UNIVERSITY OF IOANNINA

SCHOOL OF PHYSICAL SCIENCES

DEPARTMENT OF CHEMISTRY

INVESTIGATING THE FEASIBILITY OF SPECTROSCOPIC TECHNIQUES AND NON-INVASIVE SAMPLING OF BIOFLUIDS TO EXPLORE PHYSIOLOGICAL ALTERATIONS DURING PHYSICAL EXERCISE

CHRISTOFOROS CHRIMATOPOULOS

CHEMIST, MSc

DOCTORAL THESIS

2025

IOANNINA

UNIVERSITY OF IOANNINA

SCHOOL OF PHYSICAL SCIENCES

DEPARTMENT OF CHEMISTRY

INVESTIGATING THE FEASIBILITY OF SPECTROSCOPIC TECHNIQUES AND NON-INVASIVE SAMPLING OF BIOFLUIDS TO EXPLORE PHYSIOLOGICAL ALTERATIONS DURING PHYSICAL EXERCISE

CHRISTOFOROS CHRIMATOPOULOS

CHEMIST, MSc

DOCTORAL THESIS

2025

IOANNINA

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΔΙΕΡΕΥΝΗΣΗ ΑΝΑΛΥΤΙΚΩΝ ΜΕΘΟΔΟΛΟΓΙΩΝ ΒΑΣΙΣΜΕΝΩΝ ΣΕ ΜΕΘΟΔΟΥΣ ΜΗ ΕΠΕΜΒΑΤΙΚΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΒΙΟΛΟΓΙΚΩΝ ΥΓΡΩΝ ΚΑΙ ΤΕΧΝΙΚΕΣ ΦΑΣΜΑΤΟΣΚΟΠΙΑΣ ΓΙΑ ΤΗ ΜΕΛΕΤΗ ΦΥΣΙΟΛΟΓΙΚΩΝ ΠΡΟΣΑΡΜΟΓΩΝ ΚΑΤΑ ΤΗ ΣΩΜΑΤΙΚΗ ΑΣΚΗΣΗ

ΧΡΙΣΤΟΦΟΡΟΣ ΧΡΗΜΑΤΟΠΟΥΛΟΣ

ΧΗΜΙΚΟΣ, MSc

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

2025

ΙΩΑΝΝΙΝΑ

«Η έγκριση της διδακτορικής διατριβής από το Τμήμα Χημείας της Σχολής Θετικών Επιστημών, του Πανεπιστημίου Ιωαννίνων δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα Ν. 5343/32, άρθρο 202, παράγραφος 2»

Ορισμός Τριμελούς Συμβουλευτικής Επιτροπής από τη Συνέλευση: **1032/12-02-2021**

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπων:

Βασίλειος Σακκάς, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων

Μέλη:

Μάμαντος Προδρομίδης, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων

Κωνσταντίνος Σταλίκας, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων

Ημερομηνία ορισμού θέματος: 22-04-2021

Θέμα: «**Investigating the feasibility of spectroscopic techniques and non-invasive sampling of biofluids to explore physiological alterations during physical exercise**»

ΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ από τη Συνέλευση: 1145/30-06-2025

1. Βασίλειος Σακκάς, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων
2. Μάμαντος Προδρομίδης, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων
3. Κωνσταντίνος Σταλίκας, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων
4. Αθανάσιος Βλεσσίδης, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων
5. Δημοσθένης Γκιώκας, Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων
6. Νικόλαος Κουρκουμέλης, Καθηγητής, Τμήμα Ιατρικής, Πανεπιστήμιο Ιωαννίνων
7. Παντελεήμων Τάκης, Επίκουρος Καθηγητής, Τμήμα Χημείας, Πανεπιστήμιο Ιωαννίνων

Έγκριση Διδακτορικής Διατριβής με βαθμό «ΑΡΙΣΤΑ» (Δέκα-10) στις 10-07-2025

Η Πρόεδρος του Τμήματος Χημείας
Δήμητρα Χελά, Καθηγήτρια

Η Γραμματέας του Τμήματος
Ξανθή Τουτουνζόγλου

Ευχαριστίες

Η παρούσα διδακτορική διατριβή αποτελεί το επιστέγασμα μιας πολύχρονης διαδρομής, γεμάτης προκλήσεις, αμφισβητήσεις, αναζητήσεις αλλά και έντονες στιγμές δημιουργικότητας και προσωπικής εξέλιξης. Η εκπόνησή της πραγματοποιήθηκε στο Εργαστήριο Αναλυτικής Χημείας του Τμήματος Χημείας του Πανεπιστημίου Ιωαννίνων και αποτελεί την πιο ώριμη και ουσιαστική φάση της επιστημονικής μου πορείας μέχρι σήμερα.

Θα ήθελα πρώτα και πάνω από όλα να εκφράσω την ειλικρινή μου ευγνωμοσύνη στον επιβλέποντα καθηγητή μου, Καθηγητή του Τμήματος Χημείας του Πανεπιστημίου Ιωαννίνων κ. Βασίλειο Σακκά, για την πολύτιμη καθοδήγηση, τη διαρκή υποστήριξη και την εμπιστοσύνη που μου έδειξε καθ' όλη τη διάρκεια της διατριβής. Η συνέπειά του, το επιστημονικό του ήθος και η αφοσίωσή του στην έρευνα αποτέλεσαν φάρο καθοδήγησης, αλλά και πηγή έμπνευσης. Η ευκαιρία να εργαστώ υπό την επίβλεψή του με τιμά βαθύτατα.

Θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή του Τμήματος Χημείας του Πανεπιστημίου Ιωαννίνων κ. Μάμαντο Προδρομίδη. για τη ουσιαστική παρουσία του καθ' όλη τη διάρκεια της διατριβής. Με την επιστημονική του εμπειρία, την οξυδέρκεια και την ικανότητά του να εστιάζει στην ουσία των πραγμάτων, προσέφερε πολύτιμη καθοδήγηση και στήριξη. Η συμβολή του, μέσα από εύστοχες παρατηρήσεις και γόνιμες συζητήσεις, συνέβαλε καθοριστικά στην εξέλιξη και ωρίμανση του ερευνητικού μου έργου.

Θερμές ευχαριστίες οφείλω επίσης στον Καθηγητή του Τμήματος Χημείας του Πανεπιστημίου Ιωαννίνων κ. Κωνσταντίνο Σταλίκα, για τη διαρκή του διαθεσιμότητα, την ενθάρρυνση και τις τεκμηριωμένες προτάσεις του σε κρίσιμα στάδια της διατριβής. Ο συνδυασμός επιστημονικής αυστηρότητας και ευγένειας που τον διακρίνει, αποτέλεσε

για μένα παράδειγμα επιστημονικής δεοντολογίας. Η συμβολή του στην πορεία μου υπήρξε ουσιαστική και του είμαι ειλικρινά ευγνώμων.

Θα ήθελα να ευχαριστήσω ιδιαίτερα τα μέλη της εξεταστικής επιτροπής Καθηγητή του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων κ. Νικόλαο Κουρκουμέλη. και Επίκουρου Καθηγητή του Τμήματος Χημείας του Πανεπιστημίου Ιωαννίνων κ. Παντελεήμων Τάκη, με τους οποίους είχα τη χαρά και την τιμή να συνεργαστώ στενά σε επιμέρους στάδια της διατριβής. Η ενεργή τους συμμετοχή, οι διορατικές τους επισημάνσεις και η επιστημονική τους καθοδήγηση συνέβαλαν ουσιαστικά στην εμβάθυνση της έρευνάς μου. Η εμπειρία και η αφοσίωσή τους αποτελούν για μένα πολύτιμο παράδειγμα επιστημονικής στάσης και ήθους.

Ευχαριστώ επίσης εγκάρδια τα μέλη της εξεταστικής επιτροπής Καθηγητές του Τμήματος Χημείας του Πανεπιστημίου Ιωαννίνων κ. Αθανάσιο Βλεσσίδη. και κ. Δημοσθένη Γκιώκα για τη διάθεση και τον χρόνο που αφιέρωσαν για την αξιολόγηση της παρούσας διατριβής. Η πόρτα των εργαστηρίων και γραφείων τους ήταν πάντα ανοιχτή, και παρουσία τους στην επιτροπή αποτελεί για μένα ιδιαίτερη τιμή.

Ιδιαίτερη μνεία αξίζει στους αγαπητούς φίλους και συνεργάτες μου, Δρ. Ελευθέριο Παύλου, Καλλιόπη Παπαδοπούλου και Θεοδώρα Τσιλούλη, καθώς και στον αδερφό μου Γρηγόρη Χρηματόπουλο με τους οποίους μοιράστηκα απαιτητικά αλλά γόνιμα στάδια της έρευνας. Η μεταξύ μας συνεργασία υπήρξε παραγωγική, αλληλοσυμπληρούμενη και βαθιά ανθρώπινη.

Στο χώρο του εργαστηρίου, δεν ήμουν ποτέ μόνος. Συνεργάτες και φίλοι, με τη γενναιόδωρη διάθεσή τους να μοιραστούν γνώσεις και ιδέες, αποτέλεσαν το υπόβαθρο μιας συλλογικής προσπάθειας που ξεπέρασε το στενό πλαίσιο μιας διατριβής. Στην συνοδοιπόρο των πρώτων μου ερευνητικών βημάτων Dr. Cristina Jiménez-Holgado, στην Dr. Maria Laura Tummino, και στον Ελευθέριο Ηλιάδη οφείλω ένα μεγάλο ευχαριστώ για την συνεργασία και τις όμορφες στιγμές.

Θα ήθελα, επίσης, να εκφράσω την ιδιαίτερη ευγνωμοσύνη μου στην Δρ. Μερίνα Καλαμπόκα, η οποία υπήρξε καθοδηγητική μορφή κατά τη διάρκεια των μεταπτυχιακών μου σπουδών και εξακολουθεί μέχρι σήμερα να αποτελεί για μένα πολύτιμη μέντορα και προσωπικότητα ήθους. Η υπομονή της, η σαφήνεια της σκέψης της και η γενναιοδωρία με την οποία μοιράστηκε μαζί μου τις γνώσεις και την εμπειρία της, με σημάδεψαν βαθιά και με ακολουθούν σε κάθε βήμα της επιστημονικής μου πορείας. Η συμβολή της υπερβαίνει τα χρονικά όρια ενός προγράμματος σπουδών – είναι διαρκής και αληθινά εμπνευστική.

Ένα μεγάλο και ειλικρινές ευχαριστώ οφείλω σε όλους τους εθελοντές, τους αθλητές, τους προπονητές, καθώς και τις αθλητικές ομάδες και συλλόγους με τους οποίους είχα τη χαρά να συνεργαστώ στο πλαίσιο της παρούσας έρευνας. Η πολύτιμη συμμετοχή και η πρόθυμη ανταπόκρισή τους υπήρξαν καθοριστικές για την υλοποίηση ενός σημαντικού μέρους της εργασίας μου, το οποίο εκτελέστηκε εκτός του κλασικού εργαστηριακού πλαισίου. Η συνέπεια, ο επαγγελματισμός και η θετική τους διάθεση έδωσαν ουσιαστικό περιεχόμενο στα πειραματικά δεδομένα. Χωρίς τη συμβολή τους, η παρούσα διατριβή δεν θα είχε την ίδια πληρότητα και δυναμική.

