



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**  
**ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ**  
**ΥΛΙΚΩΝ**

**Ανάπτυξη Τεχνικών Επεξεργασίας Δεδομένων στην Αντικαρκινική Θεραπεία  
και τη Μεταβολομική**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

της

**ΑΡΙΣΤΕΑΣ ΙΩΑΝΝΑΣ Γ. ΦΙΛΑΝΤΙΣΗ**

Διπλωματούχου Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών  
(Εθνικό Μετσόβιο Πολυτεχνείο)

ΑΘΗΝΑ, ΜΑΡΤΙΟΣ 2018





**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**  
**ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ**  
**ΥΛΙΚΩΝ**

**Ανάπτυξη Τεχνικών Επεξεργασίας Δεδομένων στην Αντικαρκινική Θεραπεία  
και τη Μεταβολομική**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

της

**ΑΡΙΣΤΕΑΣ ΙΩΑΝΝΑΣ Γ. ΦΙΛΑΝΤΙΣΗ**

Διπλωματούχου Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών  
(Εθνικό Μετσόβιο Πολυτεχνείο)

ΑΘΗΝΑ, ΜΑΡΤΙΟΣ 2018





**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**  
**ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ**  
**ΥΛΙΚΩΝ**

**Ανάπτυξη Τεχνικών Επεξεργασίας Δεδομένων στην Αντικαρκινική Θεραπεία  
και τη Μεταβολομική**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

της

**ΑΡΙΣΤΕΑΣ ΙΩΑΝΝΑΣ Γ. ΦΙΛΑΝΤΙΣΗ**

Διπλωματούχου Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών

(Εθνικό Μετσόβιο Πολυτεχνείο)

Συμβουλευτική Επιτροπή:

Αν. Καθ. Γεώργιος Ματσόπουλος

Καθ. Νικόλαος Ουζούνογλου

Καθ. Δημήτριος Κουτσούρης

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 2<sup>η</sup> Μαρτίου 2018.

.....  
Γ. Ματσόπουλος  
Αν. Καθηγητής Ε.Μ.Π.

.....  
Ν. Ουζούνογλου  
Καθηγητής Ε.Μ.Π.

.....  
Δ. Κουτσούρης  
Καθηγητής Ε.Μ.Π.

.....  
Κ. Πολιτόπουλος  
Αν. Καθηγητής Ε.Μ.Π.

.....  
Δ. Κάβουρας  
Καθηγητής Τ.Ε.Ι. Αθήνας

.....  
Π. Ζουμπουλάκης  
Κύριος Ερευνητής Ε.Ι.Ε.

.....  
Δ. Βλαχάκης  
Επ. Καθηγητής Γ.Π.Α.

ΑΘΗΝΑ, ΜΑΡΤΙΟΣ 2018

Copyright © Αριστέας Ιωάννας Γ. Φιλντίση, 2018.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Το αντικείμενο της παρούσας διδακτορικής διατριβής είναι η διερεύνηση, η ανάπτυξη και η παρουσίαση τεχνικών επεξεργασίας αφ' ενός βιολογικών δεδομένων, με επίκεντρο τη στοχευμένη αντικαρκινική θεραπεία, αφ' ετέρου χημικών και φασματοσκοπικών δεδομένων, με επίκεντρο τη μεταβολομική. Αρχικά πραγματοποιείται μία εισαγωγή σε θέματα βιολογίας, ιατρικής και βιοπληροφορικής που αφορούν την αντικαρκινική θεραπεία. Στη συνέχεια παρουσιάζεται το πρώτο σκέλος της έρευνας, που επικεντρώθηκε στην ανάπτυξη μίας υπολογιστικής τεχνικής σχεδιασμού Συμπλεγμάτων Αντισωμάτων-Φαρμάκων, μίας κατηγορίας στοχευμένων αντικαρκινικών θεραπευτικών παραγόντων. Σκοπός της τεχνικής αυτής είναι η συμβολή στην αντιμετώπιση της δυσκολίας υπολογιστικού σχεδιασμού Συμπλεγμάτων Αντισωμάτων-Φαρμάκων με τις καθιερωμένες μεθόδους *in silico* φαρμακευτικού σχεδιασμού. Η τεχνική αξιοποιεί έναν αφινικό μετασχηματισμό, υλοποιήθηκε για δύο τρόπους σύζευξης του φαρμάκου στο αντίσωμα, και εφαρμόστηκε σε πραγματικά βιολογικά και φαρμακευτικά δεδομένα. Ακολούθως, επιχειρείται μία εισαγωγή σε θέματα μεταβολομικής με χρήση φασματοσκοπίας Πυρηνικού Μαγνητικού Συντονισμού και παρουσιάζεται το δεύτερο σκέλος της έρευνας, το οποίο επικεντρώθηκε στην ανάπτυξη ενός υπολογιστικού μεθοδολογικού σχήματος με αντικείμενο την αυτόματη ταυτοποίηση μεταβολιτών σε μονοδιάστατα φάσματα Πυρηνικού Μαγνητικού Συντονισμού πρωτονίου. Γίνεται αναφορά των μεθόδων προεπεξεργασίας και μείωσης δεδομένων που εφαρμόστηκαν και αναλύονται τα συστατικά τμήματα της μεθοδολογίας αναγνώρισης μεταβολιτών που αναπτύχθηκε. Παρουσιάζεται η διεργασία της αναζήτησης μεταβολιτών και παρατίθενται οι συναρτήσεις βαθμολόγησης των υποψήφιων συνόλων φασματικών κορυφών για τις διαφορετικές κατηγορίες κορυφών μεταβολιτών. Αναλύεται η στρατηγική επιλογής του καλύτερου συνδυασμού φασματικών κορυφών για κάθε μεταβολίτη και γίνεται αναφορά στην προτεραιοποίηση των υποψήφιων μεταβολιτών που υιοθετήθηκε στη μεθοδολογία. Η μεθοδολογία εφαρμόστηκε σε τεχνητά και πραγματικά φάσματα από διάφορες μελέτες, και συγκρίθηκε ως προς τα αποτελέσματά της με άλλες αυτόματες υπολογιστικές τεχνικές αναγνώρισης μεταβολιτών.

### Λέξεις-κλειδιά

Συμπλέγματα Αντισωμάτων-Φαρμάκων, Αντικαρκινική θεραπεία, Φαρμακευτικός σχεδιασμός, Πυρηνικός Μαγνητικός Συντονισμός, Ταυτοποίηση μεταβολιτών





## Abstract

---

The primary purpose of this doctoral thesis is the research, development and presentation of processing techniques for biological data, focused on targeted anticancer therapy, as well as chemical and spectroscopic data, focused on metabolomics. Firstly, an introduction in aspects of biology, medicine and bioinformatics that are related to anticancer therapy is performed. Subsequently, the first part of the research is presented, which is focused on the development of a technique for the computational design of Antibody-Drug conjugates, a category of targeted anticancer therapeutic agents. The purpose of this technique is to contribute to the resolution of the difficulty there is in designing Antibody-Drug Conjugates with the established *in silico* drug design techniques. The method uses an affine transformation, it has been implemented for two conjugation ways and it has been applied on real biological and pharmaceutical data. In addition, an introduction in Nuclear Magnetic Resonance based metabolomics is performed, followed by the analysis of the second part of the research, which is focused on the development of a computational methodology for the automated identification of metabolites from one-dimensional proton Nuclear Magnetic Resonance spectra. The presentation of the preprocessing and data reduction methods that have been used is followed by the analysis of the separate main parts of the metabolite identification methodology. The metabolite screening process is analysed, including the scoring functions used for the assessment of the candidate spectral peak sets for each multiplet category. The strategy for the selection of the optimal spectral peak set for each candidate metabolite that has been implemented, as well as the prioritization of the candidate metabolites, are explained. The performance of the methodology has been evaluated on artificial and real spectra from various studies and has been compared to other automated metabolite identification methods.

### Keywords

Antibody-Drug Conjugates, Anticancer therapy, Drug design, Nuclear Magnetic Resonance, Metabolite identification



## Ευχαριστίες

---

Με την ολοκλήρωση της διδακτορικής διατριβής μου θα ήθελα να ευχαριστήσω θερμά όλους όσους συνέβαλαν στη συγγραφή της.

Ευχαριστώ ιδιαίτερα τον επιβλέποντα καθηγητή μου κ. Ματσόπουλο, για τη γενικότερη επιστημονική καθοδήγηση και για την ικανότητά του να διακρίνει πάντα την ουσία στα ζητήματα που προέκυπταν. Επίσης, για την αισιοδοξία του και την εμπιστοσύνη που μου έδειξε. Είναι ιδιαίτερη τιμή και χαρά για μένα η συνεργασία μαζί του.

Ευχαριστώ επίσης ιδιαίτερα τους κ.κ. Καθηγητές κ. Ουζούνογλου και κ. Κουτσούρη.

Ακόμη οφείλω θερμές ευχαριστίες σε όλους όσους συνεργάστηκα στα πλαίσια αυτού του πονήματος: την κ. Κοσσίδα και τον κ. Βλαχάκη από το Ίδρυμα Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών, τον κ. Ζουμπουλάκη και τον κ. Φωτάκη από το Εθνικό Ίδρυμα Ερευνών, τον κ. Ασβεστά και τον κ. Κάβουρα από το Τεχνολογικό Εκπαιδευτικό Ίδρυμα Αθήνας. Θα θυμάμαι πάντα με ευγνωμοσύνη τις πολύτιμες συμβουλές και την καλοπροαίρεσή τους.

Επίσης, οφείλω να ευχαριστήσω το Ίδρυμα Κρατικών Υποτροφιών για την οικονομική στήριξη που μου παρείχε, αφού η παρούσα διδακτορική διατριβή χορηγήθηκε από το πρόγραμμα Υποτροφιών Αριστείας Ι.Κ.Υ. Μεταπτυχιακών σπουδών στην Ελλάδα – Πρόγραμμα Siemens.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου, για την απεριόριστη στήριξη, κατανόηση και δύναμη που μου προσφέρουν όλα αυτά τα χρόνια.



## Πίνακας Περιεχομένων

---

Περίληψη.....	7
Abstract .....	9
Κεφάλαιο 1. Εισαγωγή.....	15
1.1. Σκοπός διδακτορικής διατριβής .....	15
1.2. Περιεχόμενα διδακτορικής διατριβής .....	15
1.3. Αναφορές .....	17
Κεφάλαιο 2. Εισαγωγή στην Αντικαρκινική Θεραπεία και στις Εφαρμογές Βιοπληροφορικής .....	19
2.1. Ανοσολογία.....	19
2.1.1. Ανοσοποιητικό σύστημα.....	19
2.1.2. Αντισώματα .....	20
2.1.3. Τεχνικές καθορισμού τρισδιάστατης δομής πρωτεϊνών .....	21
2.2. Αντικαρκινική θεραπεία .....	22
2.2.1. Καρκίνος .....	22
2.2.2. Αντικαρκινικές θεραπείες .....	22
2.2.3. Συμπλέγματα Αντισωμάτων-Φαρμάκων .....	24
2.3. Στοιχεία Βιοπληροφορικής .....	30
2.3.1. Υπολογιστική τυποποίηση δεδομένων τρισδιάστατης δομής πρωτεϊνών .....	30
2.3.2. Βάσεις Δεδομένων .....	30
2.3.3. <i>In silico</i> σχεδιασμός φαρμάκων .....	31
2.4. Αναφορές .....	33
Κεφάλαιο 3. Υπολογιστικός Σχεδιασμός Συμπλεγμάτων Αντισωμάτων-Φαρμάκων .....	41
3.1. Εισαγωγή .....	41
3.2. Μεθοδολογία.....	42
3.2.1 Περιγραφή δεδομένων .....	42
3.2.2 Περιγραφή μεθοδολογίας.....	43
3.3. Αποτελέσματα σύνθεσης συμπλεγμάτων .....	48
3.3.1 Αποτελέσματα σύζευξης σε λυσίνες .....	48
3.3.2 Αποτελέσματα σύζευξης σε τροποποιημένες κυστεΐνες .....	51
3.4. Αναφορές .....	54
Κεφάλαιο 4. Εισαγωγή στη Μεταβολομική και τη Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού .....	57
4.1. Μεταβολομική .....	57
4.1.1. Έννοιες μεταβολομικής .....	57
4.1.2. Πλεονεκτήματα μεταβολομικής.....	57
4.1.3. Αναλυτικές τεχνικές.....	59
4.1.4. Εφαρμογές μεταβολομικής .....	60

4.1.4.1. Διάγνωση.....	60
4.1.4.2. Φαρμακολογία.....	62
4.2. Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού .....	66
4.2.1 Εισαγωγή.....	66
4.2.2 Βασικές αρχές φυσικής .....	67
4.2.3 Σήμα Πυρηνικού Μαγνητικού Συντονισμού .....	70
4.2.3.1. Σήμα στο πεδίο χρόνου .....	70
4.2.3.2. Σήμα στο πεδίο συχνότητας .....	70
4.2.3.3. Παράμετροι σήματος.....	74
4.2.4. Φασματογράφος Πυρηνικού Μαγνητικού Συντονισμού .....	77
4.2.5. Πειραματική διαδικασία .....	80
4.2.6. Επεξεργασία σήματος Πυρηνικού Μαγνητικού Συντονισμού .....	83
4.2.7. Περιστολή δεδομένων φάσματος Πυρηνικού Μαγνητικού Συντονισμού .....	86
4.2.8. Αναγνώριση και Ποσοτικοποίηση Μεταβολιτών .....	88
4.3. Αναφορές.....	91
Κεφάλαιο 5. Αυτόματη αναγνώριση μεταβολιτών σε μονοδιάστατα φάσματα <sup>1</sup> H-NMR .	101
5.1. Εισαγωγή .....	101
5.2. Μεθοδολογία.....	102
5.2.1 Περιγραφή δεδομένων .....	102
5.2.1.1. Φάσματα.....	102
5.2.1.2. Μεταβολίτες .....	102
5.2.2 Περιγραφή μεθοδολογίας.....	103
5.2.2.1. Προεπεξεργασία .....	103
5.2.2.2. Μείωση δεδομένων .....	104
5.2.2.3. Αναζήτηση μεταβολιτών .....	106
5.3. Αποτελέσματα .....	123
5.3.1 Απόδοση μεθοδολογίας .....	123
5.3.2 Συγκριτική αξιολόγηση .....	124
5.4. Σχολιασμός .....	127
5.5. Αναφορές .....	128
Κεφάλαιο 6. Επίλογος και Μελλοντικές Επεκτάσεις .....	131
Παράρτημα .....	133
Δημοσιεύσεις.....	147

## Κεφάλαιο 1. Εισαγωγή

---

### 1.1. Σκοπός διδακτορικής διατριβής

Ο συγκεκριασμός της πληροφορικής με τις επιστήμες υγείας έχει δώσει νέες προοπτικές σε αυτές, κάνοντας δυνατή την επεξεργασία πολύπλοκων δεδομένων, και κατ' επέκταση προάγοντας τη γρηγορότερη διάγνωση και την αποτελεσματικότερη θεραπεία. Ένα από τα πεδία στα οποία η συμβολή της πληροφορικής έχει αποκτήσει ιδιαίτερη σημασία είναι η μοριακή μοντελοποίηση, και ειδικότερα ο *in silico* φαρμακευτικός σχεδιασμός. Με δεδομένο ότι η διαδικασία ανάπτυξης ενός νέου φαρμάκου είναι ιδιαίτερα πολύπλοκη και απαιτητική σε χρόνο και πόρους, οι *in silico* τεχνικές φαρμακευτικού σχεδιασμού μπορούν να έχουν ιδιαίτερα χρήσιμο ρόλο, μειώνοντας σημαντικά το κόστος και τη διάρκεια ανακάλυψης ενός νέου φαρμάκου.

Μία υποσχόμενη κατηγορία αντικαρκινικών θεραπευτικών παραγόντων, τα Συμπλέγματα Αντισωμάτων-Φαρμάκων (Antibody-Drug Conjugates, ADCs) είναι σύνθετα μόρια που αποτελούνται από τρία στοιχεία, ένα αντίσωμα, ένα συνδετικό μόριο κι ένα κυτταροτοξικό μόριο, σε αντιδιαστολή με τους πιο συμβατικούς αντικαρκινικούς θεραπευτικούς παράγοντες που χρησιμοποιούνται στην χημειοθεραπεία, οι οποίοι αποτελούνται από ένα κυτταροτοξικό μόριο. Η σύνθετη δομή και το πλήθος των συστατικών μορίων των ADCs καθιστά ιδιαίτερα δύσκολο τον υπολογιστικό σχεδιασμό τους με τις καθιερωμένες *in silico* τεχνικές φαρμακευτικού σχεδιασμού.

Αρχικός στόχος της παρούσας διδακτορικής διατριβής ήταν η ανάπτυξη μίας μεθόδου για τον υπολογιστικό σχεδιασμό Συμπλεγμάτων Αντισωμάτων-Φαρμάκων, συμβάλλοντας στην αντιμετώπιση της ανάγκης για υπολογιστική μοντελοποίηση αυτής της κατηγορίας θεραπευτικών παραγόντων. Η υπολογιστική μέθοδος που αναπτύχθηκε επικεντρώθηκε σε δύο τρόπους σύζευξης του φαρμάκου στο αντίσωμα, σε λυσίνες και τροποποιημένες κυστεΐνες. Βασικά σημεία της μεθόδου είναι ο ορισμός αξόνων για το φάρμακο, το συνδέτη και το αμινοξύ στο οποίο θα πραγματοποιηθεί η σύζευξη και η εφαρμογή του αφινικού μετασχηματισμού για την περιστροφή και μετατόπιση του κυτταροτοξικού και του συνδετικού μορίου. Για την επίτευξη του στόχου της μεθόδου αναπτύχθηκαν συναρτήσεις διαχείρισης των δεδομένων τρισδιάστατης δομής των μορίων.

Η μεταβολομική είναι ένα σχετικά νέο, εξελισσόμενο πεδίο των επιστημών υγείας που έχει ως αντικείμενο τη μελέτη των μεταβολιτών ενός βιολογικού συστήματος και έχει σημαντικές εφαρμογές στο πεδίο της ιατρικής και της φαρμακευτικής ανάπτυξης. Η φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού είναι μία από τις πιο καθιερωμένες τεχνικές για την ανάλυση δειγμάτων ως προς το μεταβολικό περιεχόμενό τους. Μία από τις προκλήσεις που συναντώνται στις μελέτες μεταβολομικής είναι η αναγνώριση μεταβολιτών από τα μονοδιάστατα φάσματα Πυρηνικού Μαγνητικού Συντονισμού, αφού η προεπεξεργασία των φασμάτων και η ταυτοποίηση των μεταβολιτών που περιέχουν είναι διαδικασία που απαιτεί εξειδικευμένη γνώση και χρόνο. Φαινόμενα όπως η αλληλεπικάλυψη φασματικών κορυφών που ανήκουν σε διαφορετικούς μεταβολίτες, ή η μετατόπιση των φασματικών κορυφών στον άξονα συχνοτήτων δυσχεραίνουν την εξαγωγή χρήσιμης μεταβολικής πληροφορίας από ένα φάσμα. Η πληροφορική έχει αυξανόμενα σημαντικό ρόλο στο χώρο της μεταβολομικής, με την υπολογιστικά υποβοηθούμενη ταυτοποίηση και

ποσοτικοποίηση μεταβολιτών να έχει γίνει το αντικείμενο αρκετών ερευνητικών προσπαθειών.

Στόχος της παρούσας διδακτορικής διατριβής στάθηκε η συμβολή στην επίλυση του προβλήματος αυτού με την ανάπτυξη μίας υπολογιστικής μεθοδολογίας αυτόματης αναγνώρισης μεταβολιτών από μονοδιάστατα φάσματα Πυρηνικού Μαγνητικού Συντονισμού πρωτονίου. Στα πλαίσια της μεθοδολογίας αυτής, αρχικά γίνεται εφαρμογή μεθόδων προεπεξεργασίας και περιστολής δεδομένων. Η αναγνώριση μεταβολιτών διαμορφώθηκε ώστε να διακρίνεται σε δύο βασικά στάδια. Το πρώτο στάδιο επικεντρώθηκε στη σάρωση του φάσματος με στόχο την επιλογή συνόλων υποψήφιων φασματικών κορυφών για κάθε απλή ή πολλαπλή κορυφή ενός μεταβολίτη με τη βοήθεια κατάλληλα διαμορφωμένων συναρτήσεων βαθμολόγησης. Το δεύτερο στάδιο επικεντρώθηκε στην επιλογή του βέλτιστου συνδυασμού κορυφών για κάθε μεταβολίτη, μέσα από μία διαδικασία ελέγχου κάθε συνδυασμού ως προς συγκεκριμένα κριτήρια και μία διαδικασία βαθμολόγησης των συνδυασμών που τα ικανοποιούν.

## **1.2. Περιεχόμενα διδακτορικής διατριβής**

Στα πλαίσια της παρούσας διδακτορικής διατριβής αναπτύχθηκαν καινοτόμες τεχνικές επεξεργασίας βιολογικών και μεταβολομικών δεδομένων. Το πρώτο σκέλος της έρευνας επικεντρώθηκε στην ανάπτυξη μίας υπολογιστικής μεθόδου σχεδιασμού Συμπλεγμάτων Αντισωμάτων-Φαρμάκων, μίας κατηγορίας αντικαρκινικών θεραπευτικών παραγόντων. Η μέθοδος αυτή αποσκοπεί στην αντιμετώπιση της δυσκολίας σχεδιασμού αυτής της κατηγορίας θεραπευτικών παραγόντων με τις καθιερωμένες υπολογιστικές τεχνικές φαρμακευτικού σχεδιασμού, υλοποιήθηκε για δύο τρόπους σύζευξης του φαρμάκου στο αντίσωμα, και εφαρμόστηκε σε πραγματικά βιολογικά και φαρμακευτικά δεδομένα. Το δεύτερο σκέλος της έρευνας επικεντρώθηκε στην ανάπτυξη μίας υπολογιστικής μεθοδολογίας με αντικείμενο την αυτόματη αναγνώριση μεταβολιτών σε φάσματα Πυρηνικού Μαγνητικού Συντονισμού πρωτονίου. Η μεθοδολογία αυτή εφαρμόστηκε σε τεχνητά και πραγματικά φάσματα που είχαν ληφθεί από προηγούμενες μελέτες, και η απόδοσή της συγκρίθηκε με άλλες υπολογιστικές τεχνικές αναγνώρισης μεταβολιτών. Τα αποτελέσματα και οι παρατηρήσεις σχετικά με την επίδοση των αλγορίθμων παρουσιάζονται στα κεφάλαια που ακολουθούν.

Στο κεφάλαιο 2 πραγματοποιείται μία γενική εισαγωγή στα θέματα βιολογίας και βιοπληροφορικής που αφορούν την αντικαρκινική θεραπεία. Αρχικά γίνεται αναφορά στα κύτταρα του ανοσοποιητικού συστήματος, εστιάζοντας περισσότερο στα αντισώματα, καθώς και στις τεχνικές προσδιορισμού της τρισδιάστατης δομής των αντισωμάτων. Στη συνέχεια γίνεται επισκόπηση των σύγχρονων αντικαρκινικών θεραπειών, δίνοντας περισσότερη έμφαση στα Συμπλέγματα Αντισωμάτων-Φαρμάκων. Επίσης, παρουσιάζονται στοιχεία της βιοπληροφορικής που αφορούν το φαρμακευτικό σχεδιασμό, όπως η υπολογιστική τυποποίηση της τρισδιάστατης δομής των μακρομορίων, βάσεις δεδομένων με βιολογικά δεδομένα, καθώς και τεχνικές *in silico* φαρμακευτικού σχεδιασμού.

Στο κεφάλαιο 3 παρουσιάζεται η υπολογιστική μέθοδος που αναπτύχθηκε για τον υπολογιστικό σχεδιασμό Συμπλεγμάτων Αντισωμάτων-Φαρμάκων με δύο τρόπους σύζευξης του φαρμάκου στο αντίσωμα, σε λυσίνες και τροποποιημένες κυστεΐνες. Συγκεκριμένα, περιγράφεται η εφαρμογή του αφινικού μετασχηματισμού για την περιστροφή και



μετατόπιση του κυτταροτοξικού και του συνδετικού μορίου, ο καθορισμός των ατόμων με βάση τα οποία υπολογίστηκαν οι άξονες των μορίων και οι αποστάσεις μεταξύ τους, καθώς και οι συνθήκες με βάση τις οποίες ρυθμίστηκε η περιστροφή και μετατόπιση των μορίων. Επίσης, παρουσιάζονται αποτελέσματα της εφαρμογής της μεθόδου σε αντισώματα και φάρμακα από βάσεις δεδομένων.

Στο κεφάλαιο 4 πραγματοποιείται μία εισαγωγή στη μεταβολομική και τη φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance, NMR). Αρχικά παρουσιάζονται βασικές έννοιες της μεταβολομικής, οι αναλυτικές τεχνικές που χρησιμοποιούνται στα πλαίσιά της, καθώς και ενδεικτικές εφαρμογές της μεταβολομικής ανάλυσης στη διάγνωση και την ανάπτυξη φαρμακευτικών προϊόντων. Στη συνέχεια παρουσιάζονται οι αρχές φυσικής που διέπουν τη φασματοσκοπία NMR και τα χαρακτηριστικά του σήματος NMR. Γίνεται αναφορά στη δομή και λειτουργία του φασματογράφου NMR, στην πειραματική διαδικασία, τους τρόπους επεξεργασίας του σήματος NMR, καθώς και στις υπάρχουσες υπολογιστικές μεθόδους αναγνώρισης και ποσοτικοποίησης μεταβολιτών.

Στο κεφάλαιο 5 παρουσιάζεται το μεθοδολογικό σχήμα που αναπτύχθηκε με αντικείμενο την αυτόματη αναγνώριση μεταβολιτών από μονοδιάστατα φάσματα  $^1\text{H}$  NMR. Γίνεται αναφορά στις μεθόδους προεπεξεργασίας και μείωσης δεδομένων που εφαρμόστηκαν. Στη συνέχεια αναλύονται τα συστατικά τμήματα της μεθόδου αναγνώρισης μεταβολιτών που αναπτύχθηκε. Παρουσιάζεται η διεργασία της αναζήτησης μεταβολιτών και του τρόπου απόρριψης των μικρών κορυφών. Ακόμη, παρατίθενται οι συναρτήσεις βαθμολόγησης των υποψήφιων συνόλων φασματικών κορυφών που συντέθηκαν για τις διαφορετικές κατηγορίες κορυφών. Αναλύεται η στρατηγική επιλογής του καλύτερου συνδυασμού κορυφών για έναν μεταβολίτη και γίνεται αναφορά στην προτεραιοποίηση των υποψήφιων μεταβολιτών που υιοθετήθηκε στη μεθοδολογία. Τέλος, παρουσιάζονται τα αποτελέσματα εφαρμογής της προτεινόμενης μεθοδολογίας σε δείγματα από διάφορες μελέτες, και συγκρίνονται με τα αντίστοιχα αποτελέσματα άλλων δύο μεθοδολογιών αναγνώρισης μεταβολιτών.



## Κεφάλαιο 2. Εισαγωγή στην Αντικαρκινική Θεραπεία και στις Εφαρμογές Βιοπληροφορικής

---

### 2.1. Ανοσολογία

#### 2.1.1. Ανοσοποιητικό σύστημα

Το ανοσοποιητικό σύστημα αποτελεί ένα δίκτυο από μόρια, κύτταρα, ιστούς και όργανα που σχετίζονται με την προστασία και υπεράσπιση του οργανισμού από εξωτερικούς μολυσματικούς παράγοντες. Οι μολυσματικοί παράγοντες είναι κυρίως βακτήρια, ιοί, παράσιτα και μύκητες, και προσπαθούν να εισχωρήσουν στο ανθρώπινο σώμα γιατί αποτελεί ιδανικό περιβάλλον για την επιβίωση και τον πολλαπλασιασμό τους. Ρόλος του ανοσοποιητικού συστήματος είναι η αποτροπή της εισβολής τους ή η καταστροφή των εισβολέων (Cruse & Lewis, 2010).

Τα λεμφοκύτταρα έχουν κεντρικό ρόλο στο ανοσοποιητικό σύστημα. Όλα τα λεμφοκύτταρα προέρχονται από βλαστοκύτταρα, τα οποία πυροδοτούμενα από κατάλληλα χημικά σήματα εξελίσσονται σε πιο εξειδικευμένα ανοσοποιητικά κύτταρα, όπως τα Β λεμφοκύτταρα, τα Τ λεμφοκύτταρα, τα κύτταρα φυσικοί φονείς (natural killer cells, NK) και τα μακροφάγα. Τα Β λεμφοκύτταρα διαφοροποιούνται σε πλασματοκύτταρα και Β κύτταρα μνήμης, ενώ τα Τ λεμφοκύτταρα διακρίνονται σε κυτταροτοξικά Τ κύτταρα, βοηθητικά Τ κύτταρα και Τ κύτταρα μνήμης. (Barrett et al., 2016)

Το ανοσοποιητικό σύστημα έχει δύο κλάδους, τη φυσική και την επίκτητη ανοσία. Φυσική ανοσία είναι η προστασία του οργανισμού από μολυσματικούς παράγοντες μέσω μη εξειδικευμένων φυσικών, χημικών και μοριακών μηχανισμών. Παραδείγματα φυσικής ανοσίας είναι η προστασία από το δέρμα, οι βλεννώδεις μεμβράνες και τα στομαχικά οξέα. Σε κυτταρικό επίπεδο, η φυσική ανοσία συμπεριλαμβάνει τα κύτταρα φυσικούς φονείς (NK) και τα φαγοκύτταρα. Επίκτητη ανοσία είναι η προστασία από έναν μολυσματικό παράγοντα μετά από έκθεση του οργανισμού σε αυτόν που ρυθμίζεται από Β και Τ λεμφοκύτταρα. Χαρακτηριστικά της επίκτητης ανοσίας είναι η εξειδίκευση ως προς τον μολυσματικό παράγοντα και η ανοσολογική μνήμη. (Cruse & Lewis, 2010)

Αντιγόνο ονομάζεται μία ουσία που είναι ικανή να προκαλέσει την παραγωγή αντισωμάτων και να αντιδράσει με τα αντισώματα αυτά. Περιληπτικά, αντίσωμα είναι μία πρωτεΐνη που παράγεται από ένα Β κύτταρο και μπορεί να προσδεθεί σε ένα αντιγόνο. Πιο αναλυτικά, τα Β και Τ κύτταρα έχουν υποδοχείς που μπορούν να αναγνωρίσουν αντιγόνα. Όταν ένα Β κύτταρο συναντήσει το αντιγόνο του, πυροδοτείται η παραγωγή ενός μεγάλου αριθμού κυττάρων πλάσμα. Κάθε πλασματοκύτταρο παράγει ένα συγκεκριμένο τύπο αντισώματος. Κάθε αντίσωμα έχει την ιδιότητα της εξειδίκευσης (specificity), δηλαδή η δομή του είναι τέτοια ώστε αφ' ενός να προσδένεται σε συγκεκριμένο αντιγόνο με μεγάλη συνάφεια (affinity), αφ' ετέρου να έχει σημαντικά μικρότερη συνάφεια με άλλα αντιγόνα, και επομένως να μην προσδένεται σε αυτά. Η ιδιότητα αυτή αποτυπώνεται στο ότι η σχέση ενός αντιγόνου με το αντίσωμά του περιγράφεται συχνά με τη σχέση κλειδιού και κλειδαριάς. Όταν το αντίσωμα προσδένεται στο αντιγόνο, το αντίσωμα σηματοδοτεί το αντιγόνο ώστε στη συνέχεια να καταστραφεί από τα Τ κύτταρα. (Goldsby et al., 2002)

### 2.1.2. Αντισώματα

Τα αντισώματα είναι πρωτεΐνες που ανήκουν στην οικογένεια των ανοσοσφαιρινών (immunoglobulins) και παράγονται από το ανοσοποιητικό σύστημα στα πλαίσια του μηχανισμού εξουδετέρωσης αντιγόνων. Η βασική δομή ενός αντισώματος είναι 4 πολυπεπτιδικές αλυσίδες, ενωμένες ώστε να σχηματίζουν το σχήμα «Υ». Οι 2 αλυσίδες έχουν μοριακό βάρος τουλάχιστον 50000 kd, είναι πανομοιότυπες και ονομάζονται βαριές (heavy, H), ενώ οι υπόλοιπες 2 αλυσίδες έχουν μοριακό βάρος περίπου 25000 kd, είναι πανομοιότυπες και ονομάζονται ελαφρές (light, L). Μία ελαφριά αλυσίδα ενώνεται με μία βαριά με έναν δισουλφιδικό δεσμό, δεσμούς υδρογόνου, υδροφοβικούς και ιοντικούς δεσμούς, σχηματίζοντας ένα διμερές (H-L). Παρόμοιοι μη ομοιοπολικοί και δισουλφιδικοί δεσμοί ενώνουν τα δύο (H-L) διμερή. Ο ακριβής αριθμός και η τοποθεσία αυτών των δισουλφιδικών δεσμών διαφέρει ανάλογα με την κατηγορία αντισώματος. (Goldsby et al., 2002)

Η αλληλουχία αμινοξέων ενός αντισώματος χαρακτηρίζεται από διαστήματα που είναι σε μεγάλο ποσοστό σταθερά σε όλα τα αντισώματα και δηλώνονται ως σταθερές περιοχές (constant regions, C), καθώς και από διαστήματα που παρουσιάζουν μεγάλες διαφορές από αντίσωμα σε αντίσωμα και δηλώνονται ως μεταβλητές περιοχές (variable regions, V). Η μεταβλητή περιοχή διακρίνεται περαιτέρω σε 3 υπερμεταβλητές περιοχές (hypervariable, HV ή complementarity- determining regions, CDRs), στις οποίες κατά κύριο λόγο εντοπίζεται η έντονη μεταβλητότητα στην αλληλουχία αμινοξέων, και σε 4 πιο σταθερές (framework, FR) περιοχές, οι οποίες έχουν λιγότερο μεταβλητή αλληλουχία αμινοξέων και υποστηρίζουν δομικά τις HV περιοχές. Η ιδιότητα της εξειδίκευσης των αντισωμάτων οφείλεται στην έντονη μεταβλητότητα των περιοχών CDR, στις οποίες πραγματοποιείται η πρόσδεση με το αντιγόνο. (Goldsby et al., 2002)

Με βάση τη δομή της σταθερής περιοχής τους, τα αντισώματα διακρίνονται σε πέντε τάξεις (IgM, IgG, IgA, IgD, IgE), οι οποίες έχουν διαφορετικούς ρόλους στο ανοσοποιητικό σύστημα. Η ανοσοσφαιρίνη G (immunoglobulin G, IgG) καλύπτει τα μικρόβια, επιταχύνοντας την εξουδετέρωσή τους από άλλα ανοσοποιητικά κύτταρα. Η ανοσοσφαιρίνη M (immunoglobulin M, IgM) εξουδετερώνει τα βακτηρίδια. Η ανοσοσφαιρίνη A (immunoglobulin A, IgA) εντοπίζεται στις εξωτερικές εκκρίσεις προστατεύοντας τις εισόδους στο σώμα. Η ανοσοσφαιρίνη E (immunoglobulin E, IgE) προστατεύει από παρασιτικές μολύνσεις. Η ανοσοσφαιρίνη D (immunoglobulin D, IgD) παραμένει συνδεδεμένη στα B-κύτταρα και παίζει βασικό ρόλο στην απόκριση των B-κυττάρων. (Goldsby et al., 2002)

Μία πρωτεΐνη χαρακτηρίζεται από πρωτοταγή, δευτεροταγή, τριτοταγή και τεταρτοταγή δομή. Η πρωτοταγής δομή μίας πρωτεΐνης είναι η γραμμική ακολουθία των αμινοξέων που την απαρτίζουν. Η δευτεροταγής δομή αναφέρεται στις αναδιπλώσεις που μπορούν να έχουν τμήματα της πολυπεπτιδικής αλυσίδας είτε σε μορφή α-έλικα είτε σε μορφή β-πτυχωτής επιφάνειας. Η τριτοταγής δομή αφορά την τρισδιάστατη διάταξη που λαμβάνει μία πρωτεΐνη μετά την αναδίπλωση μεγάλων τμημάτων των πολυπεπτιδικών αλυσίδων της στο χώρο. Τέλος, η τεταρτοταγής δομή αφορά το σχήμα του πρωτεϊνικού σύμπλοκου που σχηματίζεται μετά τη συνένωση όμοιων ή διαφορετικών πολυπεπτιδικών αλυσίδων. (Goldsby et al., 2002)

Τα πολυκλωνικά αντισώματα (polyclonal antibodies) είναι ένα ετερογενές σύνολο αντισωμάτων με διαφορετικές συνάψεις πρόσδεσης, που παράγονται φυσικά από το ανοσοποιητικό σύστημα και εντοπίζονται στον ορό αίματος. Τα μονοκλωνικά αντισώματα (monoclonal antibodies, mAbs) είναι ένα ομογενές σύνολο αντισωμάτων, με πανομοιότυπη δομή και εξειδίκευση. Μπορούν να παραχθούν φυσικά σε παθολογικές περιπτώσεις, όπως το πολλαπλό μυέλωμα, ή τεχνητά με τεχνολογία υβριδώματος. Στη δεύτερη περίπτωση, η παραγωγή μονοκλωνικών αντισωμάτων επιτυγχάνεται συγχωνεύοντας ένα φυσιολογικό Β-κύτταρο με ένα μυελωματικό κύτταρο, που είναι ένα καρκινικό πλασματοκύτταρο. Προϊόν αυτής της διαδικασίας είναι ένα υβριδικό κύτταρο, το υβρίδωμα, το οποίο κατέχει τόσο την ιδιότητα ατέρμονου πολλαπλασιασμού όσο και την ιδιότητα παραγωγής συγκεκριμένου τύπου αντισωμάτων. Τα μονοκλωνικά αντισώματα χρησιμοποιούνται ως θεραπευτικοί παράγοντες (Goldsby et al., 2002; Cruse & Lewis, 2010).

### 2.1.3. Τεχνικές καθορισμού τρισδιάστατης δομής πρωτεϊνών

Οι πιο καθιερωμένες τεχνικές προσδιορισμού της τρισδιάστατης μοριακής δομής πρωτεϊνών είναι η Κρυσταλλογραφία Ακτίνων Χ (X-ray crystallography) και η Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (NMR Spectroscopy). Στην κρυσταλλογραφία ακτίνων Χ, τα τρία βασικά στοιχεία είναι ο κρύσταλλος πρωτεΐνης, μία πηγή ακτίνων Χ κι ένας ανιχνευτής. Το δείγμα της πρωτεΐνης που πρόκειται να μελετηθεί κρυσταλλοποιείται. Επιλέγεται η χρήση κρυστάλλου γιατί το σήμα από ένα μοναδικό μόριο θα ήταν πολύ αδύναμο. Τα μόρια του κρυστάλλου, που συνήθως είναι  $>10^{15}$ , διασκορπίζουν τις ακτίνες σε συγκεκριμένες κατευθύνσεις, με αποτέλεσμα το σήμα να μπορεί να καταγραφεί πριν το δείγμα καταστραφεί από την ακτινοβολία. Μία δέσμη ακτίνων Χ με μήκος κύματος 1.54 Å παράγεται επιταχύνοντας ηλεκτρόνια προς έναν χάλκινο στόχο. Η δέσμη ακτινοβολίας Χ βάλλει τον κρύσταλλο, και ένα μέρος της διέρχεται από τον κρύσταλλο, ενώ το υπόλοιπο διασκορπίζεται σε διάφορες κατευθύνσεις από τα ηλεκτρόνια του κρυστάλλου. Οι διασκορπισμένες ακτίνες ανιχνεύονται από φιλμ ακτίνων Χ ή από ηλεκτρονικό ανιχνευτή στερεάς κατάστασης. Οι εντάσεις των διασκορπισμένων ακτίνων Χ μετρώνται και συνθέτουν την εικόνα της τρισδιάστατης μοριακής δομής του κρυστάλλου (Smyth et al., 2000; Hickman et al., 1997; Berg et al., 2002).

Η φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance Spectroscopy) είναι η μόνη μέθοδος που επιτρέπει τον καθορισμό της τρισδιάστατης δομής μορίων πρωτεΐνης σε διάλυμα. Η τεχνική αυτή στηρίζεται στο γεγονός ότι ορισμένοι πυρήνες ατόμων είναι εγγενώς μαγνητικοί. Το spin ενός πρωτονίου μπορεί να πάρει μία από δύο καταστάσεις (α ή β) όταν εφαρμόζεται ένα εξωτερικό μαγνητικό πεδίο. Ένα πρωτόνιο στην κατάσταση α μπορεί να περάσει στην υψηλότερης ενέργειας κατάσταση β υπό την επίδραση ενός παλμού ηλεκτρομαγνητικής ακτινοβολίας (παλμός RF). Με δεδομένο ότι μαγνητισμός μεταφέρεται από έναν διεγερμένο πυρήνα σε μη διεγερμένο αν απέχουν λιγότερο από 5 Å, η τρισδιάστατη δομή μίας πρωτεΐνης μπορεί να ανακατασκευαστεί (Berg et al., 2001; Wuthrich et al., 1990).

## 2.2. Αντικαρκινική θεραπεία

### 2.2.1. Καρκίνος

Ο καρκίνος είναι μία κατηγορία ασθενειών που εμφανίζονται σε πολυκύτταρους οργανισμούς και αποτελεί μία από τις επικρατέστερες αιτίες θνησιμότητας στο δυτικό κόσμο. Κοινό χαρακτηριστικό των ασθενειών αυτών είναι ένα σύνολο γενετικών μεταλλάξεων που διαταράσσουν την ισορροπία του κυτταρικού κύκλου και οδηγούν στην εμφάνιση συνόλων κυττάρων με μη ελεγχόμενο ρυθμό ανάπτυξης και έλλειψη διαφοροποίησης, που εισβάλλουν σε τοπικούς ιστούς και συχνά κάνουν μετάσταση σε άλλα όργανα ή ιστούς. Αιτίες πρόκλησης καρκίνου είναι η κληρονομικότητα, οι περιβαλλοντολογικές επιδράσεις, η γήρανση και ορισμένοι μολυσματικοί παράγοντες (Ruddon et al., 2007).

Υπάρχουν δεκάδες είδη καρκίνου, που ταξινομούνται στις εξής κατηγορίες με κριτήριο το είδος των καρκινικών κυττάρων. Αυτές είναι (i) το καρκίνωμα, που αφορά τα επιθηλιακά κύτταρα, (ii) το σάρκωμα, που αφορά τα κύτταρα των συνδετικών ιστών, όπως ο μυϊκός και λιπώδης ιστός και οι τένοντες, (iii) η λευχαιμία, που αφορά τα κύτταρα του αίματος, (iv) το λέμφωμα, που αφορά τα κύτταρα του ανοσοποιητικού συστήματος, (v) το πολλαπλούν μυέλωμα, που αφορά τα πλασματοκύτταρα, και (vi) το μελάνωμα, που αφορά τα κύτταρα που πρόκειται να εξελιχθούν σε μελανοκύτταρα. (“National Cancer Institute”, <http://www.cancer.gov> )

Τα φυσιολογικά κύτταρα ωριμάζουν μέσω μίας διαδικασίας που ονομάζεται διαφοροποίηση, μετά από την οποία αρχίζουν να διεξάγουν τη λειτουργία που τους αναλογεί. Επίσης, παραμένουν ενωμένα στη σωστή περιοχή και αναπαράγονται με ελεγχόμενο ρυθμό. Δεχόμενα σήματα από άλλα κύτταρα, αυτοκαταστρέφονται όταν υποστούν σημαντική βλάβη ή γήρανση. Αντίθετα, τα καρκινικά κύτταρα διαιρούνται με ρυθμό πιο γρήγορο από τον φυσιολογικό, με αποτέλεσμα να μην προλαβαίνουν να ωριμάσουν και να μην εκτελούν τη λειτουργία που τους αναλογεί. Μπορούν να αποκολληθούν από τα γειτονικά κύτταρα και να μεταναστεύσουν σε άλλους ιστούς. Κύρια χαρακτηριστικά τους είναι ότι όχι μόνο δεν υπακούουν σε εξωτερικά σήματα, αλλά παράγουν σήματα που προωθούν την ανάπτυξή τους, με αποτέλεσμα να πολλαπλασιάζονται ατέρμονα, αποφεύγοντας τον φυσιολογικό κυτταρικό θάνατο. (“National Cancer Institute”, Ruddon et al., 2007).

Το ανοσοποιητικό σύστημα μπορεί να δράσει εναντίον των καρκινικών κυττάρων μέσω του μηχανισμού ADCC (Antibody-dependent cell-mediated cytotoxicity). Σύμφωνα με το μηχανισμό αυτό, αντισώματα (κυρίως της τάξης IgG, IgE ή IgA) προσδένονται σε αντιγόνα που βρίσκονται στην μεμβρανική επιφάνεια των κυττάρων-στόχων, δηλαδή των καρκινικών κυττάρων. Στη συνέχεια άλλα ανοσοποιητικά κύτταρα, κυρίως τα κύτταρα φυσικοί φονείς (NK), αναγνωρίζουν τα τμήματα Fc των προσδεσμένων αντισωμάτων, προσδένονται σε αυτά και προκαλούν λύση των καρκινικών κυττάρων. (Teillaud et al., 2012). Παρ’ όλα αυτά, οι όγκοι αποφεύγουν την αναγνώριση και καταστροφή τους από το ανοσοποιητικό σύστημα μέσα από διάφορους ανοσοκατασταλτικούς μηχανισμούς, με αποτέλεσμα η φυσική δράση του ανοσοποιητικού συστήματος να μην είναι αρκετή για την αντιμετώπιση του καρκίνου (Charlton et al., 2016).

## **2.2.2. Αντικαρκινικές Θεραπείες**

### **Συμβατικές Θεραπείες**

Διαδεδομένες προσεγγίσεις αντιμετώπισης του καρκίνου είναι η χειρουργική απομάκρυνση του όγκου και η ραδιοθεραπεία, που είναι τοπικές θεραπείες, δηλαδή εφαρμόζονται κυρίως σε περιπτώσεις καρκίνου εντοπισμένου σε μία σχετικά περιορισμένη περιοχή. Η ραδιοθεραπεία είναι η εντοπισμένη χρήση ιοντικής ακτινοβολίας για την εξουδετέρωση κακοήθων όγκων. Η δόση της ραδιοθεραπείας περιορίζεται από τον κίνδυνο καταστροφής των φυσιολογικών κυττάρων. Η αποτελεσματικότητα της ραδιοθεραπείας μπορεί να είναι μειωμένη σε περιπτώσεις καρκίνου με τάση ανάκαμψης, όπως ο καρκίνος στο κεφάλι ή στο λαιμό (Greenhalgh et al., 2014).

Άλλη διαδεδομένη προσέγγιση αντιμετώπισης του καρκίνου είναι η χημειοθεραπεία, δηλαδή η ενδοφλέβια συστηματική χορήγηση κυτταροτοξικών φαρμάκων, η οποία εφαρμόζεται και σε περιπτώσεις που ο καρκίνος δεν είναι εντοπισμένος σε ένα σημείο. Στόχος της χημειοθεραπείας είναι η εξουδετέρωση των καρκινικών κυττάρων, επηρεάζοντας σε μικρότερο βαθμό τα φυσιολογικά κύτταρα. Τα κυτταροτοξικά φάρμακα που χρησιμοποιούνται στη χημειοθεραπεία επιδρούν στην κυτταρική μίτωση και λειτουργούν μέσω 3 μηχανισμών: (i) διαταραχή της δομής DNA (ii) αποτροπή της σύνθεσης DNA (iii) διαταραχή των κυτταρικών μικροσωληνίσκων (Charlton et al., 2016). Κύριο μειονέκτημα της χημειοθεραπείας είναι η έλλειψη στοχευμένης τοξικότητας, αφού δέκτες της τοξικής δράσης του χορηγούμενου φαρμάκου είναι όχι μόνο τα καρκινικά αλλά και τα φυσιολογικά κύτταρα. Η έλλειψη στοχευμένης δράσης θέτει περιορισμούς ως προς τη δόση και την τοξικότητα του χορηγούμενου φαρμάκου, που έχει ως αποτέλεσμα τη μείωση της ποιότητας διαβίωσης του ασθενή, την πιθανότητα ανάπτυξης αντίστασης των καρκινικών κυττάρων στο φάρμακο, αλλά και την πιθανότητα επαναφοράς του καρκίνου (Greenhalgh et al., 2014, Sapra et al., 2013).

### **Στοχευμένες Θεραπείες**

Τα μειονεκτήματα της χημειοθεραπείας έχουν οδηγήσει στην ανάπτυξη στοχευμένων θεραπευτικών παραγόντων, που αποσκοπούν στην επιλεκτική στόχευση των καρκινικών κυττάρων με βάση συγκεκριμένα μοριακά χαρακτηριστικά. Οι στοχευμένες θεραπείες μπορούν να αναστέλλουν συγκεκριμένους μοριακούς στόχους που εμπλέκονται στον καρκίνο, αντί να επιδρούν στην κυτταρική διαίρεση ή τη σύνθεση του DNA. Οι κατηγορίες στοχευμένων θεραπειών συμπεριλαμβάνουν τους μικρομοριακούς αναστολείς, τα μονοκλωνικά αντισώματα και τα συμπλέγματα αντισωμάτων. (Charlton et al., 2016)

Οι υποδοχείς αυξητικών παραγόντων είναι πρωτεΐνες που έχουν σημαντικό ρόλο στην ανάπτυξη των καρκινικών κυττάρων. Για παράδειγμα, ο αγγειακός επιδερμικός αυξητικός παράγοντας (VEGFR) είναι υπεύθυνος για την αγγειογένεση που είναι απαραίτητη για την παροχή επαρκούς αίματος στους όγκους. Ο ανθρώπινος υποδοχέας επιδερμικού αυξητικού παράγοντα τύπου 2 (HER2) και ο υποδοχέας του επιδερμικού αυξητικού παράγοντα (EGFR) ανήκουν στην οικογένεια υποδοχέων ErbB, τα μέλη της οποίας αλληλεπιδρώντας με προσδέτες συμβάλλουν στη μεταγωγή σημάτων και τον κυτταρικό πολλαπλασιασμό. Οι αναστολείς κινάσης τυροσίνης είναι μικρά μόρια που δρουν αποτρέποντας τη μεταγωγή σημάτων από τους υποδοχείς αυξητικών παραγόντων. Παράδειγμα φαρμάκου αυτής της κατηγορίας είναι το erlotinib, που είναι αναστολέας της κινάσης της τυροσίνης του υποδοχέα

του επιδερμικού αυξητικού παράγοντα (EGFR) και χρησιμοποιείται στη θεραπεία του μη μικροκυτταρικού καρκίνου πνεύμονα (NSCLC). Το lapatinib είναι επίσης αναστολέας της κινάσης της τυροσίνης, που αναστέλλει τον ανθρώπινο υποδοχέα επιδερμικού αυξητικού παράγοντα τύπου 2 (HER2) και τον υποδοχέα του επιδερμικού αυξητικού παράγοντα (EGFR), και επιστρατεύεται για τη θεραπεία του μεταστατικού καρκίνου μαστού. Ακόμη, το sorafenib είναι αναστολέας της κινάσης της τυροσίνης του αγγειακού επιδερμικού αυξητικού παράγοντα (VEGFR) και χρησιμοποιείται για τη θεραπεία του καρκίνου νεφρού (Yap and Workman, 2012; Zhang et al., 2009; Charlton et al., 2016).

Τα μονοκλωνικά αντισώματα (monoclonal antibodies, mAbs) είναι αντισώματα που παράγονται εργαστηριακά και προέρχονται από το ίδιο Β-κύτταρο. Είναι πανομοιότυπα και επομένως προσδένονται στο ίδιο αντιγόνο-στόχο. Με βάση αυτό το χαρακτηριστικό, μονοκλωνικά αντισώματα έχουν χρησιμοποιηθεί για τη στόχευση ενός συνόλου αντιγόνων που εντοπίζονται σε μεγάλο βαθμό στην επιφάνεια συγκεκριμένων τύπων καρκινικών κυττάρων και σε πολύ μικρότερο βαθμό στην επιφάνεια φυσιολογικών κυττάρων. Τα αντισώματα μπορούν να έχουν αντικαρκινική δράση μέσω τεσσάρων μηχανισμών: ενεργοποίηση των μηχανισμών CDC (complement dependent cytotoxicity) και ADCC (antibody dependent cellular cytotoxicity), διαταραχή των σημάτων μεταξύ των καρκινικών κυττάρων, καθώς και πρόκληση επίκτητης ανοσίας (Shuptrine et al., 2012; Kubota et al., 2009; Weiner et al., 2015). Παραδείγματα μονοκλωνικών αντισωμάτων που έχουν γνωρίσει επιτυχία ως αυτόνομοι θεραπευτικοί παράγοντες είναι τα rituximab, trastuzumab και cetuximab (Iyer et al., 2011).

Παρ' όλα αυτά, τα μονοκλωνικά αντισώματα παρουσιάζουν μειονεκτήματα, όπως η ελαττωματική διείσδυση στους όγκους, με μόνο ένα μικρό ποσοστό των χορηγημένων αντισωμάτων να εντοπίζονται στον όγκο, καθώς και η αργή εξαγγείωση του φαρμάκου από την κυκλοφορία. Η παραγωγή αντισωμάτων είναι αρκετά ακριβή, αυξάνοντας την τιμή των αντίστοιχων φαρμάκων. Επιπλέον, η θεραπεία με μονοκλωνικά αντισώματα συνήθως δεν είναι επαρκής και χρειάζεται να συνδυάζεται με χημειοθεραπεία (Weiner et al., 2010; Iyer et al., 2011).

Οι περιορισμοί των μονοκλωνικών αντισωμάτων είχαν ως αποτέλεσμα την προσπάθεια για την ενίσχυση της αποτελεσματικότητάς τους προσδένοντάς τα με κάποια κυτταροτοξική ουσία. Τα συμπλέγματα αντισωμάτων είναι μία ευρεία κατηγορία θεραπευτικών παραγόντων, των οποίων η λειτουργία μπορεί να συνοψιστεί ως η στοχευμένη παράδοση μίας κυτταροτοξικής ουσίας στα καρκινικά κύτταρα, η οποία μπορεί να είναι ένα φάρμακο, ένα ραδιοϊσότοπο ή μία τοξίνη. (Weiner et al., 2015)

Τα ραδιοανοσοσυμπλέγματα (radioimmunoconjugates) είναι συμπλέγματα αντισωμάτων συνδεδεμένα με ένα ραδιοϊσότοπο. Η λειτουργία τους μπορεί να συνοψιστεί ως η στοχευμένη παράδοση του ραδιοϊσότοπου στο καρκινικό κύτταρο. Μετά τη χορήγηση, το σύμπλεγμα κυκλοφορεί στο αίμα. Όταν το αντίσωμα συναντήσει το αντιγόνο-στόχο του, προσδένεται σε αυτό με αποτέλεσμα το ραδιοϊσότοπο να παραδίδεται στο κύτταρο. Μειονέκτημα αυτής της κατηγορίας θεραπευτικών παραγόντων είναι ότι τα ραδιοϊσότοπα φθίνουν συνεχώς, προκαλώντας μη στοχευμένη βλάβη σε φυσιολογικά κύτταρα όταν το σύμπλεγμα βρίσκεται στην κυκλοφορία. Επίσης, ένα υψηλό ποσοστό της ακτινοβολίας απορροφάται από το συκώτι και τα νεφρά, με αποτέλεσμα ένα μικρό τμήμα της χορηγημένης δόσης να καταλήγει στον όγκο. Δύο ραδιοανοσοσυμπλέγματα, τα ibritumomab tiuxetan



(Witzig et al., 2002) και tositumomab (Kaminski et al., 1993), έχουν εγκριθεί ως θεραπευτικοί παράγοντες για το λέμφωμα. (Weiner et al., 2015)

Αν και τα μη τροποποιημένα μονοκλωνικά αντισώματα έχουν συγκεκριμένη εξειδίκευση για ένα μοναδικό αντιγόνο, ο μηχανισμός των πολύπλοκων ασθενειών συνήθως σχετίζεται με περισσότερα από ένα χαρακτηριστικά μόρια-στόχους. Βελτιωμένη θεραπευτική απόδοση μπορεί να σημειωθεί με πρόσδεση σε πολλαπλά μόρια, το οποίο μπορεί να επιτευχθεί τόσο με συνδυασμένη χορήγηση διαφορετικών μονοκλωνικών αντισωμάτων, όσο με τη χρήση αμφιειδικών αντισωμάτων (bispecific antibodies, bsAbs). Τα αμφιειδικά αντισώματα συνδυάζουν σε ένα μόριο τις εξειδικεύσεις δύο αντισωμάτων, επομένως έχουν τη δυνατότητα να προσδεθούν σε δύο διαφορετικούς στόχους. Μπορούν επίσης να φέρουν σε κοντινή απόσταση δύο στόχους, είτε για να υποστηρίξουν τη δημιουργία πρωτεϊνών είτε για να πυροδοτήσουν κάποια διαμοριακή αλληλεπίδραση. Το αμφιειδικό αντίσωμα Removab έχει εγκριθεί για τη θεραπεία του καρκίνου ωοθηκών (Kontermann et al., 2012; Kontermann et al., 2015).

Τα Συμπλέγματα Αντισωμάτων-Φαρμάκων (Antibody drug conjugates, ADCs) είναι μία κατηγορία θεραπευτικών παραγόντων που συντίθενται από ένα μονοκλωνικό αντίσωμα, ένα κυτταροτοξικό μόριο και ένα συνδετικό μόριο που ενώνει τα δύο προηγούμενα. Το αντίσωμα είναι επιλεγμένο ώστε να προσδέεται σε συγκεκριμένο αντιγόνο που υπάρχει σε μεγάλη συγκέντρωση στην επιφάνεια των καρκινικών κυττάρων. Το ADC κυκλοφορεί στο αίμα μέχρι το αντίσωμα να προσδεθεί στο κατάλληλο αντιγόνο, οπότε το φάρμακο απελευθερώνεται και ενεργοποιείται. Παραδείγματα ADCs που έχουν εγκριθεί ως θεραπευτικοί παράγοντες είναι το Kadcyla και το Adcetris (Charlton et al., 2016). Τα ADCs περιγράφονται πιο αναλυτικά στο επόμενο κεφάλαιο.

### **2.2.3. Συμπλέγματα Αντισωμάτων-Φαρμάκων (ADCs)**

#### **Βασικές αρχές**

Τα Συμπλέγματα Αντισωμάτων-Φαρμάκων (Antibody-Drug Conjugates, ADCs) είναι μία κατηγορία αντικαρκινικών φαρμάκων που συνδυάζει τα πλεονεκτήματα των αντισωμάτων και των κυτταροτοξικών ουσιών συνδυάζοντάς τα σε ένα μόριο. Συγκεκριμένα, ένα ADC αποτελείται από τρία στοιχεία: μία κυτταροτοξική ουσία, ή αλλιώς φάρμακο (drug), ένα αντίσωμα (antibody) ή τμήμα αντισώματος, και ένα συνδετικό μόριο, ή συνδέτη (linker), που συνδέει τα προηγούμενα δύο. Η λειτουργία ενός ADC συνοψίζεται στην παράδοση του κυτταροτοξικού μορίου απευθείας στα καρκινικά κύτταρα, περιορίζοντας τη βλάβη στους υγιείς ιστούς (Li et al., 2012; Adler et al., 2012; Goldmacher et al., 2012). Το ADC κυκλοφορεί στο αίμα μέχρι το αντίσωμα να συναντήσει το εμφανιζόμενο στα καρκινικά κύτταρα αντιγόνο για το οποίο έχει εξειδίκευση και να προσδεθεί σε αυτό. Μετά την πρόσδεση του αντισώματος στο αντιγόνο, το φάρμακο απελευθερώνεται και ενεργοποιείται, σκοτώνοντας το καρκινικό κύτταρο (Nolting et al., 2013).

Πιο αναλυτικά, η τυπική διαδρομή ενός μορίου ADC μπορεί να περιγραφεί με τα ακόλουθα βήματα. Τα ADCs χορηγούνται ενδοφλέβια για να αποτρέψουν την καταστροφή του αντισώματος από τα γαστρικά οξέα και τα πρωτεολυτικά ένζυμα. Πρώτα, το ADC κυκλοφορεί στο πλάσμα παραμένοντας αμετάβλητο, με το φάρμακο να παραμένει ανενεργό λόγω της πρόσδεσής του στο αντίσωμα. Όταν το ADC συναντήσει το αντιγόνο-στόχο, το αντίσωμα του ADC συνδέεται με αυτό, δημιουργώντας ένα σύμπλεγμα ADC-αντιγόνου.

Τυπικά, το σύμπλεγμα ADC-αντιγόνου υπόκειται σε εσωτερικήυση από το καρκινικό κύτταρο. Ακολούθως, το ADC διασπάται στα τμήματά του μέσω μίας διαδικασίας που εξαρτάται από τον τύπο του συνδέτη που χρησιμοποιείται. Σε αυτό το στάδιο, το φάρμακο ενεργοποιείται και σκοτώνει το κύτταρο, με τρόπο που εξαρτάται από τον τύπο του χρησιμοποιούμενου κυτταροτοξικού παράγοντα. Μία εναλλακτική στρατηγική των ADCs είναι η στόχευση των ενδοθηλιακών κυττάρων των καρκινικών αγγείων με σκοπό να στερήσουν το αίμα από τα καρκινικά κύτταρα (Sapra et al., 2013; Kovtun et al., 2007; Gerber et al., 2009; Perez et al., 2014; Wu et al., 2005).

Ένα ιδανικό ADC πρέπει να διατηρεί τις ωφέλιμες φαρμακοκινητικές και λειτουργικές ιδιότητες του αντισώματος, για παράδειγμα τον μεγάλο χρόνο κυκλοφορίας στο αίμα και την δυνατότητα πρόσδεσης σε συγκεκριμένο αντιγόνο. Πρέπει επίσης να αξιοποιεί την κυτταροτοξική ιδιότητα του φαρμάκου με ωφέλιμο τρόπο, διατηρώντας το ανενεργό στην κυκλοφορία και απελευθερώνοντάς το στην τοποθεσία-στόχο. Το φάρμακο πρέπει να είναι αρκετά ισχυρό ώστε να μπορεί να εξουδετερώσει το καρκινικό κύτταρο. Ένας ιδανικός συνδέτης πρέπει να είναι αρκετά σταθερός στην κυκλοφορία ώστε να μην επιτρέψει την πρόωρη αποδέσμευση του φαρμάκου, αλλά ικανός να απελευθερώσει την ενεργή μορφή του φαρμάκου όταν είναι απαραίτητο. Το αντίσωμα πρέπει να είναι επιλεγμένο ώστε να αντιστοιχεί σε αντιγόνο που υπερεκφράζεται στα καρκινικά κύτταρα (Perez et al., 2014).

### **Συστατικά μέρη**

Ένα μονοκλωνικό αντίσωμα πρέπει να πληρεί κάποιες προϋποθέσεις για να είναι συστατικό στοιχείο ενός ADC. Κατ' αρχάς, πρέπει να διαθέτει σημεία κατάλληλα για σύζευξη με το φάρμακο που να μην παρεμποδίζουν την πρόσδεση του αντισώματος στο αντιγόνο. Το αντίσωμα δεν πρέπει να είναι ανοσογόνο, δηλαδή να προκαλεί ανοσολογική αντίδραση. Τα πρώτα ADCs περιείχαν αντισώματα από ποντίκια, τα οποία ήταν ανοσογόνα. Για την αντιμετώπιση του προβλήματος αυτού, υιοθετήθηκε η τακτική χρήσης εξανθρωπισμένων (humanized) ή πλήρως ανθρώπινων (human) αντισωμάτων στα ADCs. Επίσης, το αντίσωμα πρέπει να είναι επιλεγμένο ώστε να προσδέεται με μεγάλη συνάφεια σε κάποιο από τα μόρια που υπάρχουν σε μεγάλη συγκέντρωση στα καρκινικά κύτταρα (Ryan et al., 2010; Iyer et al., 2011; Thomas et al., 2016)

Μία από τις μεγαλύτερες προκλήσεις στην ανάπτυξη των ADCs είναι η παραγωγή κατάλληλων συνδετικών μορίων (linkers) για τη σύζευξη αντισώματος και φαρμάκου. Ο ρόλος του συνδέτη είναι θεμελιώδης γιατί επηρεάζει αφ' ενός την παράδοση του φαρμάκου στο καρκινικό κύτταρο, αφ' ετέρου τη σταθερότητα του συνδέσμου μεταξύ του φαρμάκου και του αντισώματος, που είναι κρίσιμη για την αποτελεσματική λειτουργία του ADC. Σημαντικά στοιχεία που αφορούν τον συνδέτη είναι το σημείο σύζευξης με το αντίσωμα, ο μέσος αριθμός σημείων σύζευξης ανά αντίσωμα (μόριο), η δυνατότητα του συνδέτη να απελευθερώσει το φάρμακο, και η πολικότητα του συνδέτη.

Τα απαραίτητα χαρακτηριστικά που πρέπει να πληρεί ένας συνδέτης είναι τα εξής. (i) Κατ' αρχάς, πρέπει να παραμένει σταθερός μετά την ενδοφλέβια χορήγηση του ADC στο πλάσμα, ώστε το κυτταροτοξικό μόριο να μην απελευθερωθεί πρόωρα στο πλάσμα. Εάν το φάρμακο απελευθερωθεί στο πλάσμα πριν βρεθεί το αντιγόνο-στόχος, θα καταστραφούν τόσο υγιή όσο και καρκινικά κύτταρα χωρίς διάκριση, όπως στην κλασική χημειοθεραπεία, πράγμα που θα χαμηλώσει το θεραπευτικό δείκτη του ADC. (ii) Μετά την εσωτερικήυση του

ADC από το καρκινικό κύτταρο, ο συνδέτης θα πρέπει να απελευθερώνει το φάρμακο ώστε αυτό να οδηγήσει το καρκινικό κύτταρο σε θάνατο. (iii) Ο συνδέτης πρέπει να έχει φυσικοχημικά χαρακτηριστικά που σε συνδυασμό με τα φυσικοχημικά χαρακτηριστικά του φαρμάκου και του αντισώματος να συμβάλλουν στην καλή λειτουργία του ADC. Για παράδειγμα, ένας υδροφοβικός συνδέτης συνδυασμένος με ένα υδροφοβικό φάρμακο μπορεί να οδηγήσει σε συσσωμάτωση των συμπλεγμάτων, η οποία έχει συνδεθεί με αρνητικές παρενέργειες. Αντίθετα, έχει αποδειχθεί ότι ένας υδροφιλικός συνδέτης συνδυασμένος με το υδροφοβικό φάρμακο της κατηγορίας των maytansinoids παράγει ένα υδροφιλικό σύμπλεγμα συνδέτη-φαρμάκου, που ήταν πιο αποτελεσματικό από το MCC-DM1 ADC (Jain et al., 2015; Nolting et al., 2013; Alley et al., 2008; Ducry et al., 2010; Kovtun et al., 2007).

Οι συνδέτες ταξινομούνται σε δύο κατηγορίες με κριτήριο τον τρόπο απελευθέρωσης του φαρμάκου, τους διασπάσιμους (cleavable) και μη διασπάσιμους (non-cleavable) συνδέτες. Οι διασπάσιμοι συνδέτες αξιοποιούν μία εγγενή ιδιότητα του καρκινικού κυττάρου για την επιλεκτική απελευθέρωση της κυτταροτοξικής ουσίας από το ADC. Διακρίνονται περαιτέρω στους συνδέτες που είναι ασταθείς σε όξινο περιβάλλον (acid-labile linkers), τους δισουλφιδικούς συνδέτες (disulfide linkers), και τους πεπτιδικούς συνδέτες (peptide linkers) (Ducry et al., 2010, Nolting et al., 2013, Peters et al., 2015).

Οι συνδέτες που είναι ασταθείς σε όξινο περιβάλλον στηρίζουν τη λειτουργία τους στη διαφορά του pH ανάμεσα στο πλάσμα και τα λυσοσώματα. Συγκεκριμένα, είναι σχεδιασμένοι να παραμένουν σταθεροί κατά τη συστηματική κυκλοφορία του ADC στο πλάσμα (pH=7.4), και να υπόκεινται σε υδρόλυση απελευθερώνοντας το φάρμακο όταν το ADC εσωτερικεύεται από το κύτταρο, στα ενδοσώματα (pH=5–6) και λυσοσώματα (pH=4.8). Πρόκειται για μία από τις πρώτες κατηγορίες συνδετών που αναπτύχθηκαν, και έχουν συνδεθεί με μη εξειδικευμένη απελευθέρωση του φαρμάκου σε κλινικές μελέτες. Το ADC Gemtuzumab ozogamicin (Mylotarg) χρησιμοποιεί συνδέτη αυτής της κατηγορίας, και αποσύρθηκε από την αγορά το 2010 γιατί συνδέθηκε με σημαντικά υψηλότερη τοξικότητα σε σχέση με την καθιερωμένη θεραπεία (Nolting et al., 2013; Peters et al., 2015; VanDerVelden et al., 2001).

Οι δισουλφιδικοί συνδέτες βασίζουν τη λειτουργία τους στο γεγονός ότι τα καρκινικά κύτταρα έχουν αυξημένες συγκεντρώσεις μορίων θιόλης. Η διάσπαση των δισουλφιδικών δεσμών απαιτεί την παρουσία ενός μορίου που εντοπίζεται στο κυτταρόπλασμα, επομένως οι δισουλφιδικοί συνδέτες παραμένουν σταθεροί στην κυκλοφορία και απελευθερώνουν το φάρμακο μετά την εσωτερική τους από τα καρκινικά κύτταρα. Στη διάσπαση των δισουλφιδικών συνδετών μπορεί να συμβάλλει και το ένζυμο ισομεράση (Peters et al., 2015; Nolting et al., 2013; Homewood et al., 1972; Balendiran et al., 2004).

Οι πεπτιδικοί συνδέτες στηρίζουν τη λειτουργία τους στο γεγονός ότι τα πρωτεολυτικά ένζυμα που απαιτούνται για τη διάσπαση των πεπτιδικών δεσμών έχουν χαμηλή δραστηριότητα στο αίμα (Koblinksi et al., 2000). Επομένως, οι πεπτιδικοί συνδέτες παραμένουν σταθεροί στην κυκλοφορία, και απελευθερώνουν το φάρμακο όταν ο διπεπτιδικός δεσμός τους διασπάται στο όξινο περιβάλλον των λυσοσωμάτων από τις λυσοσωματικές πρωτεάσες. Τα τρία είδη συνδετών έχουν συγκριθεί σε κλινική μελέτη, και αποδείχθηκε ότι οι πεπτιδικοί έχουν μεγαλύτερη εξειδίκευση, σταθερότητα και χρόνο ημιζωής σε σχέση με τους συνδέτες που είναι ασταθείς σε όξινο περιβάλλον (Polson et al.,

2009). Το Brentuximab vedotin είναι ένα ADC που εμπεριέχει έναν πεπτιδικό συνδέτη (Doronina et al., 2008; Peters et al., 2015; Nolting et al., 2013)

Οι μη διασπάσιμοι συνδέτες διαφοροποιούνται από τους διασπάσιμους στο ότι αφ' ενός δεν αξιοποιούν κάποια διαφοροποίηση μεταξύ του πλάσματος και του κυτταροπλάσματος για να απελευθερώσουν το φάρμακο, αφ' ετέρου δεν χρειάζεται να διασπαστούν για να απελευθερώσουν το φάρμακο. Αντίθετα, το ADC εσωτερικεύεται από το καρκινικό κύτταρο και παραδίδεται στα λυσοσώματα, όπου το αντίσωμα διασπάται στα αμινοξέα του (Erickson et al., 2006). Προϊόν αυτής της διαδικασίας είναι ένα σύμπλεγμα που αποτελείται από το φάρμακο, τον συνδέτη και το αμινοξύ σύζευξης, το οποίο απελευθερώνεται στο κυτταρόπλασμα. Το σύμπλεγμα αυτό πρέπει να δρα ως ένα ενεργό κυτταροτοξικό στοιχείο ώστε το ADC να είναι αποτελεσματικό. Οι μη διασπάσιμοι συνδέτες έχουν μεγαλύτερη σταθερότητα στην κυκλοφορία σε σύγκριση με τους διασπάσιμους, πλεονέκτημα που μεταφράζεται σε μεγαλύτερους χρόνους ημιζωής και μικρότερη μη εξειδικευμένη τοξικότητα. Από την άλλη πλευρά, οι διασπάσιμοι συνδέτες μπορούν να απελευθερώσουν το φάρμακο χωρίς το ADC να εσωτερικευθεί πλήρως από το κύτταρο, ενώ οι μη διασπάσιμοι εξαρτώνται από την πλήρη εσωτερίκευση του ADC. Μη διασπάσιμος συνδέτης χρησιμοποιήθηκε στο ADC T-DM1 (Polson et al., 2010; Peters et al., 2015; Nolting et al., 2013; Erickson et al., 2006; Polson et al., 2009).

Παρά την πρόοδο που έχει σημειωθεί, δεν υπάρχουν συγκεκριμένοι κανόνες για την επιλογή του συνδέτη για όλους τους τύπους καρκίνου ή όλα τα ADCs. Η επιλογή αυτή εξαρτάται σε μεγάλο βαθμό από το αντίσωμα, το φάρμακο και τον καρκινικό στόχο. Ο σχεδιασμός και η επιλογή του συνδέτη αξιολογείται με βάση την αποτελεσματικότητα και την τοξικότητα του συγκεκριμένου αντίστοιχου ADC που τον περιέχει. (Nolting et al., 2013; Ducry et al., 2010)

Το φάρμακο που χρησιμοποιείται σε ένα ADC πρέπει να πληρεί ένα σύνολο προϋποθέσεων. Κατ' αρχάς, πρέπει να έχει υψηλή δραστηριότητα για να εξουδετερώσει τα καρκινικά κύτταρα, χαμηλό μοριακό βάρος και να μην προκαλεί ανοσολογική αντίδραση. Η μοριακή δομή του φαρμάκου πρέπει να είναι κατάλληλη ώστε να είναι εφικτή η σύζευξή του με το συνδέτη. Άλλο απαραίτητο χαρακτηριστικό είναι η επαρκής σταθερότητα και διαλυτότητα του φαρμάκου στο αίμα, γιατί τα ADCs προετοιμάζονται σε υδατικό διάλυμα και χορηγούνται ενδοφλέβια (Thomas et al., 2016, Iyer et al., 2011, Singh et al., 2009).

