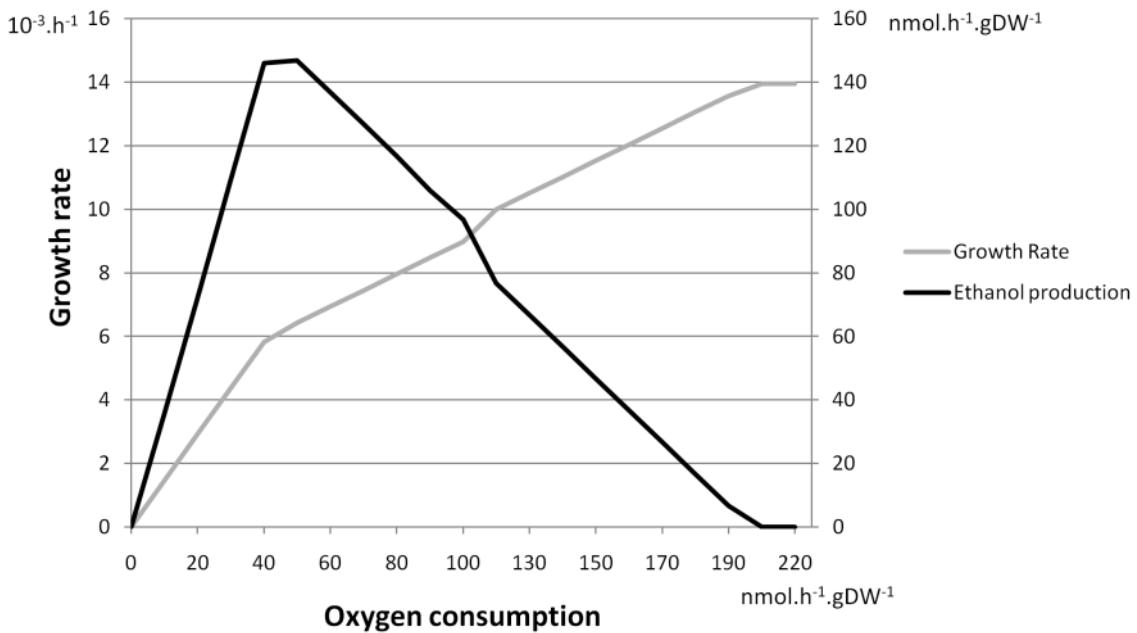
Επιπρόσθετα, πραγματοποιήθηκαν προσομοιώσεις με σκοπό την εκτίμηση της λειτουργικότητας του μοντέλου ως προς την προσομοίωση κυρίων μεταβολικών καταστάσεων. Το Σχήμα 1.4 συνοψίζει τα αποτελέσματα της προσομοίωσης αυτοτροφικών συνθηκών ανάπτυξης, δηλαδή μόνο φωτοσύνθεσης χωρίς την παροχή σακχαρόζης. Παρατηρείται ότι όσο αυξάνεται η ροή μέσα από την αντίδραση της φωτοσύνθεσης (φωτοσυστήματα I και II), αυξάνεται ανάλογα ο ρυθμός παραγωγής

βιομάζας και η πρόσληψη διοξειδίου του άνθρακα. Συνεπώς το μοντέλο παρέχει μια ρεαλιστική προσομοίωση των αυτοτροφικών συνθηκών ανάπτυξης.



**Σχήμα 1.4.** Προσομοίωση αυτοτροφικών συνθηκών ανάπτυξης. Στην προσομοίωση αυτή, έχει επιτραπεί η κατανάλωση CO<sub>2</sub>, ενώ η πρόσληψη οξυγόνου και γλυκόζης έχει περιοριστεί στο 0. Η προσομοίωση εκτελέστηκε μεταβάλλοντας την τιμή περιορισμού της αντίδρασης των φωτοσυστημάτων I και II.

Το Σχήμα 1.5 συνοψίζει τα αποτελέσματα της προσομοίωσης του ρυθμού παραγωγής βιομάζας ως προς την διαθεσιμότητα οξυγόνου. Η προσομοίωση προβλέπει την παραγωγή αιθανόλης σε συνθήκες αναεροβίωσης (μειωμένης διαθεσιμότητας οξυγόνου) εις βάρος της παραγωγής βιομάζας. Αντίθετα, σε συνθήκες άφθονης διαθεσιμότητας οξυγόνου, η παραγωγή αιθανόλης ελαχιστοποιείται, προς όφελος του ρυθμού παραγωγής βιομάζας που μεγιστοποιείται.



**Σχήμα 1.5.** Προσομοίωση περιορισμού της πρόσληψης οξυγόνου, με σταθερή ροή εισόδου γλυκόζης  $58.3 \text{ nmol.h}^{-1}.\text{gDW}^{-1}$ . Παρατηρείται η αύξηση παραγωγής αιθανόλης στις χαμηλές τιμές πρόσληψης οξυγόνου και η μείωση της σε συνθήκες μεγάλης διαθεσιμότητας οξυγόνου. Ο ρυθμός παραγωγής βιομάζας είναι υψηλότερος στις αερόβιες συνθήκες.

Συμπερασματικά, το μοντέλο είναι σε θέση να αναπαράγει κύριες μεταβολικές καταστάσεις. Η γενική λειτουργικότητα του μοντέλου αποτελεί και την πρωταρχική αποτίμηση της ορθότητάς του, ως προς την περιγραφή του μεταβολισμού της ελαιοκράμβης.

### 1.3.2 Ανάλυση Διαφοροποίησης Ροών (FVA)

Με τη μέθοδο FVA εκτιμήθηκαν τα ανώτερα και κατώτερα όρια των αντιδράσεων του μοντέλου, τα οποία παρατίθενται στον Πίνακα 1.6. Η ανάλυση αυτή διερευνά την πλαστικότητα του μεταβολικού δικτύου, εκτιμώντας την ελαστικότητα ή αντίθετα, την ακαμψία των ακμών του. Παραδείγματος χάρην, παρατηρείται ότι με την παρουσία φωτοσύνθεσης, χρησιμοποιούνται και οι δυο αφυδρογωνάσες του πυροσταφυλικού οξέος, η χλωροπλαστική και η μιτοχονδριακή. Αντιθέτως, χωρίς φωτοσύνθεση, η μεταβολική ροή ανακατευθύνεται κυρίως προς το χλωροπλαστικό ισόφορο (Πίνακας 1.5).

Όνομα αντίδρασης	Με φωτοσύνθεση		Χωρίς φωτοσύνθεση	
	Ρυθμός ανάπτυξης: 18.43 h <sup>-1</sup>	Μέγιστη ροή	Ρυθμός ανάπτυξης: 13.96 h <sup>-1</sup>	Μέγιστη ροή
ATP_Maintenance_m	0,00	1741,45	0,00	0,00
ATPsynthase_m	803,31	2000,00	892,47	892,47
EX_acetate(e)	0,00	0,00	0,00	0,00
EXAlanine(e)	-7,37	-7,37	-7,37	-7,37
EX_ammonia(e)	-51,17	-49,97	-34,14	-34,14
EX_biomass(e)	18,43	18,43	13,96	13,96
EX_co2(e)	142,00	142,00	289,35	289,35
EX_ethanol(e)	0,00	0,00	0,00	0,00
EX_formate(e)	0,00	0,00	0,00	0,00
EX_Glutamine(e)	-4,25	-4,25	-4,25	-4,25
EX_h(e)	-69,92	-69,92	-69,92	-69,92
EX_lactate(e)	0,00	0,00	0,00	0,00
EX_nitricOxide(e)	0,00	0,00	0,00	0,00
EX_no3(e)	-1,20	0,00	0,00	0,00
EX_o2(e)	-22,14	-19,74	-197,69	-197,69
EX_phosphate(e)	-11,15	-11,15	-8,45	-8,45
EX_sucrose(e)	-58,30	-58,30	-58,30	-58,30
EX_sulfate(e)	-1,69	-1,69	-1,28	-1,28
EX_UDP_glucose(e)	0,00	0,00	0,00	0,00
J_biomass	18,43	18,43	13,96	13,96
J_protein	0,00	0,00	0,00	0,00
J_starch	0,00	0,00	0,00	0,00
J_tg	12,05	12,05	9,13	9,13
JFAC16	12,06	12,06	9,14	9,14
JFAC18	11,17	11,17	8,46	8,46
JFAC20	4,78	4,78	3,62	3,62
JFAC22	3,57	3,57	2,70	2,70
Jpyr_dehydrog_complex_c	0,00	125,80	0,00	9,31
Jpyr_dehydrog_complex_p	0,00	113,51	85,96	85,96
RXN_1_1_1_1_c	0,00	0,00	0,00	0,00
RXN_1_1_1_23_p	0,90	0,90	0,68	0,68
RXN_1_1_1_23_p2	0,90	0,90	0,68	0,68
RXN_1_1_1_25_p	3,84	3,84	2,91	2,91
RXN_1_1_1_27_c	0,00	0,00	0,00	0,00
RXN_1_1_1_29_p	0,00	1160,96	0,00	0,00
RXN_1_1_1_3_p	0,00	125,80	0,00	0,00
RXN_1_1_1_37_m	-145,44	2000,00	65,54	2000,00
RXN_1_1_1_38_c	-2000,00	2000,00	-1863,32	2000,00
RXN_1_1_1_38_m	-2000,00	1160,96	-2000,00	972,48
RXN_1_1_1_38_p	-1371,26	79,65	-263,77	-263,77
RXN_1_1_1_42_m	120,41	687,32	137,75	137,75
RXN_1_1_1_44_p	-317,59	-1,05	80,48	80,48
RXN_1_1_1_49_p	-317,59	-1,05	80,48	80,48

RXN_1_1_1_85_p	4,12	4,12	3,12	3,12
RXN_1_1_1_86_p	7,33	7,33	5,55	5,55
RXN_1_1_1_86_p2	2,32	2,32	1,76	1,76
RXN_1_1_1_95_p	51,39	1338,97	77,14	77,14
RXN_1_1_99_14_p	0,00	0,00	0,00	0,00
RXN_1_10_2_2_m	357,03	888,89	396,65	396,65
RXN_1_11_1_6_p	-0,85	-0,85	-0,64	-0,64
RXN_1_14_13_39_c	0,00	0,00	0,00	0,00
RXN_1_2_1_10_c	-117,45	11,22	-2,98	8,50
RXN_1_2_1_11_p	1,91	127,71	1,45	1,45
RXN_1_2_1_12_c	-725,60	0,00	0,00	0,00
RXN_1_2_1_12_p	-725,60	0,00	0,00	0,00
RXN_1_2_1_21_p	0,00	0,00	0,00	0,00
RXN_1_2_1_24_m	0,00	1489,22	0,00	0,00
RXN_1_2_1_3_c	-4,56	111,82	-3,46	-1,28
RXN_1_2_1_38_c	0,00	2,87	0,00	2,18
RXN_1_2_1_41_p	2,65	2,65	2,01	2,01
RXN_1_2_1_43_p	0,00	0,00	0,00	0,00
RXN_1_2_1_9_c	19,58	19,58	14,83	14,83
RXN_1_3_1_26_p	1,91	1,91	1,45	1,45
RXN_1_3_1_43_p	1,32	1,32	1,00	1,00
RXN_1_3_5_1_m	-131,73	-5,93	-2,70	-2,70
RXN_1_4_1_14_c	-2000,00	1999,13	-1828,99	2000,00
RXN_1_4_1_2_m	-2000,00	2000,00	-2000,00	1863,32
RXN_1_4_1_3_c	-2000,00	2000,00	-2000,00	2000,00
RXN_1_4_3_16_p	0,00	0,00	0,00	0,00
RXN_1_4_7_1_p	-1252,14	-91,17	-170,92	-170,92
RXN_1_5_1_2_p	2,65	2,65	2,01	2,01
RXN_1_5_1_20_p	-1246,12	-85,03	-169,00	-169,00
RXN_1_5_1_3_p	0,00	0,00	0,00	0,00
RXN_1_5_7_1_p	85,03	1246,12	169,00	169,00
RXN_1_6_5_3_m	362,95	1020,62	399,36	399,36
RXN_1_7_1_1_p	0,00	1,20	0,00	0,00
RXN_1_7_7_1_p	0,00	1,20	0,00	0,00
RXN_1_8_3_1_p	-1,69	-1,69	-1,28	-1,28
RXN_1_8_4_9_p	0,00	0,00	0,00	0,00
RXN_1_8_7_1_p	1,69	1,69	1,28	1,28
RXN_1_9_3_1_m	178,51	444,44	198,33	198,33
RXN_2_1_1_13_p	0,00	0,00	0,00	0,00
RXN_2_1_1_14_p	0,90	0,90	0,68	0,68
RXN_2_1_2_1_p	45,69	411,77	72,82	72,82
RXN_2_1_2_11_p	0,00	0,00	0,00	0,00
RXN_2_1_3_3_c	2,87	548,53	2,18	2,18
RXN_2_2_1_1_p	-307,51	-35,77	-5,90	-5,90
RXN_2_2_1_1_p	-303,66	-31,93	-2,98	-2,98
RXN_2_2_1_2_p	-303,66	-31,93	-2,98	-2,98

RXN_2_2_1_6_p	7,33	7,33	5,55	5,55
RXN_2_2_1_6_p2	2,32	2,32	1,76	1,76
RXN_2_3_1_1_c	0,00	2,87	0,00	2,18
RXN_2_3_1_30_p	1,69	1,69	1,28	1,28
RXN_2_3_2_2_p	0,00	580,48	0,00	0,00
RXN_2_3_3_1_m	0,00	0,00	0,00	0,00
RXN_2_3_3_13_p	4,12	4,12	3,12	3,12
RXN_2_3_3_8_c	0,00	125,80	0,00	0,00
RXN_2_3_3_8_p	0,00	113,51	0,00	0,00
RXN_2_3_3_9_m	0,00	0,00	0,00	0,00
RXN_2_4_1_13_c	-1082,82	552,14	-288,54	-288,54
RXN_2_4_1_14_c	0,00	1024,52	230,24	230,24
RXN_2_4_2__p	0,90	0,90	0,68	0,68
RXN_2_4_2_17_p	0,90	0,90	0,68	0,68
RXN_2_4_2_18_p	0,48	0,48	0,36	0,36
RXN_2_4_2_19_p	0,00	0,00	0,00	0,00
RXN_2_5_1__p	0,90	0,90	0,68	0,68
RXN_2_5_1_15_p	0,00	0,00	0,00	0,00
RXN_2_5_1_19_p	3,84	3,84	2,91	2,91
RXN_2_5_1_47_p	1,69	1,69	1,28	1,28
RXN_2_5_1_54_p	3,84	3,84	2,91	2,91
RXN_2_5_1_72_p	0,00	0,00	0,00	0,00
RXN_2_6_1__p	1,91	1,91	1,45	1,45
RXN_2_6_1_1__m	-2000,00	2000,00	-2000,00	2000,00
RXN_2_6_1_1_c	-2000,00	2000,00	-2000,00	2000,00
RXN_2_6_1_1_p	-1998,09	2000,00	-1998,55	2000,00
RXN_2_6_1_11_c	0,00	2,87	0,00	2,18
RXN_2_6_1_13_c	-2,87	0,00	-2,18	0,00
RXN_2_6_1_19_m	0,00	2000,00	0,00	2000,00
RXN_2_6_1_19_m2	-2000,00	1489,22	-2000,00	0,00
RXN_2_6_1_2_c	-1361,90	2000,00	-188,40	2000,00
RXN_2_6_1_2_p	-1999,08	1362,82	-1998,27	190,13
RXN_2_6_1_4_p	-1495,15	-120,41	-137,75	-137,75
RXN_2_6_1_42__p	-3,20	-3,20	-2,42	-2,42
RXN_2_6_1_42_p	-2,32	-2,32	-1,76	-1,76
RXN_2_6_1_42_p2	-4,12	-4,12	-3,12	-3,12
RXN_2_6_1_45_p	0,00	1160,96	0,00	0,00
RXN_2_6_1_52_p	51,39	1338,97	77,14	77,14
RXN_2_6_1_79_p	3,37	3,37	2,55	2,55
RXN_2_6_1_9_p	0,90	0,90	0,68	0,68
RXN_2_7_1_1_c	0,00	870,72	0,00	0,00
RXN_2_7_1_1_p	0,00	580,48	0,00	0,00
RXN_2_7_1_11_c	0,00	1160,96	0,00	0,00
RXN_2_7_1_11_p	0,00	1160,96	0,00	0,00
RXN_2_7_1_19_p	97,19	595,86	91,31	91,31
RXN_2_7_1_31_p	0,00	1160,96	0,00	0,00

## Κεφάλαιο 1

## Αποτελέσματα

RXN_2_7_1_39_p	0,00	125,80	0,00	0,00
RXN_2_7_1_4_c	58,30	1082,82	288,54	288,54
RXN_2_7_1_40_c	0,00	2000,00	0,00	1934,46
RXN_2_7_1_40_p	0,00	2000,00	172,86	2000,00
RXN_2_7_1_71_p	3,84	3,84	2,91	2,91
RXN_2_7_1_90_c	-1082,97	1015,82	0,53	1011,98
RXN_2_7_1_90_p	-1997,72	132,28	-997,23	14,22
RXN_2_7_2_11_p	2,65	2,65	2,01	2,01
RXN_2_7_2_3_c	0,00	725,60	0,00	0,00
RXN_2_7_2_3_p	0,00	725,60	0,00	0,00
RXN_2_7_2_4_p	1,91	127,71	1,45	1,45
RXN_2_7_2_8_c	0,00	2,87	0,00	2,18
RXN_2_7_4_3_c	5,17	550,83	3,92	3,92
RXN_2_7_4_3_p	1,38	114,89	1,04	1,04
RXN_2_7_4_6_c	-58,30	812,42	-58,30	-58,30
RXN_2_7_4_6_p	0,00	0,00	0,00	0,00
RXN_2_7_6_1_p	1,38	1,38	1,04	1,04
RXN_2_7_6_3_p	0,00	0,00	0,00	0,00
RXN_2_7_7_18_p	0,00	0,00	0,00	0,00
RXN_2_7_7_27_c	0,00	16,49	0,00	12,49
RXN_2_7_7_27_p	0,00	16,49	0,00	12,49
RXN_2_7_7_4_p	0,00	0,00	0,00	0,00
RXN_2_7_7_9_c	-58,30	812,42	-58,30	-58,30
RXN_2381	0,48	0,48	0,36	0,36
RXN_2OXOGLUTARATEDEH	-1160,96	570,47	135,04	135,04
RXN_3_1_1_31_p	-317,59	-1,05	80,48	80,48
RXN_3_1_3_15_p	0,90	0,90	0,68	0,68
RXN_3_1_3_18_p	0,00	0,00	0,00	0,00
RXN_3_1_3_24_c	0,00	1024,52	230,24	230,24
RXN_3_1_3_3_p	51,39	1338,97	77,14	77,14
RXN_3_2_1_26_c	0,00	870,72	0,00	0,00
RXN_3_2_2_14_p	0,00	0,00	0,00	0,00
RXN_3_4_11_2_p	0,00	580,48	0,00	0,00
RXN_3_5_1_16_c	0,00	2,87	0,00	2,18
RXN_3_5_1_2_p	-1894,59	2000,00	-1828,99	2000,00
RXN_3_5_1_5_c	0,00	545,65	0,00	0,00
RXN_3_5_1_53_c	0,00	0,00	0,00	0,00
RXN_3_5_3_1_c	0,00	545,65	0,00	0,00
RXN_3_5_3_12_c	0,00	0,00	0,00	0,00
RXN_3_5_4_16_p	0,00	0,00	0,00	0,00
RXN_3_5_4_19_p	0,90	0,90	0,68	0,68
RXN_3_6_1_p	0,00	0,00	0,00	0,00
RXN_3_6_1_p2	0,00	0,00	0,00	0,00
RXN_3_6_1_1_c	-1068,94	1029,84	-1066,36	-54,91
RXN_3_6_1_1_p	0,00	2000,00	0,00	1011,45
RXN_3_6_1_31_p	0,90	0,90	0,68	0,68

RXN_4_1_1_p	0,00	0,00	0,00	0,00
RXN_4_1_1_1_c	0,00	125,80	0,00	9,31
RXN_4_1_1_15_c	0,00	1489,22	0,00	0,00
RXN_4_1_1_19_c	0,00	0,00	0,00	0,00
RXN_4_1_1_20_p	1,91	1,91	1,45	1,45
RXN_4_1_1_39_p	97,19	595,86	91,31	91,31
RXN_4_1_1_48_p	0,48	0,48	0,36	0,36
RXN_4_1_1_49_c	-154,88	2000,00	58,39	1992,85
RXN_4_1_2_13_p	-1223,67	1086,76	-997,23	14,22
RXN_4_1_2_13_c	-1013,24	1015,82	0,53	1011,98
RXN_4_1_2_25_p	0,00	0,00	0,00	0,00
RXN_4_1_2_5_p	-5,63	120,17	-4,26	-4,26
RXN_4_1_3_1_m	-813,12	-120,41	-137,75	-137,75
RXN_4_1_3_27_p	0,48	0,48	0,36	0,36
RXN_4_1_3_38_p	0,00	0,00	0,00	0,00
RXN_4_2_1_1_c	8,35	445,51	6,33	6,33
RXN_4_2_1_1_p	95,63	95,63	72,42	72,42
RXN_4_2_1_10_p	3,84	3,84	2,91	2,91
RXN_4_2_1_11_c	-706,03	19,58	14,83	14,83
RXN_4_2_1_11_p	17,19	868,59	105,47	105,47
RXN_4_2_1_19_p	0,90	0,90	0,68	0,68
RXN_4_2_1_2_m	-128,85	542,60	-0,53	-0,53
RXN_4_2_1_20_p	0,48	0,48	0,36	0,36
RXN_4_2_1_3_m	-125,80	0,00	0,00	0,00
RXN_4_2_1_3_m2	-125,80	0,00	0,00	0,00
RXN_4_2_1_33_p	4,12	4,12	3,12	3,12
RXN_4_2_1_52_p	1,91	1,91	1,45	1,45
RXN_4_2_1_9_p	2,32	2,32	1,76	1,76
RXN_4_2_1_9_p	7,33	7,33	5,55	5,55
RXN_4_2_1_91_p	2,04	2,04	1,55	1,55
RXN_4_2_3_1_p	-0,90	124,90	-0,68	-0,68
RXN_4_2_3_4_p	3,84	3,84	2,91	2,91
RXN_4_2_3_5_p	3,84	3,84	2,91	2,91
RXN_4_3_1_19_p	2,32	2,32	1,76	1,76
RXN_4_3_2_1_c	2,87	548,53	2,18	2,18
RXN_4_4_1_8_p	0,90	0,90	0,68	0,68
RXN_5_1_1_7_p	1,91	1,91	1,45	1,45
RXN_5_1_3_1_p	-611,17	-67,70	-8,88	-8,88
RXN_5_3_1_1_c	-1003,76	1359,04	-14,30	997,15
RXN_5_3_1_1_p	-1153,51	963,26	-1002,77	8,68
RXN_5_3_1_16_p	0,90	0,90	0,68	0,68
RXN_5_3_1_24_p	0,48	0,48	0,36	0,36
RXN_5_3_1_6_p	30,55	302,28	1,94	1,94
RXN_5_3_1_9_c	-1071,54	957,52	-57,77	953,68
RXN_5_3_1_9_p	-914,66	1430,93	-988,35	23,10
RXN_5_4_2_1_c	-706,03	19,58	14,83	14,83

RXN_5_4_2_1_p	17,19	868,59	105,47	105,47
RXN_5_4_2_2_c	-828,92	58,30	45,81	58,30
RXN_5_4_2_2_p	-16,49	0,00	-12,49	0,00
RXN_5_4_99_5_p	3,37	3,37	2,55	2,55
RXN_5257	-12,05	-12,05	-9,13	-9,13
RXN_6_2_1_1_p	0,00	113,51	0,00	0,00
RXN_6_2_1_5_m	-1160,96	570,47	135,04	135,04
RXN_6_3_1_2_p	0,00	1160,96	0,00	0,00
RXN_6_3_1_5_p	0,00	0,00	0,00	0,00
RXN_6_3_2_12_p	0,00	0,00	0,00	0,00
RXN_6_3_2_2_p	0,00	580,48	0,00	0,00
RXN_6_3_2_3_p	0,00	580,48	0,00	0,00
RXN_6_3_4_16_c	0,00	440,72	2,18	2,18
RXN_6_3_4_5_c	2,87	548,53	2,18	2,18
RXN_6_3_5_1_p	0,00	0,00	0,00	0,00
RXN_6_3_5_4_c	2,30	2,30	1,74	1,74
RXN_6_3_5_5_c	0,00	437,16	0,00	0,00
RXN_6_3_5_8_p	0,00	0,00	0,00	0,00
RXN_6_4_1_2_c	8,35	8,35	6,33	6,33
RXN_6_4_1_2_p	95,63	95,63	72,42	72,42
RXN_7800	4,12	4,12	3,12	3,12
RXN_GCVMULTI_p	-411,77	-45,69	-72,82	-72,82
RXN_nadkinase_c	0,00	0,00	0,00	0,00
RXN_nadkinase_p	0,00	0,00	0,00	0,00
RXN_PROLINE_MULTI	0,00	2,87	0,00	2,18
RXN_PSII_PSI_ATPs_p	156,64	423,86	0,00	0,00
RXN_SPONT_PRO_p	2,65	2,65	2,01	2,01
transp_acet_p	-1,69	111,82	-1,28	-1,28
transp_acetaldehyde_p	-120,17	5,63	4,26	4,26
transp_ADPGluc_p	0,00	16,49	0,00	12,49
transp_ala_m	-2000,00	0,00	-2000,00	0,00
transp_ala_p	-1362,82	1999,08	-190,13	1998,27
transp_atpAdp_m	839,04	2000,00	1027,52	1027,52
transp_atpAdp_p	0,00	2000,00	0,00	1827,14
transp_cit_p	-2000,00	2000,00	-2000,00	2000,00
transp_co2_m	-1360,81	2000,00	-1245,28	1727,21
transp_co2_p	253,46	1887,29	246,66	246,66
transp_cys_p	0,79	0,79	0,60	0,60
transp_dhap_p	-2000,00	2000,00	-2000,00	22,90
transp_gaba_m	-1489,22	0,00	0,00	0,00
transp_glc6p_p	-1496,19	1129,84	-907,87	116,07
transp_gln_p	-2000,00	1910,21	-1998,87	1830,12
transp_glu_m	-2000,00	2000,00	-2000,00	2000,00
transp_gluc_p	0,00	580,48	0,00	0,00
transp_gly_p	-4,79	-4,79	-3,63	-3,63
transp_glyoxylate_p2m	120,41	813,12	137,75	137,75

transp_his_p	0,90	0,90	0,68	0,68
transp_ileu_p	2,32	2,32	1,76	1,76
transp_ketoglCit_m	-2000,00	2000,00	-2000,00	2000,00
transp_ketoglSucc_m	-2000,00	2000,00	-2000,00	2000,00
transp_leu_p	4,12	4,12	3,12	3,12
transp_lys_p	1,91	1,91	1,45	1,45
transp_mal_p	-2000,00	2000,00	-2000,00	2000,00
transp_malAsp_p	-2000,00	1998,09	-2000,00	1998,55
transp_malGlu_p	-2000,00	2000,00	-2000,00	2000,00
transp_malKetog_m	-2000,00	2000,00	-2000,00	2000,00
transp_malKetog_p	-2000,00	2000,00	-2000,00	2000,00
transp_mallCit_m	-2000,00	2000,00	-2000,00	2000,00
transp_mallCit_p	-2000,00	2000,00	-2000,00	2000,00
transp_MalP_m	-2000,00	2000,00	-2000,00	2000,00
transp_met_p	0,90	0,90	0,68	0,68
transp_nh3_m	-2000,00	2000,00	-1863,32	2000,00
transp_nh3_p	-1958,22	2000,00	-1929,62	1899,37
transp_o2_m	178,51	444,44	198,33	198,33
transp_o2_p	-424,71	-157,49	-0,64	-0,64
transp_oaaAsp_m	-2000,00	2000,00	-2000,00	1873,44
transp_oaaCit_m	-2000,00	2000,00	-2000,00	2000,00
transp_oaaKetogl_m	-2000,00	2000,00	-2000,00	2000,00
transp_oaaMal_m	-2000,00	2000,00	-2000,00	2000,00
transp_oaaMal_p	-2000,00	2000,00	-1998,55	2000,00
transp_oaaSucc_m	-2000,00	2000,00	-2000,00	2000,00
transp_P_m	-1098,96	2000,00	-972,48	2000,00
transp_P_p	-0,90	1999,10	-0,68	1826,45
transp_pep_p	-1990,50	860,91	-1900,35	-73,22
transp_phe_p	2,04	2,04	1,55	1,55
transp_pro_p	2,65	2,65	2,01	2,01
transp_pyr_m	0,00	2000,00	0,00	2000,00
transp_pyr_p	0,00	2000,00	0,00	2000,00
transp_ser_p	-3,54	-3,54	-2,68	-2,68
transp_succFum_m	-548,53	-2,87	-2,18	-2,18
transp_succMal_m	-2000,00	2000,00	-2000,00	2000,00
transp_succP_m	-2000,00	2000,00	-2000,00	2000,00
transp_thr_p	2,41	2,41	1,83	1,83
transp_trp_p	0,48	0,48	0,36	0,36
transp_tyr_p	1,32	1,32	1,00	1,00
transp_val_p	3,20	3,20	2,42	2,42

**Πίνακας 1.5.** Ανάλυση Διαφοροποίησης Ροών (Flux Variability Analysis). Οι ροές εκφράζονται σε nmol.h<sup>-1</sup>.gDW<sup>-1</sup>.