Θα ήθελα να ευχαριστήσω τον Δρ. Λάμπρο Νούση, τον Καθηγητή του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων κ. Κωνσταντίνο Τσιλίδη και το Εργαστήριο Υγιεινής και Επιδημιολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων, το οποίο αποτέλεσε το εφαλτήριο της επαγγελματικής μου εμπειρίας. Εκεί στεγάστηκαν τα πρώτα μου επαγγελματικά βήματα και μου δόθηκε η πολύτιμη ευκαιρία να εργάζομαι παράλληλα με την εκπόνηση του διδακτορικού μου. Το εργαστήριο αυτό δεν αποτέλεσε μόνο έναν χώρο εργασίας, αλλά και έναν πυρήνα μάθησης, ευθύνης και καθημερινής εξέλιξης. Πέρα όμως από την επιστημονική γνώση και εμπειρία, μου χάρισε κάτι ακόμη πιο πολύτιμο: τη γνωριμία με δύο εξαιρετικούς ανθρώπους, την Χριστίνα Διαμάντη και τον Χρήστο Τσιόστα, που δεν ήταν απλώς συνεργάτες αλλά αληθινοί φίλοι, συνοδοιπόροι και στήριγμα σε πολλές στιγμές της διαδρομής αυτής.

Θα ήταν άδικο να μην ευχαριστήσω τους φίλους μου, Ανέστη Τσιαρτσιανίδη, Γιάννη Δουλιάκα και Τάσο Κωνσταντίνη, και την κοπέλα μου Ειρήνη Χατζή που στάθηκαν δίπλα μου με κατανόηση και ενθάρρυνση, προσφέροντάς μου πολύτιμες στιγμές ανάπαυλας, ειλικρινούς συντροφιάς και εσωτερικής ισορροπίας. Η παρουσία τους υπήρξε λυτρωτική στις πιο απαιτητικές περιόδους της διαδρομής αυτής.

Τέλος, η πιο ειλικρινής και βαθιά ευγνωμοσύνη μου ανήκει στην οικογένειά μου – Θεόδωρο Χρηματόπουλο και Ευτυχία Νούση, στον αδερφό μου Γρηγόρη Χρηματόπουλο και στην γιαγιά μου Ερμιόνη Νούση – καθώς και στον νονό μου Νίκο Χαλτογιαννίδη για την αγάπη, τη στήριξη και την αδιάλειπτη πίστη τους σε μένα. Η ύπαρξή τους είναι το στέρεο θεμέλιο κάθε μου προσπάθειας.

Ευχαριστώ!

Χριστόφορος Χρηματόπουλος

Ιούνιος 2025, Ιωάννινα

Table of Contents

Investigating the feasibility of spectroscopic techniques and non-invasive sampling of biofluids to explore physiological alterations during physical exercise

Christoforos Chrimatopoulos

Doctoral thesis

Summary

This PhD thesis presents the development and application of an innovative, non-invasive analytical framework for monitoring physiological and metabolic responses to physical exercise, utilizing saliva and dried blood spot (DBS) sampling.

The study drew upon a substantial cohort of 260 athletes representing a broad spectrum of sporting disciplines, encompassing both individual and team sports. Participants included competitive runners and triathlon athletes, football, basketball, volleyball and tennis players, boxers, karatekas, and muai thai athletes, as long as aerial hoops and aerobic gymnastics athletes and pole dancers, ensuring a wide representation of physiological demands and training regimens. Crucial to the success of this effort was the close collaboration with numerous sports clubs and athletic associations across the region of Epirus, Greece. Ongoing communication with coaches, trainers, and administrative staff facilitated participant recruitment, ensured adherence to protocol requirements, and strengthened the practical relevance of the study. This regional network of support not only provided access to a varied and committed participant base but also helped bridge the gap between laboratory research and applied sports science in real-world training environments.

The work is structured into four distinct yet interconnected phases, integrating analytical techniques —spectroscopic (ATR-FTIR, Vis photometry, and NMR) and chemometric (multivariate and statistical analyses)— to evaluate biochemical changes with precision and minimal invasiveness. This integrative approach addresses the increasing demand for real-time, field-deployable diagnostic tools in sports science, health monitoring, and personalized fitness assessment.

In Phase 1, attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy was applied to saliva samples from both low-level (occasional light load training) and high-level athletes (frequent heavy load training) to evaluate its potential as a non-invasive tool for physiological monitoring. The primary aim was to explore the feasibility of using salivary biochemical profiles as indicators of training status and physical conditioning. For the first time, multivariate statistical analysis of the salivary infrared data revealed clear biochemical distinctions between athletic levels, underlining the effectiveness of ATR-FTIR in capturing exercise-induced metabolic alterations. These findings highlighted the responsiveness of salivary composition to training load and demonstrated the diagnostic value of saliva in capturing cumulative physiological adaptations. Phase 1 established the proof-of-concept that infrared saliva fingerprinting can serve as a rapid, field-applicable screening method, laying the analytical and conceptual foundation for more dynamic, intensity-resolved investigations in the subsequent Phases.

Phase 2 advanced this line of inquiry by focusing on the dynamics of salivary composition during physical exercise of increasing intensity (0, 5, 10, and 15 km/h) by employing ATR-FTIR spectroscopy and advanced multivariate analysis. The novelty lies in the integration of second-derivative spectral preprocessing with extensive chemometric modeling—specifically PCA-LDA and PLS-DA—to enable fine-tuned, non-invasive classification of training loads based on salivary biochemical profiles. The innovation is expanded on the construction and rigorous validation of predictive models that achieved high classification accuracy, demonstrating the discriminative power of salivary infrared

fingerprints. In addition, alterations in salivary lactate and glucose were compared with the corresponding trends in blood, resulting an excellent harmonization in the case of lactate, in contrast to glucose, facts that align with the literature. Multivariate analysis revealed 5 spectral features (corresponding to phosphate, phospholipids, glucose, lactate, thiocyanate) where they lead to discrimination. To our surprise, thiocyanate (SCN^-) emerged as a prominent spectral feature linked with physical exercise. Phase 2 highlighted the value of spectroscopic fingerprinting in identifying metabolite-specific signatures of physical stress (such as the candidate SCN^-) and in supporting personalized training strategies.

To quantitatively validate the role of thiocyanate as a biomarker, Phase 3 introduced a robust, cost-effective photometric method tailored for high-throughput analysis of salivary SCN^- concentrations. A large-scale study involving 161 athletes confirmed, for the first time, a consistent decrease in thiocyanate levels with escalating exercise intensity. Detailed statistical evaluation, also revealed statistically significant influences of gender and smoking status on salivary thiocyanate profiles, thereby enriching our understanding of interindividual variability in metabolic responses to physical stress. This comprehensive dataset enabled precise mapping of SCN^- response trends across diverse athlete profiles and training intensities. The findings represent the first large-cohort validation of thiocyanate's utility as a non-invasive, exercise-responsive biomarker, supporting its future use in personalized monitoring frameworks within sports and exercise science.

In Phase 4, the research turned to systemic metabolic profiling using nuclear magnetic resonance (NMR) spectroscopy applied to dried blood spot (DBS) samples. This approach leveraged the minimally invasive nature of blood microsampling—achieved through DBS cards and alternative sorbent materials—making it particularly suitable for athletic monitoring in field conditions. Despite the inherent challenges of limited sample volume and matrix complexity, NMR successfully distinguished metabolic signatures associated with increasing exercise intensities. This represents a novel integration of

microsampling and high-resolution spectroscopy in sports science, where such a combination remains largely unexplored. The findings not only underscored the impact of training load on systemic metabolism but also demonstrated the feasibility of DBS-NMR workflows for real-world biomonitoring. Moreover, this Phase highlights the complementarity of different biofluids and analytical platforms in constructing a holistic, minimally invasive physiological monitoring system.

Taken together, this thesis delivers a multidimensional, minimally invasive analytical strategy for assessing biochemical and metabolic responses to physical exercise. The work underscores the synergistic utility of ATR-FTIR spectroscopy, photometric quantification of thiocyanate, and NMR metabolomics in personalizing athletic training, enhancing performance monitoring, and potentially informing broader health diagnostics. The methodologies and findings presented herein contribute significantly to the growing field of non-invasive/minimally-invasive biomonitoring and pave the way for future translational applications in sports science.

Διερεύνηση αναλυτικών μεθοδολογιών βασισμένων σε μεθόδους μη επεμβατικής δειγματοληψίας βιολογικών υγρών και τεχνικές φασματοσκοπίας για τη μελέτη φυσιολογικών προσαρμογών κατά τη σωματική άσκηση

Χριστόφορος Χρηματόπουλος

Διδακτορική διατριβή

Περίληψη

Η παρούσα διδακτορική διατριβή αναπτύσσει και εφαρμόζει ένα καινοτόμο, μη επεμβατικό αναλυτικό πλαίσιο για την παρακολούθηση φυσιολογικών και μεταβολικών αποκρίσεων κατά τη διάρκεια της σωματικής άσκησης, αξιοποιώντας τη σίελο και δείγματα ξηρού αίματος (DBS).

Η μελέτη βασίστηκε σε ένα αντιπροσωπευτικό δείγμα 260 αθλητών, οι οποίοι προέρχονταν από ένα ευρύ φάσμα αθλητικών κλάδων, περιλαμβάνοντας ατομικά (τρέξιμο, τρίαθλο, τένις, πυγμαχία, muay thai, καράτε, pole dancing, aerial hoops (στεφάνη), αερόβιες αθλοπαιδιές) και ομαδικά αθλήματα (ποδόσφαιρο, μπάσκετ, βόλεϊ). Καθοριστική υπήρξε η συνεργασία με αθλητικούς συλλόγους και προπονητικά κέντρα της ευρύτερης περιοχής της Ηπείρου, μέσω της οποίας διευκολύνθηκε η πρόσβαση στο συμμετοχικό δυναμικό και διασφαλίστηκε η εφαρμογή του ερευνητικού πρωτοκόλλου σε ρεαλιστικά περιβάλλοντα άσκησης. Η δικτύωση αυτή συνέβαλε ουσιαστικά στη γεφύρωση του χάσματος μεταξύ εργαστηριακής έρευνας και εφαρμοσμένης αθλητικής επιστήμης.

Η μελέτη οργανώνεται σε τέσσερις αλληλοσυμπληρούμενες φάσεις, στις οποίες ενσωματώνονται αναλυτικές τεχνικές —φασματοσκοπικές (ATR-FTIR, φωτομετρία

ορατού και NMR) και χημειομετρικά εργαλεία (πολυπαραμετρική και στατιστική ανάλυση)— με στόχο την ακριβή και αξιόπιστη αποτύπωση βιοχημικών μεταβολών με ελάχιστα επεμβατικό τρόπο. Το πλαίσιο αυτό ανταποκρίνεται στις σύγχρονες ανάγκες για επιτόπια και προσωποποιημένη παρακολούθηση της φυσικής κατάστασης στον χώρο της επιστήμης της άθλησης.

Στην Πρώτη Φάση, εφαρμόστηκε φασματοσκοπία υπερύθρου μετασχηματισμού Fourier με αποσβένουσα ολική ανάκλαση (ATR-FTIR) σε δείγματα σιέλου από αθλητές διαφορετικού επιπέδου (χαμηλού επιπέδου αθλητές που ασκούνται περιστασιακά και ήπια και υψηλού επιπέδου αθλητές που εκτελούν συχνές προπονήσεις υψηλής έντασης). Στόχο είχε την αξιολόγηση της σιέλου ως εναλλακτικού, μη-επεμβατικού δείγματος για την παρακολούθηση της σωματικής/αθλητικής δραστηριότητας. Για πρώτη φορά, η πολυπαραμετρική ανάλυση των υπέρυθρων φασμάτων της σιέλου ανέδειξε σαφείς διαφορές μεταξύ υψηλού και χαμηλού επιπέδου αθλητών. Η παρούσα Φάση ανέδειξε τη χρήση του υπέρυθρου αποτυπώματος της σιέλου—ως γρήγορη, φορητή και πρακτική μέθοδος αξιολόγησης—δημιουργώντας την αναλυτική βάση για επερχόμενες μελέτες σχετικά με την έντασης της άσκησης που αναπτύχθηκαν στις επόμενες Φάσεις.