Οι κυτταροτοξικές ουσίες που χρησιμοποιούνται στα ADCs διαχωρίζονται σε (i) ουσίες που διαταράσσουν τους κυτταρικούς μικροσωληνίσκους και (ii) ουσίες που στοχεύουν στη δομή του DNA. Στην πρώτη κατηγορία ανήκουν οι auristatins και οι maytansinoids. Οι auristatins είναι συνθετικά ανάλογα της ουσίας dolostatin και είναι δεκάδες φορές πιο ισχυρά από τα πιο συμβατικά κυτταροτοξικά φάρμακα vinca alkaloids και doxorubicin. Μέλη της κατηγορίας των auristatins που έχουν διερευνηθεί ως μέρη ενός ADC είναι οι monomethyl auristatin E (MMAE) και monomethyl auristatin F (MMAF). Το φάρμακο monomethyl auristatin E (MMAE) έχει χρησιμοποιηθεί στο πραγματικό ADC SGN-35 (Adcetris) (Senter et al., 2012; van de Donk et al., 2012). Οι maytansinoids είναι παράγωγα ενός φυσικού προϊόντος και χαρακτηρίζονται από πολύ υψηλή κυτταροτοξικότητα. Το πραγματικό ADC T-DM1 (Kadcyla) περιέχει maytansine (Polson et al., 2010). Οι calicheamicins χαρακτηρίζονται από μικρό θεραπευτικό δείκτη και αρκετές τοξικές παρενέργειες (Iyer et al., 2011, Thomas et al., 2016). Ένα παράγωγο αυτής της κατηγορίας χρησιμοποιήθηκε στο ADC

gemtuzumab ozogamicin (Mylotarg), το πρώτο που εγκρίθηκε ως θεραπευτικός παράγοντας. Μία πρόσθετη κατηγορία φαρμάκων είναι οι duocarmycins, για τα οποία αποδείχθηκε ότι έχουν έντονες τοξικές παρενέργειες που αποτρέπουν τη χορήγησή τους σε δόσεις ικανές να έχουν αντικαρκινική δράση (MacMillan et al., 2009).

### Τρόποι σύζευξης

Ένας από τους πιο απλούς τρόπους σύζευξης ενός συνδέτη με ένα αντίσωμα είναι σε μία από τις λυσίνες του αντισώματος. Μία ανοσοσφαιρίνη (IgG) περιέχει κατά προσέγγιση 80-100 λυσίνες, οι περισσότερες από τις οποίες είναι αρκετά προσβάσιμες από το εξωτερικό της ανοσοσφαιρίνης. Ανεξάρτητα από τη χημική αντίδραση με την οποία επιτυγχάνεται η σύζευξη, ο μεγάλος αριθμός διαθέσιμων λυσινών σε κάθε αντίσωμα (IgG) σε συνδυασμό με το μικρό λόγο φαρμάκου προς αντίσωμα (drug-to-antibody ratio, DAR) έχει ως αποτέλεσμα την πραγματοποίηση μίας στοχαστικής αντίδρασης και τη δημιουργία ενός ετερογενούς συνόλου ADCs που μπορεί να αποτελείται από περισσότερα από  $10^6$  διαφορετικά πιθανά είδη ADCs, με στατιστική διανομή του φαρμάκου στο αντίσωμα (Wang et al., 2005; Brun et al., 2013).

Ένας άλλος τρόπος σύζευξης του φαρμάκου είναι στις εγγενείς κυστεΐνες του αντισώματος, με προϋπόθεση το αντίσωμα να έχει υποστεί προεπεξεργασία για να διασπαστούν οι δισουλφιδικοί δεσμοί μεταξύ των κυστεϊνών. Η μέθοδος αυτή επιστρατεύτηκε για την κατασκευή συμπλεγμάτων μονοκλωνικών αντισωμάτων-φαρμάκων με τα φάρμακα auristatin E (AE), monomethyl auristatin E (MMAE) και τα χημικά μονοκλωνικά αντισώματα cBR96 και cAC10 (Doronina et al., 2003). Η ίδια μέθοδος χρησιμοποιήθηκε αργότερα για τη σύνθεση συμπλεγμάτων μονοκλωνικών αντισωμάτων-φαρμάκων με τα χημικά αντισώματα IgG1 BR96, AC10 και το φάρμακο auristatin MMAF (Doronina et al., 2006). Με δεδομένο ότι υπάρχουν 8 κυστεΐνες ανά αντίσωμα, αυτός ο τρόπος σύζευξης μπορεί να καταλήξει σε περισσότερα από 100 διαφορετικά είδη συμπλεγμάτων με μη καθορισμένο αριθμό και θέσεις φαρμάκων (0-8 κυτταροτοξικά μόρια ανά αντίσωμα) (Hamblett et al., 2004; Junutula et al., 2008).

Η ετερογένεια όσον αφορά τα σημεία σύζευξης και τον αριθμό μορίων φαρμάκων ανά αντίσωμα στα ADCs συνδέεται με ανεπιθύμητα *in vivo* αποτελέσματα. Επιπλέον, η διαταραχή των εγγενών δισουλφιδικών δεσμών μπορεί να αλλοιώσει την τεταρτοταγή δομή του αντισώματος, διαταράσσοντας την *in vivo* συμπεριφορά του αντισώματος (Junutula et al., 2008). Με δεδομένα αυτά τα μειονεκτήματα, έγιναν ιδιαίτερες προσπάθειες για ανακάλυψη τρόπων σύζευξης πιο εξειδικευμένων ως προς την τοποθεσία σύζευξης και του αριθμού φαρμάκων ανά αντίσωμα.

Η τροποποίηση της αλληλουχίας αμινοξέων του αντισώματος πριν τη σύζευξη με το φάρμακο είναι ένας τρόπος επίτευξης πιο ομογενών συνόλων ADCs. Μία τέτοια προσπάθεια έγινε από τους Junutula et al., στα πλαίσια της οποίας επιλεγμένα εγγενή αμινοξέα του αντισώματος αντικαταστάθηκαν από κυστεΐνες με στόχο να λειτουργήσουν ως στοχευμένα σημεία σύζευξης, επιτρέποντας στα μόρια φαρμάκων να συζευχθούν σε καθορισμένες θέσεις, χωρίς να διαταραχθούν οι εγγενείς δισουλφιδικοί δεσμοί του αντισώματος. Τα τροποποιημένα αντισώματα αυτά ονομάζονται THIOMABS (Thiol-conjugated monoclonal antibodies), και τα συμπλέγματα αυτών με μόρια φαρμάκων ονομάζονται THIOMAB-drug conjugates (TDCs). Ένα σύνολο αμινοξέων (LC-V15C, LC-V110C, LC-S114C, LC-S121C,

LC-S127C, LC-A153C, LC-S168C, LC-V205C στην ελαφριά αλυσίδα και HC-S112C, HC-S113C, HC-S115C, HC-T116C στη βαριά αλυσίδα – αρίθμηση Kabat) αντικαταστάθηκαν με κυστεΐνη και υποβλήθηκαν σε σύζευξη. Η *in vitro* και *in vivo* αξιολόγησή τους απέδειξε ότι τα τροποποιημένα αντισώματα αφ' ενός διατηρούν τις επιθυμητές ιδιότητες του αντισώματος, δηλαδή την εξειδίκευση και τη δυνατότητα πρόσδεσης σε αντιγόνο, αφ' ετέρου τα αντίστοιχα συμπλέγματα είναι σχεδόν ομογενή, με τα περισσότερα να περιέχουν 2 μόρια φαρμάκου ανά αντίσωμα. (Junutula et al., 2008).

Επακόλουθο της έρευνας αυτής ήταν η διερεύνηση του αν τα THIOMABs μπορούν να συζευχθούν με κυτταροτοξικά φάρμακα της κατηγορίας maytansinoids. Συγκεκριμένα, το αντίσωμα trastuzumab μετατράπηκε σε THIOMAB (thio-trastuzumab) αντικαθιστώντας με κυστεΐνη την αλανίνη με αριθμό ακολουθίας 114 κάθε βαριάς αλυσίδας. Το τροποποιημένο αντίσωμα που προέκυψε συνδέθηκε με το φάρμακο DM1 (drug maytansinoid 1), με αποτέλεσμα το σχηματισμό ομογενών συμπλεγμάτων με δύο μόρια φαρμάκου ανά αντίσωμα (Junutula et al., 2010).

Ένας άλλος τρόπος εξειδικευμένης σύζευξης φαρμάκου σε αντίσωμα έχει επιτευχθεί με την εισαγωγή του αμινοξέος σεληνοκυστεΐνη (Selenocysteine, Sec) στο αντίσωμα. Η σεληνοκυστεΐνη είναι το 21<sup>ο</sup> φυσικό αμινοξύ, είναι ανάλογη με τα αμινοξέα κυστεΐνη (Cys) και σερίνη (Ser) και έχει ιδιαίτερες χημικές ιδιότητες (Johansson et al., 2004; Li et al., 2014). Μία σεληνοκυστεΐνη εισήχθη σε ένα τμήμα πρωτεΐνης, το οποίο στη συνέχεια συνδέθηκε στο μικρό συνθετικό μόριο LLP2A, το οποίο έχει φαρμακολογική δραστηριότητα (Hofer et al., 2008). Ακολούθως, η παραπάνω τεχνολογία τροποποίησης σεληνοκυστεϊνών επιτεύχθηκε όχι μόνο σε τμήματα αντισωμάτων αλλά και σε ολόκληρα αντισώματα, διευρύνοντας τη χρησιμότητα της μεθόδου (Hofer et al., 2009).

Εξειδικευμένη σύνθεση ADCs είναι εφικτή και με την εισαγωγή μη φυσικών αμινοξέων στο αντίσωμα. Οι Axup et al επιχείρησαν να συμπεριλάβουν το μη φυσικό αμινοξύ p-acetylphenylalanine (pAcPhe) σε παραλλαγή του αντισώματος trastuzumab. Στη συνέχεια έγινε σύζευξη του τροποποιημένου αντισώματος με το φάρμακο C-terminal alkoxy-amine auristatin derivative. Η μέθοδος αυτή είχε ως αποτέλεσμα να παραχθούν ομογενή ADCs με άριστη αποτελεσματικότητα και φαρμακοκινητικές ιδιότητες (Axup et al, 2012). Παρόμοια προσπάθεια έγινε από τους Zimmerman et al, οι οποίοι εισήγαγαν το μη φυσικό αμινοξύ azidomethyl-l-phenylalanine (pAMF) στο αντίσωμα trastuzumab. Το τροποποιημένο αντίσωμα στη συνέχεια συζεύχθηκε με το φάρμακο DBCO-PEG-monomethyl auristatin (DBCO-PEG-MMAF) (Zimmerman et al., 2014).

## **2.3. Στοιχεία Βιοπληροφορικής**

### **2.3.1. Υπολογιστική τυποποίηση δεδομένων τρισδιάστατης δομής πρωτεϊνών**

Η Protein Data Bank (PDB) είναι μία βάση δεδομένων με πειραματικά καθορισμένες τρισδιάστατες δομές βιολογικών μακρομορίων. Τα δεδομένα που περιέχονται στη βάση δεδομένων συμπεριλαμβάνουν ατομικές συντεταγμένες, κρυσταλλογραφικούς δομικούς παράγοντες και NMR πειραματικά δεδομένα. Επίσης, κάθε καταχώρηση συμπεριλαμβάνει τα ονόματα των μορίων, πληροφορία για την πρωτοταγή και δευτεροταγή δομή, αναφορές σε ακολουθιακές βάσεις δεδομένων, πληροφορία σχετικά με πρόσδεση μορίων, λεπτομέρειες σχετικά με τη συλλογή δεδομένων και βιβλιογραφικές αναφορές.

Ένα αρχείο PDB αποτελείται από ένα σύνολο γραμμών, κάθε μία από τις οποίες αποτελείται από 80 στήλες. Οι πρώτες έξι στήλες κάθε γραμμής περιέχουν τον τύπο της εγγραφής. Υπάρχουν δεκάδες τύποι εγγραφών, οι οποίες κατηγοριοποιούνται σε 10 τμήματα με κριτήριο το περιεχόμενό τους, κάποια από τα οποία αναφέρονται ενδεικτικά. Το τμήμα τίτλου (title section) περιέχει εγγραφές σχετικά με το πείραμα και τα βιολογικά μακρομόρια που είναι παρόντα στην καταχώρηση. Το τμήμα πρωτοταγής δομής (primary structure section) περιέχει την ακολουθία αμινοξέων σε κάθε αλυσίδα του μακρομορίου. Το τμήμα ετερογενών στοιχείων (heterogen section) περιέχει την περιγραφή των non-standard αμινοξέων στην καταχώρηση. Το τμήμα δευτεροταγούς δομής (secondary structure section) περιέχει πληροφορίες για τους έλικες και τις πτυχωτές επιφάνειες που βρίσκονται στην πρωτεΐνη. Το τμήμα υποσημειώσεων συνδεσιμότητας (connectivity annotation section) προσδιορίζει την ύπαρξη και τοποθεσία των δισουλφιδικών και άλλων δεσμών. Το τμήμα συντεταγμένων (coordinate section) εμπεριέχει κυρίως τις ατομικές συντεταγμένες της πρωτεΐνης. Το τμήμα συνδεσιμότητας (connectivity section) περιέχει πληροφορίες σχετικά με τη συνδεσιμότητα των ατόμων.

Το format PDBML (Protein Data Bank Markup Language) παρέχει την αναπαράσταση των δεδομένων PDB στο format XML (Westbrook et al., 2005). Η PDB χρησιμοποιεί το format mmCIF για να περιγράψει το περιεχόμενο των καταχωρήσεων PDB. Το format PDBx/mmCIF είναι το καθιερωμένο PDB archive format από το 2014. Μεγάλες μοριακές δομές που περιέχουν >62 αλυσίδες ή και >99999 ATOM records δεν μπορούν να αναπαρασταθούν πλήρως στο PDB File Format, αλλά είναι διαθέσιμες ως μεμονωμένα PDBx/mmCIF αρχεία (“PDBx/mmCIF Dictionary Resources”).

### 2.3.2. Βάσεις δεδομένων

Οι βάσεις δεδομένων πρωτεϊνών περιέχουν τις δομές πρωτεϊνών που έχουν καθοριστεί πειραματικά και παρέχουν χρήσιμες πληροφορίες που συσχετίζουν την τρισδιάστατη δομή με τη βιολογική λειτουργία. Ένα αυξανόμενο σύνολο πειραματικών δεδομένων υποστηρίζει την ιδέα ότι η δομή μίας πρωτεΐνης αντανακλά το ρόλο της και καθορίζει τη βιολογική λειτουργία της.

Το αρχείο PDB περιέχει πληροφορίες σχετικά με τις τρισδιάστατες δομές βιολογικών μακρομορίων, όπως πρωτεϊνών και νουκλεϊκών οξέων, οι οποίες έχουν καθοριστεί πειραματικά με κρυσταλλογραφία ακτίνων X, NMR και τεχνικές ηλεκτρονικού μικροσκοπίου. Το αρχείο PDB είναι μία συλλογή αρχείων σε 3 διαφορετικά format: το PDB format, το PDB exchange format και το PDBML/XML format. Το worldwide Protein Data Bank (wwPDB) είναι η διεθνής συνεργασία που διαχειρίζεται την κατάθεση, επεξεργασία και διανομή του αρχείου PDB, μέλη της οποίας είναι οι οργανισμοί RCSB PDB (ΗΠΑ), PDBe (Ευρώπη), PDBj (Ιαπωνία) και BMRB (Biological Magnetic Resonance Data Bank, ΗΠΑ) (Berman et al., 2000; Berman et al., 2007; Berman et al., 2012).

Άλλες, πιο εξειδικευμένες βάσεις δεδομένων με δομική πληροφορία είναι η SCOP (Structural Classification of Proteins) και η εξέλιξή της, SCOP2, οι οποίες έχουν ως αντικείμενο την κατηγοριοποίηση των πρωτεϊνών με γνωστή τρισδιάστατη δομή με κριτήριο τις εξελικτικές και δομικές συσχετίσεις τους (Andreeva et al., 2008, Andreeva et al., 2014). Επίσης, οι βάσεις δεδομένων Phospho3D (Zanzoni et al., 2007), Phospho3D 2.0 (Zanzoni et

al., 2011) περιέχουν τις τρισδιάστατες δομές των θέσεων φωσφορυλίωσης των πρωτεϊνών (Zhang et al., 2011).

### 2.3.3. *In silico* σχεδιασμός φαρμάκων

Οι υπολογιστικές τεχνικές βιοπληροφορικής και χημειοπληροφορικής αξιοποιούνται πλέον σε όλα τα στάδια της ανακάλυψης και ανάπτυξης νέων φαρμάκων. Η χρήση των υπολογιστών στη φαρμακευτική έρευνα επιτρέπει τη μείωση του κόστους και του χρόνου που απαιτείται για να προκύψουν νέα υποψήφια φάρμακα με αυξημένες πιθανότητες να επιτύχουν στα στάδια των κλινικών δοκιμών.

Στο πεδίο του *in silico* σχεδιασμού φαρμάκων, μοριακός στόχος του φαρμάκου είναι ένα μακρομόριο (π.χ. μία πρωτεΐνη) του οργανισμού που συνδέεται με μία ασθένεια και προσδένεται σε ένα μικρό μόριο για να εκτελέσει μία λειτουργία. Το μακρομόριο αυτό συνήθως έχει μία καλά καθορισμένη περιοχή πρόσδεσης (binding pocket). Το φάρμακο είναι ένα μικρό μόριο που προσδένεται στην περιοχή πρόσδεσης του στόχου, με αποτέλεσμα την εκτέλεση μίας ενέργειας που συμβάλλει στην αντιμετώπιση της ασθένειας. Η περιοχή πρόσδεσης του μακρομορίου ονομάζεται αλλιώς υποδοχέας (receptor) και το μικρό μόριο προσδέτης (ligand). Ο σχεδιασμός φαρμάκων με τη βοήθεια του υπολογιστή μπορεί να πραγματοποιηθεί με την τεχνική της εικονικής διαλογής και την τεχνική του *de novo* σχεδιασμού προσδετών.

### Εικονική Διαλογή

Σύμφωνα με τη μέθοδο της εικονικής διαλογής (virtual screening), μικρά μόρια λημμένα από βάσεις δεδομένων προσαρμόζονται στην περιοχή ενδιαφέροντος, δηλαδή την περιοχή πρόσδεσης ενός μακρομορίου, και βαθμολογούνται με βάση τις προβλεπόμενες αλληλεπιδράσεις με την περιοχή αυτή. Τα μικρά μόρια για τα οποία προκύπτει ότι έχουν αυξημένη σταθερότητα πρόσδεσης με την πρωτεΐνη ενδείκνυνται για περαιτέρω διερεύνηση ως πιθανοί θεραπευτικοί παράγοντες, διαφορετικά απορρίπτονται. (Anderson, 2003; Trott et al., 2010)

Η εικονική διαλογή στηρίζεται στη μέθοδο του docking, δηλαδή την υπολογιστική διαδικασία προσαρμογής ενός προσδέτη στην περιοχή πρόσδεσης ενός μακρομορίου. Αποτέλεσμα της διαδικασίας docking είναι η βέλτιστη θέση πρόσδεσης του προσδέτη στον υποδοχέα, καθώς και η τιμή μίας συνάρτησης scoring που αντιπροσωπεύει τη σταθερότητα της πρόσδεσης. Τις τελευταίες δεκαετίες έχουν αναπτυχθεί διάφορα διαθέσιμα προγράμματα λογισμικού docking. Ένα από αυτά είναι το Autodock (Morris et al., 1996, Morris et al., 1998), το οποίο στηρίχθηκε σε έναν γενετικό αλγόριθμο και μία εμπειρική συνάρτηση πρόσδεσης (empirical binding free energy function). Το Autodock Vina αποτελεί μία εξελιγμένη και βελτιωμένη μορφή του, επιτυγχάνοντας τόσο αύξηση της ακρίβειας ως προς την πρόβλεψη της ενέργειας πρόσδεσης όσο και μείωση του υπολογιστικού χρόνου με παραλληλοποίηση (Trott et al., 2010). Άλλο πρόγραμμα είναι το Dock (Meng et al., 1992; Lang et al., 2009), το οποίο αρχικά στηρίχθηκε σε έναν αλγόριθμο γεωμετρικού ταιριάσματος (Meng et al., 1992), υπέστη διάφορες βελτιώσεις, με την τελευταία του επέκταση να αφορά στη μοντελοποίηση της πρόσδεσης μικρών μορίων σε μόρια RNA (Lang et al., 2009). Το λογισμικό FRED (McGann et al., 2011) χρησιμοποιεί μία εξαντλητική μέθοδο αναζήτησης των διαφορετικών πιθανών θέσεων του προσδέτη, αναδεικνύει τις 100 καλύτερες θέσεις και



εκτελεί μία επιπρόσθετη, πιο λεπτομερή αναζήτηση για αυτές. Επιπλέον λογισμικά docking είναι το FLEXX (Kramer et al., 1999) και το Glide (Halgren et al., 2004). Η χρήση λογισμικού docking είναι εφικτή και μέσω webservers, όπως οι SwissDock (Grosdidier et al., 2011), iScreen (Tsai et al., 2011) και HADDOCK (Vries et al., 2010).

### **De novo σχεδιασμός προσδετών**

Ο de novo σχεδιασμός προσδετών έχει ως αντικείμενο την παραγωγή νέων μοριακών δομών με συγκεκριμένες ιδιότητες που θα λειτουργήσουν ως προσδέτες σε μία πρωτεΐνη. Πιο συγκεκριμένα, με δεδομένη μία πρωτεΐνη, η τεχνική de novo σχεδιασμού εξάγει τους δομικούς περιορισμούς που επιβάλλονται από την περιοχή πρόσδεσης. Με βάση αυτά τα χαρακτηριστικά, το πρόγραμμα προσπαθεί να συνθέσει το βέλτιστο προσδέτη κατασκευάζοντας ένα πλήθος μοριακών δομών, αξιολογώντας τις και επιλέγοντας τις κατάλληλες. Στην πραγματικότητα, τα προγράμματα de novo φαρμακευτικού σχεδιασμού δεν εκτελούν εξαντλητική αναζήτηση των πιθανών μοριακών δομών, και συγκλίνουν σε μία τοπικά, αντί ολικά, βέλτιστη λύση. Επίσης, αρκετά προγράμματα αυτού του είδους δεν είναι ντετερμινιστικά. Οι υποψήφιες μοριακές δομές κατασκευάζονται χρησιμοποιώντας δομικά στοιχεία, που μπορεί να είναι είτε άτομα είτε χημικά τμήματα. Οι μέθοδοι αναζήτησης νέων δομών που έχουν υιοθετηθεί συμπεριλαμβάνουν την αναζήτηση κατά βάθος ή πλάτος, την αναζήτηση Monte Carlo, τους εξελικτικούς αλγόριθμους, και τη Βελτιστοποίηση Σμήνους Σωματιδίων (Particle Swarm Optimization) (Schneider et al., 2005; Schneider et al., 2009).

Μία πληθώρα προγραμμάτων de novo φαρμακευτικού σχεδιασμού έχουν αναπτυχθεί. Το πρόγραμμα LEA3D (Douguet et al., 2005) χρησιμοποιεί έναν γενετικό αλγόριθμο για την αναζήτηση νέων δομών και χημικά τμήματα ως δομικά στοιχεία. Το πρόγραμμα LigBuilder (Wang et al., 2000) χρησιμοποιεί τις στρατηγικές growing και linking για τη σύνθεση προσδετών, οι οποίες υλοποιούνται επίσης με γενετικό αλγόριθμο, και αντλεί δομικά στοιχεία από μία βιβλιοθήκη οργανικών τμημάτων. Το LigBuilder 2 (Yuan et al., 2011) αποτελεί την εξέλιξη του, συμπεριλαμβάνοντας βελτιώσεις του αρχικού προγράμματος, καθώς και μία νέα στρατηγική σχεδιασμού που δεν απαιτεί από τον χρήστη την παροχή μίας αρχικής δομής για τον προσδέτη. Παλαιότερα προγράμματα de novo σχεδιασμού προσδετών είναι τα CONCERTS (Pearlman et al., 1995), GrowMol (Bohacek et al., 1994), SPROUT (Gillet et al., 1993), Pro-Ligand (Clark et al., 1995), HOOK (Eisen et al., 1994).

### **Μοριακή Δυναμική**

Οι μέθοδοι μοριακής δυναμικής (Molecular Dynamics) επικεντρώνονται στην προσομοίωση της κίνησης των ατόμων ενός βιολογικού συστήματος στο χρόνο. Βασίζονται στην εξίσωση κίνησης του Νεύτωνα,  $F = ma$ , όπου  $F$  είναι η δύναμη που ασκείται σε ένα σωματίδιο,  $m$  η μάζα του και  $a$  η επιτάχυνσή του. Με δεδομένη μία αρχική διάταξη του μοριακού συστήματος, υπολογίζεται η συνολική δύναμη που ασκείται σε κάθε άτομο, με βάση την οποία υπολογίζεται η νέα θέση του. Η διαδικασία επαναλαμβάνεται για όλο το σύστημα σε χρονικά διαστήματα 1-4 femtoseconds (fs,  $10^{-15}$  s). Η δύναμη που ασκείται σε ένα άτομο υπολογίζεται διαφοροποιώντας τη δυναμική ενέργεια στη θέση του. Η δυναμική ενέργεια σε κάθε θέση εξαρτάται από όλα τα άλλα άτομα του μοριακού συστήματος (Friedman et al., 2013).

Στο πεδίο του υπολογιστικού σχεδιασμού φαρμάκων, οι προσομοιώσεις μοριακής δυναμικής χρησιμοποιούνται για τη διερεύνηση της ορθότητας των διαφορετικών τρόπων πρόσδεσης ενός μορίου φαρμάκου σε μία πρωτεΐνη, καθώς και τη διερεύνηση της αλληλεπίδρασης μεταξύ των διαφορετικών δυνάμεων στην πρόσδεση. Προγράμματα μοριακής δυναμικής είναι τα Gromacs (Abraham et al., 2015; Pronk et al., 2013; Spoel et al., 2011; Hess et al., 2008; Spoel et al., 2005; Berendsen et al., 1995), NAMD (Phillips et al., 2005), Amber (Case et al., 2005) και CHARMM (Brooks et al., 2009; Vanommeslaeghe et al., 2009), τα οποία επωφελούνται ιδιαίτερα ως προς τη χρονική απόδοσή τους από την παραλληλοποίηση των υπολογιστικών διεργασιών τους. Επίσης, προσομοιώσεις μοριακής δυναμικής είναι εφικτές και με τα πακέτα μοριακής μοντελοποίησης MOE (Chemical Computing Group), Discovery Studio (Accelrys), και SYBYL-X (Tripos).

#### 2.4. Αναφορές

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1, 19–25. <http://doi.org/10.1016/j.softx.2015.06.001>
- Adler, M. J., & Dimitrov, D. S. (2012). Therapeutic antibodies against cancer. *Hematology/oncology Clinics of North America*, 26(3), 447–81, vii. <http://doi.org/10.1016/j.hoc.2012.02.013>
- Alley, S. C., Zhang, X., Okeley, N. M., Anderson, M., Law, C.-L., Senter, P. D., & Benjamin, D. R. (2009). The pharmacologic basis for antibody-auristatin conjugate activity. *The Journal of Pharmacology and Experimental Therapeutics*, 330(3), 932–8. <http://doi.org/10.1124/jpet.109.155549>
- Anderson, A. C. (2003). The Process of Structure-Based Drug Design. *Chemistry & Biology*, 10(9), 787–797. <http://doi.org/10.1016/j.chembiol.2003.09.002>
- Andreeva, A., Howorth, D., Chandonia, J.-M., Brenner, S. E., Hubbard, T. J. P., Chothia, C., & Murzin, A. G. (2008). Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Research*, 36(Database issue), D419–25. <http://doi.org/10.1093/nar/gkm993>
- Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., & Murzin, A. G. (2014). SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Research*, 42(Database issue), D310–4. <http://doi.org/10.1093/nar/gkt1242>
- Axup, J. Y., Bajjuri, K. M., Ritland, M., Hutchins, B. M., Kim, C. H., Kazane, S. A., ... Schultz, P. G. (2012). Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, 109(40), 16101–6. <http://doi.org/10.1073/pnas.1211023109>
- Balendiran, G. K., Dabur, R., & Fraser, D. (2004). The role of glutathione in cancer. *Cell Biochemistry and Function*, 22(6), 343–352. <http://doi.org/10.1002/cbf.1149>
- Barrett, K. E., Barman, S. M., Boitano, S., & Brooks, H. L. (n.d.). *Ganong's Review of Medical Physiology* (24th ed.).
- Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1), 43–56. [http://doi.org/10.1016/0010-4655\(95\)00042-E](http://doi.org/10.1016/0010-4655(95)00042-E)

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. <http://doi.org/10.1093/nar/28.1.235>
- Berman, H., Henrick, K., Nakamura, H., & Markley, J. L. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 35(Database issue), D301–3. <http://doi.org/10.1093/nar/gkl971>
- Berman, H. M., Kleywegt, G. J., Nakamura, H., & Markley, J. L. (2012). The Protein Data Bank at 40: reflecting on the past to prepare for the future. *Structure (London, England : 1993)*, 20(3), 391–6. <http://doi.org/10.1016/j.str.2012.01.010>
- Bohacek, R. S., & McMartin, C. (1994). Multiple Highly Diverse Structures Complementary to Enzyme Binding Sites: Results of Extensive Application of a de Novo Design Method Incorporating Combinatorial Growth. *Journal of the American Chemical Society*, 116(13), 5560–5571. <http://doi.org/10.1021/ja00092a006>
- Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., ... Karplus, M. (2009). CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, 30(10), 1545–614. <http://doi.org/10.1002/jcc.21287>
- Brun, M.-P., & Gauzy-Lazo, L. (2013). Protocols for lysine conjugation. *Methods in Molecular Biology (Clifton, N.J.)*, 1045, 173–87. [http://doi.org/10.1007/978-1-62703-541-5\\_10](http://doi.org/10.1007/978-1-62703-541-5_10)
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., ... Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16), 1668–1688. <http://doi.org/10.1002/jcc.20290>
- Charlton, P., Spicer, J., Druker, B. J., Talpaz, M., Resta, D. J., al., et, ... al., et. (2016). Targeted therapy in cancer. *Medicine*, 44(1), 34–38. <http://doi.org/10.1016/j.mpmed.2015.10.012>
- Cruse, J. M., & Lewis, R. E. (Robert E. (2010). *Atlas of immunology*. CRC Press/Taylor & Francis.
- Clark, D. E., Frenkel, D., Levy, S. A., Li, J., Murray, C. W., Robson, B., ... Westhead, D. R. (1995). PRO-LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. *Journal of Computer-Aided Molecular Design*, 9(1), 13–32. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7751867>
- De Vries, S. J., van Dijk, M., & Bonvin, A. M. J. J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, 5(5), 883–897. <http://doi.org/10.1038/nprot.2010.32>
- Donk, van de. (2012). Brentuximab vedotin. *mAbs*, 4(4), 458–465. <http://doi.org/http://dx.doi.org/10.4161/mabs.20776>
- Doronina, S. O., Toki, B. E., Torgov, M. Y., Mendelsohn, B. A., Cerveny, C. G., Chace, D. F., ... Senter, P. D. (2003). Development of potent monoclonal antibody auristatin conjugates for cancer therapy. *Nature Biotechnology*, 21(7), 778–784. <http://doi.org/10.1038/nbt832>
- Doronina, S. O., Mendelsohn, B. A., Bovee, T. D., Cerveny, C. G., Alley, S. C., Meyer, D. L., ... Senter, P. D. (2006). Enhanced Activity of Monomethylauristatin F through Monoclonal Antibody Delivery: Effects of Linker Technology on Efficacy and Toxicity. *Bioconjugate Chemistry*, 17(1), 114–124. <http://doi.org/10.1021/bc0502917>

- Ducry, L., & Stump, B. (2010). Antibody-drug conjugates: linking cytotoxic payloads to monoclonal antibodies. *Bioconjugate Chemistry*, 21(1), 5–13. <http://doi.org/10.1021/bc9002019>
- Eisen, M. B., Wiley, D. C., Karplus, M., & Hubbard, R. E. (1994). HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins: Structure, Function, and Genetics*, 19(3), 199–221. <http://doi.org/10.1002/prot.340190305>
- Erickson, H. K. (2006). Antibody-Maytansinoid Conjugates Are Activated in Targeted Cancer Cells by Lysosomal Degradation and Linker-Dependent Intracellular Processing. *Cancer Research*, 66(8), 4426–4433. <http://doi.org/10.1158/0008-5472.CAN-05-4489>
- Friedman, R., Boye, K., & Flatmark, K. (2013). Molecular modelling and simulations in cancer research. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1836(1), 1–14. <http://doi.org/10.1016/j.bbcan.2013.02.001>
- Gerber, H.-P., Senter, P. D., & Grewal, I. S. (2009). Antibody drug-conjugates targeting the tumor vasculature: Current and future developments. *mAbs*, 1(3), 247–53.
- Gillet, V., Johnson, A. P., Mata, P., Sike, S., & Williams, P. (1993). SPROUT: a program for structure generation. *Journal of Computer-Aided Molecular Design*, 7(2), 127–53. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8320553>
- Goldmacher, V. S., Chittenden, T., Chari, R. V. J., Kovtun, Y. V., & Lambert, J. M. (2012). Antibody-drug conjugates for targeted cancer therapy. In M. C. Desai (Ed.), *Annual Reports in Medicinal Chemistry, Volume 47* (pp. 349–366). Academic Press. Retrieved from <https://books.google.com/books?hl=en&lr=&id=xpe9Mk50BGsC&pgis=1>
- Greenhalgh, T. A., & Symonds, R. P. (2014). Principles of chemotherapy and radiotherapy. *Obstetrics, Gynaecology & Reproductive Medicine*, 24(9), 259–265. <http://doi.org/10.1016/j.ogrm.2014.06.004>
- Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(Web Server issue), W270-7. <http://doi.org/10.1093/nar/gkr366>
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, 47(7), 1750–1759. <http://doi.org/10.1021/jm030644s>
- Hamblett, K. J., Senter, P. D., Chace, D. F., Sun, M. M. C., Lenox, J., Cervený, C. G., ... Francisco, J. A. (2004). Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 10(20), 7063–70. <http://doi.org/10.1158/1078-0432.CCR-04-0789>
- Hess, B., Kutzner, C., Spoel, D. van der, & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. <http://doi.org/10.1021/CT700301Q>
- Hickman, A. B., Davies, D. R., Hickman, A. B., & Davies, D. R. (1997). Principles of Macromolecular X-Ray Crystallography. In *Current Protocols in Protein Science* (p. 17.3.1-17.3.15). Hoboken, NJ, USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/0471140864.ps1703s10>

- Hofer, T., Skeffington, L. R., Chapman, C. M., & Rader, C. (2009). Molecularly defined antibody conjugation through a selenocysteine interface. *Biochemistry*, *48*(50), 12047–57. <http://doi.org/10.1021/bi901744t>
- Hofer, T., Thomas, J. D., Burke, T. R., & Rader, C. (2008). An engineered selenocysteine defines a unique class of antibody derivatives. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(34), 12451–6. <http://doi.org/10.1073/pnas.0800800105>
- Iyer, U., & Kadambi, V. J. (2011). Antibody drug conjugates - Trojan horses in the war on cancer. *Journal of Pharmacological and Toxicological Methods*, *64*(3), 207–12. <http://doi.org/10.1016/j.vascn.2011.07.005>
- Jain, N., Smith, S. W., Ghone, S., & Tomczuk, B. (2015). Current ADC Linker Chemistry. *Pharmaceutical Research*, *32*(11), 3526–40. <http://doi.org/10.1007/s11095-015-1657-7>
- Johansson, L., Chen, C., Thorell, J.-O., Fredriksson, A., Stone-Elander, S., Gafvelin, G., & Arnér, E. S. J. (2004). Exploiting the 21st amino acid—purifying and labeling proteins by selenolate targeting. *Nature Methods*, *1*(1), 61–66. <http://doi.org/10.1038/nmeth707>
- Joosten, R. P., te Beek, T. A. H., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R., ... Vriend, G. (2011). A series of PDB related databases for everyday needs. *Nucleic Acids Research*, *39*(Database issue), D411-9. <http://doi.org/10.1093/nar/gkq1105>
- Junutula, J. R., Flagella, K. M., Graham, R. A., Parsons, K. L., Ha, E., Raab, H., ... Sliwkowski, M. X. (2010). Engineered thio-trastuzumab-DM1 conjugate with an improved therapeutic index to target human epidermal growth factor receptor 2-positive breast cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *16*(19), 4769–78. <http://doi.org/10.1158/1078-0432.CCR-10-0987>
- Junutula, J. R., Raab, H., Clark, S., Bhakta, S., Leipold, D. D., Weir, S., ... Mallet, W. (2008). Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotechnology*, *26*(8), 925–32. <http://doi.org/10.1038/nbt.1480>
- Koblinski, J. E., Ahram, M., & Sloane, B. F. (2000). Unraveling the role of proteases in cancer. *Clinica Chimica Acta*, *291*(2), 113–135. [http://doi.org/10.1016/S0009-8981\(99\)00224-7](http://doi.org/10.1016/S0009-8981(99)00224-7)
- Kontermann, R. (2012). Dual targeting strategies with bispecific antibodies. *mAbs*, *4*(2), 182–197. <http://doi.org/10.4161/mabs.4.2.19000>
- Kontermann, R. E., & Brinkmann, U. (2015). Bispecific antibodies. *Drug Discovery Today*, *20*(7), 838–847. <http://doi.org/10.1016/j.drudis.2015.02.008>
- Kovtun, Y. V., & Goldmacher, V. S. (2007). Cell killing by antibody-drug conjugates. *Cancer Letters*, *255*(2), 232–40. <http://doi.org/10.1016/j.canlet.2007.04.010>
- Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins*, *37*(2), 228–41. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10584068>
- Kubota, T., Niwa, R., Satoh, M., Akinaga, S., Shitara, K., & Hanai, N. (2009). Engineered therapeutic antibodies with improved effector functions. *Cancer Science*, *100*(9), 1566–1572. <http://doi.org/10.1111/j.1349-7006.2009.01222.x>

- Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., ... Kuntz, I. D. (2009). DOCK 6: Combining techniques to model RNA-small molecule complexes. *RNA*, *15*(6), 1219–1230. <http://doi.org/10.1261/rna.1563609>
- Li, J., Chen, F., Cona, M. M., Feng, Y., Himmelreich, U., Oyen, R., ... Ni, Y. (2012). A review on various targeted anticancer therapies. *Targeted Oncology*, *7*(1), 69–85. <http://doi.org/10.1007/s11523-012-0212-2>
- Li, X., Yang, J., & Rader, C. (2014). Antibody conjugation via one and two C-terminal selenocysteines. *Methods (San Diego, Calif.)*, *65*(1), 133–8. <http://doi.org/10.1016/j.ymeth.2013.05.023>
- Mack, F., Ritchie, M., & Sapra, P. (2014). The next generation of antibody drug conjugates. *Seminars in Oncology*, *41*(5), 637–52. <http://doi.org/10.1053/j.seminoncol.2014.08.001>
- MacMillan, K. S., & Boger, D. L. (2009). Fundamental Relationships between Structure, Reactivity, and Biological Activity for the Duocarmycins and CC-1065. *Journal of Medicinal Chemistry*, *52*(19), 5771–5780. <http://doi.org/10.1021/jm9006214>
- McGann, M. (2011). FRED Pose Prediction and Virtual Screening Accuracy. *Journal of Chemical Information and Modeling*, *51*(3), 578–596. <http://doi.org/10.1021/ci100436p>
- Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (1992). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, *13*(4), 505–524. <http://doi.org/10.1002/jcc.540130412>
- Morris, G. M., Goodsell, D. S., Huey, R., & Olson, A. J. (1996). Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design*, *10*(4), 293–304. <http://doi.org/10.1007/BF00124499>
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. [http://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](http://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B)
- National Cancer Institute. (n.d.). Retrieved February 21, 2017, from <https://www.cancer.gov/>
- Nolting, B. (2013). Linker technologies for antibody-drug conjugates. *Methods in Molecular Biology (Clifton, N.J.)*, *1045*, 71–100. [http://doi.org/10.1007/978-1-62703-541-5\\_5](http://doi.org/10.1007/978-1-62703-541-5_5)
- PDBx/mmCIF Dictionary Resources. (n.d.). Retrieved February 21, 2017, from <http://mmcif.wwpdb.org/>
- Pearlman, D. A., & Murcko, M. A. (1996). CONCERTS: Dynamic Connection of Fragments as an Approach to *de Novo* Ligand Design. *Journal of Medicinal Chemistry*, *39*(8), 1651–1663. <http://doi.org/10.1021/jm9507921>
- Perez, H. L., Cardarelli, P. M., Deshpande, S., Gangwar, S., Schroeder, G. M., Vite, G. D., & Borzilleri, R. M. (2014). Antibody-drug conjugates: current status and future directions. *Drug Discovery Today*, *19*(7), 869–81. <http://doi.org/10.1016/j.drudis.2013.11.004>
- Peters, C., & Brown, S. (2015). Antibody-drug conjugates as novel anti-cancer chemotherapeutics. *Bioscience Reports*, *35*(4). <http://doi.org/10.1042/BSR20150089>
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–802. <http://doi.org/10.1002/jcc.20289>

- Polson, A. G., Williams, M., Gray, A. M., Fujii, R. N., Poon, K. A., McBride, J., ... Ebens, A. (2010). Anti-CD22-MCC-DM1: an antibody-drug conjugate with a stable linker for the treatment of non-Hodgkin's lymphoma. *Leukemia*, 24(9), 1566–1573. <http://doi.org/10.1038/leu.2010.141>
- Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics (Oxford, England)*, 29(7), 845–54. <http://doi.org/10.1093/bioinformatics/btt055>
- Sapra, P., & Shor, B. (2013). Monoclonal antibody-based therapies in cancer: Advances and challenges. *Pharmacology & Therapeutics*, 138(3), 452–469. <http://doi.org/10.1016/j.pharmthera.2013.03.004>
- Scott C. Jeffrey, \*, Jamie B. Andreyka, Starr X. Bernhardt, Kim M. Kissler, Toni Kline, Joel S. Lenox, ... Senter. (2006). Development and Properties of  $\beta$ -Glucuronide Linkers for Monoclonal Antibody–Drug Conjugates. <http://doi.org/10.1021/BC0600214>
- Senter, P. D., & Springer, C. J. (2001). Selective activation of anticancer prodrugs by monoclonal antibody-enzyme conjugates. *Advanced Drug Delivery Reviews*, 53(3), 247–64. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11744170>
- Shuptrine, C. W., Surana, R., & Weiner, L. M. (2012). Monoclonal antibodies for the treatment of cancer. *Seminars in Cancer Biology*, 22(1), 3–13. <http://doi.org/10.1016/j.semcancer.2011.12.009>
- Smyth, M. S., & Martin, J. H. (2000). x ray crystallography. *Molecular Pathology: MP*, 53(1), 8–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10884915>
- Teillaud, J.-L., Teillaud, & Jean-Luc. (2012). Antibody-dependent Cellular Cytotoxicity (ADCC). In *eLS*. Chichester, UK: John Wiley & Sons, Ltd. <http://doi.org/10.1002/9780470015902.a0000498.pub2>
- Thomas, A., Teicher, B. A., & Hassan, R. (2016). Antibody–drug conjugates for cancer therapy. *The Lancet Oncology*, 17(6), e254–e262. [http://doi.org/10.1016/S1470-2045\(16\)30030-4](http://doi.org/10.1016/S1470-2045(16)30030-4)
- Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), NA-NA. <http://doi.org/10.1002/jcc.21334>
- Tsai, T.-Y., Chang, K.-W., & Chen, C. Y.-C. (2011). iScreen: world's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *Journal of Computer-Aided Molecular Design*, 25(6), 525–531. <http://doi.org/10.1007/s10822-011-9438-9>
- Van der Spoel, D., & Hess, B. (2011). GROMACS-the road ahead. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5), 710–715. <http://doi.org/10.1002/wcms.50>
- Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701–1718. <http://doi.org/10.1002/jcc.20291>
- Van Der Velden, V. H., te Marvelde, J. G., Hoogeveen, P. G., Bernstein, I. D., Houtsmuller, A. B., Berger, M. S., & van Dongen, J. J. (2001). Targeting of the CD33-calicheamicin immunoconjugate Mylotarg (CMA-676) in acute myeloid leukemia: in vivo and in vitro

- saturation and internalization by leukemic and normal myeloid cells. *Blood*, 97(10), 3197–204. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11342449>
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., ... Mackerell, A. D. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4), NA-NA. <http://doi.org/10.1002/jcc.21367>
- Wang, R., Gao, Y., & Lai, L. (2000). LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *Journal of Molecular Modeling*, 6(7–8), 498–516. <http://doi.org/10.1007/s0089400060498>
- Westbrook, J., Ito, N., Nakamura, H., Henrick, K., & Berman, H. M. (2005). PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics*, 21(7), 988–992. <http://doi.org/10.1093/bioinformatics/bti082>
- Wu, A. M., & Senter, P. D. (2005). Arming antibodies: prospects and challenges for immunoconjugates. *Nature Biotechnology*, 23(9), 1137–46. <http://doi.org/10.1038/nbt1141>
- Yap, T. A., & Workman, P. (2012). Exploiting the Cancer Genome: Strategies for the Discovery and Clinical Development of Targeted Molecular Therapeutics. *Annual Review of Pharmacology and Toxicology*, 52(1), 549–573. <http://doi.org/10.1146/annurev-pharmtox-010611-134532>
- Yuan, Y., Pei, J., & Lai, L. (2011). LigBuilder 2: A Practical *de Novo* Drug Design Approach. *Journal of Chemical Information and Modeling*, 51(5), 1083–1091. <http://doi.org/10.1021/ci100350u>
- Zabriskie, J. B. (Ed.). (2009). *Essential Clinical Immunology*. Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511575266>
- Zanzoni, A., Carbajo, D., Diella, F., Gherardini, P. F., Tramontano, A., Helmer-Citterich, M., & Via, A. (2011). Phospho3D 2.0: an enhanced database of three-dimensional structures of phosphorylation sites. *Nucleic Acids Research*, 39(Database), D268–D271. <http://doi.org/10.1093/nar/gkq936>
- Zanzoni, A., Ausiello, G., Via, A., Gherardini, P. F., & Helmer-Citterich, M. (2007). Phospho3D: a database of three-dimensional structures of protein phosphorylation sites. *Nucleic Acids Research*, 35(Database issue), D229–31. <http://doi.org/10.1093/nar/gkl922>
- Zhang, J., Yang, P. L., & Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*, 9(1), 28–39. <http://doi.org/10.1038/nrc2559>
- Zimmerman, E. S., Heibeck, T. H., Gill, A., Li, X., Murray, C. J., Madlansacay, M. R., ... Sato, A. K. (2014). Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system. *Bioconjugate Chemistry*, 25(2), 351–61. <http://doi.org/10.1021/bc400490z>



## Κεφάλαιο 3. Υπολογιστικός σχεδιασμός Συμπλεγμάτων Αντισωμάτων-Φαρμάκων

---

### 3.1. Εισαγωγή

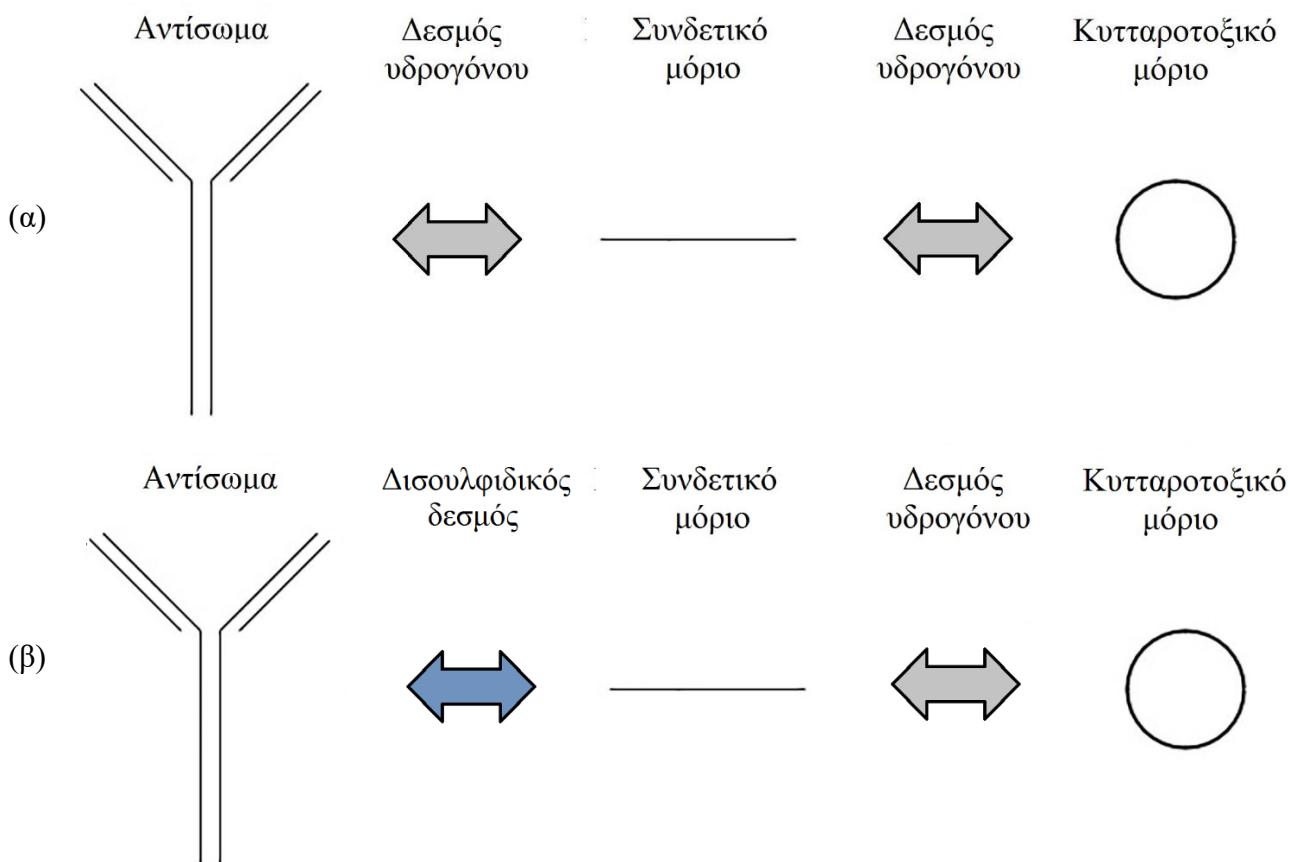
Η διαδικασία ανάπτυξης φαρμάκων είναι πολύπλοκη, με υψηλό χρονικό και οικονομικό κόστος. Κατά μέσο όρο, η ανάπτυξη ενός νέου φαρμάκου απαιτεί 12-15 χρόνια και πάνω από \$1 δις. Τα στάδια φαρμακευτικής ανάπτυξης μπορούν να συνοψιστούν ως ο προσδιορισμός και η επαλήθευση στόχου, η ανακάλυψη φαρμάκου, το screening, η προκλινική δοκιμή, οι κλινικές δοκιμές, και η μετακλινική παρακολούθηση (Hughes et al., 2011; Kessel, 2011).

Οι τεχνικές υπολογιστικού φαρμακευτικού σχεδιασμού αποσκοπούν στη διευκόλυνση της διαδικασίας ανάπτυξης νέων φαρμάκων. Η στρατηγική της εικονικής διαλογής αποσκοπεί στη διερεύνηση της πρόσδεσης ενός συνόλου υποψήφιων φαρμάκων από βάσεις δεδομένων σε μία πρωτεΐνη που σχετίζεται με μία ασθένεια. Η στρατηγική de novo σχεδιασμού αποσκοπεί στη δημιουργία νέων μοριακών δομών που διαθέτουν τις επιθυμητές ιδιότητες για να προσδεθούν στα σημεία πρόσδεσης της πρωτεΐνης. Και στις δύο προσεγγίσεις, η πρωτεΐνη έχει άμεση σχέση με την ασθένεια, και η πρόσδεση ενός φαρμάκου σε μία από τις περιοχές πρόσδεσής της έχει αποτέλεσμα την αντιμετώπιση της ασθένειας (Katsila et al., 2016; Taft et al., 2008).

Σε ένα Σύμπλεγμα Αντισωμάτων-Φαρμάκων (Antibody-Drug Conjugate, ADC), το αντίσωμα, που είναι πρωτεΐνη, έχει τον ρόλο να μεταφέρει το κυτταροτοξικό φάρμακο στα καρκινικά κύτταρα. Οι περιοχές πρόσδεσής του (Complementary-Determining Regions, CDRs), πρέπει να μείνουν ελεύθερες για πρόσδεση στο αντιγόνο-στόχο που εκφράζεται στα καρκινικά κύτταρα, άρα το κυτταροτοξικό φάρμακο πρέπει να προσδεθεί σε κάποιο άλλο σημείο του αντισώματος (Peters et al., 2015). Αποτέλεσμα αυτής της ιδιαιτερότητας είναι ο υπολογιστικός σχεδιασμός Συμπλεγμάτων Αντισωμάτων-Φαρμάκων να μην έχει γίνει εφικτός με τις προαναφερθέντες τεχνικές υπολογιστικού φαρμακευτικού σχεδιασμού.

Στο παρόν κεφάλαιο παρουσιάζεται μία μέθοδος υπολογιστικού σχεδιασμού Συμπλεγμάτων Αντισωμάτων-Φαρμάκων χρησιμοποιώντας δεδομένα από καθιερωμένες βάσεις δεδομένων. Τα αρχεία PDB εισόδου του αντισώματος, του συνδέτη και της κυτταροτοξικής ουσίας υπόκεινται σε επεξεργασία και συγχωνεύονται στο τελικό αρχείο PDB του ADC μορίου. Ειδικότερα, η διάταξη του συνδετικού και του κυτταροτοξικού μορίου μεταβάλλονται χρησιμοποιώντας έναν αφινικό μετασχηματισμό, με στόχο την κατάλληλη τοποθέτησή τους σε σχέση με το αντίσωμα. Η μέθοδος προσαρμόστηκε για δύο τρόπους σύνδεσης του συμπλέγματος συνδέτη-φαρμάκου με το αντίσωμα, συγκεκριμένα για σύζευξη σε λυσίνη και σύζευξη σε τροποποιημένη κυστεΐνη, η οποία έχει αντικαταστήσει ένα εγγενές αμινοξύ στην εξωτερική επιφάνεια του αντισώματος. Η σύνδεση του φαρμάκου με το συνδέτη γίνεται και στις δύο περιπτώσεις με ένα δεσμό υδρογόνου. Στη σύζευξη σε λυσίνες, το σύμπλεγμα συνδέτη-φαρμάκου συνδέεται με μία λυσίνη στην επιφάνεια του αντισώματος με ένα δεσμό υδρογόνου (Εικόνα 3.1.A). Στη σύζευξη σε τροποποιημένες κυστεΐνες, το σύμπλεγμα συνδέτη-φαρμάκου συνδέεται με μία τροποποιημένη κυστεΐνη στην επιφάνεια του αντισώματος με ένα δισουλφιδικό δεσμό (Εικόνα 3.1.B). Η μέθοδος υλοποιήθηκε υπολογιστικά στην προγραμματιστική γλώσσα C++. Η οπτικοποίηση των μορίων, η

προσθήκη μορίων υδρογόνου στα αρχεία αντισωμάτων, καθώς και η αντικατάσταση επιλεγμένων αμινοξέων με κυστεΐνες έγινε με το προγραμματιστικό πακέτο UCSF Chimera (Pettersen et al., 2004).



Εικόνα 3.1. Σχήμα των μορίων και των δεσμών που σχηματίζουν ένα σύμπλεγμα αντισώματος-φαρμάκου (ADC) σύμφωνα με τη μέθοδο που περιγράφεται σε αυτό το κεφάλαιο (α) Σύζευξη σε λυσίνη (β) Σύζευξη σε τροποποιημένη κυστεΐνη.

## 3.2. Μεθοδολογία

### 3.2.1. Περιγραφή δεδομένων

Όλα τα δεδομένα που χρησιμοποιήθηκαν ως είσοδοι της μεθόδου είναι αρχεία σε μορφή PDB, και περιλαμβάνουν αντισώματα ανακτημένα από την RCSB Protein Data Bank (Berman et al., 2012; Berman et al., 2009; Berman, 2007), κυτταροτοξικές ουσίες από το Open National Cancer Institute Database, καθώς και τα μόρια συνδετών  $C_{15}N$ ,  $C_{12}S$  που δημιουργήθηκαν με το λογισμικό μοριακής μοντελοποίησης UCSF Chimera (Pettersen et al., 2004).

Το αντίσωμα επιλέχθηκε ώστε να έχει εκτεθειμένα αμινοξέα τα οποία αφ' ενός δε βρίσκονται στην υπερμεταβλητή περιοχή του αντισώματος, αφ' ετέρου μπορούν να αντικατασταθούν με κυστεΐνες και να λειτουργήσουν ως σημεία σύζευξης. Το φάρμακο επιλέχθηκε ώστε να έχει τουλάχιστον ένα άτομο υδρογόνου (H) συνδεδεμένο ομοιοπολικά με ένα ηλεκτραρνητικό άτομο (EN), ώστε να μπορεί να σχηματίσει δεσμό υδρογόνου με τον συνδέτη.

Το μόριο  $C_{15}N$  επιλέχθηκε ως συνδέτης για τη σύζευξη σε λυσίνες, ενώ το μόριο  $C_{12}S$  επιλέχθηκε για τη σύζευξη σε τροποποιημένες κυστεΐνες. Οι συνδέτες σχεδιάστηκαν ώστε το σύμπλεγμα συνδέτη-φαρμάκου να μπορεί να συζευχθεί με το επιλεγμένο αμινοξύ χωρίς να συγκρούεται με τα γειτονικά αμινοξέα. Για αυτό το λόγο τα άτομα άνθρακα διαμορφώνονται σε γραμμική, αντί κυκλική, διάταξη. Το μόριο συνδέτη  $C_{15}N$  σχεδιάστηκε ώστε να μπορεί να σχηματίσει έναν δεσμό υδρογόνου με μία λυσίνη, επομένως ένα άτομο αζώτου (N) συμπεριλήφθηκε. Το μόριο συνδέτη  $C_{12}S$  σχεδιάστηκε ώστε να μπορεί να σχηματίσει δισουλφιδικό δεσμό με μία κυστεΐνη, επομένως ένα άτομο θείου (S) συμπεριλήφθηκε. Επίσης, ο συνδέτης θα πρέπει να μπορεί να απελευθερώσει το φάρμακο όταν το αντίσωμα προσδεθεί στο αντιγόνο του.

Αν και το PDB format έχει περισσότερους από 40 τύπους καταχωρήσεων, οι τρεις τύποι που είναι κρίσιμοι για την αναπαράσταση τρισδιάστατων μοριακών δομών είναι οι PDB ATOM, HETATM και CONECT. Ο τύπος καταχώρησης PDB ATOM περιέχει πληροφορίες σχετικά με τα άτομα των αμινοξέων ή νουκλεοτιδίων της μοριακής δομής. Ο τύπος καταχώρησης PDB HETATM περιέχει πληροφορία για επιπρόσθετες, μη προδιαγεγραμμένες χημικές δομές του μορίου, και ο τύπος καταχώρησης PDB CONECT περιέχει πληροφορία για τη συνδεσιμότητα μεταξύ των ατόμων του μορίου.

### 3.2.2. Περιγραφή μεθοδολογίας

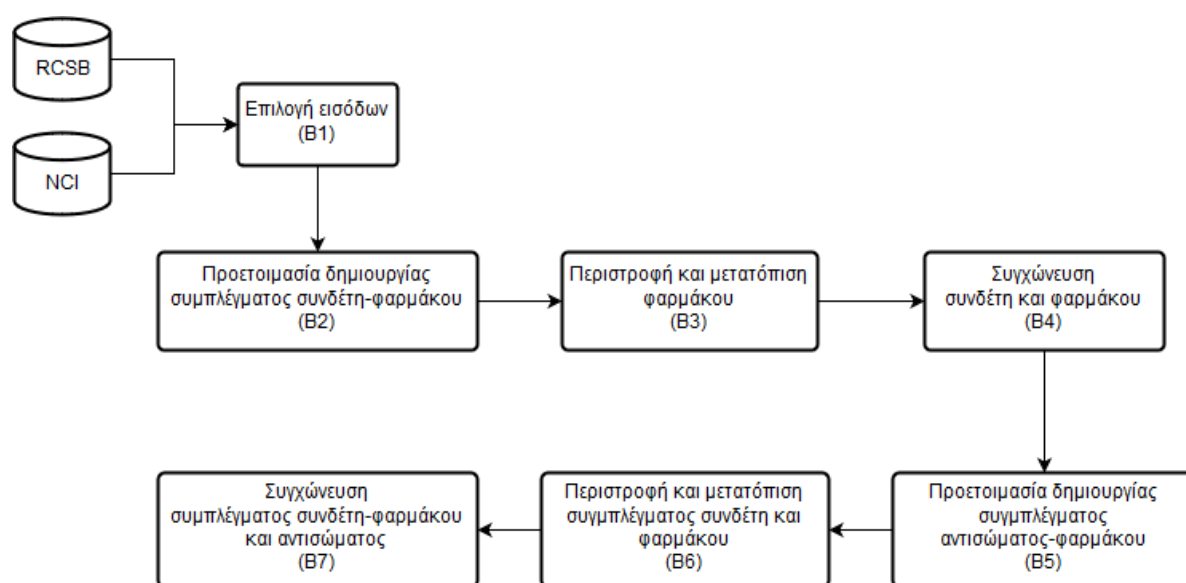
#### Βήματα μεθοδολογίας

Πρώτα, το αντίσωμα, ο κυτταροτοξικός παράγοντας και ο συνδέτης που θα συνθέσουν το σύμπλεγμα καθορίζονται ως είσοδος (βήμα B1). Όταν τα αρχεία εισόδου του αντισώματος, του συνδέτη και της κυτταροτοξικής ουσίας είναι διαθέσιμα, οι απαραίτητες καταχωρήσεις PDB φορτώνονται σε κατάλληλες δομές δεδομένων. Η υπόλοιπη διαδικασία μπορεί να χωριστεί σε δύο βασικά στάδια, που είναι η σύνθεση ενός συμπλέγματος συνδέτη-φαρμάκου από το συνδέτη και το φάρμακο (βήματα B2-B4), και η σύνθεση ενός συμπλέγματος αντισώματος-φαρμάκου από το σύμπλεγμα συνδέτη-φαρμάκου και αντισώματος (βήματα B5-B7). Πιο αναλυτικά, ενώ ο συνδέτης παραμένει σταθερός, το φάρμακο υπόκειται σε περιστροφή και μετατόπιση ώστε να έρθει σε συγκεκριμένη θέση κοντά στον συνδέτη. (βήμα B3). Στη συνέχεια, τα δύο μόρια συνδέονται με ένα δεσμό υδρογόνου και συγχωνεύονται σε ένα σύμπλεγμα συνδέτη-φαρμάκου (βήμα B4). Ακολούθως, το αντίσωμα παραμένει σταθερό και το σύμπλεγμα συνδέτη-φαρμάκου υπόκειται σε περιστροφή και μετατόπιση (βήματα B6). Τα δύο μόρια συνδέονται είτε με δεσμό υδρογόνου (στην περίπτωση της σύζευξης σε λυσίνη), είτε με δισουλφιδικό δεσμό (στην περίπτωση της σύζευξης σε τροποποιημένη κυστεΐνη) και συγχωνεύονται, διαμορφώνοντας το τελικό σύμπλεγμα αντισώματος-φαρμάκου (βήμα B7). Με δεδομένο ότι ένα μόριο φαρμάκου μπορεί να περιέχει πάνω από ένα ζευγάρι ατόμων που να μπορούν να σχηματίσουν δεσμό υδρογόνου με το συνδέτη, είναι εφικτό είτε να επιλεγθεί εξ αρχής το σημείο σύζευξης του φαρμάκου, είτε να παραχθούν περισσότερα από ένα τελικά συμπλέγματα. Τα βήματα B1-B7 της υπολογιστικής μεθόδου παρουσιάζονται παρακάτω, καθώς και στην Εικόνα 3.2.

B1. Επιλογή αντισώματος, φαρμάκου, συνδέτη

B2. Προετοιμασία δημιουργίας συμπλέγματος συνδέτη-φαρμάκου.

- B2α. Καθορισμός αξόνων φαρμάκου και συνδέτη.  
 B2β. Καθορισμός ατόμων που ορίζουν την απόσταση μεταξύ συνδέτη και φαρμάκου.  
 B3. Περιστροφή και μετατόπιση φαρμάκου  
 B4. Συγχώνευση συνδέτη και φαρμάκου  
 B5. Προετοιμασία δημιουργίας συμπλέγματος αντισώματος-φαρμάκου.  
 B5α. Καθορισμός αξόνων συμπλέγματος συνδέτη-φαρμάκου και αμινοξέος.  
 B5β. Καθορισμός ατόμων που ορίζουν την απόσταση μεταξύ συμπλέγματος συνδέτη-φαρμάκου και αμινοξέος.  
 B6. Περιστροφή και μετατόπιση συμπλέγματος συνδέτη-φαρμάκου  
 B7. Συγχώνευση συμπλέγματος συνδέτη-φαρμάκου και αντισώματος



Εικόνα 3.2. Διάγραμμα της υπολογιστικής μεθόδου σύνθεσης Συμπλεγμάτων Αντισωμάτων-Φαρμάκων.

### Ορισμός αξόνων

Ο άξονας κάθε μορίου έχει οριστεί λαμβάνοντας υπόψιν την επιθυμητή τελική διάταξη των τριών μορίων. Ο άξονας  $L_d$  του φαρμάκου ορίζεται από τα άτομα του φαρμάκου που λαμβάνουν μέρος στο δεσμό υδρογόνου με το συνδέτη, τα οποία είναι ένα άτομο υδρογόνου (H) ( $p_{d1}$ ) και το ηλεκτραρνητικό (EN) άτομο ( $p_{d2}$ ) ομοιοπολικά συνδεδεμένο με αυτό (Εξισώσεις 3.1-3.3).

$$p_{d1} = (x, y, z)_{\text{H άτομο φαρμάκου}} = (x_{Hd}, y_{Hd}, z_{Hd}) \quad (3.1)$$

$$p_{d2} = (x, y, z)_{\text{EN άτομο φαρμάκου}} = (x_{END}, y_{END}, z_{END}) \quad (3.2)$$

$$L_d: (x, y, z) = p_{d1} + t_d * (p_{d1} - p_{d2}) = (x_{Hd}, y_{Hd}, z_{Hd}) + t_d * (x_{Hd} - x_{END}, y_{Hd} - y_{END}, z_{Hd} - z_{END}), \quad -\infty < t_d < +\infty \quad (3.3)$$

Ο άξονας  $L_1$  του συνδέτη ορίζεται ως η ευθεία που συνδέει τα δύο άτομα  $p_{11}, p_{12}$  του που αφ' ενός δεν είναι υδρογόνα, αφ' ετέρου έχουν τη μεγαλύτερη απόσταση μεταξύ τους, αφού αυτή η ευθεία είναι αρκετά αντιπροσωπευτική του σχήματος του συνδέτη (Εξισώσεις

3.4-3.6). Στη σύζευξη σε λυσίνες, με δεδομένη τη χημική δομή  $C_{15}N$  του συνδέτη που χρησιμοποιήθηκε στη σύζευξη μέσω λυσινών, τα δύο άτομα που ορίζουν τον άξονα του συνδέτη είναι πάντα ένα από τα ακριανά άτομα άνθρακα (C) που είναι συνδεδεμένα με το άτομο αζώτου (N), έστω  $C_1$ , και το άτομο άνθρακα (C) με τη μεγαλύτερη απόσταση από αυτό. Με δεδομένη τη χημική δομή  $C_{12}S$  του συνδέτη που χρησιμοποιήθηκε στη σύζευξη μέσω κυστεϊνών, τα δύο άτομα που ορίζουν τον άξονα του συνδέτη είναι πάντα το άτομο θείου (S) και το άτομο άνθρακα (C) με τη μεγαλύτερη απόσταση από αυτό.

$$p_{11} = (x, y, z)_{C \text{ άτομο συνδέτη}} = (x_{C1}, y_{C1}, z_{C1}) \quad (3.4)$$

$$p_{12} = \begin{cases} (x, y, z)_{C_1 \text{ άτομο συνδέτη για σύζευξη σε Λυσίνη}} & = \begin{cases} (x_{C1}, y_{C1}, z_{C1}) \\ (x_{S1}, y_{S1}, z_{S1}) \end{cases} \\ (x, y, z)_{S \text{ άτομο συνδέτη για σύζευξη σε Κυστεΐνη}} & \end{cases} \quad (3.5)$$

$$L_1: (x, y, z) = p_{11} + t_1 * (p_{11} - p_{12}) = \begin{cases} (x_{C1}, y_{C1}, z_{C1}) + t_1 * (x_{C1} - x_{C1}, y_{C1} - y_{C1}, z_{C1} - z_{C1}) & \text{για σύζευξη σε Λυσίνη,} \\ (x_{C1}, y_{C1}, z_{C1}) + t_1 * (x_{C1} - x_{S1}, y_{C1} - y_{S1}, z_{C1} - z_{S1}) & \text{για σύζευξη σε Κυστεΐνη,} \end{cases} \quad -\infty < t_1 < +\infty \quad (3.6)$$

Παρομοίως, ο άξονας  $L_{1d}$  του συμπλέγματος συνδέτη-φαρμάκου συνδέει τα δύο πιο απομακρυσμένα άτομα  $p_{1d1}$ ,  $p_{1d2}$  του συμπλέγματος, εξαιρώντας τα υδρογόνα (Εξισώσεις 3.7-3.9).

$$p_{1d1} = (x, y, z)_{NH \text{ άτομο συνδέτη-φαρμάκου}} = (x_{NH1d}, y_{NH1d}, z_{NH1d}) \quad (3.7)$$

$$p_{1d2} = \begin{cases} (x, y, z)_{NH \text{ άτομο συνδέτη-φαρμάκου για σύζευξη σε Λυσίνη}} & = \begin{cases} (x_{NH1d}, y_{NH1d}, z_{NH1d}) \\ (x_{S1d}, y_{S1d}, z_{S1d}) \end{cases} \\ (x, y, z)_{S \text{ άτομο συνδέτη-φαρμάκου για σύζευξη σε Κυστεΐνη}} & \end{cases} \quad (3.8)$$

$$L_{1d}: (x, y, z) = p_{1d1} + t_{1d} * (p_{1d1} - p_{1d2}) = \begin{cases} (x_{NH1d}, y_{NH1d}, z_{NH1d}) + t_{1d} * (x_{NH1d} - x_{S1d}, y_{NH1d} - y_{S1d}, z_{NH1d} - z_{S1d}) & \text{για σύζευξη σε Λυσίνη} \\ (x_{NH1d}, y_{NH1d}, z_{NH1d}) + t_{1d} * (x_{NH1d} - x_{S1d}, y_{NH1d} - y_{S1d}, z_{NH1d} - z_{S1d}) & \text{για σύζευξη σε Κυστεΐνη} \end{cases} \quad -\infty < t_{1d} < +\infty \quad (3.9)$$

Ο άξονας  $L_{AA}$  του αμινοξέος στο οποίο θα γίνει η σύζευξη επιλέχθηκε ώστε να έχει κατεύθυνση προς το εξωτερικό του αμινοξέος. Συγκεκριμένα, ο άξονας της λυσίνης ορίστηκε ως η ευθεία που διέρχεται από το άτομο άλφα άνθρακα ( $C_\alpha$ ) και το άτομο αζώτου (N) της παράπλευρης αλυσίδας της λυσίνης. Ο άξονας της κυστεΐνης ορίστηκε ως η ευθεία γραμμή που διέρχεται από τα άτομα άλφα άνθρακα ( $C_\alpha$ ) και θείου (S) της παράπλευρης αλυσίδας της κυστεΐνης (Εξισώσεις 3.10-3.12).

$$p_{AA1} = (x, y, z)_{C_\alpha \text{ άτομο αμινοξέος (Lys or Cys)}} = (x_{Ca}, y_{Ca}, z_{Ca}) \quad (3.10)$$

$$p_{AA2} = \begin{cases} (x, y, z)_N \text{ άτομο Λυσίνης} & = \begin{cases} (x_{NLys}, y_{NLys}, z_{NLys}) & \text{για σύζευξη σε Λυσίνη} \\ (x_{SCys}, y_{SCys}, z_{SCys}) & \text{για σύζευξη σε Κυστεΐνη} \end{cases} \\ (x, y, z)_S \text{ άτομο Κυστεΐνης} & \end{cases} \quad (3.11)$$

$$L_{AA}: (x, y, z) = p_{AA1} + t_{AA} * (p_{AA1} - p_{AA2}) = \begin{cases} (x_{Ca}, y_{Ca}, z_{Ca}) + t_{AA} * (x_{Ca} - x_{NLys}, y_{Ca} - y_{NLys}, z_{Ca} - z_{NLys}) & \text{για σύζευξη σε Λυσίνη} \\ (x_{Ca}, y_{Ca}, z_{Ca}) + t_{AA} * (x_{Ca} - x_{SCys}, y_{Ca} - y_{SCys}, z_{Ca} - z_{SCys}) & \text{για σύζευξη σε Κυστεΐνη} \end{cases} \quad -\infty < t_{AA} < +\infty \quad (3.12)$$

## Ορισμός αποστάσεων

Λαμβάνοντας υπόψιν την επιθυμητή συνδεσιμότητα μεταξύ των τριών μορίων, η απόσταση μεταξύ δύο μορίων A και B ορίστηκε ως η απόσταση μεταξύ των ατόμων που συμμετέχουν στον επιθυμητό δεσμό μεταξύ τους. Η απόσταση  $D_{d-1}$  μεταξύ του φαρμάκου και του συνδέτη ορίστηκε ως η απόσταση μεταξύ του ατόμου υδρογόνου  $p_d$  του φαρμάκου και του ηλεκτραρνητικού ατόμου  $p_{11}$  του συνδέτη που συμμετέχουν στο δεσμό υδρογόνου μεταξύ τους (Εξισώσεις 3.13-3.15).

$$p_d = (x, y, z)_H \text{ άτομο φαρμάκου} = (x_{Hd}, y_{Hd}, z_{Hd}) = p_{d1} \quad (3.13)$$

$$p_{11} = (x, y, z)_{EN} \text{ άτομο συνδέτη} = (x_{EN1}, y_{EN1}, z_{EN1}) = (x_{Cl}, y_{Cl}, z_{Cl}) \quad (3.14)$$

$$D_{d-1} = \|p_d - p_{11}\| \quad (3.15)$$

Η απόσταση  $D_{AA-1d}$  μεταξύ του συμπλέγματος συνδέτη-φαρμάκου και του αμινοξέος σύζευξης ορίστηκε ως η απόσταση μεταξύ του ατόμου  $p_{1d}$  του συνδέτη και του ατόμου  $p_{AA}$  της παράπλευρης αλυσίδας του αμινοξέος που συμμετέχουν στο δεσμό μεταξύ τους (Εξισώσεις 3.16-3.18). Στη σύζευξη σε λυσίνες, αυτά είναι το άτομο αζώτου του συνδέτη και το άτομο αζώτου της λυσίνης, που θα συνδεθούν με δεσμό υδρογόνου. Στη σύζευξη σε τροποποιημένες κυστεΐνες, αυτά είναι το άτομο θείου του συνδέτη και το άτομο θείου της κυστεΐνης, που θα συνδεθούν με δισουλφιδικό δεσμό.

$$p_{1d} = \begin{cases} (x, y, z)_N \text{ άτομο συνδέτη για σύζευξη σε Λυσίνη} \\ (x, y, z)_S \text{ άτομο συνδέτη για σύζευξη σε Κυστεΐνη} \end{cases} = \begin{cases} (x_{N1}, y_{N1}, z_{N1}) \\ (x_{S1}, y_{S1}, z_{S1}) \end{cases} \quad (3.16)$$

$$p_{AA} = \begin{cases} (x, y, z)_N \text{ άτομο Λυσίνης για σύζευξη σε Λυσίνη} \\ (x, y, z)_S \text{ άτομο Κυστεΐνης για σύζευξη σε Κυστεΐνη} \end{cases} = \begin{cases} (x_{Nlys}, y_{Nlys}, z_{Nlys}) \\ (x_{Scys}, y_{Scys}, z_{Scys}) \end{cases} \quad (3.17)$$

$$D_{AA-1d} = \|p_{AA} - p_{1d}\| \quad (3.18)$$

## Αφινικός Μετασχηματισμός

Η τροποποίηση των ατομικών συντεταγμένων στοχεύει στην μετατροπή των σχετικών θέσεων μεταξύ των διαφορετικών μορίων. Παρ' όλα αυτά, οι ευθείες και οι αποστάσεις μεταξύ των ατόμων ενός μορίου πρέπει να παραμένουν αμετάβλητες, ώστε να μην γίνεται επέμβαση στην εγκυρότητα της χημικής δομής του. Για αυτό το λόγο η αναδιάταξη ενός μορίου στο χώρο έγινε με τη χρήση ενός αφινικού μετασχηματισμού. Ο μετασχηματισμός ενός άκαμπτου σώματος στον τρισδιάστατο χώρο μπορεί να περιγραφεί με τις ακόλουθες εξισώσεις ("Geometric Transformations"):

$$x' = R_{00} * x + R_{01} * y + R_{02} * z + R_{03} \quad (3.19)$$

$$y' = R_{10} * x + R_{11} * y + R_{12} * z + R_{13} \quad (3.20)$$

$$z' = R_{20} * x + R_{21} * y + R_{22} * z + R_{23} \quad (3.21)$$

$$R_{00} = a * \cos r_z * \cos r_y \quad (3.22)$$

$$R_{01} = a * (\cos r_z * \sin r_y * \sin r_x - \sin r_z * \cos r_x) \quad (3.23)$$

$$R_{02} = a * (\cos r_z * \sin r_y * \cos r_x + \sin r_z * \sin r_x) \quad (3.24)$$

$$R_{03} = T_x \quad (3.25)$$

$$R_{10} = a * \sin r_z * \cos r_y \quad (3.26)$$

$$R_{11} = a * (\sin r_z * \sin r_y * \sin r_x + \cos r_z * \cos r_x) \quad (3.27)$$

$$R_{12} = a * (\sin r_z * \sin r_y * \cos r_x - \cos r_z * \cos r_x) \quad (3.28)$$

$$R_{13} = T_y \quad (3.29)$$

$$R_{20} = -a * \sin r_y \quad (3.30)$$

$$R_{21} = a * \cos r_y * \sin r_x \quad (3.31)$$

$$R_{22} = a * \cos r_y * \cos r_x \quad (3.32)$$

$$R_{23} = T_z \quad (3.33)$$

όπου  $x, y, z$  είναι οι αρχικές συντεταγμένες ενός σημείου του άκαμπτου σώματος,  $x', y', z'$  είναι οι συντεταγμένες του ίδιου σημείου μετά την εφαρμογή του μετασχηματισμού,  $a$  είναι ο συντελεστής κλιμάκωσης,  $r_x, r_y, r_z$  είναι οι γωνίες περιστροφής του σώματος περί τους άξονες  $x, y, z$  αντίστοιχα, και  $T_x, T_y, T_z$  είναι οι αποστάσεις σύμφωνα με τις οποίες κάθε σημείο του σώματος θα μετατοπιστεί στην κατεύθυνση  $x, y, z$  αντίστοιχα. Από τις εξισώσεις 3.19-3.33 μπορεί να συναχθεί ότι αν εφαρμοστούν οι τιμές  $a = 1, r_x = r_y = r_z = T_x = T_y = T_z = 0$ , θα είναι  $x' = x, y' = y, z' = z$ .