### 1.3.3 Συστημική ανάλυση του χώρου ροών

Η Ανάλυση Πρωτευουσών Συνιστωσών μέσω της μεθόδου SVD δίνει πληροφορίες για τις ρυθμιστικές ιδιότητες του δικτύου των αντιδράσεων υπό διαφορετικές φυσιολογικές συνθήκες οι οποίες εκφράζονται μέσω της εφαρμογής περιορισμών (constraints) ως προς τη μεγιστοποίηση μιας αντικειμενικής συνάρτησης (Barrett et al., 2006).

Η μέθοδος αυτή εφαρμόστηκε σε 4 διαφορετικές συνθήκες, θέτοντας ως αντικειμενική συνάρτηση τη μεγιστοποίηση της παραγωγής είτε πλήρους βιομάζας, είτε μόνο των λιπιδίων. Συγκεκριμένα, οι 4 συνθήκες που αντιστοιχούν σε 4 διαφορετικές δειγματοληψίες (samplings) είναι οι παρακάτω:

- 1) Μεγιστοποίηση της παραγωγής πλήρους βιομάζας με φωτοσύνθεση (Sampling 1)
- 2) Μεγιστοποίηση της παραγωγής λιπιδίων με φωτοσύνθεση (Sampling 2)
- 3) Μεγιστοποίηση της παραγωγής πλήρους βιομάζας χωρίς φωτοσύνθεση (Sampling 3)
- 4) Μεγιστοποίηση της παραγωγής λιπιδίων χωρίς φωτοσύνθεση (Sampling 4)

Στον πίνακα 1.6 παρουσιάζονται οι 5 πρωτεύουσες συνιστώσες ρύθμισης για κάθε μια από τις παραπάνω περιπτώσεις και για ελάχιστη τιμή της αντικειμενικής συνάρτησης το 90% της μέγιστης τιμής.

Ιδιοδιάνυσμα διακύμανσης	Ποσοστό	Λειτουργική περιγραφή
<b>Δειγματοληψία 1 (Αύξηση πλήρους βιομάζας με φωτοσύνθεση)</b>		
1	61.68	Φωτοσύνθεση, αερόβια αναπνοή, γλυκόλυση
2	6.40	Μεταφορά αλανίνης (κυτταρόπλασμα/μιτοχόνδριο)
3	5.56	Κύκλος του Krebs, μεταβολισμός γλουταμικού και 4-αμινοβουτυρικού
4	3.71	Μεταβολισμός σακχαρόζης, κύκλος ουρίας, βιοσύνθεση αργινίνης
5	3.24	Μεταβολικό μονοπάτι γλυκίνης/σερίνης
<b>Δειγματοληψία 2 (Αύξηση λιπιδίων με φωτοσύνθεση)</b>		
1	52.90	Φωτοσύνθεση, αερόβια αναπνοή, γλυκόλυση

<b>2</b>	6.0	Κύκλος του Krebs, μεταβολισμός γλουταμικού και 4-αμινοβουτυρικού
<b>3</b>	5.1	Μεταφορά αλανίνης (κυτταρόπλασμα/μιτοχόνδριο)
<b>4</b>	4.38	Μεταβολισμός σακχαρόζης, κύκλος ουρίας
<b>5</b>	4.2	Μεταβολικό μονοπάτι γλυκίνης/σερίνης

**Δειγματοληψία 3 (Αύξηση πλήρους βιομάζας χωρίς φωτοσύνθεση)**

<b>1</b>	75.86	αερόβια αναπνοή, γλυκόλυση, πρόσληψη οξυγόνου, εκπομπή CO <sub>2</sub>
<b>2</b>	13.51	Μεταφορά αλανίνης (κυτταρόπλασμα/μιτοχόνδριο)
<b>3</b>	1.52	Κύκλος του Krebs, Μεταβολικό μονοπάτι γλυκίνης/σερίνης, λυάση του ισοκιτρικού
<b>4</b>	1.09	Μεταβολισμός σακχαρόζης, κύκλος ουρίας, βιοσύνθεση αργινίνης
<b>5</b>	0.78	Μεταβολισμός σακχαρόζης, αφομοίωση αμμωνίας, μετασχηματισμοί φυλλικού οξέος

**Δειγματοληψία 4 (Αύξηση λιπιδίων χωρίς φωτοσύνθεση)**

<b>1</b>	56.80	Μηλική αφυδρογονάση, αφομοίωση αμμωνίας, πρόσληψη σακχαρόζης, rubisco shunt, μονοπάτι φωσφοπεντοζών, γλυκόλυση (χλωροπλάστη), καρβοξυλάση του ακετυλο-συν. A, εκπομπή CO <sub>2</sub>
<b>2</b>	19.45	Μεταφορά αλανίνης (κυτταρόπλασμα/μιτοχόνδριο)
<b>3</b>	10.24	Μεταφορά φωσφο-ενολ-πυροσταφυλικού (κυτταρόπλασμα/χλωροπλάστης)
<b>4</b>	1.6	αερόβια αναπνοή, εκπομπή CO <sub>2</sub> , πρόσληψη οξυγόνου
<b>5</b>	1.1	μετασχηματισμοί φυλλικού οξέος, μεταφορά CO <sub>2</sub> (κυτταρόπλασμα / χλωροπλάστης), αφομοίωση αμμωνίας, γλυκόλυση, Μηλική αφυδρογονάση

**Πίνακας 1.6.** Σύνοψη των πέντε πρώτων πρωτευούσών συνιστωσών για κάθε μία από τις τέσσερεις δειγματοληψίες, οι οποίες αντιπροσωπεύουν διαφορετικά σενάρια ανάπτυξης. Το ποσοστό διακύμανσης (percentage variance) αντιπροσωπεύει τη σχετική συνεισφορά του κάθε ιδιο-διανύσματος (eigen-vector) στη συνολική διαφοροποίηση των ροών και αντιστοιχεί στην ιδιάζουσα τιμή (singular value) του ιδιο-διανύσματος στον πίνακα Σ, κανονικοποιημένο στο 100. Η λειτουργική περιγραφή των ιδιο-διανυσμάτων βασίζεται στις βιοχημικές επισημειώσεις των ενζύμων με τη μεγαλύτερη σε αυτά συνεισφορά.

Συγκρίνοντας τις πρωτεύουσες συνιστώσες των δειγματοληψιών 1 και 2, παρατηρούμε ότι δεν διακρίνεται διαφορά μεταξύ της παραγωγής πλήρους βιομάζας ή μόνο των λιπιδίων. Και στις δύο περιπτώσεις οι 5 πρώτες πρωτεύουσες συνιστώσες αφορούν τα ίδια βασικά μεταβολικά μονοπάτια. Αντίθετα, στις συνθήκες χωρίς φωτοσύνθεση (δειγματοληψίες 3 και 4) οι πρωτεύουσες συνιστώσες διαφέρουν. Συγκεκριμένα, στην παραγωγή πλήρους βιομάζας χωρίς φωτοσύνθεση (δειγματοληψία 1) οι πρώτη πρωτεύουσα συνιστώσα περιγράφει την αερόβια αναπνοή και την γλυκόλυση. Αντίθετα, στην παραγωγή μόνο λιπιδίων η πρώτη πρωτεύουσα συνιστώσα περιγράφει το μονοπάτι των φωσφορικών πεντοζών καθώς και το μονοπάτι παράκαμψης μέσω του ενζύμου οξυγενάση/καρβοξυλάση της 1,5-διφωσφορικής ριβουλόζης (rubisco shunt). Η παρατήρηση αυτή είναι σε απόλυτη συμφωνία με πειραματικές, *in vivo* παρατηρήσεις που έχουν δείξει ότι τα παραπάνω μεταβολικά μονοπάτια αποτελούν τις κύριες οδούς προς την παραγωγή λιπιδίων (Schwender et al. 2004). Σύμφωνα με τα πειράματα αυτά, η ροή άνθρακα ανακατευθύνεται προς το μονοπάτι των φωσφορικών πεντοζών και κυρίως στα ένζυμα οξυγενάση/καρβοξυλάση της 1,5-διφωσφορικής ριβουλόζης (4.1.1.39) και φωσφοριβουλοκινάσης (2.7.1.19). Η ανακατεύθυνση αυτή γίνεται εφικτή μέσω της παραγωγής φωτοσυνθετικού NADPH αλλά χωρίς τη συμβολή του κύκλου του Calvin στην πρόσληψη άνθρακα, λόγω της περιορισμένης ποσότητας φωτός που προσπίπτει στα έμβρυα της ελαιοκράμβης. Συμπερασματικά, στην προσομοίωση σε συνθήκες έλλειψης φωτοσύνθεσης το παραπάνω μονοπάτι αναδύεται ως η κύρια ρυθμιστική οδός παραγωγής λιπιδίων. Ο ρόλος του μονοπατιού αντικατοπτρίζεται στο ποσοστό 56,8% διαφοροποίησης του δικτύου (network variation).

#### **1.3.4 *In silico* διερεύνηση του ρόλου του μεταγραφικού παράγοντα WRINKLED1 (WRL1)**

Στα πλαίσια της εφαρμογής της μεθόδου SVD στην ανάλυση της μεταβολικής ρύθμισης, αλλά και ως επιβεβαίωση της λειτουργικότητας του μοντέλου, διερευνήθηκε η επίδραση του μεταγραφικού παράγοντα WRINKLED1 (WRL1). Έχει αποδειχθεί ότι ο παράγοντας WRL1 παίζει σημαντικό ρόλο στη βιοσύνθεση των λιπιδίων στους σπόρους του φυτού *Arabidopsis thaliana* (Baud et al. 2007). Συγκεκριμένα, η ισχυρή έκφραση του WRL1 είναι απαραίτητη για την συγκέντρωση υψηλής ποσότητας ελαίου στους σπόρους. Οι μεταλλάξεις knock-out για τον WRL1 προκαλούν τη δραματική μείωση της βιοσύνθεσης ελαίου, της τάξεως του 45-55%, με αποτέλεσμα οι σπόροι να αποκτούν ζαρωμένη όψη (wrinkled) (Baud et al. 2007).

Ο WRL1 ρυθμίζει την μεταγραφική έκφραση δύο τουλάχιστον γονιδίων, της χλωροπλαστικής κινάσης του πυροσταφυλικού οξέος (pyruvate kinase) και της μουτάσης του φωσφογλυκερικού οξέος (phosphoglycerate mutase). Ο WRL1 ελέγχει επίσης την παραγωγή λιποϊκού οξέως και βιοτίνης, που αποτελούν βασικούς συμπαράγοντες της δραστικότητας της αφυδρογονάσης του πυροσταφυλικού οξέος (pyruvate dehydrogenase) και της καρβοξυλάσης του ακετυλο-συνένζυμου A (acetyl-coa carboxylase) (Baud et al. 2007). Οι παραπάνω αντιδράσεις αποτελούν κύριους συνεισφορείς της πρώτης πρωτεύουσας συνιστώσας (component) της δειγματοληψίας 4 (Πίνακας 1.7).

<b>Δειγματοληψία 4 (Αντικειμενική συνάρτηση αύξησης λιπιδίων χωρίς φωτοσύνθεση)</b>			
<b>Αντίδραση</b>	<b>Συντελεστής συννεισφοράς (<math>\geq  0.1 </math>)</b>	<b>Έννυμα</b>	<b>Κυτταρικό διαμέρισμα</b>
<b>Πρωτεύουσα Συνιστώσα 1</b>			
1.1.1.38	0.19	μηλική αφυδρογονάση	χλωροπλάστης
6.3.1.2	0.15	γλουταμινική συνθετάση	χλωροπλάστης
Πρόσληψη σακχαρόζης	0.11	μεταφορέας σακχαρόζης	κυτταρική μεμβράνη
3.4.11.2	0.10	αμινοπεπτιδάση κυστεΐνυλ-γλυκίνης	χλωροπλάστης
2.3.2.2	0.10	γλουταθειόνη τρανσφεράση	χλωροπλάστης
5.1.3.1	-0.10	επιμεράση 3-φωσφορικής ριβουλόζης	χλωροπλάστης
4.1.1.39	-0.13	καρβοξυλάση/οξυγενάση της διφωσφορικής ριβουλόζης (rubisco)	χλωροπλάστης
2.7.1.19	-0.13	φωσφοριβουλοκινάση	χλωροπλάστης
5.4.2.1	-0.13	φωσφογλυκερική μουτάση	κυτταρόπλασμα
4.2.1.11	-0.13	φωσφοπυροσταφυλική υδρατάση	κυτταρόπλασμα
<b><u>5.4.2.1</u></b>	<b><u>-0.16</u></b>	<b><u>φωσφογλυκερική μουτάση</u></b>	<b><u>χλωροπλάστης</u></b>
4.2.1.11	-0.16	φωσφοπυροσταφυλική υδρατάση	χλωροπλάστης
<b><u>Pyruvate dehydrogenase</u></b>	<b><u>-0.20</u></b>	<b><u>σύμπλοκο αφυδρογονάσης πυροσταφυλικού</u></b>	<b><u>χλωροπλάστης</u></b>

<b><u>complex</u></b>			
1.1.1.49	-0.28	αφυδρογονάση της 6-φωσφορικής γλυκόζης	χλωροπλάστης
1.1.1.44	-0.28	φωσφογλυκονική αφυδρογονάση	χλωροπλάστης
3.1.1.31	-0.28	6-φωσφογλυκονογαλακτάση	χλωροπλάστης
CO <sub>2</sub> export	-0.28	μεταφορέας CO <sub>2</sub>	κυτταρική μεμβράνη
4.2.1.1	-0.29	αφυδρατάση του καρβονικού οξέος	χλωροπλάστης
<b><u>6.4.1.2</u></b>	<b><u>-0.29</u></b>	<b><u>καρβοξυλάση του ακετολο-</u></b>	<b><u>χλωροπλάστης</u></b>
		<b><u>συνενζ. Α</u></b>	
<b>Πρωτεύουσα Συνιστώσα 2</b>			
Alanine exchange	-1.00	μεταφορέας αλανίνης	μιτοχονδριακή μεβράνη
<b>Πρωτεύουσα Συνιστώσα 3</b>			
Phosphoenol pyruvate exchange	-1.00	μεταφορέας φωσφο-ενόλ- πυροσταφυλικού	μεμβράνη χλωροπλάστη
<b>Πρωτεύουσα Συνιστώσα 4</b>			
ATP synthase	0.59	ATP συνθάση	μιτοχόνδριο
ATP exchange	0.56	μεταφορέας ATP/ADP	μιτοχονδριακή μεβράνη
1.10.2.2	0.26	Αναγωγάση ουβικινόλης/κυτοχρώματος c	μιτοχόνδριο
1.6.5.3	0.22	αφυδρογονάση του NADH	μιτοχόνδριο
CO <sub>2</sub> export	0.16	Μεταφορέας CO <sub>2</sub>	κυτταρική μεμβράνη
1.9.3.1	0.13	οξειδάση του κυτοχρώματος c	μιτοχόνδριο
O <sub>2</sub> transport	0.13	μεταφορέας O <sub>2</sub>	μιτοχονδριακή μεβράνη
ATP maintenance	0.11	(ψευδο-αντίδραση)	μιτοχόνδριο
6.3.1.2	0.10	γλουταμινική συνθετάση	χλωροπλάστης
2.7.1.11	0.10	6-φωσφοφρουκτοκινάση	κυτταρόπλασμα,

O <sub>2</sub> import	-0.13	μεταφορέας O <sub>2</sub>	χλωροπλάστης κυτταρική μεμβράνη
<b>Πρωτεύουσα Συνιστώσα 5</b>			
1.5.7.1	0.41	αναγωγάση μεθυλεν- τετραϋδροφυλλικού οξ.	χλωροπλάστης
CO <sub>2</sub> exchange	0.35	μεταφορέας CO <sub>2</sub>	μεμβράνη χλωροπλάστη
6.3.1.2	0.22	γλουταμινική συνθετάση	μιτοχόνδριο
6.3.2.2	0.12	γλουταμόλ-κυττεϊνική λιγκάση	χλωροπλάστης
6.3.2.3	0.12	γλουταθιόνη συνθετάση	χλωροπλάστης
2.3.2.2	0.12	γλουταθειόνη τρανσφεράση	χλωροπλάστης
3.4.11.2	0.12	Κυττεϊνύλ-γλυκίνη αμινοπεπτιδάση	χλωροπλάστης
2.7.1.11	-0.24	6-φωσφοφρουκτοκινάση	χλωροπλάστης, κυτταρόπλασμα
1.1.1.38	-0.37	μηλική αφυδρογονάση	χλωροπλάστης
1.4.7.1	-0.41	Γλουταμική συνθάση	χλωροπλάστης
1.5.1.20	-0.41	αναγωγάση μεθυλεν- τετραϋδροφυλλικού οξ.	χλωροπλάστης

**Πίνακας 1.7.** Οι πέντε πρωτεύουσες συνιστώσες της δειγματοληγίας 4, με αντικειμενική συνάρτηση την μεγιστοποίηση παραγωγής λιπιδίων, χωρίς φωτοσύνθεση. Οι συντελεστές συνεισφοράς (loading coefficients) αντιπροσωπεύουν τη συνεισφορά της αντίδρασης στην πρωτεύουσα συνιστώσα (Σχήμα 1.2, πίνακας U). Η πρώτη πρωτεύουσα συνιστώσα αντιστοιχεί στον ρυθμιστικό ρόλο του παράγοντα WRL1 στη βιοσύνθεση των λιπιδίων, καθώς τριά από τα ενζύμα με τη μεγαλύτερη συνεισφορά αποτελούν στόχους του (υπογραμμισμένα).

Όπως φαίνεται επίσης στον πίνακα 1.8, οι παραπάνω αντιδράσεις είναι μεταξύ αυτών με το υψηλότερο ρυθμιστικό σκορ. Αφού το σκορ αυτό αποτελεί μέτρο της επίδρασης μιας αντίδρασης στην γενική κατανομή των μεταβολικών ροών, το αποτέλεσμα αυτό αναδεικνύει τον υψηλό ρυθμιστικό ρόλο του WRL1 στο μεταβολικό δίκτυο, μέσω του ελέγχου τριών τουλάχιστον ενζύμων, η δραστικότητα των οποίων καθορίζει σε μεγάλο βαθμό τη γενική κατανομή ροών.

<b>Διαγεγραμμένη αντίδραση</b>	<b>Ένζυμο</b>	<b>Ρυθμιστικός Βαθμός</b>	<b>Ποσοστό ρυθμού παραγωγής λιπιδίων</b>
Phosphoenol pyruvate transport to chloroplast	μεταφορέας φωσφο-ενόλ-πυροσταφυλικού οξ. (χλωροπλάστης)	0.99	82
Alanine transport to mitochondrion	μεταφορέας ανανίνης	0.99	100
<b>6.4.1.2 (p)</b>	<b>καρβοξυλάση</b>	<b>0.29</b>	<b>0</b>
	<b>ακετυλ-συνενζ. Α</b>		
4.2.1.1 (p)	Αφυδρατάση του καρβονικού οξέος	0.29	0
CO <sub>2</sub> export	Μεταφορέας CO <sub>2</sub>	0.28	67
3.1.1.31 (p)	6-φωσφο-γλυκονολακτονάση	0.28	0
1.1.1.44 (p)	Αφυδρογονάση φωσφο-γλυκονικού οξ.	0.28	0
1.1.1.49 (p)	αφυδρογονάση της 6-φωσφορικής γλυκόζης	0.28	0
<b>Pyruvate dehydrogenase complex (p)</b>	<b>Σύμπλοκο αφυδρογονάσης πυροσταφυλικού (p)</b>	<b>0.20</b>	<b>51</b>
1.1.1.38 (p)	μηλική αφυδρογονάση	0.19	13
4.2.1.11 (p)	φωσφοπυροσταφυλική υδρατάση	0.16	45
<b>5.4.2.1 (p)</b>	<b>φωσφογλυκερική μοντάση</b>	<b>0.16</b>	<b>45</b>

**Πίνακας 1.8.** Οι ρυθμιστικές αντιδράσεις με τα υψηλότερα σκορ της δειγματοληψίας 4, σε συνδυασμό με την επίδραση της διαγραφής τους στην παραγωγή ελαίου, εκπεφρασμένη ως ο λόγος του ρυθμού αύξησης των λιπιδίων μετά τη διαγραφή, προς τον ρυθμό αύξησης πριν τη διαγραφή. Το σκορ μετράει τη σχετική σημαντικότητα της αντίδρασης στη διαφοροποίηση των ροών του μεταβολικού δικτύου. Η διακύμανση μιας ροής με υψηλό σκορ ασκεί σημαντική επίδραση στη συνολική κατανομή ροών. Παρατηρείται ότι τρεις από τους κύριους στόχους του παράγοντα WRI1 βρίσκονται μεταξύ των αντιδράσεων με το υψηλότερο σκορ (υπογραμμισμένα).

Επιπρόσθετα, μελετήθηκε η επίδραση της διαγραφής των αντίστοιχων γονιδίων των παραπάνω ενζύμων στο ρυθμό παραγωγής λιπιδίων (Πίνακας 1.8, στήλη 4). Η διαγραφή των παραπάνω αντιδράσεων, με εξαίρεση την αντίδραση μεταφοράς αλανίνης από το κυτταρόπλασμα στο μιτοχόνδριο, έχει σημαντική επίδραση στον ρυθμό παραγωγής λιπιδίων. Η διαγραφή της καρβοξυλάσης του ακετυλο-συνενζύμου A (acetyl-coa carboxylase) (6.4.1.2) του χλωροπλάστη αναστέλλει εντελώς την παραγωγή λιπιδίων, παρά την ύπαρξη κυτταροπλασματικού ισόφορμου του ενζύμου αυτού.

Η παραγωγή ελαίου επίσης παρεμποδίζεται από την ανακοπή του οξειδωτικού κλάδου του μονοπατιού των φωσφορικών πεντοζών (3.1.1.31, 1.1.1.44, 1.1.1.49). Δεδομένου ότι το μονοπάτι αυτό δεν αποτελεί τη μοναδική πηγή αναγωγικού για τη βιοσύνθεση των λιπιδίων (Schwender et al. 2003) και σε συνδυασμό με την ολική αναστολή της βιοσύνθεσης αυτής στην προσομοίωση, συμπεραίνεται ότι η μετατροπή, έστω και σε μικρό βαθμό, της γλυκόζης-6-P σε ριβουλόζη-5-P είναι σημαντική για την δομή του μεταβολικού δίκτυου. Επιπρόσθετα, στη συγκεκριμένη προσομοίωση (δειγματοληψία 4) θεωρείται ότι δεν υπάρχει φωτοσύνθεση, η οποία αποτελεί κύρια πηγή NADPH για τη βιοσύνθεση των λιπιδίων, έτσι ώστε να δοθεί έμφαση στους υπόλοιπους παράγοντες. Ως εκ τούτου, στο *in vivo* σύστημα η απαλοιφή του οξειδωτικού κλάδου του μονοπατιού των φωσφορικών πεντοζών αναμένεται να έχει λιγότερο δραματικό αποτέλεσμα στην παραγωγή ελαίου.

Η απαλοιφή της PDC του χλωροπλάστη έχει ως αποτέλεσμα 49% μείωση του ρυθμού παραγωγής τριγλυκεριδίων, αφού το μεταβολικό δίκτυο έχει την ικανότητα να διοχετεύει μεγάλο μέρος της ροής άνθρακα στο κυτταροπλασματικό ισόφορμό της. Παρόλα αυτά, η διαγραφή της κυτταροπλασματικής PDC δεν έχει επίδραση στη βιοσύνθεση λιπιδίων. Τα αποτελέσματα αυτά δείχνουν ότι η παραγωγή των λιπιδίων εξαρτάται σε μεγαλύτερο βαθμό από τα χλωροπλαστικά ένζυμα, σε απόλυτη συμφωνία με τις πειραματικές παρατηρήσεις σε φυτά με μεταλλάξεις στον WRL1 παράγοντα.

Επιπρόσθετα, η Ανάλυση Πρωτεινουσών Συνιστωσών υποδεικνύει πιθανούς στόχους κυτταρικής ρύθμισης που δεν διακρίνονται μέσω της προσομοίωσης διαγραφής αντιδράσεων, όπως οι αντιδράσεις μεταφοράς αλανίνης και φωσφο-ενόλ-πυροσταφυλικού οξέως. Ενώ οι αντιδράσεις αυτές δεν έχουν επίδραση στο μέγιστο

ρυθμό παραγωγής λιπιδίων, έχουν μεγάλη επίδραση στην κατανομή των μεταβολικών ροών, όπως φαίνεται από τον ρυθμιστικός βαθμό τους (Πίνακας 1.8). Στο *in vivo* βιολογικό σύστημα η επίδραση αυτή αναμένεται να είναι ισχυρή, καθώς η ανακατανομή των μεταβολικών ροών σημαίνει ανακατανομή του ενεργειακού φορτίου. Συνεπώς, οι αντιδράσεις αυτές ελέγχουν έντονα τη θερμοδυναμική ισορροπία του μεταβολισμού.

## 1.4 Συμπεράσματα

Τα μεταβολικά μοντέλα μεγάλης κλίμακας αποτελούν σημαντικό εργαλείο στη μελέτη των βιοχημικών δικτύων ως συστήματα. Παρόλα αυτά, οι μέθοδοι γραμμικής βελτιστοποίησης που χρησιμοποιούνται στην FBA υπο-ορισμένων μοντέλων δεν καταλήγουν σε μια μοναδική βέλτιστη κατανομή ροών σταθερής κατάστασης. Η διερεύνηση του συνολικού χώρου λύσεων ενδείκνυται περισσότερο για την ανάλυση των ιδιοτήτων μεγάλων μεταβολικών δικτύων, ειδικά όταν δεν υπάρχουν διαθέσιμα δεδομένα που αφορούν τις κινητικές των αντιδράσεων. Τέτοια δεδομένα αφορούν πολύ μικρό αριθμό ενζύμων, συνήθως απομονωμένων, γεγονός που καθιστά αδύνατη τη χρησιμοποίησή τους στα πλαίσια ενός εκτεταμένου μεταβολικού δικτύου. Εν τούτοις, η ίδια η δομή του μεταβολικού δικτύου υπαγορεύει σε μεγάλο βαθμό ορισμένες από τις συστημικές ρυθμιστικές ιδιότητες του μεταβολισμού, όπως τη ιδιότητα της ευρωστίας (robustness), τη γονιδιακή ρύθμιση, τη πλαστικότητα καθώς έναν αριθμό πιθανών φαινοτύπων (phenotypic capabilities) (Patil and Nielsen 2005; Stelling et al. 2002). Επιπρόσθετα, τα μεγάλα στοιχειομετρικά μοντέλα μπορούν να αξιοποιήσουν θερμοδυναμικά δεδομένα, τα οποία μπορούν να δώσουν πιο ακριβή όρια των μεταβολικών ροών σύμφωνα με τη διαθεσιμότητα ενέργειας.

Στην παρούσα εργασία εγίνε επίδειξη της αποτελεσματικότητας των μεθόδων μείωσης διαστασιμότητας σε προβλήματα ρύθμισης βιοχημικών δικτύων, με σκοπό την σύνδεση της λειτουργικότητας του δικτύου με την τοπολογία του. Το βιοχημικό μοντέλου του κεντρικού μεταβολισμού της ελαιοκράμβης χρησιμοποιήθηκε με επιτυχία για την ρεαλιστική αναπαραγωγή ορισμένων λειτουργικών ιδιοτήτων του φυτικού εμβρύου, σε συμφωνία με πειραματικά δεδομένα. Συμπερασματικά, το μοντέλο αυτό, που προέκυψε από ολοκληρωμένη συσχέτιση γονιδιακών και μεταβολικών δεδομένων μπορεί να αποτελέσει τη βάση για πιο αναλυτική και συστηματική διευρεύνηση λειτουργικών πλευρών του φυτικού μεταβολισμού. Επισημαίνεται ότι το είδος της μοντελοποίησης αυτό, έχει ως πρωταρχικό σκοπό την εξαγωγή συμπερασμάτων αναδυόμενων ρυθμιστικών ιδιοτήτων που απορρέουν από τη δομή του δικτύου και έχουν φαινοτυπική επίδραση. Ως εκ τούτου, η εργασία αυτή συντελεί στην ανάπτυξη μεθοδολογιών που δίνουν έμφαση στο συστημικό χαρακτηρισμό μεταβολικών καταστάσεων, αντί για την ακριβή μέτρηση μεταβολικών ροών, η οποία απαιτεί δεδομένα ισοτοπικής σήμανσης (isotopic labeling) (Schwender et al. 2006). Η μέθοδος της ισοτοπικής σήμανσης προσφέρει ακριβείς μετρήσεις αλλά

υστερεί στην συστηματική πρόβλεψη διαφορετικών μεταβολικών καταστάσεων. Αντίθετα, στην παρούσα εργασία αποδεικνύεται ότι μια δομική μέθοδος ανάλυσης της κυτταρικής ρύθμισης μπορεί να προβλέψει τον ρόλο ενός μεταγραφικού παράγοντα (WRII) σε συμφωνία με πειραματικές παρατηρήσεις. Αν και η μέθοδος που χρησιμοποιήθηκε έχει προταθεί από τους Barrett et al., 2006, στην παρούσα διατριβή γίνεται για πρώτη φορά χρήση της για την μελέτη ενός συγκεκριμένου προβλήματος κυτταρικής ρύθμισης.