Η Δεύτερη Φάση εστίασε στην μεταβολή της σύστασης της σιέλου κατά την αύξηση της έντασης της άσκησης (0, 5, 10, 15 km/h) αξιοποιώντας φασματοσκοπία υπερύθρου (ATR-FTIR) και προηγμένες τεχνικές πολυπαραμετρικής ανάλυσης. Η καινοτομία έγκειται στον συνδυασμό φασματοσκοπίας δεύτερης παραγώγου με χημειομετρικά μοντέλα—PCA-LDA και PLS-DA—για την διάκριση του προπονητικού φορτίου μέσω της πολυπαραμετρικής ανάλυσης του φασματικού προφίλ της σιέλου Η καινοτομία επεκτείνεται και στην κατασκευή και αυστηρή επικύρωση των προβλεπτικών μοντέλων, τα οποία παρουσίασαν υψηλή ακρίβεια ταξινόμησης, αποδεικνύοντας τη διακριτική ικανότητα του υπέρυθρου φασματικού αποτυπώματος της σιέλου. Επιπλέον, οι μεταβολές του γαλακτικού οξέος και της γλυκόζης στην σίελο συγκρίθηκαν με τις αντίστοιχες τάσεις που παρουσιάζουν στο αίμα, δίνοντας εξαιρετική ταύτιση στην περίπτωση του γαλακτικού, σε αντίθεση με την γλυκόζη, γεγονότα που εναρμονίζονται

με την βιβλιογραφία. Η πολυπαραμετρική ανάλυση ανέδειξε 5 φασματικές μεταβλητές (που αντιστοιχούν στα φωσφορικά ιόντα, φωσφολιπίδια, γλυκόζη, γαλακτικό, θειοκυανικά ιόντα) στις οποίες οφείλεται η διάκριση. Προς έκπληξή μας, το θειοκυανικό ιόν (SCN^-) αναδείχθηκε ως σημαντικός μεταβολίτης συσχετισμένος με την σωματική άσκηση. Η Φάση 2 επισήμανε τη σημασία της φασματοσκοπικής παρακολούθησης της σιέλου στην ανάδειξη χαρακτηριστικών μεταβολιτών που συνδέονται με τη σωματική άσκηση (όπως την περίπτωση των θειοκυανικών ιόντων), υποστηρίζοντας παράλληλα την ανάπτυξη εξατομικευμένων στρατηγικών προπόνησης.

Η Τρίτη Φάση εξέτασε τον ρόλο του θειοκυανικού ιόντος ως έναν δυνητικό βιοδείκτη μέσω της ανάπτυξης και επικύρωσης μιας αξιόπιστης, φωτομετρικής μεθόδου ποσοτικού προσδιορισμού των SCN^- στην σίελο. Σε δείγμα 161 αθλητών παρατηρήθηκε, για πρώτη φορά στην βιβλιογραφία, σταθερή μείωση των επιπέδων SCN^- με την αύξηση της έντασης άσκησης. Η ενδελεχής στατιστική ανάλυση κατάγραψε επίσης στατιστικά σημαντικές διαφοροποιήσεις ανά φύλο και συνήθειες καπνίσματος. Η Φάση αυτή αποτελεί την πρώτη ολοκληρωμένη έρευνα μεγάλης κλίμακας της συμπεριφοράς των SCN^- σχετιζόμενη με τον αθλητισμό. Τα εξαιρετικά ευρήματα αναδεικνύουν τα θειοκυανικά ιόντα ως ένα δυνητικό μη-επεμβατικό βιοδείκτη στην παρακολούθηση—σε πραγματικό χρόνο—της φυσιολογικής απόκρισης στην άσκηση, ενισχύοντας τη δυνατότητα εφαρμογής του σε εξατομικευμένα συστήματα προπονητικής αξιολόγησης.

Στη Τέταρτη Φάση, η έρευνα επικεντρώθηκε στη αποτύπωση μεταβολικών αλλαγών μέσω φασματοσκοπίας πυρηνικού μαγνητικού συντονισμού (NMR), εφαρμοσμένης σε δείγματα ξηρής κηλίδας αίματος (DBS). Η προσέγγιση αυτή αξιοποίησε τον ελάχιστα επεμβατικό χαρακτήρα της μικροδειγματοληψίας αίματος—μέσω καρτών DBS και εναλλακτικών προσροφητικών υλικών—καθιστώντας την ιδιαίτερα κατάλληλη για παρακολούθηση αθλητών σε συνθήκες πεδίου. Παρά τις προκλήσεις που σχετίζονται με τον περιορισμένο όγκο δείγματος και την πολυπλοκότητα του υποστρώματος, η φασματοσκοπία NMR κατάφερε να προσδιορίσει πλήθος χαρακτηριστικών μεταβολιτών, και να διακρίνει μεταβολικά προφίλ που σχετίζονται με

την αύξηση της έντασης της άσκησης. Η Φάση αυτή εισάγει με επιτυχία μια καινοτόμο συνδυαστική εφαρμογή μικροδειγματοληψίας αίματος και φασματοσκοπίας υψηλής ευαισθησίας στον τομέα του αθλητισμού, όπου η εν λόγω προσέγγιση παραμένει ως επί το πλείστων ανεξερεύνητη. Επιπλέον, η Φάση αναδεικνύει τη συμπληρωματικότητα μεταξύ διαφορετικών βιολογικών υγρών και αναλυτικών μεθοδολογιών για την ανάπτυξη ενός ολιστικού και ελάχιστα επεμβατικού συστήματος φυσιολογικής παρακολούθησης.

Συνολικά, η διατριβή αυτή προτείνει μια πολυδιάστατη και μη-επεμβατική/ελάχιστα επεμβατική στρατηγική παρακολούθησης των φυσιολογικών και μεταβολικών αποκρίσεων κατά την άσκηση. Προτείνει το θειοκυανικό ως νέο βιοδείκτη έντασης άσκησης και συνδυάζει τεχνικές ATR-FTIR, φωτομετρίας ορατού και NMR για την παροχή ενός ολοκληρωμένου εργαλείου προσωποποιημένης αξιολόγησης φυσικής κατάστασης, με ευρύτατες εφαρμογές στην επιστήμη της άθλησης.

THEORITICAL FRAMEWORK

Theoretical Framework

1. The art of sports science

1.1 Overview of sports science

"The definition of sport science: as a sub-field of humankind's universal culture, it is a theoretical system representing the culture of the body by the evidence-based, systematical and generalised principles, themes, laws and rules, theories and methods. Its research aim is to enrich values of the society's culture of the body (as a subculture of the universal culture), and thus support individual and eventually the totality of societal development. It is the examination of people as biological-psychological and social units, who consciously practice physical activity" – Biróné Nagy Edit, 2011[1].

The significant advancement of sports science began in the 1950s, largely driven by the competitive rivalry between the Soviet Union and the United States[2]. Prior to this period, scientific investigations into health care had examined the effects of physical education and sports movements on the human body. However, there was a notable difference in the research focus of the two superpowers: the Soviet Union concentrated almost exclusively on professional sports, whereas the United States pursued research in recreation, rehabilitation, and specialized physical education alongside professional sports studies. The global recognition of the field grew with the introduction of sport science conferences held in conjunction with the Olympic Games starting in 1956[2]. In terms of key organizations, the International Federation of Sports Medicine (FIMS) was

established in 1928, and the International Council of Sport Science and Physical Education (ICSSPE) was founded in 1960.

The domain of sports science has evolved dramatically over the past few decades, becoming a critical area of study that integrates knowledge from various scientific disciplines to optimize human performance and health (Fig. 1). Sports science is an interdisciplinary field that encompasses physiology, biomechanics, psychology, nutrition, and medicine. It is dedicated to understanding how the human body responds to exercise, how performance can be enhanced, and how injuries can be prevented. As athletes continuously strive to break records and achieve new levels of excellence, sports science provides the foundational knowledge needed to push the limits of physical capability while safeguarding the long-term well-being of athletes. This field not only supports elite athletes in reaching peak performance but also plays a crucial role in public health by promoting physical activity and helping to combat lifestyle-related diseases[3].

Central to sports science is the study of the physiological adaptations that occur in response to physical training. These adaptations involve complex interactions between various systems in the body, including the cardiovascular, muscular, skeletal, and nervous systems[4–9]. By analyzing these interactions, sports scientists can identify the most effective training methods, tailor exercise programs to individual needs, and develop strategies to optimize recovery and prevent overtraining.

Moreover, sports science is vital for understanding the mechanisms of fatigue and the factors that limit performance. Fatigue is a multifaceted phenomenon influenced by energy depletion, metabolic by-products, neuromuscular function, and psychological state[10–13]. Through research, sports scientists have been able to develop interventions to delay the onset of fatigue, thereby enhancing endurance and overall performance[14].

Figure 1. The relationship of sport science and the main scientific fields[2] (taken from ref. 2).

Injury prevention and rehabilitation are also key aspects of sports science. The high demands placed on athletes' bodies increase their susceptibility to injuries, which can be career-threatening if not managed properly. Sports science provides the tools and knowledge to design training programs that minimize injury risk, identify early signs of

potential injuries, and implement rehabilitation protocols that ensure a safe return to competition[15].

In addition to physical training, sports science also encompasses the psychological aspects of performance. Mental toughness, motivation, and focus are critical determinants of success in sports, and sports psychologists work alongside coaches and athletes to develop these attributes. Techniques such as goal setting, visualization, and stress management are used to enhance performance under pressure and maintain mental well-being[16,17].

Furthermore, the role of nutrition in sports science cannot be overstated. Proper nutrition supports training adaptations, aids recovery, and ensures that athletes have the energy and nutrients necessary for peak performance[18,19]. Sports nutritionists design diet plans that meet the specific demands of different sports and training Phases, optimizing macronutrient intake, hydration, and supplementation[20].

The application of technology in sports science has also revolutionized the way athletes train and compete. Wearable devices, motion analysis systems, and performance tracking software provide real-time data that can be used to fine-tune training and improve performance[21–23]. These technologies have made it possible to monitor athletes with unprecedented precision, enabling personalized training programs that account for individual variability in response to exercise.

Finally, the science of physical exercise plays a crucial role in the development of youth and amateur athletes. By applying scientific principles to training and competition at all levels, sports scientists help young athletes develop their full potential while promoting lifelong habits of physical activity[24,25].

Sports science is indispensable in the modern era of athletics. It provides the knowledge and tools necessary to enhance performance, prevent injuries, and promote the overall health of athletes. As the field continues to evolve, it will undoubtedly

contribute to further advancements in both elite and recreational sports, ensuring that athletes can achieve their goals in the safest and most effective manner possible.

1.2 Recent scientific advancements in sports science

In recent years, sports science has witnessed significant advancements driven by new technologies, data analytics, and an increased understanding of human physiology. These developments have not only deepened our knowledge but also opened new avenues for enhancing athletic performance and health outcomes.

One of the most prominent areas of advancement is the integration of genomic and molecular biology into sports science[26,27]. The field of exercise genomics has gained traction as researchers seek to understand how genetic variations influence an individual's response to training, susceptibility to injuries, and potential for recovery. Studies have identified specific genes that are associated with traits such as muscle fiber composition, oxygen utilization, and inflammation[28–31], allowing for more personalized approaches to training and rehabilitation. This genetic insight is beginning to inform training programs tailored to an athlete's unique genetic profile, optimizing performance while reducing the risk of injury[32].