Με τον όρο σημείο αναφερόμαστε όχι σε ένα σταθερό σύνολο συντεταγμένων, αλλά σε ένα στοιχείο του σώματος με μεταβλητές συντεταγμένες. Στα πλαίσια της αναδιάταξης ενός μορίου, κάθε μόριο θεωρείται ως άκαμπτο σώμα. Κάθε άτομο του μορίου θεωρείται ως ένα μοναδικό σημείο, αφού σε κάθε άτομο αντιστοιχεί ένα σύνολο τρισδιάστατων καρτεσιανών συντεταγμένων, σύμφωνα με το PDB format. Οι καρτεσιανές συντεταγμένες ενός ατόμου αντιστοιχούν στις μεταβλητές  $x, y, z, x', y', z'$  των εξισώσεων 3.19-3.21. Ο συντελεστής κλιμάκωσης  $a$  τίθεται πάντα στην τιμή 1, αφού ο μετασχηματισμός της κλίμακας των μορίων θα οδηγούσε στην παραμόρφωση της τρισδιάστατης δομής τους.

### Περιστροφή και Μετατόπιση

Στη μέθοδό μας, η περιστροφή και η μετατόπιση συμβαίνουν σε δύο περιπτώσεις. Πρώτα, το φάρμακο υιοθετεί το ρόλο του μορίου που θα υποστεί περιστροφή και μετατόπιση (έστω A) και ο συνδέτης υιοθετεί το ρόλο του σταθερού μορίου (έστω B). Αργότερα, το σύμπλεγμα συνδέτη-φαρμάκου υιοθετεί το ρόλο του μορίου A και το αμινοξύ στο οποίο γίνεται η σύζευξη υιοθετεί το ρόλο του σταθερού μορίου B.

Για να πραγματοποιηθεί η περιστροφή ενός μορίου A σε σχέση με ένα σταθερό μόριο B, ο αφινικός μετασχηματισμός εφαρμόζεται στο A χωρίς να χρησιμοποιούνται οι λειτουργίες μετασχηματισμού κλίμακας και μετατόπισης, για τις τιμές παραμέτρων  $a = 1, T_x = T_y = T_z = 0$ . Το πεδίο τιμών των μεταβλητών  $r_x, r_y, r_z$  είναι το διάστημα  $[-180^\circ, +180^\circ]$ . Η περιστροφή του μορίου A εφαρμόζεται για τις τιμές των  $r_x, r_y, r_z$  που ικανοποιούν συγκεκριμένη συνθήκη, η οποία είναι η ελαχιστοποίηση της απόλυτης τιμής της γωνίας μεταξύ των αξόνων των δύο μορίων A, B (Εξισώσεις 3.34, 3.35). Στις εξισώσεις 3.34, 3.35,  $L_d, L_l$  είναι οι άξονες του φαρμάκου και του συνδέτη αντίστοιχα (Εξισώσεις 3, 6), και  $L_{ld}, L_{AA}$  είναι οι άξονες του συμπλέγματος συνδέτη-φαρμάκου και του αμινοξέος, αντίστοιχα (Εξισώσεις 3.9, 3.12).

$$r_x, r_y, r_z: \angle(L_d, L_l) = 0^\circ \quad (3.34)$$

$$r_x', r_y', r_z': \angle(L_{ld}, L_{AA}) = 0^\circ \quad (3.35)$$

Για πραγματοποιηθεί η μετατόπιση ενός μορίου A σε σχέση με ένα σταθερό μόριο B, ο αφινικός μετασχηματισμός εφαρμόζεται στο A χωρίς να χρησιμοποιούνται οι λειτουργίες μετασχηματισμού κλίμακας και περιστροφής, για τις τιμές παραμέτρων  $\alpha = 1, r_x = r_y = r_z = 0$ . Η μετατόπιση εφαρμόζεται για τις τιμές των μεταβλητών  $T_x, T_y, T_z$  που ικανοποιούν μία συγκεκριμένη συνθήκη, η οποία είναι η ελαχιστοποίηση της απόλυτης τιμής της διαφοράς μεταξύ της απόστασης των δύο μορίων και μίας σταθερής τιμής  $l$  (Εξισώσεις 3.36, 3.37). Οι τιμές των μεταβλητών  $T_x, T_y, T_z$  αναζητώνται στο διάστημα  $[-D, +D]$ . Συνήθως, η τιμή  $D = 200$  ήταν επαρκής για την επιτυχή ολοκλήρωση της αναζήτησης, σε διαφορετική περίπτωση μία μεγαλύτερη τιμή  $D$  εφαρμοζόταν. Στις εξισώσεις (3.36, 3.37),  $D_{d-1}$  είναι η απόσταση φαρμάκου και συνδέτη (Εξίσωση 3.15), και  $D_{AA-id}$  είναι η απόσταση μεταξύ φαρμάκου και συνδέτη (Εξίσωση 3.18). Η σταθερή τιμή  $l$  ορίστηκε ως το επίσημο μήκος του διαμοριακού δεσμού ( $l_{Hb} = 1.5 - 2.5 \text{ \AA}$  για το δεσμό υδρογόνου,  $l_{SSb} = 2.05 \text{ \AA}$  για το δισουλφιδικό δεσμό).

$$T_x, T_y, T_z: |D_{d-1} - l_{Hb}| = 0 \quad (3.36)$$

$$T_x', T_y', T_z': \begin{cases} |D_{AA-id} - l_{Hb}| = 0 \text{ για σύζευξη σε Λυσίνη} \\ |D_{AA-id} - l_{SSb}| = 0 \text{ για σύζευξη σε Κυστεΐνη} \end{cases} \quad (3.37)$$

### 3.3. Αποτελέσματα

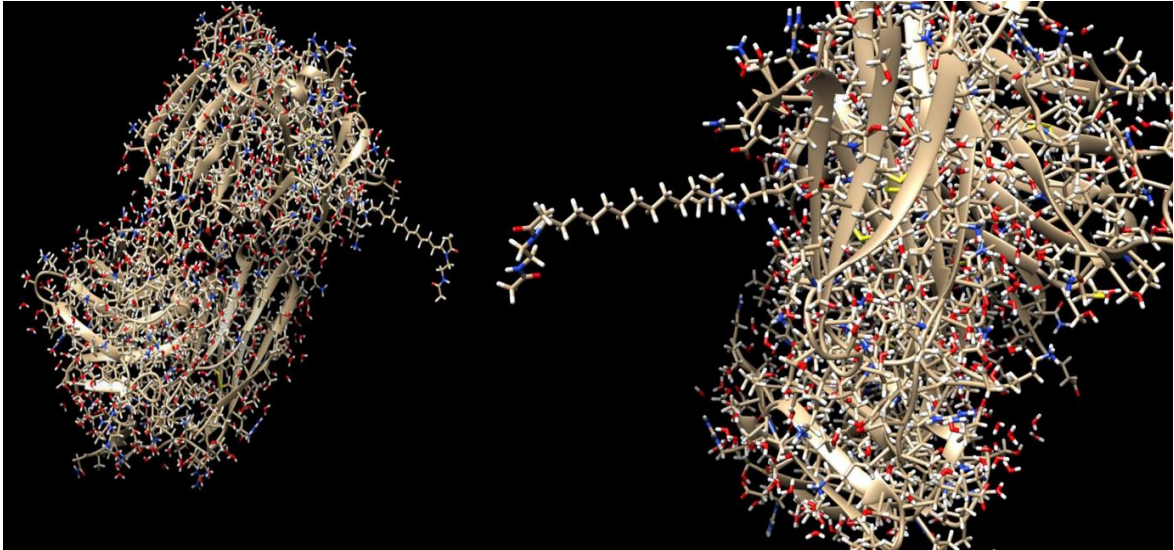
#### 3.3.1. Αποτελέσματα σύζευξης μέσω λυσινών

Η εφαρμογή της διαδικασίας που περιγράφεται στο τμήμα 3.2. είχε ως αποτέλεσμα την παραγωγή ενός αριθμού ADCs ως PDB αρχεία. Αφού το αντίσωμα, το φάρμακο και τα σημεία σύζευξης του αντισώματος και του φαρμάκου ποικίλλουν, ο αριθμός των διαφορετικών συνδυασμών που μπορούν να παράξουν ADCs είναι αρκετά μεγάλος. Ακολουθώς, τέσσερα διαφορετικά ADCs παρουσιάζονται ως αποτελέσματα της διαδικασίας.

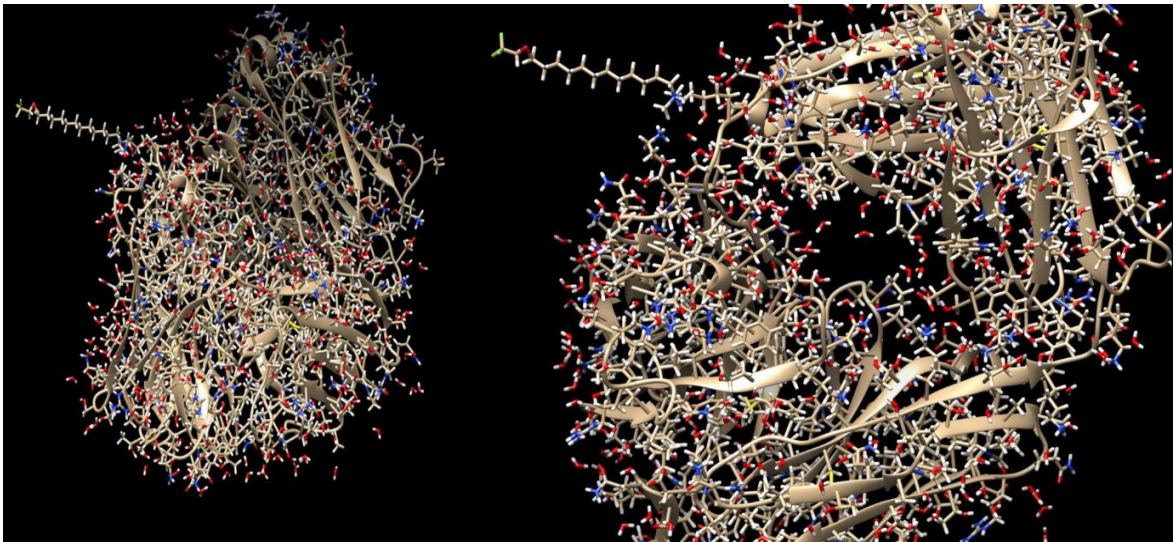
Στην Εικόνα 3.3 παρουσιάζεται το ADC που συντέθηκε από το αντίσωμα με PDB id 4GAG, τον συνδέτη  $C_{15}N$  και το φάρμακο με αριθμό ακολουθίας 600 (nci\_600.pdb). Έχει σχηματιστεί ένας δεσμός υδρογόνου ανάμεσα στο συνδέτη και το αμινοξύ λυσίνη με αριθμό ακολουθίας 147 στην ελαφριά αλυσίδα (L) του αντισώματος, κι ένας δεσμός υδρογόνου ανάμεσα στο συνδέτη και το φάρμακο. Το ίδιο ADC απεικονίζεται από διαφορετική γωνία και μικρότερη απόσταση δεξιά της εικόνας.

Στην Εικόνα 3.4 παρουσιάζεται το ADC που συντέθηκε από το αντίσωμα με pdb id 4GAG, τον συνδέτη  $C_{15}N$  και το φάρμακο με αριθμό ακολουθίας 450 (nci\_450.pdb). Ένας δεσμός υδρογόνου έχει σχηματιστεί ανάμεσα στο συνδέτη και το αμινοξύ λυσίνη με αριθμό ακολουθίας 115 στην βαριά αλυσίδα (H) του αντισώματος. Επίσης, ένας δεσμός υδρογόνου έχει σχηματιστεί ανάμεσα στο συνδέτη και το φάρμακο. Το ίδιο ADC απεικονίζεται από διαφορετική γωνία και μικρότερη απόσταση δεξιά της εικόνας.

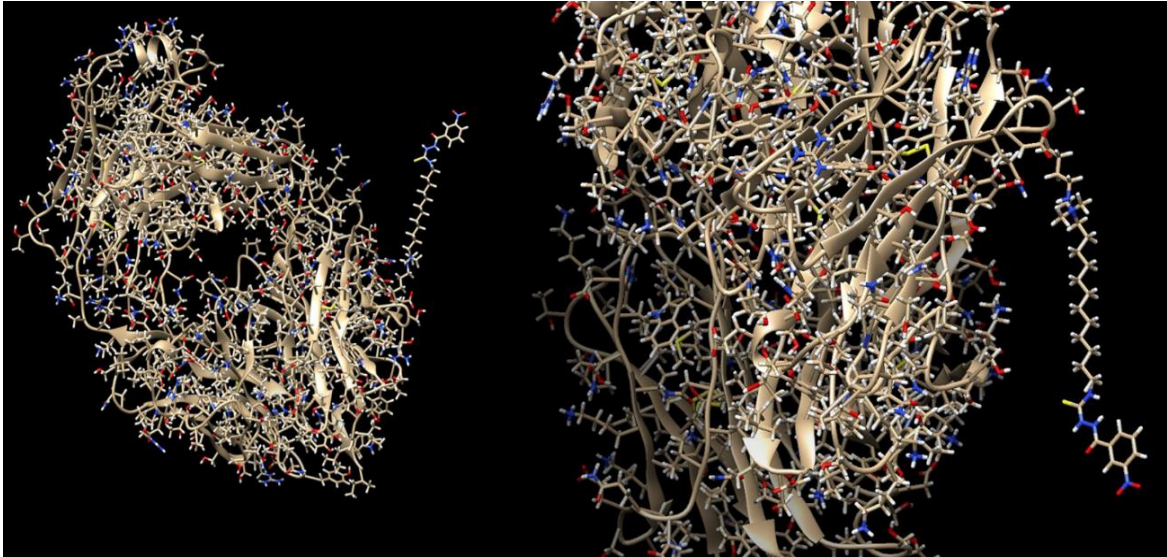




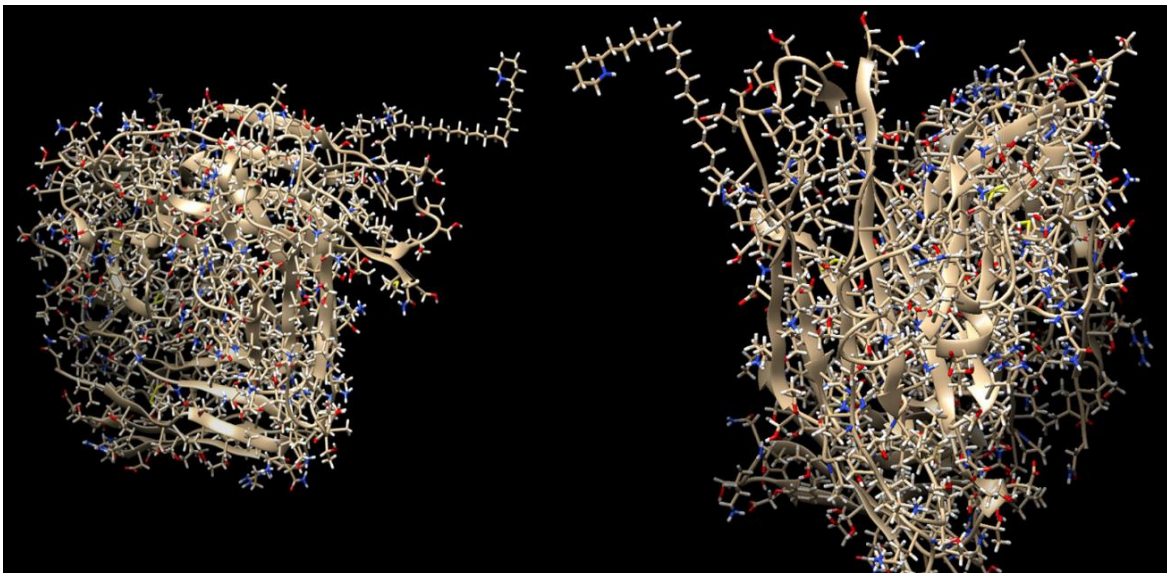
Εικόνα 3.3. Αντίσωμα: 4GAG. Συνδέτης: C<sub>15</sub>N. Φάρμακο: 600. Λυσίνη: 147-LC



Εικόνα 3.4. Αντίσωμα: 4GAG. Συνδέτης: C<sub>15</sub>N. Φάρμακο: 450. Λυσίνη: 115-HC.



Εικόνα 3.5. Αντίσωμα: 4GAJ. Συνδέτης: C<sub>15</sub>N. Φάρμακο: 700. Λυσίνη: 75-HC.



Εικόνα 3.6. Αντίσωμα: 4GAJ. Συνδέτης: C<sub>15</sub>N. Φάρμακο: 14. Λυσίνη: 209-HC.

Στην Εικόνα 3.5 παρουσιάζεται το ADC που συντέθηκε από το αντίσωμα με PDB id 4GAJ, τον συνδέτη C<sub>15</sub>N και το φάρμακο με αριθμό ακολουθίας 700 (nci\_700.pdb). Ένας δεσμός υδρογόνου έχει σχηματιστεί ανάμεσα στο συνδέτη και το αμινοξύ λυσίνη με αριθμό ακολουθίας 75 στην βαριά αλυσίδα (H) του αντισώματος, καθώς και ανάμεσα στο συνδέτη και το φάρμακο. Το ίδιο ADC απεικονίζεται από διαφορετική γωνία και μικρότερη απόσταση στο δεξί μέρος της Εικόνας 3.5.

Στην Εικόνα 3.6 απεικονίζεται το ADC που συντέθηκε από το αντίσωμα 4GAJ, τον ίδιο συνδέτη και το φάρμακο με αριθμό ακολουθίας 14 (nci\_14.pdb). Ένας δεσμός υδρογόνου έχει σχηματιστεί ανάμεσα στο συνδέτη και την λυσίνη με αριθμό ακολουθίας 209 στη βαριά αλυσίδα (H) του αντισώματος, καθώς και ανάμεσα στο συνδέτη και το φάρμακο. Το ίδιο

ADC απεικονίζεται από διαφορετική γωνία και μικρότερη απόσταση στο δεξί μέρος της Εικόνας 3.6.

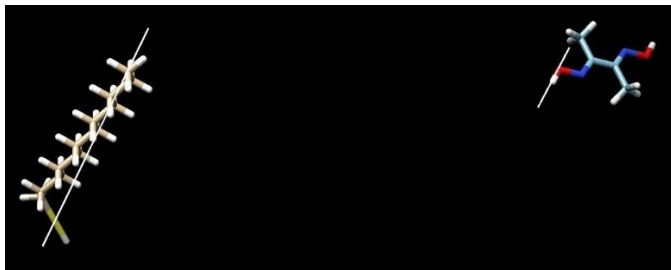
### 3.3.2. Αποτελέσματα σύζευξης μέσω τροποποιημένων κυστεϊνών

Η μέθοδος σύζευξης σε τροποποιημένες κυστεΐνες έχει εφαρμοστεί σε διάφορους συνδυασμούς αντισωμάτων και φαρμάκων. Η δημιουργία ενός TDC (THIOMAB-Drug conjugate) απαιτεί περίπου 10-13 δευτερόλεπτα. Στην εικόνα 3.7 απεικονίζονται τα βασικά στάδια δημιουργίας ενός συμπλέγματος THIOMAB-φαρμάκου. Πρώτα, το φάρμακο και ο συνδέτης υπάρχουν σε δύο διαφορετικά PDB αρχεία. Ο άξονας κάθε μορίου απεικονίζεται ως λευκή ευθεία γραμμή, και η αρχική γωνία μεταξύ τους είναι  $98.2^\circ$  (Εικόνα 3.7(α)). Το φάρμακο περιστρέφεται, και η γωνία μεταξύ των δύο αξόνων γίνεται  $0.08^\circ$  (Εικόνα 3.7(β)). Τα άτομα που θα σχηματίσουν το δεσμό υδρογόνου είναι το άτομο άνθρακα (C) του συνδέτη και το άτομο υδρογόνου (H) του φαρμάκου που συνδέεται ομοιοπολικά με ένα άτομο οξυγόνου (O). Η αρχική απόσταση μεταξύ των δύο αυτών ατόμων είναι  $28.02 \text{ \AA}$ . Το φάρμακο μετατοπίζεται και η απόσταση μεταξύ των δύο ατόμων γίνεται  $1.51 \text{ \AA}$  (Εικόνα 3.7(γ)). Ο συνδέτης έχει παραμείνει σταθερός. Τα δύο μόρια συγχωνεύονται σε ένα σύμπλεγμα συνδέτη-φαρμάκου, του οποίου ο άξονας απεικονίζεται ως λευκή γραμμή (Εικόνα 3.7(δ)). Το αμινοξύ βαλίνη με αριθμό ακολουθίας 206 στην ελαφριά αλυσίδα (LC-VAL206) του αντισώματος με PDB id 4GAG αντικαταστάθηκε από μία κυστεΐνη, της οποίας ο άξονας υπολογίστηκε (Εικόνα 3.7(ε)). Η αρχική γωνία μεταξύ τον άξονα του συμπλέγματος συνδέτη-φαρμάκου και τον άξονα της κυστεΐνης είναι  $105.6^\circ$ , και η απόσταση μεταξύ τα άτομα θείου της κυστεΐνης και του συνδέτη είναι  $54.75 \text{ \AA}$ . Το σύμπλεγμα συνδέτη-φαρμάκου περιστρέφεται και μετατοπίζεται, με αποτέλεσμα η γωνία να γίνει  $0.083^\circ$  και η απόσταση να γίνει  $2.07 \text{ \AA}$ . Τέλος, το σύμπλεγμα συνδέτη-φαρμάκου συνδέεται με την κυστεΐνη (Εικόνα 3.7(ζ)).

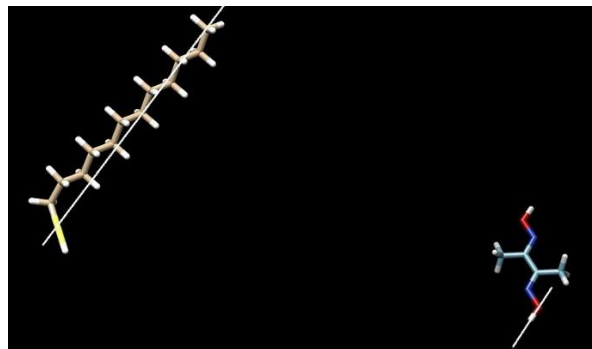
Περισσότερα παραδείγματα από παρηγμένα TDCs απεικονίζονται στην Εικόνα 3.8. Το TDC που αποτελείται από το αντίσωμα με PDB id 4GAJ, το συνδετικό μόριο  $C_{12}S$  και το φάρμακο με αριθμό ακολουθίας 5 φαίνονται στην εικόνα 3.8(α). Ένας δεσμός υδρογόνου συνδέει το φάρμακο με το συνδέτη, και ένας δισουλφιδικός δεσμός συνδέει την τροποποιημένη κυστεΐνη του αντισώματος με το σύμπλεγμα συνδέτη-φαρμάκου. Η κυστεΐνη έχει αντικαταστήσει το προσβάσιμο αμινοξύ αλανίνη με αριθμό ακολουθίας 114 στη βαριά αλυσίδα του αντισώματος (HC-ALA114). Το ίδιο TDC απεικονίζεται από κοντινότερη απόσταση στην δεξιά εικόνα. Στην Εικόνα 3.8(β) απεικονίζεται ένα παρόμοιο TDC που αποτελείται από το αντίσωμα με PDB id 4GAJ, το συνδέτη  $C_{12}S$  και το φάρμακο με αριθμό ακολουθίας 7. Η παραπάνω τεχνική μπορεί να εφαρμοστεί για τη σύζευξη TDCs με πάνω από ένα μόρια φαρμάκου ανά αντίσωμα. Η Εικόνα 3.8(γ) απεικονίζει ένα TDC με DAR 2, που αποτελείται από το αντίσωμα με PDB id 4GAG, στο οποίο δύο μόρια φαρμάκου με αριθμό ακολουθίας 1 έχουν συζευχθεί. Κυστεΐνες έχουν αντικαταστήσει την αλανίνη με αριθμό ακολουθίας 114 στη βαριά αλυσίδα (HC-ALA114) και την βαλίνη με αριθμό ακολουθίας 206 στην ελαφριά αλυσίδα (LC-VAL206) του αντισώματος, για να λειτουργήσουν ως σημεία σύζευξης. Οι περιοχές σύζευξης του ίδιου TDC απεικονίζονται από μικρότερη απόσταση στο δεξί μέρος της εικόνας. Οι συνδέτες και τα φάρμακα που χρησιμοποιήθηκαν στα παραδείγματα των εικόνων 3.3-3.8 απεικονίζονται στην εικόνα 3.9.



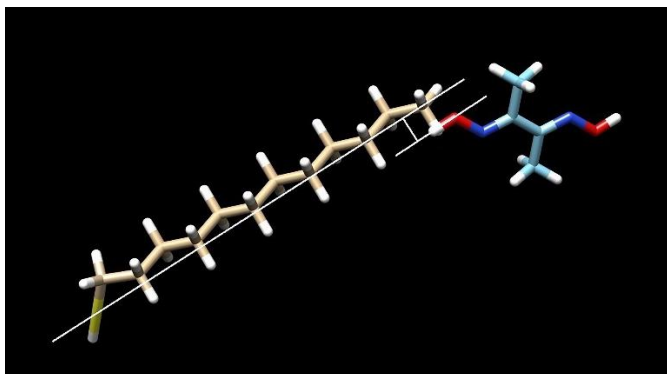
(α)



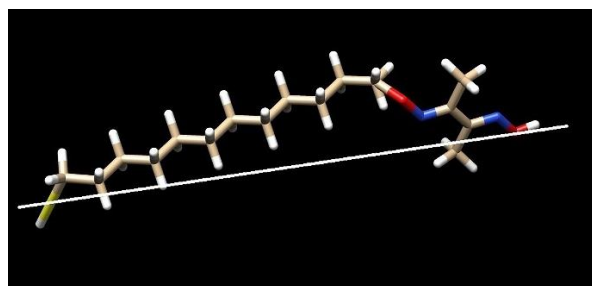
(β)



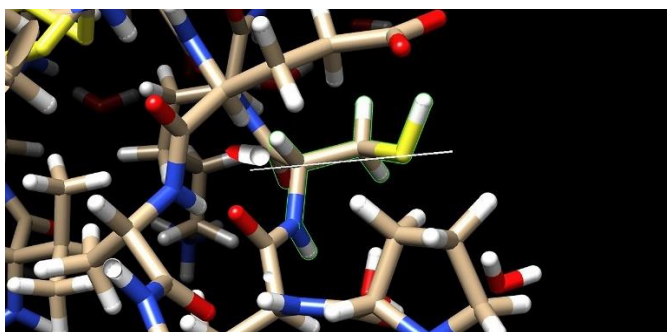
(γ)



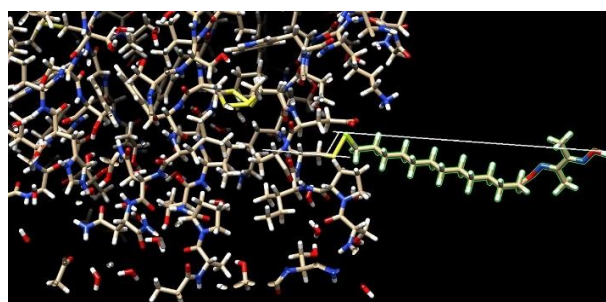
(δ)



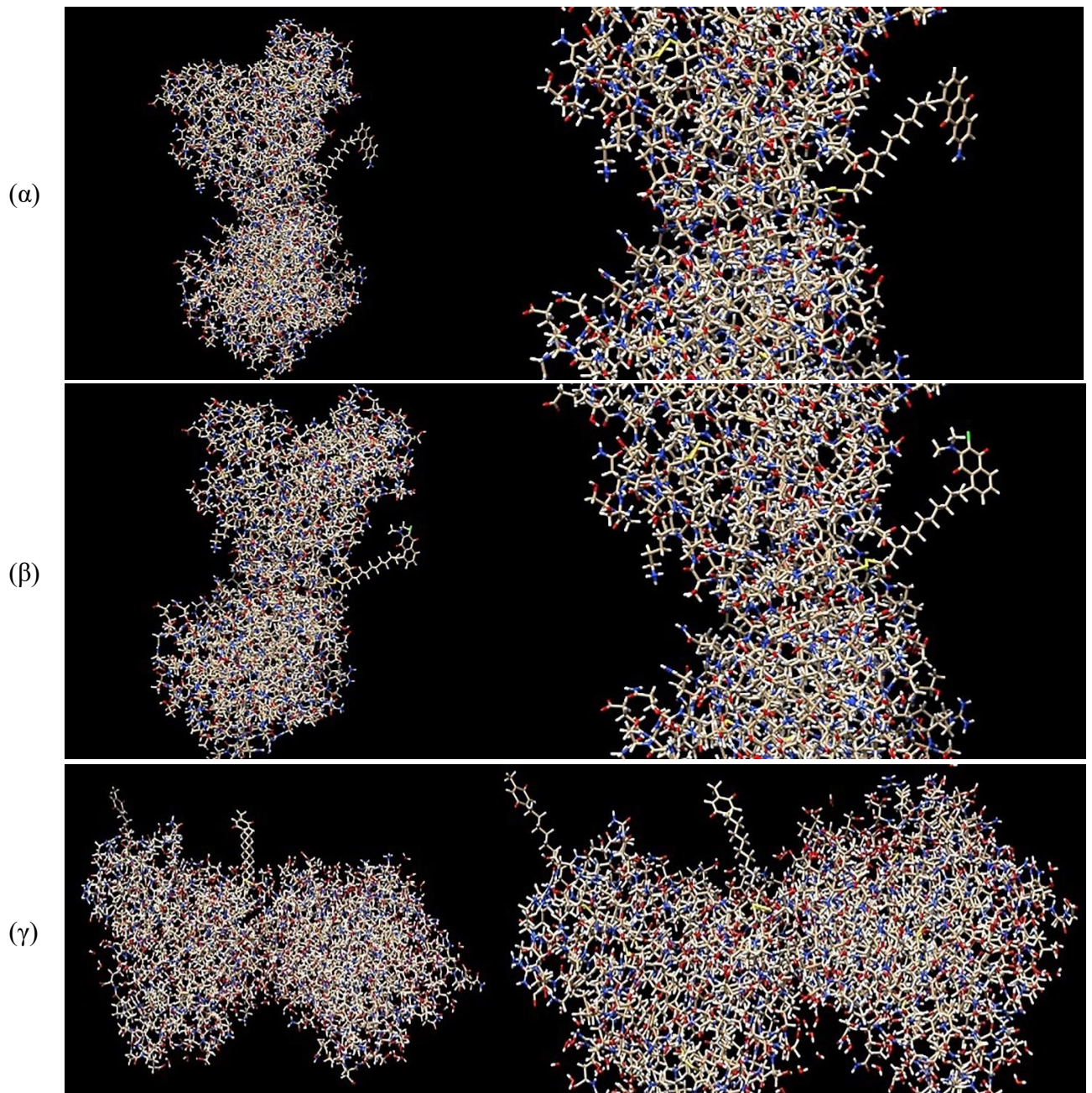
(ε)



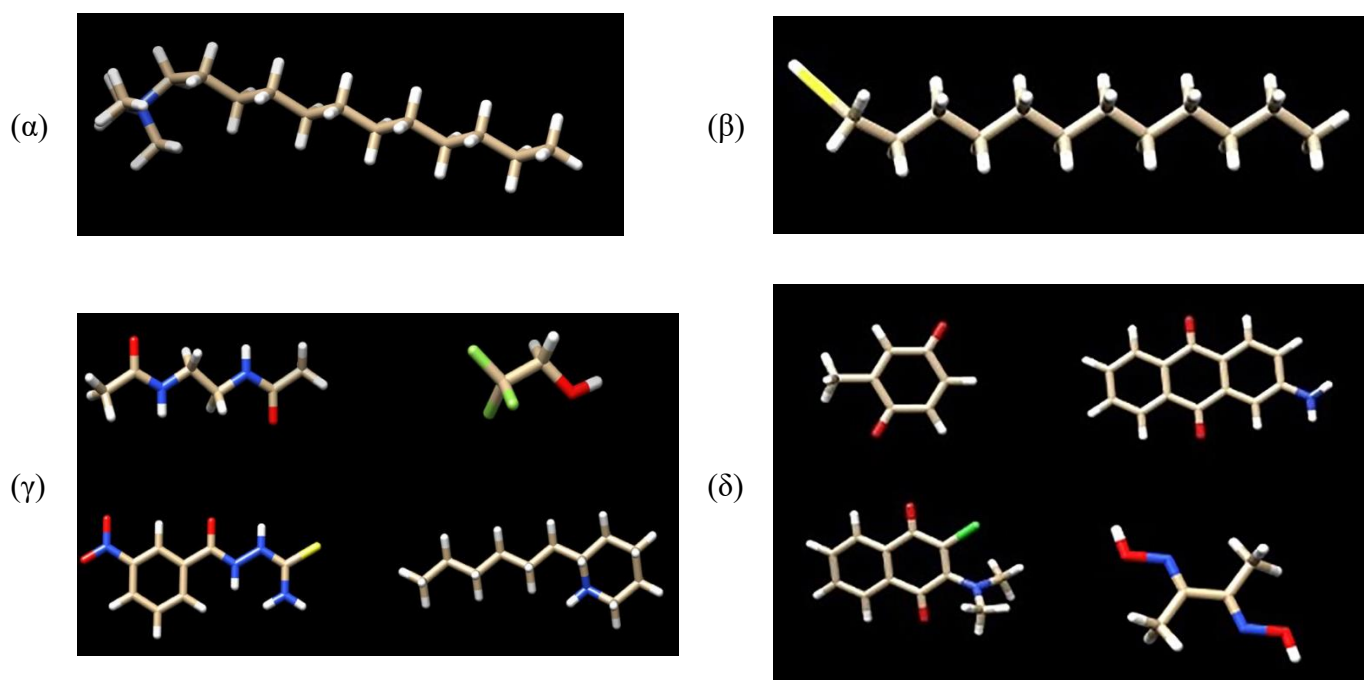
(ζ)



Εικόνα 3.7. Τα βασικά στάδια του υπολογιστικού σχεδιασμού ενός TDC. (α) Αριστερά: Το συνδετικό μόριο  $C_{12}S$ . Δεξιά: Το φάρμακο με NCI αριθμό ακολουθίας 9. Αρχική γωνία μεταξύ των δύο αξόνων:  $98.2^\circ$ . (β) Ο συνδέτης και το περιστραμμένο φάρμακο. Τελική γωνία μεταξύ των δύο αξόνων:  $0.08^\circ$ . Αρχική απόσταση μεταξύ του ατόμου H του φαρμάκου και του ατόμου C του συνδέτη:  $28.02 \text{ \AA}$ . (γ) Ο συνδέτης και το μετατοπισμένο φάρμακο. Τελική απόσταση μεταξύ του ατόμου H του φαρμάκου και του ατόμου C του συνδέτη:  $1.51 \text{ \AA}$ . (δ) Το σύμπλεγμα συνδέτη-φαρμάκου. (ε) Η τροποποιημένη κυστεΐνη που έχει αντικαταστήσει το αμινοξύ βαλίνη στη θέση 206 της ελαφριάς αλυσίδας (LC-VAL206). Αρχική γωνία μεταξύ των αξόνων της κυστεΐνης και του συμπλέγματος συνδέτη-φαρμάκου:  $105.6^\circ$ . Αρχική απόσταση μεταξύ των ατόμων θείου:  $54.75 \text{ \AA}$ . (ζ) Το σύμπλεγμα συνδέτη-φαρμάκου που έχει συζευχθεί στην κυστεΐνη του αντισώματος. Τελική γωνία μεταξύ των αξόνων της κυστεΐνης και του συμπλέγματος συνδέτη-φαρμάκου:  $0.084^\circ$ . Τελική απόσταση μεταξύ των ατόμων θείου:  $2.07 \text{ \AA}$ .



Εικόνα 3.8. (α) Αριστερά: TDC που αποτελείται από το αντίσωμα με PDB id 4GAJ, το συνδετικό μόριο  $C_{12}S$ , και το φάρμακο με αριθμό ακολουθίας 5. Το φάρμακο έχει συζευχθεί μέσω του συνδέτη στην τροποποιημένη κυστεΐνη που έχει αντικαταστήσει την αλανίνη της βαριάς αλυσίδας στη θέση 114 (HC-ALA114). Δεξιά: Το ίδιο TDC από κοντινότερη απόσταση. (β) Αριστερά: TDC που αποτελείται από το αντίσωμα με PDB id 4GAJ, το συνδετικό μόριο  $C_{12}S$ , και το φάρμακο με αριθμό ακολουθίας 7. Το φάρμακο έχει συζευχθεί μέσω του συνδέτη στην τροποποιημένη κυστεΐνη που έχει αντικαταστήσει την αλανίνη της βαριάς αλυσίδας στη θέση 114 (HC-ALA114). Δεξιά: Το ίδιο TDC από κοντινότερη απόσταση. (γ) Αριστερά: TDC που αποτελείται από το αντίσωμα με PDB id 4GAJ, δύο συνδετικά μόρια  $C_{12}S$ , και δύο φάρμακα με αριθμό ακολουθίας 1. Τα σημεία σύζευξης στο αντίσωμα είναι τροποποιημένες κυστεΐνες που έχουν αντικαταστήσει τα αμινοξέα HC-ALA114 και LC-VAL206. Δεξιά: Το ίδιο TDC από κοντινότερη απόσταση.



Εικόνα 3.9. (α) Το συνδετικό μόριο  $C_{15}N$ . (β) Το συνδετικό μόριο  $C_{12}S$  (γ). Οι κυτταροτοξικές ουσίες που χρησιμοποιήθηκαν στην παράγραφο 3.3.1. Πάνω αριστερά: Η κυτταροτοξική ουσία με αριθμό ακολουθίας 600. Πάνω δεξιά: Η κυτταροτοξική ουσία με αριθμό ακολουθίας 450. Κάτω αριστερά: Η κυτταροτοξική ουσία με αριθμό ακολουθίας 700. Κάτω δεξιά: Η κυτταροτοξική ουσία με αριθμό ακολουθίας 14. (δ) Οι κυτταροτοξικές ουσίες που χρησιμοποιήθηκαν στην παράγραφο 3.3.2. Πάνω αριστερά: Το φάρμακο με αριθμό ακολουθίας 1. Πάνω δεξιά: Το φάρμακο με αριθμό ακολουθίας 5. Κάτω αριστερά: Το φάρμακο με αριθμό ακολουθίας 7. Κάτω δεξιά: Το φάρμακο με αριθμό ακολουθίας 9.

### 3.4. Αναφορές

- Berman, H. M., Kleywegt, G. J., Nakamura, H., & Markley, J. L. (2012). The Protein Data Bank at 40: reflecting on the past to prepare for the future. *Structure (London, England : 1993)*, *20*(3), 391–6. <http://doi.org/10.1016/j.str.2012.01.010>
- Berman, H., Henrick, K., Nakamura, H., & Markley, J. L. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, *35*(Database issue), D301-3. <http://doi.org/10.1093/nar/gkl971>
- Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. <http://doi.org/10.1093/nar/28.1.235>
- Geometric Transformations. (n.d.). Retrieved February 21, 2017, from <https://www.cs.mtu.edu/~shene/COURSES/cs3621/NOTES/geometry/geo-tran.html>
- Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–49. <http://doi.org/10.1111/j.1476-538>
- Katsila, T., Spyroulias, G. A., Patrinos, G. P., & Matsoukas, M.-T. (2016). Computational approaches in target identification and drug discovery. *Computational and Structural Biotechnology Journal*, *14*, 177–184. <http://doi.org/10.1016/j.csbj.2016.04.004>
- Kessel, M. (2011). The problems with today's pharmaceutical business—an outsider's view. *Nature Biotechnology*, *29*(1), 27–33. <http://doi.org/10.1038/nbt.1748>

- Peters, C., & Brown, S. (2015). Antibody-drug conjugates as novel anti-cancer chemotherapeutics. *Bioscience Reports*, 35(4). <http://doi.org/10.1042/BSR20150089>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–12. <http://doi.org/10.1002/jcc.20084>
- Taft, C. A., da Silva, V. B., & da Silva, C. H. T. de P. (2008). Current topics in computer-aided drug design. *Journal of Pharmaceutical Sciences*, 97(3), 1089–1098. <http://doi.org/10.1002/jps.21293>





## Κεφάλαιο 4. Εισαγωγή στη Μεταβολομική και τη Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού

---

### 4.1. Μεταβολομική

#### 4.1.1. Έννοιες μεταβολομικής

Οι μεταβολίτες είναι χημικές ενώσεις χαμηλού μοριακού βάρους (<1 kDa) που αποτελούν παράγωγα χημικών αντιδράσεων, εντοπίζονται στο εσωτερικό των κυττάρων και είναι απαραίτητες για τη συντήρηση, ανάπτυξη και φυσιολογική λειτουργία των κυττάρων. Οι μεταβολίτες συμπεριλαμβάνουν τόσο οργανικές ουσίες (αμινοξέα, λιπαρά οξέα, υδατάνθρακες, βιταμίνες, λιπίδια), όσο και ανόργανες ουσίες και χημικά στοιχεία (Dunn et al., 2005; Nicholson, 2006).

Το μεταβόλωμα (metabolome) έχει οριστεί ως το σύνολο των μεταβολιτών που βρίσκονται σε ένα δεδομένο τύπο κυττάρου ή ιστού ή εκκρίνονται από αυτόν (Nicholson et al., 2003). Το μεταβόλωμα έχει αποδοθεί επίσης ως το σύνολο μεταβολιτών που είναι παρόντα σε ένα βιολογικό σύστημα σε μία δεδομένη φυσιολογική κατάσταση με ένα δεδομένο σύνολο περιβαλλοντολογικών συνθηκών (Oliver et al., 1998; Naz et al., 2014), καθώς και ως το ολικό μεταβολικό περιεχόμενο ενός βιολογικού δείγματος (Agin et al., 2016). Το μεταβόνωμα (metabonome) έχει οριστεί ως το σύνολο των κυτταρικών μεταβολωμάτων σε έναν πολυκύτταρο οργανισμό, καθώς και τα προϊόντα χημικών αντιδράσεων (Nicholson et al., 2003). Όπως υποδηλώνουν οι παραπάνω ορισμοί, οι έννοιες του μεταβόλωματος και του μεταβονώματος αρχικά ήταν διαχωρισμένες, πλέον όμως είναι συνώνυμες και αναφέρονται στο σύνολο των μεταβολιτών ενός βιολογικού συστήματος (Kosmidis et al., 2013). Το μέγεθος του μεταβόλωματος ποικίλλει ανάλογα με τον οργανισμό, με την πιο ολοκληρωμένη βάση δεδομένων για το ανθρώπινο μεταβόλωμα (Human Metabolome Database, HMDB) να περιλαμβάνει μέχρι στιγμής (Ιανουάριος 2018) περισσότερες από 114.000 καταχωρήσεις μεταβολιτών (Agin et al., 2016). Επομένως, το μεταβόλωμα αποτελεί ένα σύνολο δεκάδων χιλιάδων ουσιών με μεγάλες διακυμάνσεις ως προς τις φυσικοχημικές ιδιότητές τους (Dunn & Ellis, 2005).

Το επιστημονικό πεδίο της μεταβολομικής / μεταβονομικής είναι σχετικά πρόσφατο, με δεδομένο ότι οι πρώτες έρευνες που επικεντρώθηκαν στη μεταβολική σύνθεση κυττάρων, ιστών και βιολογικών υγρών δημοσιεύθηκαν τη δεκαετία του '80. Η μεταβολομική και η μεταβονομική είναι δύο έννοιες για τις οποίες έχουν δοθεί αρκετοί ορισμοί στη διάρκεια των χρόνων. Συγκεκριμένα, η μεταβολομική (metabolomics) έχει οριστεί ως η ολική, αμερόληπτη, ποιοτική και ποσοτική ανάλυση των μεταβολιτών σε ένα βιολογικό σύστημα (Fiehn et al., 2002; Agin et al., 2016). Ακόμη, η μεταβολομική έχει οριστεί ως η μέτρηση των συγκεντρώσεων των μεταβολιτών σε κύτταρα και ιστούς στα οποία υπάρχει μία άμεση σύνδεση μεταξύ της γενετικής, της πρωτεϊνικής και της μεταβολικής δραστηριότητας (Nicholson et al., 2003). Η μεταβολομική έχει αποδοθεί επίσης ως η αμερόληπτη αναγνώριση και ποσοτικοποίηση όλων των μεταβολιτών σε ένα βιολογικό σύστημα (Dunn & Ellis, 2005). Η μεταβονομική (metabonomics) έχει οριστεί ως η ποσοτική μέτρηση της δυναμικής πολυπαραμετρικής μεταβολικής απόκρισης των έμβιων συστημάτων σε παθοφυσιολογικά ερεθίσματα ή σε γενετικές μεταβολές (Nicholson et al., 1999). Το αντικείμενο της μεταβονομικής έχει αποδοθεί ως η ανίχνευση, η ταυτοποίηση, η ποσοτικοποίηση και η

καταγραφή των μεταβολικών αλλαγών στο χρόνο ενός βιολογικού συστήματος, παρά ενός μεμονωμένου κυττάρου, ως απόκριση σε διάφορους ενδογενείς ή εξωγενείς παράγοντες, όπως ο τρόπος ζωής και οι γενετικοί ή οι περιβαλλοντολογικοί παράγοντες (Nicholson et al., 2003). Επίσης, η μεταβολομική έχει αποδοθεί ως η αξιολόγηση των ιστών και των βιολογικών υγρών όσον αφορά τις μεταβολές στα επίπεδα των ενδογενών μεταβολιτών που απορρέουν από μία ασθένεια ή μία θεραπευτική αγωγή (Dunn & Ellis, 2005). Αν και οι έννοιες της μεταβολομικής και μεταβολομικής αρχικά ήταν διαχωρισμένες, πλέον είναι συνώνυμες και συνοπτικά αναφέρονται στην ανάλυση των μεταβολιτών ενός βιολογικού συστήματος ή δείγματος (Naz et al., 2014; Kosmidis et al., 2013).

Στο πεδίο της μεταβολομικής υπάρχουν διάφορες προσεγγίσεις όσον αφορά τη μελέτη ενός βιολογικού συστήματος. Το μεταβολικό προφίλ “metabolite ή metabolic profiling” είναι μία αναλυτική διαδικασία που επικεντρώνεται στα αποτελέσματα μίας διαταραχής σε συγκεκριμένους προκαθορισμένους μεταβολίτες ενός δείγματος, και όχι σε όλο το μεταβολισμό. Στόχος αυτής της διαδικασίας είναι η αναγνώριση και ποσοτικοποίηση των μεταβολιτών ενδιαφέροντος, οι οποίοι μπορεί να είναι μία κατηγορία μεταβολιτών (π.χ. υδατάνθρακες, λιπίδια) ή μέλη ενός βιοχημικού μονοπατιού. Με δεδομένο ότι οι μεταβολίτες υπό μελέτη είναι προκαθορισμένοι, η προετοιμασία των δειγμάτων και η λήψη δεδομένων μπορεί να διαμορφωθεί με βάση τις χημικές ιδιότητες αυτών των μεταβολιτών, μειώνοντας την επίδραση του υποστρώματος του δείγματος. Η διαδικασία του μεταβολικού προφίλ είναι χρήσιμη στην περίπτωση που είναι επιθυμητή η διερεύνηση επιλεγμένων βιοχημικών μονοπατιών, και αξιοποιείται συχνά στα πλαίσια της φαρμακευτικής έρευνας. Το μεταβολικό αποτύπωμα “metabolite ή metabolic fingerprinting” είναι μία γρήγορη διαδικασία σάρωσης υψηλής απόδοσης με στόχο την κατηγοριοποίηση των δειγμάτων προς ανάλυση, δηλαδή το διαχωρισμό τους ως προς την προέλευσή τους ή τη βιολογική κατάσταση στην οποία αντιστοιχούν. Η διαδικασία του μεταβολικού αποτύπωματος συνήθως δεν περιλαμβάνει την *a priori* αναγνώριση και ποσοτικοποίηση μεταβολιτών (παρόλο που η μέθοδος είναι ποσοτική), η προετοιμασία των δειγμάτων είναι απλή. Η στοχευμένη ανάλυση μεταβολιτών “metabolite target analysis” έχει ως αντικείμενο την ποιοτική και ποσοτική ανάλυση ενός ή περισσότερων μεταβολιτών που συνδέονται με μία συγκεκριμένη μεταβολή (π.χ. γενετική μετάλλαξη). Συμπεριλαμβάνει εκτεταμένη προετοιμασία του δείγματος και διαχωρισμό του ζητούμενου μεταβολίτη από τους υπόλοιπους. (Dunn & Ellis, 2005; Fiehn et al., 2001; Fiehn et al., 2002)

Όπως προκύπτει και από τα παραπάνω, οι μελέτες στο χώρο της μεταβολομικής μπορούν να διαχωριστούν σε δύο διαφορετικές προσεγγίσεις, τη μη στοχευμένη (untargeted) και τη στοχευμένη (targeted) μεταβολομική ανάλυση. Η μη στοχευμένη προσέγγιση αποσκοπεί στην ταυτόχρονη αποτίμηση όσο το δυνατόν περισσότερων μεταβολιτών, χωρίς εκ των προτέρων γνώση της φύσης και της ταυτότητάς τους, και δίνει τη δυνατότητα εξερεύνησης νέων περιοχών του μεταβολισμού. Αντίθετα, η στοχευμένη προσέγγιση αποσκοπεί στην ακριβή ποσοτικοποίηση γνωστών, αναμενόμενων μεταβολιτών. Σύμφωνα με αυτή τη στρατηγική, προκαθορισμένα σήματα χρησιμοποιούνται για την ακριβή αποτίμηση των συγκεντρώσεων ενός περιορισμένου αριθμού μεταβολιτών. Στη στοχευμένη μεταβολομική ανάλυση, η προετοιμασία του δείγματος μπορεί να βελτιστοποιηθεί μειώνοντας την κυριαρχία των μορίων που βρίσκονται σε αφθονία. Σε αντίθεση με τη μη στοχευμένη προσέγγιση, η στοχευμένη προσέγγιση οδηγείται από μία συγκεκριμένη υπόθεση

και αποσκοπεί στη διερεύνηση συγκεκριμένων βιοχημικών μονοπατιών (Agin et al., 2016; Roberts et al., 2012).

#### **4.1.2. Πλεονεκτήματα μεταβολομικής**

Η μεταβολομική παρουσιάζει το πλεονέκτημα ότι παρέχει ένα επίπεδο πληροφορίας για τις βιολογικές διαδικασίες που δεν μπορεί να αποκτηθεί με τις προσεγγίσεις συστημικής βιολογίας χαμηλότερου επιπέδου, δηλαδή την πρωτεομική και τη γενομική ανάλυση. Η πρωτεομική και η γενομική επικεντρώνονται στην κατανόηση του τι μπορεί να συμβεί σε έναν βιολογικό οργανισμό, δηλαδή αποκαλύπτουν το δυναμικό για ένα φαινοτυπικό αποτέλεσμα, αλλά δεν μπορούν να εκτιμήσουν το πλήρες βιολογικό αντίκτυπο της γενετικής έκφρασης και των μεταβολών στα επίπεδα πρωτεϊνών ενός οργανισμού. Αντίθετα, η μεταβολομική δίνει τη δυνατότητα για παρακολούθηση των μεταβολιτών ενός βιολογικού συστήματος. Επομένως, η μεταβολομική παρέχει πληροφορία για τη φαινοτυπική απόκριση ενός οργανισμού και μπορεί να συμβάλει στην κατανόηση των συνεπειών των γενετικών και πρωτεομικών μεταβολών. Επιπλέον, αν και η μεταβολομική ανάλυση απαλλάσσει από την ανάγκη αποσαφήνισης της προέλευσης του φαινοτύπου ενός οργανισμού, δηλαδή το σύνολο των παρατηρήσιμων φυσικών χαρακτηριστικών του, είναι δυνατή η εξαγωγή συμπερασμάτων ως προς τις αιτίες ενός φαινοτύπου (Thomas et al., 2006; Kosmidis et al., 2013).

Ένα ακόμη πλεονέκτημα της μεταβολομικής αφορά το χρόνο μέτρησης μεταβολιτών. Συγκεκριμένα, οι μεταβολές στα επίπεδα μεταβολιτών πραγματοποιούνται σε δευτερόλεπτα, σε αντίθεση με τις μεταβολές των προϊόντων του πρωτεώματος, που πραγματοποιούνται σε χρόνο που κυμαίνεται από λεπτά μέχρι ώρες. Αυτός ο μικρότερος χρόνος απόκρισης επιτρέπει τη χρήση της μεταβολομικής ως δείκτη περιβαλλοντικών διαταραχών και την καθιστά ένα υποσχόμενο εργαλείο στην κλινική συστημική βιολογία για την ανίχνευση πρώιμων μεταβολικών διαταραχών, ακόμα και πριν την εμφάνιση των συμπτωμάτων μίας ασθένειας (Smolinska et al., 2012).

Βεβαίως, η συλλογή και ανάλυση δεδομένων από όλα τα επίπεδα της βιολογίας συστημάτων (μεταβολομική, πρωτεομική, γενομική) μπορεί να αποκαλύψει τις αλληλεπιδράσεις μεταξύ του μεταβολώματος, του πρωτεώματος, του γονιδιώματος και του μεταγραφώματος, καθώς και να αποκαλύψει τις συσχετίσεις που προκύπτουν μεταξύ των διαφορετικών επιπέδων πληροφορίας, οδηγώντας σε πληρέστερη κατανόηση της λειτουργίας ενός βιολογικού οργανισμού. Οι μεταβολές στο επίπεδο της γενετικής έκφρασης και τις συγκεντρώσεις πρωτεϊνών μπορούν να συνδεθούν με τις μεταβολικές αλλαγές και να ερμηνευθούν (Kosmidis et al, 2013).

#### **4.1.3. Αναλυτικές τεχνικές**

Η μεταβολομική ανάλυση ενός δείγματος είναι εφικτή με τη χρήση ενός συνόλου διαθέσιμων αναλυτικών τεχνικών, οι πιο ευρέως χρησιμοποιούμενες από τις οποίες είναι η Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance, NMR) και η Φασματομετρία Μάζας (Mass Spectrometry, MS).

Η Φασματομετρία Μάζας (Mass Spectrometry, MS) είναι μία αναλυτική τεχνική που στηρίζεται στην αρχή ότι τα κινούμενα ιόντα εκτρέπονται από ηλεκτρικά και μαγνητικά πεδία. Βασικά στάδια στη φασματομετρία μάζας είναι (α) η δημιουργία φορτισμένων

σωματιδίων, δηλαδή ιόντων, ( $\beta$ ) ο διαχωρισμός των ιόντων σύμφωνα με το λόγο μάζας προς φορτίο ( $m/z$ ) και ( $\gamma$ ) η ανίχνευση των διαχωρισμένων ιόντων. Αν και η φασματομετρία μάζας είναι μία τεχνική με αρκετά υψηλή ταχύτητα και ευαισθησία, παρουσιάζει κάποια μειονεκτήματα, όπως το ότι είναι επιρρεπής στο φαινόμενο καταστολής ιόντων. Για το λόγο αυτό, συνήθως δε χρησιμοποιείται μόνη της αλλά σε συνδυασμό με τεχνικές χρωματογραφίας, ειδικά στην περίπτωση μελέτης πολύπλοκων βιολογικών δειγμάτων. Συνήθεις χρωματογραφικές τεχνικές που χρησιμοποιούνται στη μεταβολομική σε σύζευξη με τη φασματομετρία μάζας είναι η αέρια χρωματογραφία (Gas Chromatography, GC), η υγρή χρωματογραφία (Liquid Chromatography, LC) ή υγρή χρωματογραφία υψηλής απόδοσης (High-performance Liquid Chromatography, HPLC). Ο χρωματογραφικός διαχωρισμός των μεταβολιτών μειώνει την επίδραση του υποστρώματος, διαχωρίζει τα ισομερή και παρέχει πρόσθετα δεδομένα για το χαρακτηρισμό των μεταβολιτών. (Agin et al., 2016; Dunn & Ellis, 2005)

Η Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance, NMR) είναι μία αναλυτική τεχνική που βασίζεται στην απορρόφηση και την επανεκπομπή ηλεκτρομαγνητικής ακτινοβολίας από τους πυρήνες των ατόμων του δείγματος. Στη μεταβολομική χρησιμοποιείται κύρια αλλά όχι αποκλειστικά η φασματοσκοπία NMR πρωτονίου ( $^1\text{H}$  NMR) λόγω της αφθονίας των πυρήνων  $^1\text{H}$  στα βιολογικά δείγματα. Ως τεχνική, χαρακτηρίζεται από υψηλή ταχύτητα και αναπαραγωγιμότητα, ενώ δίνει τη δυνατότητα για ταυτόχρονη ποιοτική και ποσοτική ανάλυση. Αν και έχει χαμηλότερη ευαισθησία από τη φασματομετρία μάζας, δεν απαιτείται η συνδυαστική χρήση της με κάποια άλλη τεχνική (Agin et al., 2016; Lindon & Nicholson, 2008). Στην παράγραφο 4.2 γίνεται εκτενής περιγραφή της φασματοσκοπίας NMR.

#### **4.1.4. Εφαρμογές μεταβολομικής**

##### **4.1.4.1. Διάγνωση**

Κεντρικό ρόλο στις εφαρμογές της μεταβολομικής έχουν οι βιοδείκτες. Ως βιοδείκτης μπορεί να οριστεί ένα μετρήσιμο χαρακτηριστικό που αντικατοπτρίζει την κατάσταση μίας φυσιολογικής ή παθολογικής διαδικασίας. Πιο συγκεκριμένα, βιοδείκτης μίας ασθένειας είναι μία ουσία της οποίας η συγκέντρωση παρουσιάζει υψηλή απόκλιση μεταξύ των διαφορετικών ομάδων δειγμάτων. Λόγω του πολυπαραγοντικού υπόβαθρου, η ταυτοποίηση του βιοδείκτη μίας ασθένειας απαιτεί μια πολύ προσεκτική επιλογή του δειγματοχώρου που περιλαμβάνει ασθενείς και υγιείς με σωστά καθορισμένα κριτήρια αποκλεισμού. Οι βιοδείκτες παρέχουν πληροφορία για την έκταση και το μηχανισμό της ασθένειας, καθώς και την επίδραση ενός φαρμάκου στην πορεία της ασθένειας. Έχουν ιδιαίτερα σημαντικό ρόλο στην κλινική πρόγνωση και διάγνωση, με το μεγαλύτερο ποσοστό (>80%) των διαγνωστικών ελέγχων να βασίζονται στην ανίχνευση και μέτρησή τους (Kumar et al., 2014; Koen et al., 2015).

Με δεδομένο ότι οι μεταβολίτες μπορούν να λειτουργήσουν ως βιοδείκτες ασθενειών, η ταυτοποίηση και ανάλυση μεταβολιτών είναι ένα αξιόλογο εργαλείο για την διαγνωστική ιατρική. Πράγματι, η μεταβολομική ανάλυση έχει επιστρατευθεί για την ανίχνευση βιοδεικτών σε διάφορες κατηγορίες βιολογικών δειγμάτων, όπως ο ορός, το πλάσμα, τα ούρα και το εγκεφαλονωτιαίο υγρό, για ένα μεγάλο εύρος ασθενειών, όπως ο καρκίνος (Armitage et al., 2014; Spratlin et al., 2009), οι καρδιαγγειακές (Kordalewska & Markuszewski, 2012;

Shah et al., 2012), νευρολογικές (Hassan-Smith et al., 2012) και ψυχιατρικές ασθένειες (Kumar et al., 2014). Συνήθως, οι μεταβολομικοί βιοδείκτες εξερευνούνται προκλινικά σε κυτταρικές καλλιέργειες κι επικυρώνονται σε βιολογικά δείγματα (Spratlin et al., 2009).

### **Η Μεταβολομική στην διάγνωση του καρκίνου**

Η μεταβολομική θεωρείται ένα από τα πιο γρήγορα εξελισσόμενα πεδία στο χώρο της καρκινικής έρευνας. Η ανακάλυψη συγκεκριμένων βιοδεικτών για τα διαφορετικά είδη καρκίνου και η χρήση τους για διαγνωστικούς και προγνωστικούς σκοπούς αποτελεί βασικό στόχο που θα συμβάλλει στην εξέλιξη αυτή (Armitage et al., 2014). Από τις αρκετές ερευνητικές προσπάθειες που έχουν επικεντρωθεί στην εύρεση καρκινικών βιοδεικτών έχει προκύψει η συσχέτιση αυτής της κατηγορίας ασθενειών με ανυψωμένα επίπεδα φωσφολιπιδίων, αυξημένη γλυκολυτική ικανότητα κι αυξημένη κατανάλωση γλυκόζης (Spratlin et al., 2009). Στη συνέχεια παρουσιάζονται ενδεικτικά παραδείγματα μεταβολομικών ερευνών που έχουν διεκπεραιωθεί για συγκεκριμένα είδη καρκίνου.

Μία από τις καλύτερες εφαρμογές της μεταβολομικής έχει αποδειχθεί ότι είναι ο καρκίνος μαστού, ο οποίος έχει συνδεθεί με αυξημένα επίπεδα χολίνης, καθώς και χαμηλά επίπεδα γλυκεροφωσφοχολίνης και γλυκόζης σε σύγκριση με τους καλοήθεις όγκους και τους υγιείς ιστούς (Spratlin et al., 2009; Armitage et al., 2014). Άλλα παραδείγματα ουσιών που έχουν χαρακτηριστεί ως βιοδείκτες του καρκίνου μαστού είναι το σύνολο των λιπιδίων (Bougnoux et al., 2008), το ομοβανιλικό οξύ και η ουρία (Nam et al., 2009). Σύμφωνα με τους Lv & Yang (2012), η ανάλυση με την τεχνική GC/MS των δειγμάτων ορού από ασθενείς με καλοήθεις ή κακοήθεις όγκους στήθους και από δείγματα ελέγχου είχε ως αποτέλεσμα τη διαφοροποίηση της κακοήθειας από τις άλλες δύο κατηγορίες με βάση το παλμιτικό, το στεαρικό και το λινολεϊκό οξύ, καθώς και το συνολικό προφίλ ελεύθερων λιπαρών οξέων (Lv & Yang 2012). Επιπλέον, έχει υποστηριχθεί ότι τα αυξημένα επίπεδα μονοακόρεστων λιπιδίων, καθώς και η χαμηλή αναλογία μεταξύ των ω-6 και ω-3 λιπαρών οξέων συνδέονται με μείωση του κινδύνου εκδήλωσης καρκίνου μαστού (Bougnoux et al., 2008).

Σημαντικά ευρήματα έχουν προκύψει και από τις μεταβολομικές έρευνες με επίκεντρο τον επιθηλιακό καρκίνο ωοθηκών (Armitage et al., 2014). Η ανάλυση δειγμάτων ορού από ασθενείς και δείγματα ελέγχου με φασματοσκοπία <sup>1</sup>H-NMR είχε ως αποτέλεσμα το διαχωρισμό των δύο κατηγοριών δειγμάτων (Odunsi et al., 2005). Μεταγενέστερη έρευνα με αντικείμενο τη μεταβολομική ανάλυση δειγμάτων ούρων γυναικών με MS απέδειξε ότι ορισμένοι βιοδείκτες του καρκίνου μαστού παρουσιάζουν αυξημένη συγκέντρωση και στον καρκίνο ωοθηκών σε σχέση με τα δείγματα ελέγχου (Woo et al., 2009). Επίσης, η μεταβολομική ανάλυση με την τεχνική UPLC/MS των δειγμάτων πλάσματος από ασθενείς με είτε κακοήθη είτε καλοήθη επιθηλιακό καρκίνο ωοθηκών είχε ως αποτέλεσμα το διαχωρισμό των δύο κατηγοριών δειγμάτων και την ανάδειξη τεσσάρων βιοδεικτών (Zhang et al., 2012).

Όσον αφορά τον καρκίνο πνεύμονα, η μεταβολομική ανάλυση δειγμάτων ορού και πνευμονικού ιστού με την τεχνική GC/MS αποκάλυψε ότι τα επίπεδα διαφόρων μεταβολιτών επηρεάστηκαν από την εξέλιξη του καρκίνου πνεύμονα (Hori et al., 2011). Επίσης, η μεταβολομική ανάλυση δειγμάτων ορού από ασθενείς με καρκίνο πνεύμονα και υγιή άτομα με την τεχνική MS αποκάλυψε σημαντική διαφοροποίηση σε 141 μεταβολίτες ανάμεσα στις δύο κατηγορίες ατόμων (Guo et al., 2012).

### **Η μεταβολομική στη διάγνωση καρδιαγγειακών παθήσεων**

Οι καρδιαγγειακές ασθένειες είναι άλλη μία κατηγορία ασθενειών στις οποίες η μεταβολομική ανάλυση έχει αποδειχθεί ιδιαίτερα αξιόλογη (Kordalewska et al., 2015). Ενδεικτικά, βιοδείκτες που συνδέονται με το εγκεφαλικό επεισόδιο ταυτοποιήθηκαν μετά από μεταβολομική ανάλυση με φασματοσκοπία NMR των δειγμάτων πλάσματος και ούρων από ασθενείς με εγκεφαλικό έμφρακτο (Jung et al., 2011). Η ανάλυση δειγμάτων πλάσματος με την τεχνική MS ανέδειξε συγκεκριμένες ομάδες μεταβολιτών ως πιθανούς βιοδείκτες για την πορεία της στεφανιαίας νόσου (Shah et al., 2012). Ακόμη, η ανάλυση με την τεχνική LC-MS/MS δειγμάτων πλάσματος από ανθρώπους, καθώς και δειγμάτων πλάσματος και εγκεφαλονωτιαίου υγρού από ποντίκια αποκάλυψε ότι κατά τη διάρκεια οξέος καρδιοεμβολικού επεισοδίου παρατηρούνται μειωμένα επίπεδα βαλίνης, λευκίνης και ισολευκίνης (Kimberly et al., 2013). Τέλος, η μεταβολομική ανάλυση με GC/MS των δειγμάτων πλάσματος από νέους υπερτασικούς και υγιείς άντρες οδήγησε στην παρατήρηση διακριτών μεταβολικών αποτυπομάτων μεταξύ των δύο κατηγοριών όσον αφορά τα αμινοξέα γλυκίνη, λυσίνη και κυστεΐνη (Wang et al., 2015).

### **Η μεταβολομική στη διάγνωση νευρολογικών ασθενειών**

Οι τεχνικές μεταβολομικής ανάλυσης έχουν σταθεί χρήσιμες όσον αφορά τη διάγνωση νευρολογικών ασθενειών. Η ανακάλυψη βιοδεικτών είναι ιδιαίτερα σημαντική για τη νόσο του Parkinson, αφού τα κλινικά κριτήρια αυτής της ασθένειας δεν παρουσιάζουν ιδιαίτερα υψηλή διαγνωστική ακρίβεια. Πράγματι, η ανάλυση δειγμάτων πλάσματος από ασθενείς με τη νόσο του Parkinson και δειγμάτων ελέγχου με την τεχνική LC-MS οδήγησε στον προσδιορισμό δύο βιοδεικτών (ουρικό οξύ, γλουταθιόνη) για τη νόσο του Parkinson (Bogdanov et al., 2008). Αντίστοιχα, η ανάλυση με την τεχνική LC-MS των δειγμάτων εγκεφαλικού ιστού από ασθενείς με τη νόσο του Alzheimer έδειξε ότι σε πολύ πρώιμα στάδια της νόσου παρουσιάζεται μειωμένη συγκέντρωση σουλφιδίων και αυξημένη συγκέντρωση κεραμιδίων (Han et al., 2002). Μεταγενέστερη έρευνα επικεντρώθηκε στη μεταβολομική ανάλυση με δύο αναλυτικές πλατφόρμες των δειγμάτων ορού από ασθενείς με τη νόσο του Alzheimer, από ασθενείς με ήπια νοητική έκπτωση (mild cognitive impairment, MCI) και από υγιή άτομα, πόρισμα της οποίας ήταν η συσχέτιση της νόσου Alzheimer με φαινόμενα όπως η υποξία και το οξειδωτικό στρες (Oresik et al., 2011). Όσον αφορά την πολλαπλή σκλήρυνση, η μεταβολομική ανάλυση του εγκεφαλονωτιαίου υγρού ποντικών με δύο αναλυτικές τεχνικές (LC-MS, GC-MS) αποκάλυψε μεταβολές στα επίπεδα μεταβολιτών του εγκεφαλονωτιαίου υγρού σε συνάρτηση με το στάδιο της ασθένειας (Noga et al., 2012).