Το μοντέλο που παρουσιάζεται στην εργασία αυτή έχει προοπτική βελτίωσης μέσω του εμπλουτισμού με επιπρόσθετα δεδομένα και κυρίως την εισαγωγή κινητικών ή/και θερμοδυναμικών περιορισμών ως προς το εύρος των βιοχημικών αντιδράσεων. Επίσης μπορεί να επωφεληθεί από τη χρήση διαφορετικών μεθόδων βελτιστοποίησης (genetic algorithms, particle swarm optimization κ.α.)(Nakatsui et al. 2010) και να χρησιμοποιηθεί για τη μελέτη ακόμα πιο στοχευμένων μεταβολικών προβλημάτων. Η προσαρμογή του μοντέλου σε κάποιο συγκεκριμένο πρόβλημα μεταβολικής μηχανικής περνάει μέσω της χρήσης εξειδικευμένων περιορισμών (constraints) ώστε το πρόβλημα να εστιαστεί στο μέρος του μεταβολισμού που προορίζεται για ανάλυση. Ιδανικά, ένα στοιχειομετρικό μοντέλο μεγάλης κλίμακας μπορεί να συνδυαστεί με ειδικά σχεδιασμένα πειράματα για την εξαγωγή δεδομένων που προκύπτουν ως αντίδραση του βιολογικού συστήματος σε συστηματικές διαταραχές της ισορροπίας του (Plata et al. 2010). Η συνδυαστική χρήση υπολογιστικών μοντέλων με τα πειραματικά τους ανάλογα μπορεί να αυξήσει δραματικά την ακρίβεια και των δύο διαφορετικών προσεγγίσεων, της υπολογιστικής και της πειραματικής.

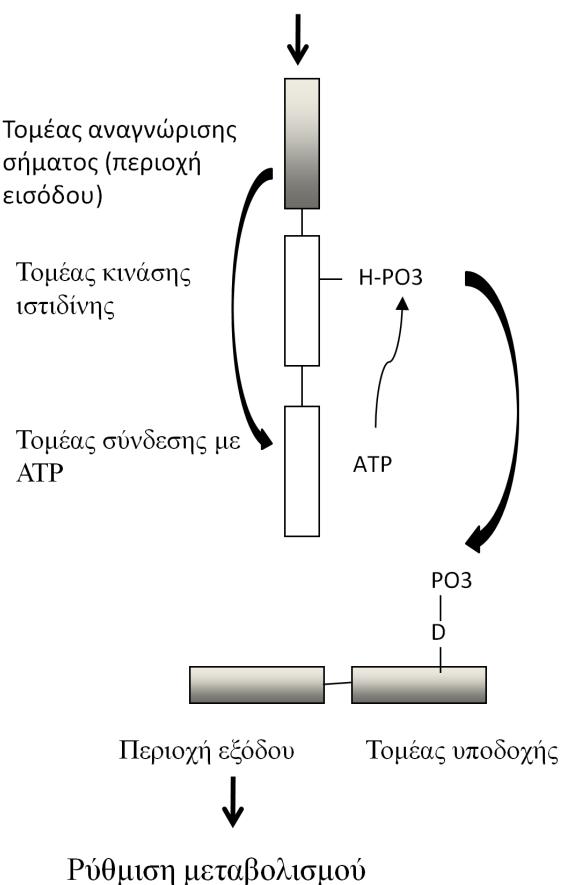
Επιπροσθέτως, η συστημική ανάλυση των ΜΜΓΚ μέσω της μεθόδου SVD προσφέρεται για την αυτοματοποιημένη διερεύνηση της δυναμικής απόκρισης μεγάλων μεταβολικών δικτύων, μέσω της συστηματικής εφαρμογής μεγάλου αριθμού τυχαίων περιορισμών και παρακολούθησης των πρωτευουσών συνιστώσων. Με τον τρόπο αυτό μπορεί να επιτυγχανθεί η διεξοδική ανάλυση της ευρωστίας του δικτύου και η ανάδειξη ρυθμιστικών κόμβων που χαρακτηρίζονται από αυξημένη ακαμψία (rigidity).

## Κεφάλαιο 2 – Βιοπληροφορική πρόβλεψη στόχων του μεταγραφικού παράγοντα AtoC στο γονιδίωμα του βακτηρίου *Escherichia coli*

### 2.1 Εισαγωγή

Τα βακτηριακά συστήματα σηματοδότησης (signal transduction systems) σχηματίζουν πολύπλοκα λειτουργικά δίκτυα που ελέγχουν τη φυσιολογία του κυττάρου. Πρωταρχικό ρόλο στα δίκτυα αυτά παίζουν τα συστήματα δυο συστατικών (Two Component Systems, TCS) (Hoch 2000) τα οποία εξασφαλίζουν την απόκριση του κυτταρικού μεταβολισμού σε εξωτερικά ερεθίσματα που δέχεται από το περιβάλλον. Η απόκριση συνίσταται στη μεταγραφική ρύθμιση γονιδίων που εξασφαλίζουν την ενεργοποίηση μηχανισμών προσαρμογής στις νέες συνθήκες που ανιχνεύονται από τα TCS. Τα TCS αποτελούνται από δυο πρωτεΐνικά μέρη, τον αισθητήρα και τον ρυθμιστή απόκρισης. Ο αισθητήρας, μια ιστιδίνη κινάση, αποτελείται από έναν τομέα αναγνώρισης σήματος (περιοχή εισόδου) που παρουσιάζει εξειδίκευση σε μοριακά σήματα και από έναν τομέα αυτοκινάσης, ο οποίος ενεργοποιείται από τη δέσμευση του σήματος στην περιοχή εισόδου. Η δέσμευση αυτή έχει ως αποτέλεσμα την υδρόλυση του ATP και την έναρξη ενός μηχανισμού φωσφομεταβίβασης που αρχίζει με τη φωσφορυλίωση μιας ιστιδίνης της αυτοκινάσης. Στη συνέχεια, η φωσφομεταβίβαση συνεχίζεται με την εξειδικευμένη φωσφορυλίωση ενός ρυθμιστή απόκρισης, ο οποίος είναι ένας μεταγραφικός παράγοντας που ελέγχει την έκφραση ενός ή περισσότερων γονιδίων με σκοπό την προσαρμογή κυτταρικών λειτουργιών ως προς το εξωτερικό ερέθισμα.

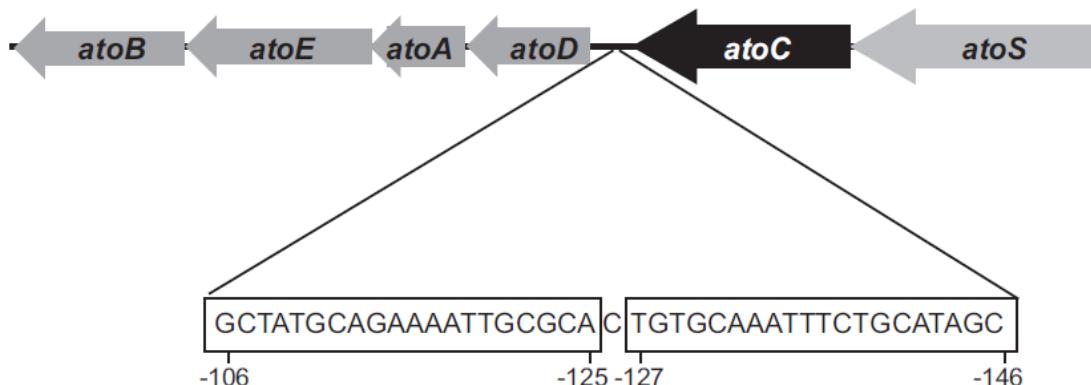
## Ερέθισμα περιβάλλοντος



**Σχήμα 2.1.** Τα συστήματα δύο συστατικών αποτελούνται από μια πρωτεΐνική κινάση ιστιδίνης που ελέγχεται από ερεθίσματα του περιβάλλοντος. Όταν η κινάση αυτή ενεργοποιηθεί από ένα εξωτερικό ερέθισμα (χημικό σήμα, μεταβολή θερμοκρασίας, ωσμωτικής πίεσης κ.α.), αυτοφωσφορυλώνεται σε μια ιστιδίνη. Η φωσφορική ομάδα στη συνέχεια μεταφέρεται σε ένα ασπαρτικό οξύ μιας πρωτεΐνης-ρυθμιστή απόκρισης και ενεργοποιεί έναν τομέα που μεσολαβεί για την απόκριση του μεταβολισμού στο εξωτερικό ερέθισμα.

Το σύστημα AtoS-AtoC (AtoSC) είναι ένα TCS που ενεργοποιεί τη μεταγραφή του οπερονίου *atoDAEB* στο βακτήριο *Escherichia coli*. Τα προϊόντα του οπερονίου *atoDAEB* εμπλέκονται στον καταβολισμό λιπαρών οξέων μικρής αλυσίδας (Kyriakidis et al. 2008; Theodorou et al. 2006; Theodorou et al. 2007). Το ακετοξικό οξύ αποτελεί επαγωγέα του συστήματος αυτού. Επιπρόσθετα, το σύστημα AtoSC εμπλέκεται σε άλλες μεταβολικές λειτουργίες του βακτηρίου *Escherichia coli*, ιδιαίτερα στη βιοσύνθεση πολύ-υδρόξυβουτυρικού οξέως (cPHB) (Theodorou et al. 2006) όπου παίζουν ρόλο η σπερμιδίνη (Theodorou et al. 2007) και η ισταμίνη (Kyriakidis et al. 2008).

Ο ρυθμιστής απόκρισης AtoC παρουσιάζει σημαντική ομολογία και χαρακτηριστικά με τους μεταγραφικούς παράγοντες NtrC-NifA οι οποίοι ενεργοποιούν την σ-54 RNA πολυμεράση (Grigoroudis et al. 2007). Ο παράγοντας AtoC έχει ένα πειραματικά προσδιορισμένο σημείο πρόσδεσης στον επαγωγέα του οπερονίου *atoDAEB*, το οποίο είναι ένα αντιστραμμένο παλίνδρομο των 40 νουκλεοτιδικών βάσεων, δηλαδή αποτελείται από δυο ίδιες αλλά αντιστραμμένες νουκλεοτιδικές αλληλουχίες των 20 βάσεων. Τα cis-ρυθμιστικά στοιχεία αυτά βρέθηκαν να ελέγχουν ταυτόχρονα την πρόσδεση του AtoC αλλά και το βαθμό ενεργοποίησης του επαγωγέα από το ακετοξικό οξύ. Επίσης, πειράματα ανοσοκατακρήμνισης της χρωματίνης (Chromatin Immunoprecipitation, ChIP) επιβεβαίωσαν την, επαγώγιμη από το ακετοξικό οξύ, σύνδεση του AtoC (Matta et al. 2007).



**Σχήμα 2.2.** Η θέση πρόσδεσης του μεταγραφικού παράγοντα AtoC στον επαγωγέα του οπερονίου *atoDAEB* αποτελείται από ένα αντιστραμμένο παλίνδρομο των 40 νουκλεοτιδικών βάσεων.

Ωστόσο, ο ρόλος του AtoSC δεν περιορίζεται μόνο στη ρύθμιση του οπερονίου *atoDAEB*. Μεταγραφωματικές (transcriptomics) και φαινοτυπικές αναλύσεις όλων των TCS στο *E. coli* υποδεικνύουν λειτουργική επικοινωνία μεταξύ αυτών (Oshima et al. 2002). Συγκεκριμένα, διαγραφή (deletion) των γονιδίων *atoSC* είχε ως αποτέλεσμα τη δραστική μεταβολή των προφίλ mRNA. Στο φαινοτυπικό επίπεδο προκάλεσε μειωμένη βακτηριακή κινητικότητα, μειωμένη ευαισθησία στα άλατα και σε ορισμένους περιβαλλοντικούς παράγοντες. Παρά την ως τώρα αποδεδειγμένη διασταυρούμενη ρύθμιση (cross-regulation) μεταξύ διαφόρων TCS του *E. Coli* (Yamamoto et al. 2005), η παρατήρηση των παραπάνω φαινοτύπων υποδεικνύει και εξειδικευμένο ρόλο του AtoSC. Συμπερασματικά, είναι επιτακτική η ανάγκη

αναγνώρισης νέων γονιδιακών στόχων του AtoSC για τη διαλεύκανση του διευρυμένου μεταβολικού ρόλου του συστήματος αυτού.

Διάφορες μεθοδολογίες υπάρχουν για την πρόβλεψη και ανίχνευση σημείων προσδέσεων μεταγραφικών παραγόντων (TFBS). Οι περισσότερες από αυτές βασίζονται σε έναν αριθμό ήδη πειραματικά προσδιορισμένων γονιδιακών στόχων, οι οποίοι περιέχουν *cis*-ρυθμιστικά στοιχεία στους επαγωγείς τους. Οι αλληλουχίες DNA των επαγωγέων τους, πριν τη μεταγραφική έναρξη (upstream), χρησιμοποιούνται ως δεδομένα εισόδου σε διάφορους μαθηματικούς αλγόριθμους (Gibbs Sampling) (Thijs et al. 2002), όπως η υπεραντιπροσώπευση νουκλεοτιδίων (van Helden et al. 1998)) με σκοπό την κατασκευή μοτίβων. Τα μοτίβα αυτά είναι μαθηματικές αναπαραστάσεις διατηρημένων υπο-αλληλουχιών σε σχέση με το μη διατηρημένο θόρυβο υποβάθρου. Ο στόχος των μεθόδων αυτών είναι η εύρεση διατηρημένων αλληλουχιών DNA σε ομάδες γονιδίων που συσχετίζονται στο μεταγραφικό επίπεδο. Τα διατηρημένα μοτίβα (patterns) τέτοιων ομάδων γονιδίων έχουν πολύ μεγάλη πιθανότητα να παίζουν κάποιο βιολογικό ρόλο, ως *cis*-ρυθμιστικά στοιχεία που ελέγχουν την πρόσδεση μεταγραφικών παραγόντων. Οι ομάδες συσχετιζόμενων γονιδίων τυπικά προσδιορίζονται με ανάλυση μεταγραφώματος που εντοπίζουν γονίδια με συσχετισμένη έκφραση σε συγκεκριμένες συνθήκες (Roth et al. 1998). Άλλες μέθοδοι περιλαμβάνουν την αξιοποίηση φυλογενετικής πληροφορίας, μέσω της συγκέντρωσης ορθόλογων μεσογονιδιακών περιοχών από διαφορετικούς οργανισμούς. Οι μέθοδοι αυτές ανξάνουν την ακρίβεια ανίχνευσης ρυθμιστικών DNA περιοχών, καθώς βασίζονται στο γεγονός ότι βιολογικά σημαντικές μεσογονιδιακές περιοχές έχουν πολύ μεγαλύτερη πιθανότητα διατήρησης μεταξύ διαφορετικών ειδών (Janky and van Helden 2008; Marchal et al. 2004; Newberg et al. 2007). Τέλος, υπάρχουν μεθοδολογίες που χρησιμοποιούν πληροφορίες γονιδιακής λειτουργίας και μεταβολικών μονοπατιών ώστε να εξαγάγουν συσχετισμούς γονιδιακής ρύθμισης (Hughes et al. 2000; McGuire et al. 2000; Mullapudi et al. 2007). Το κοινό χαρακτηριστικό όλων των παραπάνω μεθοδολογιών είναι η εκμετάλλευση ομάδων γονιδίων με κοινή ρύθμιση ή έστω η ύπαρξη ικανού όγκου δεδομένων που προέρχονται από πειράματα μεταγραφωματικής ανάλυσης, έτσι ώστε να μπορούν να βρεθούν οι αλληλοσυσχετισμοί έκφρασης γονιδίων. Στην περίπτωση του συστήματος AtoSC η χρησιμοποίηση μιας από τις προαναφερθείσες μεθόδους δεν κατέστη δυνατή, καθώς δεν υπήρχε διαθεσιμότητα ούτε εκτεταμμένων

αναλύσεων μεταγραφώματος, ούτε άλλοι γνωστοί γονιδιακοί στόχοι, εκτός του οπερονίου *atoDAEB*. Στα πλαίσια της παρούσης διατριβής, αναπτύχθηκε λοιπόν μια καινούρια μεθοδολογία ανίχνευσης στόχων πρόσδεσης του μεταγραφικού παράγοντα AtoC, προσαρμοσμένη στην επίλυση του παραπάνω προβλήματος, δηλαδή στην εύρεση γονιδιακών στόχων με βάση τη γνώση ενός μοναδικού σημείου πρόσδεσης, καθώς και τη σύνδεση των μοτίβων DNA με βιολογικά μονοπάτια. Η μέθοδος αυτή βασίζεται στον εμπλουτισμό γονιδιακών οντολογιών (Gene Ontology Term Enrichment) και χρησιμοποιήθηκε με επιτυχία για τον εντοπισμό νέων σημείων πρόσδεσης του AtoC σε όλο το γονιδίωμα του *E. coli*. Οι νέοι γονιδιακοί στόχοι επιβεβαιώθηκαν πειραματικά για την *in vivo* σύνδεση του AtoC, με την τεχνική ChIP. Ελέγχθηκε επίσης η γενική λειτουργικότητα της μεθόδου με την εφαρμογή της στον μεταγραφικό παράγοντα LexA, που διαθέτει πλήθος γνωστών γονιδιακών στόχων.

## 2.2 Μέθοδοι

### 2.2.1 Γενική περιγραφή της υπολογιστικής διαδικασίας ανίχνευσης και εκτίμησης μοτίβων

Προκειμένου να γίνει δυνατή η ανίχνευση νέων γονιδιακών στόχων του AtoC ξεκινώντας από το μοναδικό γνωστό σημείο πρόσδεσης του, αναπτύχθηκε μια υπολογιστική διαδικασία ανίχνευσης μοτίβων και εκτίμησης της βιολογικής τους σημασίας (Σχήμα 2.1).

Η διαδικασία αυτή χωρίζεται στα παρακάτω κύρια μέρη:

#### A) Κατασκευή μοτίβων μέσω επαναληπτικής δειγματοληψίας (Iterative Motif Sampling)

Στο στάδιο αυτό επιχειρείται η κατασκευή ενός μοτίβου που αντιπροσωπεύει μια συντηρημένη αλληλουχία DNA μεταξύ του γνωστού σημείου πρόσδεσης και άλλων πιθανών σημείων προσδέσης σε διαφορετικές περιοχές του βακτηριακού γονιδιώματος. Η διαδικασία περιλαμβάνει τη σύκριση του επαγωγέα *atoDAEB* με κάθε έναν από μια ομάδα επαγωγών που μπορεί να είναι ειδικά επιλεγμένη ή ακόμα και το σύνολο των επαγωγών του γονιδιώματος. Όταν βρεθεί επαγωγέας που περιλαμβάνει αλληλουχία DNA συντηρημένη σε σχέση με το γνωστό σημείο πρόσδεσης, ο επαγωγέας αυτός ομαδοποιείται με τον επαγωγέα *atoDAEB*. Η διαδικασία επαναλαμβάνεται μεταξύ των εναπομεινάντων επαγωγών για να προστεθούν και άλλοι, μεγαλώνοντας την ομάδα των επιλεγμένων επαγωγών οι οποίοι στο σύνολό τους περιλαμβάνουν μια συντηρημένη περιοχή DNA που περιγράφεται από ένα μοτίβο.

#### B) Λειτουργική ανάλυση μοτίβων και σύνδεσή τους με βιολογικές λειτουργίες και μεταβολικά μονοπάτια

Γενικά, ο γενετικός κώδικας των μεσογονιδιακών περιοχών εμφανίζει πολύ μικρή συντήρηση. Οι ρυθμιστικές θέσεις πρόσδεσης των μεταγραφικών παραγόντων δεν αποτελούν εξαίρεση, παρουσιάζοντας με τη σειρά τους πολύ μικρή ομολογία αλληλουχίας, γεγονός που καθιστά δύσκολη τη διάκρισή τους από τις γειτονικές, μη λειτουργικές θέσεις DNA. Κατά συνέπεια, τα μοτίβα χαρακτηρίζονται από αυξημένο θόρυβο και ενδέχεται να αντιπροσωπεύουν τυχαία ομοιότητα αλληλουχίας αντί για συντηρημένες λειτουργικές θέσεις. Προκειμένου να αποτιμάται η λειτουργική

σημασία των μοτίβων που προκύπτουν από την επαναληπτική δειγματοληψία, αναπτύχθηκε μια μεθοδολογία συσχέτισης μοτίβων με βιολογικές λειτουργίες, βασισμένη στον εμπλουτισμό γονιδιακών οντολογιών (Gene Ontology Term Enrichment).

Η συνολική διαδικασία περιγράφεται λεπτομερώς στη συνέχεια.

### **2.2.2 Πληροφορικά και μαθηματικά εργαλεία**

#### **Περιβάλλον προγραμματισμού**

Η υπολογιστική διαδικασία αναπτύχθηκε σε περιβάλλον Linux και στην γλώσσα προγραμματισμού Perl. Οι ρουτίνες Perl (Παράρτημα Α) συνδυάστηκαν με πακέτο INCLUSIVE (Thijs et al. 2001) που περιλαμβάνει τα προγράμματα κατασκευής και χρήσης μοτίβων MotifSampler, MotifSearch και MotifRanking. Χρησιμοποιήθηκε επίσης η γλώσσα διαχείρισης βάσεων δεδομένων MySQL για την αποθήκευση και διαχείριση όλων των δεδομένων των αλληλουχιών, των επισημειώσεων και των ενδιάμεσων αποτελεσμάτων των ρουτινών Perl.

#### **Ορισμός Μοτίβων και Motif Sampling**

Ως μοτίβο ορίζεται η μαθηματική αναπαράσταση μιας συντηρημένης δομής DNA, η οποία δίνει την πιθανότητα να υπάρχει μια συγκεκριμένη νουκλεοτιδική βάση σε κάθε θέση μιας αλληλουχίας DNA. Συνεπώς, ένα μοτίβο μήκους  $m$  αντιστοιχεί σε ένα πίνακα διαστάσεων  $m * 4$ , λόγω των τεσσάρων νουκλεοτιδικών βάσεων. Ο αλγόριθμος δειγματοληψίας (motif sampling) δέχεται ως δεδομένο εισόδου ένα σύνολο αλληλουχιών DNA και εξάγει συντηρημένες περιοχές με τη μορφή μοτίβων.

#### **Επιλογή επαγωγικών περιοχών**

Ως επαγωγικές περιοχές ορίστηκαν στο πρόγραμμα οι αλληλουχίες από τη θέση -250 ως +30 από την έναρξη της μεταγραφής (Transcription Start Site). Οι επισημειώσεις (annotations) της έναρξης της μεταγραφής και του γονιδιώματος του *Escherichia coli* K-12 εξήχθησαν από τη βάση Refseq (Pruitt et al. 2007) (refseq id: NC\_000913) με τη χρήση ρουτινών Perl και Bioperl. Μεταξύ διαφορετικών επαγωγέων δεν επιτράπηκε επικάλυψη και όλες οι μεσογονιδιακές περιοχές μικρότερες των 250 βάσεων, μεταξύ δυο γονιδίων κοδικωποιημένων αντίστοιχα στις -1 και 1 χρωματίδες (strands) εισήχθησαν αυτούσιες και θεωρήθηκαν ως επαγωγικές και των δυο αυτών γονιδίων.

## Ορισμός ομάδων επαγωγέων

Ως αρχική ομάδα επαγωγέων, AOE (initial promoter set) ορίστηκε η ομάδα όλων των επαγωγέων που θα χρησιμοποιηθούν στην επαναληπτική δειγματοληψία. Από την επιλογή της AOE εξαρτάται άμεσα το τελικό μοτίβο, καθώς διαφορετικές ομάδες επαγωγέων παρουσιάζουν διαφορετικούς συνδυασμούς συντηρημένων περιοχών DNA. Μια περιορισμένη AOE ενδέχεται να αποδώσει μοτίβα που παρουσιάζουν κάποια συσχέτιση με το κριτήριο περιορισμού.

Στον πίνακα 2.1 παρουσιάζονται οι τρεις διαφορετικές ομάδες που χρησιμοποιήθηκαν στις τρεις αντίστοιχες εκτελέσεις της διαδικασίας.

- Το σύνολο των επαγωγέων του *E. coli*.
- Οι επαγωγείς εξαρτώμενοι από τον παράγοντα sigma-54, όπως εξήχθησαν από τη βάση RegulonDB (Gama-Castro et al. 2008) και συγκεκριμένα οι επαγωγείς των παρακάτω γονιδίων: *argT, astC, chaC, dcuD, ddpX, fdhF, glmY, glnA, glnH, glnK, gltI, hycA, hydA, hyfA, hypA, ibpB, kch, nac, norV, potF, prpB, pspA, pspG, puuP, rpoH, rtcB, rtcR, rutA, yaiS, ybhK, yeaG, yfhK, ygjG, yhdW, zraP, zraS*.
- Οι επαγωγείς που περιλαμβάνουν γνωστές θέσεις πρόσδεσης άλλων ρυθμιστών απόκρισης, επισημειωμένων στην βάση Ecocyc (Keseler et al. 2005). Συγκεκριμένα: *dmsA, emrK, gadE, yca , spy, cydA, sucA, cpx, rdoA, ppiA, yqjA, acrD, ompC, mdtA, ftnB, yebE, ydeH, ycfS, efeU, yccA, aroG, ppiD, degP, ybaJ, cusR, dctA, frdA, dcuB, astC, kdpF, nrfA, nirB, hcp, napF, tppB, ompF, csgD, psiE, pstS, asr, phoH, phoB, phoA, phoE, mgtA, yrbL, mgrB, rstA, nagA, borD, hemL, ftsA, osmC, tnaA, torC, uhpT, zraS*.

Η ομάδα επαγωγέων εισόδου, ΟΕΕΣ (input promoter set) είναι η ομάδα των επαγωγέων που εισάγονται σε κάθε επανάληψη της δειγματοληψίας. Στην πρώτη επανάληψη η ΟΕΕ ταυτίζεται με την AOE και επιχειρείται δειγματοληψία μεταξύ αυτής και του επαγωγέα *atoDAEB*.

Οι επιλεγμένοι επαγωγείς αποτελούν την ομάδα επαγωγέων εξόδου, ΟΕΕΞ (qualified promoter set) είναι το αποτέλεσμα της δειγματοληψίας και περιλαμβάνει τον επαγωγέα *atoDAEB* μαζί με ένα σύνολο επαγωγέων που ορίζουν μια συντηρημένη περιοχή DNA και άρα ένα μοτίβο. Η ομάδα αυτή αυξάνεται κατά έναν επαγωγέα σε κάθε επανάληψη, αντλώντας επαγωγείς από την ΟΕΕΣ, η οποία

αντίστοιχα μειώνεται. Οι επαγωγείς που δεν περιέχουν συντηρημένη περιοχή DNA απορρίπτονται και η διαδικασία σταματάει όταν εξαντληθούν οι επαγωγείς ή δεν βρίσκεται κάποιος με συντηρημένη περιοχή DNA.

### **Όροι γονιδιακών οντολογιών**

Ο όροι οντολογιών (Ontology Terms), αποτελούν μια μέθοδο δομημένης και ιεραρχημένης αναπαράστασης της γνώσης. Κάθε όρος οντολογίας αντιστοιχεί σε μια θεμελιώδη έννοια ή οντότητα που ορίζεται χωρίς αμφισημία και άρα είναι μοναδική. Μια οντολογία αποτελεί ένα λεξιλόγιο όρων, οι οποίοι έχουν σαφή και ιεραρχημένη σχέση μεταξύ τους. Οι οντολογίες στην πληροφορική αναπτύχθηκαν ως μέσο δομημένης αναπαράστασης εννοιών, έτσι ώστε να επιτρέπει την αυτόματη εξαγωγή συμπερασμάτων. Στα πλαίσια της βιολογικής επιστήμης έχει αναπτυχθεί ένας αριθμός οντολογιών, εκ των οποίων και η *Γονιδιακή Οντολογία* (Gene Ontology) (Ashburner et al. 2000). Η γονιδιακή οντολογία αποτελεί ένα δομημένο λεξιλόγιο βιολογικών επισημειώσεων (annotations) των προϊόντων των γονιδίων. Χωρίζεται σε τρεις επιμέρους τομείς, οι οποίοι καλύπτουν τη μοριακή λειτουργία του γονιδιακού προϊόντος (molecular function), την βιολογική διαδικασία στην οποία εμπλέκεται (biological process) και τον εντοπισμό του στο κύτταρο (cellular component). Συνεπώς, η χρήση γονιδιακών οντολογιών δίνει το πλεονέκτημα της αυτοματοποιημένης διαχείρισης της γονιδιακής πληροφορίας και επιτρέπει την στατιστική ανάλυσή της σε μεγάλη κλίμακα.