Biomechanics and neuroscience have also seen remarkable progress, particularly in understanding how the brain and nervous system interact with the musculoskeletal system during exercise[33,34]. Advances in imaging technologies, such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI), have enabled scientists to study the neural mechanisms underlying movement, coordination, and motor learning in unprecedented detail[35]. This research is critical for developing interventions that enhance motion skills, prevent injuries, and rehabilitate athletes after injury[36,37].

Another significant trend in sports science is the growing emphasis on data analytics and artificial intelligence (AI)[38]. The vast amounts of data generated by wearable devices, performance tracking systems, and physiological monitoring tools have created opportunities for advanced analytics to identify patterns and predict outcomes[39]. AI and machine learning algorithms are increasingly being used to analyze these data, providing insights that can refine training regimens[40], optimize in-game strategies[41], and predict injury risks[42]. For example, AI-driven models can assess an athlete's workload and recovery patterns, helping coaches make informed decisions about training intensity and rest periods[43].

Sports nutrition has also evolved, with a deeper understanding of how macronutrients, micronutrients, and supplements affect performance and recovery[44]. Research on nutrient timing, the role of gut microbiota in health and performance, and the impact of personalized nutrition plans has led to more sophisticated dietary strategies for athletes[45–47]. Additionally, there is growing interest in the use of ergogenic aids– substances or techniques that enhance performance– ranging from traditional supplements like creatine and caffeine to novel approaches like ketone esters and personalized hydration strategies[48,49].

The focus on recovery science has intensified, recognizing that recovery is as crucial as training itself in the overall performance equation. Techniques such as cryotherapy, compression garments, massage therapy, and sleep optimization are being rigorously studied to determine their efficacy in promoting muscle repair, reducing inflammation, and restoring physiological balance[50–53]. Understanding the science behind these recovery modalities is helping athletes recover faster and perform at their best more consistently.

Lastly, sports psychology has gained prominence as mental health and well-being are increasingly recognized as integral components of athletic success. Research in this area has expanded to include the psychological impact of injuries[54] and the prevention of burnout[55]. Mental conditioning techniques are now being integrated into regular training regimens, helping athletes manage stress, enhance focus, and maintain motivation[55–57].

The recent advancements in sports science have transformed the way athletes train, compete, and recover. The integration of genetic insights, advanced biomechanics, data analytics, and personalized nutrition is leading to more effective and individualized approaches to athletic performance. As the field continues to evolve, sports science will undoubtedly play a pivotal role in pushing the boundaries of human potential while ensuring the health and well-being of athletes.

1.3 Sports biochemistry: Understanding the molecular basis of athletic performance

The field of sports biochemistry plays a pivotal role in understanding the complex biochemical processes that occur in the body during physical activity. As a sub-discipline of sports science, sports biochemistry focuses on the molecular and metabolic pathways that underpin exercise, performance, recovery, and adaptation to training. These biochemical processes not only determine an athlete's ability to perform but also govern the body's response to different types and intensities of exercise, making sports biochemistry an essential component in the scientific study of athletic performance.

During physical exertion, the human body undergoes a series of rapid biochemical changes, driven by the need to produce energy for muscle contraction[58], maintain homeostasis[59], and repair tissues[60]. The primary source of energy during exercise is adenosine triphosphate (ATP), which is generated through three main metabolic pathways: the phosphagen system (creatine phosphate (ATP-PC)), glycolysis, and oxidative phosphorylation (a series of chemical reactions that generate energy in the mitochondria)[61,62] (Fig. 2).

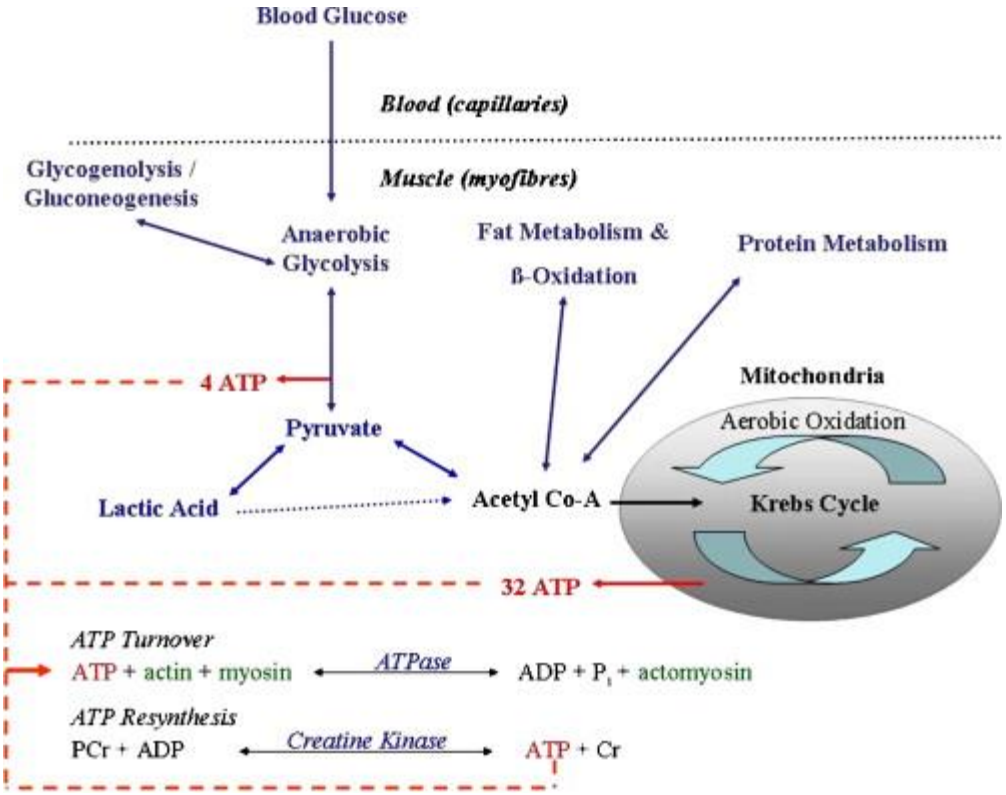


Figure 2. A general representation of aerobic, anaerobic and high energy phosphate bioenergetic pathways[62] (taken from ref. 62).

Each of these systems contributes differently depending on the intensity and duration of exercise. For example, short bursts of high-intensity activity rely heavily on the phosphagen system, which uses stored ATP and creatine phosphate in muscles[62,63]. This system can provide energy for muscles in the initial 1 to 15 seconds of high intensity exercise[64]. Longer, endurance-based activities depend more on aerobic pathways such as oxidative phosphorylation[62,63] (Fig. 3).

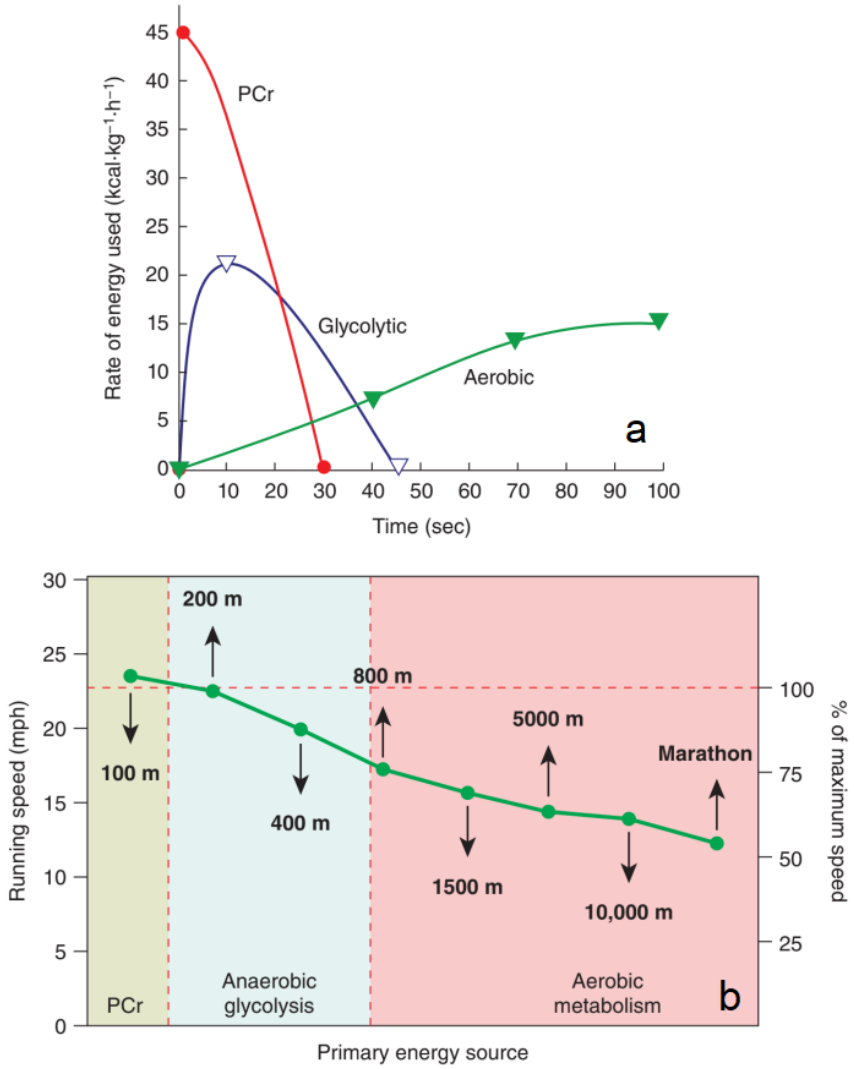


Figure 3. Energy systems and their application to training. (a) Energy continuum, (b) primary energy sources for different running distances[63] (taken from ref. 63).

The study of sports biochemistry is crucial for understanding how these energy systems are activated and regulated during exercise, as well as how the body's biochemical environment changes in response to different types of physical stress. For instance, high-intensity exercise results in the accumulation of metabolic by-products such as lactate and hydrogen ions, leading to muscle acidosis and fatigue[61]. The body's

ability to buffer and remove these by-products, as well as regenerate ATP, is a key determinant of athletic performance and endurance[61].

Furthermore, biochemistry in sports plays a significant role in nutritional strategies for athletes. The body's biochemical responses to different macronutrients—carbohydrates, fats, and proteins—are crucial for energy production, muscle repair, and recovery[65]. For example, consuming carbohydrates post-exercise helps replenish glycogen stores, while protein intake supports muscle repair and growth[66].

Sports biochemistry also examines the role of key biochemical markers in muscle damage and repair. During intense or prolonged exercise, muscle fibers undergo microscopic tears, triggering an inflammatory response[67]. This process leads to the release of various biochemical markers, such as creatine kinase (CK) and myoglobin, which are indicative of muscle damage[68,69] (Fig. 4). The body responds to this damage through a series of repair processes, involving protein synthesis, the activation of satellite cells, and the reorganization of muscle fibers[60]. Understanding these biochemical responses is essential for developing strategies to optimize recovery and minimize the risk of overtraining and injury.

The study of hormonal regulation during exercise is critical part of overtraining estimation. Hormones such as cortisol, testosterone, and insulin-like growth factor 1 (IGF-1) play significant roles in regulating metabolism, muscle growth, and recovery[70]. Cortisol, often referred to as the "stress hormone," increases in response to physical and psychological stress[71] and helps regulate energy metabolism by promoting gluconeogenesis and the mobilization of fatty acids[72]. However, chronically elevated cortisol levels, often seen in overtrained athletes, can have detrimental effects on immune function, muscle tissue, and overall recovery[73,74]. Conversely, testosterone and IGF-1 promote muscle protein synthesis and adaptation to training[75,76]. The balance between anabolic (muscle-building) hormones and catabolic (muscle-degrading) hormones is a key factor in determining an athlete's ability to build muscle and recover from training[75,77,78]. Monitoring these hormonal levels provides valuable insights into an

athlete's physiological state and can guide adjustments in training volume, intensity, and recovery periods.