#### **4.1.4.2. Φαρμακομεταβολομική**

Η ανάπτυξη νέων φαρμάκων είναι μία διαδικασία με υψηλές χρονικές και οικονομικές απαιτήσεις και με υψηλό ρίσκο για τη φαρμακευτική βιομηχανία. Το παραδοσιακό επιχειρηματικό μοντέλο των μεγάλων φαρμακευτικών εταιριών βασίζεται στον προσδιορισμό νέων υποσχόμενων φαρμάκων, τη διεξαγωγή μεγάλων, οικονομικά απαιτητικών κλινικών δοκιμών και την εμπορική προώθηση των επιτυχημένων φαρμάκων στις ανεπτυγμένες χώρες (Kessel, 2011). Όπως αναφέρθηκε και στο 3<sup>ο</sup> κεφάλαιο, κατά μέσο όρο η ανάπτυξη ενός νέου φαρμάκου απαιτεί 12-15 χρόνια και πάνω από \$1 δις, και κάθε εταιρία παράγει περίπου μία νέα χημική οντότητα κάθε χρόνο (Hughes et al., 2011; Kessel,

2011). Επίσης, υπάρχει κίνδυνος (~5%) για ένα φάρμακο που έχει διοχετευθεί στην αγορά να αποσυρθεί εξ' αιτίας αρνητικών παρενεργειών (Smith & Schmid, 2006).

Η χαμηλή απόδοση της διαδικασίας ανάπτυξης νέων φαρμάκων μπορεί να βελτιωθεί χρησιμοποιώντας διάφορες μεθόδους, μία από τις οποίες είναι η μεταβολομική. Μάλιστα, έχει εισαχθεί ο όρος της φαρμακομεταβολικής, δηλαδή το ερευνητικό πεδίο που έχει ως αντικείμενο την απόκριση ενός ασθενή σε ξενοβιοτικούς παράγοντες από την οπτική γωνία του μεταβολισμού (Baraldi et al., 2009). Ξενοβιοτικοί είναι οι παράγοντες που είναι ξένοι στο βιολογικό σύστημα, και οι οποίοι μπορεί να προέρχονται από το περιβάλλον, συμπεριλαμβάνοντας τη διαίτα και τη φαρμακευτική αγωγή. Η κατάποση των ξενοβιοτικών ουσιών από τον οργανισμό ακολουθείται από την εξουδετέρωσή τους από τον οργανισμό μέσα από ενζυματικές αντιδράσεις, μία διαδικασία που ονομάζεται ξενοβιοτικός μεταβολισμός (Koen et al., 2015).

Οι μεταβολίτες μπορούν να λειτουργήσουν όχι μόνο ως βιοδείκτες ασθενειών, αλλά και ως βιοδείκτες της αποτελεσματικότητας ενός φαρμάκου, επιτρέποντας το γρήγορο έλεγχο των νέων χημικών οντοτήτων σε ζωικά μοντέλα και *in vitro* δοκιμασίες (Kumar et al., 2014). Με τη μεταβολομική, τα φάρμακα που έχουν αυξημένες πιθανότητες να αποτύχουν λόγω τοξικότητας στα στάδια της κλινικής ανάπτυξης μπορούν να εντοπιστούν πιο εύκολα στα στάδια της προκλινικής μελέτης (Zheng et al., 2009). Η μεταβολομική δυννητικά προσφέρει ένα οικονομικό και εύκολο εργαλείο για παρακολούθηση της επίδρασης ενός φαρμάκου σε ασθενείς και έχει τη δυνατότητα να επηρεάσει τη διαδικασία ανάπτυξης νέων φαρμάκων χαμηλώνοντας το οικονομικό και χρονικό κόστος που συνδέονται με την ανάπτυξη και προώθηση μίας νέας χημικής οντότητας (Schnackenberg & Beger, 2007; Kumar et al., 2014).

Η ανίχνευση βιοδεικτών μπορεί να πραγματοποιηθεί με τη χρήση βιολογικών υγρών, απαλλάσσοντας από την ανάγκη για εκτέλεση παρεμβατικών διαδικασιών όπως η βιοψία (Koen et al., 2015). Η πιο ευρεώς χρησιμοποιούμενη κατηγορία βιολογικού υγρού στις μελέτες τοξικότητας είναι τα ούρα. Ένα πλεονέκτημα της χρήσης δειγμάτων ούρων ή πλάσματος στα πειράματα τοξικότητας είναι ότι το δείγμα μπορεί να ληφθεί μη επεμβατικά. Επίσης, πολλαπλά δείγματα μπορούν να συλλεχθούν από ένα συγκεκριμένο υποκείμενο στη διάρκεια του χρόνου, επιτρέποντας τον προσδιορισμό μίας μεταβολικής πορείας που αντανακλά την τοξική απόκριση σε ένα φάρμακο και την περίοδο ανάκαμψης του υποκειμένου. Ακόμη, η ανάλυση των μεταβολιτών στα βιολογικά υγρά επιτρέπει την τοξικολογική αξιολόγηση της κατάστασης πολλών διαφορετικών οργάνων του ίδιου υποκειμένου στο χρόνο. Επιπλέον, το ίδιο υποκείμενο μπορεί να χρησιμοποιηθεί σε αρκετές χρονικές στιγμές, με αποτέλεσμα να μειώνεται ο αριθμός των υποκειμένων που απαιτούνται για τη διεκπεραίωση μίας μελέτης τοξικότητας (Schnackenberg & Beger, 2007). Τα επίπεδα φαρμάκων και μεταβολιτών φαρμάκων μπορούν να ανιχνευθούν με τη χρήση αναλυτικών τεχνικών (Kumar et al., 2014).

### **Η μεταβολομική στη διερεύνηση της τοξικότητας φαρμάκων**

Υπάρχει πάντα μία ισορροπία μεταξύ της αποτελεσματικότητας και της ασφάλειας ενός φαρμάκου, και πολλά από τα διαθέσιμα στο εμπόριο φάρμακα πρέπει να λαμβάνονται με προσοχή προς αποφυγή των τοξικών παρενεργειών τους. Επίσης, η επίδραση ενός φαρμάκου σε διαφορετικά άτομα παρουσιάζει μεγάλη διακύμανση ανάλογα με την ηλικία, το φύλο και το σύνολο γονιδίων του, ώστε η δοσολογία που είναι ασφαλής για ένα άτομο μπορεί να αποδειχθεί μοιραία για ένα άλλο. Επομένως, στο στάδιο ανάπτυξης ενός νέου

φαρμάκου υπάρχει η επιτακτική ανάγκη για παρακολούθηση της επίδρασής του σε μεμονωμένους ασθενείς. (Kumar et al., 2014) Μία σημαντική περιοχή φαρμακευτικής έρευνας είναι η αναγνώριση βιοδεικτών τοξικότητας που μπορούν να χρησιμοποιηθούν σε προκλινικές και κλινικές μελέτες φαρμάκων. Υπάρχει έντονη ανάγκη για ανάπτυξη νέων βιοδεικτών που μπορούν να προβλέψουν με ακρίβεια την τοξικότητα των νέων χημικών οντοτήτων στα πρώιμα στάδια της φαρμακευτικής ανάπτυξης (Schnackenberg & Beger, 2007).

Αρκετές μεταβολομικές έρευνες έχουν ταυτοποιήσει καθολικούς, μη εξειδικευμένους βιοδείκτες που αφορούν τη γενική τοξικότητα των φαρμάκων και δε συνδέονται με συγκεκριμένο τύπο, μηχανισμό ή σημείο δράσης φαρμάκου (Koen et al., 2015). Οι μεταβολίτες αυτοί είναι αντανάκλαση των δευτερευόντων παρενεργειών της κατάποσης ενός φαρμάκου και συνδέονται με γενικές μεταβολές όσον αφορά τις ενεργειακές απαιτήσεις, το μεταβολισμό, και τους εντερικούς μικροοργανισμούς (Keun, 2006). Εκτός από τους βιοδείκτες γενικής τοξικότητας, η μεταβολομική δίνει επίσης τη δυνατότητα ταυτοποίησης βιοδεικτών που σχετίζονται με συγκεκριμένους μηχανισμούς φαρμάκων και συγκεκριμένες παρενέργειες σε ζωτικά όργανα. Παραδείγματα τέτοιων ερευνών αφορούν την ηπατοτοξικότητα (Schnackenberg et al., 2007) και τη νεφροτοξικότητα (Garrod et al., 2005; Lenz et al., 2000).

Αρκετές έρευνες έχουν ανακαλύψει συσχετίσεις μεταξύ διάφορων ειδών τοξικότητας με συγκεκριμένα επίπεδα μεταβολιτών χρησιμοποιώντας αναλυτικές τεχνικές (Kumar et al., 2014). Για παράδειγμα, η ανάλυση με την τεχνική HPLC-MS των δειγμάτων ούρων από αρουραίους είχε ως αποτέλεσμα η οφειλόμενη στην ουσία σερίνη νεφροτοξικότητα να συσχετιστεί με αυξημένα επίπεδα συγκεκριμένων μεταβολιτών (λακτικό οξύ, φαινυλαλανίνη, τρυπτοφάνη, τυροσίνη, βαλίνη) (Williams et al., 2005). Αντίστοιχα, η οφειλόμενη στην ουσία γενταμικίνη νεφροτοξικότητα έχει συσχετιστεί με αυξημένα επίπεδα γλυκόζης και μειωμένα επίπεδα ξανθουρενικού και κυνουρενικού οξέος σε ούρα (Lenz et al., 2005).

Ένα παράδειγμα μελέτης με επίκεντρο την οφειλόμενη σε φάρμακα τοξικότητα είχε ως αντικείμενο την ανάλυση φασμάτων  $^1\text{H-NMR}$  από δείγματα ούρων και ιστό ήπατος ποντικών που είχαν δεχθεί θεραπεία με συγκεκριμένο συστατικό, συσχετισμένο με ηπατοτοξικότητα, και ποντικών που δεν είχαν δεχθεί τη συγκεκριμένη θεραπεία. Η μεταβολομική ανάλυση των δειγμάτων είχε ως αποτέλεσμα το διαχωρισμό των δύο ομάδων με βάση συγκεκριμένους μεταβολίτες και συνέβαλε στη διαλεύκανση του μηχανισμού της τοξικότητας του συγκεκριμένου φαρμάκου. (Mortishire-Smith et al., 2004)

Μεταβολομικές προσεγγίσεις έχουν χρησιμοποιηθεί και για την εκτίμηση τοξικότητας σε *in vitro* δείγματα. Για παράδειγμα, έχει εκτιμηθεί η τοξικότητα των πεπτιδίων σε κύτταρα ωοθηκών κινεζικών κρικητών (Chinese hamster ovary - CHO) (Kilk et al., 2009), η τοξικότητα του καπνού από τσιγάρο στα επιθηλιακά κύτταρα πνεύμονα (Vulimiri et al., 2009), καθώς και η νευροτοξικότητα του υδραργύρου και της καφεΐνης σε καλλιέργειες εγκεφαλικών κυττάρων (van Vliet et al., 2008) (Robertson et al., 2011).

### **Η μεταβολομική στην κατανόηση θεραπευτικής απόκρισης**

Εκτός από τη διαλεύκανση των μηχανισμών λειτουργίας των φαρμάκων, η μεταβολομική μπορεί να συμβάλει στην κατανόηση της απόκρισης ενός οργανισμού σε μία φαρμακευτική αγωγή. Ένα παράδειγμα μίας σχετικής ερευνητικής μελέτης αφορά τις στατίνες, που χορηγούνται ως φαρμακευτική αγωγή σε ασθενείς με αυξημένα επίπεδα LDL



χοληστερόλης και υψηλό ρίσκο εκδήλωσης καρδιαγγειακών ασθενειών, αν και η απόκριση των ασθενών σε αυτές παρουσιάζει σημαντική διακύμανση. Το φαινόμενο αυτό μελετήθηκε από τους Trupp et al. (2012), οι οποίοι ανέλυσαν με την τεχνική GC-MS δείγματα πλάσματος ασθενών λημμένα πριν και 6 εβδομάδες μετά από την εκκίνηση της θεραπείας, επιτυγχάνοντας την ταυτοποίηση των μεταβολιτών που ευθύνονται για τη διακύμανση της απόκρισης στη θεραπεία. (Trupp et al., 2012; Koen et al., 2015).

Το αντικείμενο μίας άλλης μελέτης ήταν η κατασκευή ενός προγνωστικού μοντέλου με βάση τέσσερις ταυτοποιημένους βιοδείκτες ορού, με στόχο την πρόβλεψη του αποτελέσματος της εισαγωγικής χημειοθεραπείας καρκίνου μαστού. Το μοντέλο αυτό κατάφερε να προβλέψει το βαθμό ανταπόκρισης στη θεραπεία (πλήρης ανταπόκριση ή μείωση του όγκου μικρότερη του 50%) με εξειδίκευση 100% και ευαισθησία 80%. (Wei et al., 2013; Koen et al., 2015)

Τέλος, οι Das et al. (2015) διερεύνησαν την επιρροή της θεραπείας φυματίωσης στα μεταβολικά προφίλ ασθενών με φυματίωση σε συγκεκριμένα χρονικά διαστήματα. Η ανάλυση των δειγμάτων ούρων των ασθενών απέδειξε την ύπαρξη συσχέτισης μεταξύ των μεταβολικών προφίλ των ασθενών και του στάδιου θεραπείας, αφού τα μεταβολικά προφίλ των ασθενών μετά από κάθε θεραπεία προσέγγιζαν τα μεταβολικά προφίλ των υγιών ατόμων. (Das et al., 2015; Koen et al., 2015)

### **Η μεταβολομική στην κατανόηση των μηχανισμών δράσης φαρμάκων**

Η κατανόηση του μεταβολισμού και του μηχανισμού δράσης ενός φαρμάκου είναι απαραίτητη πριν τη χορήγησή του, και μπορεί να γίνει εφικτή με τη βοήθεια της φαρμακομεταβολομικής. Η φαρμακομεταβολομική μπορεί να συμβάλλει στην κατανόηση του μηχανισμού δράσης ενός φαρμάκου με διάφορους τρόπους, όπως τη σύγκριση των μεταβολικών προφίλ μίας ομάδας ασθενών στους οποίους χορηγείται ένα εικονικό φάρμακο με τα μεταβολικά προφίλ μίας ομάδας ασθενών στους οποίους χορηγείται η ξενοβιοτική θεραπεία, καθώς και τη σύγκριση των μεταβολικών προφίλ των άγριου τύπου ζώων και των γενετικά τροποποιημένων ζώων που λαμβάνουν την ξενοβιοτική θεραπεία υπό εξέταση (Lan & Jia, 2010; Koen et al., 2015).

Ένα παράδειγμα διερεύνησης του μηχανισμού δράσης ενός φαρμάκου με τη βοήθεια της φαρμακομεταβολομικής είναι η μελέτη των Rozen et al. (2005). Σύμφωνα με αυτή την έρευνα, η διαφοροποίηση μεταξύ ασθενών με νόσο κινητικού νευρώνα (motor neuron disease, MND) και υγιών ατόμων, καθώς και η διαφοροποίηση μεταξύ ασθενών με MND με ή χωρίς θεραπεία με το φάρμακο Riluzole, έγινε εφικτή χρησιμοποιώντας τους μεταβολίτες πλάσματος των ατόμων αυτών. Συγκεκριμένα, οι ασθενείς που ακολουθούσαν θεραπεία με το Riluzole συσχετίστηκαν με δύο μεταβολίτες, οι οποίοι δεν αποτελούν παράγωγα του φαρμάκου αλλά αντικατοπτρίζουν αλλαγές στο μεταβόλωμα του ασθενή ως απόρροια του μεταβολισμού του φαρμάκου (Rozen et al., 2005; Koen et al., 2015).

Η φαρμακομεταβολομική μπορεί να βοηθήσει στη διερεύνηση των διακυμάνσεων στο μεταβολισμό φαρμάκων εξ' αιτίας γενετικών παραγόντων, ανατροφοδοτώντας μελέτες φαρμακογενετικής. Για παράδειγμα, οι Ji et al. (2011) διερεύνησαν τους μεταβολίτες στα ούρα ατόμων με μείζονα καταθλιπτική διαταραχή (major depressive disorder) που υπόκεινται σε θεραπεία με εκλεκτικούς αναστολείς επαναπρόσληψης σεροτονίνης (Selective serotonin reuptake inhibitors, SSRIs). Κατά μέσο όρο, το 40% των ασθενών δεν ανταποκρίνονται σε αυτή τη θεραπεία και προηγούμενες φαρμακογενετικές μελέτες απέτυχαν στο να

προσδιορίσουν πιθανούς γενετικούς παράγοντες που θα μπορούσαν να χρησιμοποιηθούν για την πρόβλεψη του αποτελέσματος της θεραπείας με SSRIs. Από τα μεταβολομικά προφίλ που αποκτήθηκαν στα πλαίσια της εν λόγω μελέτης, η μειωμένη απόκριση στη θεραπεία συσχετίστηκε με αυξημένα επίπεδα γλυκίνης. Το συμπέρασμα αυτό πυροδότησε νέες φαρμακογενομικές μελέτες, οι οποίες προσδιόρισαν πολυμορφισμούς σε γονίδιο που συνδέεται με τη γλυκίνη. Η παραπάνω έρευνα δείχνει ότι η μεταβολομική μπορεί να παίζει σημαντικό ρόλο στην υποστήριξη ή στην πυροδότηση φαρμακογενομικών μελετών, συμβάλλοντας στον προσδιορισμό γενετικών παραγόντων που σχετίζονται με ατομικές διακυμάνσεις στο μεταβολισμό φαρμάκων. (Ji et al., 2011; Koen et al., 2015)

## **4.2. Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού**

### **4.2.1 Εισαγωγή**

Η Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance, NMR) είναι μία αναλυτική τεχνική που αξιοποιεί τις μαγνητικές ιδιότητες συγκεκριμένων πυρήνων ατόμων με ασύζευκτα πυρηνικά spin. Έτσι, σε αντίθεση με τις περισσότερες φασματοσκοπικές τεχνικές αφορά στους πυρήνες των ατόμων του δείγματος υπό μελέτη, και όχι τα ηλεκτρόνιά τους. Συχνά μελετούμενοι πυρήνες με αυτή την τεχνική είναι οι  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ ,  $^{31}\text{P}$ . Τα βασικά σημεία ενός πειράματος NMR είναι τα εξής. Το δείγμα εισάγεται σε ομογενές στατικό μαγνητικό πεδίο, με αποτέλεσμα οι μαγνητικές ροπές των πυρήνων του δείγματος να παίρνουν συγκεκριμένους προσανατολισμούς ως προς την ένταση του μαγνητικού πεδίου και να έρχονται σε κατάσταση ισορροπίας. Στη συνέχεια, εφαρμόζονται παλμοί ραδιοσυχνοτήτων, με αποτέλεσμα οι πυρήνες του δείγματος να απορροφούν ηλεκτρομαγνητική ενέργεια και να πραγματοποιούν μεταβάσεις μεταξύ των ενεργειακών επιπέδων. Με το τέλος των παλμών ραδιοσυχνοτήτων, το δείγμα επανέρχεται στην κατάσταση ισορροπίας εκπέμποντας την ηλεκτρομαγνητική ακτινοβολία που απορρόφησε, από την οποία προκύπτει το σήμα NMR.

Η φασματοσκοπία NMR έχει διάφορες εφαρμογές, η κυριότερη από τις οποίες είναι η ταυτοποίηση της μοριακής δομής των χημικών ενώσεων. Επίσης, αξιοποιείται για τη μελέτη της κινητικής, δυναμικής και θερμοδυναμικής των χημικών συστημάτων, τον προσδιορισμό των προϊόντων των χημικών αντιδράσεων, καθώς και τον προσδιορισμό των χημικών ουσιών σε ένα δείγμα. (Bothwell & Griffin, 2011; Keeler, 2010)

Η φασματοσκοπία NMR παρουσιάζει αρκετά πλεονεκτήματα που την έχουν καθιερώσει ως μία από τις κύριες τεχνικές για τον προσδιορισμό της δομής οργανικών ενώσεων. Κατ' αρχάς, η τεχνική αυτή είναι μη καταστρεπτική γιατί η ενέργεια που απαιτείται για τη διέγερση ενός πυρήνα είναι αρκετά μικρή. Επιπλέον, η τεχνική αυτή παρέχει τη δυνατότητα για ανάλυση ολόκληρου του φάσματος συχνοτήτων. Όπως προαναφέρθηκε, η φασματοσκοπία NMR μπορεί να χρησιμοποιηθεί αυτόνομα, χωρίς να είναι απαραίτητη η συνδυαστική χρήση με κάποια άλλη τεχνική. Ακόμη, είναι μία αναπαραγώγιμη και αξιόπιστη μέθοδος, αφού με δεδομένο ότι τηρείται ένα τυπικό λειτουργικό πρωτόκολλο, τυχόν αποκλίσεις μεταξύ δειγμάτων προέρχονται από βιολογικές, και όχι οργανικές διακυμάνσεις. Τα κύρια μειονεκτήματα της φασματοσκοπίας NMR είναι η χαμηλή ευαισθησία και η χαμηλή χωρική ανάλυση. (Bothwell & Griffin, 2011)

Υπάρχουν διάφορες κατηγορίες πειραμάτων NMR, με κυριότερο κριτήριο διαφοροποίησης τον αριθμό διαστάσεων του σήματος NMR. Τα μονοδιάστατα (1D)

πειράματα έχουν ως έξοδο ένα σήμα στο πεδίο χρόνου, από το μετασχηματισμό Fourier του οποίου προκύπτει ένα φάσμα συχνοτήτων με έναν άξονα συχνότητας ( $x$ ) και έναν άξονα έντασης ( $y$ ) του φάσματος. Τα διδιάστατα (2D) πειράματα έχουν ως έξοδο ένα δισδιάστατο σήμα στο πεδίο χρόνου, από το δισδιάστατο μετασχηματισμό Fourier του οποίου προκύπτει ένα φάσμα συχνοτήτων με δύο άξονες συχνοτήτων ( $x,y$ ) και έναν τρίτο άξονα έντασης ( $z$ ) του σήματος. (Blümich, 2016)

Κάποιοι από τους σημαντικότερους ιστορικούς σταθμούς που συνδέονται με τη φασματοσκοπία NMR είναι οι εξής. Το 1896, ο P. Zeeman παρατήρησε το διαχωρισμό των φασματικών γραμμών από ένα ισχυρό μαγνητικό πεδίο, ένα έργο για το οποίο έλαβε το βραβείο Nobel Φυσικής το 1902. Το 1938, ο I.I. Rabi πραγματοποίησε ακριβείς μετρήσεις πυρηνικών μαγνητικών ροπών, ένα έργο για το οποίο έλαβε το βραβείο Nobel Φυσικής το 1944. Το 1946, ο F. Bloch και ο E.M. Purcell, ανεξάρτητα ο ένας από τον άλλο, επέδειξαν επιτυχώς το φαινόμενο του Πυρηνικού Μαγνητικού Συντονισμού για συμπυκνωμένη ύλη, σηματοδοτώντας την έναρξη του NMR (Bloch et al., 1946; Purcell et al., 1945). Για το έργο αυτό μοιράστηκαν το βραβείο Nobel Φυσικής το 1952. Το 1952 διατέθηκε στην αγορά ο πρώτος εμπορικός φασματογράφος (Varian 30 Hz). Το 1961 διατέθηκε στην αγορά ο φασματογράφος Varian A-60, ο οποίος χρησιμοποιήθηκε για τη μελέτη πρωτονίου NMR στα 60 MHz και αποδείχθηκε ο πρώτος εμπορικός φασματογράφος NMR με αναπαράξιμα αποτελέσματα. Την επόμενη χρονιά διατέθηκε ο πρώτος φασματογράφος NMR με υπεραγωγίμο μαγνήτη (Varian 220 MHz). Το 1985, ο K. Wüthrich προσδιόρισε για πρώτη φορά την τρισδιάστατη δομή πρωτεΐνης χρησιμοποιώντας φασματοσκοπία NMR, και τα επόμενα χρόνια συνέχισε με τον προσδιορισμό της δομής άλλων πρωτεϊνών. Για το έργο αυτό βραβεύθηκε με Nobel Φυσικής το 2002. (Emsley & Feeney, 2007)

#### 4.2.2. Βασικές αρχές φυσικής

Η κατανόηση της λειτουργίας της φασματοσκοπίας NMR απαιτεί την κατανόηση συγκεκριμένων εννοιών και φαινομένων φυσικής. Στο επίκεντρο της φασματοσκοπίας NMR βρίσκεται ο πυρήνας, ένα φορτισμένο σωματίδιο που διαθέτει ορισμένες ιδιότητες. Ο πυρήνας συντίθεται από πρωτόνια και νετρόνια, τα οποία συλλογικά ονομάζονται νουκλεόνια. Μαζικός αριθμός ( $A$ ) του πυρήνα είναι ο αριθμός των νουκλεονίων του, και ατομικός αριθμός ( $Z$ ) είναι ο αριθμός των πρωτονίων του. Κάθε πυρήνας έχει στροφομή (angular momentum,  $p$ ), ένα διανυσματικό μέγεθος που χαρακτηρίζει την περιστροφική κίνηση και είναι ανάλογο με το μέγεθος της ορμής (momentum) που χαρακτηρίζει τη γραμμική κίνηση. Όταν ένας πυρήνας βρίσκεται μέσα σε ένα μαγνητικό πεδίο συμπεριφέρεται ως μαγνητικό δίπολο, το μέγεθος του οποίου αντιπροσωπεύεται από την ιδιότητα της μαγνητικής ροπής (magnetic moment,  $\mu$ ). Επίσης, κάθε νουκλεόνιο έχει την ιδιότητα της ιδιοστροφομής ή σπιν (spin,  $I$ ), που είναι μία μορφή γωνιακής ορμής. Το σπιν ενός πυρήνα είναι η συνισταμένη των σπιν των νουκλεονίων που τον αποτελούν, είναι κβαντισμένος αριθμός κι εξαρτάται από το μαζικό και ατομικό αριθμό του πυρήνα. Το σπιν αναφέρεται στο γεγονός ότι ο μαγνητισμός ενός πυρήνα μπορεί να θεωρηθεί ότι προκαλείται από την περιστρεφόμενη κίνηση του ηλεκτρικού του φορτίου. Οι πυρήνες με άρτιο μαζικό αριθμό και άρτιο ατομικό αριθμό έχουν μηδενικό σπιν ( $I=0$ ). Οι πυρήνες με άρτιο μαζικό αριθμό, περιττό ατομικό αριθμό έχουν ακέραιο σπιν ( $I=1,2,3,..$ ), και οι πυρήνες με περιττό μαζικό και ατομικό αριθμό έχουν ημιακέραιο σπιν ( $I=1/2, 3/2, 5/2,..$ ). Είναι αξιοσημείωτο

ότι μόνο οι πυρήνες με μη μηδενικό σπιν δίνουν ανιχνεύσιμο σήμα NMR, επομένως μόνο αυτοί οι πυρήνες μελετώνται με φασματοσκοπία NMR. (Bothwell & Griffin, 2011; Mlynarik, 2017; Keeler, 2010)

Κεντρικό φαινόμενο στη φασματοσκοπία NMR είναι η αλληλεπίδραση των πυρήνων με ένα εξωτερικό εφαρμοζόμενο μαγνητικό πεδίο (έστω  $\vec{B}_0$ ). Έστω πυρήνες με μη μηδενική ιδιοστροφορμή και μαγνητική ροπή ( $I \neq 0, \mu \neq 0$ ). Όταν δεν υπάρχει εξωτερικό μαγνητικό πεδίο ( $\vec{B}_0 = 0$ ), όλοι οι πυρήνες βρίσκονται στο ίδιο ενεργειακό επίπεδο, έχουν τυχαίο προσανατολισμό και πραγματοποιούν περιστροφική κίνηση γύρω από τον άξονά τους. Όταν εφαρμόζεται εξωτερικό ομογενές μαγνητικό πεδίο ( $\vec{B}_0 \neq 0$ ), έχει παρατηρηθεί ότι οι πυρήνες μπορούν να βρεθούν σε συγκεκριμένα ενεργειακά επίπεδα, τα οποία περιγράφονται από την εξίσωση  $E_m = -\gamma\hbar m B_0$ , όπου  $m = -I, -I + 1, \dots, I - 1, I$ . Στην απλή περίπτωση ενός πυρήνα με  $I = 1/2$  υπάρχουν 2 επιτρεπτά ενεργειακά επίπεδα, τα  $E_{-1/2} = E_\beta = +\frac{1}{2}\gamma\hbar B_0, E_{+1/2} = E_\alpha = -\frac{1}{2}\gamma\hbar B_0$ , τα οποία αντιστοιχούν σε προσανατολισμό του πυρήνα αντιπαράλληλα και παράλληλα με το πεδίο  $\vec{B}_0$ , αντίστοιχα. Η συμπεριφορά αυτή των πυρήνων υπό την επίδραση εξωτερικού μαγνητικού πεδίου ονομάζεται φαινόμενο Zeeman. Επίσης, υπό την επίδραση εξωτερικού μαγνητικού πεδίου ( $\vec{B}_0 \neq 0$ ) οι πυρήνες εκτελούν συγκεκριμένη περιστροφική κίνηση που ονομάζεται μετάπτωση Larmor (Larmor precession), στην οποία ο άξονας της μαγνητικής ροπής  $\mu$  περιστρέφεται γύρω από το  $\vec{B}_0$  με κυκλική συχνότητα  $\omega_L = -\gamma B_0$  (σε rad/s),  $\nu_L = -\frac{1}{2\pi}\gamma B_0$  (σε Hz), η οποία ονομάζεται συχνότητα Larmor. (Keeler, 2010)

Έχει αποδειχθεί ότι η κατανομή ενός συνόλου πυρήνων με  $I = 1/2$  στα ενεργειακά επίπεδα  $E_\alpha, E_\beta$  υπακούει στο νόμο κατανομής Boltzmann  $n^\beta/n^\alpha = e^{\frac{-\Delta E}{kT}} \times 0.9999382$ , όπου  $n^\beta$  ο πληθυσμός πυρήνων του υψηλότερου ενεργειακού επιπέδου ( $E_\beta$ ),  $n^\alpha$  ο πληθυσμός πυρήνων του χαμηλότερου ενεργειακού επιπέδου ( $E_\alpha$ ),  $\Delta E$  η ενεργειακή διαφορά μεταξύ των δύο ενεργειακών επιπέδων,  $k$  η σταθερά Boltzmann και  $T$  η θερμοκρασία δωματίου. Σύμφωνα με το νόμο αυτό, υπάρχει περίσσειμα πυρήνων στο χαμηλότερο ενεργειακό επίπεδο ( $E_\alpha$ ) με προσανατολισμό παράλληλο προς το πεδίο  $\vec{B}_0$ , κι επομένως η πιθανότητα παρατήρησης ενός πυρήνα στο χαμηλότερο ενεργειακό επίπεδο  $E_\alpha$  είναι λίγο μεγαλύτερη από ό,τι στο υψηλότερο ενεργειακό επίπεδο  $E_\beta$ . Σε αυτή την πληθυσμιακή διαφορά μεταξύ των ενεργειακών επιπέδων στηρίζεται η φασματοσκοπία NMR. Συγκεκριμένα, αν αυξηθεί η ένταση πεδίου  $B_0$ , θα αυξηθεί η ενεργειακή διαφορά  $\Delta E = E_\beta - E_\alpha = \gamma\hbar B_0$ , και κατ' επέκταση ο λόγος  $n^\beta/n^\alpha$ . Με δεδομένο ότι το άθροισμα  $n^\beta + n^\alpha$  είναι σταθερό, προκύπτει ότι θα αυξηθεί επίσης η πληθυσμιακή διαφορά  $n^\beta - n^\alpha$ . Επομένως, η αύξηση της έντασης πεδίου  $B_0$  συνεπάγεται την αύξηση της ευαισθησίας ενός πειράματος NMR. (Keeler, 2010)

Σε ένα πείραμα NMR παρατηρείται ένας μεγάλος αριθμός πυρήνων (π.χ.  $10^{20}$ ), οπότε είναι χρήσιμη η εξέταση της συνολικής συμπεριφορά τους. Αν οι μαγνητικές ροπές των επιμέρους πυρήνων του δείγματος ήταν προσανατολισμένες σε τυχαίες κατευθύνσεις, τότε τα μικρά μαγνητικά πεδία των επιμέρους πυρήνων θα αλληλομηδενίζονταν και δε θα υπήρχε παρατηρήσιμο συνολικό αποτέλεσμα. Αντιθέτως, στην κατάσταση ισορροπίας οι μαγνητικές ροπές των πυρήνων έχουν συγκεκριμένους προσανατολισμούς, τέτοιους ώστε το διανυσματικό άθροισμά τους είναι παράλληλο με το εξωτερικό πεδίο  $B_0$ . Ολική μαγνήτιση

( $\vec{M} = \vec{M}_x + \vec{M}_y + \vec{M}_z$ ) ενός δείγματος ονομάζεται λοιπόν το διανυσματικό άθροισμα των μαγνητικών ροπών των πυρήνων του δείγματος, από την οπτική γωνία της οποίας έχει ενδιαφέρον η εξέταση της αλληλεπίδρασης του δείγματος πυρήνων με ένα εξωτερικό μαγνητικό πεδίο (Keeler, 2010).

Όταν δεν υπάρχει εφαρμοζόμενο εξωτερικό μαγνητικό πεδίο, η μαγνήτιση του δείγματος είναι μηδενική ( $\vec{B}_0 = 0 \Rightarrow \vec{M} = 0$ ). Όταν υπάρχει εξωτερικό σταθερό ομογενές μαγνητικό πεδίο, το δείγμα έρχεται σε κατάσταση ισορροπίας, με τη μαγνήτιση  $\vec{M}$  του δείγματος να είναι παράλληλη με το εξωτερικό πεδίο ( $\vec{B}_0 \neq 0 \Rightarrow \angle(\vec{M}, \vec{B}_0) = 0 \Rightarrow \vec{M} = \vec{M}_z = \vec{M}_0 \uparrow \uparrow \vec{B}_0, \vec{M}_{xy} = \vec{M}_x + \vec{M}_y = 0$ ). Αν το δείγμα πυρήνων απορροφήσει επιπλέον ακτινοβολία, θα πραγματοποιηθούν μεταβάσεις των πυρήνων σε διαφορετικά ενεργειακά επίπεδα. Για να πραγματοποιηθεί η μετάβαση ενός πυρήνα σε ένα άλλο ενεργειακό επίπεδο από αυτό στο οποίο βρίσκεται στην κατάσταση ισορροπίας, πρέπει το ποσό της απορροφούμενης ενέργειας να είναι ίσο με την ενεργειακή διαφορά μεταξύ του αρχικού και του νέου ενεργειακού επιπέδου. Το παραπάνω φαινόμενο ονομάστηκε Πυρηνικός Μαγνητικός Συντονισμός και, όπως προαναφέρθηκε στην παράγραφο 4.2.1, ανακαλύφθηκε το 1946 από τους Purcell και Bloch (Purcell et al., 1946; Bloch et al., 1946). Σε μακροσκοπικό επίπεδο, η απορρόφηση ενέργειας έχει ως αποτέλεσμα το δείγμα να φύγει από την κατάσταση ισορροπίας. Συγκεκριμένα, η μαγνήτιση του δείγματος θα διαταραχθεί ( $\angle(\vec{M}, \vec{B}_0) \neq 0, \vec{M}_{xy} = \vec{M}_x + \vec{M}_y \neq 0$ ) και θα εκτελέσει περιστροφική κίνηση ως προς την κατεύθυνση του μαγνητικού πεδίου  $\vec{B}_0$  σχηματίζοντας έναν κώνο με σταθερή γωνία. Η κίνηση αυτή ονομάζεται μετάπτωση Larmor (Larmor precession), και η κυκλική συχνότητά της ονομάζεται συχνότητα μετάπτωσης Larmor  $\omega_L = -\gamma B_0$  (σε rad/s),  $\nu_L = -\frac{1}{2\pi}\gamma B_0$  (σε Hz). Η συχνότητα μετάπτωσης της μαγνήτισης γύρω από το πεδίο  $\vec{B}_0$  είναι η συχνότητα της γραμμής που θα εμφανιστεί στο φάσμα συχνοτήτων. Στη φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού, η διαταραχή της μαγνήτισης από την κατάσταση ισορροπίας σε άλλη θέση επιτυγχάνεται με την εφαρμογή ενός πεδίου ραδιοσυχνοτήτων (RF) που εκτελεί ταλάντωση επί του άξονα x στη συχνότητα Larmor. Το πεδίο αυτό (έστω  $\vec{B}_1$ ), αν και είναι αρκετά μικρό σε σχέση με το πεδίο  $\vec{B}_0$ , είναι ικανό να στρέψει τη μαγνήτιση εκτός του άξονα z. (Mlynarik, 2016; Keeler, 2010)

Όταν διακοπεί η ακτινοβολία του δείγματος, το δείγμα έρχεται στη λεγόμενη διαδικασία εφησυχασμού ή αποκατάστασης (relaxation), κατά την οποία οι διεγερμένοι πυρήνες επανέρχονται στα προηγούμενα ενεργειακά επίπεδά τους και η μαγνήτιση  $\vec{M}$  του δείγματος θα αρχίσει να επανέρχεται στη θέση ισορροπίας της στον άξονα z ( $\vec{M} = \vec{M}_z = \vec{M}_0 \uparrow \uparrow \vec{B}_0, \vec{M}_{xy} = \vec{M}_x + \vec{M}_y = 0$ ). Πιο συγκεκριμένα, η διαδικασία εφησυχασμού διακρίνεται στη διαμήκη μαγνητική επαναφορά (longitudinal magnetic relaxation), η οποία αναφέρεται στην επαναφορά της συνιστώσας  $\vec{M}_z$  στην τιμή ισορροπίας της ( $\vec{M}_z = \vec{M}_0$ ), και την εγκάρσια μαγνητική επαναφορά (transverse magnetic relaxation), η οποία αναφέρεται στην επαναφορά της συνιστώσας  $\vec{M}_{xy}$  στην τιμή ισορροπίας της ( $\vec{M}_{xy} = 0$ ). Οι χρόνοι  $T_1, T_2$  είναι οι χρόνοι διαμήκους και εγκάρσιας επαναφοράς, αντίστοιχα. Συνήθως ισχύει  $T_2 < T_1$ , γιατί η επαναφορά της συνιστώσας  $\vec{M}_z$  προκαλεί αυτόματα μείωση της  $\vec{M}_{xy}$ , και γιατί ο εφησυχασμός  $\vec{M}_{xy}$

επηρεάζεται όχι μόνο από το περιβάλλον, αλλά και από άλλους παράγοντες, όπως την αμοιβαία ανταλλαγή πυρηνικών σπιν ομοειδών πυρήνων. Οι παράγοντες που συμβάλλουν στον εφηρευσχασμό της μαγνήτισης σχετίζονται κυρίως με τη φύση του μορίου και ιδιαίτερα με τη δυνατότητα ανταλλαγής ευκίνητων πυρήνων, καθώς και με τις συνθήκες διεξαγωγής του πειράματος NMR (Mlynarik, 2016; Keeler, 2010)

Στη φασματοσκοπία NMR, το ανιχνεύσιμο σήμα είναι η μετάπτωση της μαγνήτισης  $\vec{M}$ . Η ανίχνευση του σήματος επιτυγχάνεται τοποθετώντας ένα μικρό πηνίο γύρω από το δείγμα, με τον άξονα του πηνίου στο επίπεδο xy. Το διάνυσμα μαγνήτισης «κόβει» το πηνίο, με αποτέλεσμα το πηνίο να ανιχνεύει τη συνιστώσα x της μαγνήτισης. Το σήμα που ανιχνεύεται είναι το λεγόμενο σήμα ελεύθερης επαγωγικής απόσβεσης, που στη συνέχεια ενισχύεται και καταγράφεται. (Keeler, 2010)

### 4.2.3. Σήμα Πυρηνικού Μαγνητικού Συντονισμού

#### 4.2.3.1. Σήμα στο πεδίο χρόνου

Το σήμα που ανιχνεύεται και καταγράφεται από το φασματογράφο NMR είναι το λεγόμενο σήμα ελεύθερης επαγωγικής απόσβεσης (free induction decay, FID). Αν υποθέσουμε ότι τη χρονική στιγμή που διακόπτονται οι παλμοί ραδιοσυχνότητας η μαγνήτιση βρίσκεται στον άξονα x ( $\beta=90^\circ$ ), οι συνιστώσες της μαγνήτισης  $M_x$ ,  $M_y$  εκφράζονται μαθηματικά από τις εξισώσεις  $M_x = M_0 \cos \omega t$ ,  $M_y = M_0 \sin \omega t$ , τότε το αντίστοιχο σήμα στο πεδίο χρόνου (έστω  $S(t)$ ) εκφράζεται από τις εξισώσεις  $S_x(t) = S_0 \cos \omega t$ ,  $S_y(t) = S_0 \sin \omega t \Rightarrow S(t) = S_x(t) + iS_y(t) = S_0 \cos \omega t + iS_0 \sin \omega t = S_0 e^{i\omega t}$ . Λαμβάνοντας υπόψιν και ότι η εγκάρσια μαγνήτιση μειώνεται εκθετικά στο χρόνο με σταθερά χρόνου  $T_2$ , το αντίστοιχο σήμα στο πεδίο χρόνου είναι  $S(t) = S_0 e^{i\omega t} e^{-\frac{t}{T_2}} = S_0 e^{i\omega t} e^{-R_2 t}$ ,  $R_2 = -\frac{1}{T_2}$  όπου  $R_2 = -\frac{1}{T_2}$  η σταθερά ταχύτητας. Όσο μικρότερος ο χρόνος  $T_2$ , τόσο πιο γρήγορα το σήμα μειώνεται. (Εικόνα 4.1(α)). (Keeler, 2010)

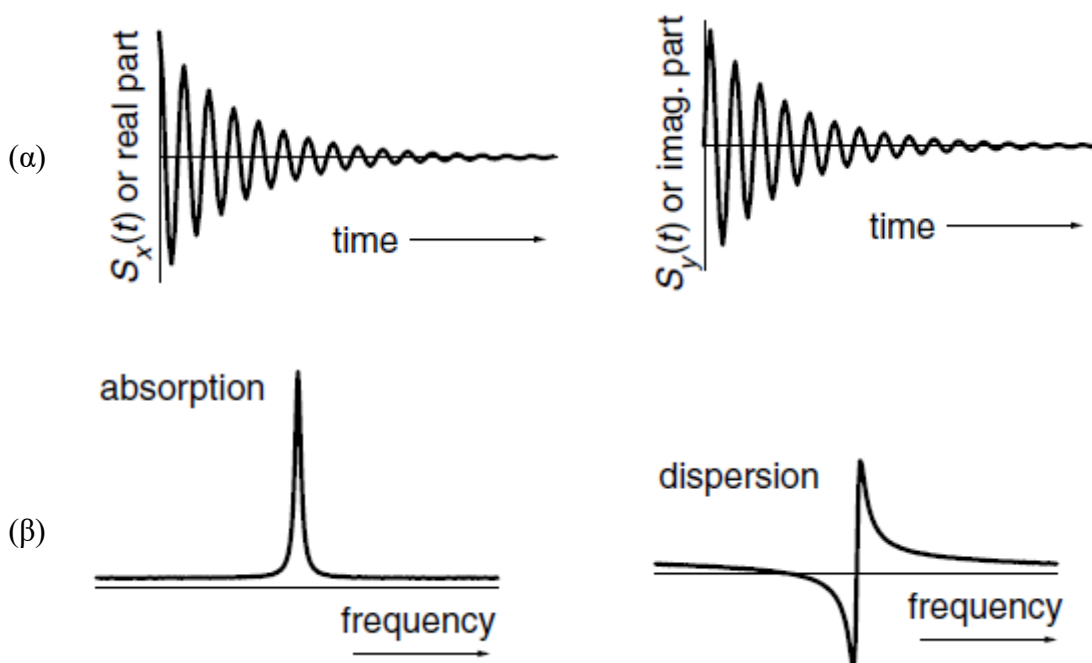
#### 4.2.3.2. Σήμα στο πεδίο συχνότητας

Το σήμα στο πεδίο συχνότητας (έστω  $s(\omega)$ ), ή αλλιώς φάσμα συχνοτήτων, προκύπτει μετά από την εφαρμογή του μετασχηματισμού Fourier στο σήμα ελεύθερης επαγωγής ( $S(t)$ ). Όπως και το σήμα ελεύθερης επαγωγής, το φάσμα έχει πραγματικό (έστω  $s_x(\omega)$ ) και φανταστικό μέρος (έστω  $s_y(\omega)$ ).

Το πραγματικό μέρος του φάσματος ( $s_x(\omega)$ ) διαμορφώνεται ως μία γραμμή σε absorption mode, η οποία παίρνει μόνο θετικές τιμές (Εικόνα 4.1(β)). Για σήμα στο πεδίο χρόνου που φθίνει εκθετικά ( $S(t) = S_0 e^{i\omega t} e^{-\frac{t}{T_2}}$ ), η φασματική γραμμή  $s_x(\omega)$  έχει τη μορφή μίας λορεντζιανής κορυφής. Το πλάτος της κορυφής αυτής στο μισό ύψος της ισούται με  $(1/\pi T_2)$  Hz, που συνεπάγεται ότι όσο πιο γρήγορη η μείωση του FID, τόσο μεγαλύτερο πλάτος θα έχει η κορυφή  $s_x(\omega)$ . Επίσης, η περιοχή ολοκλήρωσης κάτω από την κορυφή παραμένει σταθερή, επομένως η αύξηση του πλάτους της συνεπάγεται τη μείωση του ύψους της. Αν το ύψος του σήματος  $S(t)$  αυξηθεί, θα αυξηθεί ανάλογα και το ύψος της αντίστοιχης φασματικής κορυφής. Επομένως, ολοκληρώνοντας την περιοχή κάτω από τις φασματικές κορυφές μπορούμε να προσδιορίσουμε τον αριθμό πρωτονίων που συνεισφέρουν σε αυτές.

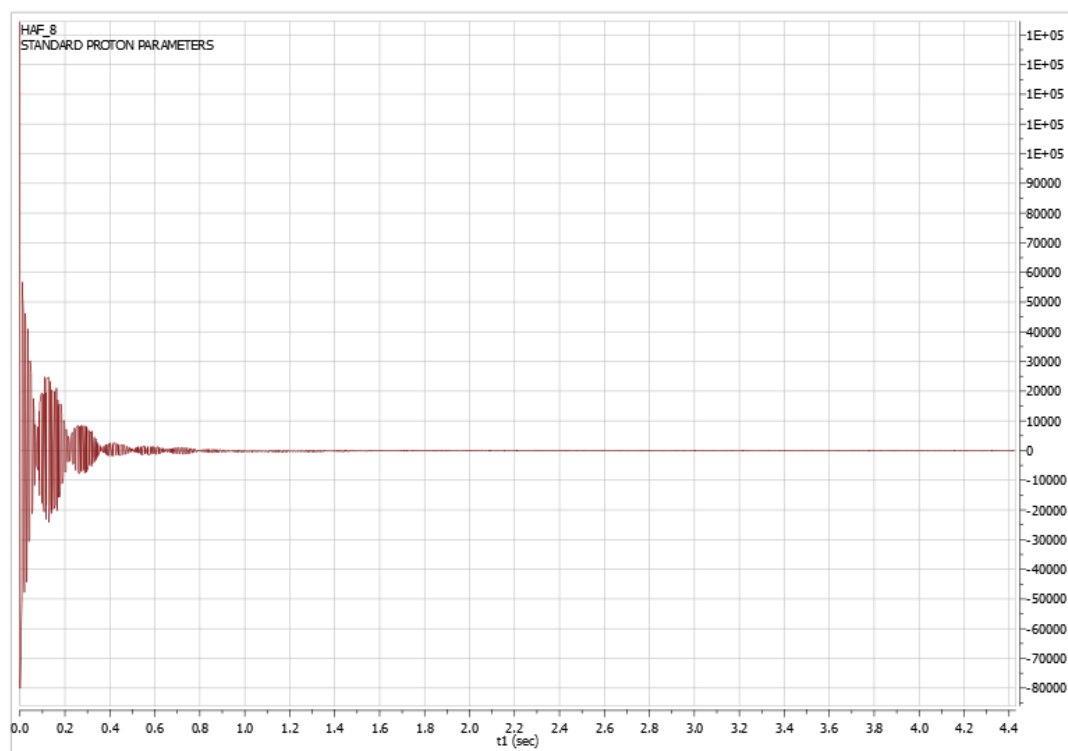
Το φανταστικό μέρος του φάσματος ( $S_y(\omega)$ ) διαμορφώνει μία γραμμή σε dispersion mode, που είναι επίσης λορεντζιανή κορυφή (Εικόνα 4.1(β)). Η γραμμή αυτής της μορφής παρουσιάζει κάποια μειονεκτήματα που αποτρέπουν τη χρήση της. Κατ' αρχάς, έχει μεγαλύτερο πλάτος από ό,τι η γραμμή του πραγματικού μέρους του φάσματος. Επίσης, έχει τόσο θετικές όσο και αρνητικές τιμές, οι οποίες σε ένα σύνθετο φάσμα μπορούν να αλληλοακυρωθούν προκαλώντας σύγχυση. (Keeler, 2010)

Τα σήματα FID και τα αντίστοιχα φάσματα συχνοτήτων που προκύπτουν από την πειραματική ανάλυση σύνθετων δειγμάτων είναι αρκετά πιο πολύπλοκα. Κάθε ουσία έχει ένα χαρακτηριστικό φάσμα, αποτελούμενο από ένα σύνολο συγκεκριμένων συχνοτήτων. Το φάσμα ενός δείγματος που περιέχει ένα σύνολο ουσιών αποτελεί το άθροισμα των φασμάτων των επιμέρους αυτών ουσιών, επομένως μπορεί να έχει αρκετά πολύπλοκη μορφή. Η Εικόνα 4.2 περιέχει το σήμα FID και το φάσμα συχνοτήτων ενός δείγματος αμνιακού υγρού και ενός δείγματος ορού.

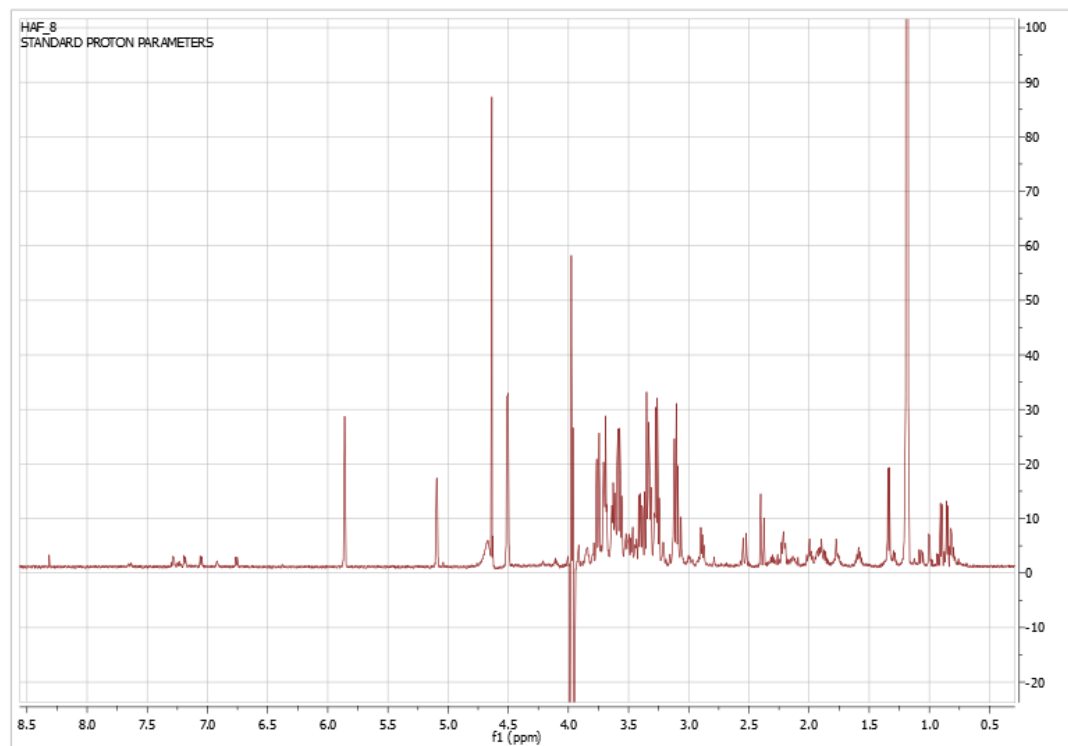


Εικόνα 4.1. (α) Απεικόνιση του πραγματικού και φανταστικού μέρους του σήματος FID  $S(t) = S_x(t) + iS_y(t) = S_0 \cos \omega t + iS_0 \sin \omega t = S_0 e^{i\omega t}$ . (β) Απεικόνιση μίας φασματικής γραμμής σε absorption και dispersion mode. (Πηγή: Keeler, 2010)

( $\alpha$ )

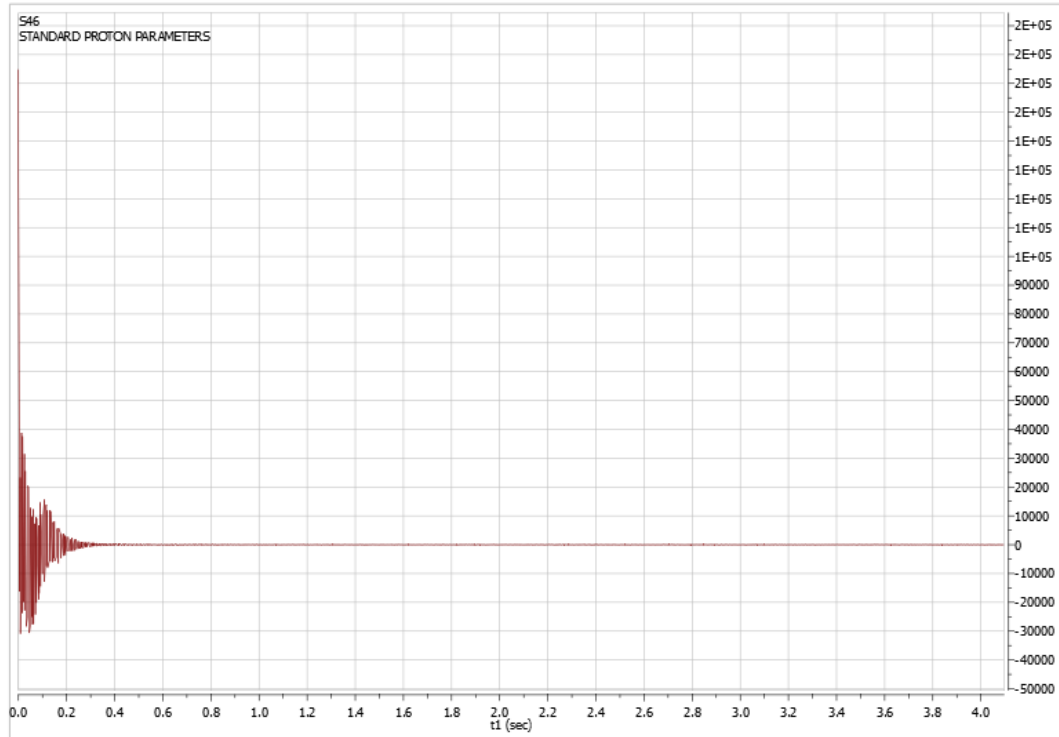


( $\beta$ )

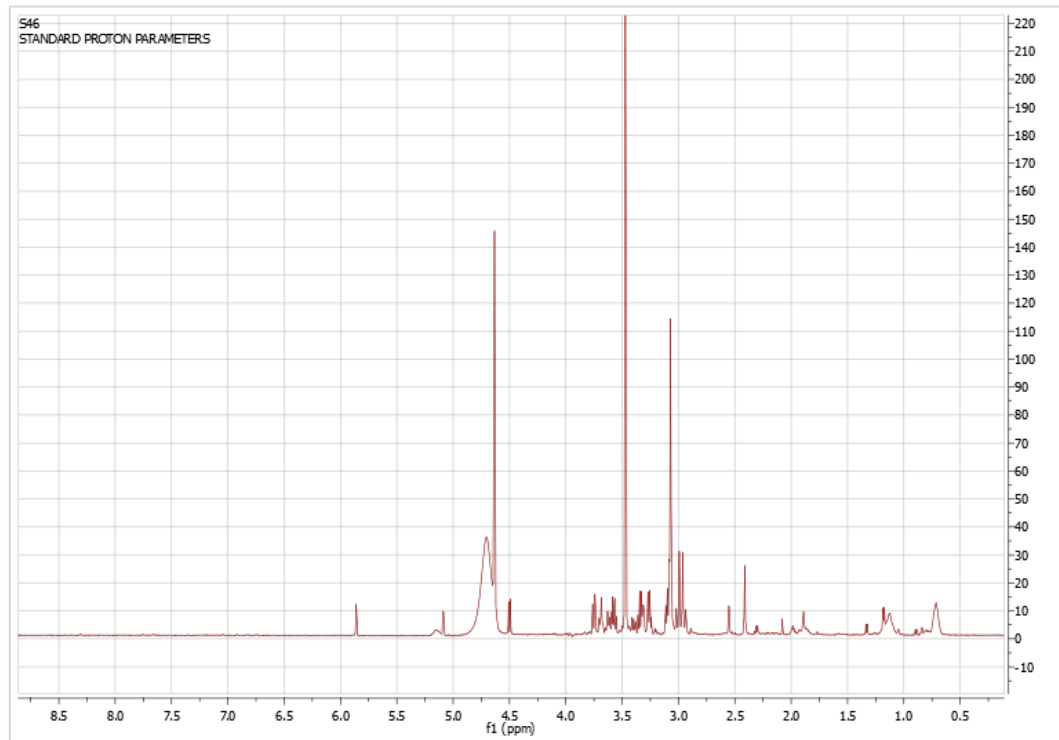




(γ)



(δ)



Εικόνα 4.2. Σήματα στο πεδίο χρόνου και στο πεδίο συχνότητας δύο βιολογικών δειγμάτων, όπως απεικονίζονται στο λογισμικό MestReNova. (α) Το πραγματικό μονοδιάστατο σήμα FID ενός δείγματος αμνιακού υγρού. (β) Τμήμα του μονοδιάστατου φάσματος συχνότητων που προέκυψε από το μετασχηματισμό Fourier του σήματος της εικόνας 4.2.(α). (γ) Το πραγματικό μονοδιάστατο σήμα FID ενός δείγματος ορού. (δ) Τμήμα του μονοδιάστατου φάσματος συχνότητων που προέκυψε από το μετασχηματισμό Fourier του σήματος της εικόνας 4.2.(γ).

### 4.2.3.3. Παράμετροι σήματος

#### Χημική μετατόπιση

Έχει παρατηρηθεί ότι οι πυρήνες του ίδιου ατόμου μπορούν να παρουσιάσουν διαφοροποιήσεις ως προς τη συχνότητα απορρόφησης. Με δεδομένο ένα στατικό εξωτερικό μαγνητικό πεδίο, τα ηλεκτρόνια που περιβάλλουν έναν πυρήνα παράγουν μικρά τοπικά μαγνητικά πεδία που αντιτίθενται στο εξωτερικό πεδίο, με αποτέλεσμα να μειώνουν την επίδρασή του στον πυρήνα. Το φαινόμενο αυτό ονομάζεται θωράκιση του πυρήνα, είναι ανάλογη του μαγνητικού πεδίου και εξαρτάται από το χημικό περιβάλλον του πυρήνα, που σημαίνει ότι είναι χαρακτηριστική της μοριακής δομής μίας χημικής ένωσης. (Mlynarik, 2016)

Το φαινόμενο θωράκισης είναι η αιτία για το ότι η συχνότητα απορρόφησης μίας χημικής ένωσης είναι ανάλογη με την ένταση του μαγνητικού πεδίου. Αν η φασματική κορυφή από την ένωση τετραμεθυλοσιλάνιο (tetramethylsilane, TMS) προκύπτει στα 400 MHz σε ένα φασματογράφο, η διπλασίαση του μαγνητικού πεδίου του φασματογράφου θα έχει ως αποτέλεσμα η αντίστοιχη φασματική κορυφή να προκύψει στα 800 MHz.

Για την αντιμετώπιση αυτού του προβλήματος, έχει εισαχθεί το μέγεθος της χημικής μετατόπισης (chemical shift), με βάση το οποίο γίνεται αναφορά στη συχνότητα μίας φασματικής κορυφής αναφορικά με τη συχνότητα μίας ένωσης αναφοράς. Πιο συγκεκριμένα, ο όρος της χημικής μετατόπισης αναφέρεται στη διαφοροποίηση των συχνοτήτων συντονισμού των πυρηνικών αυτοστροφορμών, η οποία προκαλείται από το χημικό περιβάλλον γύρω από τον πυρήνα. Η χημική μετατόπιση είναι ένας αριθμός που εκφράζει τη θέση της συχνότητας  $\nu$  μίας φασματικής κορυφής (σε Hz) σε σχέση με τη συχνότητα  $\nu_{\text{ref}}$  της ένωσης αναφοράς (σε Hz), και ορίζεται ως  $\delta = \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}}$ . Επίσης, η χημική μετατόπιση έχει οριστεί και ως  $\delta = \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{spec}}}$  όπου  $\nu_{\text{spec}}$  η συχνότητα του φασματογράφου. Η χημική μετατόπιση είναι μέγεθος αδιάστατο και ανεξάρτητο της έντασης του μαγνητικού πεδίου. Συνήθως πολλαπλασιάζεται με την ποσότητα  $10^6$  και εκφράζεται σε parts per million (ppm). Στη περίπτωση της φασματοσκοπίας NMR, η ένωση αναφοράς είναι το TMS. (Keeler, 2010; Mlynarik, 2016; “Chapter 13: Spectroscopy”, n.d.; “5-HMR-2 Chemical Shift”, 2017).

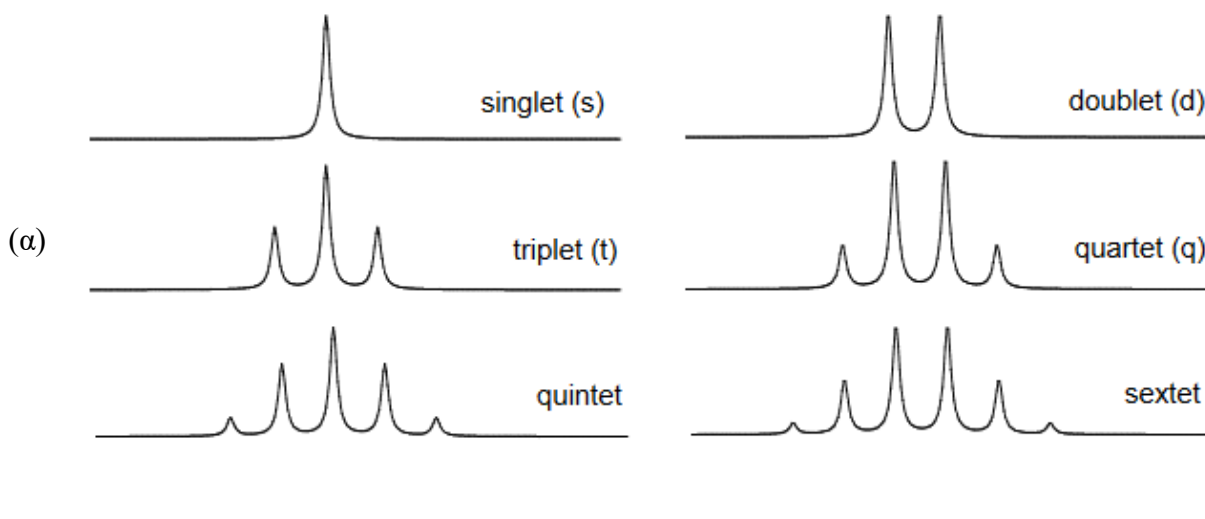
#### Σταθερά σύζευξης

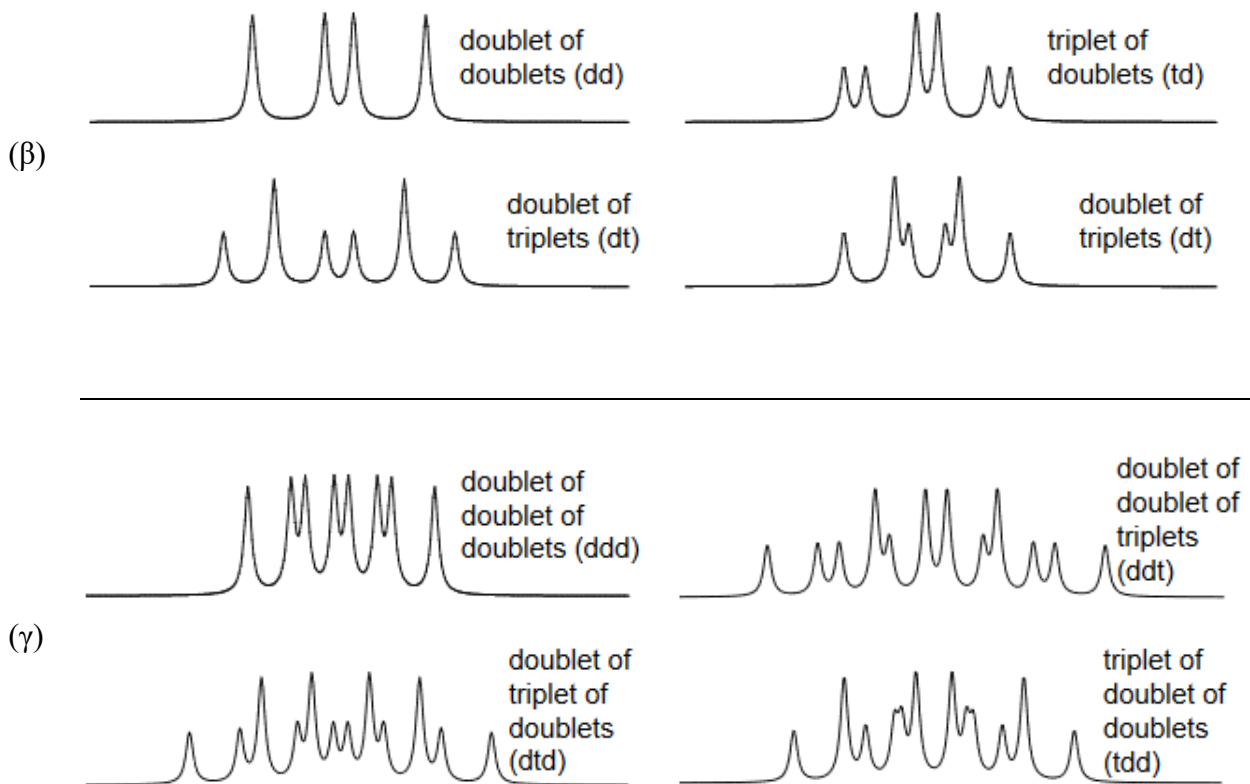
Ένα σημαντικό φαινόμενο που επηρεάζει την εμφάνιση των φασμάτων NMR είναι η σταθερά σύζευξης (J-coupling), που ευθύνεται για το διαχωρισμό των φασματικών γραμμών σε πολλαπλές κορυφές (multiplets). Η σύζευξη δύο πυρήνων ή ομάδων πυρήνων οφείλεται στους διαφορετικούς προσανατολισμούς των πυρηνικών μαγνητικών ορμών σε σχέση με το μαγνητικό πεδίο. Ο προσανατολισμός της μαγνητικής ροπής ενός πυρήνα παράλληλα με το μαγνητικό πεδίο μπορεί να αυξήσει την πυκνότητα ηλεκτρονίων σε ένα γειτονικό πυρήνα, ενώ ο προσανατολισμός της μαγνητικής ροπής αντιπαράλληλα με το μαγνητικό πεδίο μπορεί να την μειώσει. Το μέγεθος αυτής της αλληλεπίδρασης ονομάζεται σταθερά σύζευξης J κι εκφράζεται σε Hz. (Mlynarik, 2016)

Με βάση τον αριθμό των γειτονικών πυρήνων παρατηρούνται διαφορετικά μοτίβα διαχωρισμού των φασματικών γραμμών, με αποτέλεσμα να υπάρχουν πολλαπλές κορυφές A, B ή ανώτερης τάξης. Στην περίπτωση που ένα άτομο είναι συζευγμένο με ένα σύνολο n ατόμων με τιμή σταθεράς σύζευξης J, η φασματική κορυφή του ατόμου αυτού θα διαχωριστεί

σε  $n+1$  υποκορυφές, διαμορφώνοντας μία πολλαπλή κορυφή A τάξης. Οι πολλαπλές κορυφές A τάξης χαρακτηρίζονται από το ότι οι αναλογίες υψών των υποκορυφών είναι σύμφωνες με τις τιμές του τριγώνου Pascal και οι αποστάσεις μεταξύ των διαδοχικών υποκορυφών είναι ίσες με την τιμή σταθεράς σύζευξης  $J$ . Για παράδειγμα, αν ένα άτομο δεν είναι συζευγμένο με άλλα άτομα, η φασματική κορυφή του δε θα διασπαστεί, διαμορφώνοντας μία απλή κορυφή (singlet). Αν το άτομο είναι συζευγμένο με ένα άτομο με σταθερά σύζευξης  $J_1$ , η φασματική κορυφή του θα διασπαστεί σε δύο υποκορυφές που θα απέχουν κατά  $J_1$  Hz και τα ύψη τους θα έχουν την αναλογία που ορίζεται από τη δεύτερη γραμμή του τριγώνου Pascal (1:1), διαμορφώνοντας ένα doublet (d). Αν το άτομο είναι συζευγμένο με δύο άτομα με σταθερά σύζευξης  $J_2$ , η φασματική κορυφή του θα διασπαστεί σε τρεις υποκορυφές, η κάθε μία από τις οποίες θα απέχει κατά  $J_2$  Hz από τη διπλανή της, και τα ύψη τους θα έχουν περίπου την αναλογία που ορίζεται από την τρίτη γραμμή του τριγώνου Pascal (1:2:1), διαμορφώνοντας ένα triplet (t). Η Εικόνα 4.3(α) περιέχει παραδείγματα από πολλαπλές κορυφές A τάξης για  $n = 0, 1, \dots, 5$ .

Στην περίπτωση που ένα άτομο είναι συζευγμένο με ένα σύνολο  $n$  ατόμων με τιμή σταθεράς σύζευξης  $J_1$  και με ένα άλλο σύνολο  $m$  ατόμων με τιμή  $J_2 \neq J_1$ , τότε η φασματική κορυφή του θα διαχωριστεί σε  $(n+1)(m+1)$  υποκορυφές, διαμορφώνοντας μία πολλαπλή κορυφή B τάξης. Κατ' επέκταση, όσο αυξάνεται το πλήθος των ομάδων ατόμων με τις οποίες ένα άτομο είναι συζευγμένο για διαφορετικές τιμές σταθεράς σύζευξης, αυξάνεται η τάξη της αντίστοιχης πολλαπλής κορυφής. Η δομή των φασματικών γραμμών και οι σταθερές σύζευξης είναι χρήσιμες παράμετροι για τη δομική ανάλυση των χημικών ενώσεων και είναι ανεξάρτητες από το στατικό μαγνητικό πεδίο. Η Εικόνα 4.3(β) περιέχει παραδείγματα από πολλαπλές κορυφές B τάξης, και η εικόνα 4.3(γ) περιέχει παραδείγματα από πολλαπλές κορυφές C τάξης. (Mlynarik, 2016)





Εικόνα 4.3. Σχηματική απεικόνιση των multiplets διαφόρων κατηγοριών. (α) Πολλαπλές κορυφές (Multiplets) Α τάξης, συμπεριλαμβανομένης και της απλής κορυφής (singlet). (β) Πολλαπλές κορυφές Β τάξης (γ) Πολλαπλές κορυφές C τάξης. (Πηγή: Keeler, 2010)

### Χρόνοι αποκατάστασης ή εφησυχασμού

Όπως αναφέρθηκε στην παράγραφο 4.2.2, όταν διακοπεί η ακτινοβολήση ενός δείγματος, το δείγμα έρχεται σε διαδικασία εφησυχασμού, κατά την οποία οι διεγερμένοι πυρήνες επανέρχονται στα προηγούμενα ενεργειακά επίπεδά τους και η μαγνήτιση  $\vec{M}$  του δείγματος επανέρχεται στη θέση ισορροπίας της. Η διαδικασία αποκατάστασης χαρακτηρίζεται από τους χρόνους χαλάρωσης  $T_1$ ,  $T_2$ .

Ο χρόνος  $T_1$  είναι ο χρόνος διαμήκους επαναφοράς, ή αλλιώς χρόνος χαλάρωσης σπιν-πλέγματος. Αναφέρεται στην αποκατάσταση του πληθυσμού ισορροπίας των καταστάσεων σπιν κατά μήκος του πεδίου  $B_0$ , και επομένως στην αποκατάσταση της μαγνήτισης στον άξονα  $z$ , η οποία πραγματοποιείται κυρίως μέσω αλληλεπιδράσεων των πυρήνων με το περιβάλλον τους. Ο χρόνος  $T_1$  επηρεάζει το ρυθμό εφαρμογής των παλμών ραδιοσυχνότητας σε ένα πείραμα NMR. Αυτό συμβαίνει γιατί εξ' ορισμού ο  $T_1$  καθορίζει το βαθμό στον οποίο ένα δείγμα μπορεί να κορεστεί κατά τη διάρκεια της απόκτησης δεδομένων, επομένως όσο μεγαλύτερος είναι ο χρόνος  $T_1$ , τόσο μεγαλύτερη καθυστέρηση χαλάρωσης απαιτείται μεταξύ των παλμών.