### **2.2.3 Επαναληπτική διαδικασία de novo ανίχνευσης συντηρημένης περιοχής DNA και κατασκευή μοτίβων**

Η διαδικασία ανίχνευσης συντηρημένων περιοχών DNA πραγματοποιείται με επαναληπτικό τρόπο και χρησιμοποιεί τον αλγόριθμο δειγματοληψίας Gibbs (Gibbs sampling) υλοποιημένου στο πακέτο προγραμμάτων INCLUSIVE (Thijs et al. 2001).

Κατά την πρώτη επανάληψη πραγματοποιείται δειγματοληψία μεταξύ του επαγωγέα *atoDAEB* και κάθε έναν από τους επαγωγείς της AOE ξεχωριστά, χρησιμοποιώντας τις εξής παραμέτρους: εκτελέσεις δειγματοληψίας 100, μήκος μοτίβου 20, θέση μοτίβου στον επαγωγέα *atoDAEB* -184 ως -165 (ως προς την έναρξη της μεταγραφής), μοντέλο υποβάθρου 3 (χρησιμοποιήθηκε το διαθέσιμο μοντέλο του προγράμματος INCLUSIVE για το *E. coli* K-12), αρχική πιθανότητα (prior probability) 0,5.

Γενικότερα, σε κάθε επανάληψη εφαρμόζεται ο εξής αλγόριθμος:

- Ένα log-likelihood score (LLS) υπολογίζεται για κάθε μοτίβο που προκύπτει από δειγματοληψία μεταξύ κάθε ενός ξεχωριστά επαγωγέα της ΟΕΕΣ και του ΟΕΕΞ που έχει προκύψει από τις προηγούμενες επαναλήψεις. Σε κάθε εκτέλεση παράγεται μεγάλος αριθμός μοτίβων τα οποία δεν αντιστοιχούν απαραίτητα στις θέσεις πρόσδεσης του AtoC. Κατά συνέπεια το πρόγραμμα λαμβάνει υπόψη μόνο τα μοτίβα που αντιστοιχεί στην ορισμένη θέση πρόσδεσης στον επαγωγέα *atDAEB*. Το LLS που αντιστοιχεί στη θέση αυτή ορίζεται ως Position Specific Log-likelihood Score (PSLLS).
- Ο επαγωγέας της ΟΕΕΣ που αποδίδει το μοτίβο με το μέγιστο PSLLS παρακρατείται και προστίθεται στην ΟΕΕΞ. Συνεπώς, η ΟΕΕΞ αυξάνεται σε κάθε επανάληψη με τη συγκέντρωση όλων των επαγωγέων που παρουσιάζουν συντηρημένη αλληλουχία DNA που αντιστοιχεί στην γνωστή θέση πρόσδεσης του AtoC.
- Οι επαγωγέις που δεν αποδίδουν μοτίβο ή αποδίδουν με PSLLS χαμηλότερο από μια ορισμένη τιμή απορρίπτονται στο τέλος της επανάληψης. Ως ελάχιστη τιμή ορίστηκε το 30ό εκατοστημόριο του συνόλου των PSLLS. Η απόρριψη αυτή είναι απαραίτητη για τη μείωση του χρόνου εκτέλεσης του προγράμματος αλλά και για την αποφυγή υπερβολικής αύξησης του θορύβου στην ΟΕΕΞ λόγω της πρόσθεσης μεγάλου αριθμού επαγωγέων με μικρή συντήρηση αλληλουχίας DNA.

Στο τέλος της διαδικασίας αποδίδεται ένα μοτίβο με μέγιστο PSLLS, το οποίο αντιπροσωπεύει μια συντηρημένη σε πολλούς επαγωγέις αλληλουχία DNA που αντιστοιχεί στη γνωστή θέση πρόσδεσης του AtoC. Συνεπώς, το μοτίβο αυτό υποδεικνύει ομόλογες θέσεις πρόσδεσης σε διαφορετικά σημεία του γονιδιώματος, όπου ο μεταγραφικός παράγοντας AtoC ενδέχεται να έχει ρυθμιστικό ρόλο.

## 2.2.4 Λειτουργική ανάλυση μοτίβων

Η λειτουργική ανάλυση των μοτίβων περιλαμβάνει δυο συζευγμένα στάδια, την αναζήτηση θέσεων του μοτίβου πάνω στο γονιδίωμα (Motif Searching, MS) και την ανάλυση των γονιδιακών οντολογιών (Gene Ontology Term Analysis, GOT).

- **Αναζήτηση θέσεων του μοτίβου σε ολόκληρο το γονιδίωμα (Genome-wide Motif Searching, GWMS)**

Οι PSPM των μοτίβων που προέκυψαν από την επαναληπτική διαδικασία χρησιμοποιούνται στο στάδιο αυτό για την αναζήτηση των στόχων τους, δηλαδή των θέσεων DNA που παρουσιάζουν το αντίστοιχο πρότυπο (pattern). Η διαδικασία GWMS εκτελείται δυο φορές, μία στους επαγωγείς (μεσογονιδιακές περιοχές) και μια στις ενδογονιδιακές περιοχές, με αρχική πιθανότητα (prior probability) 0.7. Για την οπτικοποίηση των αλληλουχιών χρησιμοποιήθηκε το πρόγραμμα *WebLogo* (Crooks et al. 2004).

- **Ανάλυση γονιδιακών οντολογιών (Gene Ontology Terms, GOT)**

Προκειμένου να εκτιμηθεί ο λειτουργικός ρόλος των μοτίβων, οι στόχοι τους αντιστοιχίζονται σε γονίδια, επειδή η συντηρημένη θέση που το μοτίβο περιγράφει βρίσκεται είτε στον επαγωγέα τους, είτε πάνω στην ίδια τη γονιδιακή περιοχή. Στη συνέχεια, εφαρμόζεται το υπεργεωμετρικό τεστ για την εύρεση υπερ-αντιπροσώπευσης GOT μεταξύ των γονιδίων αυτών. Σε περίπτωση υπερ-αντιπροσώπευσης ενός GOT, συνάγεται η στατιστικά σημαντική συσχέτιση του μοτίβου με την αντίστοιχη βιολογική λειτουργία. Για να καθοριστεί αν ένας GOT βρίσκεται στις επισημειώσεις μιας ομάδας γονιδίων σε συχνότητα μεγαλύτερη από την τυχαία αναμενόμενη, υπολογίζεται μια τιμή *p-value* χρησιμοποιώντας την Υπεργεωμετρική Κατανομή (Boyle et al. 2004):

$$P=1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{i}}$$

Όπου  $N$ , ο συνολικός αριθμός γονιδίων στην κατανομή υποβάθρου,  $M$  ο αριθμός των γονιδίων/επαγωγέων της κατανομής που είναι επισημειωμένα στον GOT ενδιαφέροντος,  $n$  το μέγεθος της λίστας των γονιδίων/επαγωγέων ενδιαφέροντος και

*k* ο αριθμός των γονιδίων μέσα στη λίστα που είναι επισημειωμένα με τον συγκεκριμένο GOT. Ως κατανομή υποβάθρου χρησιμοποιήθηκαν όλα τα γονίδια του *E. coli* K-12 περιέχονταν στο αρχείο επισημειώσεων. Στο υπεργεωμετρικό τεστ εφαρμόζεται δειγματοληψία χωρίς επανατοποθέτηση.

Το τεστ υλοποιήθηκε με τη χρήση του προγράμματος Perl GO::TermFinder (Boyle et al. 2004). Οι επισημειώσεις των GOT εξήχθησαν από τη βάση EBI Gene Ontology Annotation (GOA) (Camon et al. 2003) για το *E. coli* K-12 (84.4 % κάλυψη γονιδιώματος). Ως ελάχιστη p-value για το υπεργεωμετρικό τεστ ορίστηκε το 0.1.

### **2.2.5 Πειραματική επιβεβαίωση επιλεγμένων στόχων των μοτίβων για *in vivo* σύνδεση με τον ρυθμιστή απόκρισης AtoC**

Οι μέθοδοι της πειραματικής επιβεβαίωσης παρατίθενται στο Παράρτημα B.

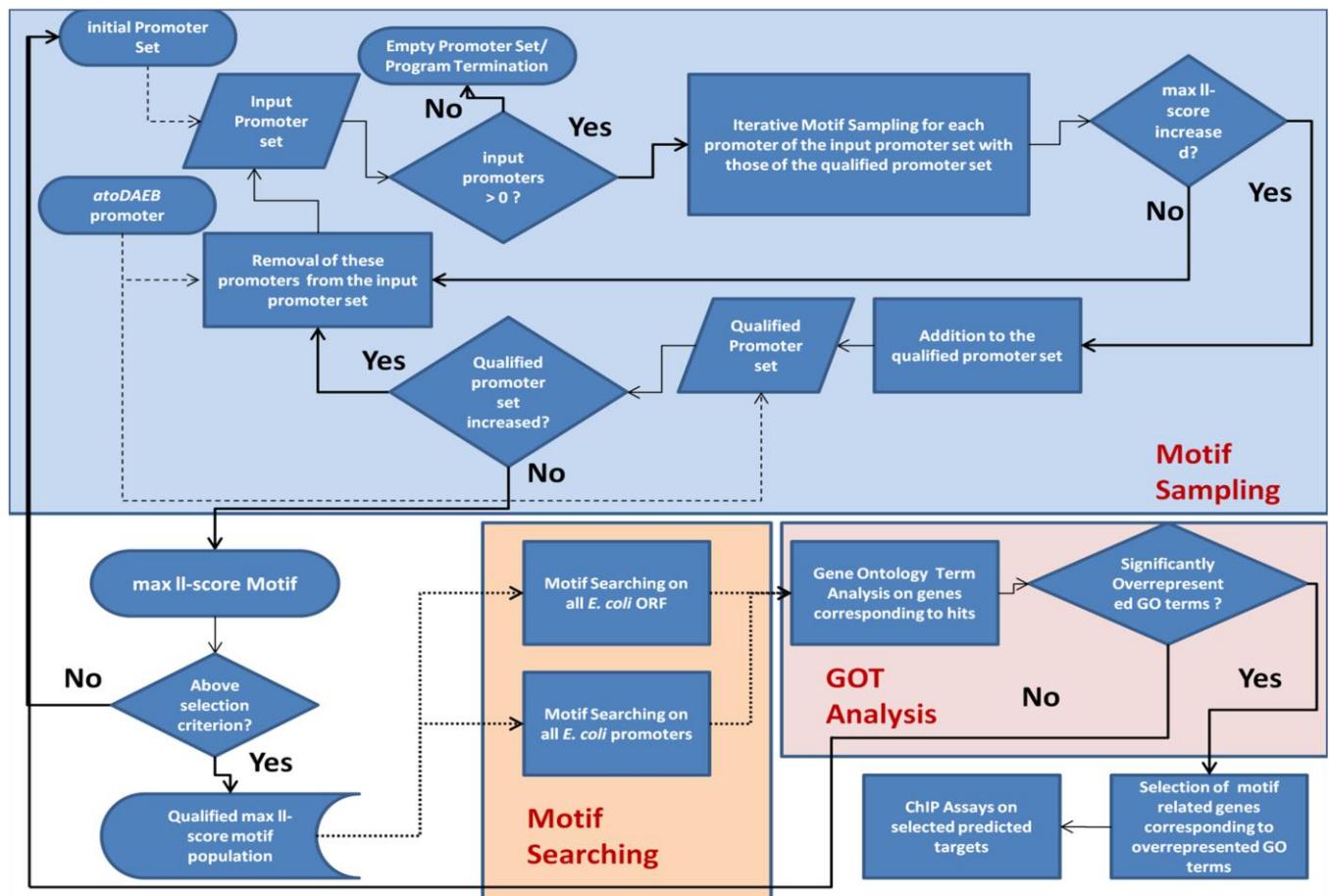
## 2.3 Αποτελέσματα

### 2.3.1 Εφαρμογή της υπολογιστικής διαδικασίας πρόβλεψης στόχων του μεταγραφικού παράγοντα AtoC

Στα πλαίσια της διατριβής αναπτύχθηκε μια νέα υπολογιστική διαδικασία πρόβλεψης θέσεων πρόσδεσης του μεταγραφικού παράγοντα AtoC σε ολόκληρο το γονιδίωμα του βακτηρίου *Escherichia coli*, μέσω της *de novo* κατασκευής μοτίβων, τα οποία αντιπροσωπεύουν συντηρημένες αλληλουχίες DNA. Η διαδικασία αυτή, χρησιμοποιώντας ως δεδομένο εισόδου τη μοναδική γνωστή θέση πρόσδεσης του AtoC, ανιχνεύει αντίστοιχες συντηρημένες περιοχές DNA και κατασκευάζει ένα μοτίβο, το οποίο αξιολογεί λειτουργικά, μέσω μια στατιστικής ανάλυσης βασισμένης σε γονιδιακές οντολογίες (GOT).

Λόγω του γεγονότος ότι ο αλγόριθμος δειγματοληψίας (MotifSampler) είναι ευριστικός, τα αποτελέσματα που αποδίδει είναι εξαιρετικά ευαίσθητα στην αρχική παραμετροποίηση. Ακόμα και με μικρή τροποποίηση μιας από τις αρχικές παραμέτρους, όπως για παράδειγμα το μήκος του μοτίβου, οι θέσεις μεγιστοποίησης του SPLS και η αρχική ομάδα επαγωγέων, η διαδικασία αποδίδει διαφορετικά μοτίβα. Τα διαφορετικά μοτίβα ενδέχεται να αντιπροσωπεύουν διαφορετικούς συνδυασμούς συντήρησης πρωτοταγών δομών DNA μεταξύ επαγωγέων. Σε κάθε περίπτωση, η συντήρηση των αλληλουχιών είναι σχετικά μικρή, αφού η πρόσδεση ενός μεταγραφικού παράγοντα στον επαγωγέα δεν εξαρτάται μόνο από την πρωτοταγή δομή του DNA αλλά και από ένα πλήθος άλλων παραγόντων, όπως η δομή της περιβάλλουσας χρωματίνης και ο σχηματισμός συμπλόκων μεταγραφής με ταυτόχρονη πρόσδεση άλλων μεταγραφικών παραγόντων. Ως εκ τούτου και σε συνδυασμό με το γεγονός της αυξημένης τυχαιότητας ασθενούς συντήρησης ανάμεσα σε διαφορετικές ομάδες συντηρημένων αλληλουχιών, μόνο κάποιες θα είχαν λειτουργική βάση. Το πρόβλημα αυτό αντιμετωπίστηκε μέσω πολλαπλών εκτελέσεων της υπολογιστικής διαδικασίας τροποποιώντας κάθε φορά τις αρχικές παραμέτρους και σε συνδυασμό με την ανάλυση GOT για την αυστηρή στατιστική αξιολόγηση των μοτίβων.

A



B

Symbol	Name	Description
	Process	An operation or action step
	Decision	A branch in the process
	Data Input/Output	Indicates data inputs and outputs to and from a process
	Initialization /Termination	A start or stop point in a process
	Stored data	Data storage

**Σχήμα 2.3.** Διάγραμμα ροής της υπολογιστικής διαδικασίας Α) Το διάγραμμα περιγράφει τον επαναληπτικό αλγόριθμο της διαδικασίας κατασκευής και ανάλυσης μοτίβων. Οι συνεχόμενες γραμμές αντιπροσωπεύουν υποχρεωτικά στάδια του αλγορίθμου, ενώ οι διακεκομμένες γραμμές αντιπροσωπεύουν εναλλακτικά σενάρια. Η διαδικασία περιγράφεται λεπτομερώς στον τομέα Μέθοδοι. Β) Περιγραφή των συμβόλων του διαγράμματος ροής.

Συνολικά, τρεις εκτελέσεις απέδωσαν μοτίβα που βρέθηκαν στατιστικά σημαντικά, ως προς πιθανή βιολογική λειτουργία (Πίνακας 2.1).

Εκτέλεση	Θέσεις αναζήτησης	Αρχική ομάδα επαγωγέων	Ομάδα επαγωγέων εξόδου
ως προς την έναρξη μεταγραφής			
1	-184 ως -165	Σύνολο επαγωγέων <i>E. coli</i>	atoD, galU, yhjR-bcsE, narZ, puuP, mdlA, dmsA, crr
2	-184 ως -165	Επαγωγείς εξαρτώμενοι από τον παράγοντα Sigma-54	atoD, rtcR, ibpB
3	-184 ως -165	Επαγωγείς με θέσεις πρόσδεσης για ρυθμιστές απόκρισης	atoD, dmsA, borD, csgD, pstS, glnH, dctA, yebE, nirB, torC, asr, tppB, acrD,

**Πίνακας 2.1.** Περιγραφή των τριών εκτελέσεων της υπολογιστικής διαδικασίας, οι οποίες απέδωσαν μοτίβα συσχετισμένα με στατιστικά σημαντικές βιολογικές λειτουργίες

Και στις τρεις επιτυχημένες εκτελέσεις χρησιμοποιήθηκαν οι θέσεις -184 ως -165 από την έναρξη της μεταγραφής (transcription start site, TSS), οι οποίες είναι οι ακριβείς γνωστές θέσεις πρόσδεσης του AtoC (Matta et al. 2007). Επιχειρήθηκε η χρησιμοποίηση πιο χαλαρού κριτηρίου θέσης μοτίβου με τη συμπερίληψη και γειτονικών θέσεων, ωστόσο η απόπειρα αυτή δεν έδωσε θετικό αποτέλεσμα. Από κάθε μια από τις τρεις παραπάνω εκτελέσεις κατασκευάστηκε ένα αντιπροσωπευτικό μοτίβο με το μέγιστο PSLLS (πίνακας 2.2), το οποίο προωθήθηκε στο στάδιο λειτουργικής ανάλυσης (βλέπε κεφάλαιο Μέθοδοι). Εκτός από τα τρία κύρια μοτίβα (μοτίβα 1,2,3) που έχουν το μέγιστο PSLLS, αξιολογήθηκαν λειτουργικά και μοτίβα με χαμηλότερο PSLLS. Από αυτά, μόνο ένα προκρίθηκε στο στάδιο λειτουργικής ανάλυσης, το οποίο ονομάστηκε Motif 4 και παρουσιάζεται εκτενώς παρακάτω.

Motíbo	PSLLS	Consensus	Logo
Motif 1	117.6	GCkATrCrrnnATTTrCwCA	
Motif 2	51.8	GnGnAAnTTyCTGCAwAGCC	
Motif 3	168.6	wnnnwrsAwAwmrnknmAnr	
Motif 4	66.3	nCknATnmGrnnnTnACGCM	

**Πίνακας 2.2.** Οι οπτικές αναπαραστάσεις των μοτίβων. Τα μοτίβα 1-3 είναι αυτά με το υψηλότερο σκορ που αποδόθηκαν από τις τρεις αντίστοιχες εκτελέσεις της υπολογιστικής διαδικασίας. Το μοτίβο 4 είναι ένα ενδιάμεσο μοτίβο που ανιχνεύει μια εξαιρετικά συντηρημένη μεσογονιδιακή περιοχή.

Η λειτουργική ανάλυση ξεκινάει με τη χρησιμοποίηση του μοτίβου για την αναζήτηση στόχων του σε ολόκληρο το γονιδίωμα (GWMS). Στον πίνακα 2.3 συνοψίζεται ο αριθμός των στόχων κάθε μοτίβου. Οι ενδογονιδιακές περιοχές είναι πιο συντηρημένες από τις μεσογονιδιακές, αφού αντιστοιχούν σε πρωτεϊνικούς τομείς. Συνεπώς ο αριθμός στόχων είναι μικρότερος στις επαγγελματικές περιοχές, εκτός από το μοτίβο 4, που ανιχνεύει μια εξαιρετικά συντηρημένη μεσογονιδιακή περιοχή.

Μοτίβο	Στόχοι σε επαγωγείς	Στόχοι σε ενδογονιδιακές περιοχές
1	55	234
2	153	694
3	331	688
4	346	266

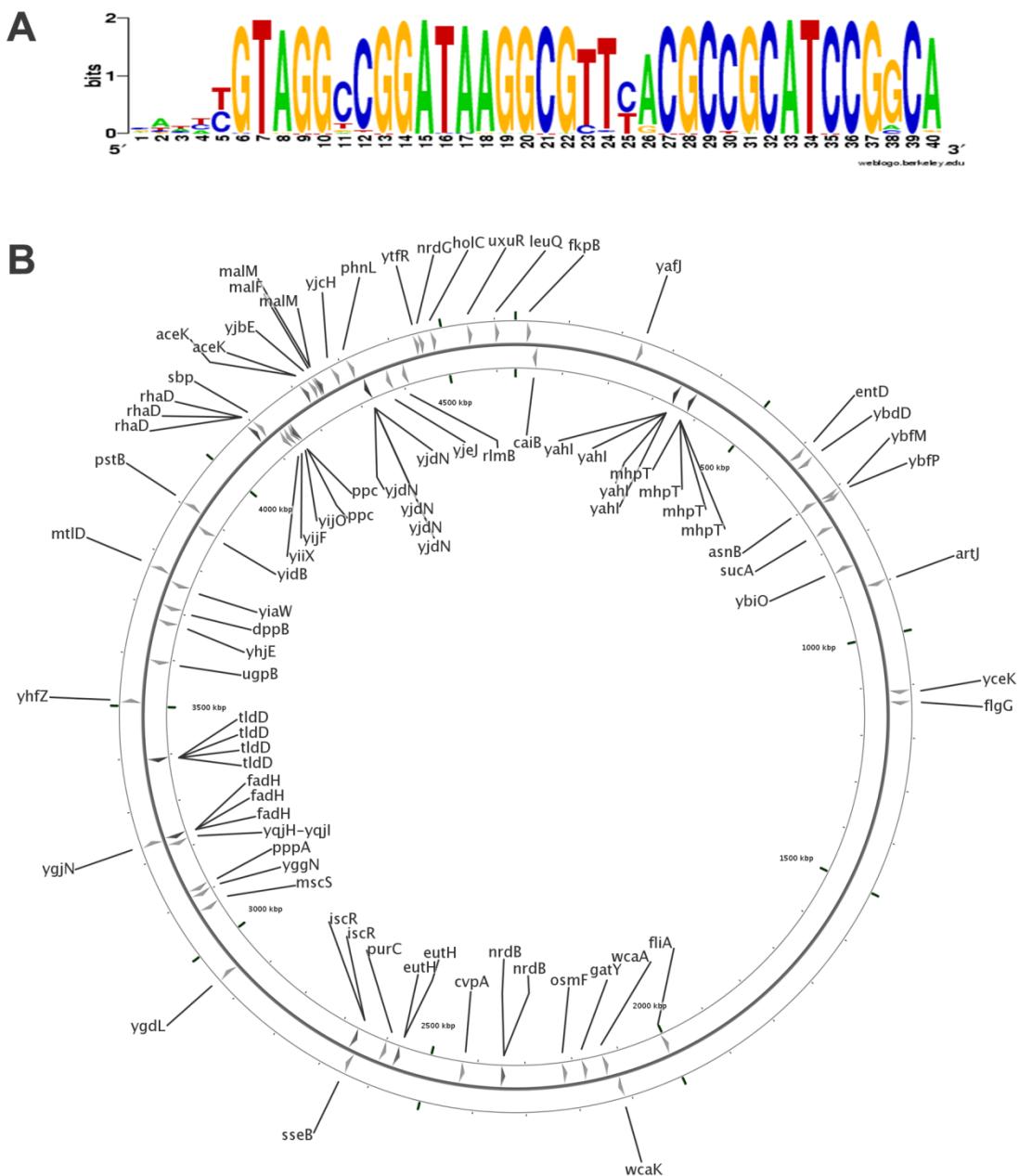
**Πίνακας 2.3** Αριθμός στόχων για κάθε μοτίβο. Τα μοτίβα 1-3 έχουν περισσότερους στόχους σε ενδογονιδιακές περιοχές, ενώ το μοτίβο 4 αντιστοιχεί σε μια εξαιρετικά συντηρημένη μεσογονιδιακή περιοχή.

Η ανάλυση GOT χρησιμοποιεί ως πλεονέκτημα το γεγονός της ασθενούς συντήρησης των αλληλουχιών DNA, διότι στο υπεργεωμετρικό τεστ γονιδιακών οντολογιών, μόνο με μεγάλο αριθμό αλληλουχιών είναι δυνατή η εξαγωγή υπεραντιπροσώπευσης λειτουργιών. Συνεπώς, το στάδιο GWMS είναι απαραίτητο αντί για τη χρησιμοποίηση μόνο της ομάδας επαγωγέων εξόδου που αποτελεί ένα μικρό μέρος των στόχων των μοτίβων.

### **2.3.3 Αξιολόγηση ενδιάμεσων μοτίβων**

Η υπολογιστική διαδικασία σχεδιάστηκε με σκοπό την απόδοση ενός μοτίβου με μεγιστοποιημένο PSLLS. Η κατασκευή μοτίβων με υψηλό score θεωρήθηκε πιο κατάλληλη για την εφαρμογή της λειτουργικής αξιολόγησης των στόχων, αφού το log-likelihood score λαμβάνει υπόψη όχι μόνο τη συντήρηση αλληλουχίας αλλά και τον αριθμό των γονιδίων όπου παρατηρείται η συντήρηση αυτή (occurrency). Παρόλα αυτά, το κριτήριο του μέγιστου PSLLS παραμένει αρκετά αυστηρό και συνεπάγεται την απόρριψη μεγάλου αριθμού «ενδιάμεσων» μοτίβων της διαδικασίας, δηλαδή μοτίβων που δημιουργήθηκαν σε μια δειγματοληψία και χαρακτηρίστηκαν από χαμηλότερο PSLLS σε σχέση με το μέγιστο που επετεύχθη στο τέλος της διαδικασίας. Για την παράκαμψη του αυστηρού κριτηρίου μέγιστου PSLLS και τον έλεγχο των ενδιάμεσων μοτίβων, εφαρμόστηκε μια παράλληλη διαδικασία υποβολής των μοτίβων αυτών στη λειτουργική ανάλυση. Γενικά, όπως ήταν αναμενόμενο, τα μοτίβα αυτά δεν εμφάνισαν συσχέτιση με βιολογικές λειτουργίες, καθώς στην πλειοψηφία τους αποτέλεσαν υπο-περιπτώσεις των μοτίβων με το μέγιστο PSLLS και επικαλύπτονταν από αυτά. Εξαίρεση αποτέλεσε το Μοτίβο 4, ένα μοτίβο που δημιουργήθηκε κατά τη δεύτερη εκτέλεση της διαδικασίας και συσχετίστηκε πολύ ισχυρά με τη λειτουργία της μοριακής μεταφοράς (transporter activity). Το μοτίβο αυτό αναδεικνύει μια εξαιρετικά συντηρημένη περιοχή DNA που εντοπίζεται σε συνολικά 346 επαγωγείς (Πίνακας 2.3).

Το Μοτίβο 4 εκτός από την αλληλουχία των 20 νουκλεοτιδικών βάσεων (base pairs, bp) που αντιστοιχούν στο μήκος του, υποδεικνύει μια ευρύτερη περιοχή DNA των 40 bp που παρουσιάζει πρωτοφανή συντήρηση. Ένα επιπρόσθετο μοτίβο με διπλάσιο μήκος (Μοτίβο 5) κατασκευάστηκε επίσης για την αντιπροσώπευση της περιοχής αυτής (Σχήμα 2.4.A).



**Σχήμα 2.4** Α) *Logo* αναπαράστασης του μοτίβου 5, που αντιπροσωπεύει ένα συντηρημένο τμήμα DNA των 40 bp. Η συντήρηση 40 νουκλεοτιδικών βάσεων σε 85 επαγωγείς είναι αξιοσημείωτη, καθώς είναι στατιστικά απίθανο να ωφείλεται στην τυχαιότητα. Β) Η κατανομή των 85 εμφανίσεων του μοτίβου 5 υποδεικνύει κάποιον σημαντικό ρόλο, καθώς βρίσκεται μόνο σε μεσογονιδιακές, επαγωγικές περιοχές και σε πολλές περιπτώσεις φαίνεται να έχει υποστεί downstream των επαγωγικών περιοχών.

	1	CACTTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA
flgG					
phnL	CATTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
nrdG	CAAATGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
ma1M	CAAATGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yjdN	CATCCGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
mhpT	CATCCGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
fliA	TTTTTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yafJ	CTGGCGTAGG	CCGGATAAAGA	CGTTTACGCC	GCATCCGGCA	
yjcH	CGAACGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yahI	CATTCTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yahI	CATTCTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
ybdD	CATTCTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
sseB	AATTCTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yahI	CACTCTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yjdN	CCTCCGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yjdN	CCTCCGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
yjdN	CCTCCGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
ybfM	CTTTTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	
fkpB	AAATCTGTAGG	CCGGATAAAGG	CGTTTACGCC	GCATCCGGCA	

**Σχήμα 2.5.** Στοίχιση των 20 πρώτων στόχων του μοτίβου 5, που αναδεικνύει την υψηλή συντήρηση αλληλουχίας.

Η εφαρμογή GWMS για το Μοτίβο 5 έδειξε ότι έχει στόχο 85 μεσογονιδιακές περιοχές, ενώ δεν έχει καμία ενδογονιδιακή (Σχήμα 2.4). Η αυστηρή αυτή κατανομή, σε συνδυασμό με την εξαιρετική συντήρηση αλληλουχίας μεσογονιδιακού DNA, υποδεικνύει τον ενδεχόμενο σημαντικό της ρόλο στο βακτηριακό γονιδίωμα (Εικόνα 2.3). Ενδεικτικά, 6 από τις εμφανίσεις (occurrences) του μοτίβου περιέχουν 38/40 ταύτιση αλληλουχίας. Η πιθανότητα να παρατηρηθεί τυχαία δυο φορές μια αλληλουχία των 38 bp είναι  $1/0.25^{38}$ , ενώ το γονιδίωμα του E.coli έχει μήκος  $4.6 \times 10^6$  bp. Σημειώνεται ότι η αλληλουχία αυτή παρατηρήθηκε σε μεσογονιδιακές περιοχές άλλων *Enterobacteria* και συγκεκριμένα στα *Shigella flexneri*, *Salmonella enterica*, *Klebsiella pneumoniae* and *Citrobacter coseri* (τα δεδομένα δεν παρέχονται).