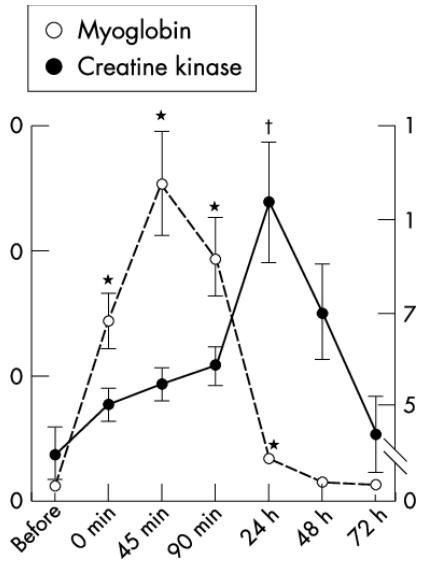


Figure 4. Changes in plasma myoglobin concentration and creatine kinase activity after the rugby matches. Values are mean (SE) (n = 14). *,† Significantly different from resting state within the same subjects ($p<0.05$, Wilcoxon signed ranks test)[79] (taken from ref. 70).

The applied science of biochemistry in physical exercise is also critical for understanding fatigue and recovery. For example, elevated levels of cortisol and creatine kinase, combined with a decrease in immunoglobulin A (IgA), can indicate that an athlete is not fully recovering between training sessions, increasing the risk of overtraining syndrome[80]. This way enables coaches and sports scientists to intervene before an athlete reaches a state of chronic fatigue or injury.

For both amateur and professional athletes, preventing injury is essential, especially with the increasing number of competitions that result in more intense training[81]. Health experts warn that excessive training and competition can lead to overtraining, posing significant health risks[82]. Molecular markers mentioned previously can help optimize

training, assess an athlete's functional condition, and provide evidence-based guidance for high-performance athletes[82]. Metabolomics, which measures low molecular weight metabolites, offers an objective way to analyze the molecular effects of exercise, reducing injury risk, adjusting training, and speeding up recovery[83]. The use of -omics data provides a comprehensive view of the biological processes behind athletic performance, helping to identify new intervention strategies. The term "sportomics" has been coined to refer to the application of -omics sciences to understand metabolic changes caused by physical activity[84] (Fig. 5).

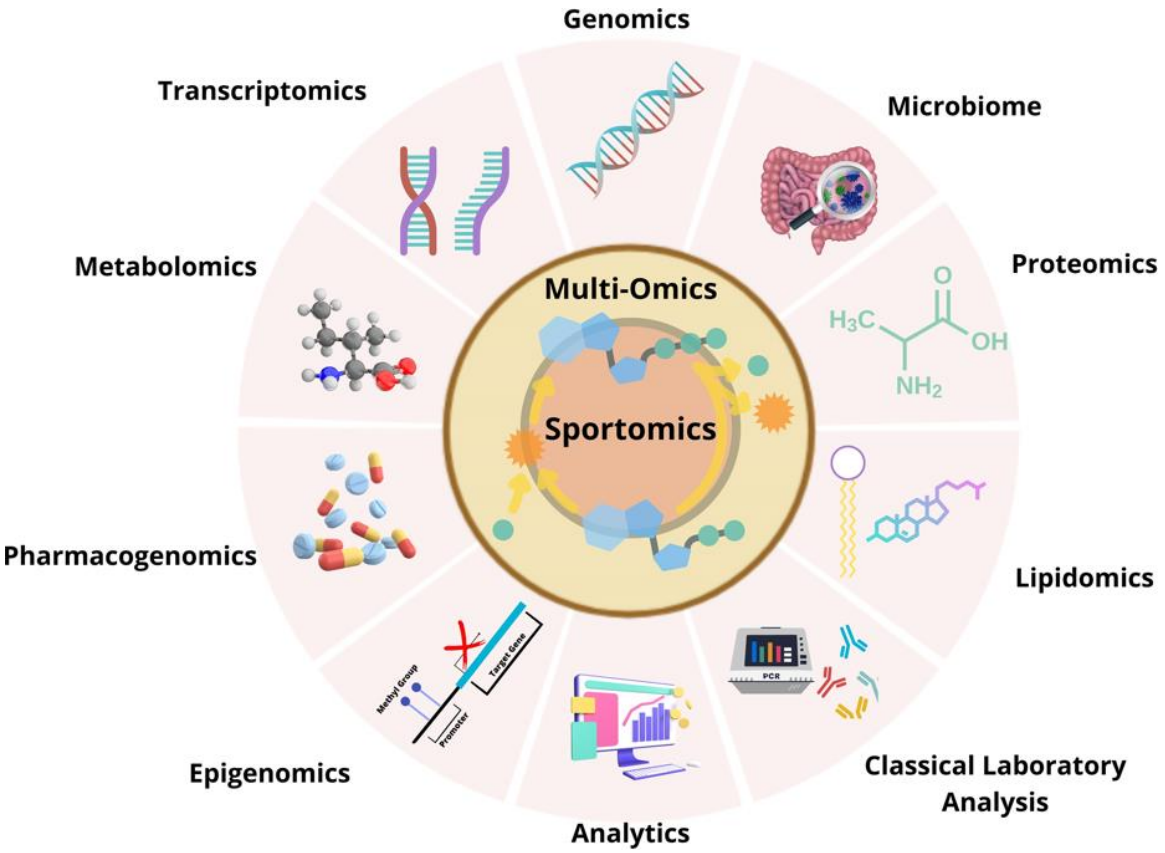


Figure 5. An overview of sportomics and the correlated –omics sciences[85] (taken from ref. 85).

Sports biochemistry serves as the critical link between sportomics and athletic state monitoring, providing the molecular and metabolic insights necessary to optimize athletic performance. By understanding the biochemical processes that occur during exercise, recovery, and adaptation, sports scientists can develop more effective training protocols, monitor athlete health in real-time, and prevent overtraining and injury. Biofluid analysis, in turn, allows for the practical application of these biochemical principles, offering a efficient way to track an athlete's physiological status and adjust their training accordingly. The science of sports metabolomics via biofluid monitoring represent powerful tool in the optimization of human performance.

2. Biofluids: The way the human body talks

2.1 Biofluid monitoring in sports science

In the quest to enhance athletic performance and safeguard the health of athletes, monitoring physiological and biochemical changes during physical activity is paramount. This is where biofluid monitoring becomes integral to sportomics. As the body's biochemical environment fluctuates during and after exercise, these changes can be detected and measured through various biofluids, such as blood, urine, sweat, and saliva[86]. These biofluds provide a window into the body's internal environment, reflecting various metabolic and physiological states induced by exercise. They contain a wealth of biochemical markers that reflect the body's metabolic, hormonal, and immune responses to physical activity[87]. By analyzing these markers, sports scientists and clinicians can gain real-time insights into an athlete's physiological status, allowing for precise adjustments to training and recovery protocols.

Blood has traditionally been the gold standard in physiological monitoring, offering detailed insights into a wide array of biomarkers, including hormones, metabolites, and electrolytes[88–90]. For instance, blood is commonly used to measure lactate levels, hormone concentrations, and inflammatory markers, providing a detailed snapshot of the body's biochemical state during and after exercise[91–93]. A recent study demonstrated succinate, pantothenate, glucose-6-phosphate, and niacinamide increment in plasma after physical exercise[94], while on metabolic changes were also detected in serum[95]. Blood sampling allows for the precise measurement of parameters such as lactate levels, which indicate anaerobic metabolism[61], and cortisol levels, which reflect stress and recovery[71]. However, the invasive nature of blood sampling poses challenges, particularly in field settings or during continuous monitoring. The need for trained personnel, the

potential discomfort for athletes, and the logistical difficulties of frequent sampling[96] have spurred the exploration of alternative biofluids.

Urine analysis is another widely used method in sports science, providing valuable information on hydration status, electrolyte balance, and kidney function, and the excretion of metabolic waste products[97,98]. It has been instrumental in assessing hydration levels, particularly in endurance sports, where maintaining fluid balance is crucial[99]. For example, urine analysis after physical activity, provided remarkable findings[100,101], where valine, isoleucine, succinate, citrate, trimethylamine, trimethylamine N-oxide, tyrosine, and formate remain decreased for few hours after exercise[102].Urine sampling is less invasive than blood sampling and can be performed more frequently. However, like blood, urine sampling typically requires controlled conditions, and the results can be influenced by factors such as fluid intake and timing, which may limit its utility for real-time monitoring[103].

Sweat is an increasingly popular biofluid for monitoring athletes, in the context of thermoregulation and electrolyte balance, especially during prolonged exercise in hot environments[97,104]. Sweat contains a variety of electrolytes, such as sodium, potassium, and chloride[105,106], as well as metabolic by-products like lactate[90,107–109]. The analysis of sweat composition can provide insights into an athlete's hydration status, electrolyte losses, and overall metabolic activity during exercise[90]. Advances in wearable technology have facilitated the collection and real-time analysis of sweat, allowing for continuous monitoring during physical activity[107,110–113]. These developments are particularly beneficial in hot or humid environments, where maintaining electrolyte balance is essential for preventing dehydration and heat-related illnesses[104,114].

Despite the advantages offered by blood, urine, and sweat, each of these biofluids has its limitations. Blood sampling is invasive and not always practical for frequent monitoring[96]. Urine analysis, while less invasive, can be affected by external factors like hydration and fluid intake[103], and sweat analysis, though promising, is still limited by

variability in sweat production and composition[109,115]. These challenges have led to increased interest in saliva as a biofluid for monitoring athletes.

Saliva is an attractive alternative for several reasons. It can be collected non-invasively and repeatedly with minimal discomfort, making it suitable for frequent monitoring in both laboratory and field settings[116,117]. Saliva contains a wide range of biomarkers[118], including hormones (such as cortisol and testosterone[119,120]), immune markers (like immunoglobulins[121,122]), and metabolic enzymes[123], many of which correlate with blood levels[124]. This makes saliva a valuable surrogate, in a non-invasive manner and a friendly approach for the volunteers, for assessing physiological responses to exercise, stress, and recovery[125], making it ideal for frequent sampling in both training and competitive settings[126].

2.2 Microsampling techniques in blood collection

Invasive intravascular access has been the standard method for blood sampling for decades and remains widely used in healthcare and disease assessment, but it comes with several limitations[127]. Collecting blood using a hypodermic needle (typically requiring more than 1 mL of blood) demands a skilled phlebotomist and a sterile environment[128]. This invasive, centralized approach often leads to issues such as discomfort, anxiety, pain, and phobias, which may lower patient compliance[128]. Indeed, the rather large amount of sample required makes them less useful for repeated sampling in tight intervals or vulnerable groups, such as infants or the elderly. Improper venipuncture techniques can cause haemoconcentration or haemolysis, making the samples unusable and forcing patients to undergo additional blood draws[129]. Healthcare workers also face risks of sharp injuries and exposure to bloodborne pathogens, and complications like hematomas, infections, nerve damage, and iatrogenic anemia may occur, leading to physical, emotional, and financial burdens[130–133]. These factors can delay or prevent medical procedures and reduce participation in clinical research[134]. Additionally, blood samples

require time-consuming and costly processing to limit pre-analytical variability[134], with around 75% needing centrifugation to separate plasma or serum. This increases costs and can slow down laboratory workflows[135]. Furthermore, wet blood samples must be stored and shipped under cold-chain conditions to avoid degradation and bacterial contamination[136].