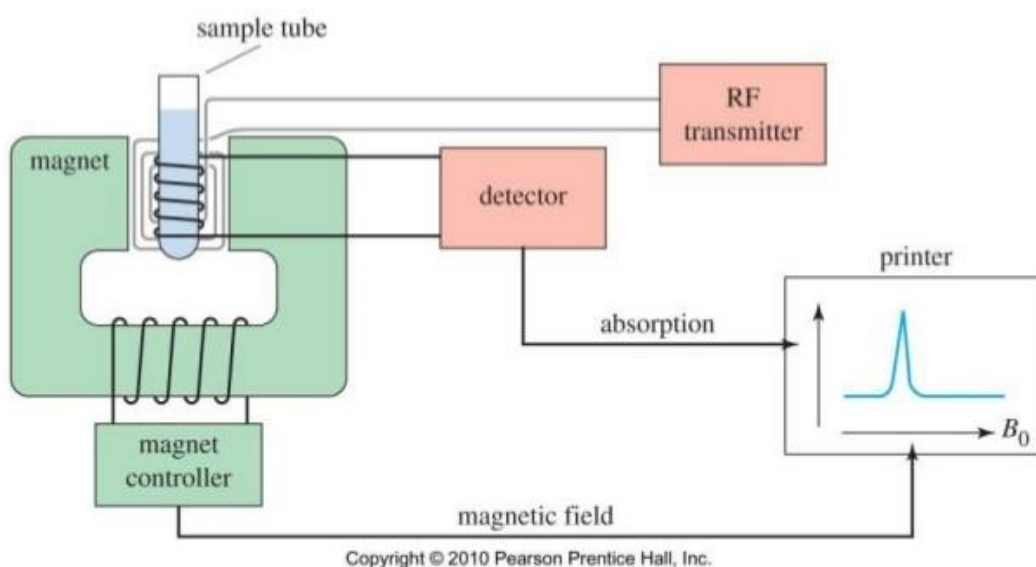
Ο χρόνος  $T_2$  είναι ο χρόνος εγκάρσιας επαναφοράς, ή αλλιώς χρόνος χαλάρωσης σπιν-σπιν. Αναφέρεται στην απώλεια μαγνήτισης στο επίπεδο  $xy$ . Ο χρόνος  $T_2$  καθορίζει το εύρος των συντονισμών που εμφανίζονται στο φάσμα NMR. Το πραγματικό μέρος του φάσματος ( $s_x(\omega)$ ) παίρνει θετικές τιμές. Έχει αποδειχθεί ότι το πλάτος της φασματικής κορυφής στο

μισό της ύψος ισούται με  $1/\pi T_2$ , επομένως η μείωση του χρόνου  $T_2$  σημαίνει αύξηση του πλάτους της φασματικής κορυφής. Με δεδομένο ότι η περιοχή κάτω από την φασματική κορυφή είναι σταθερή, η μείωση του χρόνου  $T_2$  συνεπάγεται και μείωση του ύψους της κορυφής. Συνήθως ισχύει  $T_2 < T_1$ , γιατί η επαναφορά της συνιστώσας  $\vec{M}_z$  προκαλεί αυτόματα μείωση της  $\vec{M}_{xy}$ . Οι τιμές των χρόνων  $T_1$ ,  $T_2$  συνήθως κυμαίνονται στο διάστημα 100-1000 ms (Mlynarik, 2016; Keeler, 2010).

#### 4.2.4. Φασματογράφος Πυρηνικού Μαγνητικού Συντονισμού

Ο φασματογράφος Πυρηνικού Μαγνητικού Συντονισμού (NMR) είναι ένα πολύπλοκο σύστημα η λειτουργία του οποίου στηρίζεται στην αρχή του Πυρηνικού Μαγνητικού Συντονισμού. Ο φασματογράφος είναι σχεδιασμένος ώστε να δέχεται ένα δείγμα σε σωλήνα, να δημιουργεί ένα σταθερό μαγνητικό πεδίο για να φέρει το δείγμα σε κατάσταση ισορροπίας, να παράγει παλμούς ραδιοσυχνότητας για τη διέγερση του δείγματος, καθώς και να καταγράφει το σήμα FID που εκπέμπεται μετά τη διακοπή των παλμών.

Συνοπτικά, τα βασικά μέρη ενός φασματογράφου NMR είναι ένας μαγνήτης που παράγει ένα ομογενές, σταθερό μαγνητικό πεδίο (πεδίο  $B_0$ ), ένας αισθητήρας που μετακινεί τα πηνία που διεγείρουν και ανιχνεύουν το σήμα κοντά στο δείγμα, ένας πομπός ραδιοσυχνότητας με τη δυνατότητα να παράγει παλμούς (πεδίο  $B_1$ ), κι ένας δέκτης υψηλής ευαισθησίας που ενισχύει τα σήματα NMR. Σημαντικά τμήματα του φασματογράφου είναι επίσης ο Αναλογικός/Ψηφιακός μετατροπέας που μετατρέπει το σήμα NMR από αναλογική σε ψηφιακή μορφή, ο προγραμματιστής παλμών που παράγει χρονικούς παλμούς και καθυστερήσεις, καθώς και ο υπολογιστής που ελέγχει τη διαδικασία κι επεξεργάζεται τα δεδομένα (Keeler, 2010). Τα τμήματα του φασματογράφου NMR αναλύονται στη συνέχεια της παραγράφου. Στην εικόνα 4.4 απεικονίζεται ένα διάγραμμα του φασματογράφου Πυρηνικού Μαγνητικού Συντονισμού.



Εικόνα 4.4. Σχεδιάγραμμα του φασματογράφου Πυρηνικού Μαγνητικού Συντονισμού. (Πηγή: Wade, 2010)

## Μαγνήτης

Οι σύγχρονοι φασματογράφοι NMR χρησιμοποιούν υπεραγωγίμους μαγνήτες για να παράξουν το μαγνητικό πεδίο  $B_0$ . Κεντρικό στοιχείο του μαγνήτη είναι το πηνίο, από το οποίο διέρχεται ηλεκτρικό ρεύμα παράγοντας το σταθερό μαγνητικό πεδίο. Το πηνίο αποτελείται από υπεραγωγίμο υλικό, συνήθως από κράμα νιόβιου (Nb) και κασσίτερου (Sn), που σημαίνει ότι η αντίσταση του πηνίου τείνει στο μηδέν σε χαμηλές θερμοκρασίες (<6 K). Επομένως, εφ'όσον το ρεύμα αρχίσει να διαρρέει το πηνίο, συνεχίζει επ' άπειρον, παράγοντας το μαγνητικό πεδίο χωρίς περαιτέρω κατανάλωση ηλεκτρικής ενέργειας.

Για να διατηρηθεί σε υπεραγωγίμη κατάσταση, το πηνίο περιβάλλεται από ένα σύστημα ψύξης. Συγκεκριμένα, το πηνίο περιβάλλεται από φιάλες υγρού ηλίου με σταθερή θερμοκρασία στους ~6K (-269 °C). Οι φιάλες ηλίου συνήθως περικλείονται από φιάλες υγρού αζώτου με θερμοκρασία στους 77K (-196 °C), οι οποίες μειώνουν την ποσότητα του ηλίου που εξατμίζεται στο περιβάλλον. Επιπλέον, το σύστημα αζώτου και ηλίου περιβάλλεται από μία φιάλη κενού για περαιτέρω μείωση της διαρροής θερμότητας στο περιβάλλον. Επομένως, το κόστος διατήρησης του μαγνητικού πεδίου είναι κυρίως το κόστος του αρκετά ακριβού υγρού ηλίου και του φθηνότερου υγρού αζώτου.

Ο σωληνίσκος δείγματος (bore tube) του μαγνήτη είναι ένας κάθετος σωλήνας στον οποίο τοποθετείται το δείγμα, ο οποίος διέρχεται από το μαγνήτη με κατεύθυνση ίδια με το μαγνητικό πεδίο. Πρόκειται για ένα εξάρτημα που είναι προσβάσιμο στο εξωτερικό περιβάλλον του φασματογράφου και βρίσκεται σε θερμοκρασία δωματίου.

Ο φασματογράφος συμπεριλαμβάνει ένα σύνολο από πηνία ομογενοποίησης (shim coils) που περιβάλλουν το δείγμα και αντισταθμίζουν τις ανομοιογένειες του μαγνητικού πεδίου  $B_0$ . Το μαγνητικό πεδίο  $B_0$  πρέπει να είναι όσο το δυνατόν πιο ομοιογενές γιατί τυχόν ανομοιογένειες επηρεάζουν αρνητικά την ανάλυση του φάσματος. Παρ' όλα αυτά, κανένας υπεραγωγίμος μαγνήτης δεν μπορεί να παράξει ένα απόλυτα ομογενές μαγνητικό πεδίο. Κάθε ένα από τα πηνία ομογενοποίησης παράγει ένα μαγνητικό πεδίο μικροσκοπικής έντασης με συγκεκριμένο χωρικό προφίλ, η συνισταμένη των οποίων ακυρώνει τις ανομοιογένειες του κύριου μαγνητικού πεδίου. Το αποτέλεσμα αυτό επιτυγχάνεται με τη ρύθμιση του ρεύματος που διέρχεται από κάθε ένα από αυτά μέχρι το μαγνητικό πεδίο να έχει την απαιτούμενη ομοιογένεια (Keeler, 2010)

## Αισθητήρας

Ο αισθητήρας (probe) του φασματογράφου NMR είναι ένας κυλινδρικός μεταλλικός σωλήνας που εισάγεται στο σωληνίσκο δείγματος του μαγνήτη και δίνει τη δυνατότητα τα πηνία που χρησιμοποιούνται για τη διέγερση και ανίχνευση του σήματος να τοποθετηθούν κοντά στο δείγμα.

Βασικό τμήμα του αισθητήρα είναι ένα μικρό πηνίο, ο ρόλος του οποίου είναι η διέγερση και η ανίχνευση της μαγνήτισης. Για να είναι βέλτιστη η ευαισθησία του, πρέπει να είναι όσο το δυνατόν πιο κοντά στο δείγμα, αλλά και το δείγμα να μπορεί να κατέβει από την κορυφή του μαγνήτη μέσα στο πηνίο. Μέρος του αισθητήρα είναι επίσης ένας πυκνωτής, που μαζί με το πηνίο αποτελούν ένα συντονισμένο κύκλωμα. Η αυτεπαγωγή του πηνίου και η χωρητικότητα του πυκνωτή ρυθμίζονται ώστε το κύκλωμα που σχηματίζουν να είναι συντονισμένο στη συχνότητα Larmor. Επίσης, στοιχείο του αισθητήρα είναι ένας δεύτερος πυκνωτής, του οποίου η χωρητικότητα προσαρμόζεται με στόχο τη βελτιστοποίηση της μεταφοράς ενέργειας στον πομπό και δέκτη. (Keeler, 2010)

## Πομπός

Ο πομπός (transmitter) είναι το τμήμα του φασματογράφου που παράγει τους παλμούς ραδιοσυχνοτήτων (RF). Βασικό στοιχείο του είναι η πηγή των παλμών ραδιοσυχνοτήτων (πηγή RF), η οποία πρέπει να έχει τη δυνατότητα να παράγει σήμα ραδιοσυχνοτήτων με σταθερή συχνότητα και να μπορεί να ρυθμιστεί επακριβώς. Συνήθως ως πηγή ραδιοσυχνοτήτων χρησιμοποιείται ένας συνθέτης συχνοτήτων (frequency synthesizer), αφού έχει όλες τις επιθυμητές ιδιότητες, μπορεί να ελεγχθεί από έναν υπολογιστή, και έχει τη δυνατότητα μετακίνησης φάσης της εξόδου του.

Στοιχείο του πομπού είναι επίσης μία πύλη (gate), που ακολουθεί την πηγή RF και φράσσει την έξοδο της δημιουργώντας τους παλμούς RF. Η λειτουργία της πύλης, και κατ' επέκταση τα χαρακτηριστικά του παλμού, ελέγχονται από ηλεκτρονικό υπολογιστή.

Η πηγή RF και η πύλη ακολουθούνται από έναν εξασθενητή (attenuator), ο ρόλος του οποίου είναι η μείωση του σήματος RF και κατ' επέκταση ο έλεγχος του πεδίου  $B_1$ . Όταν η ισχύς RF εφαρμόζεται στο συντονισμένο κύκλωμα δημιουργούνται υψηλές τάσεις, οι οποίες εμπεριέχουν τον κίνδυνο δημιουργίας ηλεκτρικής έκρηξης, αφ' ενός καταστρέφοντας το πηνίο και τον πυκνωτή, αφ' ετέρου δημιουργώντας μη προβλέψιμο πεδίο  $B_1$ . Ο συνήθης τρόπος ελέγχου του πεδίου  $B_1$  είναι με τον εξασθενητή, ο οποίος μειώνει το σήμα καθώς αυτό διέρχεται. Συνήθως ο κατασκευαστής του φασματογράφου δηλώνει το επίπεδο ισχύος που είναι ασφαλές για ένα συγκεκριμένο αισθητήρα.

Μέρος του πομπού είναι ακόμη ένας ενισχυτής (amplifier). Με δεδομένο ότι η πηγή ραδιοσυχνοτήτων έχει συνήθως χαμηλή ισχύ (κάποια mW), είναι απαραίτητη η ενίσχυση του σήματός της για να παραχθεί ένα αξιοποιήσιμο πεδίο  $B_1$ . Ο ενισχυτής του πομπού ενισχύει το μικρό αυτό σήμα σε ισχύ των 100 W ή μεγαλύτερη. Βεβαίως, το μέγεθος της ισχύος που μπορεί να εφαρμοστεί περιορίζεται από τις υψηλές τάσεις που δημιουργούνται στον αισθητήρα. (Keeler, 2010)

## Δέκτης

Ο δέκτης (receiver) του φασματογράφου NMR είναι σχεδιασμένος ώστε να εισάγει τον ελάχιστο δυνατό θόρυβο. Μέρος του δέκτη είναι ένας προενισχυτής (pre-amplifier) που ενισχύει το αδύναμο σήμα που προκύπτει από τον αισθητήρα (της τάξης των  $\mu V$ ) πριν σταλεί στην κονσόλα του φασματογράφου και τοποθετείται όσο το δυνατόν πιο κοντά στον αισθητήρα. Σημαντικό στοιχείο του δέκτη είναι επίσης ο διπλέκτης (diplexer), του οποίου ο ρόλος είναι να προστατεύει το σύστημα διαχωρίζοντας τον πομπό και το δέκτη. Το πηνίο στον αισθητήρα χρησιμοποιείται αφ' ενός για τη διέγερση των πυρήνων, ένα στάδιο στο οποίο εφαρμόζεται σήμα RF εκατοντάδων Watts, αφ' ετέρου για την ανίχνευση του σήματος, ένα στάδιο στο οποίο ανιχνεύεται ένα σήμα λίγων  $\mu V$ . Ο παλμός υψηλής ισχύος είναι απαραίτητο να μην καταλήξει στον ευαίσθητο δέκτη ώστε να μην τον καταστρέψει. Επομένως, όταν εκπέμπεται ο παλμός RF υψηλής ισχύος, ο δέκτης αποσυνδέεται ή γειώνεται με τη βοήθεια του διπλέκτη. Αντίστοιχα, όταν ο παλμός RF διακόπτεται, ο δέκτης συνδέεται με τον αισθητήρα και ο πομπός αποσυνδέεται. (Keeler, 2010)

## Αναλογικός-Ψηφιακός Μετατροπέας

Μέρος του φασματογράφου NMR είναι ένας μετατροπέας αναλογικού σήματος σε ψηφιακό (analogue to digital converter, ADC) που χρησιμοποιείται για να μετατρέψει το σήμα NMR από τάση σε δυαδικό αριθμό. Ο μετατροπέας κάνει δειγματοληψία του σήματος σε τακτικά διαστήματα, με αποτέλεσμα το σήμα στο πεδίο χρόνου να αναπαριστάται ως ένα

σύνολο δεδομένων. Με δεδομένο ότι οι τιμές εξόδου του μετατροπέα είναι πεπερασμένες, για μία δεδομένη τιμή εισόδου ο μετατροπέας δίνει ως έξοδο την τιμή που είναι πιο κοντά σε αυτή. Επομένως, το σήμα εξόδου του αποτελεί προσέγγιση της πραγματικής κυματομορφής. Η μέγιστη τιμή εξόδου του εξαρτάται από την ανάλυσή του, δηλαδή τον αριθμό των δυαδικών ψηφίων (bits) που χρησιμοποιεί. Οι σύγχρονοι φασματογράφοι συνήθως συμπεριλαμβάνουν μετατροπέα με ανάλυση 16-32 ψηφίων.

Η βασικότερη συνέπεια της διαδικασίας προσέγγισης του μετατροπέα είναι η παραγωγή ενός συνόλου μικρών κορυφών περιστροφής, που ονομάζονται *digitization sidebands*, γύρω από τις βάσεις των κορυφών στο φάσμα, τα οποία συνήθως δεν αποτελούν πρόβλημα γιατί επικαλύπτονται από θόρυβο. Παρ' όλα αυτά, αν το φάσμα εμπεριέχει μία κορυφή πολύ υψηλής έντασης, τα *sidebands* που της αντιστοιχούν μπορούν να επικαλύψουν μία κοντινή κορυφή χαμηλότερης έντασης. Η μείωση των *digitization sidebands* επιτυγχάνεται με τη βελτίωση της ανάλυσης του αναλογικού-ψηφιακού μετατροπέα. (Keeler, 2010)

### **Προγραμματιστής παλμών**

Ο προγραμματιστής παλμών (*pulse programmer*) είναι ένα αρκετά προηγμένο τμήμα υπολογιστικού υλικού που ελέγχει τις λειτουργίες του φασματογράφου. Παράγει επακριβώς χρονολογημένα γεγονότα, συχνά σε γρήγορη ακολουθία. Τυπικά, αφού το πρόγραμμα παλμών καθοριστεί στον κυρίως υπολογιστή, μετά την εκκίνηση του πειράματος, οι οδηγίες φορτώνονται στον προγραμματιστή παλμών και εκτελούνται εκεί. Επίσης, ο προγραμματιστής παλμών διαχειρίζεται τη συλλογή δεδομένων χωριστά από τον κυρίως υπολογιστή. (Keeler, 2010)

### **4.2.5. Πειραματική διαδικασία**

Η διαδικασία διεξαγωγής ενός πειράματος NMR μπορεί να διακριθεί στο στάδιο της προετοιμασίας δείγματος, το κυρίως πείραμα και το στάδιο συλλογής δεδομένων.

#### **Προετοιμασία δείγματος**

Η φασματοσκοπία NMR έχει τη δυνατότητα να αποσπάσει πληροφορία από μία ευρεία γκάμα δειγμάτων. Τα δείγματα χαρακτηρίζονται ως *ex vivo* αν πρόκειται για απομονωμένα όργανα ή άθικτα δείγματα ιστών που μελετώνται προσωρινά εκτός του φυσικού περιβάλλοντός τους και ως *in vitro* αν προέρχονται από ομογενή αποσπάσματα ιστών που μελετώνται σε συνθήκες εργαστηρίου. Συνήθη *ex vivo* βιολογικά δείγματα που μελετώνται με τη φασματοσκοπία NMR είναι ο ορός αίματος, τα ούρα, το εγκεφαλονωτιαίο υγρό και το αμνιακό υγρό. Τα δείγματα αυτά συνήθως βρίσκονται ήδη σε μορφή διαλύματος και μπορούν να μελετηθούν χωρίς ιδιαίτερη προετοιμασία, με εξαίρεση τις ενέργειες που εκτελούνται για να αποφευχθούν τυχόν αρνητικές παρενέργειες ανάλογα με τα χαρακτηριστικά κάθε κατηγορίας δείγματος. Για παράδειγμα, το αίμα μελετάται είτε ως ορός είτε ως πλάσμα μετά από αφαίρεση των κυττάρων που περιέχουν αιμοσφαιρίνη. Αυτό συμβαίνει γιατί η αιμοσφαιρίνη έχει παραμαγνητικές ιδιότητες, με αποτέλεσμα να μπορεί να διαταράξει το ομογενές σταθερό πεδίο  $B_0$  του φασματογράφου, οδηγώντας σε μη επιθυμητή διεύρυνση των φασματικών κορυφών. Επίσης, τα μεταλλικά ιόντα που εντοπίζονται στα ούρα μπορούν να δεσμεύσουν μεταβολίτες, διευρύνοντας τις φασματικές κορυφές τους, φαινόμενο που αποφεύγεται χρησιμοποιώντας κατάλληλες ουσίες που δεσμεύουν τα ιόντα αυτά (Nicholson, Buckingham & Sadler, 1983; Bothwell & Griffin, 2011)



Στα περισσότερα πειράματα NMR προστίθεται στο δείγμα υπό μελέτη μία χημική ένωση αναφοράς (chemical shift reference) που συμβάλλει στον ακριβέστερο ορισμό των χημικών μετατοπίσεων. Στα πειράματα  $^1\text{H}$  NMR συνήθως χρησιμοποιούνται οι ουσίες TMS (Tetramethylsilane), TSP (Trimethylsilylpropanoic acid), DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid), Sodium malaete ως ενώσεις αναφοράς (Harris et al., 2002; Bothwell & Griffin, 2011).

Αφού ο ιστός υπό μελέτη προετοιμαστεί, συνήθως ως διάλυμα σε έναν γυάλινο σωλήνα, τοποθετείται στο σωληνίσκο δείγματος του φασματογράφου και εισέρχεται στο ισχυρό μαγνητικό πεδίο  $B_0$  που δημιουργείται από το μεγάλο πηνίο του μαγνήτη. Στη συνέχεια, το δείγμα τοποθετείται στο μικρότερο πηνίο του αισθητήρα του φασματογράφου για να ξεκινήσει το πείραμα NMR.

### **Πείραμα**

Μετά την προετοιμασία, το δείγμα τοποθετείται στον ηλεκτρομαγνήτη, του οποίου το πηνίο παράγει το ισχυρό στατικό μαγνητικό πεδίο  $B_0$ , και έρχεται σε κατάσταση ισορροπίας. Μέσα στον μαγνήτη, το δείγμα τοποθετείται σε ένα δεύτερο, μικρότερο ηλεκτρομαγνητικό πηνίο, γνωστό ως πηνίο αισθητήρα (probe coil). Το δείγμα ακτινοβολείται με παλμούς ραδιοσυχνότητας (πεδίο  $B_1$ ), οι οποίοι κάποιες φορές συγκεντρώνονται σε ομάδες που ονομάζονται ακολουθίες παλμών. Η εφαρμογή του πεδίου  $B_1$  έχει ως αποτέλεσμα οι πυρήνες του δείγματος να διεγείρονται, δηλαδή να μεταφέρονται σε άλλο ενεργειακό επίπεδο, και η μαγνήτιση του δείγματος να μεταφέρεται από τον άξονα  $z$  στο επίπεδο  $xy$ . Όσο μεγαλύτερη η ισχύς των κυμάτων ραδιοσυχνότητας, τόσο περισσότεροι πυρήνες αλλάζουν ενεργειακό επίπεδο. (Bothwell & Griffin, 2011; Keeler, 2010)

Στη συνέχεια, οι παλμοί ραδιοσυχνότητας (πεδίο  $B_1$ ) διακόπτονται. Οι πυρήνες μπαίνουν σε διαδικασία επαναφοράς στην κατάσταση ισορροπίας τους. Στο στάδιο αυτό οι πυρήνες εκλύουν ηλεκτρομαγνητική ακτινοβολία που καταγράφεται ως σήμα FID, οι παράμετροι της οποίας περιέχουν πληροφορία για το δείγμα. (Bothwell & Griffin, 2011)

Μόλις διακοπεί η λειτουργία του πομπού τίθεται σε λειτουργία ο δέκτης, ο οποίος, συνήθως για ένα χρονικό διάστημα λίγων δευτερολέπτων, συλλέγει την απόσβεση της εγκάρσιας μαγνήτισης. Το πείραμα επαναλαμβάνεται μερικές φορές για να ενισχυθεί το σήμα, αφήνοντας κάποιο χρονικό διάστημα για την επαναδιέγερση του δείγματος. (Keeler, 2010)

### **Συλλογή δεδομένων**

Το σήμα που εκλύεται από τους πυρήνες συλλέγεται στο πεδίο του χρόνου ως σήμα ελεύθερης επαγωγικής απόσβεσης (Free Induction Decay, FID). Το FID απλοποιείται μαθηματικά, μετατρέπεται στο πεδίο συχνότητας με μετασχηματισμό Fourier και παρουσιάζεται με τη μορφή μονοδιάστατου (1D) ή διδιάστατου (2D) φάσματος που περιέχει ένα σύνολο συχνοτήτων. Το FID και το αντίστοιχο φάσμα περιέχουν παραμέτρους που εμπεριέχουν πληροφορίες για το δείγμα, οι οποίες είναι η ένταση, οι χημικές μετατοπίσεις, οι χρόνοι χαλάρωσης και η φάση.

Όσον αφορά την ένταση του σήματος, ζητούμενο είναι να αυξηθεί ο λόγος σήματος προς θόρυβο (signal-to-noise ratio, SNR). Ένας τρόπος βελτίωσης του SNR είναι η αρκετά ακριβή μέθοδος της αύξησης του πεδίου  $B_0$ , καθώς αυτή θα επιφέρει την αύξηση της πληθυσμιακής διαφοράς μεταξύ των επιπέδων ενέργειας. Το SNR μπορεί ακόμη να βελτιωθεί με την ανάμειξη πολλών σημάτων FID (signal averaging). Ενώ η ένταση του σήματος

αυξάνεται ως συνάρτηση του αριθμού ανεξάρτητων δειγμάτων (έστω  $n$ ), ο θόρυβος αυξάνεται ως συνάρτηση του  $\sqrt{n}$ , με αποτέλεσμα το SNR να αυξάνεται ως συνάρτηση του  $\sqrt{n}$ . Επίσης, το SNR μπορεί να βελτιωθεί με μείωση της θερμοκρασίας του δείγματος ή του εξοπλισμού. Στην πρώτη περίπτωση, η ψύξη του δείγματος συνεπάγεται τη μείωση της θερμικής κίνησης των μορίων του δείγματος, με αποτέλεσμα περισσότεροι πυρήνες να βρίσκονται στα χαμηλότερα ενεργειακά επίπεδα και να γίνεται μεγαλύτερη η πληθυσμιακή διαφορά μεταξύ των ενεργειακών επιπέδων. Παρ'όλα αυτά, υπάρχει ο περιορισμός της θερμοκρασίας στην οποία παγώνει ο διαλύτης. Στη δεύτερη περίπτωση γίνεται χρήση κρυογονικών probes που αποσκοπούν στη μείωση της θερμοκρασίας του εξοπλισμού, με αποτέλεσμα τη μείωση του ηλεκτρονικού και θερμικού θορύβου, που θεωρητικά μπορεί να επιφέρει μέχρι και τετραπλάσια αύξηση του SNR.

Ο δεύτερος παράγοντας του φάσματος είναι το σύνολο των χημικών μετατοπίσεων. Η χημική μετατόπιση είναι πιο χρήσιμη ως πληροφορία όταν οι φασματικές κορυφές που αντιστοιχούν σε διαφορετικές χημικές ουσίες είναι καλά διαχωρισμένες και καλά διακριτές. Ο διαχωρισμός των φασματικών κορυφών επηρεάζεται από την ομοιογένεια του πεδίου  $B_0$  και μπορεί να βελτιωθεί με διάφορους τρόπους, ένας από οποίους είναι η ομογενοποίηση του πεδίου στο φασματογράφο. Κανένας ηλεκτρομαγνήτης, ανεξάρτητα από το κόστος, δεν μπορεί να δημιουργήσει ένα απόλυτα ομοιογενές πεδίο, γιατί το δείγμα από μόνο του διαταράσσει το πεδίο  $B_0$ . Το πρόβλημα αυτό επιλύεται με τη χρήση πηνίων ομογενοποίησης που αποτελούν μέρος του μαγνήτη του φασματογράφου (παράγραφος 4.2.5). Αν και στο παρελθόν η διαδικασία αυτή ήταν χρονοβόρα, πλέον γίνεται αυτοματοποιημένα στις σύγχρονες μηχανές. Επίσης, παλαιότεροι μαγνήτες απαιτούσαν περιστροφή του δείγματος, συνήθως σε συχνότητα 20 Hz, για τη μείωση οποιασδήποτε υπολειπόμενης ανομοιογένειας του πεδίου στο οριζόντιο επίπεδο ( $xy$ ). Στους σύγχρονους φασματογράφους η διαδικασία της περιστροφής του δείγματος δεν είναι τόσο απαραίτητη. Ένας άλλος τρόπος βελτίωσης του διαχωρισμού των φασματικών γραμμών είναι η μείωση της έντασης πεδίου  $B_0$ , γιατί ένα ομοιογενές πεδίο είναι πιο εύκολο να επιτευχθεί με μικρή ένταση.

Η βελτίωση της φασματικής ανάλυσης μπορεί να επιτευχθεί με τη χρήση διαλυτών στα δείγματα υπό εξέταση. Σε ένα ετερογενές δείγμα, διαφορετικές περιοχές δειγμάτων θα βιώσουν διαφορετικά μαγνητικά περιβάλλοντα. Οι τοπικές στρεβλώσεις στο πεδίο  $B_0$  έχουν ως αποτέλεσμα την εμφάνιση μη ευδιάκριτων φασματικών κορυφών. Αντίθετα, στο εσωτερικό ενός διαλύματος, η ελεύθερη περιστροφή των διαλυμένων ουσιών σημαίνει ότι τα χημικά περιβάλλοντα αυτών των ουσιών είναι πιο όμοια. Επομένως, η χρήση διαλυμάτων μειώνει τη διεύρυνση των φασματικών γραμμών και συμβάλλει στην εμφάνιση μη επικαλυπτόμενων, πιο ξεκάθαρων συντονισμών στο φάσμα.

Τέλος, ο διαχωρισμός των φασματικών γραμμών μπορεί να βελτιωθεί με τη μέθοδο της μαγνητικής περιστροφής στη «μαγική» γωνία των  $54.7^\circ$ . Στα δείγματα στερεών ιστών οι ενώσεις δεν μπορούν να «ανατραπούν» τόσο εύκολα όσο στα δείγματα σε διάλυμα, με αποτέλεσμα να παρατηρείται διεύρυνση των γραμμών απορροφήσεων φάσματος. Οι χαμηλοί ρυθμοί χαλάρωσης σε συνδυασμό με τα αποτελέσματα της μαγνητικής ευαισθησίας και της ανισοτροπίας χημικής μετατόπισης τείνουν να παράγουν έναν ευρύ άμορφο συντονισμό για οποιοδήποτε στερεό εξετάζεται με NMR. Έχει παρατηρηθεί ότι πολλές από τις επιπτώσεις διεύρυνσης των γραμμών στα στερεά εξαρτώνται από τη γωνία στην οποία τοποθετείται το δείγμα σε σχέση με το πεδίο  $B_0$ , και έχει αποδειχθεί ότι πως περιστρέφοντας ένα στερεό

δείγμα στη λεγόμενη «μαγική γωνία» των  $54.7^\circ$  έχει ως αποτέλεσμα την παραγωγή φασμάτων με σημαντικά μειωμένα εύρη γραμμών. Ο συνδυασμός των φυσικών επιδράσεων που προκαλούν το παραπάνω φαινόμενο είναι ιδιαίτερα περίπλοκος, γι' αυτό χρησιμοποιείται ο όρος «μαγική» γωνία.

Στη φασματοσκοπία NMR μπορεί να προσδιοριστεί η φάση του σήματος, πληροφορία σημαντική που αξιοποιείται στον μετασχηματισμό Fourier. Ο μετασχηματισμός Fourier, ο οποίος μετατρέπει το σήμα FID από το πεδίο χρόνου στο πεδίο συχνότητας, επιτάχυνε σημαντικά τη φασματοσκοπία αφού επέτρεπε την απόκτηση ενός εύρους συχνοτήτων σε ένα μόνο παλμό, αντί να απαιτεί μία χρονοβόρα σάρωση σε ένα φάσμα συχνοτήτων. Επίσης, η πληροφορία της φάσης αξιοποιείται στη χρήση της ακολουθίας παλμών και της πολυδιάστατης φασματοσκοπίας.

Τέλος, σημαντικές παράμετροι του συλλεγόμενου σήματος είναι οι χρόνοι χαλάρωσης  $T_1$ ,  $T_2$  που μπορούν να δώσουν πληροφορίες σχετικά με το περιβάλλον των μορίων. Ο χρόνος χαλάρωσης  $T_1$ , αναφέρεται στην αποκατάσταση της μαγνήτισης κατά μήκος του πεδίου  $B_0$ . Ο  $T_1$  ονομάζεται αλλιώς χρόνος χαλάρωσης σπιν-πλέγματος, γιατί η αποκατάσταση στην οποία αναφέρεται πραγματοποιείται κυρίως μέσω της αλληλεπίδρασης των πυρήνων με το περιβάλλον τους. Ο χρόνος  $T_1$  επηρεάζει το ρυθμό εφαρμογής των παλμών ραδιοσυχνοτήτων ( $B_1$ ). Εξ' ορισμού ο  $T_1$  καθορίζει το βαθμό στον οποίο μπορεί ένα δείγμα να κορεστεί κατά τη διάρκεια της απόκτησης δεδομένων, οπότε εάν ο  $T_1$  είναι μεγάλος, απαιτείται μεγάλη καθυστέρηση χαλάρωσης μεταξύ των παλμών. Ο χρόνος  $T_2$ , που ονομάζεται και χρόνος χαλάρωσης σπιν-σπιν, αναφέρεται στην απώλεια μαγνήτισης στο επίπεδο  $xy$ . Ο  $T_2$  καθορίζει το εύρος των συντονισμών που εμφανίζονται στο φάσμα NMR και συνήθως ισχύει  $T_2 < T_1$ . Οι τιμές των  $T_1$ ,  $T_2$  συνήθως κυμαίνονται στο διάστημα 100-1000 ms. (Bothwell & Griffin, 2011)

#### **4.2.6. Επεξεργασία σήματος Πυρηνικού Μαγνητικού Συντονισμού**

##### **Προσθήκη μηδενικών**

Με τη δειγματοληψία, το αναλογικό σήμα FID μετατρέπεται σε ένα ψηφιακό σύνολο δεδομένων. Από το μετασχηματισμό Fourier του ψηφιακού σήματος FID προκύπτει το φάσμα συχνοτήτων του, που αναπαριστάται επίσης ως ένα σύνολο σημείων δεδομένων. Η προσθήκη μηδενικών είναι η διαδικασία κατά την οποία προστίθεται ένα σύνολο μηδενικών τιμών στο FID σήμα, με στόχο να αυξηθούν τα σημεία δεδομένων του FID και επομένως να αυξηθούν τα σημεία δεδομένων του φάσματος που θα προκύψει από το μετασχηματισμό Fourier. Επειδή ο μετασχηματισμός Fourier συνήθως εφαρμόζεται σε σύνολα δεδομένων το πλήθος των οποίων είναι δύναμη του 2, συνηθίζεται να προστίθενται μηδενικές τιμές στο σήμα FID μέχρι το πλήθος των σημείων δεδομένων να είναι δύναμη του 2.

Η διαδικασία της προσθήκης μηδενικών ορίζει καλύτερα τις φασματικές γραμμές, βελτιώνοντας αισθητικά το φάσμα. Δεν είναι δαπανηρή διαδικασία, με την έννοια ότι είναι μία υπολογιστική χειραγώγηση που δεν συμπεριλαμβάνει την πειραματική συλλογή επιπλέον δεδομένων. Παρ' όλα αυτά, δε βελτιώνει την ανάλυση του φάσματος, καθώς το σήμα που συλλέγεται πειραματικά παραμένει το ίδιο. (Keeler, 2010)

##### **Βελτίωση ευαισθησίας**

Στη φασματοσκοπία NMR υπάρχουν διάφορες πηγές θορύβου, όπως ο θερμικός θόρυβος από το πηνίο που ανιχνεύει το σήμα, καθώς και ο θόρυβος που προέρχεται από τους

ενισχυτές και τα άλλα ηλεκτρονικά στοιχεία του φασματογράφου. Με δεδομένο ότι η φασματοσκοπία NMR δεν είναι ευαίσθητη τεχνική, είναι χρήσιμες οι ενέργειες βελτίωσης του λόγου σήματος προς θόρυβο (signal-to-noise ratio, SNR) στο φάσμα. Η αύξηση του SNR μπορεί να επιτευχθεί αφ' ενός με τη μείωση του χρόνου καταγραφής του σήματος FID, αφ' ετέρου με την εφαρμογή συναρτήσεων βάρους.

Η βελτίωση του SNR με τη μείωση του χρόνου καταγραφής του σήματος FID στηρίζεται στο ότι το FID μειώνεται στο χρόνο αλλά ο θόρυβος παραμένει στο ίδιο επίπεδο, επομένως το χρήσιμο τμήμα του FID εντοπίζεται στην αρχή του. Αν η καταγραφή δεδομένων συνεχιστεί αφού η ένταση του σήματος FID έχει μειωθεί, θα μετρηθεί θόρυβος και όχι σήμα, με αποτέλεσμα το φάσμα που θα προκύψει να έχει χαμηλό SNR. Η μείωση του χρόνου καταγραφής του σήματος θα βελτιώσει το SNR, αφού το σήμα περιέχεται στο αρχικό τμήμα του FID. Βεβαίως, ο χρόνος καταγραφής δεν πρέπει να μειωθεί τόσο ώστε να χαθεί τμήμα του χρήσιμου σήματος, καθώς κάτι τέτοιο θα μειώσει το SNR.

Παρομοίως, η βελτίωση του SNR με την εφαρμογή συναρτήσεων βάρους στηρίζεται στο ότι το αρχικό τμήμα του σήματος είναι ισχυρότερο και πιο σημαντικό. Το χαρακτηριστικό αυτό αξιοποιείται πολλαπλασιάζοντας το FID με μία φθίνουσα συνάρτηση βάρους με μέγιστη τιμή το 1, η οποία θα κρατήσει μόνο τα πρώτα χρήσιμα τμήματα του σήματος. Συνήθως η συνάρτηση βάρους που επιλέγεται είναι η φθίνουσα εκθετική συνάρτηση  $W(t) = e^{-R_{LB}t}$ , όπου  $R_{LB}$  είναι μία σταθερά. Η εφαρμογή της συνάρτησης  $W(t) = e^{-R_{LB}t}$  έχει ως αποτέλεσμα το σήμα FID να φθίνει ταχύτερα, και κατ' επέκταση οι φασματικές κορυφές να διευρύνονται. Με δεδομένο ότι η περιοχή ολοκλήρωσης κάτω από μία φασματική κορυφή παραμένει σταθερή, η διεύρυνση μίας φασματικής κορυφής συνεπάγεται τη μείωση ύψους της, και κατ' επέκταση τη μείωση του SNR. Επομένως, υπάρχει ένας συμβιβασμός όσον αφορά την εφαρμογή της φθίνουσας συνάρτησης, αφού όσο πιο γρήγορα φθίνει η  $W(t)$ , αφ' ενός τόσο περισσότερο μειώνεται ο θόρυβος στο τέλος του FID αυξάνοντας το SNR, αφ' ετέρου τόσο περισσότερο αυξάνεται το πλάτος και μειώνεται το ύψος των φασματικών κορυφών, μειώνοντας το SNR. Έχει αποδειχθεί ότι υπάρχει μία βέλτιστη συνάρτηση βάρους, που μπορεί να προσδιοριστεί. (Keeler, 2010)

### **Βελτίωση ανάλυσης**

Όπως προαναφέρθηκε, όσο πιο γρήγορα το σήμα FID φθίνει, τόσο περισσότερο διευρύνονται οι φασματικές γραμμές, με αποτέλεσμα τη μείωση ανάλυσης του φάσματος. Επομένως, μία φθίνουσα συνάρτηση βάρους που είναι σχεδιασμένη να βελτιώνει το SNR αναπόφευκτα οδηγεί σε διεύρυνση των φασματικών γραμμών και κατ' επέκταση σε μείωση της φασματικής ανάλυσης.

Η ανάλυση του φάσματος μπορεί να βελτιωθεί εφαρμόζοντας μία συνάρτηση βάρους που είναι σχεδιασμένη να αναχαιτίζει τη μείωση του σήματος FID και κατ' επέκταση να μειώνει το εύρος των φασματικών γραμμών. Η συνάρτηση βάρους αυτή είναι αύξουσα, όπως για παράδειγμα η εκθετική συνάρτηση  $W(t) = e^{R_{RE}t}$ ,  $R_{RE} > 0$ . Παρ' όλα αυτά, η εφαρμογή μίας αύξουσας συνάρτησης βάρους αφ' ενός επιτυγχάνει τη βελτίωση ανάλυσης σήματος, αφ' ετέρου ενισχύει το θόρυβο στο τέλος του σήματος FID, με αποτέλεσμα τη μη επιθυμητή μείωση του SNR.

Για την αντιστάθμιση του παραπάνω προβλήματος, εφαρμόζεται μία δεύτερη, φθίνουσα συνάρτηση βάρους που αποκόβει το θόρυβο στο τέλος του FID. Συνήθως, η συνάρτηση αυτή επιλέγεται ώστε να φθίνει με αργό ρυθμό στο μεγαλύτερο μέρος του FID

και με γρήγορο ρυθμό προς το τέλος του, όπως η Γκαουσιανή συνάρτηση  $W(t) = e^{-at^2}$ , όπου  $a$  μία παράμετρος που καθορίζει το ρυθμό μείωσης. (Keeler, 2010)

### Διόρθωση φάσης

Η περίπτωση που στην αρχή του FID ( $t=0$ ) το  $S_x(t)$  έχει μέγιστη τιμή και το  $S_y(t)$  παίρνει μηδενική τιμή θεωρείται ιδανική γιατί στο πεδίο της συχνότητας το πραγματικό μέρος του φάσματος βρίσκεται σε absorption mode και το φανταστικό μέρος του φάσματος σε dispersion mode, γεγονός που αντικατοπτρίζεται σε βέλτιστη φασματική ανάλυση.

Στην πραγματικότητα, ο άξονας στον οποίο εμφανίζεται το σήμα δεν μπορεί να προκαθοριστεί, με αποτέλεσμα συνήθως να υπάρχει ένα άγνωστο σφάλμα φάσης που οφείλεται στη λειτουργία του φασματογράφου. Στο πεδίο συχνότητας, ένα τέτοιο σφάλμα φάσης σημαίνει ότι το πραγματικό μέρος του φάσματος δεν βρίσκεται σε πλήρες absorption mode, γεγονός που αντικατοπτρίζεται σε χαμηλότερη φασματική ανάλυση. Το σφάλμα φάσης ονομάζεται μηδενικής τάξης όταν είναι ανεξάρτητο από τη συχνότητα, δηλαδή όταν φασματικές κορυφές διαφορετικών συχνοτήτων έχουν το ίδιο σφάλμα φάσης.

Συμπεριλαμβάνοντας το σφάλμα φάσης, το σήμα αναπαριστάται μαθηματικά ως  $S(t) = S_0 e^{i\varphi} e^{i\Omega t} e^{-t/T_2}$ . Η διόρθωση φάσης έχει στόχο την επαναφορά του φάσματος σε absorption mode και μαθηματικά είναι είναι ο πολλαπλασιασμός του FID με έναν παράγοντα  $e^{i\varphi_{\text{corr}}}$  ( $e^{i\varphi_{\text{corr}}} S(t) = S_0 e^{i\varphi_{\text{corr}}} e^{i\varphi} e^{i\Omega t} e^{-t/T_2} \xrightarrow{\varphi_{\text{corr}} = -\varphi} e^{i\varphi_{\text{corr}}} S(t) = S_0 e^{i\Omega t} e^{-t/T_2}$ ). Επομένως, αρκεί η εύρεση του σωστού  $\varphi_{\text{corr}}$ . Η διόρθωση φάσης μπορεί να εφαρμοστεί στο φάσμα συχνότητας  $S(\omega)$  όπως στο FID, υπολογίζοντας το  $e^{i\varphi_{\text{corr}}} S(\omega)$ . Στην πράξη, η διόρθωση φάσης πραγματοποιείται κάνοντας μετασχηματισμό Fourier του FID, απεικονίζοντας το πραγματικό μέρος του φάσματος και προσαρμόζοντας τη φάση  $\varphi_{\text{corr}}$  μέχρι το φάσμα να βρεθεί στην επιθυμητή μορφή. Η προσαρμογή αυτή συνήθως επιτυγχάνεται με μία εύκολη τεχνικά διαδικασία, όπως το πάτημα ενός κουμπιού, που πρέπει να γίνεται κάθε φορά που καταγράφεται ένα φάσμα. (Keeler, 2010)

Επιπλέον, υπάρχει το σφάλμα φάσης που οφείλεται στην ανάμιξη των συνιστωσών μαγνήτισης  $M_x$ ,  $M_y$  και εξαρτάται από τη συχνότητα, δηλαδή σε διαφορετικές φασματικές κορυφές αντιστοιχεί διαφορετική τιμή φάσης. Αν το σφάλμα φάσης είναι ανάλογο της συχνότητας ονομάζεται πρώτης τάξης. Σε αυτή την περίπτωση, η διόρθωση φάσης πραγματοποιείται με την προσαρμογή του ρυθμού μεταβολής της φάσης με τη συχνότητα μέχρι το φάσμα να βρίσκεται στην επιθυμητή μορφή, διαδικασία που διευκολύνεται από τα κατάλληλα λογισμικά. Στην πράξη, η διόρθωση φάσης ενός φάσματος απαιτεί μία επανάληψη διορθώσεων μηδενικής και πρώτης τάξης. (Keeler, 2010)

### Διόρθωση γραμμής βάσης

Η αλλοίωση των πρώτων σημείων δεδομένων του σήματος FID μπορεί να προκαλέσει παραμόρφωση της γραμμής βάσης στα μονοδιάστατα φάσματα NMR. Η παραμόρφωση της γραμμής βάσης συνεπάγεται αλλοιώσεις των τιμών έντασης του φάσματος, που μπορούν να προκαλέσουν σφάλματα στην ταυτοποίηση και ποσοτικοποίηση των μεταβολιτών του φάσματος. Τα φάσματα NMR συμπεριλαμβάνουν και κορυφές χαμηλής έντασης που είναι επιρρεπείς στις αλλοιώσεις της γραμμής βάσης αλλά δεν υστερούν σε σημασία, και η εσφαλμένη ποσοτικοποίηση των κορυφών αυτών μπορεί να οδηγήσει σε λάθος συμπεράσματα όσον αφορά τον εντοπισμό πιθανών βιοδεικτών. Επομένως, η διόρθωση της γραμμής βάσης είναι ένα απαραίτητο βήμα επεξεργασίας των φασμάτων NMR. (Xi & Rocke,

2008) Η διόρθωση βασικής γραμμής επιτυγχάνεται με διάφορες μεθόδους, ανάμεσα στις οποίες είναι η χρήση του φίλτρου Whittaker (Whittaker, 1922; Eilers, 2003), και του πολυωνόμου Bernstein (Marco et al., 2010).

#### **4.2.7. Περιστολή δεδομένων φάσματος Πυρηνικού Μαγνητικού Συντονισμού**

Το φάσμα συχνοτήτων αποτελείται από ένα σύνολο χιλιάδων ζευγών συχνότητας και έντασης. Ο μεγάλος όγκος του φάσματος αφ' ενός το καθιστά δύσκολα διαχειρίσιμο, αφ' ετέρου δυσχεραίνει την εξαγωγή χρήσιμης πληροφορίας από αυτό. Για αυτούς τους λόγους είναι συνήθης τακτική το φάσμα να υπόκειται σε περιστολή δεδομένων, κατά κύριο λόγο με τη μέθοδο του binning ή αλλιώς bucketing. Σύμφωνα με τη μέθοδο του binning, το φάσμα χωρίζεται σε διαστήματα, τα λεγόμενα bins, κάθε ένα από τα οποία αντιπροσωπεύεται από μία συχνότητα και μία τιμή έντασης. Ιδανικά, κάθε bin αντιστοιχεί σε μία μοναδική φασματική κορυφή και κάθε φασματική κορυφή αντιστοιχεί σε ένα bin. Αποτέλεσμα αυτής της διαδικασίας είναι το φάσμα να αντιπροσωπεύεται από ένα μικρότερο σύνολο ζευγών συχνότητας και έντασης που είναι πιο εύκολα επεξεργάσιμο.

Το ομοιόμορφο binning είναι η πρώτη, απλούστερη μέθοδος μείωσης δεδομένων, σύμφωνα με την οποία το φάσμα χωρίζεται σε ομοιόμορφα διαστήματα σταθερού εύρους, συνήθως 0.04 ppm (Davis et al., 2007), το οποίο μπορεί να επιλέξει ο χρήστης. Κάθε bin αντιπροσωπεύεται από μία τιμή συχνότητας, για παράδειγμα η μέση συχνότητα του bin, και μία τιμή έντασης, για παράδειγμα η τιμή που προκύπτει από την ολοκλήρωση του φάσματος στα πλαίσια του bin. Αρνητικό χαρακτηριστικό του ομοιόμορφου binning είναι ότι τα bins διαμορφώνονται «τυφλά», χωρίς να υπάρχει απαραίτητα αντιστοιχία ενός bin με μία φασματική κορυφή. Επομένως, ένα bin μπορεί να μην αντιστοιχεί σε καμία φασματική κορυφή ή να αντιστοιχεί σε περισσότερες από μία. Παρά τα μειονεκτήματά της, η μέθοδος του ομοιόμορφου binning είναι ενσωματωμένη σε προγράμματα επεξεργασίας φασμάτων NMR όπως το MestReNova, και έχει χρησιμοποιηθεί σε διάφορες μελέτες μεταβολομικής (Lenz et al., 2003; Nicholls et al., 2003; Jang-Eun et al., 2009; Agnolet et al., 2010).

Τα τελευταία χρόνια έχουν αναπτυχθεί διάφοροι μέθοδοι binning που αποσκοπούν στην πιο έξυπνη επιλογή των ορίων των bins, ώστε να είναι πιο ενδεικτικά των φασματικών κορυφών. Μία από αυτές είναι η μέθοδος Adaptive binning, που έχει ως αντικείμενο την ανίχνευση κορυφών σε ένα φάσμα NMR εισόδου βασιζόμενη στον μετασχηματισμό wavelet. Αν ο αριθμός φασμάτων εισόδου είναι μεγαλύτερος από ένα, επιλέγεται ένα φάσμα αναφοράς με βάση το οποίο θα γίνει το binning. Οι συγγραφείς επέλεξαν ως φάσμα αναφοράς αυτό που προκύπτει από τις μέγιστες τιμές έντασης όλων των φασμάτων εισόδου για όλες τις συχνότητες, ώστε να αποφευχθεί η απώλεια κάποιας κορυφής που υπάρχει σε κάποια μόνο από τα φάσματα εισόδου. Σύμφωνα με τη μέθοδο του adaptive binning, ο μετασχηματισμός wavelet χρησιμοποιείται για αποθρούβωση του φάσματος αναφοράς, για την εύρεση των τοπικών ελαχίστων του, και κατ' επέκταση την εύρεση της αρχής και του τέλους κάθε φασματικής κορυφής του. Στη συνέχεια, κάθε φάσμα μετασχηματίζεται αθροίζοντας τις εντάσεις μεταξύ των τοπικών ελαχίστων. Ο μετασχηματισμός wavelet εξομαλύνει το φάσμα αναφοράς, ώστε οι κορυφές που παρουσιάζουν μικρές μετατοπίσεις στα διάφορα φάσματα εισόδου συνδυάζονται σε μία κορυφή. Τα όρια των bins τοποθετούνται στα τοπικά ελάχιστα εκατέρωθεν της κορυφής. Ο βαθμός εξομάλυνσης εξαρτάται από τον αριθμό επιπέδων του μετασχηματισμού wavelet και μπορεί να προσαρμοστεί ανάλογα με την ανάλυση των

δεδομένων και τις μετατοπίσεις που αναμένονται μεταξύ των δειγμάτων. Επομένως, περισσότερα επίπεδα μετασχηματισμού μπορεί να απαιτούνται για πιο πολύπλοκα υποστρώματα, όπως τα ούρα, αφού αυτά συνήθως παρουσιάζουν μεγαλύτερες χημικές μετατοπίσεις, που αντικατοπτρίζονται σε ευρύτερες κορυφές στο φάσμα αναφοράς. Αρνητικό χαρακτηριστικό της μεθόδου είναι ότι απαιτείται από το χρήστη ο καθορισμός του αριθμού επιπέδων μετασχηματισμού wavelet. (Davis et al., 2007)

Η μέθοδος Gaussian binning έχει ως χαρακτηριστικό τη δυνατότητα να παράγει αλληλεπικαλυπτόμενα bins. Στηρίζεται στη χρήση μίας μάσκας (kernel), η οποία αποδίδει βάρη στα σημεία δεδομένων του κάθε bin. Για κάθε σημείο δεδομένων, η μάσκα πολλαπλασιάζεται με το σήμα και η τιμή που προκύπτει αντικαθιστά την αρχική τιμή του σημείου. Ως μάσκα επιλέχθηκε η γκαουσιανή συνάρτηση πυκνότητας πιθανότητας, που αναθέτει υψηλό βάρος στα σημεία δεδομένων κοντά στο κέντρο του bin, και χαμηλότερο βάρος στα σημεία δεδομένων κοντά στα όρια του bin. Το μέσο της κατανομής τοποθετείται στο κέντρο του bin. Η γκαουσιανή μάσκα έχει δύο προσαρμόσιμες παραμέτρους, αφ' ενός την απόσταση μεταξύ των κέντρων δύο διαδοχικών μασκών, δηλαδή το βήμα, αφ' ετέρου το πλάτος που αναπαριστάται από την τυπική απόκλιση της γκαουσιανής κατανομής. Οι δύο αυτές παράμετροι καθορίζουν το βαθμό επικάλυψης και τον αριθμό των bins που προκύπτουν για ένα δεδομένο φασματικό εύρος. Το σήμα μετασχηματίζεται πολλαπλασιάζοντας κάθε σημείο με τη γκαουσιανή συνάρτηση πυκνότητας πιθανότητας και η μετασχηματισμένη καμπύλη ολοκληρώνεται σε κάθε bin. (Anderson et al., 2008)

Άλλος ένας αλγόριθμος που προτάθηκε ως βελτίωση της συμβατικής διαδικασίας ομοιόμορφου binning είναι το Optimized bucketing. Σύμφωνα με αυτή τη μέθοδο, αν ο αριθμός φασμάτων εισόδου είναι μεγαλύτερος από ένα, επιλέγεται το μέσο φάσμα εισόδου ως φάσμα αναφοράς. Η μέθοδος του Optimized bucketing αποσκοπεί στη δημιουργία bins με μεταβλητό μέγεθος που είναι κοινά για το σύνολο των φασμάτων εισόδου, και προσδιορίζονται με βάση τα τοπικά ελάχιστα του μέσου φάσματος εισόδου. Ο χρήστης καλείται να ορίσει δύο παραμέτρους, το αρχικό πλάτος των bins σε rpm, και τη χαλαρότητα ("slackness"), δηλαδή το περιθώριο μετακίνησης των ορίων των bins στην αναζήτηση των τοπικών ελαχίστων του μέσου φάσματος. Ο αλγόριθμος αρχικά δημιουργεί ομοιόμορφα bins με μέγεθος το αρχικό πλάτος που όρισε ο χρήστης και στη συνέχεια τα τροποποιεί, οδηγώντας τα όριά τους στα τοπικά ελάχιστα του φάσματος, με περιορισμό τη χαλαρότητα που δόθηκε ως παράμετρος. Αρνητικό χαρακτηριστικό της μεθόδου είναι ότι απαιτείται από το χρήστη ο καθορισμός δύο παραμέτρων που έχουν σημαντικό ρόλο στα αποτελέσματα της μεθόδου (Sousa et al., 2013).

Η μέθοδος dynamic adaptive binning έχει ως αντικείμενο τον καθορισμό των ορίων bin με δυναμικό προγραμματισμό, βελτιστοποιώντας μία συνάρτηση που αξιολογεί την ποιότητα των διαμορφωμένων bins. Ο αλγόριθμος αυτός προσδιορίζει τη βέλτιστη διαμόρφωση των bins σε δύο βήματα. Αρχικά καθορίζει τις θέσεις των κορυφών σε κάθε φάσμα, χρησιμοποιώντας το μετασχηματισμό wavelet. Στη συνέχεια, βρίσκει τα βέλτιστα όρια των bins με βάση τη συνάρτηση αξιολόγησης, ακολουθώντας μία στρατηγική δυναμικού προγραμματισμού. Η μέθοδος dynamic adaptive binning απαιτεί τον ορισμό δύο παραμέτρων από το χρήστη, τη μέγιστη απόσταση μεταξύ παρατηρούμενων κορυφών στο ίδιο bin και την ελάχιστη απόσταση μεταξύ μίας παρατηρούμενης κορυφής και ενός ορίου bin. (Anderson et al., 2011)

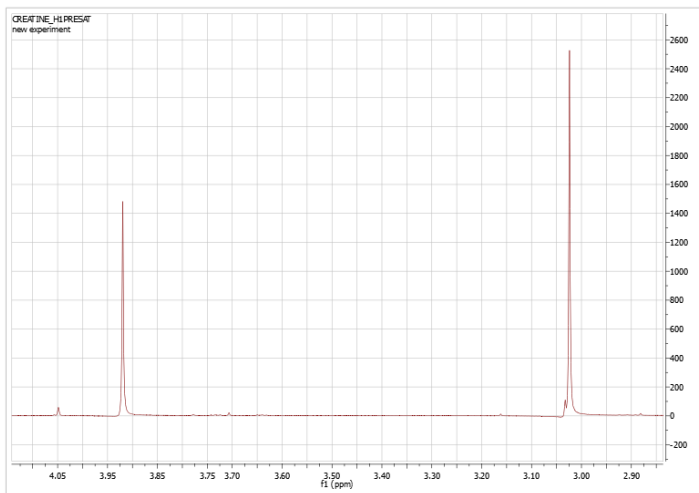
Ακόμη μία μέθοδος μείωσης δεδομένων είναι η ERVA, που συνίσταται από την συνέλιξη του φάσματος με τη δεύτερη παράγωγο της λορεντζιανής συνάρτησης. Η μέθοδος αυτή συμπεριλαμβάνει δύο παραμέτρους, και τα αποτελέσματά της εξαρτώνται αρκετά από το επίπεδο αποθορύβωσης που εφαρμόζεται στο φάσμα. Επίσης, τα bins που δημιουργούνται μπορούν να αφήνουν κάποια τμήματα του φάσματος εκτός, πράγμα που δεν αποτελεί απαραίτητα πλεονέκτημα όσον αφορά την περαιτέρω επεξεργασία του φάσματος. (Jacob et al., 2013)

Ο αλγόριθμος Adaptive Intelligent binning (AI) χρησιμοποιεί bins με μεταβλητά μεγέθη και αποφασίζει αυτόματα πότε θα σταματήσει τη διαδικασία binning των φασμάτων εισόδου. Μία ακμή bin (bin edge) είναι ένα σημείο στον άξονα συχνότητας που διαχωρίζει ένα bin σε δύο νέα. Σε κάθε βήμα του αλγορίθμου οι ακμές των bins βρίσκονται στις ίδιες συχνότητες για όλα τα φάσματα εισόδου. Στην αρχή της διαδικασίας, ο αλγόριθμος θεωρεί όλο το φάσμα ως ένα bin. Καθώς η διαδικασία εξελίσσεται, κάθε σημείο στο εύρος συχνοτήτων ενός συγκεκριμένου bin αξιολογείται ως μία υποψήφια νέα ακμή. Η υποψήφια αυτή ακμή bin διαχωρίζει το τρέχον bin σε δύο νέα, η ποιότητα των οποίων συγκρίνεται με την ποιότητα του τρέχοντος bin. Η ποιότητα ενός bin αξιολογείται με συγκεκριμένο μέγεθος, την τιμή bin (bin value). Η καλύτερη υποψήφια ακμή bin επιλέγεται και το bin διασπάται σε δύο νέα bins. Η διαδικασία επαναλαμβάνεται για κάθε bin όσο το άθροισμα των τιμών των νέων bins ξεπερνά την τιμή του παλιού. Βασικό πλεονέκτημα της μεθόδου Adaptive Intelligent binning είναι ότι δεν απαιτεί τον ορισμό παραμέτρων από το χρήστη. (De Meyer, 2008)

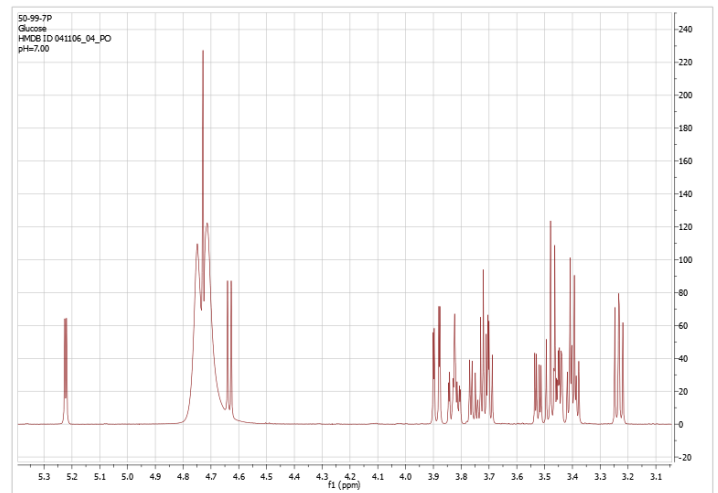
#### **4.2.8. Αναγνώριση και Ποσοτικοποίηση μεταβολιτών**

Η αναγνώριση και ποσοτικοποίηση των μεταβολιτών που βρίσκονται σε ένα δείγμα είναι ουσιώδης στόχος στο πεδίο της μεταβολομικής, καθώς αποτελεί το επίκεντρο όλων των διαγνωστικών και φαρμακολογικών εφαρμογών της. Η εργασία αυτή είναι ιδιαίτερα απαιτητική σε χρόνο και προϋποθέτει την ύπαρξη εξειδικευμένων γνώσεων. Για αυτό το λόγο, η υπολογιστικά υποβοηθούμενη αναγνώριση και ποσοτικοποίηση μεταβολιτών σε μονοδιάστατα φάσματα  $^1\text{H}$  NMR έχει υπάρξει το αντικείμενο αρκετών ερευνητικών προσπαθειών. Μία τακτική εξαγωγής μεταβολομικής πληροφορίας από φάσματα NMR είναι η επεξεργασία του φάσματος με μεθόδους πολυμεταβλητής στατιστικής ανάλυσης, όπως η μέθοδος ανάλυσης κύριων συνιστωσών (principal component analysis, PCA) και η ανάλυση μερικών ελαχίστων τετραγώνων (partial least squares, PLS) (Dona et al., 2016; Lindon et al., 2001). Άλλη προσέγγιση είναι η δημιουργία μεθόδων με αντικείμενο την αυτόματη ή ημιαυτόματη υπολογιστική αναγνώριση μεταβολιτών, οι οποίες αξιοποιούν κατάλληλες βάσεις δεδομένων όπως η Human Metabolome Database (Wishart et al, 2013), που περιέχουν πληροφορίες για τα φάσματα μεμονωμένων μεταβολιτών (Εικόνα 4.5). Οι υπολογιστικές μέθοδοι που παρουσιάζονται στη συνέχεια έχουν υιοθετήσει την προσέγγιση αυτή.





(α)



(β)

Εικόνα 4.5. Μονοδιάστατα φάσματα συχνότητας δύο μεταβολιτών, όπως απεικονίζονται από στο λογισμικό MestReNova. Τα αρχεία του σήματος FID των μεμονωμένων αυτών μεταβολιτών αποκτήθηκαν από τη βάση δεδομένων HMDB. (α) Κρεατίνη (β) D-Γλυκόζη.

Το BQuant είναι ένα λογισμικό με αντικείμενο την αυτόματη αναγνώριση και ποσοτικοποίηση μεταβολιτών σε περιοχές φασμάτων  $^1\text{H}$  NMR, το οποίο στηρίζεται σε μπαγεσιανή πιθανοτική μοντελοποίηση. Το BQuant λαμβάνει ως είσοδο ένα σύνολο φασμάτων NMR, ένα εύρος συχνοτήτων ενδιαφέροντος (σε ppm) και μία βάση δεδομένων με φάσματα μεταβολιτών. Το φάσμα εισόδου μοντελοποιείται ως ένας γραμμικός συνδυασμός των φασμάτων της βάσης δεδομένων με άγνωστα βάρη. Οι μεταβολίτες της βάσης δεδομένων αντιμετωπίζονται ως υποψήφιες μεταβλητές και το πρόβλημα της αναγνώρισης μεταβολιτών ανάγεται σε ένα πρόβλημα επιλογής μεταβλητών. Οι πιθανοί μεταβολίτες που βρίσκονται στο φάσμα εισόδου προσδιορίζονται με βάση ένα στοχαστικό σχήμα αναζήτησης μεταβλητών. Η έξοδος του BQuant είναι το σύνολο των ταυτοποιημένων μεταβολιτών, οι πιθανότητες παρουσίας τους στα φάσματα εισόδου και οι συγκεντρώσεις τους. Το BQuant είναι υλοποιημένο στη γλώσσα προγραμματισμού R (Zheng et al., 2011).

Το BATMAN είναι ένα λογισμικό με αντικείμενο την ημιαυτόματη ανάλυση μεταβολιτών σε φάσματα  $^1\text{H}$  NMR που στηρίζεται σε μπαγεσιανή μοντελοποίηση. Ο χρήστης δίνει ως είσοδο ένα φάσμα  $^1\text{H}$  NMR και προσδιορίζει μία περιοχή συχνοτήτων ενδιαφέροντος. Στη συνέχεια, το πρόγραμμα πραγματοποιεί αποσυνέλιξη των κορυφών στη δεδομένη περιοχή, τις αναθέτει σε συγκεκριμένους μεταβολίτες από μία λίστα υποψήφιων μεταβολιτών ενδιαφέροντος και κάνει εκτιμήσεις των συγκεντρώσεών τους. Το BATMAN στηρίζεται στην εφαρμογή ενός αλγορίθμου Markov chain Monte Carlo (MCMC) και είναι υλοποιημένο στη γλώσσα προγραμματισμού R (Astle et al., 2012; Hao et al., 2012; Hao et al., 2014).

Το MetaboHunter είναι μία διαδικτυακή εφαρμογή για την αναγνώριση μεταβολιτών σε μονοδιάστατα φάσματα  $^1\text{H}$ -NMR αξιοποιώντας δύο βάσεις δεδομένων μεταβολιτών (Human Metabolome Database, Madison Metabolome Consortium Database). Το MetaboHunter δέχεται ως είσοδο ένα μονοδιάστατο φάσμα  $^1\text{H}$  NMR ή μία λίστα φασματικών κορυφών. Ο χρήστης δίνει ως παραμέτρους το κατώφλι θορύβου, καθώς και το shift tolerance και το confidence threshold, η χρησιμότητα των οποίων εξηγείται παρακάτω. Το φάσμα ή λίστα κορυφών εισόδου υπόκειται σε κατωφλίωση με βάση το κατώφλι θορύβου που όρισε

ο χρήστης. Ο χρήστης επιλέγει μία από τρεις διαθέσιμες μεθόδους αναζήτησης (MH1, MH2, MH3). Η πρώτη μέθοδος (MH1) έχει ως αντικείμενο την ακριβή αντιστοίχιση των κορυφών του φάσματος εισόδου με τις φασματικές κορυφές των μεταβολιτών της βάσης δεδομένων. Για κάθε μεταβολίτη της επιλεγμένης βάσης δεδομένων υπολογίζεται μία βαθμολογία ( $\text{significance score} = \frac{\text{total number of matched peaks}}{1 + \text{total number of peaks}}$ ). Οι μεταβολίτες με βαθμολογία μεγαλύτερη ή ίση του confidence threshold δίνονται ως έξοδος. Η δεύτερη μέθοδος αναζήτησης (MH2) διαφοροποιείται από την MH1 στο ότι επιτρέπει την πιο ελαστική αντιστοίχιση των κορυφών του φάσματος εισόδου με τις κορυφές των μεταβολιτών της βάσης δεδομένων. Συγκεκριμένα, δύο κορυφές θεωρείται ότι ταιριάζουν αν οι χημικές μετατοπίσεις τους απέχουν το πολύ κατά την τιμή της παραμέτρου shift tolerance που ορίζεται από το χρήστη, επομένως οι βαθμολογίες που υπολογίζονται σύμφωνα με τη μέθοδο MH2 είναι κατά κανόνα υψηλότερες από ό,τι στην MH1. Η τρίτη μέθοδος (MH3) συνιστά μία άπληστη μέθοδο αναζήτησης που επιβάλλει τον αποκλεισμό των κορυφών μέσω μίας επαναληπτικής διαδικασίας βαθμολόγησης μεταβολιτών. Παρόμοια με τις MH1 και MH2, στην MH3 υπολογίζονται οι βαθμολογίες των μεταβολιτών. Ο μεταβολίτης με την υψηλότερη βαθμολογία και τον υψηλότερο αριθμό κορυφών επιλέγεται και οι κορυφές που του αντιστοιχούν δεσμεύονται, δηλαδή αφαιρούνται από το σύνολο φασματικών κορυφών που είναι διαθέσιμες προς αντιστοίχιση με τη βάση δεδομένων. Η διαδικασία επαναλαμβάνεται όσο υπάρχουν διαθέσιμες φασματικές κορυφές και υποψήφιοι μεταβολίτες προς αντιστοίχιση. Η μέθοδος MH3 δίνει πλεονέκτημα στους μεταβολίτες με υψηλότερη βαθμολογία και υψηλότερο αριθμό κορυφών, και μειώνει τα ψευδή θετικά αποτελέσματα με τον κίνδυνο αύξησης των ψευδών αρνητικών αποτελεσμάτων. Η έξοδος του MetaboHunter είναι μία λίστα πιθανών μεταβολιτών (Tulpan et al., 2011).

Το Bayesil είναι ένα διαδικτυακό λογισμικό με αντικείμενο την αυτόματη αναγνώριση και ποσοτικοποίηση μεταβολιτών σε μονοδιάστατα φάσματα  $^1\text{H NMR}$ . Το Bayesil δέχεται ως είσοδο ένα σήμα ελεύθερης επαγωγικής απόσβεσης. Αρχικά πραγματοποιείται αυτόματα μία σειρά βημάτων προεπεξεργασίας, όπως μετασχηματισμός Fourier και Hilbert, διόρθωση φάσης, διόρθωση βασικής γραμμής και αποθορύβωση. Η προεπεξεργασία του φάσματος ακολουθείται από τη διαδικασία αναγνώρισης και ποσοτικοποίησης μεταβολιτών, σύμφωνα με την οποία το φάσμα διαχωρίζεται σε ανεξάρτητες γειτνιάζουσες περιοχές, οι συσχετίσεις μεταξύ των οποίων αναπαριστώνται με ένα πιθανοτικό γραφικό μοντέλο. Το πιο πιθανό μεταβολικό προφίλ για το φάσμα εισόδου επιλέγεται με συνεπαγωγή (Ravanbakhsh et al., 2015).

Το Autofit είναι ένας αλγόριθμος που επικεντρώνεται στην αυτόματη αναγνώριση και ποσοτικοποίηση μεταβολιτών από μονοδιάστατα φάσματα  $^1\text{H NMR}$  και έχει ενσωματωθεί στο λογισμικό Chenomx. Απαιτεί ως είσοδο ένα μονοδιάστατο φάσμα  $^1\text{H NMR}$ , καθώς και πληροφορίες σχετικά με το pH του φάσματος εισόδου και τη συγκέντρωση της ένωσης αναφοράς. Αρχικά εντοπίζεται η ένωση αναφοράς (DSS, TSP ή φορμικό οξύ). Στη συνέχεια, η βιβλιοθήκη φασμάτων αναφοράς προσαρμόζεται με βάση τη θέση, την ένταση και το εύρος της κορυφής αναφοράς. Το φάσμα εισόδου ανακατασκευάζεται ως γραμμικός συνδυασμός των φασμάτων μεταβολιτών αναφοράς με βάση ένα λορεντζιανό μοντέλο. Ο αλγόριθμος αξιοποιεί ένα συνδυασμό προσομοιωμένης απόπτωσης κι ενός γενετικού αλγορίθμου (Mercier et al., 2011).