### 2.3.4 Συσχέτιση μοτίβων με γονιδιακές λειτουργίες και μεταβολικά μονοπάτια

Η λειτουργική αξιολόγηση των μοτίβων επετεύχθη μέσω μιας αλγορίθμικής διαδικασίας (βλέπε κεφάλαιο Μέθοδοι, Λειτουργική Ανάλυση Μοτίβων), στην οποία το κάθε μοτίβο χρησιμοποιείται για GWMS. Κατά αυτόν τον τρόπο, τα γονίδια στον

επαγωγέα των οποίων ανιχνεύεται αλληλουχία-στόχος εισάγονται στην ανάλυση γονιδιακών οντολογίων (GOTA). Από τη στιγμή που θα βρεθεί υπερ-αντιπροσώπευση ενός συγκεκριμένου GOT, υπάρχει στατιστικά σημαντική ένδειξη ότι η ομάδα των γονιδίων συσχετίζεται υπό έναν κοινό λειτουργικό παράγοντα και κατά συνέπεια δεν είναι τυχαία δομημένη.

Τα τέσσερα μοτίβα που εισήχθησαν στην GOTA παρουσίασαν υπερ-αντιπροσώπευση σε συγκεκριμένους GOT και συσχετίστηκαν με βιολογικές λειτουργίες (Πίνακας 2.4).

GO id	GO Term	p-value (corrected)	FDR	Αναλογία όρου στους στόχους	Ποσοστό όρου στους στόχους (%)	Αναλογία όρου στο γονιδίωμα	Ποσοστό όρου στο γονιδίωμα (%)
<b>Μοτίβο 1</b>							
GO:0016638	oxidoreductase activity, acting on the CH-NH2 group of donors	0.066	12.00%	3/67	4.47	10/4339	0.23
<b>Μοτίβο 2</b>							
GO:0015749	monosaccharide transport	0.045	10.00%	4/154	2.59	891101	0.23
GO:0022891	substrate-specific transmembrane transporter activity	0.006	0.00%	22/154	14.28	254/4339	5.85
GO:0043169	cation binding	0.012	0.00%	37/154	24.02	570/4339	13.13
GO:0008324	cation transmembrane transporter activity	0.076	2.33%	13/154	8.44	128/4339	2.94
<b>Μοτίβο 3</b>							
GO:0045449	regulation of transcription	0.001	0.00%	53/324	16.35	372/4339	8.57

**Κεφάλαιο 2** Αποτελέσματα

GO:0007155	cell adhesion	0.027	0.00%	10/324	3.08	34/4339	0.78
GO:0001101	response to acid	0.042	0.12%	4/324	1.23	5/4339	0.11
GO:0009289	fimbrium	0.032	0.00%	8/324	2.46	28/4339	0.64
GO:0003700	transcription factor activity	0.01	0.00%	32/324	9.87	208/4339	4.79

**Μοτίβο 4**

GO:0051234	establishment of localization	7.82E-04	0.00%	40/110	36	800/4339	18
GO:0005215	transporter activity	3.91E-03	0.00%	30/110	27	561/4339	13

**Πίνακας 2.4.** Αποτελέσματα λειτουργικής ανάλυσης μοτίβων. Κάθε μοτίβο συσχετίζεται με τις βιολογικές λειτουργίες που βρίσκονται υπερ-αντιπροσωπευμένες μεταξύ των στόχων του. Παρατηρείται ότι αναδεικνύονται γνωστές λειτουργίες του AtoC, όπως η απόκριση σε όξινο περιβάλλον, ο μεταγραφικός έλεγχος, η μοριακή μεταφορά και η κινητικότητα.

Στο σημείο αυτό πρέπει να επισημανθεί ότι οι στόχοι των μοτίβων υποδεικνύουν πιθανές θέσεις πρόσδεσης του AtoC λόγω συντήρησης της πρωτοταγούς δομής DNA. Οι θέσεις αυτές χαρακτηρίζονται από ένα δυναμικό σύνδεσης με τον AtoC, αλλά όχι απαραίτητα παρατηρήσιμο σε όλες τις χρονικές στιγμές και υπό όλες τις κυτταρικές ή φυσιολογικές συνθήκες. Η στατιστική όμως συσχέτισή τους με γονιδιακές λειτουργίες, ενισχύει την πιθανότητα παρατήρησης μιας λειτουργικής σύνδεσης και τη διαχωρίζει από την απλή τυχαία ομοιότητα αλληλουχίας DNA, η οποία είναι ιδιαίτερα συχνή στους βακτηριακούς επαγωγείς.

### **2.3.5 Πειραματική επιβεβαίωση επιλεγμένων στόχων των μοτίβων για *in vivo* σύνδεση με τον ρυθμιστή απόκρισης AtoC**

Προκειμένου να επιβεβαιωθεί πειραματικά η *in vivo* σύνδεση του ρυθμιστή απόκρισης AtoC σε πιθανούς στόχους μοτίβων που προέκυψαν από την υπολογιστική διαδικασία, χρησιμοποιήθηκε η μέθοδος της ανοσοκατακρήμνισης χρωματίνης (Chromatin Immunoprecipitation, ChIP). Δοκιμάστηκαν συνολικά 29 πιθανοί στόχοι, εκ των οποίων βρέθηκαν θετικοί οι 24, συμπεριλαμβανομένων των δύο γνωστών στόχων του AtoC στον επαγωγέα *atoDAEB*. Τα αποτελέσματα της πειραματικής επιβεβαίωσης συνομιζονται στον Πίνακα 2.5. Σημειώνεται ότι έξι από τους στόχους

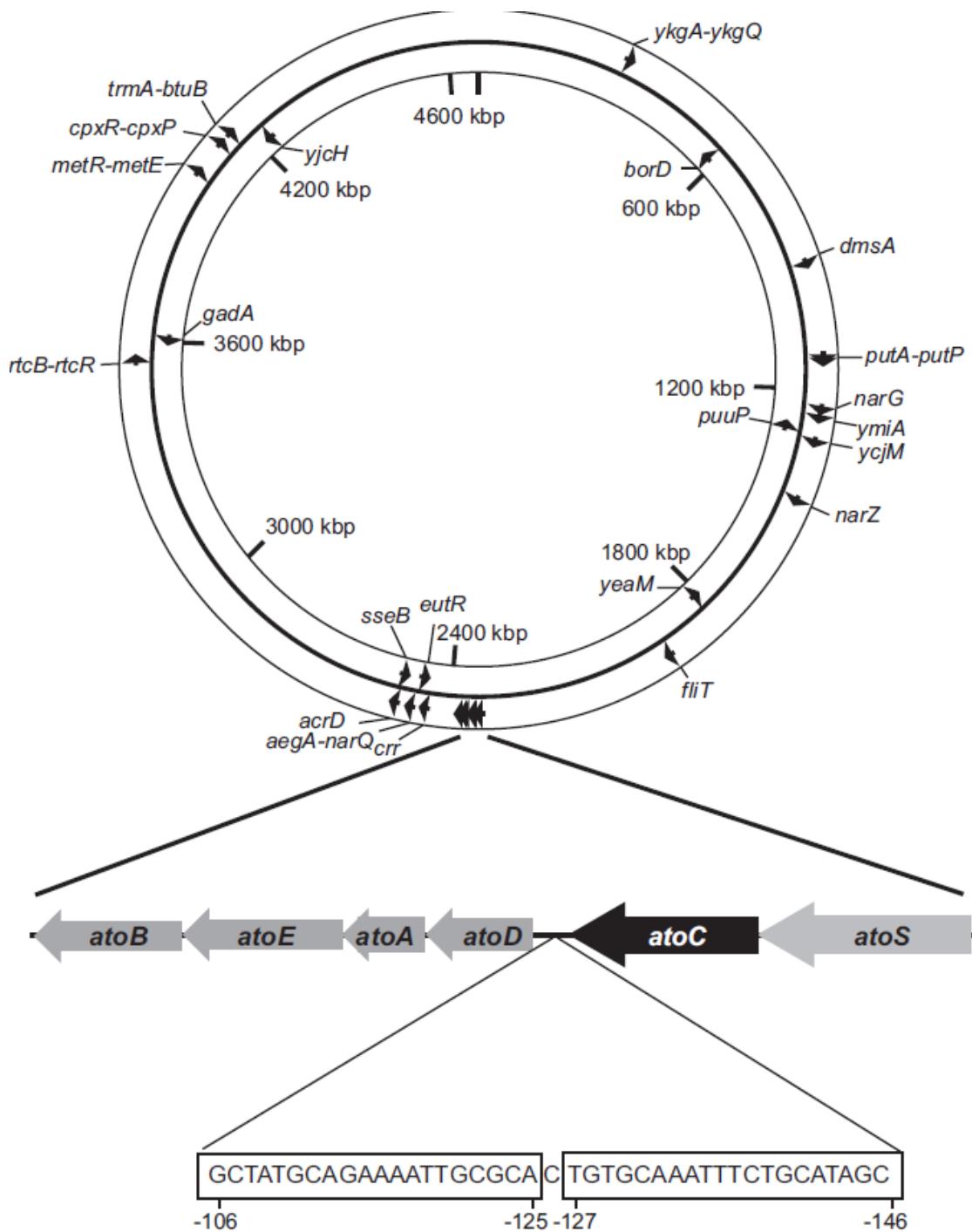
βρίσκονται σε ενδογονιδιακές περιοχές και οι υπόλοιποι σε μεσογονιδιακές (επαγωγείς). Οι θέσεις των 24 σημείων πρόσδεσης συνοψίζονται στο Σχήμα 2.6.

Αποτέλεσμα	Αλληλουχία στόχου	Γονίδιο	Μοτίβο	Απόσταση από έναρξη μεταγραφής (Refseq)				Προϊόν γονιδίου
				Χρωματίδα (strand)	(+)	-184	-165	
+	GCTATGCAGAAATTGACACA	atoD	1,2,3,4	(+)				acetyl-CoA: acetoacetyl-CoA transferase, subunit
+	GCTATGCAGAAAATTGCGCA	atoD	1,2,4	(-)		-163	-144	acetyl-CoA: acetoacetyl-CoA transferase, subunit
+	GCTATAGAAATAATTACACA	dmsA	1,2	(-)		-69	-50	dimethyl sulfoxide reductase, anaerobic, subunit A
+	GCCATGCGGGGATTAAATCA	puuP	1	(+)		-148	-129	putrescine importer
+	GCCGTCCAGATGTTACACA*	metR- metE	1	(-)		-87,	-68,	transcriptional activator
						-170	-151	
+	GCGATTGTAGGGATTGCTCA*	trmA- btuB	1,(2)	(+)		-49,	-30,	tRNA (uracil-5)- methyltransferase/ outer membrane transporter
						-340	-321	
+	ACTATACGGAAAATTCCACT*	ykgA- ykgQ	1,2	(-)		-235,	-216,	Putative transcriptional activator
						-55	-36	
+	GTGATGGATTATTGATCT*	aegA- narQ	1	(-)		-147,	-128,	predicted fused oxidoreductase: FeS binding subunit and
						-79	-60	sensory HK in TCS with NarP
+	GCCTCGGAGGTATTAAACA*	cpxR- cpxP	1	(-)		-21,	-2,	response regulator in TCS with CpxA/ periplasmic protein;

						-149	-130	combats stress
+	GCGAGGCGGGTAATTAGACA	ymiA	1	(-)	-199	-180	small predicted membrane protein	
+	GGCCCAATTACTGCTTAGG	crr	1,2	(+)	-28	-9	glucose-specific PTS component	
+	GCGGTGCAGGAGATTGCACA	<u>fliT</u>	1	(+)	<u>127</u>	<u>146</u>	predicted chaperone (flagellum)	
+	CCGGTCCAGGAATTACTCA	<u>narG</u>	1	(-)	<u>2</u>	<u>21</u>	nitrate reductase 1, alpha subunit	
+	GCGAACGATTAATTACACA	<u>ycjM</u>	1	(+)	<u>94</u>	<u>113</u>	glucosyltransferase predicted	
+	ACTGTACAGAAAGTTGCTCA	<u>yeaM</u>	1	(-)	<u>668</u>	<u>687</u>	transcription regulator (predicted)	
+	GCGGGTCAGGTAATTGCACA	<u>gadA</u>	1	(-)	<u>959</u>	<u>978</u>	glutamate decarboxylase A, PLP-dependent	
+	GACATGCAGTTGATTACACA	<u>eutR</u>	1	(+)	<u>783</u>	<u>802</u>	predicted transcription regulator	
+	GCGATCCAAAAGTTACTCA	<u>narZ</u>	1	(+)	<u>0</u>	+ 19	nitrate reductase 2 alpha subunit	
+	CGGAGATTCCCGCAAAGCC*	rtcB- rtcR	2	(+)	-145, -64	-126, -45	sigma-54-dependent transcriptional regulator of rtcBA	
+	CGGCAAGTTCGACATTGCC*	putA- putP	2	(+)	-398, -45	-379, -26	fused transcriptional regulator/proline dehydrogenase/ carboxproline:sodium symporter	
-	GGGCATTTCCTGCAAAACC*	rhaT- sodA	3	(+)	-167, -138	-148, -119	L-rhamnose:proton symporter/superoxide dismutase	
-	GGATGATGTTCTGCATAGCA	adhP	3	(+)	-35	-16	ethanol-active	

								dehydrogenase/acetalde hyde-active reductase
-	TTTATAGAAATAGATGCACG	nirB	4	(-)	-205	-186	nitrite reductase, large subunit, NAD(P)H- binding	
+	TATGTGTAGAAAATTAAACA	borD	4,3	(+)	-59	-40	DLP12 prophage; predicted lipoprotein	
+	GTTCAGTATAAAAGGGCATG	acrD	2	(+)	-128	-109	aminoglycoside/multidr ug efflux system	
-	GAAATAGAAATAGTTGAAAG	ykgE	4	(-)	-472	-453	predicted oxidoreductase	
+	CCGGATAAGGCCTTACGCC	yjcH	4	(+),(-)	-114, -122	-95, -103	conserved inner membrane protein/ acetate transport	
-	CCGGATAAGACGTTACGCC	yafJ	4	(+)	-193	-174	Unknown function, similarity to type II amidotransfase	
+	CCGGATAAGGCCTTACGCC	sseB	4	(+)	-125	-106	rhodanese-like enzyme	

**Πίνακας 2.5.** Στόχοι μοτίβων πειραματικά ελεγμένοι με τη μέθοδο ChIP για *in vivo* σύνδεση με τον παράγοντα AtoC, παρουσία ακετοξικού οξέος. Όπως προέβλεψε η ανάλυση GOTs πολλά από τα γονίδια-στόχους αφορούν μεταγραφικούς παράγοντες, μεταφορείς και ένζυμα οξειδοαναγωγής. Η υπογράμμιση υποδεικνύει στόχους σε ενδογονιδιακές περιοχές (gene-encoding). \*Οι στόχοι αυτοί αντιστοιχούν σε μία μοναδική θέση ανά ζεύγος που συσχετίζεται και με τα δυο περιβάλλοντα γονίδια.



**Σχήμα 2.6.** Οι θέσεις των 24 πειραματικά επιβεβαιωμένων στόχων του AtoC στο γονιδίωμα του *E. coli*. Συμπεριλαμβάνεται η διπλή θέση πρόσδεσης του AtoC στον επαγωγέα του οπερονίου *atoDAEB*.

### 2.3.6 Εφαρμογή της υπολογιστικής διαδικασίας στον μεταγραφικό παράγοντα LexA

Προκειμένου να αποτιμηθεί η γενική λειτουργικότητα της υπολογιστικής διαδικασίας, εφαρμόστηκε σε έναν μεταγραφικό παράγοντα με πολλούς γνωστούς γοδιακούς στόχους, τον παράγοντα LexA. Ο LexA είναι ένας καλά χαρακτηρισμένος μεταγραφικός παράγοντας που ελέγχει την έκφραση πολλών γονιδίων που εμπλέκονται στην κυτταρική απόκριση σε βλάβες DNA (SOS response) (Kelley 2006). Στη βιβλιογραφία υπάρχει μια πληθώρα πειραματικά επιβεβαιωμένων σημείων πρόσδεσής του. Η υπολογιστική διαδικασία εφαρμόστηκε με τη χρήση μόνο ενός από αυτά, με σκοπό τον έλεγχο των αποτελεσμάτων ως προς την πρόβλεψη ήδη γνωστών στόχων.

GO id	Term	p-value	FDR	Αναλογία α στους στόχους	Αναλογία στο γονιδίωμα	Γονίδια που αντιστοιχούν στο όρο GO
		(corrected)				
GO:0009432	SOS response	0,001	0,00%	7/82	22/4339	<a href="#">yebG,lexA,uvrD,ydjM,ssb,uvrA,ruvA</a>
	response to DNA damage					<a href="#">yebG,ydjM,recN,lexA,uvrD,hokE,uvrA,ssb,ruvA</a>
GO:0006974	stimulus	0,001	0,00%	10/82	82/4339	mutH
	response to					<a href="#">yebG,psiE,ydjM,recN,lexA,uvrD,hokE,rpoE,uvr</a>
GO:0006950	stress	0,053	0,80%	13/82	230/4339	<a href="#">A,ssb,ecpD,ruvA,mutH</a>

**Πίνακας 2.6.** Ανάλυση όρων γονιδιακής οντολογίας για τον παράγοντα LexA. Η υπολογιστική διαδικασία εφαρμόστηκε με σημείο έναρξης μια γνωστή θέση πρόσδεσης του μεταγραφικού παράγοντα. Με μοναδικό δεδομένο εισόδου τη θέση αυτή, η ανάλυση συσχετίζει τον LexA με τα βιοχημικά μονοπάτια απόκρισης στις βλάβες DNA (SOS response), τα οποία αποτελούν και την κύρια γνωστή λειτουργία του. Τα υπογραμμισμένα γονίδια είναι πειραματικά επιβεβαιωμένοι (από τη βιβλιογραφία) στόχοι του LexA.

Η εφαρμογή της υπολογιστικής διαδικασίας οδήγησε στη δημιουργία ενός μοτίβου που αναγνωρίζει 13 γνωστά σημεία πρόσδεσης, επιβεβαιώνοντας την αποτελεσματικότητα της μεθόδου σε παράγοντες εκτός του AtoC. Επιπρόσθετα, η ανάλυση γονιδιακών οντολογιών ανέδειξε τις λειτουργίες GO:0006974, “response to DNA damage stimulus” και GO:0009432, “SOS response”, γεγονός που είναι σε απόλυτη συμφωνία με τον βιολογικό ρόλο του LexA (πίνακας 2.6). Τα αποτελέσματα παρουσιάζονται στον Πίνακα 2.7.

**A) Στόχοι μοτίβου LexA**

Αναφορά στόχου και κατηγορία	Στόχος	Score	Γονίδιο	Απόσταση από έναρξη μεταγραφής	
*Ecocyc uvrD-Site 1 (C)	AAATCTGTATATATAACCCAG	80139800	uvrD	-78	-59
*Ecocyc lexA-Site 3 (C, BPP)	ATAACTGTATATACACCCAG	10701300	lexA	-28	-9
*Ecocyc uvrD-Site 1 (C)	AAAGCTGGTATATACAG	5847590	uvrD	-74	-55
	ATATAGGTACATAGACACAT	4226710	yfjJ	-62	-43
Ecocyc lexA-Site 2 (C, BPP)	TATGCTGTGAGTATATACAG	830614	lexA	-45	-26
	ATAGACCCACATACATTCA	516081	alaW-yfeC	-212,-44	-193,-25
	CTATCGGTGCGTGTATGCAG	495036	hupA	-91	-72
	AAATTTCGAGTATTCCCAG	238230	bisC-yiaD	-72,-101	-53,-82
**Ecocyc lexA-Site 2 (C, BPP) (site that initiated the procedure)	TTTGCTGTATATACTCACAG	190427	lexA	-49	-30
	CTTGATGGGTGTATTCTCAG	29289,6	yjhU	-220	-201
	ATAGATGGGTTTATTACAT	14309,7	cadC	-85	-66
	AATTATGCAATTATACGG	8746,46	yqeH	-37	-18
	ATAGCGCTAACGTATATAT	8122,87	yciZ	-140	-121
*Ecocyc ruvA-Site 2 (C, BPP)	AATGCTGGATAGATATCCAG	6445,8	ruvA-yebB	-68,-227	-49,-208
	AAATGGGCATTTCTACACAG	6337,23	rpoE-nadB	-215,-213	-196,-194
*Ecocyc dinD-Site 1 (C, GEA)	GTAACTGTATTTATACAG	6135,15	dinD	-61	-42
	ATATCTGTATGGATTCA	4960,38	ydfU	-41	-22
	ATTGCCCCCTATATTCCAG	4604,08	insC	2437	2456
	ATTGCCCCCTATATTCCAG	4604,08	yaiX-insC	-305,280	-286,299
	TAATATGTAATAAACCAT	4174,84	cadC	-89	-70
	TTTTCTGAACATACATGCAG	3953,4	kch	-68	-49
	TTTTTGCACTTAGATACAG	3790,25	malP-malT	-217,-415	-198,-396
	ATTACTGAGAATAACACCAT	3129,75	yjhU	-216	-197
	AATATTGTATATATTACAT	3124,23	torY	-232	-213
	AAATCTTTACTTATTCACAG	2997,96	infA	-77	-58
	TTACCTGTAATTAGATACAG	2696,68	pgaA-ycdT	-55,-552	-36,-533
	TGATAGGTGTATGCACCCAG	2586,09	insF	535	554
	TATGGTGCCTGCACACAT	2173,62	potG	-39	-20
*Ecocyc dinJ-Site 1 (C)	CAAGCTGAATAATACAG	2071,6	dinJ-yafL	-37,-193	-18,-174

	TATGTGCCGAATATCGCAG	2063,68	ygdT- mutH	-152,-54	-133,-35
	CTATTCGACGTATTCCAA	1978,57	ykgP- eaeH	-221,-351	-202,-332
*Ecocyc uvrA-Site 4 (C, BCE)	AATACTGTATATTCAATTAG	1901,86	uvrA- ssb	-106,-168	-87,-149
	TAATCGGCAAATACTCTAG	1693,61	zipA- cysZ	-71,-179	-52,-160
	ATAGCCCTATTACACCGAT	1460,81	hns- tdk	-183,-442	-164,-423
	ATAAAATGCCGTATAAACAG	1426,64	yqeI	-306	-287
	CCATTGTTACGTATAACCGAG	1406,05	ftsB	-124	-105
	CTAACTCGGAATTCCCAG	1287,4	modA	-64	-45
	TAAGTTGCCTTGTTCACAG	1153,13	rplU- ispB	-225,-54	-206,-35
	CTACCGGGCCGTAGACCCGG	1144,91	yddL	-252	-233
	CTAAGTGTAAATACATCAAT	1074,83	yjjK- slt	-200,-31	-181,-12
	TTATCTCTACTTACTCAGAG	1065,9	ybhP- ybhQ	-114,-39	-95,-20
*Ecocyc hokE-Site 1 (C)	ATAGCTGTTATTACAG	1021,32	ybdK- hokE	-279,-194	-260,-175
	TTAGCGGGAAATACTCCAA	977.885	brnQ	-107	-88
	ATATCTGAAGGTATTCAG	949.328	yijP	-75	-56
	CTATCCGTATAATTGCAT	914.858	yqeH	-33	-14
	AAAGCTGTGCTTGTATAAT	829.308	ivbL	-52	-33
	ATTATCGCTATGCATACAG	794.397	accB	-98	-79
	AAAACGCCACGTGTTACGT	760.549	rfe	-224	-205
	CTTCTCGATTGCTCTCAG	727.237	proV	-183	-164
	CTTGTGGCAAATGGACACGT	708.293	yhfG	-36	-17
	CTACGGTAGGTATATCCGG	694.405	yjbT- psiE	-155,-135	-136,-116
	ATTGATGTATTACACTTAG	666.775	yjjK- slt	-200,-31	-181,-12
*Ecocyc yebG-Site 1 (C, GEA)	ATAACTGTGATTACAG	574.974	yebG- purT	-36,-118	-17,-99
	ATTATGCATATTCTCTCAG	530.093	mdoD	-142	-123
	CTCTGTCTGTGTTACAG	519.456	mdtM- yjiP	-153,-346	-134,-327
	AAATCTCCATGTATATAACG	501.618	tyrS	-23	-4
	AAAAGTGAACGTACTCTCAT	498.834	yjbJ	-47	-28
	AAATGTGTGCTTAGTCCCTT	486.838	ravA- kup	-199,-44	-180,-25
	ATTGGCTGAAGTATACCCAT	464.462	narX- narK	-104,-255	-85,-236
*Ecocyc dinD-Site 1 (C, GEA)	ACAACTGTATATAACAG	460.333	dinD	-65	-46
	CTTGTGTGCTTATACGAT	450.441	mdoC- mdoG	-357,-57	-338,-38

*Ecocyc ydjM-Site 2 (C, GEA, BPP)	TATACTGTACGTATCGACAG	450.09	ydjM	-24	-5
	CTATCCGCCAATAAACCCGT	415.019	fliE-fliF	-100,-135	-81,-116
	TTAAATCCACACGTATCCAG	409.216	glyQ	-32	-13
	ATTAATGTGAATATATACAA	377.338	torY	-228	-209
	AAAACGCCACGTAAACACGT	370.587	rfe	-216	-197
	CACACTGTCTATACTTACAT	355.698	otsB	-76	-57
	CAATCGAAACATATATTCAT	339.014	rhsC	-145	-126
*Ecocyc recN-Sitew 1 (C, BPP)	TAAACTGGTTTATATACAG	336.08	recN	-66	-47
	TTAAATGTAAATAACAGCCAT	320.034	yfdX-ypdI	-253,-63	-234,-44
	TATGATGTGAGCATATTCA	318.001	kch	-180	-161
	ATATCGAGAAGTATAACGTAG	315.054	yjiK	-49	-30
	TATTCTACGAATATTCCGG	315.039	hypF	-117	-98
	TAATCGGAAAGTACATTGGG	266.658	agaB	-144	-125
	AATTGTGCAACTGCACGCAG	250.707	yaiP	-94	-75
	ATTATTCCCTGTATATTCA	221.946	eepD	-74	-55
	CATTCGCTATATATTGTCA	218.976	yghW	-30	-11
	CTATATAGGCTTGTATACAT	206.414	dmsA	-95	-76
	TAAGTGGTCAACATTCCAG	103.213	pssA	-105	-86
	AAATATGCCTGTAGATCAGG	890.283	ynfL-ynfM	-93,-34	-74,-15
	CATTCGTGTATATTCTG	87.867	yhdU	-53	-34
	AACTTGGTACGTAAACGCAT	641.978	alaX	-24	-5
	TAAGTTATGCGTATTCTCGT	599.613	mreC	-20	-1
	TAATTAGAATATAGATACAG	560.128	rplT	-31	-12
	ATATTCCAATTTATCAAG	398.441	citT	-40	-21

### B) Κατηγορία πειραματικής απόδειξης

Κατηγορία απόδειξης (Evidence)	Κωδικός
Computational/inference based on similarity to consensus sequence	C
Gene expression analysis	GEA
Binding of cellular extracts	BCE
Binding of purified proteins	BPP

**Πίνακας 2.7. A)** Αποτελέσματα εφαρμογής της υπολογιστικής διαδικασίας στον μεταγραφικό παράγοντα LexA. \*Γνωστές θέσεις πρόσδεσης του LexA που ανιχνεύθηκαν από την υπολογιστική διαδικασία. \*\*Θέση πρόσδεσης που χρησιμοποιήθηκε ως έναρξη της διαδικασίας. **B)** Αντιστοιχία κωδικών αξιοπιστίας των γνωστών από τη βιβλιογραφία στόχων.