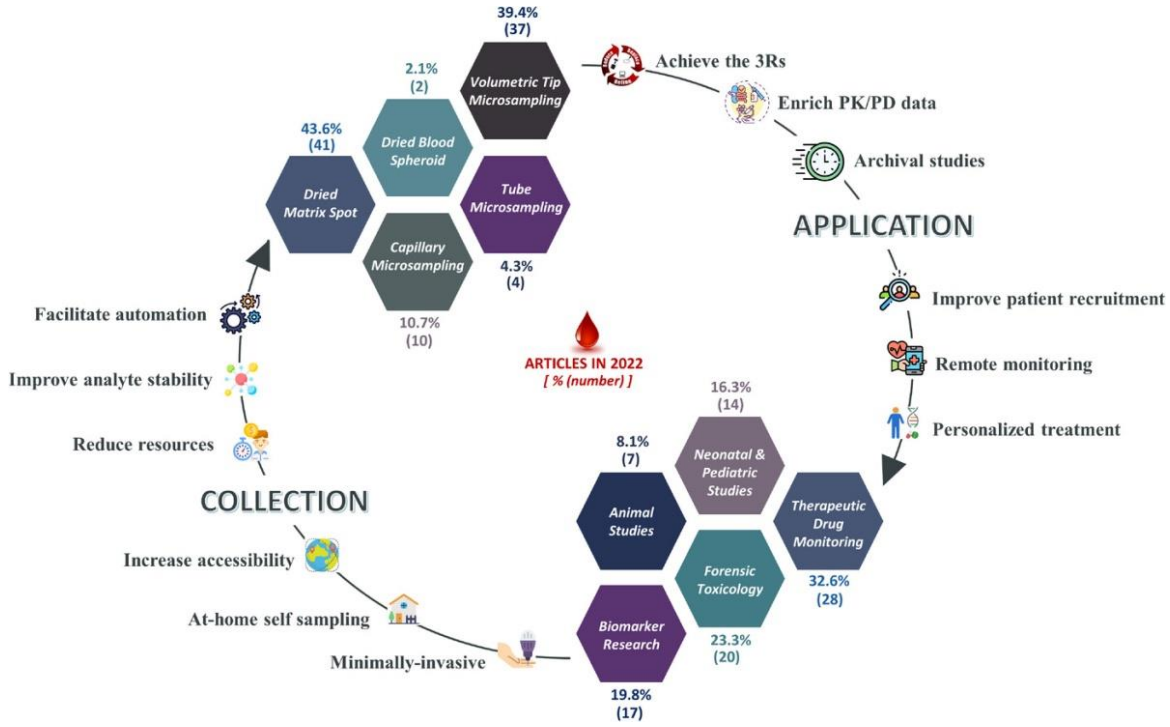


Figure 6. Graphical summary of the benefits of microsampling and 2022 publication statistics of microsampling technologies and applications[129] (taken from ref. 129).

In recent years, microsampling techniques have emerged as innovative solutions to the challenges associated with traditional blood sampling, particularly in contexts where frequent, minimally invasive, and low-volume sampling is required. These techniques enable the collection of small amounts of blood—often in the microliter range —making them particularly suitable for monitoring athletes, pediatric patients, or individuals in

remote settings. Among the most commonly employed microsampling techniques are Dried Blood Spot (DBS) sampling, Volumetric Absorptive Microsampling (VAMS) and capillary microsampling, microneedle sampling, each offering distinct advantages and facing unique challenges. Figure 6 summarizes some of the blood microsampling techniques utilized in 2022 clinical studies.

2.2.1 Dried Blood Spots (DBS)

Dried Blood Spot (DBS) sampling is one of the most widely utilized microsampling techniques, particularly in clinical and remote settings[137]. Dried blood spots (DBSs) have been used in newborn screening (NBS) since 1961 with Robert Guthrie and his test to detect phenylketonuria (PKU) at an early stage in newborn blood[138]. Nowadays, standard DBS cards are still called Guthrie cards and are commonly used for NBS. In 2022, the USA, Europe and Latin America were the regions with the highest percentage of newborns screened (100, 78 and 32%, respectively)[139]. In DBS sampling, a small drop of capillary blood is collected via a finger prick (or heel prick in newborns) with the lancet and applied to a specially designed filter paper card[140] (Fig. 7). The blood is allowed to be adsorbed onto the card and dry at room temperature for 4 hours, after which the card is stored and transported for laboratory analysis[140]. The minimal volume of blood required—often just a few microliters, typically <100 µL[129]—adds to the procedure's appeal, particularly for frequent sampling[129]. The simplicity of DBS sampling makes it highly advantageous; it requires minimal training and equipment, facilitating its use in field settings or even self-collection by patients[96,129,140]. Furthermore, once dried, the blood spots are highly stable for many metabolites, which is helpful for blood banks[141] and can be transported at ambient temperatures without the need for refrigeration, thereby reducing logistical challenges. Before analysis, fixed-diameter spots (usually 3 or 6 mm) are punched out of the paper substrate, and an extraction protocol is executed using appropriate buffers[137].

Figure 7. Illustration of dried blood spots collected from fingertip and heel on paper Whatman® 903 Protein Saver Card (Cytiva, Global).

However, the technique often suffers from variations in sample quantification and analysis due to technical errors and biological fluctuations, such as hematocrit (Hct)[137]. The Hct is the volume percentage of red blood cellsfluctuates between 36 and 50% based on factors such as race, sex, age, fluid intake, and overall health[142]. When spotting a fixed volume of blood sample on the sample collection card, the Hct affects the resultant size of spot[143,144]. The lower the Hct, the lower the viscosity: the blood will spread faster through paper fibres and will make the blood spot large, colourless and less homogenous. With a high Hct, the blood spot will be smaller, more intense in colour and more homogenous[145,146]. Punching disks of the same size for blood spots with varying Hct levels results in different volumes, leading to significant measurement and quantification errors[146]. On a traditional DBS card, the spot is punched to remove it from the card. Multiple punches can be made in one DBS spot. But due to the haematocrit effect, depending on where the punch is made, the sample may be different. The Hct effect can introduce variation in analysis, such as the amount of analytes in the sample[147].

Further analytical biases are introduced by the inherent component of DBS technology that enables the drying of blood: the filter paper. Filter paper properties determine the maximum loading capacity, blood spreadability, chromatographic effects, analyte stability and recovery[129]. During the formation of DBS, the content of blood droplets may undergo a chromatographic effect or coffee-ring effect due to differential diffusion across the filter paper[129]. Besides Hct, other factors such as humidity, drying conditions and material of the filter paper also contribute to the uneven distribution of analytes[148]. Additionally, the process of extracting analytes from dried blood spots can be complex, with some substances potentially degrading during drying, which may compromise the reliability of results. Despite these limitations, DBS remains a valuable tool in microsampling, especially for applications where sample stability and ease of collection are paramount[96].

Although different modifications to traditional DBS cards exist to minimize the Hct effects, the easiest approach to eliminate the Hct bias related to spot size and inhomogeneity is to analyze the complete DBS spot formed from a volumetric application of blood[129]. Volumetric DBS can be obtained either by punching the entire DBS after the volumetric application of blood or by volumetrically applying blood on pre-punched discs[143]. Accurate volumes of blood for application can be procured using a micropipette or microfluidic channels. However, effective pipetting requires skilled personnel, which limits its scope of application due to the reduced feasibility of self-sampling[129].

2.2.2 Volumetric Absorptive Microsampling (VAMS)

Another solution to overcome the haematocrit effect is the development of quantitative devices to collect an exact volume of capillary blood. Volumetric Absorptive Microsampling (VAMS) represents a more recent advancement in microsampling technology, addressing some of the limitations associated with DBS. VAMS devices utilize

a specially designed porous hydrophilic absorbent tip that collects a fixed, precise volume of blood—typically around 10 to 20 microliters[137]—from a finger prick (Fig. 8).

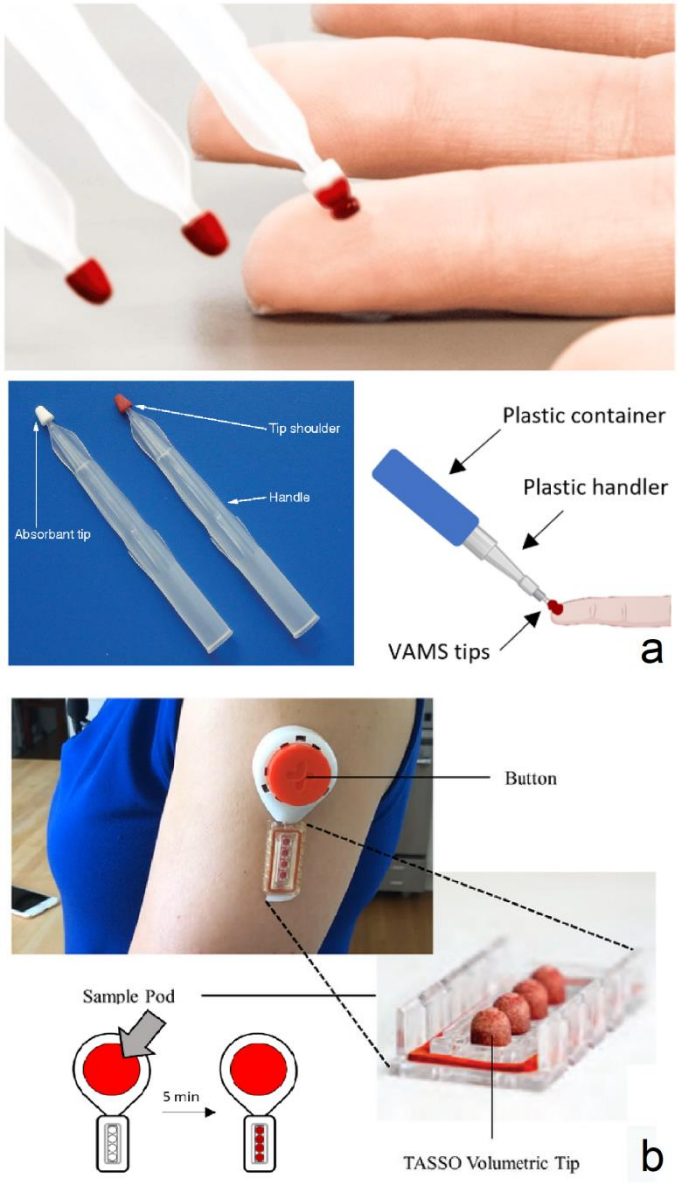


Figure 8. Volumetric absorptive microsampling technology. (a) Mitra® device (Trajan Scientific and Medical, Melbourne, VIC, Australia) and (b) TASSO (Tasso Inc., Seattle, WA, USA) M-20.

The absorbent tip is then dried and sent for analysis, similar to the DBS method. The key advantage of VAMS lies in its precision[147]; it ensures that a consistent blood volume is collected with each sample, thereby improving the accuracy and reproducibility (<4% RSD) of analytical results[149]. Like DBS, VAMS is simple to perform and can be used in non-clinical settings, making it highly versatile. Nevertheless, VAMS devices are generally more expensive than traditional DBS cards, which can be a consideration in large-scale studies. While the technique is user-friendly, proper training is essential to ensure that the correct volume is absorbed, especially in settings where self-collection is involved. As with DBS, the processing and extraction of samples from VAMS devices can be complex, requiring specialized protocols for different analytes[150].