Το FOCUS είναι ένα λογισμικό του οποίου το αντικείμενο συμπεριλαμβάνει την αναγνώριση μεταβολιτών. Περιληπτικά, η μεθοδολογία του FOCUS μπορεί να περιγραφεί ως κατάτμηση και ευθυγράμμιση των φασμάτων εισόδου, ανίχνευση κορυφών και αναγνώριση μεταβολιτών. Αρχικά, τα ανεπεξέργαστα μονοδιάστατα φάσματα διαιρούνται αυτόματα σε ίσα αλληλοεπικαλυπτόμενα τμήματα, αρκετά πλατιά ώστε να περιέχουν μία ή περισσότερες κορυφές. Στη συνέχεια, πραγματοποιείται ευθυγράμμιση των φασμάτων με βάση τον αλγόριθμο RUNAS (Recursive Unreferenced Alignment of Spectra), ο οποίος δεν απαιτεί τη χρήση φάσματος αναφοράς, καθώς και ανίχνευση κορυφών στα ευθυγραμμισμένα φάσματα. Όσον αφορά την αναγνώριση μεταβολιτών, το FOCUS βασίζεται στο γεγονός ότι οι κοντινές κορυφές στο φάσμα ενός μεταβολίτη προκύπτουν από το ίδιο πρωτόνιο και παρουσιάζουν μικρότερες διακυμάνσεις στον άξονα συχνοτήτων από ό,τι οι κορυφές που προκύπτουν από διαφορετικά πρωτόνια. Η διαδικασία αναγνώρισης μεταβολιτών του FOCUS μπορεί να περιγραφεί με βάση τρεις βαθμολογίες. Το πρώτο βήμα αφορά την ταυτοποίηση των κορυφών που αντιστοιχούν στο ίδιο πρωτόνιο, η οποία επιτυγχάνεται με τον υπολογισμό ενός intracluster score. Το δεύτερο βήμα αφορά την ταυτοποίηση των κορυφών που αντιστοιχούν στα υπόλοιπα πρωτόνια ενός μεταβολίτη, η οποία επιτυγχάνεται με τον υπολογισμό ενός intercluster score. Τέλος, αν συσχετισμένες κορυφές βρίσκονται εκτός των παραθύρων που ορίζονται γύρω από τις κορυφές των πρωτονίων μεταβολιτών, υπολογίζεται ένα penalization score. Για κάθε ανιχνευμένη κορυφή το FOCUS αποκτά μία λίστα υποψήφιων αναγνωρισμένων μεταβολιτών, σε κάθε έναν από τους οποίους ανατίθεται μία βαθμολογία (το μέσο των intracluster, intercluster και penalization scores). (Alonso et al., 2013)

Ακόμη μία μεθοδολογία επεξεργασίας φασμάτων  $^1\text{H}$  NMR και αναγνώρισης μεταβολιτών έχει αναπτυχθεί από τους Jacob et al. (2013). Τα στάδιά της μπορούν να συνοψιστούν ως προεπεξεργασία και μείωση δεδομένων του φάσματος, ομαδοποίηση των κορυφών με βάση την απόσταση μεταξύ τους και αντιστοίχιση των ομάδων μεταβλητών με τους μεταβολίτες από μία βάση δεδομένων. Οι συγγραφείς χρησιμοποιούν έξι συναρτήσεις βαθμολόγησης, μία από τις οποίες είναι η συνάρτηση βαθμολόγησης που προτάθηκε από τους Tulpan et al (2011) για το λογισμικό MetaboHunter. (Jacob et al., 2013)

### 4.3 Αναφορές

- Agin, A., Heintz, D., Ruhland, E., Chao de la Barca, J. M., Zumsteg, J., Moal, V., ... Namer, I. J. (2016). Metabolomics—an overview. From basic principles to potential biomarkers (part 1). *Médecine Nucléaire*, 40(1), 4–10. <http://doi.org/10.1016/j.mednuc.2015.12.006>
- Agnolet, S., Jaroszewski, J. W., Verpoorte, R., & Staerk, D. (2010). H NMR-based metabolomics combined with HPLC-PDA-MS-SPE-NMR for investigation of standardized Ginkgo biloba preparations. *Metabolomics: Official Journal of the Metabolomic Society*, 6(2), 292–302. <http://doi.org/10.1007/s11306-009-0195-x>
- Alonso, A., Rodríguez, M. A., Vinaixa, M., Tortosa, R., Correig, X., Julià, A., & Marsal, S. (2014). Focus: A Robust Workflow for One-Dimensional NMR Spectral Analysis. *Analytical Chemistry*, 86(2), 1160–1169. <http://doi.org/10.1021/ac403110u>
- Anderson, P. E., Mahle, D. A., Doom, T. E., Reo, N. V., DelRaso, N. J., & Raymer, M. L. (2011). Dynamic adaptive binning: an improved quantification technique for NMR

- spectroscopic data. *Metabolomics*, 7(2), 179–190. <http://doi.org/10.1007/s11306-010-0242-7>
- Anderson, P. E., Reo, N. V., DelRaso, N. J., Doom, T. E., & Raymer, M. L. (2008). Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics. *Metabolomics*, 4(3), 261–272. <http://doi.org/10.1007/s11306-008-0117-3>
- Armitage, E. G., & Barbas, C. (2014). Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *Journal of Pharmaceutical and Biomedical Analysis*, 87, 1–11. <http://doi.org/10.1016/j.jpba.2013.08.041>
- Astle, W., De Iorio, M., Richardson, S., Stephens, D., & Ebbels, T. (2012). A Bayesian Model of NMR Spectra for the Deconvolution and Quantification of Metabolites in Complex Biological Mixtures. *Journal of the American Statistical Association*, 107(500), 1259–1271. <http://doi.org/10.1080/01621459.2012.695661>
- Baraldi, E., Carraro, S., Giordano, G., Reniero, F., Perilongo, G., & Zacchello, F. (2009). Metabolomics: moving towards personalized medicine. *Italian Journal of Pediatrics*, 35(1), 30. <http://doi.org/10.1186/1824-7288-35-30>
- Bloch, F., Hansen, W. W., & Packard, M. (1946). The Nuclear Induction Experiment. *Physical Review*, 70(7–8), 474–485. <http://doi.org/10.1103/PhysRev.70.474>
- Blümich, B. (2016). Introduction to compact NMR: A review of methods. *TrAC Trends in Analytical Chemistry*, 83, 2–11. <http://doi.org/10.1016/j.trac.2015.12.012>
- Bogdanov, M., Matson, W. R., Wang, L., Matson, T., Saunders-Pullman, R., Bressman, S. S., & Beal, M. F. (2008). Metabolomic profiling to develop blood biomarkers for Parkinson's disease. *Brain*, 131(2), 389–396. <http://doi.org/10.1093/brain/awm304>
- Bothwell, J. H. F., & Griffin, J. L. (2011). An introduction to biological nuclear magnetic resonance spectroscopy. *Biological Reviews of the Cambridge Philosophical Society*, 86(2), 493–510. <http://doi.org/10.1111/j.1469-185X.2010.00157.x>
- Bougnoux, P., Hajjaji, N., & Couet, C. (2008). The lipidome as a composite biomarker of the modifiable part of the risk of breast cancer. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 79(3–5), 93–96. <http://doi.org/10.1016/j.plefa.2008.09.004>
- Chapter 13: Spectroscopy. (n.d.). Retrieved from <http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch13/ch13-nmr-2.html>
- Chen, C., Gonzalez, F. J., & Idle, J. R. (2007). LC-MS-Based Metabolomics in Drug Metabolism. *Drug Metabolism Reviews*, 39(2–3), 581–597. <http://doi.org/10.1080/03602530701497804>
- Das, M. K., Bishwal, S. C., Das, A., Dabral, D., Badireddy, V. K., Pandit, B., ... Nanda, R. K. (2015). Deregulated Tyrosine–Phenylalanine Metabolism in Pulmonary Tuberculosis Patients. *Journal of Proteome Research*, 14(4), 1947–1956. <http://doi.org/10.1021/acs.jproteome.5b00016>
- Davis, R. A., Charlton, A. J., Godward, J., Jones, S. A., Harrison, M., & Wilson, J. C. (2007). Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemometrics and Intelligent Laboratory Systems*, 85(1), 144–154. <http://doi.org/10.1016/j.chemolab.2006.08.014>
- De Meyer, T., Sinnave, D., Van Gasse, B., Tsiportkova, E., Rietzschel, E. R., De Buyzere, M. L., ... Van Criekinge, W. (2008). NMR-Based Characterization of Metabolic

- Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. *Analytical Chemistry*, 80(10), 3783–3790. <http://doi.org/10.1021/ac7025964>
- Dona, A. C., Kyriakides, M., Scott, F., Shephard, E. A., Varshavi, D., Veselkov, K., & Everett, J. R. (2016). A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Computational and Structural Biotechnology Journal*, 14, 135–153.
- Dunn, W. B., & Ellis, D. I. (2005). Metabolomics: Current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry*, 24(4), 285–294. <http://doi.org/10.1016/J.TRAC.2004.11.021>
- Dunn, W. B., Bailey, N. J. C., & Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *The Analyst*, 130(5), 606. <http://doi.org/10.1039/b418288j>
- Eilers, P. H. C. (2003). A Perfect Smoother. *Analytical Chemistry*, 75(14), 3631–3636. <http://doi.org/10.1021/ac034173t>
- Emsley, J. W., & Feeney, J. (2007). Forty years of Progress in Nuclear Magnetic Resonance Spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 50(4), 179–198. <http://doi.org/10.1016/J.PNMRS.2007.01.002>
- Fiehn, O. (2001). Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comparative and Functional Genomics*, 2(3), 155–168. <http://doi.org/10.1002/cfg.82>
- Fiehn, O. (2002). Metabolomics--the link between genotypes and phenotypes. *Plant Molecular Biology*, 48(1–2), 155–71. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11860207>
- Garrod, S., Bollard, M. E., Nicholls, A. W., Connor, S. C., Connelly, J., Nicholson, J. K., & Elaine Holmes. (2005). Integrated Metabonomic Analysis of the Multiorgan Effects of Hydrazine Toxicity in the Rat. <http://doi.org/10.1021/TX0498915>
- Gorter, C. J., & Broer, L. J. F. (1942). Negative result of an attempt to observe nuclear magnetic resonance in solids. *Physica*, 9(6), 591–596. [http://doi.org/10.1016/S0031-8914\(42\)80073-7](http://doi.org/10.1016/S0031-8914(42)80073-7)
- Guo, Y., Wang, X., Qiu, L., Qin, X., Liu, H., Wang, Y., ... Li, Z. (2012). Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using Fourier transform ion cyclotron resonance mass spectrometry. *Clinica Chimica Acta*, 414, 135–141. <http://doi.org/10.1016/j.cca.2012.08.010>
- Hassan-Smith, G., Wallace, G. R., Douglas, M. R., & Sinclair, A. J. (2012). The role of metabolomics in neurological disease. *Journal of Neuroimmunology*, 248(1–2), 48–52. <http://doi.org/10.1016/J.JNEUROIM.2012.01.009>
- Hao, J., Astle, W., De Iorio, M., & Ebbels, T. M. D. (2012). BATMAN--an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics (Oxford, England)*, 28(15), 2088–90. <http://doi.org/10.1093/bioinformatics/bts308>
- Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. D. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature Protocols*, 9(6), 1416–27. <http://doi.org/10.1038/nprot.2014.090>

- Harris, R. K., Becker, E. D., Cabral de Menezes, S. M., Goodfellow, R., & Granger, P. (2002). NMR nomenclature: nuclear spin properties and conventions for chemical shifts. IUPAC Recommendations 2001. International Union of Pure and Applied Chemistry. Physical Chemistry Division. Commission on Molecular Structure and Spectroscopy. *Magnetic Resonance in Chemistry*, 40(7), 489–505. <http://doi.org/10.1002/mrc.1042>
- Hanson, L. G. (2008). Is quantum mechanics necessary for understanding magnetic resonance? *Concepts in Magnetic Resonance Part A*, 32A(5), 329–340. <http://doi.org/10.1002/cmr.a.20123>
- Han, X., M. Holtzman, D., W. McKeel, D., Kelley, J., & Morris, J. C. (2002). Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer's disease: potential role in disease pathogenesis. *Journal of Neurochemistry*, 82(4), 809–818. <http://doi.org/10.1046/j.1471-4159.2002.00997.x>
- Hori, S., Nishiumi, S., Kobayashi, K., Shinohara, M., Hatakeyama, Y., Kotani, Y., ... Yoshida, M. (2011). A metabolomic approach to lung cancer. *Lung Cancer*, 74(2), 284–292. <http://doi.org/10.1016/j.lungcan.2011.02.008>
- Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6), 1239–49. <http://doi.org/10.1111/j.1476-5381.2010.01127.x>
- Jacob, D., Deborde, C., & Moing, A. (2013). An efficient spectra processing method for metabolite identification from <sup>1</sup>H-NMR metabolomics data. *Analytical and Bioanalytical Chemistry*, 405(15), 5049–61. <http://doi.org/10.1007/s00216-013-6852-y>
- Ji, Y., Hebring, S., Zhu, H., Jenkins, G. D., Biernacka, J., Snyder, K., ... Weinshilboum, R. M. (2011). Glycine and a Glycine Dehydrogenase (GLDC) SNP as Citalopram/Escitalopram Response Biomarkers in Depression: Pharmacometabolomics-Informed Pharmacogenomics. *Clinical Pharmacology & Therapeutics*, 89(1), 97–104. <http://doi.org/10.1038/clpt.2010.250>
- Jung, J. Y., Lee, H.-S., Kang, D.-G., Kim, N. S., Cha, M. H., Bang, O.-S., ... Hwang, G.-S. (2011). 1H-NMR-based metabolomics study of cerebral infarction. *Stroke; a Journal of Cerebral Circulation*, 42(5), 1282–8. <http://doi.org/10.1161/STROKEAHA.110.598789>
- Keeler, J. (2010). *Understanding NMR spectroscopy*. John Wiley, Chichester, UK. Retrieved from [https://books.google.gr/books?hl=en&lr=&id=WUmCpq30pygC&oi=fnd&pg=PT13&dq=Understanding+NMR+Spectroscopy+keeler&ots=SqFGUbgPWS&sig=QVZncvi65\\_uIRf9YCRTVvBwx\\_Bw&redir\\_esc=y#v=onepage&q=Understanding+NMR+Spectroscopy+keeler&f=false](https://books.google.gr/books?hl=en&lr=&id=WUmCpq30pygC&oi=fnd&pg=PT13&dq=Understanding+NMR+Spectroscopy+keeler&ots=SqFGUbgPWS&sig=QVZncvi65_uIRf9YCRTVvBwx_Bw&redir_esc=y#v=onepage&q=Understanding+NMR+Spectroscopy+keeler&f=false)
- Kessel, M. (2011). The problems with today's pharmaceutical business—an outsider's view. *Nature Biotechnology*, 29(1), 27–33. <http://doi.org/10.1038/nbt.1748>
- Keun, H. C. (2006). Metabonomic modeling of drug toxicity. *Pharmacology & Therapeutics*, 109(1–2), 92–106. <http://doi.org/10.1016/j.pharmthera.2005.06.008>
- Kilk, K., Mahlapuu, R., Soomets, U., & Langel, Ü. (2009). Analysis of in vitro toxicity of five cell-penetrating peptides by metabolic profiling. *Toxicology*, 265(3), 87–95. <http://doi.org/10.1016/J.TOX.2009.09.016>

- Kimberly, W. T., Wang, Y., Pham, L., Furie, K. L., & Gerszten, R. E. (2013). Metabolite Profiling Identifies a Branched Chain Amino Acid Signature in Acute Cardioembolic Stroke. *Stroke*, *44*(5), 1389–1395. <http://doi.org/10.1161/STROKEAHA.111.000397>
- Koen, N., Du Preez, I., & Loots, D. T. (2015). Metabolomics and Personalized Medicine. In *Advances in Protein Chemistry and Structural Biology*. <http://doi.org/10.1016/bs.apcsb.2015.09.003>
- Kordalewska, M., & Markuszewski, M. J. (2015). Metabolomics in cardiovascular diseases. *Journal of Pharmaceutical and Biomedical Analysis*, *113*, 121–136. <http://doi.org/10.1016/j.jpba.2015.04.021>
- Kosmidis, A. K., Kamisoglu, K., Calvano, S. E., Corbett, S. A., & Androulakis, I. P. (2013). Metabolomic fingerprinting: challenges and opportunities. *Critical Reviews in Biomedical Engineering*, *41*(3), 205–21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24579644>
- Kumar, B., Prakash, A., Ruhela, R. K., & Medhi, B. (2014). Potential of metabolomics in preclinical and clinical drug development. *Pharmacological Reports : PR*, *66*(6), 956–63. <http://doi.org/10.1016/j.pharep.2014.06.010>
- Lan, K., & Jia, W. (2010). An integrated metabolomics and pharmacokinetics strategy for multi-component drugs evaluation. *Current Drug Metabolism*, *11*(1), 105–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20302568>
- Lee, J.-E., Hwang, G.-S., Van Den Berg, F., Lee, C.-H., & Hong, Y.-S. (2009). Evidence of vintage effects on grape wines using <sup>1</sup>H NMR-based metabolomic study. *Analytica Chimica Acta*, *648*(1), 71–76. <http://doi.org/10.1016/J.ACA.2009.06.039>
- Lenz, E. M., Bright, J., Knight, R., Westwood, F. R., Davies, D., Major, H., & Wilson, I. D. (2005). Metabonomics with <sup>1</sup>H-NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. *Biomarkers*, *10*(2–3), 173–187. <http://doi.org/10.1080/13547500500094034>
- Lenz, E. M., Bright, J., Wilson, I. D., Morgan, S. R., & Nash, A. F. P. (2003). A <sup>1</sup>H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. *Journal of Pharmaceutical and Biomedical Analysis*, *33*(5), 1103–1115. [http://doi.org/10.1016/S0731-7085\(03\)00410-2](http://doi.org/10.1016/S0731-7085(03)00410-2)
- Lindon, J. C., Holmes, E., & Nicholson, J. K. (2001). Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *39*(1), 1–40. [http://doi.org/10.1016/S0079-6565\(00\)00036-4](http://doi.org/10.1016/S0079-6565(00)00036-4)
- Lindon, J. C., & Nicholson, J. K. (2008). Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *TrAC Trends in Analytical Chemistry*, *27*(3), 194–204. <http://doi.org/10.1016/j.trac.2007.08.009>
- Lindon, J. C., Keun, H. C., Ebbels, T. M., Pearce, J. M., Holmes, E., & Nicholson, J. K. (2005). The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics*, *6*(7), 691–699. <http://doi.org/10.2217/14622416.6.7.691>
- Lv, W., & Yang, T. (2012). Identification of possible biomarkers for breast cancer from free fatty acid profiles determined by GC-MS and multivariate statistical analysis. *Clinical Biochemistry*, *45*(1–2), 127–133. <http://doi.org/10.1016/j.clinbiochem.2011.10.011>

- Marco, A., & Martí'nez, J.-J. (2010). Polynomial least squares fitting in the Bernstein basis. *Linear Algebra and Its Applications*, 433(7), 1254–1264. <http://doi.org/10.1016/J.LAA.2010.06.031>
- Mercier, P., Lewis, M. J., Chang, D., Baker, D., & Wishart, D. S. (2011). Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *Journal of Biomolecular NMR*, 49(3–4), 307–23. <http://doi.org/10.1007/s10858-011-9480-x>
- Mlynárik, V. (2016). Introduction to nuclear magnetic resonance. *Analytical Biochemistry*. <http://doi.org/10.1016/j.ab.2016.05.006>
- Mortishire-Smith, R. J., Skiles, G. L., Lawrence, J. W., Spence, S., Nicholls, A. W., Johnson, B. A., & Nicholson, J. K. (2004). Use of Metabonomics to Identify Impaired Fatty Acid Metabolism as the Mechanism of a Drug-Induced Toxicity. *Chemical Research in Toxicology*, 17(2), 165–173. <http://doi.org/10.1021/tx034123j>
- Nam, H., Chung, B. C., Kim, Y., Lee, K., & Lee, D. (2009). Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. *Bioinformatics*, 25(23), 3151–3157. <http://doi.org/10.1093/bioinformatics/btp558>
- Naz, S., Vallejo, M., García, A., & Barbas, C. (2014). Method validation strategies involved in non-targeted metabolomics. *Journal of Chromatography A*, 1353, 99–105. <http://doi.org/10.1016/j.chroma.2014.04.071>
- Nicholls, A. W., Russell J. Mortishire-Smith, & Nicholson†, J. K. (2003). NMR Spectroscopic-Based Metabonomic Studies of Urinary Metabolite Variation in Acclimatizing Germ-Free Rats. <http://doi.org/10.1021/TX0340293>
- Nicholson, J. K., & Wilson, I. D. (2003). Opinion: understanding “global” systems biology: metabonomics and the continuum of metabolism. *Nature Reviews. Drug Discovery*, 2(8), 668–76. <http://doi.org/10.1038/nrd1157>
- Nicholson, J. K., Buckingham, M. J., & Sadler, P. J. (1983). High resolution 1 H n.m.r. studies of vertebrate blood and plasma. *Biochemical Journal*, 211(3), 605–615. <http://doi.org/10.1042/bj2110605>
- Nicholson, J. K. (2006). Global systems biology, personalized medicine and molecular epidemiology. *Molecular Systems Biology*, 2, 52. <http://doi.org/10.1038/msb4100095>
- Nicholson, J. K., Lindon, J. C., & Holmes, E. (1999). “Metabonomics”: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29(11), 1181–1189. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/004982599238047>
- Noga, M. J., Dane, A., Shi, S., Attali, A., van Aken, H., Suidgeest, E., ... Hankemeier, T. (2012). Metabolomics of cerebrospinal fluid reveals changes in the central nervous system metabolism in a rat model of multiple sclerosis. *Metabolomics : Official Journal of the Metabolomic Society*, 8(2), 253–263. <http://doi.org/10.1007/s11306-011-0306-3>
- Odunsi, K., Wollman, R. M., Ambrosone, C. B., Hutson, A., McCann, S. E., Tammela, J., ... Alderfer, J. L. (2005). Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *International Journal of Cancer*, 113(5), 782–788. <http://doi.org/10.1002/ijc.20651>
- Oliver, S. G., Winson, M. K., Kell, D. B., & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16(9), 373–8. [http://doi.org/10.1016/S0167-7799\(98\)01214-1](http://doi.org/10.1016/S0167-7799(98)01214-1)



- Orešič, M., Hyötyläinen, T., Herukka, S.-K., Sysi-Aho, M., Mattila, I., Seppänen-Laakso, T., ... Soininen, H. (2011). Metabolome in progression to Alzheimer's disease. *Translational Psychiatry*, *1*(12), e57. <http://doi.org/10.1038/tp.2011.55>
- Purcell, E. M., Torrey, H. C., & Pound, R. V. (1946). Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, *69*(1–2), 37–38. <http://doi.org/10.1103/PhysRev.69.37>
- Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., ... Wishart, D. S. (2015). Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PLOS ONE*, *10*(5), e0124219. <http://doi.org/10.1371/journal.pone.0124219>
- Roberts, L. D., Souza, A. L., Gerszten, R. E., & Clish, C. B. (2012). Targeted metabolomics. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.], Chapter 30*, Unit 30.2.1-24. <http://doi.org/10.1002/0471142727.mb3002s98>
- Robertson, D. G., Watkins, P. B., & Reily, M. D. (2011). Metabolomics in Toxicology: Preclinical and Clinical Applications. *Toxicological Sciences*, *120*(Supplement 1), S146–S170. <http://doi.org/10.1093/toxsci/kfq358>
- Rozen, S., Cudkowicz, M. E., Bogdanov, M., Matson, W. R., Kristal, B. S., Beecher, C., ... Kaddurah-Daouk, R. (2005). Metabolomic analysis and signatures in motor neuron disease. *Metabolomics*, *1*(2), 101–108. <http://doi.org/10.1007/s11306-005-4810-1>
- Schnackenberg, L. K., & Beger, R. D. (2007). Metabolomic biomarkers: their role in the critical path. *Drug Discovery Today: Technologies*, *4*(1), 13–16. <http://doi.org/10.1016/j.ddtec.2007.10.012>
- Shah, S. H., Kraus, W. E., & Newgard, C. B. (2012). Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation*, *126*(9), 1110–20. <http://doi.org/10.1161/CIRCULATIONAHA.111.060368>
- Shah, S. H., Sun, J.-L., Stevens, R. D., Bain, J. R., Muehlbauer, M. J., Pieper, K. S., ... Newby, L. K. (2012). Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *American Heart Journal*, *163*(5), 844–850.e1. <http://doi.org/10.1016/j.ahj.2012.02.005>
- Smith, D. A., & Schmid, E. F. (2006). Drug withdrawals and the lessons within. *Current Opinion in Drug Discovery & Development*, *9*(1), 38–46. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16445116>
- Smolinska, A., Blanchet, L., Buydens, L. M. C., & Wijmenga, S. S. (2012). NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Analytica Chimica Acta*, *750*, 82–97. <http://doi.org/10.1016/j.aca.2012.05.049>
- Sousa, S. A. A., Magalhães, A., & Ferreira, M. M. C. (2013). Optimized bucketing for NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory Systems*, *122*, 93–102. <http://doi.org/10.1016/j.chemolab.2013.01.006>
- Spratlin, J. L., Serkova, N. J., & Eckhardt, S. G. (2009). Clinical applications of metabolomics in oncology: a review. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *15*(2), 431–40. <http://doi.org/10.1158/1078-0432.CCR-08-1059>

- Thomas, C. E., & Ganji, G. (2006). Integration of genomic and metabonomic data in systems biology--are we "there" yet? *Current Opinion in Drug Discovery & Development*, 9(1), 92–100. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16445121>
- Trupp, M., Zhu, H., Wikoff, W. R., Baillie, R. A., Zeng, Z.-B., Karp, P. D., ... Kaddurah-Daouk, R. (2012). Metabolomics Reveals Amino Acids Contribute to Variation in Response to Simvastatin Treatment. *PLoS ONE*, 7(7), e38386. <http://doi.org/10.1371/journal.pone.0038386>
- Tulpan, D., Léger, S., Belliveau, L., Culf, A., & Cuperlović-Culf, M. (2011). MetaboHunter: an automatic approach for identification of metabolites from 1H-NMR spectra of complex mixtures. *BMC Bioinformatics*, 12, 400. <http://doi.org/10.1186/1471-2105-12-400>
- van Vliet, E., Morath, S., Eskes, C., Linge, J., Rappsilber, J., Honegger, P., ... Coecke, S. (2008). A novel in vitro metabolomics approach for neurotoxicity testing, proof of principle for methyl mercury chloride and caffeine. *NeuroToxicology*, 29(1), 1–12. <http://doi.org/10.1016/J.NEURO.2007.09.007>
- Vulimiri, S. V., Misra, M., Hamm, J. T., Mitchell, M., & Berger, A. (2009). Effects of Mainstream Cigarette Smoke on the Global Metabolome of Human Lung Epithelial Cells. *Chemical Research in Toxicology*, 22(3), 492–503. <http://doi.org/10.1021/tx8003246>
- Wade, L.G., Jr. (2010). *Organic Chemistry*, 7th Edition. Pearson Prentice Hall.
- Wang, J., Li, Z., Chen, J., Zhao, H., Luo, L., Chen, C., ... Wang, W. (2013). Metabolomic identification of diagnostic plasma biomarkers in humans with chronic heart failure. *Molecular bioSystems*, 9(11), 2618–26. <http://doi.org/10.1039/c3mb70227h>
- Wei, S., Liu, L., Zhang, J., Bowers, J., Gowda, G. A. N., Seeger, H., ... Raftery, D. (2013). Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. *Molecular Oncology*, 7(3), 297–307. <http://doi.org/10.1016/j.molonc.2012.10.003>
- Whittaker, E. T. (1922). On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75. <http://doi.org/10.1017/s0013091500077853>
- Williams, R. E., Major, H., Lock, E. A., Lenz, E. M., & Wilson, I. D. (2005). d-Serine-induced nephrotoxicity: a HPLC–TOF/MS-based metabonomics approach. *Toxicology*, 207(2), 179–190. <http://doi.org/10.1016/J.TOX.2004.08.023>
- Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., ... Scalbert, A. (2013). HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Research*, 41(Database issue), D801-7. <http://doi.org/10.1093/nar/gks1065>
- Woo, H. M., Kim, K. M., Choi, M. H., Jung, B. H., Lee, J., Kong, G., ... Chung, B. C. (2009). Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. *Clinica Chimica Acta*, 400(1–2), 63–69. <http://doi.org/10.1016/j.cca.2008.10.014>
- Xi, Y., & Rocke, D. M. (2008). Baseline correction for NMR spectroscopic metabolomics data analysis. *BMC Bioinformatics*, 9, 324. <http://doi.org/10.1186/1471-2105-9-324>
- Zhang, T., Wu, X., Yin, M., Fan, L., Zhang, H., Zhao, F., ... Li, K. (2012). Discrimination between malignant and benign ovarian tumors by plasma metabolomic profiling using

ultra performance liquid chromatography/mass spectrometry. *Clinica Chimica Acta*, 413(9–10), 861–868. <http://doi.org/10.1016/j.cca.2012.01.026>

Zheng, M., Luo, X., Shen, Q., Wang, Y., Du, Y., Zhu, W., & Jiang, H. (2009). Site of metabolism prediction for six biotransformations mediated by cytochromes P450. *Bioinformatics*, 25(10), 1251–1258. <http://doi.org/10.1093/bioinformatics/btp140>

Zheng, C., Zhang, S., Ragg, S., Raftery, D., & Vitek, O. (2011). Identification and quantification of metabolites in (1)H NMR spectra by Bayesian model selection. *Bioinformatics (Oxford, England)*, 27(12), 1637–44. <http://doi.org/10.1093/bioinformatics/btr118>

5-HMR-2 Chemical Shift. (19 November, 2017). <https://www.chem.wisc.edu/areas/reich/nmr/05-hmr-02-delta.htm>



## Κεφάλαιο 5. Αυτόματη αναγνώριση μεταβολιτών σε μονοδιάστατα φάσματα $^1\text{H-NMR}$

---

### 5.1. Εισαγωγή

Όπως αναλύθηκε στο κεφάλαιο 4, η Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance, NMR) έχει αναδειχθεί ως μία από τις επικρατέστερες τεχνολογίες για την μεταβολομική ανάλυση, αφού δίνει τη δυνατότητα εξέτασης διαφόρων βιολογικών δειγμάτων και αποσαφήνισης βιοδεικτών ασθενειών και φαρμάκων. Κεντρικό ρόλο σε αυτές τις μελέτες έχει η αναγνώριση μεταβολιτών, αφού η διαλεύκανση των φασμάτων  $^1\text{H NMR}$  είναι μία απαιτητική και χρονοβόρα διαδικασία εξ' αιτίας της πολυπλοκότητας των βιολογικών υποστρωμάτων. Τα μονοδιάστατα φάσματα  $^1\text{H NMR}$  παρουσιάζουν αρκετές δυσκολίες. Κορυφές που ανήκουν σε διαφορετικούς μεταβολίτες μπορεί να αλληλεπικαλύπτονται με αποτέλεσμα την παραμόρφωση του σχήματός τους, ή να μετατοπίζονται στον άξονα συχνοτήτων εξ' αιτίας διακυμάνσεων στη συγκέντρωση των μεταβολιτών και στο pH του δείγματος (Larive et al. 2015; Smolinska et al. 2012; Kordalewska and Markuszewski 2015; Nagana Gowda, 2015).

Η υπολογιστική αναγνώριση μεταβολιτών έχει υπάρξει το αντικείμενο αρκετών ερευνητικών μελετών, οι οποίες παρουσιάστηκαν στο προηγούμενο κεφάλαιο (παράγραφος 4.2.7). Παρά τον όγκο και την ποιότητα των αντίστοιχων μελετών, το πρόβλημα της αυτόματης αναγνώρισης μεταβολιτών δεν έχει ακόμη επιλυθεί. Πολλά από τα λογισμικά που έχουν ως αντικείμενο την αναγνώριση μεταβολιτών, όπως το Chenomx και το BATMAN, λειτουργούν είτε με χειρωνακτικό είτε με ημιαυτόματο τρόπο, και συχνά απαιτούν τον ορισμό παραμέτρων από το χρήστη. Η τακτική αυτή αφ' ενός προϋποθέτει την ύπαρξη εξειδικευμένων γνώσεων, αφ' ετέρου κάνει τα αποτελέσματα εξαρτώμενα από το χρήστη. Άλλο χαρακτηριστικό που μπορεί να διαφοροποιείται από λογισμικό σε λογισμικό είναι ο τύπος εισόδου, που μπορεί να είναι σήμα στο πεδίο του χρόνου, μονοδιάστατο ή διδιάστατο φάσμα συχνοτήτων. Επίσης, οι δυσκολίες που προκύπτουν συμπεριλαμβάνουν τεχνικά προβλήματα όπως η χρήση παλιότερων εκδοχών γλωσσών προγραμματισμού ή σφάλματα λειτουργικού χρόνου που μπορούν να αναχαιτίσουν τη χρήση τους από χρήστες που δεν είναι εξοικωμένοι με αυτά τα εργαλεία λογισμικού. (Tulpan et al., 2011; Zheng et al., 2011; Mercier et al., 2011; Astle et al., 2012; Hao et al., 2014; Alonso et al., 2013; Ravanbakhsh et al., 2015; Jacob et al., 2013)

Στο παρόν κεφάλαιο παρουσιάζεται το νέο μεθοδολογικό σχήμα που έχει προταθεί για την αυτόματη αναγνώριση μεταβολιτών σε μονοδιάστατα φάσματα  $^1\text{H-NMR}$ . Η μεθοδολογία μπορεί να περιγραφεί επιγραμματικά ως η αντιστοίχιση ενός συνόλου υποψήφιων μεταβολιτών με το φάσμα εισόδου, βασισμένη σε μεταβολική πληροφορία από τη βάση δεδομένων Human Metabolome Database (HMDB) (Wishart et al. 2013). Η προτεινόμενη μεθοδολογία παίρνει ως είσοδο ένα μονοδιάστατο φάσμα  $^1\text{H-NMR}$ . Αρχικά εκτελείται αυτόματα ένα σύνολο βημάτων προεπεξεργασίας του φάσματος εισόδου, όπως η κατωφλίωση, η αποθορύβωση και η μείωση δεδομένων. Στη συνέχεια ξεκινά η αυτόματη διαδικασία αναγνώρισης μεταβολιτών. Βασικό στοιχείο της διαδικασίας αναγνώρισης μεταβολιτών είναι ότι λαμβάνει υπόψιν τις τιμές σταθερών σύζευξης, τις αναλογίες υψών για τις επί μέρους κορυφές κάθε πολλαπλής κορυφής (multiplet), τις αναλογίες υψών για τις

κορυφές ολόκληρου του μεταβολίτη, καθώς και τις σχετικές αποστάσεις μεταξύ των απλών ή πολλαπλών κορυφών ενός μεταβολίτη.

Το προτεινόμενο μεθοδολογικό σχήμα έχει ελεγχθεί σε (α) ένα τεχνητό μίγμα αμινοξέων, (β) ένα δείγμα ορού στο οποίο έχει προστεθεί το μίγμα αμινοξέων (γ) 40 in-house βιολογικά δείγματα (20 δείγματα ορού και 20 δείγματα αμνιακού υγρού), και (δ) 160 βιολογικά δείγματα ορού, διαθέσιμα από τη βάση δεδομένων MetaboLights (Kale et al. 2016). Η απόδοση της μεθοδολογίας αξιολογήθηκε ως προς την ακρίβεια, την ευαισθησία και την ειδικότητα των αποτελεσμάτων και συγκρίθηκε με τα λογισμικά αυτόματης αναγνώρισης μεταβολιτών MetaboHunter (Tulpan et al., 2011), Bayesil (Ravanbakhsh et al., 2015), Autofit (Mercier et al., 2011). Στον Πίνακα Π.1 γίνεται μία συγκριτική παρουσίαση της προτεινόμενης μεθοδολογίας και των υπόλοιπων δημοσιευμένων μεθοδολογιών αναγνώρισης μεταβολιτών ως προς τα βασικά μεθοδολογικά χαρακτηριστικά τους και τα φάσματα που χρησιμοποιήθηκαν για την αξιολόγησή τους.

## **5.2. Μεθοδολογία**

### **5.2.1. Περιγραφή δεδομένων**

#### **5.2.1.1. Φάσματα**

##### **Τεχνητά δείγματα**

Το πρώτο δείγμα (S1) στο οποίο εφαρμόστηκε η ανεπτυγμένη μεθοδολογία δημιουργήθηκε *in vitro* και είναι ένα διάλυμα αποτελούμενο από τα αμινοξέα L-Αλανίνη (L-Alanine), L-Βαλίνη (L-Valine), L-Λευκίνη (L-Leucine), L-Ισολευκίνη (L-Isoleucine), L-Γλουταμικό οξύ (L-Glutamic acid), L-Μεθειονίνη (L-Methionine), L-Αργινίνη (L-Arginine), L-Προλίνη (L-Proline) και Τριγονελλίνη (Trigonelline). Το δεύτερο δείγμα (S2) δημιουργήθηκε από την πρόσμιξη του δείγματος S1 με ένα δείγμα ορού. Το δείγμα αυτό επιλέχθηκε για να ελεγχθεί η ικανότητα του λογισμικού να ταυτοποιεί τους μεταβολίτες σε σύνθετα βιολογικά υποστρώματα. Το πρότυπο αναφοράς Trimethylsilylpropanoic acid (TSP) χρησιμοποιήθηκε για τα δείγματα S1, S2.

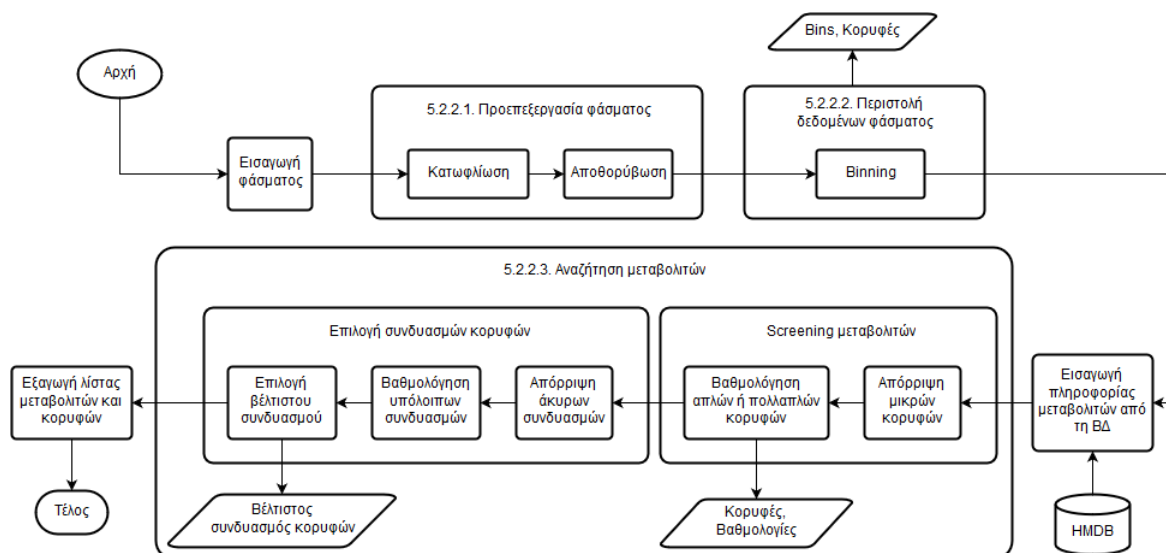
##### **Βιολογικά δείγματα**

Η μεθοδολογία εφαρμόστηκε ακόμη σε μονοδιάστατα φάσματα CPMG NMR από 20 δείγματα αμνιακού υγρού (S3-S22) και 20 δείγματα ορού (S23-S42) λημμένα από προηγούμενες μελέτες (Fotakis et al., 2016). Τα φάσματα κανονικοποιήθηκαν στην περιοχή της ένωσης αναφοράς και μετατράπηκαν σε ASCII format, χρησιμοποιώντας το λογισμικό MestreNOVA (<http://mestrelab.com>). Επιπλέον, η μεθοδολογία επαληθεύτηκε σε 18 δείγματα ορού (MTBLS174 1-18) από τους Wruck et al. (2015), 50 δείγματα ορού (MTBLS424 1-50) από τους Hart et al. (2017), 50 δείγματα ορού (MTBLS242 1-50) από τους Gralka et al. (2015) και 42 δείγματα ορού (MTBLS326 1-42) από τους Singh et al. (2017), τα οποία είναι διαθέσιμα στη βάση δεδομένων MetaboLights (Kale et al., 2016). Τα δείγματα S3-S42 περιέχουν το πρότυπο αναφοράς Sodium Maleate, ενώ τα δείγματα των ερευνών MTBLS174, MTBLS424, MTBLS242, MTBLS326 περιέχουν το πρότυπο αναφοράς TSP.

#### **5.2.1.2. Μεταβολίτες**

Ένα αρχείο που περιεχει μεταβολική πληροφορία για 850 μεταβολίτες συντέθηκε με βάση τα δημόσια διαθέσιμα αρχεία της Human Metabolome Database (HMDB) (Wishart et al., 2013). Καταχωρήσεις στην HMDB που έχουν ακατάλληλη μορφοποίηση ή είναι

διπλότυπες έχουν παραλειφθεί. Επιπλέον, συντέθηκε ένα μικρότερο αρχείο με 45 μεταβολίτες που χαρακτηρίζονται από έντονη παρουσία σε διαφορετικά βιολογικά υγρά, με στόχο να προσαρμοστεί η απόδοση της μεθοδολογίας με επίκεντρο αυτούς τους μεταβολίτες (Psychogios et al., 2011). Ένα δείγμα του εν λόγω αρχείου δίνεται στον Πίνακα Π.2 (Παράρτημα).



Εικόνα 5.1. Σχεδιάγραμμα του προτεινόμενου μεθοδολογικού σχήματος.

## 5.2.2. Περιγραφή μεθοδολογίας

Όπως προαναφέρθηκε, η μεθοδολογία ξεκινάει με τη διαδοχική εφαρμογή των βημάτων της κατωφλίωσης, της αποθορύβωσης και του binning. Στη συνέχεια, εφαρμόζεται η διαδικασία αναγνώρισης μεταβολιτών, η οποία αποτελείται από το στάδιο του screening μεταβολιτών και το στάδιο της επιλογής κορυφών. Στην Εικόνα 5.1 απεικονίζεται το σχεδιάγραμμα της μεθοδολογίας.

### 5.2.2.1. Προεπεξεργασία

#### Κατωφλίωση

Το φάσμα εισόδου αρχικά υπόκειται σε κατωφλίωση με στόχο την απομάκρυνση των κορυφών χαμηλής έντασης που αποτελούν θόρυβο. Το κατώφλι θορύβου υπολογίζεται αυτόματα ως η μέγιστη τιμή έντασης του φάσματος στην περιοχή συχνοτήτων 9.4-10 ppm, αφού σε αυτή την περιοχή δεν υπάρχουν κορυφές μεταβολιτών στα υπό εξέταση βιολογικά δείγματα σε υδαρή μέσα. Για αυτό το λόγο, η συλλογή δεδομένων για τα φάσματα εισόδου θα πρέπει να εκτείνεται τουλάχιστον μέχρι τα 10 ppm. Η επίδραση της κατωφλίωσης θορύβου στη φασματική περιοχή 7.24-7.4 ppm του δείγματος ορού S25 φαίνεται στην Εικόνα 5.2 (α).

#### Αποθορύβωση

Στη συνέχεια, το φάσμα εισόδου υπόκειται σε αποθορύβωση, δηλαδή συνέλιξη με το γκαουσιανό φίλτρο (εξίσωση (5.1)) (Haddad & Akansu, 1991). Η παράμετρος αποθορύβωσης  $N_G$ , όπου  $2N_G + 1$  είναι το μήκος γκαουσιανής μάσκας, τίθεται στην τιμή 4. Η τιμή της παραμέτρου αποθορύβωσης  $\sigma_G$ , που είναι η τυπική απόκλιση της γκαουσιανής

κατανομής, υπολογίζεται αυτόματα πραγματοποιώντας μία δοκιμαστική διαδικασία μείωσης δεδομένων και λαμβάνοντας υπόψιν τον αριθμό κορυφών που προκύπτουν από αυτές. Πιο συγκεκριμένα, το  $\sigma_G$  υπολογίζεται σύμφωνα με την εξίσωση (5.2), όπου το σύμβολο peaks συμβολίζει τον αριθμό κορυφών που προέκυψαν από τη δοκιμαστική διαδικασία μείωσης δεδομένων. Η επίδραση της αποθορύβωσης στη φασματική περιοχή 7.24-7.4 ppm του δείγματος ορού S25 φαίνεται στην Εικόνα 5.2 (β).

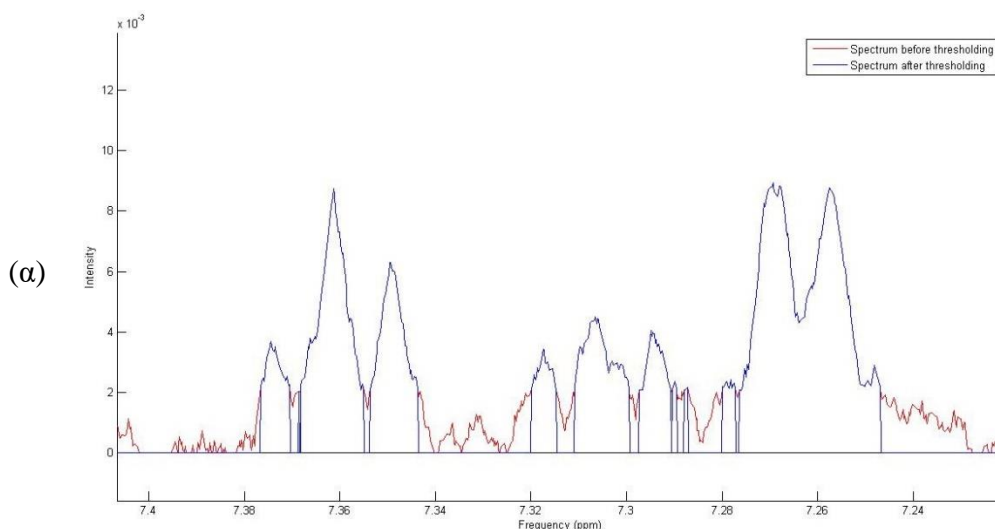
$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_G} e^{\frac{-x^2}{2\sigma_G^2}} \quad (5.1)$$

$$\sigma_G = 0.5 + 0.25 * ((\text{peaks}) \text{ div } 500) \quad (5.2)$$

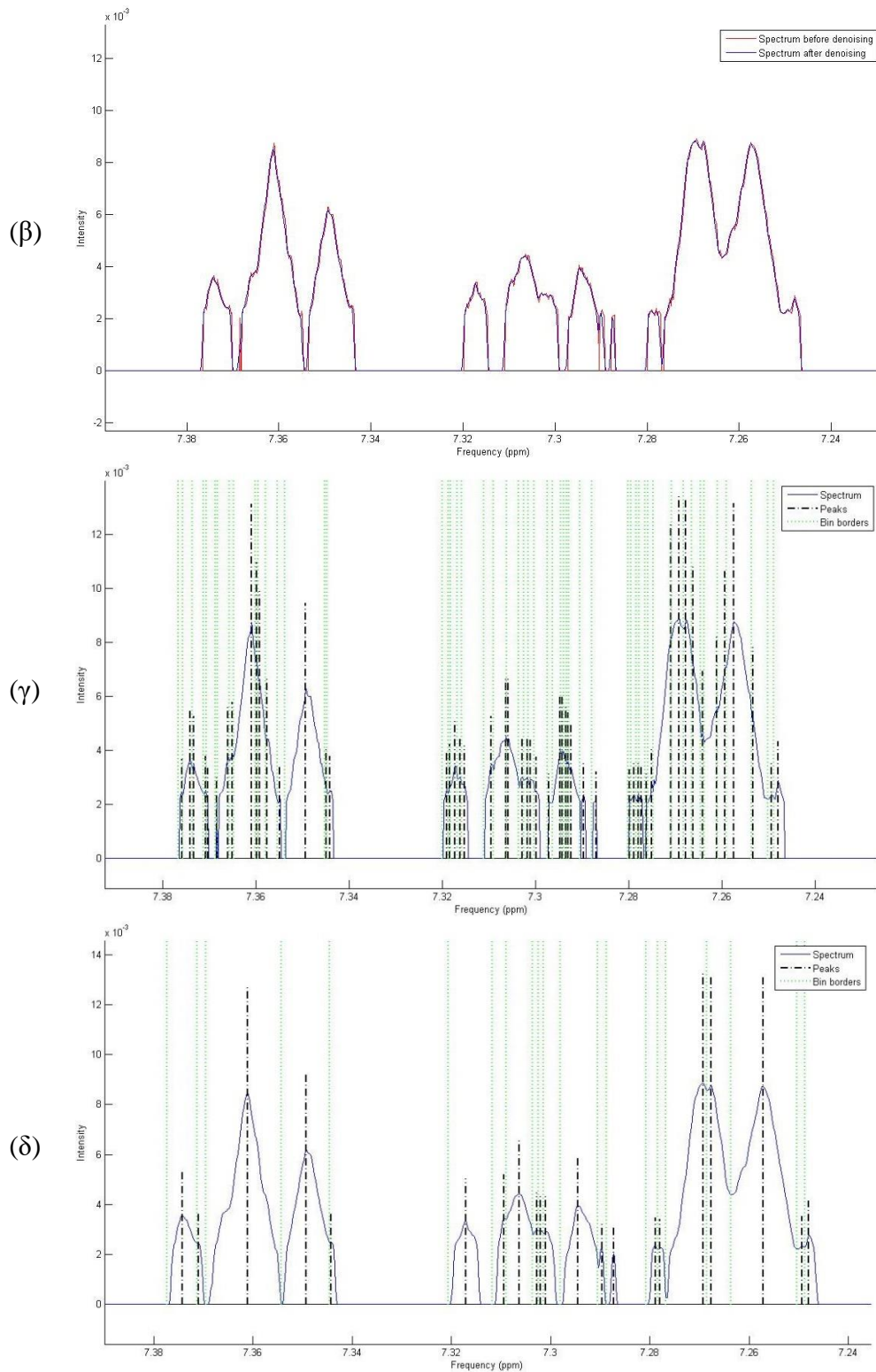
### 5.2.2.2. Περιστολή δεδομένων

Μετά την κατωφλίωση και αποθορύβωση, το φάσμα υπόκειται σε διαδικασία περιστολής δεδομένων μέσω binning, με στόχο να αναδειχθούν οι φασματικές κορυφές και ο όγκος δεδομένων του φάσματος να γίνει πιο διαχειρίσιμος. Μετά από συγκριτική εξέταση των δημοσιευμένων μεθόδων binning, επιλέχθηκε η μέθοδος Adaptive Intelligent (AI) binning (De Meyer et al., 2008; Anderson et al., 2008; Anderson et al., 2010; Davis et al., 2007; Sousa et al., 2013; Jacob et al., 2013). Αν και η μέθοδος αυτή τείνει να παράγει περιττά bins, επιλέχθηκε έναντι των άλλων μεθόδων binning γιατί εντοπίζει τις πραγματικές φασματικές κορυφές χωρίς να απαιτεί από το χρήστη τον ορισμό κάποιων τιμών παραμέτρων.

Στην Εικόνα 5.2 (γ) παρουσιάζεται το αποτέλεσμα της διαδικασίας περιστολής δεδομένων στο φάσμα που έχει υποστεί μόνο κατωφλίωση και όχι αποθορύβωση, με τις συχνότητες των ορίων των bins και των κορυφών να επισημαίνονται με κάθετες διακεκομμένες πράσινες και μαύρες γραμμές, αντίστοιχα. Στην Εικόνα 5.2 (δ) φαίνεται το αποτέλεσμα της διαδικασίας μείωσης δεδομένων στην ίδια περιοχή του φάσματος, το οποίο έχει υποστεί και κατωφλίωση και αποθορύβωση.







Εικόνα 5.2. Αποτέλεσμα της κατωφλίωσης, της αποθορύβωσης και της μείωσης δεδομένων στο φάσμα του δείγματος S25. (α) Κατωφλίωση (κατώφλι θορύβου=0.00203426). (β) Γκαουσιανή αποθορύβωση ( $\sigma_G=1.25$ ,  $N_G=4$ ) του κατωφλιωμένου φάσματος. (γ) Περιστολή δεδομένων στο κατωφλιωμένο φάσμα. Οι συχνότητες των ορίων των bins και των κορυφών επισημαίνονται με κάθετες διακεκομμένες πράσινες και μαύρες γραμμές, αντίστοιχα. (δ) Περιστολή δεδομένων στο κατωφλιωμένο και αποθορυβωμένο φάσμα.

### 5.2.2.3. Αναζήτηση μεταβολιτών

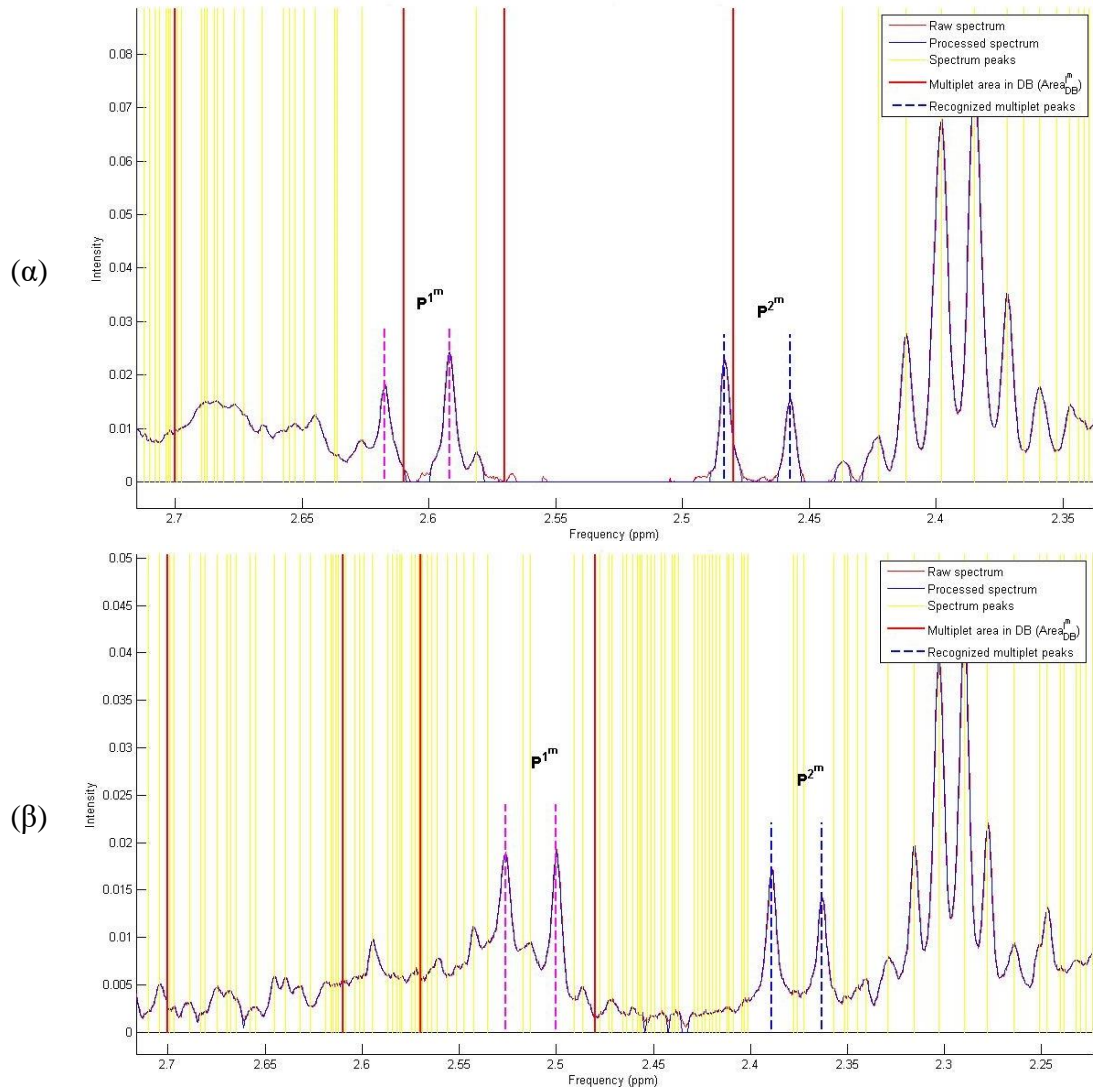
Εφ' όσον έχει οριστεί ένα σύνολο bins και κορυφών για το φάσμα εισόδου, η βιβλιοθήκη μεταβολιτών φορτώνεται και η αναζήτηση μεταβολιτών ξεκινά. Στο εξής, η μεταβλητή  $m = 1, \dots, M$  αναφέρεται σε ένα μεταβολίτη, όπου  $M$  είναι ο συνολικός αριθμός των υποψήφιων μεταβολιτών. Η μεταβλητή  $l^m = 1, \dots, L^m$  αναφέρεται σε μία απλή (singlet) ή πολλαπλή κορυφή (multiplet) ενός μεταβολίτη  $m$ , όπου  $L^m$  είναι ο συνολικός αριθμός των κορυφών του  $m$ .  $N = N^{l^m}$  είναι ο αριθμός των επί μέρους κορυφών μίας πολλαπλής κορυφής (multiplet)  $l^m$ , και  $N^m$  είναι το πλήθος των επί μέρους κορυφών ενός μεταβολίτη. Η περιοχή συχνοτήτων που ορίζεται στη βάση δεδομένων για μία κορυφή  $l^m$  συμβολίζεται ως  $area_{DB}^{l^m}$ . Ο Πίνακας Π.3 (Παράρτημα) περιέχει τη συνοπτική περιγραφή των πιο σημαντικών μεταβλητών και σταθερών που χρησιμοποιούνται στην περιγραφή του παρουσιαζόμενου μεθοδολογικού σχήματος.

### Screening μεταβολιτών

Επιγραμματικά, το αντικείμενο της διαδικασίας του screening μεταβολιτών είναι να εξετάζεται ένας αριθμός ( $K$ ) πιθανών συνόλων κορυφών και να υπολογίζεται η αντίστοιχη βαθμολογία του κάθε συνόλου, ξεχωριστά για κάθε απλή ή πολλαπλή κορυφή  $l^m$  ενός μεταβολίτη  $m$ , χωρίς να λαμβάνεται απόφαση για τις τελικές κορυφές που θα αποδοθούν στον μεταβολίτη.

Σε ένα φάσμα  $^1H$  NMR, οι κορυφές ενός μεταβολίτη μπορούν να υποστούν μεταβολές ως προς τη συχνότητά τους λόγω του pH του μίγματος ή τη συγκέντρωση του μεταβολίτη, και ως αποτέλεσμα μπορεί να βρίσκονται εκτός της περιοχής συχνοτήτων που ορίζεται στη βάση δεδομένων για αυτά ( $area_{DB}^{l^m}$ ). Ένα παράδειγμα αυτού του φαινομένου παρουσιάζεται στην Εικόνα 5.3, όπου βλέπουμε ότι οι κορυφές του Κιτρικού οξέος (Citric acid) σε δύο διαφορετικά δείγματα ορού (S25, S39) παρουσιάζουν διαφορές ως προς τις χημικές μετατοπίσεις τους.

Με δεδομένο το παραπάνω φαινόμενο, η διαδικασία του screening μεταβολιτών διαμορφώθηκε με το στόχο μία απλή ή πολλαπλή κορυφή  $l^m$  να αναζητείται χωριστά σε έναν αριθμό ( $K$ ) περιοχών ( $area_k^{l^m}$ ), οι οποίες έχουν το ίδιο εύρος με την περιοχή  $area_{DB}^{l^m}$ , αλλά διαφορετικό κέντρο συχνότητας (εξίσωση (5.3)). Η μεταβλητή  $offset^{a,K}$  συμβολίζει την ποσότητα κατά την οποία η αρχική περιοχή  $area_{DB}^{l^m}$  τροποποιείται (εξίσωση (5.4)). Η διαφορά συχνότητας (frequency difference) μεταξύ δύο διαδοχικών περιοχών είναι ο σταθερός αριθμός  $fd^a$  (εξίσωση (5.5)), που σημαίνει ότι δύο διαδοχικές περιοχές έχουν δεδομένη απόσταση μεταξύ τους. Το ολικό εύρος συχνοτήτων (frequency range) στο οποίο μία απλή ή πολλαπλή κορυφή αναζητείται συμβολίζεται ως  $fr^{a,K}$  (εξίσωση (5.6)). Η σταθερά  $width_\mu$  είναι το μέσο εύρος περιοχής κορυφής, υπολογισμένο από τους 850 μεταβολίτες που συμπεριλαμβάνονται στο αρχείο που περιγράφεται στην παράγραφο 5.2.1.2 ( $width_\mu = 0.07509$  ppm). Η αρχική προσέγγιση, που αποδείχθηκε αναποτελεσματική, ήταν να χρησιμοποιηθεί διαφορετική διαφορά συχνότητας  $fd^a$  για κάθε κορυφή  $l^m$ , χρησιμοποιώντας ένα ποσοστό της  $area_{DB}^{l^m}$  αντί της σταθεράς  $width_\mu$ , έτσι ώστε μία κορυφή με ευρύτερη  $area_{DB}^{l^m}$  να έχει μεγαλύτερη  $fd^a$  από μία κορυφή με μικρότερη  $area_{DB}^{l^m}$ .



Εικόνα 5.3. Μία φασματική περιοχή των δειγμάτων ορού S25 (α) και S39 (β), στην οποία οι διπλές κορυφές ( $P^{1^m}, P^{2^m}$ ) του Κιτρικού οξέος βρίσκονται εν μέρει έξω από τις αντίστοιχες περιοχές ( $area_{DB}^{1^m}, area_{DB}^{2^m}$ ) που ορίζονται στη βάση δεδομένων HMDB. Οι επί μέρους κορυφές των διπλών κορυφών ( $P^{1^m}, P^{2^m}$ ) επισημαίνονται με κάθετες διακεκομμένες γραμμές, ενώ οι περιοχές που ορίζονται στη βάση δεδομένων ( $area_{DB}^{1^m}, area_{DB}^{2^m}$ ) επισημαίνονται με κόκκινες κάθετες γραμμές.

$area_{DB}^{1^m} = 2.61 - 2.7 \text{ ppm}, area_{DB}^{2^m} = 2.48 - 2.57 \text{ ppm}$ . S25:  $P^{1^m} = \{2.59, 2.62\} \text{ ppm}, P^{2^m} = \{2.46, 2.49\} \text{ ppm}$ . S39:  $P^{1^m} = \{2.5, 2.53\} \text{ ppm}, P^{2^m} = \{2.36, 2.39\} \text{ ppm}$

$$area_k^{1^m} = area_{DB}^{1^m} + \left(-\frac{K}{2} + k\right) \times \alpha \times width_{\mu}, k = 1, 2, \dots, K \quad (5.3)$$

$$offset^{a,K} = \frac{K}{2} \times a \times width_{\mu} = \frac{K}{2} \times fd^a \quad (5.4)$$

$$\begin{aligned} fd^a &= \max(area_{k+1}^{1^m}) - \max(area_k^{1^m}) = \\ &= \min(area_{k+1}^{1^m}) - \min(area_k^{1^m}) = a \times width_{\mu} \end{aligned} \quad (5.5)$$

$$\begin{aligned} fr^{a,K} &= \max(area_k^{1^m}) - \min(area_1^{1^m}) = \\ &= \max\left(area_{DB}^{1^m} + \left(\frac{K}{2}\right) \times \alpha \times width_{\mu}\right) - \min\left(area_{DB}^{1^m} + \left(-\frac{K}{2} + 1\right) \times \alpha \times width_{\mu}\right) = \end{aligned}$$

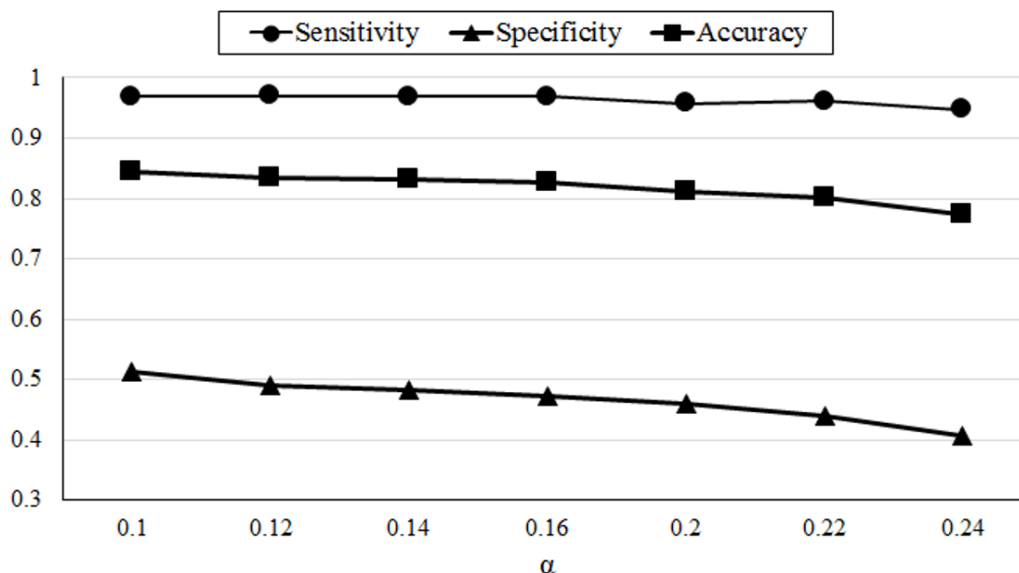
$$\begin{aligned}
&= \max(\text{area}_{\text{DB}}^{1^m}) - \min(\text{area}_{\text{DB}}^{1^m}) + (K - 1) \times \alpha \times \text{width}_\mu = \\
&= \max(\text{area}_{\text{DB}}^{1^m}) - \min(\text{area}_{\text{DB}}^{1^m}) + 2 \times |\text{offset}^{a,K}| - \text{fd}^a \quad (5.6)
\end{aligned}$$

Η μεταβλητή  $\text{offset}^{a,K}$  αντικατοπτρίζεται στο ολικό εύρος συχνοτήτων  $\text{fr}^{a,K}$  που καλύπτεται για κάθε μεταβολίτη (εξίσωση (5.6)). Οι τιμές  $\text{offset}^{a,K}$ ,  $\text{fr}^{a,K}$  χρειάζεται να είναι αρκετά υψηλές ώστε ο αλγόριθμος να αναγνωρίσει μεταβολίτες σε ποικίλα δείγματα με διάφορες χημικές μετατοπίσεις. Η μείωση της τιμής  $\text{offset}^{a,K}$  αυξάνει τον κίνδυνο να χαθούν κορυφές μεταβολιτών που βρίσκονται εκτός της  $\text{fr}^{a,K}$ . Από την άλλη πλευρά, η αύξηση του  $\text{offset}^{a,K}$  εντείνει τον κίνδυνο να αποδοθεί σε έναν μεταβολίτη ένα λάθος σύνολο κορυφών που έχει παρόμοια μορφολογικά χαρακτηριστικά (π.χ. παρόμοιες τιμές σταθεράς σύζευξης και αναλογίες υψών) με το σωστό σύνολο κορυφών, αφού σε ένα φάσμα διαφορετικά multiplets με παρόμοια χαρακτηριστικά μπορούν να βρεθούν σε σχετικά κοντινές συχνότητες. Για τους μεταβολίτες που περιέχουν μόνο απλές κορυφές χρησιμοποιήθηκε μία επιπρόσθετη μεταβλητή  $\text{offset}^{a,K'} < \text{offset}^{a,K}$  με στόχο να μειωθούν τα ψευδή θετικά αποτελέσματα (false positives), αφού οι απλές κορυφές συνήθως έχουν μικρότερες μεταβολές ως προς τη χημική μετατόπισή τους σε σύγκριση με τις πολλαπλές κορυφές, και δεν διαθέτουν χαρακτηριστικά όπως σταθερές σύζευξης, που θα βοηθούσαν στον εντοπισμό τους. Ένας αριθμός τιμών  $\text{offset}^{a,K}$ ,  $\text{offset}^{a,K'}$  δοκιμάστηκαν στο εύρος [0.09 0.19]. Μετά από πειραματισμό, οι βέλτιστες τιμές για τις μεταβλητές  $\text{offset}^{a,K}$ ,  $\text{offset}^{a,K'}$  βρέθηκαν να είναι περίπου 0.17 και 0.105, αντίστοιχα (Πίνακας Π.4).

Η παράμετρος  $\alpha$  ελέγχει τη διαφορά συχνότητας  $\text{fd}^a$  μεταξύ δύο οποιοδήποτε διαδοχικών περιοχών  $\text{area}_k^{1^m}$ ,  $\text{area}_{k+1}^{1^m}$  μίας κορυφής  $1^m$ . Η αύξηση (ή μείωση) του  $\alpha$  επιφέρει αύξηση (ή μείωση) στις τιμές των μεταβλητών  $\text{fd}^a$ ,  $\text{offset}^{a,K}$ ,  $\text{fr}^{a,K}$  (εξισώσεις (5.4-5.6)). Όσο πιο μικρή είναι η τιμή του  $\alpha$ , τόσο πιο ενδελεχές είναι το στάδιο του screening μεταβολιτών, κι επομένως τόσο πιο απίθανο είναι να παρακαμφθεί μία φασματική κορυφή κατά την αναζήτηση μίας κορυφής. Παρ' όλα αυτά, για να καλυφθεί ένα δεδομένο εύρος συχνοτήτων  $\text{fr}^{a,K}$ , η μείωση του  $\alpha$  σημαίνει ότι απαιτείται μία μεγαλύτερη τιμή  $K$ , και κατ' επέκταση μεγαλύτερος χρόνος εκτέλεσης. Επιπλέον, όσο το  $\alpha$  μειώνεται, αυξάνεται η πιθανότητα να προκύψουν τα ίδια αποτελέσματα  $P_k^{1^m}$ ,  $\text{Score}_k^{1^m}$  για διαδοχικές τιμές  $k$ . Από την άλλη πλευρά, για μία μεγαλύτερη τιμή  $\alpha$ , ένα δεδομένο εύρος συχνοτήτων  $\text{fr}^{a,K}$  μπορεί να καλυφθεί για μικρότερη τιμή  $K$ , και κατ' επέκταση μικρότερο υπολογιστικό χρόνο. Παρ' όλα αυτά, η αύξηση του  $\alpha$  πάνω από κάποιο όριο (π.χ.  $\alpha > 0.16$ ) έχει ως αποτέλεσμα να αυξάνεται ο κίνδυνος παράλειψης φασματικών κορυφών κατά την αναζήτηση μίας κορυφής.

Στα πλαίσια της αναζήτησης της βέλτιστης τιμής του  $\alpha$ , ένα σύνολο τιμών δοκιμάστηκε στο εύρος τιμών [0.10 0.24] με βήμα 0.02. Στην Εικόνα 5.4 απεικονίζονται η μέση ευαισθησία, ειδικότητα και ακρίβεια που επιτεύχθηκαν στα δείγματα S1-S42 σε συνάρτηση με την τιμή της παραμέτρου  $\alpha$ . Σύμφωνα με τα πειραματικά αποτελέσματα, μία τιμή  $\alpha \geq 0.2$  δεν ενδείκνυται, αφού μπορεί να παράγει σημαντικά σφάλματα εξ' αιτίας της παράλειψης φασματικών κορυφών στο στάδιο του screening μεταβολιτών. Από την άλλη πλευρά, μία τιμή  $\alpha < 0.1$  είναι μάλλον χαμηλή, αφού απαιτεί περισσότερο υπολογιστικό χρόνο για να ολοκληρώσει την αναζήτηση μεταβολιτών χωρίς κέρδος ως προς την ακρίβεια σε ένα τυχαίο φάσμα. Επιπλέον, μπορεί να παρατηρηθεί ότι για  $0.10 \leq \alpha \leq 0.14$  η απόδοση

της προτεινόμενης μεθοδολογίας δεν είχε σημαντικές διακυμάνσεις. Η υψηλότερη μέση ακρίβεια (81.89%) σημειώθηκε για  $\alpha = 0.1$ . Παρ' όλα αυτά, το 0.12 επιλέχθηκε ως η βέλτιστη τιμή  $\alpha$ , αφού αντιστοιχεί σε μικρότερο υπολογιστικό χρόνο για μία ελαφρώς χαμηλότερη μέση ακρίβεια (80.75%).



Εικόνα 5.4. Μέση τιμή ευαισθησίας, ειδικότητας και ακρίβειας της προτεινόμενης μεθοδολογίας στα δείγματα S1-S42 σε συνάρτηση με την τιμή της παραμέτρου  $\alpha$ .

Η παράμετρος  $K$  αντιπροσωπεύει τον αριθμό των διαφορετικών περιοχών  $area_k^{1^m}$ ,  $k = 1, \dots, K$  στις οποίες μία κορυφή  $1^m$  ενός μεταβολίτη  $m$  αναζητείται. Με δεδομένη μία τιμή του  $\alpha$ , η αύξηση (ή μείωση) του  $K$  αντικατοπτρίζεται σε μία αύξηση (ή μείωση) των τιμών  $offset^{a,K}$  και  $fr^{a,K}$ , όπως και σε μία αύξηση (ή μείωση) του υπολογιστικού χρόνου (εξισώσεις (5.4), (5.6)). Αφού απαιτείται η συνθήκη  $area_k^{1^m} = area_{DB}^{1^m}$  να είναι αληθής για μία ακέραια τιμή  $k$  στο διάστημα  $[1, K]$  για να υπάρχει συμβατότητα με την HMDB, το  $K$  τέθηκε μόνο σε ζυγούς αριθμούς. Στον Πίνακα Π.4 περιέχονται αντιπροσωπευτικοί συνδυασμοί τιμών των παραμέτρων ( $\alpha, K, K'$ ) που δοκιμάστηκαν, καθώς και οι αντίστοιχες τιμές των μεταβλητών  $fd^a, offset^{a,K}, offset^{a,K'}$ .

Με δεδομένο το ότι οι παράμετροι κατωφλίωσης και αποθορύβωσης (κατώφλι θορύβου,  $\sigma_G$ ) τίθενται αυτόματα (Παράγραφος 5.2.2.1), καθώς και το γεγονός ότι οι βέλτιστες τιμές των  $\alpha, offset^{a,K}, offset^{a,K'}$  έχουν προσδιοριστεί, ο χρήστης δεν καλείται να πάρει αποφάσεις σχετικά με τις παραμέτρους μεθοδολογίας. Επομένως, η προτεινόμενη μεθοδολογία είναι πλήρως αυτοματοποιημένη.

Συνοψίζοντας το στάδιο του screening μεταβολιτών, σε κάθε περιοχή συχνοτήτων  $area_k^{1^m}$ , ένα πλήθος συνόλων κορυφών εξετάζονται ως προς την καταλληλότητά τους για κάθε κορυφή  $1^m$  και βαθμολογούνται σύμφωνα με τις εξισώσεις (5.12)-(5.26). Το σύνολο κορυφών  $P_k^{1^m}$  με τη βέλτιστη βαθμολογία  $Score_k^{1^m}$  επιλέγεται και αποθηκεύεται για κάθε συνδυασμό τιμών  $1^m, k$ . Επομένως, το αποτέλεσμα της διαδικασίας screening μεταβολιτών για έναν μεταβολίτη  $m$  είναι ένα σύνολο κορυφών  $P_k^{1^m}$  (εξίσωση (5.7)) και το αντίστοιχο

$Score_k^{lm}$  (εξίσωση (5.8)) για κάθε απλή ή πολλαπλή κορυφή  $l^m$  και κάθε περιοχή  $area_k^{lm}$ ,  $k = 1, \dots, K$ .

$$P_k^{lm} = \{p_{k,1}^{lm}, \dots, p_{k,n}^{lm}, \dots, p_{k,N}^{lm}\} \quad (5.7)$$

$$Score_k^{lm} = \begin{cases} Score_{S_k}^{lm}, & \text{αν το } l^m \text{ είναι απλή κορυφή (εξίσωση 5.12)} \\ Score_{A_k}^{lm}, & \text{αν το } l^m \text{ είναι πολλαπλή κορυφή } A \text{ τάξης (εξίσωση 5.17)} \\ Score_{B_k}^{lm}, & \text{αν το } l^m \text{ είναι πολλαπλή κορυφή } B \text{ τάξης (εξίσωση 5.20)} \\ Score_{C_k}^{lm}, & \text{αν το } l^m \text{ είναι πολλαπλή κορυφή } C \text{ τάξης (εξίσωση 5.20)} \\ Score_{D_k}^{lm}, & \text{αν το } l^m \text{ είναι πολλαπλή κορυφή } D \text{ τάξης (εξίσωση 5.20)} \\ Score_{M_k}^{lm}, & \text{αν το } l^m \text{ είναι πολλαπλή κορυφή χωρίς κανόνες (εξίσωση 5.24)} \end{cases} \quad (5.8)$$

$$l = l^m = 1, \dots, L^m, n = 1, \dots, N^{l^m}, k = 1, \dots, K$$

### Απόρριψη μικρών κορυφών

Η αποθορύβωση του φάσματος είναι ένα χρήσιμο βήμα προεπεξεργασίας αλλά δεν εξαλείφει όλες τις ανεπιθύμητες κορυφές θορύβου. Επίσης, η ισχυρή αποθορύβωση ενός φάσματος μπορεί να εξαφανίσει σημαντικές φασματικές κορυφές που αντιστοιχούν σε μεταβολίτες. Επομένως, η συνάρτηση  $V_{bp}$  εισήχθηκε για να λειτουργήσει ως μέτρο διαφοροποίησης μεταξύ των ευδιάκριτων και δυσδιάκριτων κορυφών και χωρίς να βασίζεται στην αποθορύβωση, και χρησιμοποιείται για την απόρριψη συνόλων κορυφών με τιμή  $V_{bp}$  χαμηλότερη ενός κατωφλίου.