## 2.4 Συμπεράσματα

### 2.4.1 Αναγνώριση νέων στόχων του AtoC σε ολόκληρο το γονιδίωμα του *E. coli*.

Το σύστημα μεταφοράς σήματος δύο συστατικών AtoSC ελέγχει την έκφραση του οπερονίου *atoDAEB* που κωδικοποιεί πρωτεΐνες που υπεισέρχονται στον καταβολισμό λιπαρών οξέων μικρής αλυσίδας. Το ακετοξικό οξύ είναι επαγωγέας του συστήματος AtoSC και άρα επάγει έμμεσα την έκφραση των γονιδίων του οπερονίου *atoDAEB*. Εκτός από τον ρόλο του ως μεταγραφικό επαγωγέα, ο AtoC έχει μια δεύτερη γνωστή λειτουργία, αυτή του μετα-μεταφραστικού αναστολέα του ενζύμου ορνιθίνη αποκαρβοξυλάση που είναι το ένζυμο-κλειδί της σύνθεσης πολυαμινών (Canellakis et al. 1993). Ο διπλός αυτός ρόλος υποδεικνύει την πολυπλοκότητα των δικτύων που ρυθμίζουν τα μεταβολικά ομοιοστατικά μονοπάτια στα βακτήρια. Επιπρόσθετα, υπάρχει ένας αριθμός ενδείξεων για τον ρόλο του AtoSC σε άλλες λειτουργίες εκτός από τις προαναφερθείσες. Η διαγραφή της γονιδιακής περιοχής *atoSC* έχει ως αποτέλεσμα, εκτός από την ανικανότητα του *E. coli* να καταβολίσει τα λιπαρά οξέα μικρής αλυσίδας, άλλες βλάβες όπως την αναστολή βιοσύνθεσης του κυτταρικού μαστιγίου, προβλήματα κινητικότητας και χημειόταξης, ανικανότητα χρήσης glucuramide ως πηγή άνθρακα και αυξημένη ευαισθησία στην υψηλή οσμωτική πίεση και στα αμινογλυκοσικά αντιβιοτικά (Oshima et al. 2002). Στα πλειοτροπικά φαινόμενα αυτά, πιθανότατα υπεισέρχεται η φωσφορυλίωση νέων ρυθμιστών απόκρισης από την AtoS κινάση, καθώς και η ρύθμιση της έκφρασης επιπρόσθετων γονιδίων από τον AtoC. Στη μελέτη αυτή αναγνωρίστηκαν νέοι πιθανοί γονιδιακοί στόχοι του AtoC που εμπλέκονται σε βιολογικές λειτουργίες σχετικές με τους παρατηρημένους φαινότυπους ελεγχόμενους από το σύστημα AtoSC. Συγκεκριμένα, οι στόχοι περιλαμβάνουν το γονίδιο fliT που κωδικοποιεί μια πρωτεΐνη του συστήματος εξαγωγής μαστιγίου (flagellar export system) (Fraser et al. 1999) και το γονίδιο fliA που κωδικοποιεί την πολυμεράση σίγμα-28, η οποία ενεργοποιεί την μεταγραφή χημειοτακτικών και μαστιγιακών γονιδίων, όπως τα tar, tsr και το οπερόνιο flaA (Arnosti and Chamberlin 1989; Kameda 1986; Liu and Matsumura 1995). Οι στόχοι επίσης περιλαμβάνουν το γονίδιο acrD που κωδικοποιεί ένα συστατικό του συστήματος μεταφοράς AcrAD-TolC (multidrug transport system), υπεύθυνο για την βακτηριακή απόκριση σε aminoglycosides (Rosenberg et al. 2000), το γονίδιο ruuP που κωδικοποιεί έναν μεταφορέα εισόδου της

πουτρεσκίνης (Kurihara et al. 2009), η οποία αποτελεί το ενζυμικό προϊόν της αποκαρβοξυλάσης της ορνιθίνης και, τέλος, το γονίδιο yjcH που κωδικοποιεί μια διαμεμβρανική πρωτεΐνη-μεταφορέα οξικού οξέος (Gimenez et al. 2003). Συμπερασματικά, η υπολογιστική διαδικασία με τη χρήση ανάλυσης οντολογιών αναγνώρισε στόχους του παράγοντα AtoC που εμπλέκονται σε βιολογικές λειτουργίες ήδη έμμεσα συσχετισμένες με αυτόν. Αν και ο ακριβής ρόλος του AtoC στις λειτουργίες αυτές μένει να διερευνηθεί περαιτέρω, η εύρεση των νέων στόχων του σε ολόκληρο το γονιδίωμα αναμένεται να δώσει ώθηση στη μελέτη των ιδιοτήτων του σηματοδοτικού συστήματος AtoSC.

#### **2.4.2 Ανάδειξη cis-ρυθμιστικών στοιχείων με πολύ χαμηλή συντήρηση πρωτοταγούς δομής μέσω της στατιστικής ανάλυσης γονιδιακών οντολογιών**

Η αλληλουχία DNA των περισσότερων cis-ρυθμιστικών στοιχείων είναι ελάχιστα διατηρημένη, καθιστώντας την εύρεση αλληλουχίας consensus εξαιρετικά δύσκολη. Στην παρούσα εργασία αναγνωρίστηκαν 22 νέα σημεία πρόσδεσης του AtoC, τα οποία δεν παρουσιάζουν ιδιαίτερη ομολογία πρωτοταγούς δομής και χαρακτηρίζονται από 4 διαφορετικά μοτίβα οι στόχοι των οποίων δεν επικαλύπτονται. Προηγούμενες μελέτες έχουν δείξει ότι οι βακτηριακοί μεταγραφικοί παράγοντες μπορούν να αναγνωρίσουν πολλά μη κανονικά (non canonical) σημεία πρόσδεσης, τα οποία δεν υπόκεινται σε αλληλουχία consensus (Laub et al. 2002; Wade et al. 2005). Η μέθοδος που αναπτύχθηκε στην παρούσα διατριβή αντιμετωπίζει αποτελεσματικά το φαινόμενο αυτό, καθώς συσχετίζει βιολογικά, ελάχιστα διατηρημένες αλληλουχίες που όμως αποτελούν σημεία πρόσδεσης του ίδιου μεταγραφικού παράγοντα, όπως επιβεβαιώθηκε και πειραματικά. Η επιτυχής πρόβλεψη γονιδιακών στόχων του AtoC αλλά και του LexA, μέσω της αναλύσης οντολογιών σε επίπεδο ολόκληρου γονιδιώματος, αποτελεί τη βάση για την περαιτέρω ανάπτυξη της μεθοδολογίας για συστηματική μελέτη μεγαλύτερου αριθμού μεταγραφικών παραγόντων. Επιπρόσθετα, αν και η μέθοδος αυτή αξιοποιεί μία θέση πρόσδεσης ως μοναδικό δεδομένο, θα μπορούσε να γίνει ακόμα πιο αποτελεσματική συνδυάζοντας και άλλα δεδομένα όπως αναλύσεις μικροσυστοιχιών και πληροφορίες μεταβολικών μονοπατιών.

#### **2.4.3 Προοπτικές ανάπτυξης της υπολογιστικής διαδικασίας**

Η σύνδεση μεταγραφικών παραγόντων σε συγκεκριμένες θέσεις του γονιδιώματος αποτελεί τη θεμελιώδη βάση των γονιδιακών ρυθμιστικών δικτύων. Σύγχρονες μέθοδοι υψηλής τροφοδοσίας/απόδοσης δεδομένων (high throughput), όπως η

μέθοδος αλληλούχησης ανοσοκατακρημνισμένης χρωματίνης (ChIP-seq) (Hawkins et al. 2010) έχουν χρησιμοποιηθεί πολύ πρόσφατα για να παράσχουν μεγάλο αριθμό αλληλουχιών-στόχων διαφόρων μεταγραφικών παραγόντων (Nielsen et al. 2008). Οι θέσεις πρόσδεσης ενός και μόνο μεταγραφικού παράγοντα σε ευκαρυωτικό κύτταρο μπορούν σε μια δεδομένη στιγμή να ανέρχονται από εκατοντάδες σε μερικές χιλιάδες. Τα ευρύματα αυτά οδήγησαν στο συμπέρασμα ότι τα γεγονότα πρόσδεσης υπερβαίνουν σε μεγάλο βαθμό τους πιθανούς άμεσους γονιδιακούς στόχους (MacQuarrie et al. 2011). Ως εκ τούτου, η έννοια της μεταγραφικής ρύθμισης μέσω πρόσδεσης μεταγραφικών παραγόντων επαναπροσδιορίστηκε με την εισαγωγή της έννοιας των πολύπλοκων δικτύων ανάδρασης, οι ιδιότητες των οποίων καθορίζονται όχι μόνο από τις ενεργές προσδέσεις μεταγραφικών παραγόντων αλλά και από τις αλληλεπιδράσεις των προϊόντων των αντίστοιχων γονιδίων. Οι αλληλεπιδράσεις αυτές οδηγούν στην εκδήλωση πολύπλοκων βιοχημικών φαινοτύπων, με αποτέλεσμα τη δύσκολη αποτίμηση της λειτουργικότητας των επιμέρους cis-ρυθμιστικών αλληλουχιών. Συμπερασματικά, προκύπτει το εύλογο ερώτημα κατά πόσο μια σύνδεση μεταγραφικού παράγοντα στο DNA που δεν έχει άμεση επίδραση στη μεταγραφή ενός γονιδίου, δεν είναι στην πραγματικότητα μια έμμεσα λειτουργική θέση με σημαντικό ρόλο στο ρυθμιστικό δίκτυο μέσω της αλληλεπιδρασης με μη γειτονικά γονίδια ή με τα προϊόντα τους (MacQuarrie et al. 2011). Το πρόβλημα αυτό έχει αντιμετωπιστεί με την προσπάθεια εφαρμογής μεθοδολογιών κατασκευής μοτίβων, με σκοπό την κατηγοριοποίηση των θέσεων πρόσδεσης και την αναγωγή κοινών δομών (patterns) DNA σε λειτουργικούς ρόλους (Nielsen et al. 2008). Η μεθοδολογία κατασκευής και λειτουργικής ανάλυσης μοτίβων που αναπτύχθηκε στην παρούσα διατριβή προσφέρει ένα νέο εργαλείο προς την κατεύθυνση αυτή. Η στατιστική ανάλυση υπεραντιπροσώπευσης γονιδιακών οντολογιών σε στόχους μοτίβων, με την κατάλληλη προσαρμογή για δεδομένα υψηλής τροφοδοσίας/απόδοσης σε προκαρυωτικά αλλά και ευκαρυωτικά γονιδιώματα, έχει την δυναμική να συντελέσει στη μείωση της πολυπλοκότητας των ρυθμιστικών δικτύων, μέσω της εισαγωγής στατιστικού κριτηρίου λειτουργικότητας σε ομάδες ρυθμιστικών στοιχείων που παρουσιάζουν μεταξύ τους μικρή συντήρηση πρωτοταγούς δομής.

## Βιβλιογραφία

- Alper H, Miyaoku K, Stephanopoulos G. 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 23(5):612-6.
- Amouyal M. 2005. Gene regulation at-a-distance in *E. coli*: new insights. *C R Biol* 328(1):1-9.
- Arnosti DN, Chamberlin MJ. 1989. Secondary sigma factor controls transcription of flagellar and chemotaxis genes in *Escherichia coli*. *Proc Natl Acad Sci U S A* 86(3):830-4.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT and others. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25-9.
- Barrett CL, Price ND, Palsson BO. 2006. Network-level analysis of metabolic regulation in the human red blood cell using random sampling and singular value decomposition. *BMC Bioinformatics* 7:132.
- Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B. 2007. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. *Plant J* 50(5):825-38.
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2(3):727-38.
- Bonifer C. 1999. Long-distance chromatin mechanisms controlling tissue-specific gene locus activation. *Gene* 238(2):277-89.
- Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. 2004. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20(18):3710-5.
- Boyle NR, Morgan JA. 2009. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst Biol* 3:4.
- Broun P, Gettner S, Somerville C. 1999. Genetic engineering of plant lipids. *Annu Rev Nutr* 19:197-216.
- Camon E, Magrane M, Barrell D, Binns D, Fleischmann W, Kersey P, Mulder N, Oinn T, Maslen J, Cox A and others. 2003. The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Res* 13(4):662-72.
- Canellakis ES, Paterakis AA, Huang SC, Panagiotidis CA, Kyriakidis DA. 1993. Identification, cloning, and nucleotide sequencing of the ornithine decarboxylase antizyme gene of *Escherichia coli*. *Proc Natl Acad Sci U S A* 90(15):7129-33.
- Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res* 37(Database issue):D588-92.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-90.

- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104(6):1777-82.
- Famili I, Palsson BO. 2003. Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices. *J Theor Biol* 224(1):87-96.
- Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3:121.
- Kitano H. 2001. Foundations of systems biology. Cambridge, Mass.: MIT Press. xv, 297 p.
- Fernandez De Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, Woodgate R. 2000. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol Microbiol* 35(6):1560-72.
- Fraser GM, Bennett JC, Hughes C. 1999. Substrate-specific binding of hook-associated proteins by FlgN and FliT, putative chaperones for flagellum assembly. *Mol Microbiol* 32(3):569-80.
- Galtier N, Gouy M, Gautier C. 1996. SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12(6):543-8.
- Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muniz-Rascado L, Martinez-Flores I, Salgado H and others. 2008. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* 36(Database issue):D120-4.
- Gimenez R, Nunez MF, Badia J, Aguilar J, Baldoma L. 2003. The gene yjcG, cotranscribed with the gene acs, encodes an acetate permease in *Escherichia coli*. *J Bacteriol* 185(21):6448-55.
- Grafahrend-Belau E, Schreiber F, Koschutzki D, Junker BH. 2009. Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. *Plant Physiol* 149(1):585-98.
- Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* 36(Web Server issue):W181-4.
- Grigoroudis AI, Panagiotidis CA, Lioliou EE, Vlassi M, Kyriakidis DA. 2007. Molecular modeling and functional analysis of the AtoS-AtoC two-component signal transduction system of *Escherichia coli*. *Biochim Biophys Acta* 1770(8):1248-58.
- Hawkins RD, Hon GC, Ren B. 2010. Next-generation genomics: an integrative approach. *Nat Rev Genet* 11(7):476-86.
- Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V. 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys J* 90(4):1453-61.
- Hoch JA. 2000. Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* 3(2):165-70.
- Hughes JD, Estep PW, Tavazoie S, Church GM. 2000. Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J Mol Biol* 296(5):1205-14.

- Janký R, van Helden J. 2008. Evaluation of phylogenetic footprint discovery for predicting bacterial cis-regulatory elements and revealing their evolution. *BMC Bioinformatics* 9:37.
- Kelley WL. 2006. Lex marks the spot: the virulent side of SOS and a closer look at the LexA regulon. *Mol Microbiol* 62(5):1228-38.
- Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, Peralta-Gil M, Karp PD. 2005. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res* 33(Database issue):D334-7.
- Kitano H. 2002. Systems biology: a brief overview. *Science* 295(5560):1662-4.
- Komeda Y. 1986. Transcriptional control of flagellar genes in *Escherichia coli* K-12. *J Bacteriol* 168(3):1315-8.
- Kurihara S, Tsuboi Y, Oda S, Kim HG, Kumagai H, Suzuki H. 2009. Putrescine Importer PuuP of *Escherichia coli* K-12. *J Bacteriol*.
- Kyriakidis DA, Theodorou MC, Filippou PS, Kyriakidis KD, Tiligada E. 2008. Effect of histamine on the signal transduction of the AtoS-AtoC two component system and involvement in poly-(R)-3-hydroxybutyrate biosynthesis in *Escherichia coli*. *Amino Acids* 35(1):45-52.
- Laub MT, Chen SL, Shapiro L, McAdams HH. 2002. Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proc Natl Acad Sci U S A* 99(7):4632-7.
- Lee JM, Gianchandani EP, Papin JA. 2006. Flux balance analysis in the era of metabolomics. *Brief Bioinform* 7(2):140-50.
- Lee KH, Park JH, Kim TY, Kim HU, Lee SY. 2007. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 3:149.
- Liolioú EE, Mimitou EP, Grigoroudis AI, Panagiotidis CH, Panagiotidis CA, Kyriakidis DA. 2005. Phosphorylation activity of the response regulator of the two-component signal transduction system AtoS-AtoC in *E. coli*. *Biochim Biophys Acta* 1725(3):257-68.
- Liu X, Matsumura P. 1995. An alternative sigma factor controls transcription of flagellar class-III operons in *Escherichia coli*: gene sequence, overproduction, purification and characterization. *Gene* 164(1):81-4.
- Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. 2007. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* 3:135.
- MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ. 2011. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* 27(4):141-8.
- Mahadevan R, Schilling CH. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5(4):264-76.
- Marchal K, De Keersmaecker S, Monsieurs P, van Boxel N, Lemmens K, Thijs G, Vanderleyden J, De Moor B. 2004. In silico identification and experimental validation of PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. *Genome Biol* 5(2):R9.
- Matta MK, Liolioú EE, Panagiotidis CH, Kyriakidis DA, Panagiotidis CA. 2007. Interactions of the antizyme AtoC with regulatory elements of the *Escherichia coli* atoDAEB operon. *J Bacteriol* 189(17):6324-32.
- McGuire AM, Hughes JD, Church GM. 2000. Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res* 10(6):744-57.
- Mueller LA, Zhang P, Rhee SY. 2003. AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol* 132(2):453-60.

- Mullapudi N, Lancto CA, Abrahamsen MS, Kissinger JC. 2007. Identification of putative cis-regulatory elements in *Cryptosporidium parvum* by de novo pattern finding. *BMC Genomics* 8:13.
- Nakatsui M, Horimoto K, Okamoto M, Tokumoto Y, Miyake J. 2010. Parameter optimization by using differential elimination: a general approach for introducing constraints into objective functions. *BMC Syst Biol* 4 Suppl 2:S9.
- Newberg LA, Thompson WA, Conlan S, Smith TM, McCue LA, Lawrence CE. 2007. A phylogenetic Gibbs sampler that yields centroid solutions for cis-regulatory site prediction. *Bioinformatics* 23(14):1718-27.
- Nielsen R, Pedersen TA, Hagenbeek D, Moulou P, Siersbaek R, Megens E, Denissov S, Borgesen M, Francoijis KJ, Mandrup S and others. 2008. Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev* 22(21):2953-67.
- Oberhardt MA, Palsson BO, Papin JA. 2009. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 5:320.
- Orchard K, May GE. 1993. An EMSA-based method for determining the molecular weight of a protein--DNA complex. *Nucleic Acids Res* 21(14):3335-6.
- Oshima T, Aiba H, Masuda Y, Kanaya S, Sugiura M, Wanner BL, Mori H, Mizuno T. 2002. Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. *Mol Microbiol* 46(1):281-91.
- Palsson B, Knovel (Firm). 2006. Systems biology: properties of reconstructed networks. Cambridge ; New York: Cambridge University Press. xii, 322 p.
- Park JH, Lee KH, Kim TY, Lee SY. 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A* 104(19):7797-802.
- Patil KR, Nielsen J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 102(8):2685-9.
- Plata G, Hsiao TL, Olszewski KL, Llinas M, Vitkup D. 2010. Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. *Mol Syst Biol* 6:408.
- Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35(Database issue):D61-5.
- Rosenberg EY, Ma D, Nikaido H. 2000. AcrD of *Escherichia coli* is an aminoglycoside efflux pump. *J Bacteriol* 182(6):1754-6.
- Roth FP, Hughes JD, Estep PW, Church GM. 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol* 16(10):939-45.
- Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. .
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432(7018):779-82.
- Schwender J, Ohlrogge JB. 2002. Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing *Brassica napus* embryos. *Plant Physiol* 130(1):347-61.
- Schwender J, Ohlrogge JB, Shachar-Hill Y. 2003. A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos. *J Biol Chem* 278(32):29442-53.

- Schwender J, Shachar-Hill Y, Ohlrogge JB. 2006. Mitochondrial metabolism in developing embryos of *Brassica napus*. *J Biol Chem* 281(45):34040-7.
- Shimada T, Ishihama A, Busby SJ, Grainger DC. 2008. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res* 36(12):3950-5.
- Smallbone K, Simeonidis E, Broomhead DS, Kell DB. 2007. Something from nothing: bridging the gap between constraint-based and kinetic modelling. *Febs J* 274(21):5576-85.
- Spencer W, Siam R, Ouimet MC, Bastedo DP, Marczynski GT. 2009. CtrA, a global response regulator, uses a distinct second category of weak DNA binding sites for cell cycle transcription control in *Caulobacter crescentus*. *J Bacteriol* 191(17):5458-70.
- Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420(6912):190-3.
- Stephanopoulos G, Alper H, Moxley J. 2004. Exploiting biological complexity for strain improvement through systems biology. *Nat Biotechnol* 22(10):1261-7.
- Thelen JJ, Ohlrogge JB. 2002. Metabolic engineering of fatty acid biosynthesis in plants. *Metab Eng* 4(1):12-21.
- Theodorou MC, Panagiotidis CA, Panagiotidis CH, Pantazaki AA, Kyriakidis DA. 2006. Involvement of the AtoS-AtoC signal transduction system in poly-(R)-3-hydroxybutyrate biosynthesis in *Escherichia coli*. *Biochim Biophys Acta* 1760(6):896-906.
- Theodorou MC, Theodorou EC, Panagiotidis CA, Kyriakidis DA. 2007. Spermidine triggering effect to the signal transduction through the AtoS-AtoC/Az two-component system in *Escherichia coli*. *Biochim Biophys Acta* 1770(8):1104-14.
- Thiele I, Palsson BO. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5(1):93-121.
- Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P, Moreau Y. 2001. A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics* 17(12):1113-22.
- Thijs G, Marchal K, Lescot M, Rombauts S, De Moor B, Rouze P, Moreau Y. 2002. A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. *J Comput Biol* 9(2):447-64.
- van Helden J, Andre B, Collado-Vides J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol* 281(5):827-42.
- Varma A, Palsson BO. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 60(10):3724-31.
- Wade JT, Reppas NB, Church GM, Struhl K. 2005. Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites. *Genes Dev* 19(21):2619-30.
- Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. 2005. Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*. *J Biol Chem* 280(2):1448-56.
- Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY. 2005. MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol* 138(1):27-37.



## Παράρτημα Α

### 1. Κώδικας Κεφαλαίου 1

**Poutína βελτιστοποίησης Cobra toolbox (Matlab)**

```
model = readCbModel('model',2000);

model=addReaction(model,'EX_Al(e)','L-alanine[e] <=>');

model=addReaction(model,'EX_Gln(e)','L-glutamine[e] <=>');

model=addReaction(model,'EX_suc(e)','sucrose[e] <=>');

model=addReaction(model,'EX_o2(e)','O2[e] <=>');

model=addReaction(model,'EX_h(e)','H[e] <=>');

model=addReaction(model,'EX_h2o(c)','H2O[c] <=>');

model=addReaction(model,'EX_co2(e)','CO2[c] <=>');

model=addReaction(model,'EX_light(e)','light[e] <=>');

model=addReaction(model,'EX_for(e)','formate[e] <=>');

model=addReaction(model,'EX_ac(e)','acetate[e] <=>');

model=addReaction(model,'EX_eth(e)','ethanol[e] <=>');

model=addReaction(model,'EX_pi(e)','phosphate[e] <=>');

model=addReaction(model,'EX_UDP_gluc(e)','UDP_D_glucose[c] <=>');

model=addReaction(model,'EX_biomass(e)','biomass1g[e] <=>');

model=addReaction(model,'EX_lac(e)','lactate[e] <=>');

model=addReaction(model,'EX_no3(e)','NO3[e] <=>');

model=addReaction(model,'EX_sulfate(e)','sulfate[e] <=>');

model=addReaction(model,'EX_nitricOxide(e)','nitricOxide[e] <=>');

model=addReaction(model,'EX_ammonia(e)','ammonia[e] <=>');

model=addReaction(model,'EX_red_cytc(e)','red-cytc[m] <=>');

model=addReaction(model,'EX_UQH2(m)','UQH2[m] <=>');

model=addReaction(model,'EX_coa(c)','coenzyme-A[c] <=>');
```

```

model=addReaction(model,'EX_coa(p)','coenzyme-A[p] <=>');
model=addReaction(model,'EX_coa(m)','coenzyme-A[m] <=>');
model=addReaction(model,'EX_GDP(p)','GDP[p] <=>');
model=addReaction(model,'EX_UDP(c)','UDP[c] <=>');
model=addReaction(model,'EX_ADP(c)','ADP[c] <=>');
model=addReaction(model,'EX_ADP(p)','ADP[p] <=>');
model=addReaction(model,'EX_ADP(m)','ADP[m] <=>');
model=addReaction(model,'EX_NADP(p)','NADP[p]<=>');
model=addReaction(model,'EX_NADP(c)','NADP[c] <=>');

model=addReaction(model,'EX_nicotinamide','nicotinamide[p] ->');
model=addReaction(model,'EX_5_10_m_THF_p','5-10-methylene-THF[p] ->');
model=addReaction(model,'EX_5_m_THF_p','5-methyl-THF[p] ->');
model=addReaction(model,'EX_ornithine','L-ornithine[c] ->');
model=addReaction(model,'EX_dhpantoate_p','2-dehydropantoate[p] ->');
model=addReaction(model,'EX_putr_c','putrescine[c] ->');
model=addReaction(model,'EX_thfolate_p','tetrahydrofolate[p] ->');
model=addReaction(model,'EX_urea','urea[c] ->');
model=addReaction(model,'EX_glut_disulf_p','glutathione-disulfide[p] ->');
model=addReaction(model,'EX_oxferr_p','oxferredoxin[p] ->');
model=addReaction(model,'EX_quinolinate_p','quinolinate[p] ->');
model=addReaction(model,'EX_gaba','4-aminobutyrate[c] ->');
model=addReaction(model,'EX_glyox','glyoxylate[m] ->');

%Constraints

model=changeRxnBounds(model,['EX_suc(e)'],[-58.3],l);
model=changeRxnBounds(model,['EX_Al(e)'],[-7.36],l);
model=changeRxnBounds(model,['EX_Gln(e)'],[-4.25],l);
model=changeRxnBounds(model,['RXN_1_7_1_1_p'],[1.2],u);

```

```
model=changeRxnBounds(model,['EX_co2(e)',[142],'l');  
model=changeRxnBounds(model,['RXN_4_1_1__p'],[0],'b');
```

```
model = changeObjective(model,'J_biomass');  
sol = optimizeCbModel(model);  
printFluxVector(model,sol.x,true,true)
```

### ***Flux Variability Analysis***

```
[minFlux,maxFlux] = fluxVariability(model,90);  
fva=[model.rxns num2cell(minFlux) num2cell(maxFlux)];
```

### ***Routίνα δειγματοληψίας και SVD***

```
[modelSampling1,samples1] = sampleCbModel(model,'sampling1');  
[U1,S1,V1] = svd(samples1);  
mapcaplot(samples1)  
diagS1=diag(S1);  
varS1=diagS1.*diagS1;  
eigenval=diagS1/sum(diagS1(1:size(diagS1)));
```

## **1. Κώδικας κεφαλαίου 2**

### ***Routίνες Perl***

```
# Routines.pm  
  
package Routines;  
  
sub subseq_from_large_fasta {  
  
    use Bio::Perl;  
  
    my($start, $end, $genome)=@_;  
  
    ## Fetches a sequence from a large fasta between start and end  
  
    my $seqio = new Bio::SeqIO(-format => 'largefasta',
```

```

        -file => $genome);

my $pseq = $seqio->next_seq();

if ($start < $end){

    $seqFragment = $pseq->subseq($start,$end);

}

if ($start > $end){

    $seqFragment = $pseq->subseq($end,$start);

    $seqFragment=revcom($seqFragment) -> seq;

}

return($seqFragment);

}

#####
sub search_feature{

#Returns a table of genes with a particular feature

#Usage: search_feature('file.gbff','refseq_id', 'feature')

use Bio::SeqIO;

use Bio::DB::GenBank;

my ($gbString, $seqId)=@_;

$gb = Bio::DB::GenBank->new($gbString) ;

$seq = $gb->get_Seq_by_id($seqId);

my $annot= $seq->annotation ;

my $geneName;

my $product;

my @genes;

for my $feat_object ($seq->get_SeqFeatures) {

    if ($feat_object->primary_tag eq "CDS"){

        for my $tag ($feat_object->get_all_tags) {

            if ($tag eq "gene"){

                for my $value ($feat_object->get_tag_values($tag)) {

                    $geneName=$value;

```

```

        }

    }

    if ($tag eq "product"){

        for my $value ($feat_object->get_tag_values($tag)) {

            $product=$value;

        }

    }

    if ($product=~/response regulator/){

        push @genes,$geneName;

    }

}

return (@genes);

}

#####
## Fetches the intergenic sequence before $gene from database $db, bp1 base pairs of the promoter and bp2 base pairs
## of the gene

sub cut_promoter2{

    my ($gene,$db,$bp1,$bp2)=@_;

    my $dsn = "DBI:mysql:$db:localhost";
    my $db_user_name = 'root';
    my $db_password = '12345';
    my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
    my $seq;
    my $strand,$locus,$index;
    my $query="select strand,locus from annotation where gene_name ='$gene';";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    ($strand,$locus)= $sth->fetchrow_array();
}