2.2.3 Capillary microsampling (CMS)

Capillary microsampling (CMS) involves the collection of small volumes of blood—typically 1-35 microliters[150]— from exact-volume capillaries, coated or not with anticoagulants, to take up blood obtained via finger prick through capillary forces[129] (Fig. 9). The collected blood is then stored in a capillary tube or microcontainer sealed with wax for subsequent analysis. From a matrix point of view, no major differences are expected in the sample by the conventional way, since capillary microsamples are liquid samples[151]. CMS's primary advantage is its requirement for only a tiny amount of blood[151], making it minimally invasive and suitable for repeated measurements[152]. Additionally, the blood can be analyzed immediately on-site using portable devices, which is ideal for real-time monitoring[129]. However, capillary microsampling is not without challenges. Unlike DBS or VAMS, capillary blood stored in microtubes requires proper storage conditions to prevent degradation, typically requiring refrigeration or immediate analysis. For unstable analytes, a stabilizer is added to the sample within a few seconds of the collection. But for the analytes, which possess high affinity of binding to glass, CMS can be a constraint[153,154].

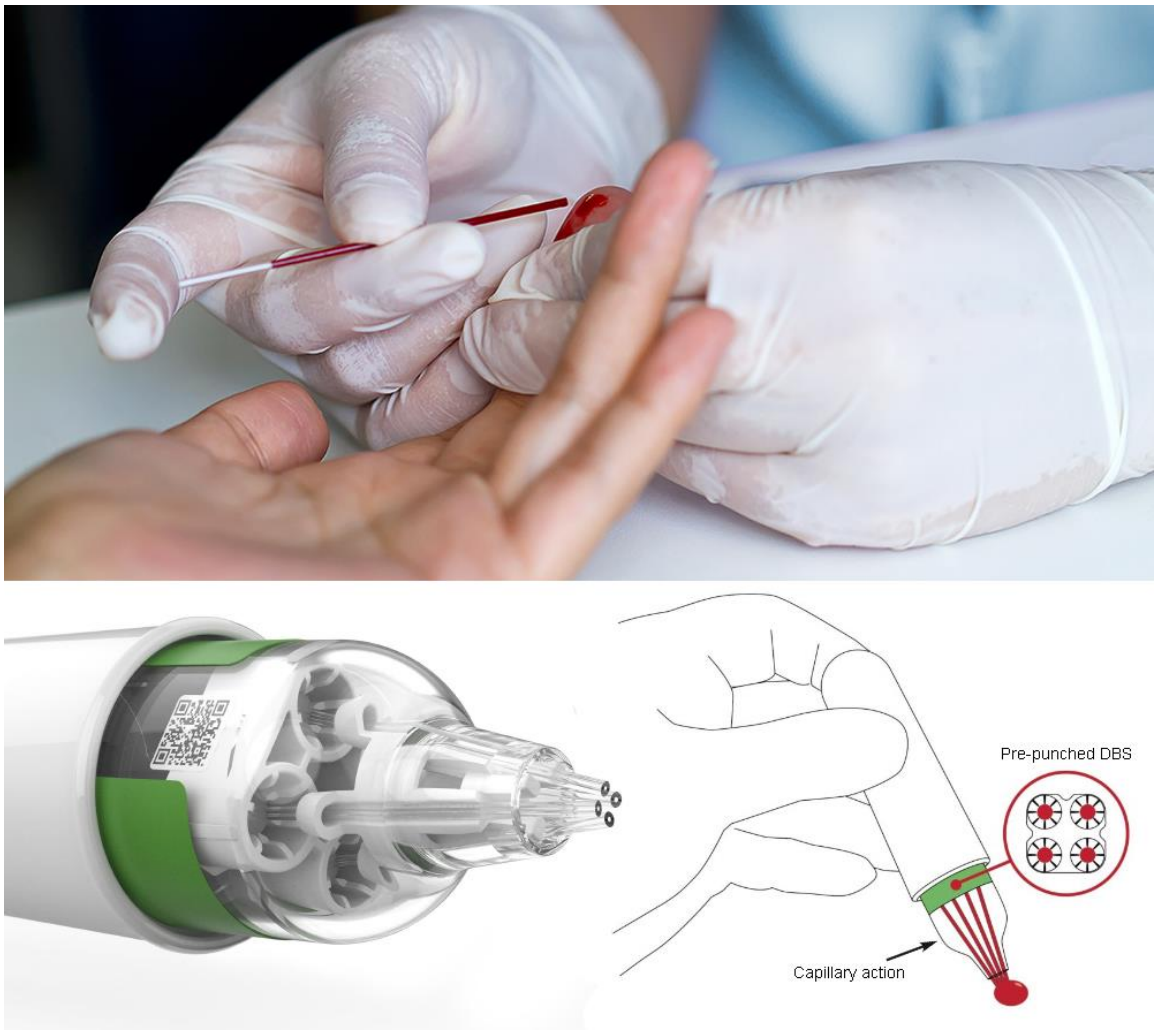


Figure 9. Typical capillary microsampling and hemaPEN® device (Trajan Scientific and Medical, Melbourne, VIC, Australia).

Although these microsampling devices provide a robust approach for collecting blood samples to overcome the DBS limitations, they fail to eliminate the impact of Hct on the extraction efficiency of analytes[129]. Despite the inconveniences caused by Hct, several advancements in the DBS technology have been made to improve existing features and develop new functionalities. Figure 10 offers an overview of the frequently used and commercially available microsampling devices for dried blood collection. Above

all, DBS is the most established blood microsampling technique, adopted by 43.6% of studies using microsampling technologies in 2022 (Fig. 10)[129].

Figure 10. A summary of commercially available of blood microsampling techniques[140] (taken from ref. 140).

2.2.4 Microneedles

Unlike previously mentioned and established techniques, microneedles are capable of penetrating the skin and sampling at the same time without the use of a blood lancet. Microneedle sampling is an emerging technology that uses an array of tiny needles, often no more than a few hundred micrometers in length[155] (Fig. 11a,b), to penetrate the skin's

outer layer and access the interstitial fluid or capillary blood[96]. These microneedles are designed to be pain-free and minimally invasive, making them particularly appealing for frequent monitoring. Early microneedle designs usually possessed a solid body and were primarily designed for cosmetic or therapeutic purposes, such as drug delivery[156]. Recent advances in rapid prototyping techniques, for example 3D printing and laser cutting has enabled microsampling with microneedles for clinical research purposes[157,158]. In most cases, a patch filled with microneedles was applied to the skin, and blood samples were continuously extracted through microfluidic channels to the back of the patch for analysis[159–161] (Fig. 11c). The primary advantage of microneedle sampling is its pain-free collection, which significantly improves comfort for the individual, especially for those requiring regular sampling. Moreover, the minimal invasiveness of microneedles reduces the risk of infection and tissue damage, making it a safer option for continuous monitoring. Microneedles also hold the potential for integration into wearable devices, enabling continuous or semi-continuous monitoring of blood or interstitial fluid biomarkers[96]. However, the volume of blood that can be collected using microneedles is extremely small (about 30 ± 5 µL of blood sample could be collected by the microneedle sampling on a rabbit model in 3 min[96]), which may limit the range of analyses that can be performed. Additionally, the cost and complexity of microneedle devices remain significant barriers to widespread adoption, as the technology is still relatively new and can be expensive both in terms of the devices themselves[162] and the infrastructure needed to process the samples.

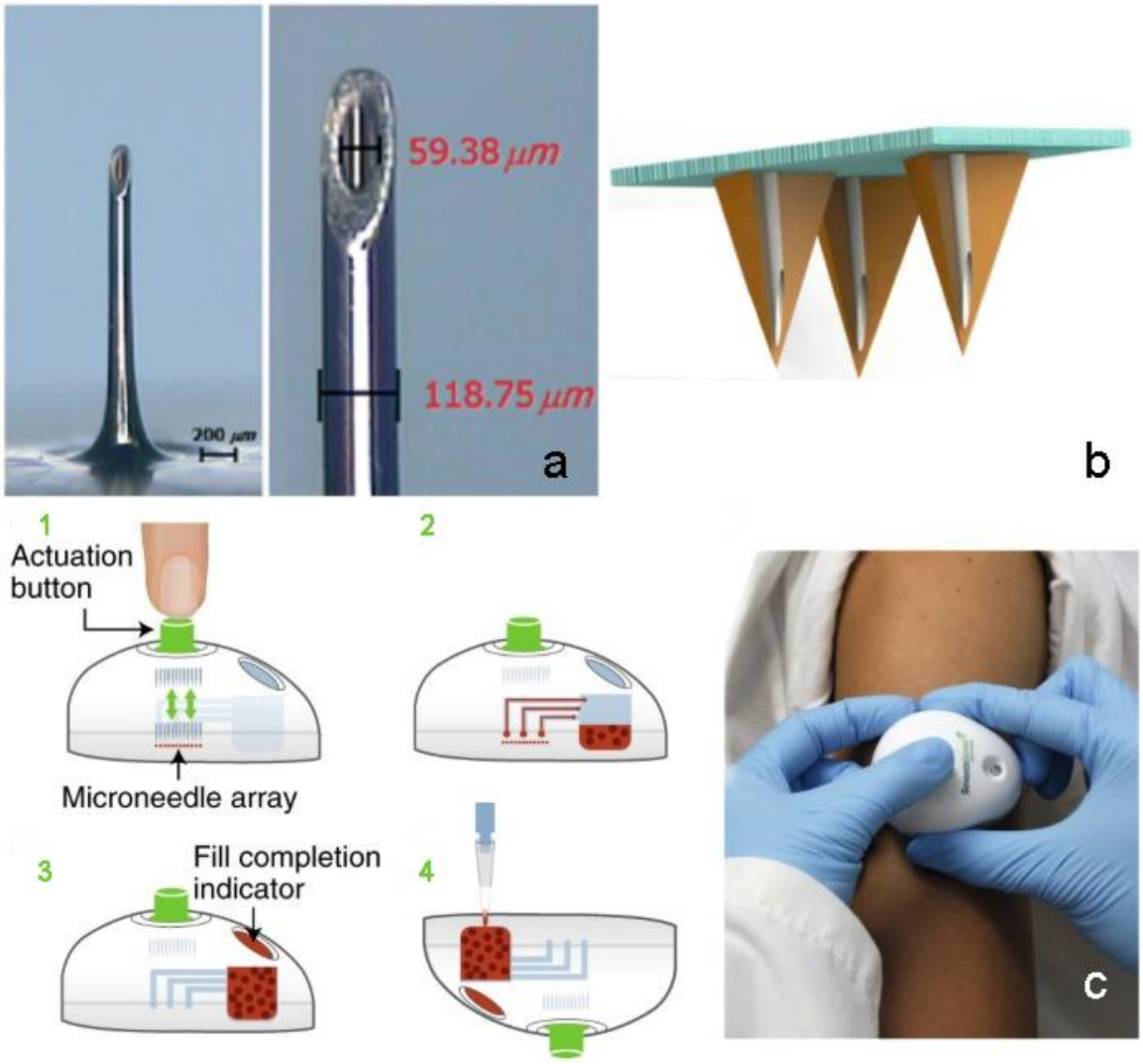


Figure 11. (a) The reported dimension of hollow microneedle[163], (b) schematic representation of the hollow microneedle array[164], and (c) microneedle TAP device (YourBio Health, Medford, MA, USA) (taken from ref. 163 and 164).

2.3 Saliva sampling: A non-invasive tool for athlete monitoring

Saliva is a hypotonic fluid primarily made up of water, electrolytes, and organic molecules like amino acids, proteins, and lipids[165]. The water in saliva mainly comes from the local capillary bed through intracellular diffusion, aquaporin water channels, and

extracellular pathways[166,167]. Small neutral molecules from the bloodstream enter the saliva through passive diffusion from the dense capillary networks surrounding the salivary glands. Electrolytes enter saliva due to osmotic gradients, with their concentration regulated by the rate of secretion, the type of stimulus, and circulating mineralocorticoid levels[165]. The organic components of saliva are largely produced through protein synthesis and stored as granules in acinar cells[168].