Η συνάρτηση  $V_{bp}$  αντιπροσωπεύει την ποσοστιαία διαφορά ανάμεσα στη μέση ένταση των κορυφών ενός multiplet και στη μέση ένταση των αντίστοιχων ορίων bin. Στην εξίσωση (5.9),  $h_n$  είναι το ύψος της κορυφής  $p_n$ , και  $h_{n\ b1}, h_{n\ b2}$ , είναι τα ύψη των αντίστοιχων ορίων bin. Το σύμβολο  $\mu = \mu(\bar{x})$  συμβολίζει τη μέση τιμή ενός συνόλου τιμών  $\bar{x}$ . Η ποσοστιαία διαφορά (percent difference, %diff) και το ποσοστιαίο σφάλμα (percent error, %error) ορίζονται στις εξισώσεις (5.10), (5.11) αντίστοιχα, όπου τα  $E, E_1, E_2$  συμβολίζουν πειραματικές τιμές, και το  $T$  συμβολίζει μία αληθή τιμή (Experimental errors and uncertainty, 2002).

$$V_{bp} = \%diff(\mu([h_1, \dots, h_n, \dots, h_N]), \mu([h_{1\ b1}, h_{1\ b2}, \dots, h_{n\ b1}, h_{n\ b2}, \dots, h_{N\ b1}, h_{N\ b2}])), \quad n = 1, \dots, N \quad (5.9)$$

$$\%diff(E_1, E_2) = \left| \frac{E_1 - E_2}{\frac{1}{2}(E_1 + E_2)} \right| \times 100 \quad (5.10)$$

$$\%error(T, E) = \left| \frac{T - E}{T} \right| \times 100 \quad (5.11)$$

Για να δοθεί προτεραιότητα στα υποψήφια σύνολα κορυφών με υψηλότερη τιμή  $V_{bp}$ , η αναζήτηση για κάθε μεταβολίτη ενσωματώνεται σε ένα επαναληπτικό βρόχο στο οποίο το κατώφλι  $V_{bp}$  έχει προκαθορισμένη αρχική και τελική τιμή. Οι μεταβολίτες με προτεραιότητα έχουν αρχική τιμή  $V_{bp}$  κατωφλίου 1 και τελική τιμή 0.1, ενώ οι υπόλοιποι έχουν αρχική τιμή 0.5 και τελική τιμή 0.1. Ο βρόχος αναζήτησης ενός μεταβολίτη τερματίζεται όταν βρεθεί ο

μεταβολίτης ή όταν ολοκληρωθεί ο μέγιστος αριθμός επαναλήψεων (5) και το κατώφλι  $V_{bp}$  πάρει την ελάχιστη τιμή του.

### **Βαθμολόγηση των κορυφών**

Η διαδικασία εύρεσης του βέλτιστου συνόλου κορυφών για μία απλή ή πολλαπλή κορυφή  $I^m$  σε μία περιοχή συχνοτήτων  $area^{I^m}$  μπορεί να περιγραφεί ως (i) η εφαρμογή μίας καινοτόμας συνάρτησης βαθμολόγησης  $Score^{I^m}$  σε κάθε υποψήφιο σύνολο κορυφών  $P^{I^m}$  για να εκτιμηθεί η καταλληλότητά του για το  $I^m$ , και (ii) η επιλογή του υποψήφιου συνόλου κορυφών με το μικρότερο  $Score^{I^m}$  ως το καταλληλότερο για το  $I^m$  στην  $area^{I^m}$ .

Η συνάρτηση βαθμολόγησης που εφαρμόζεται σε ένα υποψήφιο σύνολο φασματικών κορυφών  $P^{I^m}$  για να εκτιμηθεί η καταλληλότητά του για την απλή ή πολλαπλή κορυφή  $I^m$  ενός μεταβολίτη διαφέρει ανάλογα με την κατηγορία της κορυφής  $I^m$ , με τις διαφορετικές κατηγορίες να είναι οι απλές κορυφές (singlets), οι πολλαπλές κορυφές (multiplets) πρώτης (A) τάξης, οι πολλαπλές κορυφές δεύτερης (B) ή υψηλότερης (C, D, ... συμβολιζόμενα ως X) τάξης και οι πολλαπλές κορυφές που δεν υπακούουν σε συγκεκριμένους κανόνες. Οι συναρτήσεις βαθμολόγησης για τις πολλαπλές κορυφές πρώτης ή μεγαλύτερης τάξης έχουν σχεδιαστεί ώστε να συμπεριλαμβάνουν χαρακτηριστικά όπως οι τιμές σταθεράς σύζευξης (J coupling) και οι αναλογίες υψών που ορίζονται στην HMDB για μία δεδομένη πολλαπλή κορυφή  $I^m$ . Αντίθετα, με δεδομένο ότι οι απλές κορυφές δε συσχετίζονται με τιμές σταθεράς σύζευξης ή λόγους υψών, η συνάρτηση βαθμολόγησης για αυτές έχει σχεδιαστεί ώστε να λαμβάνει υπόψιν τις οριζόντιες αποστάσεις και τις διαφορές υψών μίας υποψήφιας φασματικής κορυφής με τις γειτονικές φασματικές κορυφές. Η συνάρτηση βαθμολόγησης για μία πολλαπλή κορυφή χωρίς κανόνες λαμβάνει υπόψιν τις αναλογίες υψών που ορίζονται στην HMDB για αυτό, καθώς και τη μονοτονία υψών του υποψήφιου συνόλου κορυφών, που είναι τα πιο εμφανή χαρακτηριστικά αυτής της κατηγορίας. Στη συνέχεια παρουσιάζονται πιο αναλυτικά οι συναρτήσεις βαθμολόγησης για κάθε κατηγορία απλής ή πολλαπλής κορυφής. Παραδείγματα βαθμολόγησης για κάθε κατηγορία κορυφής παρουσιάζονται στην Εικόνα 5.5.

### **Απλές κορυφές (Singlets)**

Οι απλές κορυφές είναι μοναδικές κορυφές που δε συσχετίζονται με τιμές σταθεράς σύζευξης ή λόγους υψών. Επομένως, μία φασματική κορυφή που είναι υποψήφια για μία απλή κορυφή  $I^m$  αποτιμάται με βάση τις οριζόντιες αποστάσεις και τις διαφορές υψών με τις γειτονικές φασματικές κορυφές. Στις παρακάτω εξισώσεις, το σύνολο κορυφών  $P^{I^m} = \{p_1, \dots, p_n, \dots, p_N\}$  αποτελείται από τις φασματικές κορυφές μίας περιοχής  $area_k^{I^m}$ , κάθε μία από τις οποίες είναι υποψήφια για την απλή κορυφή  $I^m$ . Οι μεταβλητές  $f_n$  και  $h_n$  αντιστοιχούν στη συχνότητα και την ένταση μίας κορυφής  $p_n$  υπό εξέταση. Ένα  $Score_S$  υπολογίζεται για κάθε κορυφή  $p_n$ , και η κορυφή με το ελάχιστο  $Score_S$  θεωρείται το καταλληλότερο για την απλή κορυφή υπό εξέταση στο σύνολο  $P^{I^m}$  (εξίσωση (5.12)). Η Εικόνα 5.5 (α) περιέχει ένα παράδειγμα βαθμολόγησης μίας φασματικής κορυφής που είναι υποψήφια για μία απλή κορυφή.

$$Score_S = -w_{d1} \times Score_{d1} - w_{d2} \times Score_{d2} - w_{h1} \times Score_{h1} - w_{h2} \times Score_{h2} \quad (5.12)$$

$$Score_{d1} = \%diff(|f_n - f_{n-1}|, |f_n - f_{n+1}|), n = 2, \dots, N - 1 \quad (5.13)$$

$$Score_{d2} = \mu(|f_n - f_{n-1}|, |f_n - f_{n+1}|), n = 2, \dots, N - 1 \quad (5.14)$$

$$\text{Score}_{h1} = \%diff(h_{n-1}, h_{n+1}), n = 2, \dots, N - 1 \quad (5.15)$$

$$\text{Score}_{h2} = \mu \left( \frac{h_n - h_{n-1}}{\mu(h_n, h_{n-1})}, \frac{h_n - h_{n+1}}{\mu(h_n, h_{n+1})} \right), n = 2, \dots, N - 1 \quad (5.16)$$

### Πολλαπλές κορυφές Α τάξης

Μία πολλαπλή κορυφή  $I^m$  πρώτης (Α) τάξης αποτελείται από  $N_A = N_A^{I^m}$  κορυφές. Η τιμή σταθεράς σύζευξης  $j_A^{I^m} = j_A$  είναι η οριζόντια απόσταση μεταξύ των διαδοχικών κορυφών του που ορίζεται στη βάση δεδομένων. Οι ιδανικοί λόγοι υψών  $H_{DB}^{I^m} = [h_{DB1}, h_{DB2}, \dots, h_{DBN}]$  για το  $I^m$  συνήθως ορίζονται στη βάση δεδομένων, αλλιώς τους ανατίθενται οι τιμές της  $n_A^{th}$  γραμμής του τριγώνου Pascal. Για παράδειγμα, στον Πίνακα Π.2 ισχύουν οι ακόλουθες σχέσεις για την τριπλή κορυφή Α τάξης (triplet, t,  $I^m = 2$ ) του μεταβολίτη L-Μεθειονίνη:  $N^{I^m} = 3, \text{area}_{DB}^{I^m} = 2.593 - 2.673 \text{ ppm}, j_A^{I^m} = 7.587 \text{ Hz}, H_{DB}^{I^m} = [0.60280 \ 1.02610 \ 0.52760]$ .

Ένα υποψήφιο σύνολο κορυφών  $P^{I^m} = \{p_1, \dots, p_{n_A}, \dots, p_{N_A}\}$  για μία πολλαπλή κορυφή  $I^m$  Α τάξης αποτελείται από  $N_A = N_A^{I^m}$  κορυφές. Η συχνότητα, η ένταση και ο ιδανικός λόγος ύψους που αντιστοιχούν στην κορυφή  $p_{n_A}$  είναι  $f_{n_A}, h_{n_A}, h_{DB_{n_A}}$  αντίστοιχα. Το  $P^{I^m}$  αποτιμάται σύμφωνα με την εξίσωση (5.17). Το  $\text{Score}_{j_A}$  εκτιμά πόσο καλά το  $P^{I^m}$  είναι συμβατό με την τιμή σταθεράς σύζευξης  $j_A$ , που είναι η ιδανική οριζόντια απόσταση μεταξύ των διαδοχικών κορυφών  $p_{n_A}, p_{n_A+1}$ , του  $I^m$ ,  $n_A = 1, \dots, N_A - 1$  (εξίσωση (5.18)). Το  $\text{Score}_{h_A}$  αντιπροσωπεύει πόσο καλά το σύνολο κορυφών  $P^{I^m}$  συμφωνεί με τους ιδανικούς λόγους υψών  $h_{DB_{n_A}}, n_A = 1, \dots, N_A$  για το  $I^m$  (εξίσωση (5.19)). Ανάμεσα σε όλα τα υποψήφια σύνολα κορυφών για μία πολλαπλή κορυφή  $I^m$ , το σύνολο κορυφών με το μικρότερο  $\text{Score}_A$  θεωρείται αυτό με τη βέλτιστη εφαρμογή. Η Εικόνα 5.5 (β) περιέχει ένα παράδειγμα βαθμολόγησης ενός συνόλου φασματικών κορυφών που είναι υποψήφιο για μία διπλή κορυφή Α τάξης (doublet).

$$\text{Score}_A = w_{j_A} \times \text{Score}_{j_A} + w_{h_A} \times \text{Score}_{h_A} \quad (5.17)$$

$$\text{Score}_{j_A} = \mu(\%error(|f_{n_A} - f_{n_A+1}|, j_A)), n_A = 1, \dots, N_A - 1 \quad (5.18)$$

$$\text{Score}_{h_A} = \mu \left( \begin{array}{l} \%error\left(\frac{h_{n_A}}{h_1}, \frac{h_{DB_{n_A}}}{h_{DB_1}}\right), n_A = 2, \dots, N_A \\ \%error\left(\frac{h_{n_A}}{h_{n_A-1}}, \frac{h_{DB_{n_A}}}{h_{DB_{n_A-1}}}\right), n_A = 3, \dots, N_A \\ \%error\left(\frac{h_{n_A}}{h_{n_A-2}}, \frac{h_{DB_{n_A}}}{h_{DB_{n_A-2}}}\right), n_A = 4, \dots, N_A \end{array} \right) \quad (5.19)$$

Οι συναρτήσεις βαθμολόγησης των πολλαπλών κορυφών είναι ένας μηχανισμός όχι μόνο για τη αποτίμηση ενός υποψήφιου συνόλου κορυφών, αλλά επίσης για την απόρριψη των ακατάλληλων συνόλων κορυφών χρησιμοποιώντας κατάλληλα κατώφλια. Για παράδειγμα, μία υποψήφια διπλή κορυφή (doublet) με  $\text{Score}_{j_A} = 0.5$  έχει μία τιμή σταθεράς σύζευξης που είναι 50% διαφορετική από την ιδανική τιμή που ορίζεται στη βάση δεδομένων. Λαμβάνοντας υπόψιν ότι οι τιμές σταθεράς σύζευξης δεν επηρεάζονται από διακυμάνσεις του pH, αυτή η απόκλιση είναι αρκετά μεγάλη για μία διπλή κορυφή, κι επομένως το



υποψήφιο σύνολο κορυφών είναι μάλλον ακατάλληλο για αυτήν. Οι τιμές κατωφλίου που επιλέχθηκαν για το  $\text{Score}_{j_A}$  είναι 0.25 για διπλές κορυφές και 0.3 για όλες τις πολλαπλές κορυφές Α τάξης με περισσότερες από 2 επιμέρους κορυφές.

### Πολλαπλές κορυφές X τάξης

Μία πολλαπλή κορυφή X τάξης  $I^m$  (όπου το X μπορεί να είναι B, C, D, κλπ.) περιέχει  $N_X = N_X^{I^m}$  πολλαπλές κορυφές X-1 τάξης, κάθε ένα από τα οποία περιέχει  $N_{X-1}$  πολλαπλές κορυφές X-2 τάξης με  $N_{X-2}$  κορυφές το κάθε ένα. Επομένως, μία πολλαπλή κορυφή X τάξης περιέχει  $N = N_X \times N_{X-1} \times N_{X-2}$  συνολικές κορυφές. Μία πολλαπλή κορυφή X τάξης μπορεί να έχει x τιμές σταθεράς σύζευξης ( $j^{I^m} = [j_A^{I^m} j_B^{I^m} \dots j_X^{I^m}] = [j_A j_B \dots j_X]$ ), όπου  $j_X = j_X^{I^m}$  είναι η οριζόντια απόσταση μεταξύ των διαδοχικών πολλαπλών κορυφών X-1 τάξης σύμφωνα με τη βάση δεδομένων. Για παράδειγμα, στον Πίνακα Π.2 ισχύουν οι ακόλουθες σχέσεις για την πολλαπλή κορυφή B τάξης (dd,  $I^m = 3$ ) της L-Μεθειονίνης:  $N = N_A \times N_B = 2 \times 2$ ,  $\text{area}_{DB} = 3.818 - 3.888 \text{ ppm}$ ,  $j_A = 5.382 \text{ Hz}$ ,  $j_B = 7.1 \text{ Hz}$ ,  $H_{DB} = [0.2772 \ 0.3262 \ 0.3290 \ 0.2721]$ .

Ένα υποψήφιο σύνολο κορυφών  $P^{I^m} = \{p_1, \dots, p_n, \dots, p_N\} = \{[p_{1,1}, \dots, p_{1,N_{X-1}}], \dots, [p_{N_X,1}, \dots, p_{N_X,N_{X-1}}]\} = \{P_1, \dots, P_{N_X}, \dots, P_{N_X}\}$ , όπου  $N = N_{X-1} \times N_X$ , για μία πολλαπλή κορυφή X τάξης  $I^m$ , περιέχει μικρότερα σύνολα κορυφών  $P_{n_X} = [p_{n_X,1}, \dots, p_{n_X,N_{X-1}}]$ , που αντιστοιχούν στις πολλαπλές κορυφές X-1 τάξης.  $F_{n_X}$ ,  $H_{n_X}$  είναι το σύνολο συχνοτήτων και υψών που αντιστοιχούν στο  $P_{n_X}$ , αντίστοιχα.

Το  $P^{I^m}$  βαθμολογείται σύμφωνα με την εξίσωση (5.20). Το σύνολο κορυφών με το μικρότερο  $\text{Score}_X$  θεωρείται αυτό με την καλύτερη εφαρμογή για το  $I^m$ . Το  $\text{Score}_{m,X-1}$  προκύπτει από τα scores των  $N_X$  μεμονωμένων πολλαπλών κορυφών X-1 τάξης (εξίσωση (5.21)). Το  $\text{Score}_{j_X}$  εκτιμά πόσο καλά το υποψήφιο σύνολο κορυφών συμφωνεί με τη σταθερά σύζευξης  $j_X$  (εξίσωση (5.22)). Το  $\text{Score}_{h_X}$  εκτιμά πόσο καλά το υποψήφιο σύνολο κορυφών συμφωνεί με τους ιδανικούς λόγους υψών  $h_{DB_{n_X}}$ ,  $n_X = 1, \dots, N_X$  (εξίσωση (5.23)). Για το  $\text{Score}_{j_X}$  επιλέχθηκε η τιμή κατωφλίου 0.3. Η Εικόνα 5.5 (γ) περιέχει ένα παράδειγμα βαθμολόγησης ενός συνόλου φασματικών κορυφών που είναι υποψήφιο για μία πολλαπλή κορυφή B τάξης, συγκεκριμένα ένα double doublet.

$$\text{Score}_X = w_{m_{X-1}} \times \text{Score}_{m_{X-1}} + w_{j_X} \times \text{Score}_{j_X} + w_{h_X} \times \text{Score}_{h_X} \quad (5.20)$$

$$\text{Score}_{m_{X-1}} = \mu \left( \text{Scores}_{X-1_{n_X}} \right) \quad (5.21)$$

$$\text{Score}_{j_X} = \mu \left( \%error(|\mu(F_{n_X}) - \mu(F_{n_X+1})|, j_X) \right) \quad (5.22)$$

$$\text{Score}_{h_X} = \mu \left( \left[ \begin{array}{l} \%error\left(\frac{h_{n_X}}{h_1}, \frac{h_{DB_{n_X}}}{h_{DB_1}}\right), n_X = 2, \dots, N_X \\ \%error\left(\frac{h_{n_X}}{h_{n_X-1}}, \frac{h_{DB_{n_X}}}{h_{DB_{n_X-1}}}\right), n_X = 3, \dots, N_X \\ \%error\left(\frac{h_{n_X}}{h_{n_X-2}}, \frac{h_{DB_{n_X}}}{h_{DB_{n_X-2}}}\right), n_X = 4, \dots, N_X \end{array} \right] \right) \quad (5.23)$$

$$n_X = 1, 2, \dots, N_X - 1$$

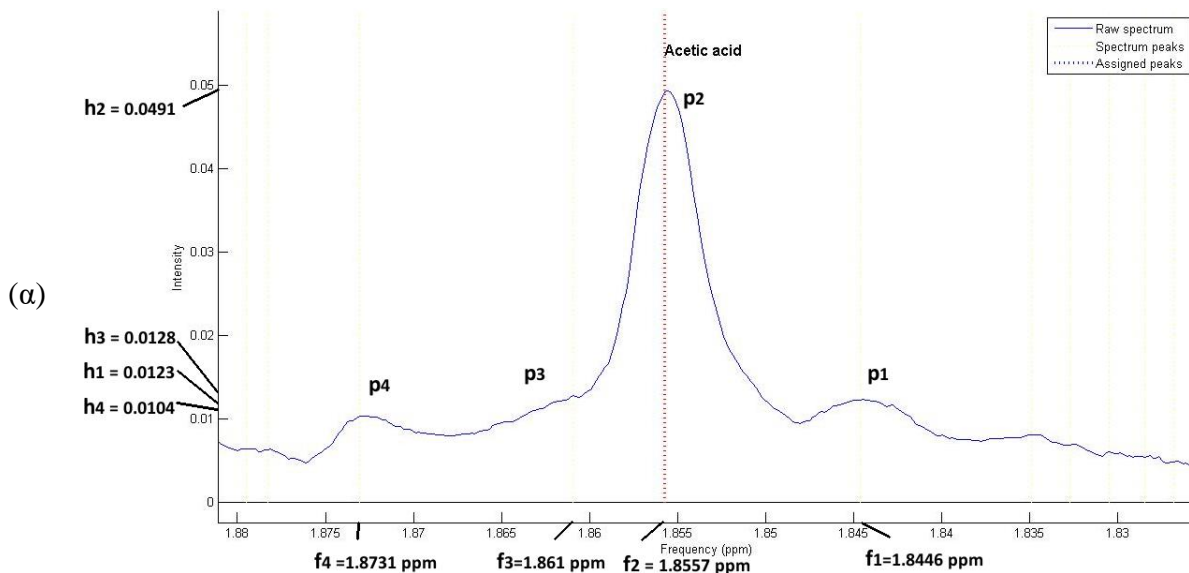
## Πολλαπλές κορυφές χωρίς κανόνες

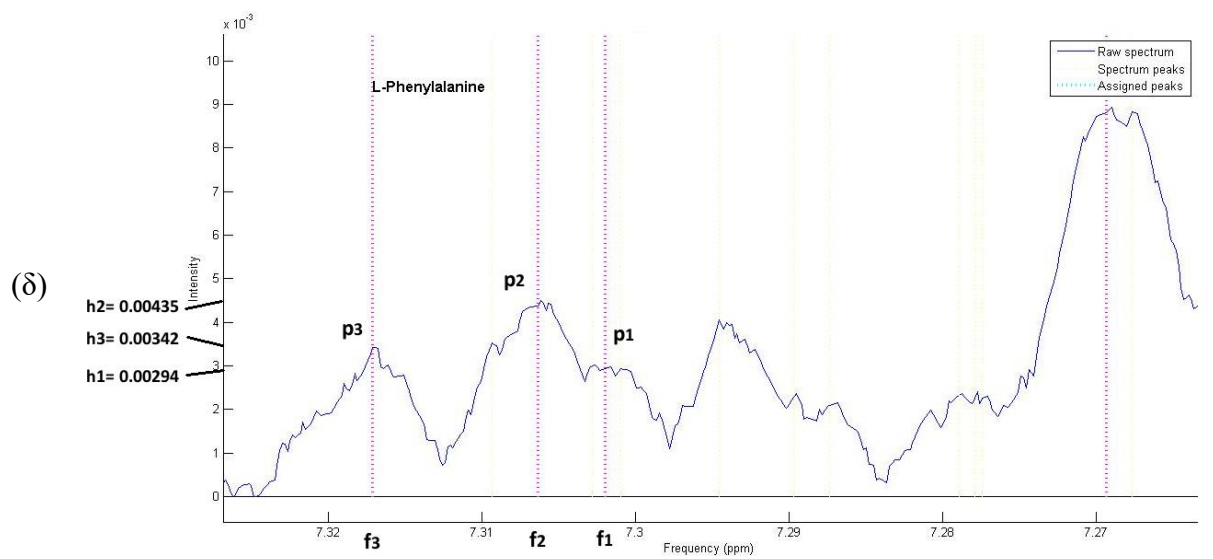
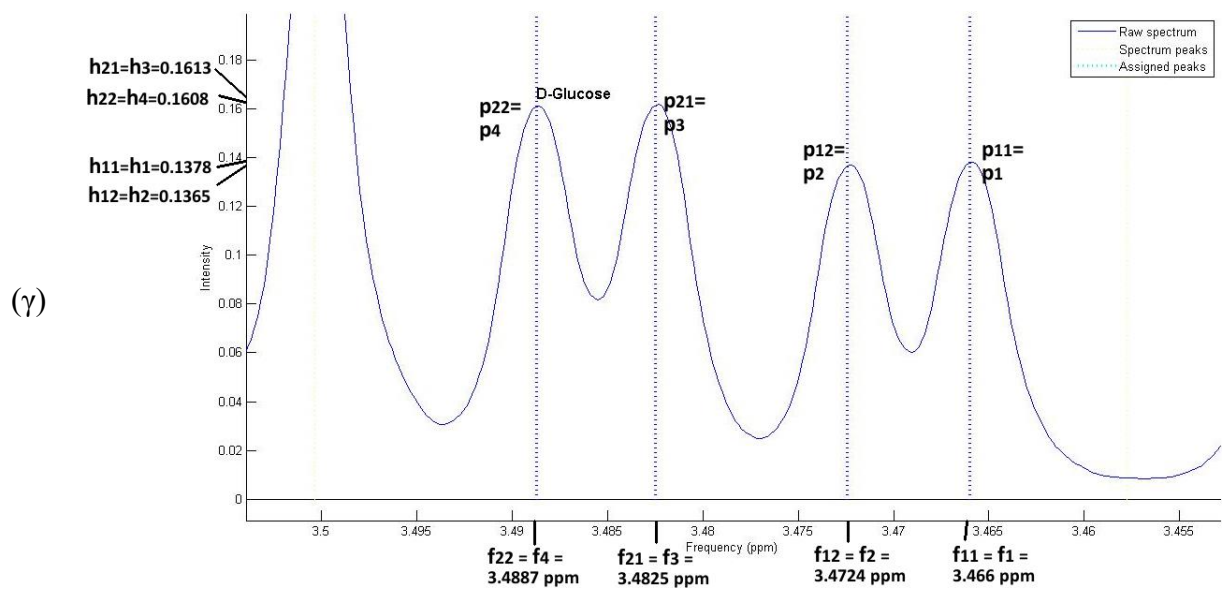
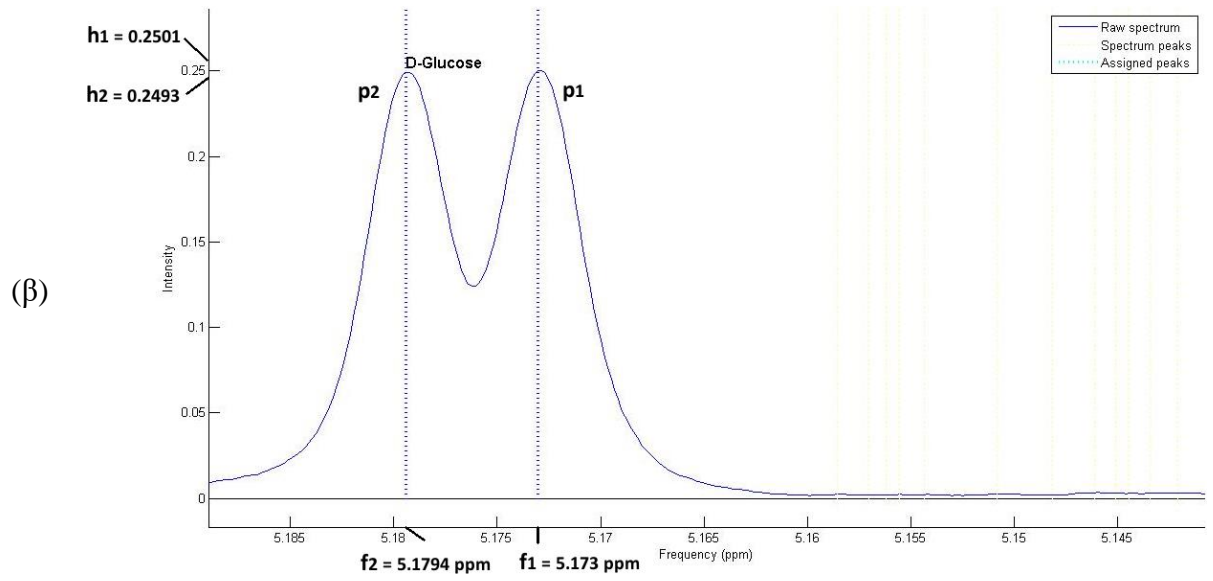
Οι πολλαπλές κορυφές χωρίς κανόνες συνήθως χαρακτηρίζονται από μεγαλύτερο πλήθος υποκορυφών (π.χ.  $N^{Im} > 5$ ), καθώς και μονοτονία που αρχικά αυξάνεται και στη συνέχεια μειώνεται. Επίσης, οι πολλαπλές κορυφές αυτές συνήθως δε συσχετίζονται με συγκεκριμένες τιμές σταθεράς σύζευξης. Ένα υποψήφιο σύνολο κορυφών  $P^{Im} = \{p_1, \dots, p_n, \dots, p_N\}$  για μία πολλαπλή κορυφή χωρίς κανόνες  $I^m$  βαθμολογείται σύμφωνα με την εξίσωση (5.24). Το  $Score_{h_M}$  εκτιμά τη μονοτονία των υψών (εξίσωση (5.25)). Το  $Score_{h_A}$  αντιπροσωπεύει πόσο καλά το σύνολο κορυφών  $P^{Im}$  συμφωνεί με τους ιδανικούς λόγους υψών  $h_{DB_n}$ ,  $n = 1, \dots, N$  που ορίζονται για το  $I^m$  (εξίσωση (5.26)). Το σύνολο κορυφών με το μικρότερο  $Score_M$  θεωρείται το βέλτιστο. Η Εικόνα 5.5 (δ) περιέχει ένα παράδειγμα βαθμολόγησης ενός συνόλου φασματικών κορυφών που είναι υποψήφιο για μία πολλαπλή κορυφή χωρίς κανόνες.

$$Score_M = w_{h_M} \times Score_{h_M} + w_{h_A} \times Score_{h_A} \quad (5.24)$$

$$Score_{h_M} = \left| \sum_{n=1}^{N/2} \frac{h_{n+1} - h_n}{\mu(h_{n+1}, h_n)} - \sum_{n=\frac{N}{2}+1}^{N-1} \frac{h_n - h_{n+1}}{\mu(h_{n+1}, h_n)} \right| \quad (5.25)$$

$$Score_{h_A} = \mu \left( \begin{array}{l} \%error\left(\frac{h_n}{h_1}, \frac{h_{DB_n}}{h_{DB_1}}\right), n = 2, \dots, N \\ \%error\left(\frac{h_n}{h_{n-1}}, \frac{h_{DB_n}}{h_{DB_{n-1}}}\right), n = 3, \dots, N \\ \%error\left(\frac{h_n}{h_{n-2}}, \frac{h_{DB_n}}{h_{DB_{n-2}}}\right), n = 4, \dots, N \end{array} \right) \quad (5.26)$$





Εικόνα 5.5. Παραδείγματα βαθμολόγησης για κάθε κατηγορία κορυφής.

(α) Παράδειγμα βαθμολόγησης για μία απλή κορυφή (singlet).

$$\text{Score}_{d1} = \%diff(|f_2 - f_1|, |f_2 - f_3|) = 0.7073, \text{Score}_{d2} = \mu(|f_2 - f_1|, |f_2 - f_3|) = 0.0082$$

$$\text{Score}_{h1} = \%diff(h_1, h_3) = 0.0398, \text{Score}_{h2} = \mu\left(\frac{h_2-h_1}{\mu(h_2, h_1)}, \frac{h_2-h_3}{\mu(h_2, h_3)}\right) = 1.1858$$

$$\text{Score}_{S(p2)} = -0.2 * \text{Score}_{d1} - 0.2 * \text{Score}_{d2} - 0.2 * \text{Score}_{h1} - 0.4 * \text{Score}_{h2} = -0.6254$$

(β) Παράδειγμα βαθμολόγησης για μία πολλαπλή κορυφή Α τάξης (doublet).

$$j_A = 3.667 \text{ Hz} = 0.00613 \text{ ppm} \text{ (για φασματογράφο συχνότητας 600MHz)}$$

$$h_{DB1} = 2.2275, h_{DB2} = 2.2192$$

$$\text{Score}_{jA} = \%error(|f_2 - f_1|, j_A) = 0.044046, \text{Score}_{hA} = \%error\left(\frac{h_2}{h_1}, \frac{h_{DB2}}{h_{DB1}}\right) = 0.000502,$$

$$\text{Score}_A = 0.6 * \text{Score}_{jA} + 0.4 * \text{Score}_{hA}$$

(γ) Παράδειγμα βαθμολόγησης για μία πολλαπλή κορυφή Β τάξης (double doublet).

$$j_A = 3.73 \text{ Hz} = 0.006217 \text{ ppm} \text{ (για φασματογράφο συχνότητας 600MHz)}$$

$$j_B = 9.823 \text{ Hz} = 0.0163717 \text{ ppm} \text{ (για φασματογράφο συχνότητας 600MHz)}$$

$$h_{DB11} = 1.1778, h_{DB12} = 1.1978, h_{DB21} = 1.4069, h_{DB22} = 1.4134$$

$$\text{Score}_{jA1} = \%error(|f_{12} - f_{11}|, j_A) = 0.0295, \text{Score}_{jA2} = \%error(|f_{22} - f_{21}|, j_A) = 0.0027$$

$$\text{Score}_{hA1} = \%error\left(\frac{h_{12}}{h_{11}}, \frac{h_{DB12}}{h_{DB11}}\right) = 0.026, \quad \text{Score}_{hA2} = \%error\left(\frac{h_{22}}{h_{21}}, \frac{h_{DB22}}{h_{DB21}}\right) = 0.0077$$

$$\text{Score}_{mA} = \mu(0.6 * \text{Score}_{jA1} + 0.4 * \text{Score}_{hA1} + 0.6 * \text{Score}_{jA2} + 0.4 * \text{Score}_{hA2}) = 0.0164$$

$$|\mu(F_2) - \mu(F_1)| = |\mu(f_{21}, f_{22}) - \mu(f_{12}, f_{11})| = 0.0164 \text{ ppm}$$

$$\text{Score}_{jB} = \%error(|\mu(F_2) - \mu(F_1)|, j_B) = 0.0017$$

$$\text{Score}_{hB} = \mu\left(\begin{array}{l} \%error\left(\frac{h_2}{h_1}, \frac{h_{DB2}}{h_{DB1}}\right), \%error\left(\frac{h_3}{h_1}, \frac{h_{DB3}}{h_{DB1}}\right), \%error\left(\frac{h_4}{h_1}, \frac{h_{DB4}}{h_{DB1}}\right), \\ \%error\left(\frac{h_3}{h_2}, \frac{h_{DB3}}{h_{DB2}}\right), \%error\left(\frac{h_4}{h_3}, \frac{h_{DB4}}{h_{DB3}}\right), \%error\left(\frac{h_4}{h_2}, \frac{h_{DB4}}{h_{DB2}}\right) \end{array}\right) = 0.0148$$

$$\text{Score}_B = 0.3 * \text{Score}_{mA} + 0.4 * \text{Score}_{jB} + 0.3 * \text{Score}_{hB} = 0.0101$$

(δ) Παράδειγμα βαθμολόγησης για μία πολλαπλή κορυφή χωρίς κανόνες.

$$\text{Score}_{hM} = \left| \frac{h_2 - h_1}{\mu(h_2, h_1)} - \frac{h_3 - h_2}{\mu(h_3, h_2)} \right| = 0.1474$$

$$h_{DB1} = 0.142, h_{DB2} = 0.466, h_{DB3} = 0.4042$$

$$\text{Score}_{hA} = \mu\left(\%error\left(\frac{h_2}{h_1}, \frac{h_{DB2}}{h_{DB1}}\right), \%error\left(\frac{h_3}{h_1}, \frac{h_{DB3}}{h_{DB1}}\right), \%error\left(\frac{h_3}{h_2}, \frac{h_{DB3}}{h_{DB2}}\right)\right) = 0.4114$$

$$\text{Score}_M = 0.6 * \text{Score}_{hM} + 0.4 * \text{Score}_{hA} = 0.253$$

### Επιλογή συνδυασμών κορυφών

Το στάδιο επιλογής συνδυασμών συνίσταται από την επεξεργασία των αποτελεσμάτων της διαδικασίας screening μεταβολιτών και την επιλογή ενός τελικού συνόλου κορυφών για κάθε μεταβολίτη m. Πιο συγκεκριμένα, στο στάδιο αυτό ένας συνδυασμός  $kc^m$  από τιμές k πρέπει να επιλεγεί ως ο καταλληλότερος για έναν μεταβολίτη m, που θα είναι ο συνδυασμός με τη μικρότερη βαθμολογία ανάμεσα σε αυτούς τους συνδυασμούς που επιβίωσαν από μία διαδικασία αποκλεισμού. Πρώτα, όλοι οι δυνατοί συνδυασμοί  $kc^m$  για το μεταβολίτη m υπό εξέταση διαμορφώνονται. Ακολούθως, όλοι οι συνδυασμοί σαρώνονται και όσοι δεν ικανοποιούν συγκεκριμένα κριτήρια απορρίπτονται. Ακολούθως, όλοι οι συνδυασμοί αποτιμούνται και ο συνδυασμός με τη βέλτιστη (ελάχιστη) βαθμολογία (score) επιλέγεται ως αυτός με την καλύτερη εφαρμογή για το μεταβολίτη m.

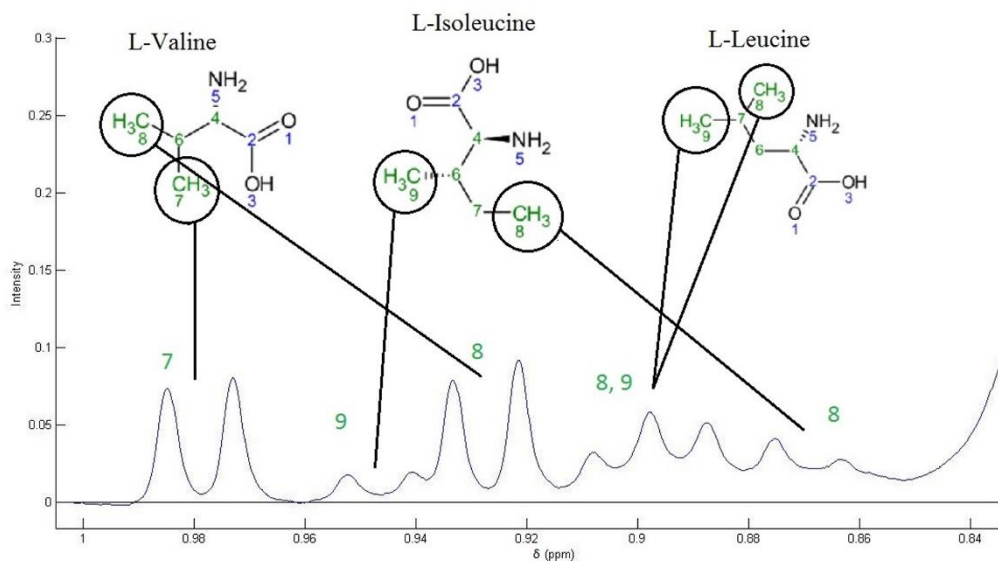
Οι συνδυασμοί  $kc^m$  διαμορφώνονται εξαντλητικά, με δεδομένο τον ακόλουθο περιορισμό. Κάθε συνδυασμός μπορεί να περιέχει τιμές σε ένα παράθυρο  $w = w_k^m = [k \ k +$

$kd^m$ ] με μήκος  $kd^m$ . Η μεταβλητή  $kd^m$  ελέγχει το εύρος τιμών της μεταβλητής  $k$  που ένας συνδυασμός  $kc^m$  έχει κι εξαρτάται από τον αριθμό  $L^m$  των απλών ή πολλαπλών κορυφών ενός μεταβολίτη  $m$ . Αρχικά, η μεταβλητή  $kd^m$  τέθηκε στις τιμές 4-5. Παρ' όλα αυτά, οι τιμές αυτές οδήγησαν σε έναν υπερβολικά μεγάλο αριθμό συνδυασμών  $kc^m$ , πολλοί από τους οποίους ήταν ακατάλληλοι εξ' αιτίας της μεγάλης διαστρέβλωσης των αποστάσεων μεταξύ των κορυφών ενός μεταβολίτη. Επομένως, το  $kd^m$  ορίστηκε σύμφωνα με την εξίσωση (5.27).

$$kd^m = \begin{cases} 0, & \text{if } L^m \leq 5 \\ 1, & \text{if } 5 < L^m \leq 8 \\ 2, & \text{if } L^m > 8 \end{cases} \quad (5.27)$$

Ακολούθως, όλοι οι συνδυασμοί  $kc^m$  υπόκεινται σε μία διαδικασία ελέγχου, κατά την οποία οι υποψήφιοι συνδυασμοί κορυφών ενός μεταβολίτη εξετάζονται ως προς συγκεκριμένα κριτήρια και αν δεν τα πλοιρούν απορρίπτονται. Το σημαντικότερο κριτήριο ελέγχου είναι η παρουσία ή απουσία προτεινόμενων συνόλων κορυφών για τις κορυφές των μεταβολιτών. Η αρχική προσέγγιση της μεθοδολογίας ήταν να απαιτείται ένα προτεινόμενο σύνολο κορυφών για κάθε απλή ή πολλαπλή κορυφή ενός μεταβολίτη ώστε αυτός να θεωρηθεί παρόν στο φάσμα. Παρ' όλα αυτά, έγινε εμφανές ότι μερικές κορυφές μπορεί να μην είναι αναγνωρίσιμες γιατί είναι επικαλυμμένες από τις κορυφές άλλων μεταβολιτών, ένα φαινόμενο αρκετά συνηθισμένο στην περιοχή 3-4 ppm. Επομένως, για να θεωρηθεί ένας μεταβολίτης παρόν στο φάσμα υπό εξέταση, δεν είναι υποχρεωτικό να υπάρχει προτεινόμενο σύνολο φασματικών κορυφών για όλες τις απλές ή πολλαπλές κορυφές του, εκτός και αν ο μεταβολίτης αποτελείται μόνο από μία. Είναι όμως απαραίτητο να αναγνωριστούν οι πιο εμφανείς κορυφές του, οι οποίες συνήθως εντοπίζονται εκτός της περιοχής συχνοτήτων 3-4 ppm στα φάσματα βιολογικών δειγμάτων. Οι συνδυασμοί στους οποίους δε βρέθηκαν υποψήφιες φασματικές κορυφές για αυτές τις κορυφές ενός μεταβολίτη θεωρούνται ακατάλληλοι κι απορρίπτονται. Άκυροι συνδυασμοί, όπως συνδυασμοί με μη αποδεκτές επικαλύψεις κορυφών, απορρίπτονται.

Επιπλέον, με κίνητρο τη σταθεροποίηση των αποτελεσμάτων ταυτοποίησης, ενσωματώθηκε ένας μικρός αριθμός κανόνων σχετικά με τη σχετική θέση συγκεκριμένων πρωτονίων στις χημικές δομές των μεταβολιτών στη μεθοδολογία. Όλοι οι συνδυασμοί  $kc^m$  ενός μεταβολίτη  $m$  εξετάζονται και, αν δεν ικανοποιούν τους αντίστοιχους κανόνες, απορρίπτονται. Για παράδειγμα, είναι γνωστό ότι το triplet της L-Λευκίνης εντοπίζεται σε μικρότερες συχνότητες από τα doublets της L-Βαλίνης. Στην Εικόνα 5.6 απεικονίζονται οι χημικές δομές των μεταβολιτών L-Βαλίνη, L-Ισολευκίνη και L-Λευκίνη, οι κορυφές τους στο φασματικό εύρος 0.84-1 ppm ενός δείγματος ορού, καθώς και τα πρωτόνια που αντιστοιχούν στις απεικονιζόμενες κορυφές.



Εικόνα 5.6. Οι χημικές δομές των μεταβολιτών L-Βαλίνη (L-Valine), L-Ισολευκίνη (L-Isoleucine), L-Λευκίνη (L-Leucine), οι κορυφές τους στο φασματικό εύρος 0.84-1 ppm ενός δείγματος ορού και τα αντίστοιχα πρωτόνια τους.

Οι εναπομείναντες συνδυασμοί βαθμολογούνται σύμφωνα με την εξίσωση (5.28). Το  $Score_{\text{Multiplets}}^{kc^m}$  αντιπροσωπεύει το μέσο των βαθμολογιών των κορυφών του μεταβολίτη, που υπολογίζονται σύμφωνα με τις εξισώσεις (5.12)-(5.26). Το  $Score_{\text{Heights}}^{kc^m}$  αντιπροσωπεύει το πόσο καλά τα ύψη όλων των κορυφών του  $kc^m$  συμφωνούν με τις ιδανικές αναλογίες υψών των κορυφών του μεταβολίτη (εξίσωση (5.30)). Ο συνδυασμός  $kc^m$  με την ελάχιστη τιμή  $Score^{kc^m}$  επιλέγεται ως ο καταλληλότερος για το μεταβολίτη  $m$ .

$$Score^{kc^m} = w_M \times Score_{\text{Multiplets}}^{kc^m} + w_H \times Score_{\text{Heights}}^{kc^m} \quad (5.28)$$

$$Score_{\text{Multiplets}}^{kc^m} = \mu \left( \begin{array}{l} \left( \begin{array}{l} Score_S^{kc^m, l^m}, \\ Score_A^{kc^m, l^m}, \\ Score_B^{kc^m, l^m}, \\ \dots \end{array} \right. \text{αν το } l^m \text{ είναι singlet (εξίσωση (5.12))} \\ \left. \begin{array}{l} \text{αν το } l^m \text{ είναι multiplet A τάξης (εξίσωση (5.17))} \\ \text{αν το } l^m \text{ είναι multiplet B τάξης (εξίσωση (5.20))} \\ \dots \end{array} \right) \\ Score_M^{kc^m, l^m}, \text{ αν το } l^m \text{ είναι multiplet χωρίς κανόνες (εξίσωση (5.24))} \end{array} \right) \quad (5.29)$$

$$Score_{\text{Heights}}^{kc^m} = \mu \left( \begin{array}{l} \%error \left( \frac{h_n^{kc^m}}{h_1^{kc^m}}, \frac{h_{DBn}}{h_{DB1}} \right), \quad n = 2, \dots, N^m \\ \%error \left( \frac{h_n^{kc^m}}{h_{n-1}^{kc^m}}, \frac{h_{DBn}}{h_{DBn-1}} \right) \quad n = 3, \dots, N^m \\ \%error \left( \frac{h_n^{kc^m}}{h_{n-2}^{kc^m}}, \frac{h_{DBn}}{h_{DBn-2}} \right) \quad n = 4, \dots, N^m \end{array} \right) \quad (5.30)$$

Η Εικόνα 5.7 περιέχει παραδείγματα της διαδικασίας του screening μεταβολιτών (α)-(δ), της επιλογής συνδυασμών κορυφών (ε)-(ζ) και των αποτελεσμάτων αναγνώρισης μεταβολιτών (η)-(θ) στο φάσμα του δείγματος S25. Η Εικόνα 5.7 (α) περιέχει ένα παράδειγμα

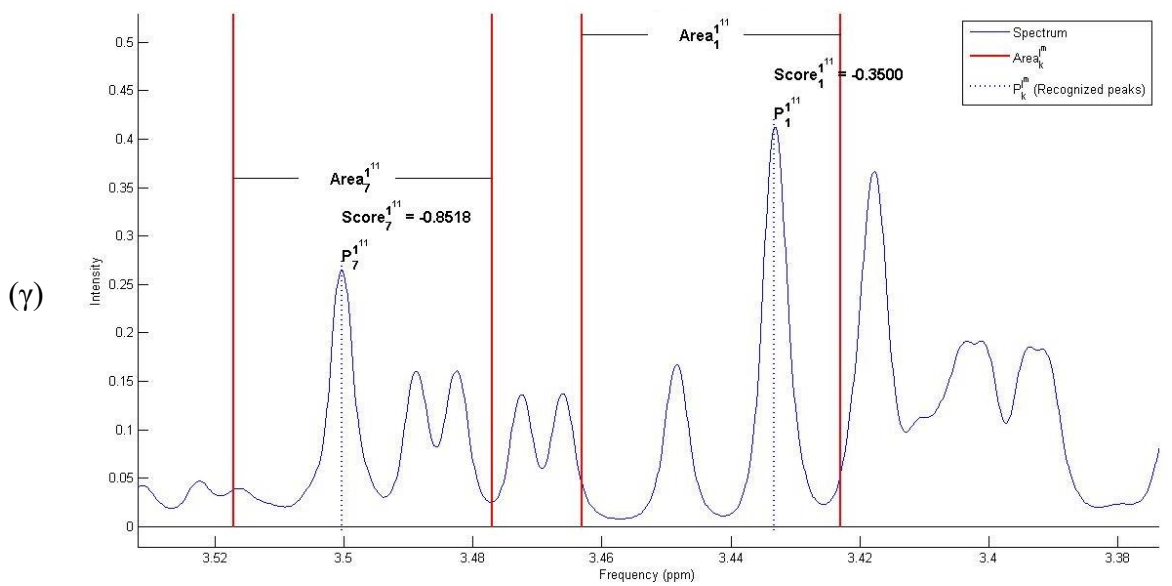
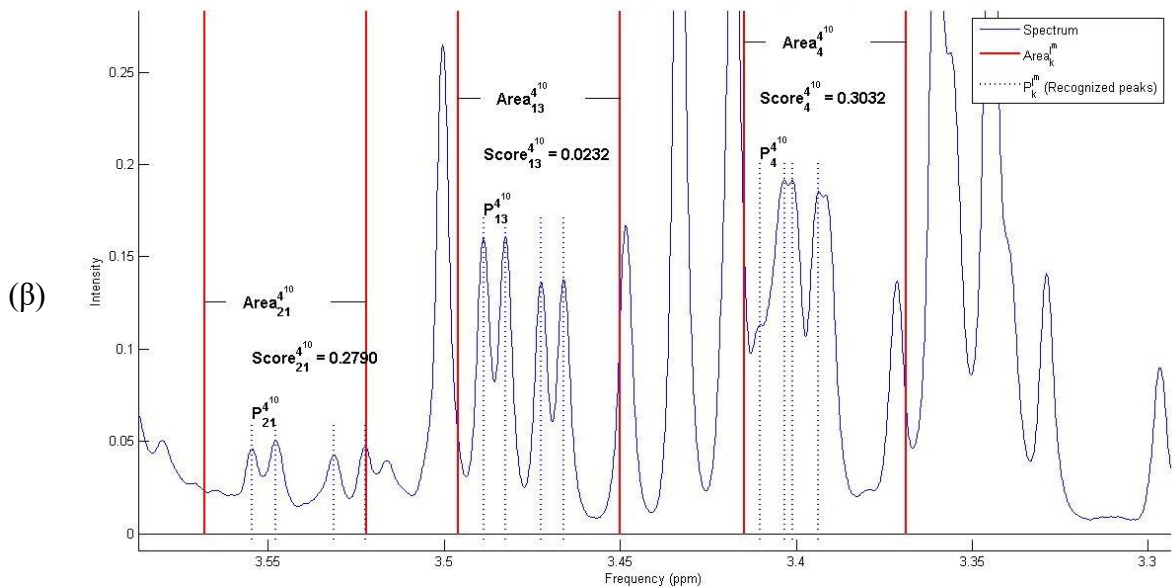
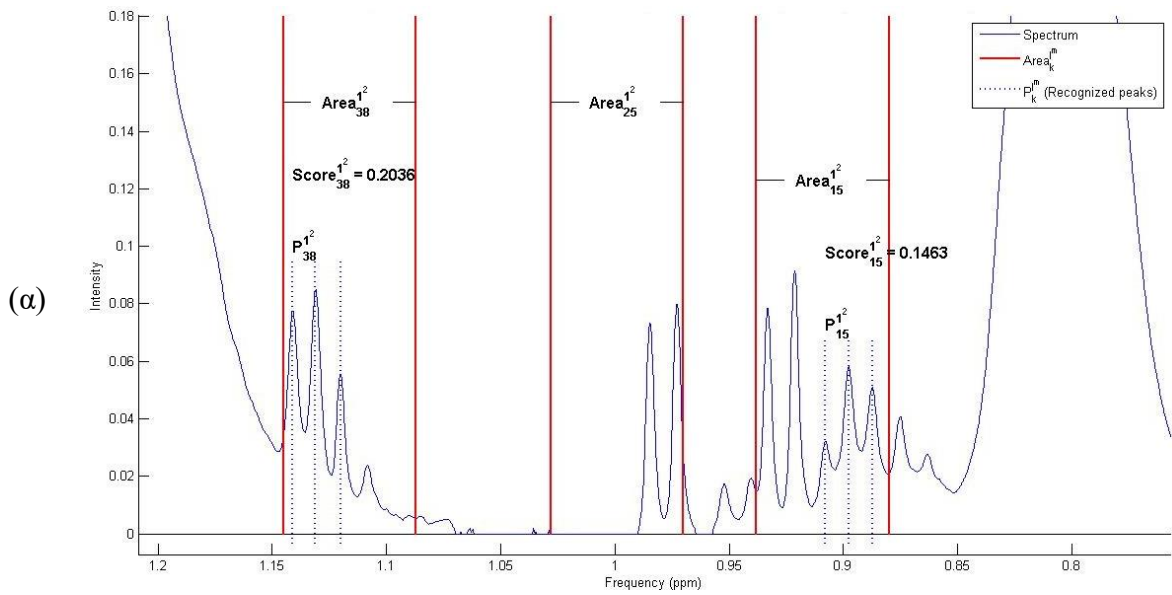
της διαδικασίας screening μεταβολιτών για μία πολλαπλή κορυφή A τάξης, συγκεκριμένα το triplet ( $l=1$ ) της L-Λευκίνης ( $m=2$ ). Τα σύνολα κορυφών  $P_{15}^{1^2}, P_{25}^{1^2}, P_{38}^{1^2}$  που έχουν επιλεγεί ως βέλτιστα για τις περιοχές  $area_{15}^{1^2}, area_{25}^{1^2}, area_{38}^{1^2}$  ( $k=15, 25, 38$ ) αντίστοιχα επισημαίνονται με κάθετες διαδοδομένες γραμμές, μαζί με τις αντίστοιχες βαθμολογίες  $Score_{15}^{1^2}, Score_{25}^{1^2}, Score_{38}^{1^2}$ . Για  $k=25$  δεν έχει επιλεγεί κάποιο σύνολο κορυφών ( $P_{25}^{1^2} = \emptyset$ ) γιατί η περιοχή  $area_{25}^{1^2}$  περιέχει μόνο δύο φασματικές κορυφές. Το πραγματικό triplet της L-Λευκίνης ταυτίζεται με το σύνολο κορυφών  $P_{15}^{1^2}$ , που έχει καλύτερη βαθμολογία από το σύνολο κορυφών  $P_{38}^{1^2}$  ( $Score_{15}^{1^2} < Score_{38}^{1^2}$ ).

Η Εικόνα 5.7 (β) περιέχει ένα παρόμοιο παράδειγμα για μία πολλαπλή κορυφή B τάξης, συγκεκριμένα ένα double doublet ( $l=4$ ) της D-Γλυκόζης ( $m=10$ ). Απεικονίζονται τα σύνολα κορυφών  $P_4^{4^{10}}, P_{13}^{4^{10}}, P_{21}^{4^{10}}$  που έχουν επιλεγεί ως βέλτιστα για τις περιοχές  $area_4^{4^{10}}, area_{13}^{4^{10}}, area_{21}^{4^{10}}$  ( $k=4,13,21$ ) αντίστοιχα, καθώς και οι βαθμολογίες τους  $Score_4^{4^{10}}, Score_{13}^{4^{10}}, Score_{21}^{4^{10}}$ . Το πραγματικό double doublet της D-Γλυκόζης ταυτίζεται με το σύνολο κορυφών  $P_{13}^{4^{10}}$ , που έχει καλύτερη βαθμολογία από τα σύνολα  $P_4^{4^{10}}, P_{21}^{4^{10}}$  ( $Score_{13}^{4^{10}} < Score_4^{4^{10}}, Score_{13}^{4^{10}} < Score_{21}^{4^{10}}$ ).

Η Εικόνα 5.7 (γ) περιέχει τα αποτελέσματα της διαδικασίας screening μεταβολιτών για την απλή κορυφή ( $l=1$ ) της Γλυκίνης ( $m=11$ ) και  $k=1,7$ . Η κορυφή  $P_7^{1^{11}}$  στην περιοχή  $area_7^{1^{11}}$  έχει καλύτερη βαθμολογία από την κορυφή  $P_1^{1^{11}}$  στην περιοχή  $area_1^{1^{11}}$  ( $Score_7^{1^{11}} < Score_1^{1^{11}}$ ). Το γεγονός ότι οι τέσσερις φασματικές κορυφές στο εύρος 3.46-3.5 ppm έχουν ήδη δεσμευθεί ως ένα double doublet της D-Γλυκόζης, αφού η D-Γλυκόζη έχει μεγαλύτερη προτεραιότητα από τη Γλυκίνη, έχει συμβάλει σε αυτό αυξάνοντας τα  $Score_{d1}, Score_{d2}$  της υποψήφιας κορυφής  $P_7^{1^{11}}$  (Εξισώσεις (5.13,5.14)).

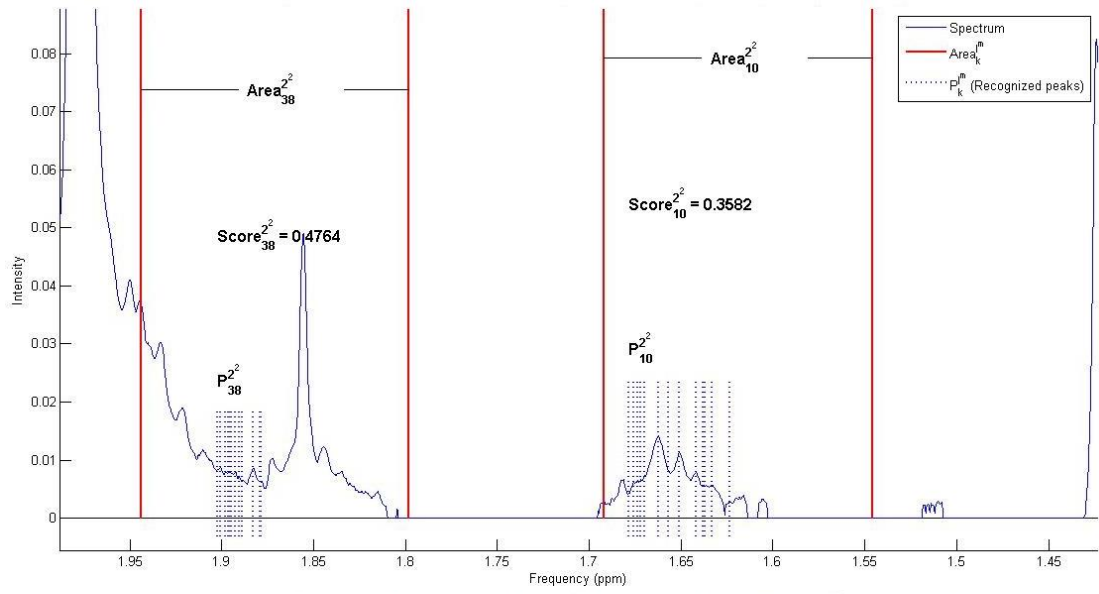
Η Εικόνα 5.7 (δ) περιέχει ένα παρόμοιο παράδειγμα για την πολλαπλή κορυφή χωρίς κανόνες ( $l=2$ ) της L-Λευκίνης ( $m=2$ ). Το σύνολο κορυφών  $P_{10}^{2^2}$  στην περιοχή  $area_{10}^{2^2}$  έχει πιο κατάλληλη μονοτονία από το σύνολο  $P_{38}^{2^2}$  της περιοχής  $area_{38}^{2^2}$ , που αντικατοπτρίζεται στη βαθμολογία τους ( $Score_{10}^{2^2} < Score_{38}^{2^2}$ ). Πρέπει να σημειωθεί ότι η τελική επιλογή των κορυφών για όλες τις παραπάνω κορυφές γίνεται στο μεταγενέστερο στάδιο της επιλογής συνδυασμών.

Οι Εικόνες 5.7 (ε)-(ζ) περιέχουν παραδείγματα από τη διαδικασία επιλογής συνδυασμών στο φάσμα ορού S25. Η Εικόνα 5.7 (ε) απεικονίζει τμήμα ενός συνδυασμού κορυφών ( $kc=6$ ), στον οποίο είναι  $k=6$  για δύο doublets ( $l=1,2$ ) της L-Βαλίνης ( $m=3$ ). Αυτός ο συνδυασμός δε θεωρείται έγκυρος, γιατί δεν έχουν βρεθεί κορυφές για το doublet  $l=1$  ( $P_6^{1^3} = \emptyset$ ). Η Εικόνα 5.7 (ζ) απεικονίζει μέρος ενός άλλου συνδυασμού κορυφών ( $kc=14$ ) στον οποίο είναι  $k=11$  για τα ίδια doublets της L-Βαλίνης. Σε αυτό το στάδιο, ο συνδυασμός θεωρείται έγκυρος, γιατί έχει βρεθεί ένα σύνολο κορυφών για κάθε ένα από τα δύο doublets ( $P_{11}^{1^3}, P_{11}^{2^3} \neq \emptyset$ ). Τέλος, η Εικόνα 5.7 (η) απεικονίζει μεταβολίτες που αναγνωρίστηκαν από την προτεινόμενη μεθοδολογία στη φασματική περιοχή 0.8-1.7 ppm, και η Εικόνα 5.7 (θ) απεικονίζει τις απλές κορυφές που αναγνωρίστηκαν στη φασματική περιοχή 1.8-3.6 ppm. Τα ονόματα των ταυτοποιημένων μεταβολιτών επισημαίνονται πάνω από κάθε αντίστοιχο σύνολο κορυφών.

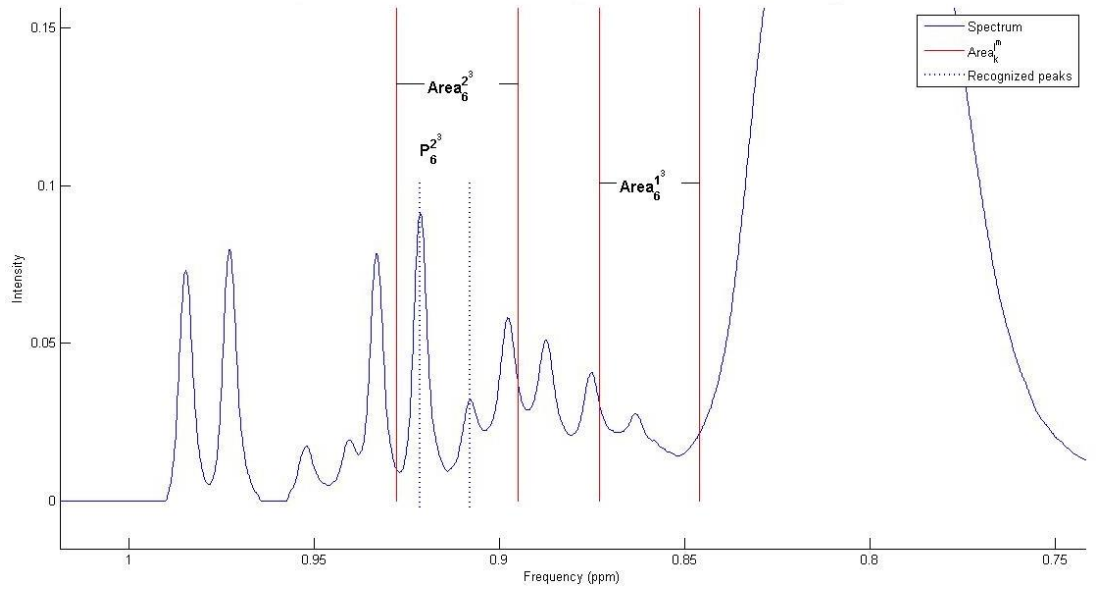




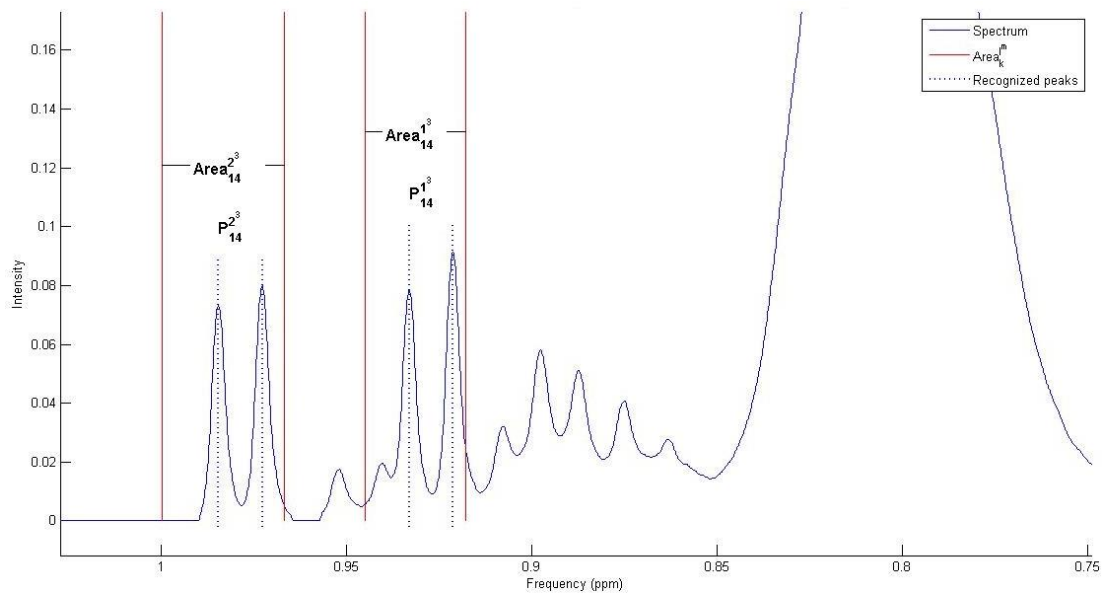
(δ)

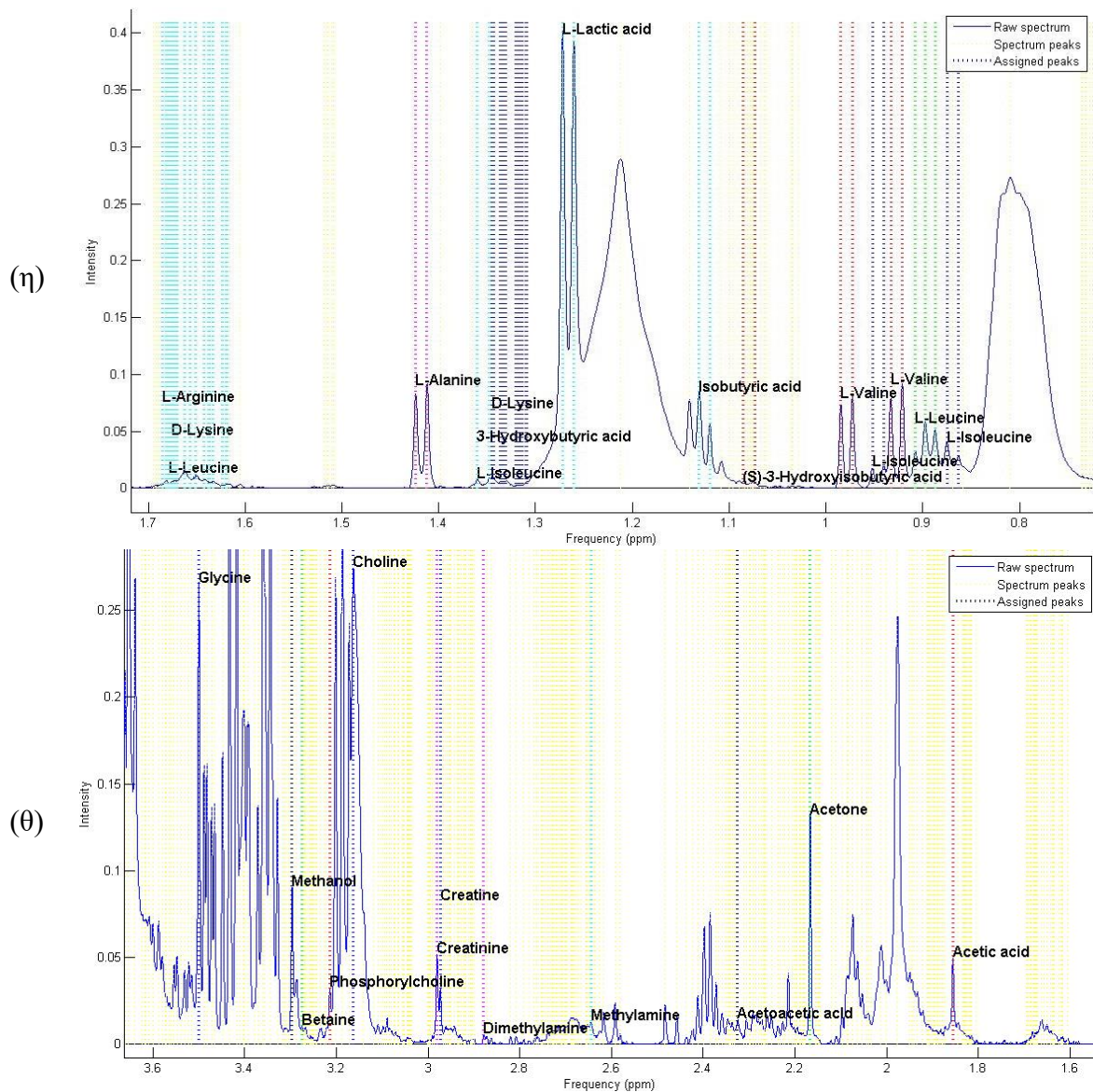


(ε)



(ζ)





Εικόνα 5.7. Απεικόνιση της διαδικασίας αναζήτησης μεταβολιτών στο φάσμα του δείγματος ορού S25. (α) Screening για το triplet του μεταβολίτη L-Λευκίνη (L-Leucine). (β) Screening για ένα double doublet του μεταβολίτη D-Γλυκόζη (D-Glucose). (γ) Screening για το singlet του μεταβολίτη Γλυκίνη (Glycine). (δ) Screening για την πολλαπλή κορυφή χωρίς κανόνες του μεταβολίτη L-Λευκίνη. (ε) Μέρος ενός άκυρου συνδυασμού κορυφών για το μεταβολίτη L-Βαλίνη (L-Valine). (ζ) Μέρος ενός έγκυρου συνδυασμού κορυφών για το μεταβολίτη L-Βαλίνη. (η) Μεταβολίτες που ταυτοποιήθηκαν από την προτεινόμενη μεθοδολογία στο φασματικό εύρος 0.8-1.7 ppm. (θ) Απλές κορυφές που ταυτοποιήθηκαν από την προτεινόμενη μεθοδολογία στο φασματικό εύρος 1.8-3.6 ppm.

### Ιεράρχηση μεταβολιτών

Μέχρι στιγμής, ο ορός αίματος ή το πλάσμα είναι το πιο ευρέως μελετημένο βιολογικό υγρό και το μεταβόλωμά του έχει υπάρξει το αντικείμενο ενός συνόλου μελετών (Fotakis et al., 2016; Psychogios et al., 2011). Σε ένα μικρότερο βαθμό, άλλες μελέτες έχουν επικεντρωθεί στο μεταβολικό προφίλ του αμνιακού υγρού (Orczyk-Pawilowicz et al., 2016). Αυτές οι μελέτες έχουν αναδείξει συγκεκριμένους μεταβολίτες που συνήθως βρίσκονται σε υψηλότερες συγκεντρώσεις. Επιπλέον, μεταβολίτες με κορυφές που εντοπίζονται εκτός της

περιοχής 3-4 ppm συνήθως είναι πιο ευδιάκριτοι σε σύγκριση με μεταβολίτες με κορυφές στο εύρος συχνοτήτων 3-4 ppm. Με δεδομένα τα παραπάνω, ο διαχωρισμός των υποψήφιων μεταβολιτών σε ομάδες διαφορετικών προτεραιοτήτων αποδείχθηκε αρκετά χρήσιμος. Το φάσμα εισόδου πρώτα σαρώνεται για τους μεταβολίτες υψηλότερης προτεραιότητας, που συμπεριλαμβάνουν τους Κιτρικό οξύ, L-Αλανίνη, L-Λακτικό οξύ, L-Βαλίνη, L-Λευκίνη, L-Ισολευκίνη, D-Γλυκόζη, Φορμικό οξύ, L-Φαινυλαλανίνη, L-Τυροσίνη, L-Ιστιδίνη, Ακετικό οξύ, Κρεατίνη, Κρεατινίνη, Γλυκίνη.

Η ομάδα χαμηλότερης προτεραιότητας συμπεριλαμβάνει τους εναπομείναντες μεταβολίτες που συνίστανται από λιγότερο ευδιάκριτες κορυφές στο εύρος συχνοτήτων 3-4 ppm, όπως οι Αιθανολαμίνη, Σερίνη και Γλυκερόλη. Αφού αυτοί οι μεταβολίτες είναι συνήθως λιγότερο διακριτοί, η απόφαση σχετικά με τις κορυφές τους είναι λιγότερο ασφαλής και επομένως λαμβάνεται σε μεταγενέστερο στάδιο.

### **5.3. Αποτελέσματα**

#### **5.3.1 Απόδοση μεθοδολογίας**

Η έξοδος της προτεινόμενης μεθοδολογίας είναι μία λίστα μεταβολιτών, καθώς και οι κορυφές που ανατέθηκαν σε κάθε μεταβολίτη. Τα αποτελέσματα εκτιμήθηκαν ως προς την παρουσία ή απουσία των εξεταζόμενων μεταβολιτών αλλά και την ορθότητα των προτεινόμενων κορυφών. Ο Πίνακας 5.1 περιλαμβάνει τη μέση ακρίβεια, ευαισθησία και ειδικότητα της προτεινόμενης μεθοδολογίας στις ομάδες δειγμάτων S1-S2, S3-S22, S23-S42, MTBLS174 1-18, MTBLS424 1-50. Οι Πίνακες Π.5, Π.6 περιέχουν τις τιμές ακρίβειας, ευαισθησίας και ειδικότητας της προτεινόμενης μεθοδολογίας για  $a=0.12$  για τα δείγματα S1-S2, S3-S22, S23-S42, MTBLS174 1-18, MTBLS424 1-50, MTBLS242 1-50, MTBLS326 1-42. Οι τιμές ακρίβειας, ευαισθησίας και ειδικότητας που επιτεύχθηκαν από την προτεινόμενη μεθοδολογία στο μείγμα αμινοξέων (S1) είναι 64.29%, 100%, και 54.59% αντίστοιχα. Οι τιμές ακρίβειας, ευαισθησίας και ειδικότητας που επιτεύχθηκαν στο δείγμα spiked ορού (S2) είναι 81.82%, 96.55% και 53.33% αντίστοιχα. Επιπλέον, η προτεινόμενη μεθοδολογία σημείωσε μέση ακρίβεια 81.44% στα δείγματα αμνιακού υγρού (S3-S22), 77.27% στα δείγματα ορού (S23-S42), 88.49% στα δείγματα MTBLS174, 76.18% στα δείγματα MTBLS424, 75.38% στα δείγματα MTBLS242 και 76.19% στα δείγματα MTBLS326. Η μεθοδολογία βελτιστοποιήθηκε για εφαρμογή σε ορό και αμνιακό υγρό, παρουσιάζοντας αποτελέσματα σε αυτά τα υποστρώματα με μέσες τιμές ακρίβειας πάνω από 75%.