```

```

my $query="select ind from intergenic where locus ='$locus';";
my $sth = $dbh->prepare($query);
$sth->execute();
($index)= $sth->fetchrow_array();
if ($strand eq -1){
    my $query="select substr(seq,1,$bp1) from intergenic where ind =$index;";
    my $query2="select substr(seq,-$bp2) from gene_seq where locus ='$locus';";
    #print $query2;
    my $sth = $dbh->prepare($query);
    $sth->execute();
    ($intergenic_seq)= $sth->fetchrow_array();
    my $sth2 = $dbh->prepare($query2);
    $sth2->execute();
    ($gene_seq)= $sth2->fetchrow_array();
    $seq=$gene_seq.$intergenic_seq;
}
elsif ($strand eq 1){
    $index=$index-1;
    my $query_check_length="select length(seq) from intergenic where ind = $index;";
    my $sth3 = $dbh->prepare($query_check_length); $sth3->execute();
    my ($interg_length)= $sth3->fetchrow_array();
    my $query;
    if ($interg_length < $bp1){
        $query="select seq from intergenic where ind = $index;";
    } else {
        $query="select substr(seq,-$bp1) from intergenic where ind = $index;";
    }
    my $query2="select substr(seq,1,$bp2) from gene_seq where locus ='$locus';";
    #print $query2;
    my $sth = $dbh->prepare($query); $sth->execute();
    ($intergenic_seq)= $sth->fetchrow_array();
}

```

```

my $sth2 = $dbh->prepare($query2);
$sth2->execute();
($gene_seq)= $sth2->fetchrow_array();
$seq=$intergenic_seq.$gene_seq;
}

return ($seq);
}

#####
## fetches intergenic region before $gene from database $db
sub cut_promoter{
my ($gene,$db)=@_;
my $dsn = "DBI:mysql:$db:localhost";
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $strand,$locus,$index;
my $query="select strand,locus from annotation where gene_name ='$gene';";
my $sth = $dbh->prepare($query);
$sth->execute();
($strand,$locus)= $sth->fetchrow_array();
my $query="select ind from intergenic where locus ='$locus';";
my $sth = $dbh->prepare($query);
$sth->execute();
($index)= $sth->fetchrow_array();

if ($strand eq -1){
    my $query="select seq from intergenic where ind ='$index';";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    ($seq)= $sth->fetchrow_array();
}

```

```

elsif ($strand eq 1){

    my $query="select seq from intergenic where ind =\'$index\' -1 ;";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    ($seq)= $sth->fetchrow_array();

}

return ($seq);

#####
#####



#####
##fetches intergenic with index @_


sub cut_intergenic{

    my ($index,$database)=@_ ;
    my $dsn = "DBI:mysql:$database:localhost";
    my $db_user_name = 'root';
    my $db_password = '12345';
    my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
    my $query="select seq from intergenic where ind =$index;";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    (my $seq)= $sth->fetchrow_array();

    return ($seq);

}

#####
##


sub get_intergenic_ind{

    my ($database,$index)=@_ ;
    my $dsn = "DBI:mysql:$database:localhost";
    my $db_user_name = 'root';

```

```

my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $query="select ind from intergenic where ind not like $index;";
my $sth = $dbh->prepare($query);
$sth->execute();
my @indexes;
while (my @results = $sth->fetchrow()){
    my ($index)=@results;
    push @indexes,$index;
}
return (@indexes);

#####
#



sub len_promoter{
my ($gene,$db)=@_;
my $dsn = "DBI:mysql:$db:localhost";
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $strand,$index;
my $locus;
my $query="select strand from annotation where gene_name ='$gene';";
my $sth = $dbh->prepare($query);
$sth->execute();
($strand)= $sth->fetchrow_array();
my $query="select locus from annotation where gene_name ='$gene';";
my $sth = $dbh->prepare($query);
$sth->execute();
($locus)= $sth->fetchrow_array();
my $query="select ind from intergenic where locus ='$locus';";

```

```

my $sth = $dbh->prepare($query);
$sth->execute();
($index)= $sth->fetchrow_array();
if ($strand eq -1){
    my $query="select length(seq) from intergenic where ind ='$index';";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    ($len_seq)= $sth->fetchrow_array();
}
elsif ($strand eq 1){
    my $query="select length(seq) from intergenic where ind ='$index' -1 ;";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    ($len_seq)= $sth->fetchrow_array();
}
return ($len_seq);
}

#####
## fetches gene annotations from database
sub annotate_gene{
use DBI;
my $dsn = 'DBI:mysql:ecoli:localhost';
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my ($gene)=@_;
my $query="select gene_name,locus,gene_start,gene_end,strand,product,func,note,ind,uniprot from annotation
where gene_name = '$gene'";
my $sth = $dbh->prepare($query);
$sth->execute();

```

```

my ($gene_name,$locus,$gene_start,$gene_end,$gene_strand,$product,$func,$note,$ind,$uniprot_id)= $sth->fetchrow_array();

return ($gene_name,$locus,$gene_start,$gene_end,$gene_strand,$product,$func,$note,$ind,$uniprot_id);

}

#####
## returns max from array

sub max{

my (@num_array)=@_ ;

my $max;

$max=@num_array[0];

foreach my $number(@num_array){

    if ($number > $max){$max=$number; }

}

return ($max);

}

#####

## returns min from array

sub min{

my (@num_array)=@_ ;

my $min;

$min=@num_array[0];

foreach my $number(@num_array){

    if ($number < $min){$min=$number; }

}

return ($min);

}

#####

### Deletes a word from array

sub del{

```

```

my ($word,@array1)=@_ ;
my @array2;
foreach my $w(@array1){
    if ($w ne $word){
        push @array2,$w ;
    }
}
return (@array2);
}

#####
### Returns .mtrx from motif id (ex. 'box_4_1_yTTnnnwCwGwnnnAwrnwT' and motif filesub get_mtrx2{
my ($motif_id,$database,$file)=@_ ;
my $motif_length= length((split "_",$motif_id)[3]);
open (FILE, "<$file") or die "open failed: $!" ;
my $number;
my %hash_lines=();
my $index=1;
while (defined(my $line=<FILE>)){
    chomp $line ;
    $hash_lines{ $index }=$line;
    if ($line =~ /$motif_id/){$number=$index;}
    $index++;
}
for (my $i=$number; $i<=$number + $motif_length + 4 ; $i++){
    print $hash_lines{$i},"\n";
}
close(FILE);
}

#####
##### Fetches intergenic regions as putative promoters

```

```

sub get_unique_promoters{
    my ($db)=@_;
    use DBI;
    my $gene;
    my $strand;
    my $dsn = "DBI:mysql:$db:localhost";
    my $db_user_name = 'root';
    my $db_password = '12345';
    my $dbh = DBI->connect($dsn,$db_user_name,$db_password);
    my @genes;
    for (my $i=1 ; $i<=4466;$i++){
        my $query1="select gene_name,strand from annotation where ind=$i;";
        my $sth = $dbh->prepare($query1);
        $sth->execute();
        my ($gene,$strand)=$sth->fetchrow_array();
        if ($strand eq '1'){
            push @genes,$gene;
        }
    }
    return(@genes);
}

sub get_unique_promoters2{
    my ($db,$bp)=@_;
    use DBI;
    my $strand;
    my $dsn = "DBI:mysql:$db:localhost";
    my $db_user_name = 'root';
    my $db_password = '12345';
    my $dbh = DBI->connect($dsn,$db_user_name,$db_password);
    my @loci_to_del,@genes_to_del;
}

```

```

my $query1="SELECT locus,gene_name,strand,ind FROM annotation ;";
my $sth = $dbh->prepare($query1); $sth->execute();
while (my @results = $sth->fetchrow()) {
    my ($prev_locus,$gene_name,$strand,$ind)=@results;
    if ($strand eq -1){
        my $query2="SELECT strand FROM annotation where ind=$ind+1;";
        my $sth2 = $dbh->prepare($query2); $sth2->execute();
        my ($next_strand)= $sth2->fetchrow_array();
        if ($next_strand eq 1){
            my $query3="SELECT length(seq) FROM intergenic where locus=\'$prev_locus\'";
            my $sth3 = $dbh->prepare($query3); $sth3->execute();
            my ($intergenic_length)=$sth3->fetchrow_array();
            if ($intergenic_length < 2*$bp){
                #print "$intergenic_length\n";
                push @loci_to_del,$prev_locus;
            }
        }
    }
}
foreach my $l(@loci_to_del){
    my $query4="SELECT gene_name FROM annotation where locus=\'$l\'";
    my $sth4 = $dbh->prepare($query4); $sth4->execute();
    my ($gene)= $sth4->fetchrow_array();
    push @genes_to_del,$gene;
}
return(@genes_to_del);
}

##### Finds all gene_names (db)
sub get_all_genes{
    my ($db)=@_;

```

```

use DBI;

my $gene;
my $strand;
my $dsn = "DBI:mysql:$db:localhost";
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn,$db_user_name,$db_password);
my @genes;
my $query1="SELECT gene_name FROM annotation;";
#my $query1="SELECT locus FROM intergenic;";
my $sth = $dbh->prepare($query1);
$sth->execute();
while (my @results = $sth->fetchrow()) {
    my ($gene)=@results;
    push @genes,$gene;
}
return(@genes);
}

##### MotifSampler parser

sub parse_motifsampler{
#####
# Usage: parse_motifsampler.out.pl all.out(motifsampler.out.gff) database
# output:      -
#####
my ($filein,$database_name)=@_;
use DBI;
use lib '/home/leyteris/perl';
use Routines;
my $dsn = "DBI:mysql:$database_name:localhost";
my $db_user_name = 'root';

```

```
my $db_password = '12345';

my $dbh = DBI->connect($dsn, $db_user_name, $db_password);

open (FILEIN, "<$filein") or die "open failed: $!" ;

my %hash_of_lines;

my $line_index;

my @motif_indexes;

$line_index=1;

while (defined(my $line=<FILEIN>)) {

    chomp $line ;

    $hash_of_lines {$line_index} = $line;

    if ($line =~ /#id/){

        push @motif_indexes,$line_index;

    }

    $line_index ++;

}

push @motif_indexes,$line_index;

my $times=scalar(@motif_indexes)-1;

for (my $i=0; $i<=$times; $i ++){

    unless ($i == $times){

        my @indexes=$motif_indexes[$i]..$motif_indexes[$i+1]-1;

        my $id_line=$hash_of_lines{$indexes[0]};

        my $id=(split ":",((split "\t",$id_line)[0]))[1] ;

        my $consensus=(split ":",((split "\t",$id_line)[1]))[1] ;

        my $sequences=(split ":",((split "\t",$id_line)[2]))[1] ;

        my $instances=(split ":",((split "\t",$id_line)[3]))[1] ;

        my $cs=(split ":",((split "\t",$id_line)[4]))[1] ;

        my $ic=(split ":",((split "\t",$id_line)[5]))[1] ;

        my $ll=(split ":",((split "\t",$id_line)[6]))[1] ;

        my $query4="insert into motif (motif_id,consensus,sequences,instances,cs,ic,ll) values ('$id','$consensus','$sequences','$instances','$cs','$ic','$ll')";;

        my $sth = $dbh->prepare($query4);
```

```

$sth->execute();

my $indexes_len=scalar(@indexes)-1;

my $query_motif_ind="select ind from motif where motif_id=\'$id\' order by ind desc limit 1 ;";

my $sth = $dbh->prepare($query_motif_ind);

$sth->execute();

my $motif_ind=$sth->fetchrow_array();

for (my $k=1; $k<= $indexes_len; $k ++){

    my $hit_line= $hash_of_lines{$indexes[$k]};

    my $gene_name=(split "\t",$hit_line)[0];

    my $motif_start=(split "\t",$hit_line)[3];

    my $motif_end=(split "\t",$hit_line)[4];

    my $score=(split "\t",$hit_line)[5];

    my $strand=(split "\t",$hit_line)[6];

    my $site=(split "'",((split "\t",$hit_line)[8]))[3];

    my $len_promoter=Routines::len_promoter($gene_name,'ecoli') ;

    my $motif_position1=$len_promoter-$motif_start;

    my $motif_position2=$len_promoter-$motif_end;

    if ($gene_name !~ m/^#/){

        my $query5="insert into hits
(motif_id,motif_ind,gene_name,hit_start,hit_end,score,strand,position1,position2,site) values
(''$id'',''$motif_ind'',''$gene_name'',''$motif_start'',''$motif_end'',''$score'',''$strand'',''$motif_position1'',''$motif_posi
tion2'',''$site'');";

        my $sth = $dbh->prepare($query5);

        $sth->execute();

    }

}

}

}

#####
##### sub create_tables_motif_hits{

```

```
###Usage: create_tables_motif_hits("hits_temp")

my ($database_name)=@_;
my $dsn = "DBI:mysql:$database_name:localhost";
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $query1="DROP TABLE if exists motif;";
my $sth = $dbh->prepare($query1);
$sth->execute();

my $query2="CREATE TABLE motif(motif_id VARCHAR(50),consensus VARCHAR(50),sequences
INTEGER, instances INTEGER,cs DOUBLE,ic DOUBLE,il DOUBLE,ind INTEGER UNSIGNED NOT NULL
AUTO_INCREMENT,PRIMARY KEY(ind),INDEX(motif_id));";

my $sth = $dbh->prepare($query2);
$sth->execute();

my $query6="DROP TABLE if exists hits;";
my $sth = $dbh->prepare($query6);
$sth->execute();

my $query3="CREATE TABLE hits(motif_id VARCHAR(50),motif_ind INTEGER,gene_name
VARCHAR(50),hit_start INTEGER,hit_end INTEGER,score DOUBLE,strand VARCHAR(10),position1
INTEGER,position2 INTEGER,site VARCHAR(250),ind INTEGER UNSIGNED NOT NULL
AUTO_INCREMENT,PRIMARY KEY(ind),INDEX(motif_id,gene_name));";

my $sth = $dbh->prepare($query3);
$sth->execute();
}

1;
```

### **Κεντρική διαδικασία**

Αρχείο Iteration.pl

```
#!/usr/bin/perl

#use strict;

##### Usage: iteration.pl motiflength gene database
my $initial_gene=$ARGV[1];
my $motif_length=$ARGV[0];
```

```
my $db=$ARGV[2];
$cmd="mkdir ~/motif/$db/data";
system($cmd);
$cmd = "mysql --skip-column-names --silent -u root --password=***** -e 'create database $db;'";
system($cmd);
$cmd="mysql --skip-column-names --silent -u root --password=***** -D $db --local-infile=1 < ~/perl/database.sql ";
system($cmd);
my $background=~"/motif/backgroundNC_000913_3.bg";
my $sampler_runs=20;
my $percentile_value1=100;
my $percentile_value2=30;
my $criterion='max';
my $deletion='yes';
my $deletion_percentile='no';
my $specific_positions="yes";
my $delete_non_annotated="no";
my $p1=77; my $p2=61;
my $promoter_bp=250; my $gene_bp=100;
my $prior=0.2;
my $temp_db="temp1";
my $subdir="data";
use DBI;
my $dsn = "DBI:mysql:$db:localhost";
my $dsn_temp = "DBI:mysql:$temp_db:localhost";
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $dbh_temp = DBI->connect($dsn_temp, $db_user_name, $db_password);
use lib '/home/leyteris/perl';
use Routines;
use List::Uniq ':all';
```

```

use List::Util;
use Bio::SeqIO;
use Statistics::Descriptive;
open (OUT,>">times.txt")
my @geneNames1=($initial_gene);
print @geneNames1,"\\n";
Routines::create_tables_motif_hits("$db");
my @geneNames2=Routines::get_unique_promoters($db);
my @loci_to_del=Routines::get_unique_promoters2($db,$promoter_bp);
##### Sigma54-related genes
#my
@geneNames2=('argT','astC','chaC','dcuD','ddpX','fdhF','glmY','glnA','glnH','glnK','gltI','hycA','hydA','hyfA','hypA','ib
pB','kch','nac','norV','potF','prpB','pspA','pspG','puuP','rpoH','rtcB','rtcR','rutA','yaiS','ybhK','yeaG','yfhK','ygjG','yhdW','
zraP','zraS');
@geneNames2=uniq(@geneNames2);
foreach my $l(@loci_to_del){
    @geneNames2=Routines::del($l,@geneNames2);
}
foreach my $ig(@geneNames1){
    @geneNames2=Routines::del($ig,@geneNames2);
}
#####
Delete non annotated genes
if ($delete_non_annotated eq "nai100"){
    my @geneNamesNonAnnotated;
    my $annotated=0;
    foreach my $g2(@geneNames2){
        my $query="select ind from go_terms where uniprot=(select uniprot from annotation where
gene_name=\$g2\ limit 1);";
        my $sth = $dbh->prepare($query); $sth->execute();
        my $annotation_exists= $sth->fetchrow_array();
        if ($annotation_exists eq ""){push @geneNamesNonAnnotated,$g2;}
        else {$annotated ++}
    }
}

```

```

foreach my $na(@geneNamesNonAnnotated){ @geneNames2=Routines::del($na,@geneNames2);}

print OUT "Annotated genes: $annotated\n";

}

@geneNames2=uniq(@geneNames2);

print scalar(@geneNames2);

print OUT (join "\\",\",",@geneNames2);

print OUT "\n";

my $turns=1;

my $max_ll_global=0;

sub iterate{

    my ($number_of_geneNames1,@total_genes)=@_;

    my @genes1=@total_genes[0..$number_of_geneNames1-1];

    my @genes2=@total_genes[$number_of_geneNames1..scalar(@total_genes-1)];

    my $seq1_initial=Routines::cut_promoter2($initial_gene,$db,$promoter_bp,$gene_bp);

    my $seq_obj1 = Bio::Seq->new(-seq => $seq1_initial,
                                    -display_id => $initial_gene,
                                    -desc => "promoter-wholeseq",
                                    -alphabet => "dna" );

    my %max_lls;

    my @genes;

    foreach my $gene2(@genes2){

        my $seqio_obj = Bio::SeqIO->new(-file => ">$subdir/$gene2.promoters.fasta", -format => 'fasta' );

        foreach my $gene1(@genes1){

            my $seq1=Routines::cut_promoter2($gene1,$db,$promoter_bp,$gene_bp);

            my $seq_obj1 = Bio::Seq->new(-seq => $seq1,
                                         -display_id => $gene1,
                                         -desc => "promoter-wholeseq",
                                         -alphabet => "dna" );

            $seqio_obj->write_seq($seq_obj1);

        }

    }

}

```

```

my $seq2=Routines::cut_promoter2($gene2,$db,$promoter_bp,$gene_bp);
if ($seq2 eq " "){next;}
my $seq_obj2 = Bio::Seq->new(-seq => $seq2,
                               -display_id => $gene2,
                               -desc => "promoter-wholeseq",
                               -alphabet => "dna" );
if (length($seq_obj2->seq) < $motif_length){
    @geneNames2=Routines::del($gene2,@geneNames2);
}
if (length($seq_obj2->seq) >= $motif_length){
    $seqio_obj->write_seq($seq_obj2);
    my $cmd="MotifSampler -f $subdir/$gene2.promoters.fasta -b $background -r $sampler_runs -w
$motif_length -x 0 -p $prior -o $subdir/$gene2.sampler$turns.out -m $subdir/$gene2.motifs$turns.mtrx";
    system($cmd);
    my $cmd5="rm $subdir/*promoters.fasta";
    system($cmd5);
    Routines::parse_motifsampler("$subdir/$gene2.sampler$turns.out", "$db");
    Routines::create_tables_motif_hits("$temp_db");
    Routines::parse_motifsampler("$subdir/$gene2.sampler$turns.out", "$temp_db");
    my $cmd2="cat $subdir/$gene2.motifs$turns.mtrx >> all.mtrx";
    system($cmd2);
    my $cmd3="rm $subdir/$gene2.motifs$turns.mtrx";
    system($cmd3);
    my $cmd4="cat $subdir/$gene2.sampler$turns.out >> all.out";
    system($cmd4);
    my $cmd4="rm $subdir/$gene2.sampler$turns.out";
    system($cmd4);
    my $query2;
    if ($specific_positions eq "yes"){
        $query2="SELECT $criterion(ll) FROM motif m,hits h where m.ind=h.motif_ind and
h.gene_name='\$initial_gene' and h.position1<=$p1 and h.position2>=$p2 ;"
    }
}

```

```

elsif($specific_positions eq "no") {

    $query2="SELECT $criterion(ll) FROM motif m,hits h where m.ind=h.motif_ind and
h.gene_name=\'$initial_gene\';";

}

my $sth = $dbh_temp->prepare($query2); $sth->execute();

my $max_ll= $sth->fetchrow_array();

if ($max_ll ne ""){$max_lls{ $gene2 }=$max_ll; }

}

@genes=uniq(@genes);

return(%max_lls);

}

my $max=0;

open(FILE1,>"results.txt");

print (FILE1 "size array2: \t",scalar(@geneNames2),"\n");

my $times_run=1;

my $incr=1;

for (my $i=1; $i <= scalar(@geneNames2) ; $i++){

my (%max_lls)=iterate($incr,(@geneNames1,@geneNames2));

$incr ++;

$stat = Statistics::Descriptive::Full->new();

while ( my ($hash_gene, $hash_ll) = each(%max_lls)) {

    #print "$hash_gene \t $hash_ll \n ";

    print OUT "$hash_gene\t$hash_ll\n";

    if ($hash_ll > $max_ll){

        $max_ll=$hash_ll;

        $gene_max_ll=$hash_gene;

    }

    $stat->add_data($hash_ll);

}

my $percentile1= $stat->percentile($percentile_value1);

```

```

my $percentile2= $stat->percentile($percentile_value2);
my @genesOverPercentile;
while ( my ($hash_gene, $hash_ll) = each(%max_lls)){
    if ($hash_ll >= $percentile1){
        push @genesOverPercentile,$hash_gene;
    }
##### Delete genes not increasing $max_ll_global
    if ($hash_ll < $max_ll_global and $deletion eq 'yes'){
        @geneNames2=Routines::del($hash_gene,@geneNames2);
    }
#####
##### Delete genes below percentile2
    if ($hash_ll < $percentile2 and $deletion_percentile eq 'yes'){
        @geneNames2=Routines::del($hash_gene,@geneNames2);
    }
#####
}
print OUT "percentile1:\t$percentile1\n ";
print OUT "percentile2:\t$percentile2\n ";
print OUT $gene_max_ll," \t",$max_ll," \t",$max_ll_global," \n";
print OUT "$incr-1","\t","\n";
print OUT "\n";
#####
##### percentile      #####
if ($percentile1 > $max_ll_global){
    $max_ll_global=$percentile1;
    foreach my $i(@genesOverPercentile){push @geneNames1,$i;}
    foreach my $i(@genesOverPercentile){ @geneNames2=Routines::del($i,@geneNames2);}
    @geneNames1=uniq(@geneNames1);
    @geneNames2=uniq(@geneNames2);
    print FILE1 "geneNames1: ";
    foreach my $g1(@geneNames1){

```

```

        print FILE1 "\$g1\\,\";

    }

    print FILE1 "\n";

    print FILE1 "geneNames2: ",scalar(@geneNames2),"\n";
    print FILE1 "\n\n";
    print FILE1 "geneNames2: ",join("\t",@geneNames2);
    print FILE1 "\n";
}

else{break;
}

print (FILE1 join("\t",@geneNames1),"\n");
close(FILE1);
close(OUT);

```

### ***MotifScanner output file parser***

```

#!/usr/bin/perl

#####
# Usage: parse_motifscanner.out motifscanner.out.gff database
# output:      > results annotated
#             ...gff.reg_genes.txt (list of regulated genes)
#             ...gff.reg_genes.txt.terms (HyperG test on regulated genes)
#             list of genes found in HyperG test
#####

use DBI;

use lib '/home/leyteris/perl';

use Routines;

use List::Uniq ':all';

use Bio::SeqIO;

my $filein=$ARGV[0];

my $db=$ARGV[1];

open (FILEIN, "<$filein") or die "open failed: $!" ;


```

```

my $line;

my $previous_gene,$next_gene,$motif_start,$motif_end,$score,$motif_strand,$site;
my $locus,$product,$func,$note,$ind;
my $gene_start,$gene_end,$gene_strand1,$gene_strand2;
my $previous_gene_id, $next_gene_id;
my $dsn = "DBI:mysql:$db:localhost";
my $db_user_name = 'root';
my $db_password = '12345';
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $query1="DROP TABLE if exists hit;";

my $query2="CREATE TABLE hit(gene_name VARCHAR(50),hit_start INTEGER,hit_end INTEGER,score
DOUBLE,strand VARCHAR(10),motif_id VARCHAR(250),site VARCHAR(250),ind INTEGER UNSIGNED NOT
NULL AUTO_INCREMENT,PRIMARY KEY(ind),INDEX(gene_name));";

my $sth = $dbh->prepare($query1);$sth->execute();

my $sth = $dbh->prepare($query2);$sth->execute();

my $query3="DROP TABLE if exists scan_hit;";

my $query4="CREATE TABLE scan_hit(site VARCHAR(50),motif_start INTEGER,motif_end INTEGER,score
DOUBLE,motif_strand VARCHAR(10),gene VARCHAR(250),locus VARCHAR(50),gene_start
VARCHAR(50),gene_end VARCHAR(50), gene_strand VARCHAR(20),product VARCHAR(250),function
VARCHAR(250),note VARCHAR(250), hit_start_genome VARCHAR(50),hit_end_genome VARCHAR(50),tss1
VARCHAR(50),tss2 VARCHAR(50),ind INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,PRIMARY
KEY(ind));";

my $sth = $dbh->prepare($query3);$sth->execute();

my $sth = $dbh->prepare($query4);$sth->execute();

while (defined($line=<FILEIN>)){
    chomp $line ;
    if ($line !~ /#/){
        my @words= split /\t/, $line ;
        $previous_gene=@words[0];
        $motif_start=@words[3];
        $motif_end=@words[4];
        $score=@words[5];
        $motif_strand=@words[6];
        my $id_site=@words[8];
    }
}

```

```

my @id_site_table=split /site/, $id_site;
my @id=split /\/*\//, $id_site_table[0];
my @site= split /\/*\//, $id_site_table[1];
my $query3="insert into hit (gene_name,hit_start,hit_end,score,strand,motif_id,site) values
(\$previous_gene\',\$motif_start\',\$motif_end\',\$score\',\$motif_strand\',\$id[1]\',\$site[1]\');";
my $sth = $dbh->prepare($query3);
$sth->execute();
}

}

#### Temporary query result file

my $temp_file ="temp.query.result.txt";

$cmd = "mysql --skip-column-names --silent -u root --password=12345 -D $db -e 'select * from hit order by score
DESC;' > $temp_file";
system($cmd);
close(FILEIN);

my @geneList;

open (SITES, ">sites1.txt") or die "open failed: $!" ;
open(MOTIFGENE,">motif_genes") or die "open failed: $!";
open (FILETEMP, "<$temp_file") or die "open failed: $!" ;
while (defined($line=<FILETEMP>)){
    chomp $line ;
    my @words= split /\t/, $line ;
    $previous_gene=@words[0];
    $motif_start=@words[1];
    $motif_end=@words[2];
    $score=@words[3];
    $motif_strand=@words[4];
    my $id=@words[5];
    my $site=@words[6];
    ###### Previous gene annotation #####
    my $query="select locus,gene_start,gene_end,strand,product,func,note,ind,uniprot from annotation where
    gene_name =\'$previous_gene\' ";

```

```

my $sth = $dbh->prepare($query);
$sth->execute();

($p_locus,$p_gene_start,$p_gene_end,$p_gene_strand1,$p_product,$p_func,$p_note,$p_ind,$previous_gene_id)
= $sth->fetchrow_array();

if ($p_product eq ""){$p_product="NA";}
if ($p_func eq ""){$p_func="NA";}
if ($p_note eq ""){$p_note="NA";}

#####
# Intergenic Region Annotation #####
#####

my $query5="select length(seq),position1,position2 from intergenic where locus ='$p_locus' ";
my $sth5 = $dbh->prepare($query5);$sth5->execute();

my ($intergenic_length,$intergenic_start,$intergenic_end)=$sth5->fetchrow_array();

#####
# Hit #####
#####

open(FILE,>"$filein.".reg_genes1.txt");
#####
# Hit #####
#####

open(FILE,>"$large_fasta.".intergenic");

my $ind2=$p_ind +1;

print SITES ">$p_ind","\n",$site,"\n";
#####
# Next gene annotation #####
#####

my $query2="select gene_name,locus,gene_start,gene_end,strand,product,func,note,ind,uniprot from annotation
where ind = $ind2 ";

my $sth2 = $dbh->prepare($query2);$sth2->execute();

($next_gene,$n_locus,$n_gene_start,$n_gene_end,$n_gene_strand2,$n_product,$n_func,$n_note,$n_ind,$next_g
ene_id)= $sth2->fetchrow_array();

if ($n_product eq ""){$n_product="NA";}
if ($n_func eq ""){$n_func="NA";}
if ($n_note eq ""){$n_note="NA";}

## Hit <---- x <-----
if ($p_gene_strand1 eq "-1" and $n_gene_strand2 eq "-1"){

    my $hit_start_genome=$intergenic_start + $motif_start;
    my $hit_end_genome=$intergenic_start + $motif_end;
    my $position_TSS1= -$motif_end;
    my $position_TSS2= -$motif_start;
}

```

```

print $site,"\\t",$motif_start,"\\t",$motif_end,"\\t",$score,"\\t",$motif_strand,"\\t";
print
$previous_gene,"\\t",$p_locus,"\\t",$p_gene_start,"\\t",$p_gene_end,"\\t",$p_gene_strand1,"\\t",$p_product,"\\t",$p_func,"\\t",
$p_note,"\\t";
print "$hit_start_genome\\t$hit_end_genome\\t$position_TSS1\\t$position_TSS2\\n";
my $query6="insert into scan_hit (site,motif_start,motif_end ,score ,motif_strand ,gene ,locus ,gene_start
,gene_end, gene_strand, hit_start_genome ,hit_end_genome,tss1,tss2) values
(''$site\\'','$motif_start\\'','$motif_end\\'','$score\\'','$motif_strand\\'','$previous_gene\\'','$p_locus\\'','$p_gene_start\\'','$p_ge
ne_end\\'','$p_gene_strand1\\'','$hit_start_genome\\'','$hit_end_genome\\'','$position_TSS1\\'','$position_TSS2\\');");
my $sth = $dbh->prepare($query6);$sth->execute();
push @geneList, $previous_gene_id;
print MOTIFGENE "$id\\t$previous_gene\\n";
}

## Hit -----> x ----->

if ($p_gene_strand1 eq '1' and $n_gene_strand2 eq '1'){
    my $hit_start_genome=$intergenic_start + $motif_start;
    my $hit_end_genome=$intergenic_start + $motif_end;
    my $position_TSS1= -($intergenic_length - $motif_start);
    my $position_TSS2= -($intergenic_length - $motif_end);
    print $site,"\\t",$motif_start,"\\t",$motif_end,"\\t",$score,"\\t",$motif_strand,"\\t";
    print
$nnext_gene,"\\t",$n_locus,"\\t",$n_gene_start,"\\t",$n_gene_end,"\\t",$n_gene_strand2,"\\t",$n_product,"\\t",$n_func,"\\t",
$n_note,"\\t";
    print "$hit_start_genome\\t$hit_end_genome\\t$position_TSS1\\t$position_TSS2\\n";
    my $query7="insert into scan_hit (site,motif_start,motif_end ,score ,motif_strand ,gene ,locus ,gene_start
,gene_end, gene_strand, hit_start_genome ,hit_end_genome,tss1,tss2) values ('$site\\'','$mo
tif_start\\'','$motif_end\\'','$score\\'','$motif_strand\\'','$next_gene\\'','$n_locus\\'','$n_gene_start\\'','$n_gene_end\\'','$n_
gene_strand2\\'','$hit_start_genome\\'','$hit_end_genome\\'','$position_TSS1\\'','$position_TSS2\\')");
my $sth = $dbh->prepare($query7);$sth->execute();
#print "putative regulator of: \"$next_gene,\"\\n\" ;
push @geneList, $next_gene_id;
print MOTIFGENE "$id\\t$next_gene\\n";
}

## Hit <---- x ----->

if ($p_gene_strand1 eq '-1' and $n_gene_strand2 eq '1'){

```

```

#print "putative regulator of both genes $previous_gene\t$next_gene","\n" ;
my $position_TSS1_1= -$motif_end;
my $position_TSS2_1= -$motif_start;
my $position_TSS1_2= -($intergenic_length - $motif_start);
my $position_TSS2_2= -($intergenic_length - $motif_end);
my $hit_start_genome=$intergenic_start + $motif_start;
my $hit_end_genome=$intergenic_start + $motif_end;
print $site,"\\t",$motif_start,"\\t",$motif_end,"\\t",$score,"\\t",$motif_strand,"\\t";
print "$previous_gene-$next_gene\\t$p_locus-
$n_locus\\t$p_gene_start\\,$n_gene_start\\t$p_gene_end\\,$n_gene_end\\t$p_gene_strand1\\,$n_gene_strand2\\t$p_product\\
-\\-\\-$n_product\\t$p_func\\-\\-\\-$n_func\\t$p_note\\-\\-
$n_note\\t$hit_start_genome\\t$hit_end_genome\\t$position_TSS1_1\\,$position_TSS1_2\\t$position_TSS2_1\\,$position_
TSS2_2\\n";
print MOTIFGENE "$id\\t$previous_gene-$next_gene\\n";
my $query8="insert into scan_hit (site,motif_start,motif_end ,score ,motif_strand ,gene ,locus ,gene_start
,gene_end, gene_strand, hit_start_genome ,hit_end_genome,tss1,tss2) values
(''$site'',''$motif_start'',''$motif_end'',''$score'',''$motif_strand'',''$previous_gene$next_gene'',''$p_locus$n_locus'',''$p
_gene_start$n_gene_start'',''$p_gene_end$n_gene_end'',''$p_gene_strand1$n_gene_strand2'',''$hit_start_genome'',''$h
it_end_genome'',''$position_TSS1_1$position_TSS1_2\\'',''$position_TSS2_1$position_TSS2_2\\')";;
my $sth = $dbh->prepare($query8);$sth->execute();
push @geneList, $previous_gene_id;
push @geneList, $next_gene_id;
}
}

close(FILETEMP);

##### Write list of putative regulated genes for hypergeometric test

@geneList=uniq(@geneList);
open(FILE,>".$filein.".reg_genes1.txt");
foreach my $regulated_gene(@geneList){print (FILE $regulated_gene,"\\n");}
close(FILE);

##### hypergeometric test

$cmd = "analyze.pl ~/perl/18.E_coli_K12.goa 4339 ~/perl/gene_ontology_edit.obo $filein".$reg_genes1.txt";
system($cmd);

$HG_test_results="$filein".$reg_genes.txt.terms";

```

```
my %hash_lines=();
my $index=1;
open(FILE,"<$HG_test_results");
while      (defined(my $line=<FILE>)){
    chomp $line ;
    $hash_lines{ $index }=$line;
    $index++;
}

my $genes_comma_sep;
my @annotated_genes;
my @id_list;
while ( my ($hash_index, $hash_line) = each(%hash_lines) ){
    if ($hash_line =~ /^The\sgenes\sannotated/){
        $genes_comma_sep=$hash_lines{$hash_index+1};
        @annotated_genes= split /\n/, $genes_comma_sep;
        foreach my $g(@annotated_genes){push @id_list,$g ;}
        @id_list=uniq(@id_list);
    }
}

my @gene_names;
foreach my $id(@id_list){
    my $query="select gene_name from annotation where uniprot ='$id';";
    my $sth = $dbh->prepare($query);
    $sth->execute();
    $aa= $sth->fetchrow_array();
    push @gene_names, $aa;
}

foreach my $name(@gene_names){print "gene: $name \n";}
my $gene_names_len= scalar(@gene_names);
my @gene_names_toDel;
my @final_gene_names;
```

```

for (my $i=0; $i < $gene_names_len ; $i++){
    for (my $j=0; $j < $gene_names_len ; $j++){
        my $query="select ind,strand from annotation where gene_name =\'$gene_names[$i]\';";
        my $sth = $dbh->prepare($query);
        $sth->execute();
        (my $gene1_index,$strand1)= $sth->fetchrow_array();
        my $query="select ind,strand from annotation where gene_name =\'$gene_names[$j]\';";
        my $sth = $dbh->prepare($query);
        $sth->execute();
        (my $gene2_index,$strand2)= $sth->fetchrow_array();
        if ($gene1_index eq $gene2_index + 1 and $strand1 ne $strand2){
            print "to del: " ,$gene_names[$i],"\n";
            push @gene_names_toDel,$gene_names[$i];
        }
    }
}

my $count;
foreach my $g1(@gene_names){
    $count=0;
    foreach my $g2(@gene_names_toDel){
        if ($g1 eq $g2){$count ++;}
    }
    if ($count == 0) { push @final_gene_names,$g1 }
}
close( MOTIFGENE );

```

### ***MotifSampler output file parser***

```

#!/usr/bin/perl

use strict;

#####
# Usage: parse_motifsampler.out.pl all.out(motifsampler.out.gff) motif

```

```
# output: - (database entry)
#####
use DBI;
use lib '/home/leyteris/perl';
use Routines;
my $filein=$ARGV[0];
open (FILEIN, "<$filein") or die "open failed: $!" ;
my $database_name=$ARGV[1];
my %hash_of_lines;
my $line_index;
my @motif_indexes;
$line_index=1;
while (defined(my $line=<FILEIN>)){
    chomp $line ;
    $hash_of_lines { $line_index } = $line;
    if ($line =~ /^#id/){
        push @motif_indexes,$line_index;
    }
    $line_index++;
}
push @motif_indexes,$line_index; ### end of final motif
my $dsn = "DBI:mysql:$database_name:localhost";
my $db_user_name = 'root';
my $db_password = "";
my $dbh = DBI->connect($dsn, $db_user_name, $db_password);
my $query1="DROP TABLE if exists motif;";
my $sth = $dbh->prepare($query1);
$sth->execute();
my $query2="CREATE TABLE motif(motif_id VARCHAR(50),consensus VARCHAR(50),sequences INTEGER,
instances INTEGER,cs DOUBLE,ic DOUBLE,ll DOUBLE,ind INTEGER UNSIGNED NOT NULL
AUTO_INCREMENT,PRIMARY KEY(ind),INDEX(motif_id));";
my $sth = $dbh->prepare($query2);
```

```

$sth->execute();

my $query6="DROP TABLE if exists hits;";

my $sth = $dbh->prepare($query6);

$sth->execute();

my $query3="CREATE TABLE hits(motif_id VARCHAR(50),motif_ind INTEGER,gene_name
VARCHAR(50),hit_start INTEGER,hit_end INTEGER,score DOUBLE,strand VARCHAR(10),position1
INTEGER,position2 INTEGER,site VARCHAR(250),ind INTEGER UNSIGNED NOT NULL
AUTO_INCREMENT,PRIMARY KEY(ind),INDEX(motif_id,gene_name));";

my $sth = $dbh->prepare($query3);

$sth->execute();

my $times=scalar(@motif_indexes)-1;

for (my $i=0; $i<=$times; $i ++){

unless ($i == $times){

    my @indexes=$motif_indexes[$i]..$motif_indexes[$i+1]-1;

    my $id_line=$hash_of_lines{$indexes[0]};

    my $id=(split ":",((split "\t",$id_line)[0]))[1] ;

    my $consensus=(split ":",((split "\t",$id_line)[1]))[1] ;

    my $sequences=(split ":",((split "\t",$id_line)[2]))[1] ;

    my $instances=(split ":",((split "\t",$id_line)[3]))[1] ;

    my $cs=(split ":",((split "\t",$id_line)[4]))[1] ;

    my $ic=(split ":",((split "\t",$id_line)[5]))[1] ;

    my $ll=(split ":",((split "\t",$id_line)[6]))[1] ;

    my $query4="insert into motif (motif_id,consensus,sequences,instances,cs,ic,ll) values
('.$id.', '$consensus', '$sequences', '$instances', '$cs', '$ic', '$ll')";;

    my $sth = $dbh->prepare($query4);

    $sth->execute();

    my $indexes_len=scalar(@indexes)-1;

    my $query_motif_ind="select ind from motif where motif_id='.$id.';";

    my $sth = $dbh->prepare($query_motif_ind);

    $sth->execute();

    my $motif_ind=$sth->fetchrow_array();

    for (my $k=1; $k<= $indexes_len; $k ++){

        my $hit_line= $hash_of_lines{$indexes[$k]};


```

```
my $gene_name=(split "\t",$hit_line)[0];
my $motif_start=(split "\t",$hit_line)[3];
my $motif_end=(split "\t",$hit_line)[4];
my $score=(split "\t",$hit_line)[5];
my $strand=(split "\t",$hit_line)[6];
my $site=(split "'",((split "\t",$hit_line)[8]))[3];
my $len_promoter=Routines::len_promoter($gene_name,$database_name) ;
my $motif_position1=$len_promoter-$motif_start;
my $motif_position2=$len_promoter-$motif_end;
if ($gene_name !~ m/^#/){
    my $query5="insert into hits
(motif_id,motif_ind,gene_name,hit_start,hit_end,score,strand,position1,position2,site) values
(''$id'\''吱$motif_ind\''吱$gene_name\''吱$motif_start\''吱$motif_end\''吱$score\''吱$strand\''吱$motif_position1\''吱$motif_posi
tion2\''吱$site\');";
    my $sth = $dbh->prepare($query5);
    $sth->execute();
}
}
}
```



## Παράρτημα Β

### Πειραματική επιβεβαίωση στόχων του παράγοντα AtoC

#### Συνθήκες καλλιέργειας

Τα κύτταρα αναπτύχθηκαν στην εκθετική φάση στους 37°C σε 10 ml μέσου M9 με προσθήκη 0.1 mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>, 0.4% (w/v) glucose, 1 µg thiamine/ml και 80 µg proline/ml παρουσία και απουσία 10 mM ακετοξικό οξύ. Κύτταρα E. coli BL21[DE3] με πλασμίδιο pHis10-AtoC αναπτύχθηκαν στους 37°C σε ελάχιστο μέσο που περιείχε 100 µg/ml ampicillin. Η επαγωγή της έκφρασης ανασυνδυασμένου AtoC επετεύχθη με την πρόσθεση IPTG (0.25 mM) στις καλλιέργειες, όταν η πυκνότητα OD<sub>600</sub> έφτασε στο 0.25 για 4 ώρες. Το ακετοξικό οξύ που είναι γνωστός επαγωγέας του AtoC προστέθηκε επίσης κατά τη διάρκεια της φάσης επαγωγής της καλλιέργειας, σε συγκέντρωση 10 mM.

#### Αναλύσεις ανοσοκατακρήμνισης χρωματίνης (ChIP assays)

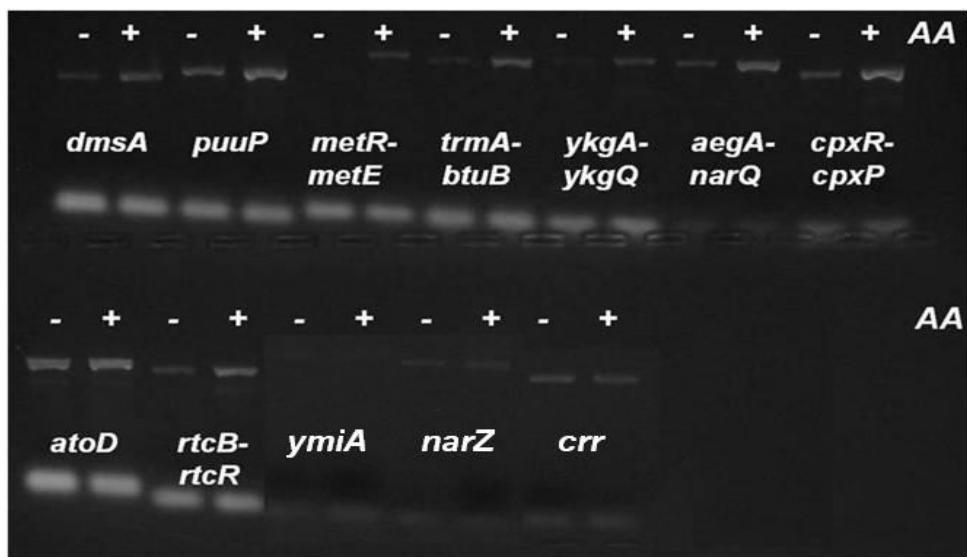
Οι αναλύσεις ChIP εκτελέστηκαν με τη χρησιμοποίηση χρωματίνης από στέλεχος E. coli BL21 [DE3] που υπερεκφράζει His-tagged AtoC, παρουσία και απουσία ακετοξικού οξέος. Χρησιμοποιήθηκε κιτ Magna ChIP™ A Millipore. *In vivo* cross-linking των νουκλεοπρωτεϊνών πραγματοποιήθηκε με την πρόσθεση φορμαλδεΰδης (1%) κατευθείαν στις καλλιέργειες (των 10 ml) όταν η πυκνότητα OD<sub>600</sub> έφτασε στο 0.6. Οι αντιδράσεις cross-linking διεκόπηκαν 20 λεπτά αργότερα με την προσθήκη γλυκίνης και τα κύτταρα διαλύθηκαν με φυγοκέντρηση και ξεπλύθηκαν με παγωμένο PBS και διάλυμα αναστολέα πρωτεασών. Οι κυτταρικές λύσεις εκτελέστηκαν σύμφωνα με το εγχειρίδιο του κιτ και το κυτταρικό DNA διασπάστηκε με υπερήχους σε τμήματα με μέσο όρο μεγέθους 500 με 1000 bps. Τα κυτταρικά υπολείμματα απομακρύνθηκαν με φυγοκέντρηση και το υπερκείμενο διάλυμα χρησιμοποιήθηκε ως είσοδος χρωματίνης στις αντιδράσεις ανοσοκατακρήμνισης, μετά από ένα αρχικό στάδιο επώασης με μαγνητικά σφαιρίδια πρωτεΐνης A για να αφαιρεθούν οι μη ειδικά συνδεδεμένες πρωτεΐνες. Οι αντιδράσεις ChIP πραγματοποιήθηκαν με την προσθήκη πολυκλωνικού αντισώματος His-probe™ Santa Cruz rabbit και μαγνητικά σφαιρίδια πρωτεΐνης A στο υπερκείμενο (supernatant) και αφήνοντάς το όλη τη νύχτα σε περιστροφή στους 4°C. Το DNA ελήφθη από τα ανοσοκατακρημνισμένα σύμπλοκα DNA/πρωτεΐνών και η παρουσία των επαγωγικών αλληλουχιών-στόχων ανιχνεύθηκε με PCR. Πραγματοποιήθηκαν παράλληλες ψευδο-κατακρημνίσεις χωρίς προσθήκη αντισώματος, ως αρνητικός έλεγχος για κάθε αντίδραση.

### **Electrophoretic Mobility Shift Assays**

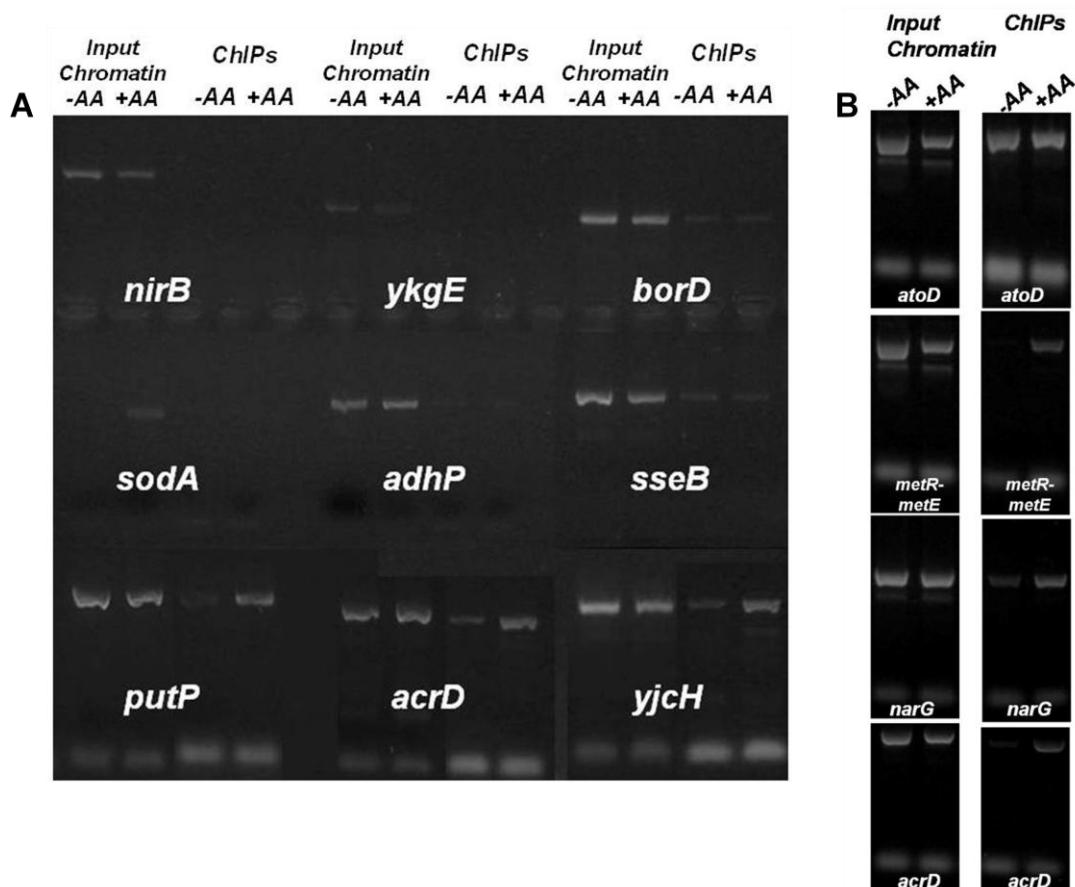
Με τη μέθοδο ηλεκτροφορητικής μετατόπισης κινητικότητας διαπιστώθηκε ο σχηματισμός συμπλόκων πρωτεΐνων/DNA. Η πρωτεΐνη His<sub>10</sub>-AtoC συνδέθηκε σε ολιγονουκλεοτίδια διπλής έλικας που αντιστοιχούσαν στους αντιπροσωπευτικούς επαγωγείς-στόχους *atoD*, *metR-metE* και *acrD*, καθώς και στο γονίδιο *narG*. Αντίστοιχες ποσότητες από κάθε ένα από τα τέσσερα σωματίδια συνδυάστηκε με αυξημένες ποσότητες His<sub>10</sub>-AtoC (1, 125, 5 και 10 µg) σε 25 µl ρυθμιστικού διαλύματος σύνδεσης, 10 mM Tris (pH 7.6), 10 mM MgCl<sub>2</sub>, 50 mM NaCl, 5 % (v/v) glycerol και επωάστηκε στον πάγο για 25 min, μετά από 5 min προ-επώασης σε θερμοκρασία δωματίου, πριν από την πρόσθεση της πρωτεΐνης. Δείγματα χωρίς πρωτεΐνη αναμείχθηκαν με 10× Gel loading buffer (Takara) και εφαρμόσθηκαν σε 100 V prerun polyacrylamide gel που περιείχε 0.5× TBE buffer, προετοιμασμένο με 6.0 % (w/v) polyacrylamide (ratio of acrylamide to bisacrylamide, 29:1). Μετά από ηλεκτροφόρηση (120 V, 120 min), τα gels χρωματίστηκαν με Gel-Red dye (Biotium) και μετά με Coomassie blue. Για τα πειράματα DNA-binding, 10 ng (3 nM) of χρησιμοποιήθηκε biotinylated *atoD* probe σε ρυθμιστικό διάλυμα αποτελούμενο από 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 1 mM EDTA, 4 mM DTT and 5% (v/v) glycerol. Ανταγωνιστικό non-biotinylated *atoD* προστέθηκε στην ίδια συγκέντρωση (μέχρι 80 nM το καθένα) με τα *metRmetE*, *acrD* και *narG*. Κάθε αντίδραση πραγματοποιήθηκε σε τελικό όγκο 20 µl, αποτελούμενο από 0.2 mg/ml BSA και 2 µg sonicated calf thymus competitive DNA. Όλα τα συστατικά προστέθηκαν και προ-επωάσθηκαν για 5 min σε θερμοκρασία δωματίου, πριν από την πρόσθεση 0.35 µM (2 µg) πρωτεΐνης και μεταφέρθηκαν σε πάγο για 30 min. Στο τέλος των αντιδράσεων τα δείγματα φορτώθηκαν σε 4% (w/v) acrylamide gel σε 0.25× TBE buffer, στα 120 V, με 30 min prerun στα 100 V. Το DNA μεταφέρθηκε σε μεμβράνες Biodyne B (Pall) και ανιχνεύθηκε με τη χρήση Phototope-Star detection kit (NewEngland Biolabs).

### **Αποτελέσματα πειραματικής επιβεβαίωσης**

Τα απολέσματα συνοψίζονται στις εικόνες Γ.1, Γ2 και Γ.3.



Εικόνα Γ.1. In vivo σύνδεση ανασυνδυασμένου DNA στους στόχους του μοτίβου 1 που προβλέφθηκαν από την υπολογιστική διαδικασία, σε μεσογονιδιακές περιοχές. Αποτελέσματα της ανάλυσης ChIP σε ηλεκτροφόρηση σε τζελ αγαρόζης (2% w/v) από PCR με primers που αντιστοιχούν στο μοτίβο 1 (Πίνακες 2.1 και 2.5). Όλες οι αναλύσεις πραγματοποιήθηκαν με επαγωγή (+AA) ή χωρίς επαγωγή (-AA) από ακετοξικό οξύ, που είναι ο κύριος επαγωγέας του AtoSC, με 25 κύκλους PCR.



Εικόνα Γ.2. Παρακολούθηση της επίδρασης του ακετοξικού οξέος (AA) στην in vivo σύνδεση του AtoC σε στόχους που προβλέφθηκαν από την υπολογιστική διαδικασία. A) Ανάλυση ChIP

στους στόχους των μοτίβων 2, 3 και 4 (Πίνακας 2.5) που έδωσαν θετικό σήμα στο DNA του γονιδιώματος. -AA: χωρίς acetoacetate. +AA: με επαγωγή ακετοξικού οξέος.

B) Τρεις επιλεγμένοι στόχοι, μαζί με την αρχική γνωστή θέση σύνδεσης στον επαγωγέα atoDAEB



Εικόνα Γ.3. Αποτελέσματα ChIP για τους στόχους που βρίσκονται σε αντίστοιχες ενδογονιδιακές περιοχές (gene-encoding).

## Παράρτημα Γ

### Δημοσιεύσεις διατριβής σε διεθνή περιοδικά

- An in silico compartmentalized metabolic model of Brassica napus enables the systemic study of regulatory aspects of plant central metabolism. **E. Pilalis, A. Chatzioannou, B. Thomasset, F. Kolisis** **Biotechnol Bioeng.** **2011 Jul;108(7):1673-82.** doi: 10.1002/bit.23107. Epub 2011 Mar 11.
- Escherichia coli Genome-Wide Promoter Analysis: Identification of Additional AtoC Binding Target Elements. **Pilalis E, Chatzioannou AA, Grigoroudis AI, Panagiotidis CA, Kolisis FN, Kyriakidis DA.** **BMC Genomics.** **2011 May 13;12(1):238.** [Epub ahead of print]

### Άλλες δημοσιεύσεις του Ελευθέριου Πιλάλη

- Microarray analysis of survival pathways in human PC-3 prostate cancer cells. Tenta, R., Katopodis, H., Chatzioannou, A., **Pilalis, E.**, Calvo, E., Van Luu-The, Labrie, F., Kolisis, Koutsilieris, M. **Cancer Genomics and Proteomics** **4 (4), 2007, pp. 309-317.**

### Ανακοινώσεις σε διεθνή συνέδρια

- A comparative microarray data analysis of the gene expression induced by survival factors in PC-3 human prostate cancer cells. Tenta R, Chatzioannou A, **Pilalis E**, Calvo E, Labrie F, Kolisis F, Koutsilieris M. Poster presentation, **3<sup>rd</sup> EMBL Biennial Symposium: From Functional Genomics to Systems Biology**, Heilderberg, 2006
- The role of survival factors in PC-3 human prostate cancer cells by a comparative microarray data analysis of the gene expression. R. Tenta, A. Chatzioannou, **E. Pilalis**, E. Calvo, F. Labrie, F. Kolisis, M. Koutsilieris. **Proceedings of the 12th International conference on Hormonal Steroids & Hormones and Cancer, October 2007, Athens, Greece.**
- In silico modeling of metabolic pathways in oilseed plants. A.Idrissi Taghki , **E. Pilalis**, A. Chatzioannou, A. Friboulet, B. Thomasset, F. Kolisis. **5th /GERLI Lipidomics Congress, October 2008, Compiegne, France.**
- Escherichia coli Genome-Wide Promoter Analysis: Search for Potential AtoC Binding Target Elements. **E.Pilalis, A.Chatziioannou, A.I. Grigoroudis, C.A. Panagiotidis, F.N.Kolisis, D.A. Kyriakidis.** **Proceedings of the 33rd FEBS Congress & 11th IUBMB Conference 2008, June 28-July 3, Athens, Greece.**
- GRISSOM a web based Grid portal and repository for interpretation and storage of DNA microarray experiments. A. Chatzioannou, I. Kanaris, C. Doukas, P. Moulos, **E. Pilalis, I. Maglogiannis, F.N. Kolisis,** **Proceedings of the 4th**

**EGEE User Forum/OGF25 & OGF-Europe's 2nd International Event in Catania, Sicily, Italy, 2-6 March 2009.**

- A compartmented in silico model of Rapeseed central metabolism. **Pilalis E.**, Chatzioannou A., Kolisis F.N. **Plant Bioinformatics, Systems and Synthetic Biology Summer School, 27-31 July 2009 - University of Nottingham, UK, European Science Foundation. (Poster, 2nd conference award)**
- Identification of additional Escherichia coli AtoC binding target elements gives new insight to bacterial regulatory networks: molecular modeling of AtoC-DNA binding domain amino acids. Asterios I. Grigoroudis, Christos A. Panagiotidis, **Eleftherios Pilalis**, Aristotelis Chatzioannou, Fragiskos Kolisis, Papadopoulos Georgios, and Dimitrios A. Kyriakidis. **Amino Acids 37 (Suppl 1), pp. S102-S103, 2009.** 11th International Congress on Amino Acids Peptides and Proteins, Vienna, Austria , August 2009.
- GRISSOM web based Grid portal: Exploiting the power of Grid infrastructure for the interpretation and storage of DNA microarray experiments. A. Chatzioannou, I. Kanaris, I. Maglogiannis, C. Doukas, P. Moulos, E. Pilalis, F.N. Kolisis, **Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB), November 5-7, 2009, Larnaca, Cyprus.**