Πίνακας 5.1. Τα μέσα αποτελέσματα της προτεινόμενης μεθοδολογίας ( $\alpha=0.12$ ), καθώς και των λογισμικών MetaboHunter και Bayesil στις ομάδες δειγμάτων S1-S2, S3-S22, S23-S42, MTBLS174 1-18, MTBLS424 1-50, MTBLS242 1-50, MTBLS326 1-42.

Μέθοδος	Μετρικές (%)	Ομάδες δειγμάτων						
		S1-S2	S3-S22	S23-S42	MTBLS174 1-18	MTBLS424 1-50	MTBLS242 1-50	MTBLS326 1-42
Προτεινόμενη Μεθοδολογία	Ακρίβεια	73.052	81.444	77.273	88.492	76.182	75.375	76.191
	Ευαισθησία	98.276	94.998	95.941	95.598	90.576	89.57	85.952
	Ειδικότητα	53.939	55.951	49.503	..**	2.486	0*	..**
MetaboHunter	Ακρίβεια	39.404	50	47.727	33.333	90.152	26.891	80
	Ευαισθησία	88.194	95	94.737	83.333	95.198	69.388	100
	Ειδικότητα	17.278	12.5	12	0*	0*	0*	0*
Bayesil	Ακρίβεια	42.362	70.539	61.934	67.378	74.762	68.947	78.571
	Ευαισθησία	83.333	78.191	65.822	67.378	74.843	68.947	78.571
	Ειδικότητα	4.546	32.095	43.981	..**	..**	..**	..**

\* Η μηδενική ειδικότητα οφείλεται στο γεγονός ότι το πλήθος των αληθών αρνητικών αποτελεσμάτων είναι μηδέν. Αυτό είναι αναμενόμενο γιατί οι εξεταζόμενοι μεταβολίτες αναμένεται να βρίσκονται στα φάσματα.

\*\* Η απουσία αποτελέσματος ειδικότητας οφείλεται στο γεγονός ότι τόσο το πλήθος των αληθών αρνητικών αποτελεσμάτων όσο και το πλήθος των ψευδών θετικών αποτελεσμάτων είναι μηδενικά.

### 5.3.2 Συγκριτική Αξιολόγηση

Η απόδοση της προτεινόμενης μεθοδολογίας συγκρίθηκε με την απόδοση των λογισμικών αναγνώρισης μεταβολιτών MetaboHunter και Bayesil. Ο Πίνακας 5.1 περιλαμβάνει τα αποτελέσματα μέσης ακρίβειας, ευαισθησίας και ειδικότητας του MetaboHunter και Bayesil σε κάθε ομάδα δειγμάτων. Οι Πίνακες Π.5, Π.6 περιέχουν τα αποτελέσματα ακρίβειας, ευαισθησίας και ειδικότητας του MetaboHunter και Bayesil για κάθε δείγμα ξεχωριστά.

Το MetaboHunter χρησιμοποιεί μία συνάρτηση βαθμολόγησης και μία από τρεις μεθόδους αναζήτησης για να καταλήξει σε ένα σύνολο μεταβολιτών για ένα φάσμα εισόδου. Με δεδομένο ένα φάσμα εισόδου, το MetaboHunter παράγει ως έξοδο μία λίστα μεταβολιτών, καθώς και τις κορυφές που ανατέθηκαν σε κάθε μεταβολίτη. Για την εφαρμογή του MetaboHunter, τα φάσματα εισόδου που χρησιμοποιήθηκαν υποβλήθηκαν σε διόρθωση βασικής γραμμής και φάσης, και το κατώφλι θορύβου που ορίστηκε ως παράμετρος για κάθε φάσμα ήταν παρόμοιο με αυτό που υπολογίστηκε από την προτεινόμενη μεθοδολογία. Χρησιμοποιήθηκε η μέθοδος MH2 του MetaboHunter, με τιμή παραμέτρου shift tolerance 0.1 ppm και τιμή παραμέτρου confidence threshold 0.5, αφού ένα χαμηλότερο confidence threshold θα επέφερε περισσότερα ψευδή θετικά (false positive) αποτελέσματα (π.χ. Τριγονελλίνη), ενώ μία υψηλότερη τιμή θα σήμαινε λιγότερα αληθή αρνητικά (true positive) αποτελέσματα. Αν και η λίστα μεταβολιτών εξόδου του MetaboHunter συνήθως περιλαμβάνει εκατοντάδες μεταβολίτες, μόνο οι εξεταζόμενοι μεταβολίτες αξιολογήθηκαν

(45 μεταβολίτες στα δείγματα S1-S42, 14 μεταβολίτες στα δείγματα MTBLS174, 22 μεταβολίτες στα δείγματα MTBLS424). Τα αποτελέσματα εκτιμήθηκαν ως προς την παρουσία ή απουσία μεταβολιτών, καθώς και την ορθότητα των κορυφών που προτείνονται για κάθε μεταβολίτη. Αυτό σημαίνει ότι ακόμα και όταν ένας μεταβολίτης χαρακτηρίζεται σωστά ως θετικός (positive), δηλαδή παρόν στο φάσμα υπό εξέταση, οι κορυφές που του ανατέθηκαν μπορεί να είναι διαφορετικές από αυτές που αντιστοιχούν στο μεταβολίτη. Πιο ειδικά, οι κορυφές που ανατίθενται σε έναν συγκεκριμένο μεταβολίτη από το MetaboHunter έχουν ανάλυση 2 δεκαδικών ψηφίων και πάντα ανήκουν στο ίδιο σύνολο κορυφών, ανεξάρτητα από το φάσμα εισόδου. Για παράδειγμα, σύμφωνα με το MetaboHunter, οι κορυφές που αντιστοιχούν στον μεταβολίτη Κιτρικό οξύ είναι πάντα ένα υποσύνολο των συχνοτήτων {2.51, 2.54, 2.64, 2.67} ppm, πράγμα που δεν ισχύει σε όλα τα φάσματα, όπως φαίνεται στην Εικόνα 5.3, και οι κορυφές που αντιστοιχούν στην Κρεατίνη είναι πάντα ένα υποσύνολο του {3.02, 3.92} ppm. Περισσότερα παραδείγματα δίνονται στην Εικόνα 5.8, όπου μπορεί να παρατηρηθεί ότι οι κορυφές που ανατίθενται από το MetaboHunter στο φάσμα MTBLS174\_1 για 4 μεταβολίτες δεν αντιστοιχούν στις πραγματικές συχνότητες των κορυφών αυτών των μεταβολιτών. Η χαμηλότερη ως προς την ακρίβεια απόδοση του MetaboHunter στις περισσότερες ομάδες δειγμάτων μπορεί να αποδοθεί στο ότι δεν λαμβάνει υπόψιν τη μετατόπιση κι αλληλεπικάλυψη κορυφών. Το MetaboHunter σημείωσε μέση ακρίβεια 50% στα δείγματα αμνιακού υγρού (S3-S22), 47.73% στα δείγματα ορού (S23-S42), 33.33% στα δείγματα ορού MTBLS174, 90.15% στα δείγματα ορού MTBLS424, 80% στα δείγματα ορού MTBLS326 και 26.89% στα δείγματα ορού MTBLS242.

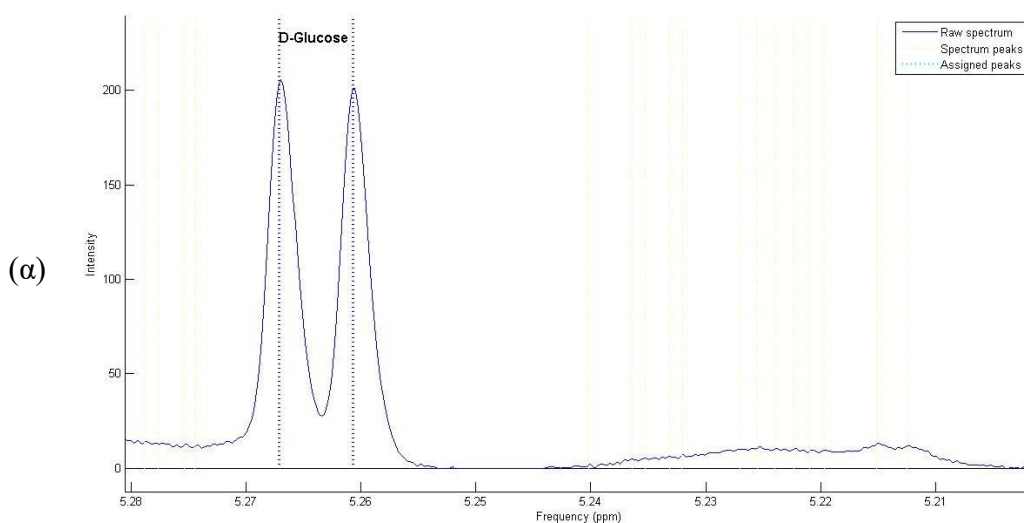
Το Bayesil δέχεται ως είσοδο το ανεπεξέργαστο αρχείο ελεύθερης επαγωγικής απόσβεσης ενός φάσματος. Πραγματοποιεί προεπεξεργασία, χωρίζει το φάσμα σε τμήματα και χρησιμοποιεί ένα πιθανοτικό γραφικό μοντέλο για να διαμορφώσει το μεταβολικό προφίλ του, παράγοντας ως έξοδο μία λίστα μεταβολιτών και τις συγκεντρώσεις τους. Τα αποτελέσματά του αξιολογήθηκαν λαμβάνοντας υπόψιν μόνο την παρουσία ή απουσία μεταβολιτών και όχι την ακρίβεια κορυφών ή συγκέντρωσης, αφού το Bayesil δεν αναθέτει συγκεκριμένες κορυφές σε κάθε μεταβολίτη. Το Bayesil προορίζεται κυρίως για δείγματα ορού, πλάσμα και εγκεφαλονωτιαίου υγρού, αλλά σημείωσε καλή απόδοση και στα δείγματα αμνιακού υγρού. Η μέση ακρίβειά του ήταν 70.54% στα δείγματα αμνιακού υγρού (S3-S22), 61.93% στα δείγματα ορού (S23-S42), 67.38% στα δείγματα ορού MTBLS174, 74.76% στα δείγματα ορού MTBLS424, 78.57% στα δείγματα ορού MTBLS326 και 68.95% στα δείγματα ορού MTBLS242.

Η ακρίβεια της μεθόδου Autofit (που βρίσκεται ενσωματωμένη στο εμπορικό πακέτο λογισμικού Chenomx) εκτιμήθηκε με βάση τους μεταβολίτες που εκτιμήθηκαν αυτόματα. Το Autofit κατάφερε να ταυτοποιήσει μόνο μέχρι 6 μεταβολίτες σε κάθε δοκιμασμένο βιολογικό δείγμα (Πίνακας Π.7, Παράρτημα). Η χαμηλή απόδοση του αλγορίθμου αυτού μπορεί να αποδοθεί στην ευαισθησία στο pH του δείγματος εισόδου. Το Autofit έχει βελτιστοποιηθεί για να διαχειρίζεται αποτελεσματικά φάσματα που έχουν αποκτηθεί με συγκεκριμένο τρόπο προετοιμασίας (Θερμοκρασία 25 °C, Χρόνος καταγραφής σήματος 4 s, Καθυστέρηση χαλάρωσης = 1 s, Φασματικό εύρος = 12 ppm, ακολουθίες παλμών NOESY με χρόνο tmix 100 ms). Παρόμοια αποτελέσματα του Autofit με ακρίβεια 54% έχουν σημειωθεί σε συνθετικά δείγματα ούρων (Tardival et al., 2017).

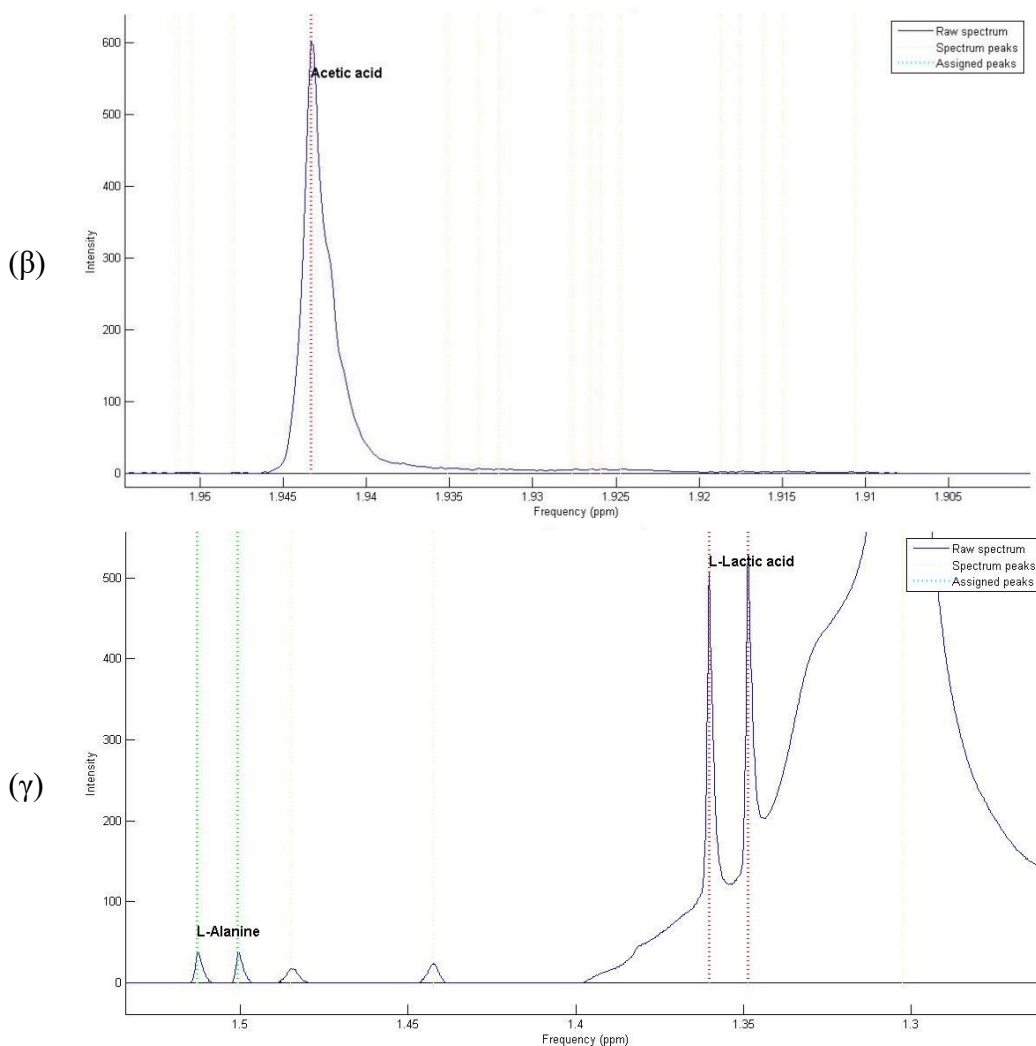
Η προτεινόμενη μεθοδολογία σημείωσε παρόμοιες τιμές ακρίβειας στις ομάδες δειγμάτων αμνιακού υγρού και ορού. Υψηλότερες τιμές ακρίβειας επιτεύχθηκαν στα δείγματα αμνιακού υγρού παρά στα δείγματα ορού, εξ' αιτίας της παρουσίας ευρέων κορυφών λιπιδίων στον ορό, που μπορούν να επικαλύψουν τις κορυφές άλλων μεταβολιτών. Παρ' όλα αυτά, σε περιπτώσεις που η κορυφή ενός λιπιδίου δεν καλύπτει εντελώς μία κορυφή, η προτεινόμενη μεθοδολογία είναι ικανή να αναγνωρίσει την κορυφή αυτή (π.χ. το Λακτικό οξύ στην Εικόνα 5.8 (γ)).

Η προτεινόμενη μεθοδολογία σημείωσε υψηλότερες τιμές ακρίβειας από το MetaboHunter και το Bayesil σε όλες τις ομάδες δειγμάτων, αποδεικνύοντας ότι είναι ένα ακόμη αξιόπιστο εργαλείο αναγνώρισης μεταβολιτών. Στους Πίνακες Π.5, Π.6 μπορεί να παρατηρηθεί ότι οι τιμές ειδικότητας είναι σταθερά χαμηλότερες από τις τιμές ακρίβειας και ευαισθησίας. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι οι τρεις μεθοδολογίες εξετάζονται για ένα σύνολο μεταβολιτών οι περισσότεροι από τους οποίους αναμένονται να υπάρχουν σε ένα βιολογικό δείγμα, επομένως ο αριθμός των αληθών αρνητικών αποτελεσμάτων είναι συνήθως χαμηλότερος ή παρόμοιος με τον αριθμό των ψευδών θετικών αποτελεσμάτων.

Πρέπει να σημειωθεί ότι η απόδοση των τριών μεθόδων στα φάσματα S1-S42 εκτιμήθηκε λαμβάνοντας υπόψη και τους 45 μεταβολίτες, ενώ στα δείγματα των μελετών MTBLS174, MTBLS424, MTBLS326, MTBLS242 η απόδοση εκτιμήθηκε λαμβάνοντας υπόψη μόνο τους μεταβολίτες που ταυτοποιήθηκαν από τους συγγραφείς τους.







Εικόνα 5.8. Παραδείγματα ανάθεσης κορυφών σε μεταβολίτες από την προτεινόμενη μεθοδολογία στο φάσμα MTBLS174\_1 από τη βάση δεδομένων Metabolights για τους μεταβολίτες (α) D-Γλυκόζη, (β) Ακετικό οξύ, (γ) L-Αλανίνη και L-Λακτικό οξύ. Οι αντίστοιχες κορυφές που ανατίθενται σε αυτούς του μεταβολίτες από το MetaboHunter είναι: (α) 5.22, 5.23 ppm (D-Γλυκόζη), (β) 1.91 ppm (Ακετικό οξύ), (γ) 1.45, 1.47 ppm (L-Αλανίνη), 1.31, 1.32 ppm (L-Λακτικό οξύ).

#### 5.4. Σχολιασμός

Στα περισσότερα δείγματα που εξετάστηκαν η ανεπτυγμένη μεθοδολογία ταυτοποίησε με ακρίβεια τους ευδιάκριτους μεταβολίτες, όπως οι D-Γλυκόζη, L-Βαλίνη, L-Ισολευκίνη, L-Λευκίνη, κλπ, καθώς και μεταβολίτες των οποίων η ταυτοποίηση μπορεί να υποβοηθηθεί λαμβάνοντας υπόψιν τις χημικές μετατοπίσεις άλλων, ήδη ταυτοποιημένων μεταβολιτών, όπως η περίπτωση της απλής κορυφής της χολίνης σε σύγκριση με την τριπλή κορυφή της D-Γλυκόζης. Από την άλλη πλευρά, η μεθοδολογία τείνει να αναγνωρίζει ψευδείς θετικές απλές κορυφές που δεν είναι απαραίτητα παρούσες στο φάσμα, όπως η Μεθανόλη, πράγμα που συμβαίνει εξ' αιτίας της έλλειψης αντικειμενικού λόγου απόρριψής τους. Για αυτό το λόγο, η αναφορά που δίνεται ως έξοδος μπορεί μελλοντικά να συμπεριλαμβάνει μία προειδοποίηση σχετικά με αυτά τα ενδεχόμενα ψευδή θετικά αποτελέσματα. Πρέπει να

διευκρινιστεί ότι προς το παρόν η μεθοδολογία είναι πλήρως αυτοματοποιημένη για τα φάσματα δειγμάτων ορού και αμνιακού υγρού.

Για να αποφευχθεί η εισαγωγή πρόσθετων ψευδών θετικών ή ψευδών αρνητικών αποτελεσμάτων υπήρξε συμβιβασμός με το γεγονός ότι η διαδικασία αναζήτησης μεταβολιτών δεν αντιλαμβάνεται φασματικές κορυφές που δεν έχουν αναγνωριστεί κατά τη διαδικασία μείωσης δεδομένων. Για παράδειγμα, αν στην περιοχή ενός quartet μόνο τρεις κορυφές έχουν αναγνωριστεί αντί για τέσσερις, ο αλγόριθμος δεν θα υποθέσει την παρουσία τέταρτης κορυφής. Αν και αυτό θα μπορούσε να ξεπεραστεί επιτρέποντας λιγότερες από το ιδανικό κορυφές να ανατεθούν σε μία πολλαπλή κορυφή, αυτό το χαρακτηριστικό δε συμπεριλήφθηκε γιατί σε άλλες περιπτώσεις θα μπορούσε να προκαλέσει πρόσθετα ψευδή θετικά αποτελέσματα. Για τον ίδιο λόγο η αποθορύβωση εφαρμόζεται φειδωλά στα φάσματα εισόδου. Αν και ισχυρότερη αποθορύβωση των φασμάτων εισόδου θα ήταν ωφέλιμη ως προς το χρόνο εκτέλεσης, θα μπορούσε επίσης να προκαλέσει την εξαφάνιση σημαντικών κορυφών, εισάγοντας πρόσθετα ψευδή αρνητικά αποτελέσματα.

Ο χρόνος εκτέλεσης κυμαίνεται από 200 σε 4,000 δευτερόλεπτα, ανάλογα με την πολυπλοκότητα του φάσματος εισόδου. Τα πειράματα διεξήχθησαν σε υπολογιστή με επεξεργαστή x64-based Intel i5 στα 2.5 GHz και 4 GB RAM. Το πρόγραμμα αναπτύχθηκε στο προγραμματιστικό περιβάλλον MATLAB 7.10.0 (MATLAB Version R2010a).

## 5.5. Αναφορές

- Anderson, P. E., Mahle, D. A., Doom, T. E., Reo, N. V., DelRaso, N. J., & Raymer, M. L. (2011). Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. *Metabolomics*, 7(2), 179–190. <http://doi.org/10.1007/s11306-010-0242-7>
- Anderson, P. E., Reo, N. V., DelRaso, N. J., Doom, T. E., & Raymer, M. L. (2008). Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics. *Metabolomics*, 4(3), 261–272. <http://doi.org/10.1007/s11306-008-0117-3>
- Experimental Errors and Uncertainty. (2002). Retrieved January 16, 2017, from [http://www.ece.rochester.edu/courses/ECE111/error\\_uncertainty.pdf](http://www.ece.rochester.edu/courses/ECE111/error_uncertainty.pdf)
- Davis, R. A., Charlton, A. J., Godward, J., Jones, S. A., Harrison, M., & Wilson, J. C. (2007). Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemometrics and Intelligent Laboratory Systems*, 85(1), 144–154. <http://doi.org/10.1016/j.chemolab.2006.08.014>
- De Meyer, T., Sinnaeve, D., Van Gasse, B., Tsiporkova, E., Rietzschel, E. R., De Buyzere, M. L., ... Van Criekinge, W. (2008). NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Analytical Chemistry*, 80(10), 3783–90. <http://doi.org/10.1021/ac7025964>
- Fotakis, C., Zoga, M., Baskakis, C., Tsiaka, T., Boutsikou, T., Briana, D. D., ... Proietto, J. (2016). Investigating the metabolic fingerprint of term infants with normal and increased fetal growth. *RSC Adv.*, 6(83), 79325–79334. <http://doi.org/10.1039/C6RA12403H>
- Gralka, E., Luchinat, C., Tenori, L., Ernst, B., Thurnheer, M., & Schultes, B. (2015). Metabolomic fingerprint of severe obesity is dynamically affected by bariatric surgery



- in a procedure-dependent manner. *American Journal of Clinical Nutrition*, 102(6), 1313–1322. <http://doi.org/10.3945/ajcn.115.110536>
- Haddad, R. A., & Akansu, A. N. (1991). A class of fast Gaussian binomial filters for speech and image processing. *IEEE Transactions on Signal Processing*, 39(3), 723–727. <http://doi.org/10.1109/78.80892>
- Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. D. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature Protocols*, 9(6), 1416–27. <http://doi.org/10.1038/nprot.2014.090>
- Hart, C. D., Vignoli, A., Tenori, L., Uy, G. L., Van To, T., Adebamowo, C., ... Di Leo, A. (2017). Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, 23(6), 1422–1431. <http://doi.org/10.1158/1078-0432.CCR-16-1153>
- Jacob, D., Deborde, C., & Moing, A. (2013). An efficient spectra processing method for metabolite identification from 1H-NMR metabolomics data. *Analytical and Bioanalytical Chemistry*, 405(15), 5049–61. <http://doi.org/10.1007/s00216-013-6852-y>
- Kale, N. S., Haug, K., Conesa, P., Jayseelan, K., Moreno, P., Rocca-Serra, P., ... Steinbeck, C. (2016). MetaboLights: An Open-Access Database Repository for Metabolomics Data. In *Current Protocols in Bioinformatics* (p. 14.13.1-14.13.18). Hoboken, NJ, USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/0471250953.bi1413s53>
- Kordalewska, M., & Markuszewski, M. J. (2015). Metabolomics in cardiovascular diseases. *Journal of Pharmaceutical and Biomedical Analysis*, 113, 121–136. <http://doi.org/10.1016/j.jpba.2015.04.021>
- Larive, C. K., Barding, G. A., & Dinges, M. M. (2015). NMR spectroscopy for metabolomics and metabolic profiling. *Analytical Chemistry*, 87(1), 133–46. <http://doi.org/10.1021/ac504075g>
- Mercier, P., Lewis, M. J., Chang, D., Baker, D., & Wishart, D. S. (2011). Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *Journal of Biomolecular NMR*, 49(3–4), 307–23. <http://doi.org/10.1007/s10858-011-9480-x>
- Nagana Gowda, G. A., Gowda, Y. N., & Raftery, D. (2015). Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. *Analytical Chemistry*, 87(1), 706–15. <http://doi.org/10.1021/ac503651e>
- Orczyk-Pawilowicz, M., Jawien, E., Deja, S., Hirnle, L., Zabek, A., & Mlynarz, P. (2016). Metabolomics of Human Amniotic Fluid and Maternal Plasma during Normal Pregnancy. *PloS One*, 11(4), e0152740. <http://doi.org/10.1371/journal.pone.0152740>
- Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., ... Wishart, D. S. (2011). The human serum metabolome. *PloS One*, 6(2), e16957. <http://doi.org/10.1371/journal.pone.0016957>
- Ravanbakhsh, S., Liu, P., Bjorndahl, T. C., Bjordahl, T. C., Mandal, R., Grant, J. R., ... Wishart, D. S. (2015). Accurate, fully-automated NMR spectral profiling for metabolomics. *PloS One*, 10(5), e0124219. <http://doi.org/10.1371/journal.pone.0124219>
- Singh, A., Sharma, R. K., Chagtoo, M., Agarwal, G., George, N., Sinha, N., & Godbole, M. M. (2017). 1H NMR Metabolomics Reveals Association of High Expression of Inositol

- 1, 4, 5 Trisphosphate Receptor and Metabolites in Breast Cancer Patients. *PloS One*, 12(1), e0169330. <http://doi.org/10.1371/journal.pone.0169330>
- Smolinska, A., Blanchet, L., Buydens, L. M. C., & Wijmenga, S. S. (2012). NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Analytica Chimica Acta*, 750, 82–97. <http://doi.org/10.1016/j.aca.2012.05.049>
- Sousa, S. A. A., Magalhães, A., & Ferreira, M. M. C. (2013). Optimized bucketing for NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory Systems*, 122, 93–102. <http://doi.org/10.1016/j.chemolab.2013.01.006>
- Tardivel, P. J. C., Canlet, C., Lefort, G., Tremblay-Franco, M., Debrauwer, L., Concordet, D., & Servien, R. (2017). ASICS: an automatic method for identification and quantification of metabolites in complex 1D <sup>1</sup>H NMR spectra. *Metabolomics*, 13(10), 109. <http://doi.org/10.1007/s11306-017-1244-5>
- Tulpan, D., Léger, S., Belliveau, L., Culf, A., & Cuperlović-Culf, M. (2011). MetaboHunter: an automatic approach for identification of metabolites from <sup>1</sup>H-NMR spectra of complex mixtures. *BMC Bioinformatics*, 12, 400. <http://doi.org/10.1186/1471-2105-12-400>
- Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., ... Scalbert, A. (2013). HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Research*, 41(Database issue), D801-7. <http://doi.org/10.1093/nar/gks1065>
- Wruck, W., Kashofer, K., Rehman, S., Daskalaki, A., Berg, D., Gralka, E., ... Ayonrinde, O. T. (2015). Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Scientific Data*, 2, 150068. <http://doi.org/10.1038/sdata.2015.68>
- Zheng, C., Zhang, S., Ragg, S., Raftery, D., & Vitek, O. (2011). Identification and quantification of metabolites in (<sup>1</sup>H) NMR spectra by Bayesian model selection. *Bioinformatics (Oxford, England)*, 27(12), 1637–44. <http://doi.org/10.1093/bioinformatics/btr118>

## Κεφάλαιο 6. Επίλογος και Μελλοντικές Επεκτάσεις

---

Το βασικό αντικείμενο της παρούσας διδακτορικής διατριβής ήταν η ανάπτυξη τεχνικών επεξεργασίας βιολογικών και χημικών δεδομένων για την επίλυση προβλημάτων στο χώρο της αντικαρκινικής θεραπείας και της μεταβολομικής. Στα πλαίσια της διδακτορικής έρευνας αναπτύχθηκε μία μεθοδολογία που αξιοποιεί ανοσολογικά και φαρμακολογικά δεδομένα με στόχο την υπολογιστική σύνθεση συμπλεγμάτων αντισωμάτων-φαρμάκων. Επίσης, αναπτύχθηκε ένα μεθοδολογικό σχήμα με αντικείμενο την αυτόματη αναγνώριση μεταβολιτών σε μονοδιάστατα φάσματα Πυρηνικού Μαγνητικού Συντονισμού πρωτονίου ( $^1\text{H}$ - NMR).

Στο κεφάλαιο 2 πραγματοποιήθηκε μία εισαγωγή στην αντικαρκινική θεραπεία και στις σχετικές εφαρμογές βιοπληροφορικής. Αρχικά, έγινε παρουσίαση κάποιων βασικών στοιχείων ανοσολογίας. Έγινε αναφορά στο ανοσοποιητικό σύστημα, στη δομή και λειτουργία των αντισωμάτων, καθώς και στις τεχνικές καθορισμού της τρισδιάστατης δομής πρωτεϊνών. Στη συνέχεια έγινε αναφορά σε βασικά χαρακτηριστικά του καρκίνου. Παρουσιάστηκαν οι κατηγορίες αντικαρκινικών θεραπειών, συμπεριλαμβανομένων των συμβατικών θεραπειών και των πιο στοχευμένων. Πραγματοποιήθηκε ειδική ανάλυση των συμπλεγμάτων αντισωμάτων-φαρμάκων, μίας κατηγορίας στοχευμένης αντικαρκινικής θεραπείας. Ακόμη, παρουσιάστηκαν στοιχεία βιοπληροφορικής που συνδέονται με την ανοσολογία και την αντικαρκινική θεραπεία, όπως η υπολογιστική τυποποίηση δεδομένων τρισδιάστατης δομής πρωτεϊνών και τις τεχνικές *in silico* φαρμακευτικού σχεδιασμού.

Στο κεφάλαιο 3 παρουσιάστηκε η μεθοδολογία που αναπτύχθηκε με αντικείμενο τον υπολογιστικό σχεδιασμό συμπλεγμάτων αντισωμάτων-φαρμάκων χρησιμοποιώντας δεδομένα από καθιερωμένες βάσεις δεδομένων. Η μέθοδος προσαρμόστηκε για δύο τρόπους σύνδεσης του συμπλέγματος συνδέτη-φαρμάκου με το αντίσωμα, συγκεκριμένα για σύζευξη σε λυσίνη και σύζευξη σε τροποποιημένη κυστεΐνη, η οποία έχει αντικαταστήσει ένα εγγενές αμινοξύ στην εξωτερική επιφάνεια του αντισώματος. Παρουσιάστηκαν τα αποτελέσματα της εφαρμογής της μεθόδου σε συγκεκριμένα αντισώματα και κυτταροτοξικές ουσίες.

Στο κεφάλαιο 4 πραγματοποιήθηκε εισαγωγή στη μεταβολομική και τη φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού. Αρχικά παρουσιάστηκαν βασικές έννοιες της μεταβολομικής, έγινε αναφορά στα πλεονεκτήματα της μεταβολομικής και τις αναλυτικές τεχνικές που χρησιμοποιούνται στα πλαίσιά της. Ακόμη έγινε αναφορά στις εφαρμογές της μεταβολομικής ανάλυσης όσον αφορά τη διάγνωση και την ανάπτυξη φαρμακευτικών προϊόντων. Στη συνέχεια παρουσιάστηκαν οι αρχές φυσικής που χαρακτηρίζουν τη φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού και τα χαρακτηριστικά του σήματος Πυρηνικού Μαγνητικού Συντονισμού. Έγινε αναφορά στη δομή του φασματογράφου Πυρηνικού Μαγνητικού Συντονισμού, στην πειραματική διαδικασία, τους τρόπους επεξεργασίας του σήματος NMR, καθώς και στις υπάρχουσες υπολογιστικές μεθόδους αναγνώρισης και ποσοτικοποίησης μεταβολιτών.

Στο κεφάλαιο 5 παρουσιάστηκε το μεθοδολογικό σχήμα που αναπτύχθηκε με αντικείμενο την αυτόματη αναγνώριση μεταβολιτών από μονοδιάστατα φάσματα  $^1\text{H}$  NMR. Έγινε αναφορά στις μεθόδους προεπεξεργασίας και μείωσης δεδομένων που εφαρμόστηκαν. Στη συνέχεια αναλύθηκαν τα συστατικά τμήματα της μεθόδου αναγνώρισης μεταβολιτών

που αναπτύχθηκε. Παρουσιάστηκε η διεργασία του screening μεταβολιτών και του τρόπου απόρριψης των μικρών κορυφών. Παρουσιάστηκαν οι συναρτήσεις βαθμολόγησης των υποψήφιων συνόλων φασματικών κορυφών που συντέθηκαν για τις διαφορετικές κατηγορίες κορυφών. Αναλύθηκε ο τρόπος επιλογής του καλύτερου συνδυασμού κορυφών για έναν μεταβολίτη και έγινε αναφορά στην προτεραιοποίηση των υποψήφιων μεταβολιτών που υιοθετήθηκε στη μεθοδολογία. Τέλος, παρουσιάστηκαν τα αποτελέσματα εφαρμογής της προτεινόμενης μεθοδολογίας σε 202 δείγματα, σε σύγκριση με άλλες δύο αντίστοιχες υπάρχουσες μεθοδολογίες.

Η εργασία που πραγματοποιήθηκε στα πλαίσια της διδακτορικής διατριβής έχει δώσει έναυσμα για να τεθούν επιπλέον ερευνητικοί στόχοι. Όσον αφορά το ερευνητικό μέρος που αφορά την αντικαρκινική θεραπεία, η μέθοδος που αναπτύχθηκε θα μπορούσε να επεκταθεί για να συμπεριλάβει τη χρήση εναλλακτικών μορφών αντισωμάτων, καθώς και άλλων συνδεδετικών μορίων. Άλλη ενδιαφέρουσα εξέλιξη θα ήταν η αυτόματη αναγνώριση των αμινοξέων ενός αντισώματος που προσφέρονται για σύζευξη, καθώς και η εκτίμηση των καλύτερων κυτταροτοξικών ουσιών και συνδεδετικών μορίων για ένα σημείο σύζευξης.

Όσον αφορά το ερευνητικό μέρος που αφορά τη μεταβολομική, σημαντικός στόχος είναι η επέκταση του λογισμικού αναγνώρισης μεταβολιτών σε μία ολοκληρωμένη διαδικτυακή πλατφόρμα επεξεργασίας φασμάτων  $^1\text{H NMR}$ , που θα συμπεριλαμβάνει τόσο την υλοποίηση της μεθοδολογίας που παρουσιάστηκε εδώ όσο και πρόσθετα βήματα προεπεξεργασίας των φασμάτων εισόδου, όπως ο μετασχηματισμός Fourier, η διόρθωση φάσης και η διόρθωση βασικής γραμμής. Άλλος στόχος είναι η ενσωμάτωση της μεθόδου ποσοτικοποίησης μεταβολιτών, αφού αυτή ολοκληρωθεί με την διεκπεραίωση περισσότερων πειραμάτων NMR. Επιπλέον, άμεσος στόχος είναι η εφαρμογή της μεθοδολογίας σε περισσότερες κατηγορίες βιολογικών δειγμάτων, όπως το μητρικό γάλα και το εγκεφαλονωτιαίο υγρό.

## Παράρτημα

Πίνακας Π.1. Συγκριτική παρουσίαση της προτεινόμενης μεθοδολογίας και των υπόλοιπων μεθοδολογιών αυτόματης αναγνώρισης μεταβολιτών ως προς τα βασικά μεθοδολογικά χαρακτηριστικά τους και τα φάσματα που χρησιμοποιήθηκαν για την αξιολόγησή τους.

Μεθοδολογία	Προεπεξεργασία	Μείωση δεδομένων	Κόρια σημεία μεθοδολογίας	Αξιολόγηση
Προτεινόμενη μεθοδολογία (Filntisi et al., 2017)*	i) Κατωφλίωση, ii) Αποθορύβωση, iii) Εντοπισμός ένωσης αναφοράς	Ενσωματωμένο Adaptive Intelligent binning	i) Screening μεταβολιτών ii) Βαθμολόγηση κορυφών (Εξισώσεις 5.12-5.26) iii) Βαθμολόγηση μεταβολιτών (Εξισώσεις 5.28-5.30) iv) Επιλογή συνδυασμών κορυφών	i) 1 δείγμα κατασκευασμένο in vitro ii) 1 spiked βιολογικό δείγμα (ορός) iii) 40 in-house βιολογικά δείγματα (ορός, αμνιακό υγρό) iv) 160 δημόσια βιολογικά δείγματα (ορός)
MetaboHunter (Tulpan et al., 2011) **	i) Κατωφλίωση (το κατώφλι θορύβου τίθεται από το χρήστη)	-	i) 1 συνάρτηση βαθμολόγησης: $\text{significance score} = \frac{\text{total number of matched peaks}}{1 + \text{total number of peaks}}$ ii) 3 μέθοδοι αναζήτησης	i) 15 συνθετικά σύνολα δεδομένων ii) 1 spiked δείγμα (ούρα) iii) 1 δείγμα κατασκευασμένο in vitro (5 μεταβολίτες)
Bayesil (Ravanbakhsh et al., 2015) **	i) Προσθήκη μηδενικών, ii) Μετασχηματισμός Fourier & Hilbert, iii) Διόρθωση φάσης, iv) Διόρθωση βασικής γραμμής, v) Αποθορύβωση, vi) Εντοπισμός ένωσης αναφοράς	-	i) Κατάτμηση του φάσματος ii) Χρήση ενός πιθανοτικού γραφικού μοντέλου iii) Χρήση μίας μεθόδου επαγωγής Monte Carlo	i) 10 δείγματα κατασκευασμένα in silico (ορός, CSF) ii) 15 δείγματα κατασκευασμένα in vitro (ορός, εγκεφαλονωτιαίο υγρό) iii) 55 βιολογικά δείγματα (ορός, CSF)
BQuant (Zheng et al., 2011) **	i) Διόρθωση βασικής γραμμής, ii) Εντοπισμός ένωσης αναφοράς iii) Κανονικοποίηση φασμάτων,	Ενσωματωμένο Adaptive Intelligent binning, Peak picking	i) Μπαγεσιανή μοντελοποίηση των φασμάτων εισόδου ii) Η αναγνώριση μεταβολιτών ανάγεται σε πρόβλημα επιλογής μεταβλητών iii) Μπαγεσιανή μοντελοποίηση, Δειγματοληψία Gibbs	i) 1 συνθετικό σύνολο δεδομένων κατασκευασμένο in silico ii) 1 spike-in δείγμα (ούρα) iii) 46 βιολογικά δείγματα (πλάσμα)

	iv) Ευθυγράμμιση φασμάτων			
Autofit (Mercier et al., 2011) **	-	-	i) Χρήση βιβλιοθηκών μεταβολιτών βασισμένων στο pH ii) Ανακατασκευή φάσματος με ένα λορεντζιανό μοντέλο iii) Συνδυασμός προσομοιωμένης ανόπτησης και γενετικού αλγορίθμου iv) Βελτιστοποίηση ελαχίστων τετραγώνων Levenberg–Marquardt	i) 100 συνθετικά δείγματα (εγκεφαλονωτιαίο υγρό, ούρα) κατασκευασμένα in silico ii) 104 βιολογικά δείγματα (ορός, εγκεφαλονωτιαίο υγρό, ούρα)
FOCUS (Alonso et al., 2014) **	(i) Κατάτμηση φάσματος (ii) Ευθυγράμμιση φάσματος	-	(i) Ανίχνευση κορυφών (ii) Αναγνώριση μεταβολιτών με intracluster, intercluster, penalization scores	i) 60 δείγματα ούρων ii) 120 υδατικά δείγματα ήπατος
(Jacob et al., 2013)**	(i) Διόρθωση βασικής γραμμής (ii) Αποθορύβωση (iii) Ευθυγράμμιση	ERVA	(i) Ομαδοποίηση κορυφών με κριτήριο την απόσταση (ii) Αντιστοίχιση κορυφών με μεταβολίτες από βάση δεδομένων	(i) 6 συνθετικά δείγματα (ii) 54 δείγματα τομάτας
* Η πλήρης αναφορά δίνεται στο τμήμα των Δημοσιεύσεων. **Οι πλήρεις αναφορές δίνονται στα κεφάλαια 4,5.				

Πίνακας Π.2. Δείγμα του αρχείου μεταβολικής πληροφορίας που συντέθηκε στα πλαίσια της παρούσας εργασίας.

Όνομα μεταβολίτη	Κατηγορία κορυφής	H	Φασματικό εύρος (ppm)	J (Hz)	Λόγοι υψών	Πλήθος υποκορυφών
L-Isoleucine	d	4	[0.894 0.957]	[7.414]	[1.76420 3.97430 2.03550]	2
	d	4	[0.973 1.027]	[7.001]	[3.87840 4.00000]	2
	m	1	[1.198 1.297]		[0.07540 0.22270 0.27060 0.27570 0.31470 0.29140 0.29380 0.34440 0.27400 0.07960]	10
	m	1	[1.400 1.506]		[0.08650 0.10940 0.26850 0.31150 0.30610 0.37020 0.26410 0.26450 0.21750 0.08610 0.06770]	11

	m	1	[1.899 2.024]		[0.04130 0.08320 0.14210 0.24230 0.24520 0.28870 0.30490 0.30070 0.28680 0.24490 0.23090 0.13900 0.08110 0.03950]	14
	d	1	[3.639 3.681]	[3.969]	[1.36840 1.39080 ]	2
L-Methionine	m	31	[2.055 2.241]		[0.05040 0.13930 0.23690 0.39190 4.00000 0.12290 0.10320 0.11580 0.20870 0.22000 0.15130 0.14900 0.09150 0.06710 0.05530 ]	15
	t	13	[2.593 2.673]	[7.587 ]	[0.60280 1.02610 0.52760 ]	3
	dd	6	[3.818 3.888]	[7.100 5.382 ]	[0.27720 0.32620 0.32900 0.27210 ]	[2 2]

Πίνακας Π.3. Ευρετήριο των πιο σημαντικών μεταβλητών και σταθερών που χρησιμοποιούνται στην περιγραφή του παρουσιαζόμενου μεθοδολογικού σχήματος.

Μεταβλητή/ Σταθερά	Περιγραφή
m	Η μεταβλητή που αντιστοιχεί σε ένα μεταβολίτη.
$L^m$	Ο συνολικός αριθμός των απλών ή πολλαπλών κορυφών (multiplets) ενός μεταβολίτη m.
$l^m$	Μία απλή ή πολλαπλή κορυφή ενός μεταβολίτη m, όπου $l^m = 1, \dots, L^m$
$N^{l^m}$	Το πλήθος των επί μέρους κορυφών μίας πολλαπλής κορυφής $l^m$
$N^m$	Το πλήθος των επί μέρους κορυφών ενός μεταβολίτη m
$area_{DB}^{l^m}$	Η περιοχή συχνοτήτων που ορίζεται για μία κορυφή $l^m$ στη βάση δεδομένων.
$area_k^{l^m}$	Μία περιοχή συχνοτήτων με το ίδιο πλάτος όπως η $area_{DB}^{l^m}$ , αλλά διαφορετικό κέντρο συχνότητας, στην οποία αναζητείται μία κορυφή $l^m$ .
$width_\mu$	Το μέσο εύρος περιοχής συχνοτήτων μίας κορυφής, όπως υπολογίζεται από 850 μεταβολίτες. ( $width_\mu = 0.07509ppm$ )
a	Το ποσοστό του $width_\mu$ που θα είναι η διαφορά συχνότητας $fd^a$ μεταξύ διαδοχικών περιοχών συχνότητας $area_k^{l^m}$ , $area_{k+1}^{l^m}$ .
K	Ο αριθμός περιοχών συχνοτήτων στις οποίες αναζητείται μία κορυφή $l^m$ ( $N^{l^m} > 1$ ).
K'	Ο αριθμός περιοχών συχνοτήτων στις οποίες αναζητείται μία απλή κορυφή (singlet) $l^m$ ( $N^{l^m} = 1$ ).
k	Μία μεταβλητή που παίρνει ακέραιες τιμές στο εύρος [1 K].
$offset^{a,K}$	Η ποσότητα που αναπαριστά την τροποποίηση της περιοχής συχνοτήτων $area_{DB}^{l^m}$ μίας κορυφής $l^m$ ( $N^{l^m} > 1$ ) για $k = K$ .
$offset^{a,K'}$	Η ποσότητα που αναπαριστά την τροποποίηση της περιοχής συχνοτήτων $area_{DB}^{l^m}$ μίας απλής κορυφής (singlet) $l^m$ ( $N^{l^m} = 1$ ) για $k = K'$ .
$fr^{a,K}$	Το ολικό εύρος συχνοτήτων στο οποίο αναζητείται μία κορυφή $l^m$ .
$fd^a$	Η διαφορά συχνότητας μεταξύ δύο διαδοχικών περιοχών συχνότητας $area_k^{l^m}$ , $area_{k+1}^{l^m}$

$V_{bp}$	Η ποσοστιαία διαφορά μεταξύ της μέσης έντασης των επί μέρους κορυφών μίας κορυφής $I^m$ και της μέσης έντασης των ορίων $bip$ που αντιστοιχούν σε αυτές.
$P_k^{I^m}$	Το σύνολο κορυφών που επιλέγεται για μία κορυφή $I^m$ στην περιοχή συχνοτήτων $area_k^{I^m}$ .
$Score_k^{I^m}$	Η βαθμολογία του συνόλου κορυφών $P_k^{I^m}$ .
$p_n$	Μία επιμέρους κορυφή της κορυφής $I^m$ .
$f_n$	Η συχνότητα της κορυφής $p_n$ .
$h_n$	Η ένταση της κορυφής $p_n$ .
$h_{DBn}$	Ο ιδανικός λόγος ύψους που ορίζεται στη βάση δεδομένων για την κορυφή $p_n$
$Score_S$	Η συνάρτηση βαθμολόγησης που χρησιμοποιείται για την εκτίμηση της καταλληλότητας μίας υποψήφιας κορυφής για μία απλή κορυφή (singlet).
$Score_A$	Η συνάρτηση βαθμολόγησης που χρησιμοποιείται για την εκτίμηση της καταλληλότητας ενός υποψήφιου συνόλου κορυφών για μία πολλαπλή κορυφή (multiplet) Α τάξης.
$Score_X$	Η συνάρτηση βαθμολόγησης που χρησιμοποιείται για την εκτίμηση της καταλληλότητας ενός υποψήφιου συνόλου κορυφών για μία πολλαπλή κορυφή (multiplet) Χ τάξης.
$Score_M$	Η συνάρτηση βαθμολόγησης που χρησιμοποιείται για την εκτίμηση της καταλληλότητας ενός υποψήφιου συνόλου κορυφών για μία πολλαπλή κορυφή (multiplet) χωρίς κανόνες.
$kc^m$	Ένας συνδυασμός τιμών $k$ για ένα μεταβολίτη $m$ .
$kd^m$	Η μέγιστη διαφορά μεταξύ οποιωνδήποτε δύο τιμών $k$ που μπορεί να έχει ένας συνδυασμός κορυφών $kc^m$ .
$Score^{kc^m}$	Η συνάρτηση βαθμολόγησης που χρησιμοποιείται για την εκτίμηση της καταλληλότητας ενός υποψήφιου συνδυασμού κορυφών $kc^m$ για ένα μεταβολίτη $m$ .

Πίνακας Π.4. Δοκιμασμένοι συνδυασμοί τιμών παραμέτρων και οι αντίστοιχες τιμές  $fd^a$ ,  $offset^{a,K}$ ,  $offset^{a,K'}$ .

$\alpha$	$K$	$K'$	$fd^a$	$offset^{a,K}$	$offset^{a,K'}$
0.10	46	28	0.0075089	0.1727047	0.1051246
0.12	38	24	0.0090108	0.17120292	0.1081282
0.14	32	20	0.01051246	0.16819936	0.1051246
0.16	28	18	0.01201424	0.16819936	0.1081282
0.18	24	16	0.01351602	0.16219224	0.1081282
0.20	22	14	0.0150178	0.1651958	0.1051246
0.22	20	12	0.0165196	0.1651958	0.0991175
0.24	18	12	0.0180214	0.16219224	0.1081282



Πίνακας Π.5. Τα αποτελέσματα ακρίβειας, ευαισθησίας και ειδικότητας της προτεινόμενης μεθοδολογίας ( $\alpha=0.12$ ), του MetaboHunter και του Bayesil στα δείγματα S1-S42.

Δείγμα	Προτεινόμενη Μεθοδολογία			MetaboHunter			Bayesil		
	Ακρίβεια (%)	Ευαισθησία (%)	Ειδικότητα (%)	Ακρίβεια (%)	Ευαισθησία (%)	Ειδικότητα (%)	Ακρίβεια (%)	Ευαισθησία (%)	Ειδικότητα (%)
S1	64.286	100	54.545	42.222	88.889	30.556	26.83	100	9.09
S2	81.818	96.552	53.333	36.585	87.5	4	58.54	68.57	0
S3	88.889	96.875	69.231	50	95	12.5	71.79	81.82	16.67
S4	86.667	93.939	66.667	50	95	12.5	63.16	70.97	28.57
S5	86.667	96.875	61.538				73.68	83.87	28.57
S6	82.222	96.774	50				79.49	90.63	28.57
S7	84.444	100	53.333				73.68	81.25	33.33
S8	80	100	52.632				75.68	83.87	33.33
S9	75.556	96.429	41.176				71.05	77.42	42.86
S10	82.222	96.667	53.333	50	95	12.5	65.79	71.88	33.33
S11	77.778	96.429	47.059	50	95	12.5	68.42	75	33.33
S12	73.333	92.308	47.368	50	95	12.5	67.57	74.19	33.33
S13	77.778	86.207	62.5				73.68	80.65	42.86
S14	84.444	87.5	76.923				73.68	81.25	33.33
S15	84.444	96.774	57.143				71.05	80.65	28.57
S16	84.444	96.667	60				65.79	75	16.67
S17	84.444	93.548	64.286				72.22	77.42	40
S18	80	93.103	56.25				70.27	77.42	33.33
S19	77.778	96.552	43.75				62.16	70	28.57
S20	73.333	96.429	35.294				70.27	75	40
S21	82.222	96.552	56.25				70.27	77.42	33.33
S22	82.222	90.323	64.286				71.05	78.13	33.33
S23	88.636	100	61.538	47.727	94.737	0.12	67.5	78.13	25
S24	77.273	96	52.632	47.727	94.737	0.12			
S25	77.273	100	47.368	47.727	94.737	0.12	56.1	60	45.45
S26	84.091	100	50				58.54	51.72	75
S27	81.818	100	50						
S28	77.273	100	44.444						
S29	79.545	100	47.059				46.34	46.67	45.45
S30	81.818	100	50	47.727	94.737	0.12	78.05	93.55	30
S31	77.273	96.296	47.059	47.727	94.737	0.12			
S32	70.455	100	38.095						
S33	75	96.154	44.444						
S34	79.545	96	57.895						

S35	75	100	38.889						
S36	77.273	88.889	58.824				60.98	54.84	80
S37	79.545	100	52.632				56.1	58.62	50
S38	72.727	92	47.368				56.1	64.52	30
S39	81.818	96.429	56.25				75	90.32	22.22
S40	75	84.615	61.111				63.41	74.19	30
S41	68.182	91.667	40				60.98	58.06	70
S42	65.909	80.769	44.444				63.41	60	72.73

\*Η εκτέλεση του Bayesil για το φάσμα εισόδου των δειγμάτων S24, S27-S28, S31-S35 απέτυχε επανειλημμένα.

Πίνακας Π.6. Τα αποτελέσματα ακρίβειας, ευαισθησίας και ειδικότητας της προτεινόμενης μεθοδολογίας ( $\alpha=0.12$ ), του MetaboHunter και του Bayesil στα δείγματα των μελετών MTBLS174, MTBLS424, MTBLS326, MTBLS242.

Δείγμα	Προτεινόμενη Μεθοδολογία			MetaboHunter			Bayesil		
	Ακρίβεια (%)	Ευαισθησία (%)	Ειδικότητα (%)	Ακρίβεια (%)	Ευαισθησία (%)	Ειδικότητα (%)	Ακρίβεια (%)	Ευαισθησία (%)	Ειδικότητα (%)
MTBLS174_1	92.857	100	0	33.333	83.333	0	81.25	81.25	-
MTBLS174_2	100	100	-	33.333	83.333	0	75	75	-
MTBLS174_3	100	100	-	33.333	83.333	0	81.25	81.25	-
MTBLS174_4	100	100	-	33.333	83.333	0	81.25	81.25	-
MTBLS174_5	85.714	85.714	-	33.333	83.333	0	86.667	86.667	-
MTBLS174_6	92.857	100	0				80	80	-
MTBLS174_7	78.571	84.615	0				60	60	-
MTBLS174_8	100	100	-				80	80	-
MTBLS174_9	85.714	100	0				86.667	86.667	-
MTBLS174_10	92.857	92.857	-				60	60	-
MTBLS174_11	71.429	90.909	0				53.333	53.333	-
MTBLS174_12	64.286	75	0				26.667	26.667	-
MTBLS174_13	85.714	100	0				46.667	46.667	-
MTBLS174_14	100	100	-				60	60	-
MTBLS174_15	92.857	100	0				80	80	-
MTBLS174_16	85.714	100	0				53.333	53.333	-
MTBLS174_17	78.571	91.667	0				53.333	53.333	-
MTBLS174_18	85.714	100	33.333				81.25	81.25	-

<b>MTBLS424_1 (BC_SE_15002)</b>	81.818	90	0	86.364	95	0	71.429	71.429	-
<b>MTBLS424_2 (BC_SE_15003)</b>	90.909	95.238	0	90.909	95.238	0	47.619	47.619	-
<b>MTBLS424_3 (BC_SE_15004)</b>	81.818	94.737	0	90.909	95.238	0	80.952	85	0
<b>MTBLS424_4 (BC_SE_15005)</b>	72.727	88.889	0				66.667	66.667	-
<b>MTBLS424_5 (BC_SE_15006)</b>	86.364	90.476	0				80.952	80.952	-
<b>MTBLS424_6 (BC_SE_15007)</b>	90.909	95.238	0				85.714	85.714	-
<b>MTBLS424_7 (BC_SE_15008)</b>	72.727	88.235	20				90.476	90.476	-
<b>MTBLS424_8 (BC_SE_15009)</b>	72.727	94.118	0				61.905	61.905	-
<b>MTBLS424_9 (BC_SE_15011)</b>	72.727	88.889	0				66.667	66.667	-
<b>MTBLS424_10 (BC_SE_15012)</b>	86.364	95	0				80.952	80.952	-
<b>MTBLS424_11 (BC_SE_15013)</b>	77.273	88.235	40				42.857	42.857	-
<b>MTBLS424_12 (BC_SE_15014)</b>	77.273	89.474	0				66.667	66.667	-
<b>MTBLS424_13 (BC_SE_15015)</b>	68.182	87.5	16.667				66.667	66.667	-
<b>MTBLS424_14 (BC_SE_15016)</b>	86.364	95	0				66.667	66.667	-
<b>MTBLS424_15 (BC_SE_15017)</b>	72.727	88.889	0				76.19	76.19	-
<b>MTBLS424_16 (BC_SE_15018)</b>	81.818	90	0				80.952	80.952	-
<b>MTBLS424_17 (BC_SE_15019)</b>	63.636	82.353	0				76.19	76.19	-
<b>MTBLS424_18 (BC_SE_15020)</b>	77.273	100	0				76.19	76.19	-
<b>MTBLS424_19 (BC_SE_15021)</b>	81.818	90	0				47.619	47.619	-
<b>MTBLS424_20 (BC_SE_15022)</b>	68.182	78.947	0				66.667	66.667	-
<b>MTBLS424_21 (BC_SE_15023)</b>	68.182	83.333	0				61.905	61.905	-
<b>MTBLS424_22 (BC_SE_15024)</b>	77.273	89.474	0				80.952	80.952	-
<b>MTBLS424_23 (BC_SE_15025)</b>	86.364	95	0				95.238	95.238	-
<b>MTBLS424_24 (BC_SE_15026)</b>	81.818	94.737	0				76.19	76.19	-
<b>MTBLS424_25 (BC_SE_15030)</b>	81.818	94.737	0				80.952	80.952	-
<b>MTBLS424_26 (BC_SE_15031)</b>	86.364	95	0	90.909	95.238	0	80.952	80.952	-
<b>MTBLS424_27 (BC_SE_15032)</b>	72.727	94.118	0	90.909	95.238	0	76.19	76.19	-

<b>MTBLS424_28 (BC_SE_15034)</b>	77.273	89.474	0	90.909	95.238	0	80.952	80.952	-
<b>MTBLS424_29 (BC_SE_15035)</b>	77.273	89.474	0				71.429	71.429	-
<b>MTBLS424_30 (BC_SE_15039)</b>	68.182	88.235	0				71.429	71.429	-
<b>MTBLS424_31 (BC_SE_15040)</b>	72.727	94.118	0				66.667	66.667	-
<b>MTBLS424_32 (BC_SE_15042)</b>	77.273	94.444	0				71.429	71.429	-
<b>MTBLS424_33 (BC_SE_15043)</b>	77.273	94.444	0				85.714	85.714	-
<b>MTBLS424_34 (BC_SE_15044)</b>	59.091	86.667	0				80.952	80.952	-
<b>MTBLS424_35 (BC_SE_15046)</b>	81.818	89.474	33.333				85.714	85.714	-
<b>MTBLS424_36 (BC_SE_15047)</b>	63.636	87.5	0				80.952	80.952	-
<b>MTBLS424_37 (BC_SE_15049)</b>	63.636	86.667	14.286				90.476	90.476	-
<b>MTBLS424_38 (BC_SE_15050)</b>	72.727	88.889	0				80.952	80.952	-
<b>MTBLS424_39 (BC_SE_15051)</b>	72.727	88.889	0				76.19	76.19	-
<b>MTBLS424_40 (BC_SE_15052)</b>	68.182	83.333	0				71.429	71.429	-
<b>MTBLS424_41 (BC_SE_15053)</b>	81.818	94.737	0				80.952	80.952	-
<b>MTBLS424_42 (BC_SE_15054)</b>	72.727	88.889	0				76.19	76.19	-
<b>MTBLS424_43 (BC_SE_15055)</b>	72.727	94.118	0				95.238	95.238	-
<b>MTBLS424_44 (BC_SE_15056)</b>	77.273	89.474	0				85.714	85.714	-
<b>MTBLS424_45 (BC_SE_15057)</b>	68.182	75	0				57.143	57.143	-
<b>MTBLS424_46 (BC_SE_15058)</b>	72.727	94.118	0				71.429	71.429	-
<b>MTBLS424_47 (BC_SE_15059)</b>	81.818	94.737	0				71.429	71.429	-
<b>MTBLS424_48 (BC_SE_15060)</b>	81.818	94.737	0				76.19	76.19	-
<b>MTBLS424_49 (BC_SE_15061)</b>	68.182	93.75	0				76.19	76.19	-
<b>MTBLS424_50 (BC_SE_15062)</b>	81.818	90	0				80.952	80.952	-
<b>MTBLS326_1 (bc_1)</b>	83.333	83.333	-	83.333	100	0	100	100	-
<b>MTBLS326_2 (bc_2)</b>	83.333	100	0	83.333	100	0	83.333	83.333	-
<b>MTBLS326_3 (bc_3)</b>	83.333	80	100	66.67	100	0	100	100	-

MTBLS326_4 (bc_4)	66.667	80	0				83.333	83.333	-
MTBLS326_5 (bc_5)	83.333	83.333	-				83.333	83.333	-
MTBLS326_6 (bc_6)	83.333	100	0				100	100	-
MTBLS326_7 (bc_7)	83.333	83.333	-				66.667	66.667	-
MTBLS326_8 (bc_8)	83.333	83.333	-				66.667	66.667	-
MTBLS326_9 (bc_9)	83.333	100	0				66.667	66.667	-
MTBLS326_10 (bc_10)	66.667	66.667	-				66.667	66.667	-
MTBLS326_11 (bc_11)	66.667	80	0				83.333	83.333	-
MTBLS326_12 (bc_12)	66.667	80	0				50	50	-
MTBLS326_13 (bc_13)	83.333	100	0				66.667	66.667	-
MTBLS326_14 (bc_14)	83.333	83.333	-				66.667	66.667	-
MTBLS326_15 (bc_15)	83.333	100	0				66.667	66.667	-
MTBLS326_16 (bc_16)	83.333	83.333	-				83.333	83.333	-
MTBLS326_17 (bc_17)	83.333	83.333	-				83.333	83.333	-
MTBLS326_18 (bc_18)	83.333	100	0				50	50	-
MTBLS326_19 (bc_19)	83.333	100	0				83.333	83.333	-
MTBLS326_20 (bc_20)	83.333	83.333	-				83.333	83.333	-
MTBLS326_21 (bc_21)	50	50	-				66.667	66.667	-
MTBLS326_22 (bc_22)	83.333	100	0				100	100	-
MTBLS326_23 (bc_23)	66.667	80	0				83.333	83.333	-
MTBLS326_24 (bc_24)	66.667	100	0				83.333	83.333	-
MTBLS326_25 (bc_25)	66.667	80	0				100	100	-
MTBLS326_26 (bc_26)	83.333	100	0				83.333	83.333	-
MTBLS326_27 (bc_27)	83.333	100	0				66.667	66.667	-
MTBLS326_28 (c_1)	66.667	80	0	83.333	100	0	83.333	83.333	-
MTBLS326_29 (c_2)	66.667	80	0	83.333	100	0	83.333	83.333	-
MTBLS326_30 (c_3)	83.333	100	0				83.333	83.333	-

MTBLS326_31 (c_4)	83.333	100	0				100	100	-
MTBLS326_32 (c_5)	66.667	80	0				50	50	-
MTBLS326_33 (c_6)	66.667	66.667	-				50	50	-
MTBLS326_34 (c_7)	83.333	100	0				66.667	66.667	-
MTBLS326_35 (c_8)	66.667	66.667	-				100	100	-
MTBLS326_36 (c_9)	83.333	100	0				66.667	66.667	-
MTBLS326_37 (c_10)	66.667	80	0				83.333	83.333	-
MTBLS326_38 (c_11)	66.667	80	0				66.667	66.667	-
MTBLS326_39 (c_12)	66.667	66.667	-				100	100	-
MTBLS326_40 (c_13)	83.333	83.333	-				83.333	83.333	-
MTBLS326_41 (c_14)	66.667	80	0				83.333	83.333	-
MTBLS326_42 (c_15)	83.333	83.333	-				83.333	83.333	-
MTBLS242_1 (Obs0_0001s)	81.25	92.857	0	29.412	71.429	0	63.158	63.158	-
MTBLS242_2 (Obs0_0002s)	81.25	92.857	0	29.412	71.429	0	89.474	89.474	-
MTBLS242_3 (Obs0_0003s)	81.25	86.667	0	23.529	66.667	0	57.895	57.895	-
MTBLS242_4 (Obs0_0004s)	68.75	84.615	0	29.412	71.429	0	78.947	78.947	-
MTBLS242_5 (Obs0_0008s)	75	85.714	0	29.412	71.429	0	68.421	68.421	-
MTBLS242_6 (Obs0_0009s)	81.25	92.857	0				94.737	94.737	-
MTBLS242_7 (Obs0_0012s)	87.5	100	0				78.947	78.947	-
MTBLS242_8 (Obs0_0014s)	75	85.714	0				57.895	57.895	-
MTBLS242_9 (Obs0_0025s)	75	80	0				73.684	73.684	-
MTBLS242_10 (Obs0_0027s)	75	92.308	0				68.421	68.421	-
MTBLS242_11 (Obs0_0029s)	68.75	84.615	0				63.158	63.158	-
MTBLS242_12 (Obs0_0030s)	62.5	83.333	0	23.529	66.667	0	78.947	78.947	-
MTBLS242_13 (Obs0_0033s)	81.25	92.857	0	23.529	66.667	0	73.684	73.684	-
MTBLS242_14 (Obs0_0037s)	75	85.714	0				68.421	68.421	-

MTBLS242_15 (Obs0_0038s)	75	92.308	0				73.684	73.684	-
MTBLS242_16 (Obs0_0041s)	75	80	0				52.632	52.632	-
MTBLS242_17 (Obs0_0045s)	75	92.308	0				57.895	57.895	-
MTBLS242_18 (Obs0_0046s)	75	85.714	0				84.211	84.211	-
MTBLS242_19 (Obs0_0052s)	87.5	93.333	0				57.895	57.895	-
MTBLS242_20 (Obs0_0053s)	68.75	84.615	0				68.421	68.421	-
MTBLS242_21 (Obs0_0054s)	81.25	100	0				68.421	68.421	-
MTBLS242_22 (Obs0_0055s)	87.5	93.333	0				78.947	78.947	-
MTBLS242_23 (Obs0_0058s)	68.75	84.615	0				68.421	68.421	-
MTBLS242_24 (Obs0_0063s)	68.75	91.667	0				63.158	63.158	-
MTBLS242_25 (Obs0_0069s)	68.75	91.667	0				73.684	73.684	-
MTBLS242_26 (Obs0_0073s)	75	85.714	0				73.684	73.684	-
MTBLS242_27 (Obs0_0079s)	75	92.308	0				73.684	73.684	-
MTBLS242_28 (Obs0_0081s)	68.75	84.615	0				63.158	63.158	-
MTBLS242_29 (Obs0_0089s)	68.75	78.571	0				68.421	68.421	-
MTBLS242_30 (Obs0_0092s)	75	85.714	0				78.947	78.947	-
MTBLS242_31 (Obs0_0095s)	81.25	86.667	0				57.895	57.895	-
MTBLS242_32 (Obs0_0099s)	62.5	90.909	0				63.158	63.158	-
MTBLS242_33 (Obs0_0100s)	75	92.308	0				68.421	68.421	-
MTBLS242_34 (Obs0_0103s)	81.25	100	0				68.421	68.421	-
MTBLS242_35 (Obs0_0106s)	75	100	0				73.684	73.684	-
MTBLS242_36 (Obs0_0112s)	81.25	100	0				68.421	68.421	-
MTBLS242_37 (Obs0_0114s)	75	100	0				78.947	78.947	-
MTBLS242_38 (Obs0_0116s)	75	92.308	0				63.158	63.158	-
MTBLS242_39 (Obs0_0117s)	75	85.714	0				63.158	63.158	-
MTBLS242_40 (Obs0_0118s)	75	92.308	0				57.895	57.895	-
MTBLS242_41 (Obs0_0134s)	75	92.308	0				73.684	73.684	-

MTBLS242_42 (Obs0_0135s)	75	85.714	0				78.947	78.947	-
MTBLS242_43 (Obs0_0139s)	75	85.714	0				63.158	63.158	-
MTBLS242_44 (Obs0_0144s)	81.25	86.667	0				68.421	68.421	-
MTBLS242_45 (Obs0_0153s)	56.25	81.818	0				63.158	63.158	-
MTBLS242_46 (Obs0_0154s)	75	85.714	0				68.421	68.421	-
MTBLS242_47 (Obs0_0155s)	75	92.308	0				73.684	73.684	-
MTBLS242_48 (Obs0_0161s)	81.25	92.857	0				68.421	68.421	-
MTBLS242_49 (Obs0_0163s)	75	85.714	0				52.632	52.632	-
MTBLS242_50 (Obs0_0170s)	81.25	92.857	0				52.632	52.632	-

Πίνακας Π.7. Αποτελέσματα της μεθόδου Autofit που συμπεριλαμβάνεται στο λογισμικό Chenomx 8.3.

Δείγματα		Αριθμός ταυτοποιημένων μεταβολιτών / Αριθμός μεταβολιτών προς ταυτοποίηση
<b>S1 (Δείγμα αμινοξέων)</b>		5/9
<b>MTBLS174</b>	<b>1</b>	6/14
	<b>2</b>	5/14
	<b>3</b>	4/14
	<b>4</b>	3/14
	<b>5</b>	2/14
	<b>6</b>	6/14
	<b>7</b>	4/14
	<b>8</b>	1/14
	<b>9</b>	3/14
	<b>10</b>	0/14
	<b>11</b>	3/14
	<b>12</b>	1/14
	<b>13</b>	2/14
	<b>14</b>	1/14
	<b>15</b>	2/14
	<b>16</b>	2/14
	<b>17</b>	1/14
	<b>18</b>	1/14
	<b>15012</b>	0/22
	<b>15013</b>	0/22
	<b>65002</b>	1/22



	<b>75002</b>	2/22
	<b>85005</b>	2/22
	<b>330006</b>	2/22
	<b>15064</b>	2/22
	<b>16083</b>	2/22
	<b>309009</b>	4/22
	<b>25029</b>	5/22
	<b>15002</b>	3/22
	<b>15003</b>	6/22
	<b>15004</b>	4/22
	<b>15005</b>	0/22
	<b>15006</b>	0/22
	<b>15007</b>	3/22
	<b>15008</b>	1/22
	<b>15009</b>	2/22
	<b>15011</b>	4/22
	<b>15014</b>	1/22
	<b>15016</b>	1/22
	<b>15017</b>	0/22
	<b>15018</b>	2/22
	<b>15019</b>	0/22
	<b>15020</b>	0/22
<b>MTBLS242</b>	<b>Obs0_0012s</b>	2/16
	<b>Obs0_0013s</b>	0/16
	<b>Obs0_0064s</b>	1/16
	<b>Obs0_0083s</b>	2/16
	<b>Obs0_0011s</b>	4/16
	<b>Obs0_0001s</b>	0/16
	<b>Obs0_0002s</b>	1/16
	<b>Obs0_0003s</b>	0/16
	<b>Obs0_0004s</b>	1/16
	<b>Obs0_0008s</b>	1/16
	<b>Obs0_0009s</b>	0/16
	<b>Obs0_0025s</b>	0/16
	<b>Obs0_0027s</b>	1/16
	<b>Obs0_0029s</b>	1/16
	<b>Obs0_0030s</b>	1/16
	<b>Obs1_0010s</b>	1/16
	<b>Obs1_0248s</b>	1/16
	<b>Obs1_0268s</b>	2/16



## Δημοσιεύσεις

---

### Διεθνή περιοδικά

Filntisi, A., Fotakis, C., Asvestas, P., Matsopoulos, G. K., Zoumpoulakis, P., & Cavouras, D. (2017). Automated metabolite identification from biological fluid 1H NMR spectra. *Metabolomics*, 13(12), 146. <http://doi.org/10.1007/s11306-017-1286-8>

Filntisi, A., Vlachakis, D., Matsopoulos, G. K., & Kossida, S. (2014). Computational Construction of Antibody-Drug Conjugates Using Surface Lysines as the Antibody Conjugation Site and a Non-cleavable Linker. *Cancer Informatics*, 13, 179–86. <http://doi.org/10.4137/CIN.S19222>

Filntisi, A., Vlachakis, D., Matsopoulos, G., & Kossida, S. (2013). 3D Structural Bioinformatics of Proteins and Antibodies: State of the Art Perspectives and Challenges. *International Journal of Systems Biology and Biomedical Technologies*, 2(3), 67–74. <http://doi.org/10.4018/ijsbbt.2013070105>

Filntisi, A., Papangelopoulos, N., Bencurova, E., Kasampalidis, I., Matsopoulos, G., Vlachakis, D., & Kossida, S. (2013). State-of-the-Art Neural Networks Applications in Biology. *International Journal of Systems Biology and Biomedical Technologies*, 2(4), 63–85. <http://doi.org/10.4018/ijsbbt.2013100105>

Loukatou, S., Papageorgiou, L., Fakourelis, P., Filntisi, A., Polychronidou, E., Bassis, I., ... Kossida, S. (2014, June 30). Molecular dynamics simulations through GPU video games technologies. *Journal of Molecular Biochemistry*. Retrieved from <http://www.jmolbiochem.com/index.php/JmolBiochem/article/view/122>

Papangelopoulos, N., Vlachakis, D., Filntisi, A., Fakourelis, P., Papageorgiou, L., Megalooikonomou, V., & Kossida, S. (2013). State-of-the-Art GPGPU Applications in Bioinformatics. *International Journal of Systems Biology and Biomedical Technologies*, 2(4), 24–48. <http://doi.org/10.4018/ijsbbt.2013100103>

### Κεφάλαια βιβλίων

Filntisi, A., Vlachakis, D., & Matsopoulos, G. K. (2016). A Bioinformatics Method for the Production of Antibody-Drug Conjugates Through Site-Specific Cysteine Conjugation. In *Bioinformatics - Updated Features and Applications*. InTech. <http://doi.org/10.5772/62747>

### Διεθνή Συνέδρια

Filntisi, A., Asvestas, P., Fotakis, C., Zoumpoulakis, P., Matsopoulos, G., & Cavouras, D. (2016). Automated Identification of Molecular Structures for NMR Based Metabolomics. *BIOTECHNO 2016: The Eighth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies*. Lisbon, Portugal, June 26 - 30, 2016. (pp. 35-36).

Filntisi, A., Vlachakis, D., Matsopoulos, G., Kossida, S. (2015). Computational Construction of Antibody-Drug Conjugates with Lysine and Cysteine Conjugation [Presentation]. BIOTECHNO 2015: The Seventh International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies. Rome, Italy, May 24 - 29, 2015.