One of the key advantages of saliva over other biofluids is its ability to reflect the acute responses of the body to exercise[125]. For instance, the concentration of cortisol in saliva can rapidly increase in response to physical or psychological stress, providing a real-time indication of the body's stress levels. Similarly, changes in salivary immunoglobulin levels can indicate alterations in immune function, which are critical for understanding the impact of intense training on an athlete's susceptibility to infections.

Salivary analysis also offers practical benefits, particularly in field settings. Unlike blood or urine, saliva collection does not require specialized equipment or trained personnel, and it can be performed by the athletes themselves. This ease of collection facilitates frequent monitoring, enabling coaches and sports scientists to track changes in an athlete's physiological state throughout training and competition. The following sections outline the most widely used saliva sampling methods, highlighting their principles, advantages, and limitations.

2.3.1 Unstimulated – passive drooling method

Unstimulated whole saliva is the mixture of secretions that enters the mouth in the absence of exogenous stimuli and depends on the daily basal salivary flow rate in the oral cavity[169]. The composition of unstimulated saliva can be affected by the degree of hydration, position of head during collection, body posture, light exposure, drugs and circadian rhythm[170]. The passive drooling method, practiced since 1934[169], is the most

commonly employed technique for saliva sampling due to its simplicity and efficiency. In this method, the participant allows saliva to accumulate naturally in the mouth and then dribbles it into a sterile collection tube[169] (Fig. 12). Passive drool requires no external stimulation, ensuring that the sample reflects the baseline composition of saliva without the interference of materials and substances used to sample or stimulate the salivary flow that could alter its content[171].

Figure 12. Example of the spitting method into a polypropylene centrifuge tube.

This method offers several advantages. First, passive drool is a true reflection of unstimulated, whole saliva, making it ideal for measuring biomarkers that are sensitive to changes in saliva flow rate. It is particularly effective for assessing hormones such as cortisol, testosterone, and immunoglobulins, which can be affected by physical exercise[125,172]. Second, it is a straightforward and non-invasive method that can be performed by the participants themselves, reducing the need for specialized personnel

and equipment[173]. This ease of collection makes passive drool particularly suitable for field-based studies and large-scale research projects where repeated sampling is required[174].

However, there are limitations to passive drool collection. The process can be time-consuming, as participants must produce an adequate volume of saliva, typically between 1 and 5 milliliters, for most assays[171]. Additionally, some individuals may find it difficult to produce sufficient saliva[175], especially under stressful conditions or in dehydrated states, which can affect the sample's quality and volume. Finally, while passive drool provides a reliable baseline measure, it may not be appropriate for studies requiring rapid or stimulated saliva production.

2.3.2 Stimulated method

Stimulated saliva is physiologically secreted in response to either masticatory or gustatory stimulations during food intake[169]. Stimulated saliva collection is used when larger volumes of saliva are required or when it is important to enhance saliva flow[176,177] to ensure timely collection. In this method, saliva production is stimulated by external means, such as chewing on inert substances (e.g., paraffin wax, unflavored chewing gum base, cotton puff and rubber bands) or applying citric acid to the tongue[169]. The increased saliva flow allows for quicker sample collection. It was found that salivary flow peaked (2.7 ± 0.52 mL/min) within the first 2 min of stimulation, and then gradually decreased to the level of the initial unstimulated flow rate (0.39 ± 0.16 mL/min) at 25 min to reach reached a plateau[177], while other studies report that under stimulation, the flow rate may increases up to about 4 mL/min[178].

Stimulated saliva collection allows for rapid collection of sufficient saliva volumes, making it useful for assays requiring a larger quantity of biofluid. Furthermore, because it promotes consistent saliva flow, this method can reduce the variability in flow rates seen

in passive collection methods, potentially leading to more reproducible results in some contexts. Stimulated saliva is often preferred when analyzing salivary electrolytes[179,180], enzymes[181], or antimicrobial proteins[180], as their concentrations can increase with elevated flow rates, providing a more robust measure of salivary gland function.

Despite these advantages, stimulated saliva collection has notable limitations. The stimulation process can alter the composition of the saliva, particularly for biomarkers like hormones that are sensitive to changes in flow rate[169]. For instance, researches indicated that sucrose, homovanillic acid and 3-methoxy-4-hydroxyphenylglycol concentrations were altered by stimulation[182,183]. Stimulated saliva is generally more dilute than unstimulated saliva, which can affect the concentration of certain analytes and may reduce the accuracy of results for specific assays, as in the case of uric acid where the concentration decreased from 70 ± 20 µg/mL to 30 ± 10 µg/mL[184]. Additionally, the choice of stimulation method can introduce further variability (e.g. volume variation observed when paraffin stimulation was used[185]. A study indicated that the higher flow rate can increase the pH of saliva from 6.7 ± 0.24 initially up to 7.35 ± 0.22 pH units after stimulation[177], potentially impacting the stability of pH-sensitive biomarkers. During stress, saliva volume and composition which are regulated by the sympathetic and parasympathetic nervous systems, can be altered. This alteration is often expressed as dry mouth (xerostomia)[186,187]. This reduction in saliva secretion leads to a decrease in the bicarbonate (alkaline) content of saliva, causing increased acidity and a drop in oral pH[188]. Lower oral pH may contribute to the dysregulation of other stress-related saliva biomarkers, affecting components such as cortisol[189], sIgA[190], and alpha-amylase[190]. As a result, pH could play a key role in the stress-induced imbalance of saliva biomarkers[187].

2.3.3 Swab-based sampling

Salivary swabs are another popular method for saliva sampling, particularly in cases where the volume of saliva needed is relatively small. Swabbing techniques can be

divided in two subgroups. In the first case, a stick with fibers frayed at one end was used to swab the total target area, while applying medium pressure, at an angle relative to the substrate (to assure that a large area of the swab is in contact with the substrate) and rotating the swab continually, followed by extraction in buffer[191] (Fig. 13a). On the other hand, a synthetic or cotton (cellulosic) swab is placed in the mouth between a tongue and the cheek for a specified period—typically 1 to 2 minutes—to absorb saliva[169]. The swab is then placed in a sterile tube and centrifuged to extract the collected saliva for analysis (Fig. 13c).

Swab collection offers several practical advantages. It is a fast and convenient method that requires minimal cooperation from the participant, making it ideal for research involving children, elderly individuals[169], or those unable to provide passive drool samples due to exercise-induced dehydration status[192]. It is also highly useful in situations where time is a limiting factor, as the collection process is quicker compared to passive drool. Additionally, swabs are small, portable, and easy to use in field settings, making them a practical choice for monitoring athletes during competitions or training sessions[193].

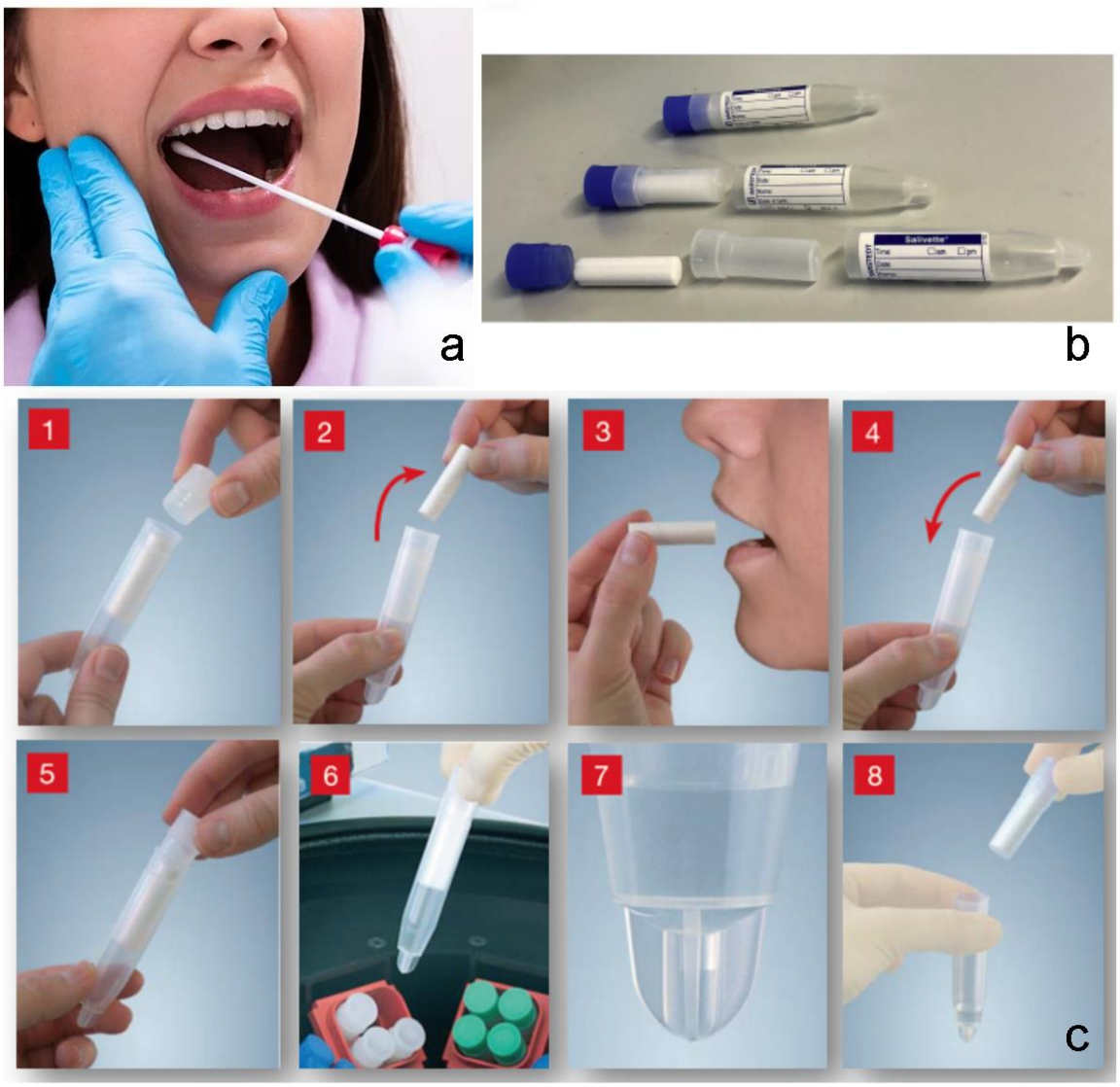


Figure 13. (a) One-end swab saliva sampling procedure, (b) Salivette® (blue cap, Sarstedt, Nümbrecht, Germany) and (c) cotton swab placed in the mouth sampling and handling procedure.

However, there are some important limitations to this method. The material of the swab can influence the composition of the saliva sample[185,194,195], with some types of swabs retaining or altering the concentration of specific biomarkers[196]. For instance, cortisol and other steroid hormones tend to adhere to polyurethane-tip applicators,

potentially leading to lower measured concentrations compared to passive drool[196]. To mitigate this issue, cotton or synthetic swabs made from inert materials such as polyester are often preferred[196]. Additionally, the recovery efficiency is also affected by the lack of uniformity in swabbing techniques. There is no clear indication of how wet or moist such a swab needs to be, and there is no consensus regarding other parameters, including swabbing time, swabbing angle, and applied pressure on the swab[191]. However, in addition to the recovery efficiency of the used swab itself, the sampling skills (including swabbing technique) of the investigator substantially influence the obtained overall efficiency. Surprisingly, the influence of the investigator's skills is frequently underestimated, with only a limited number of articles addressing this crucial aspect[197–199]. Last but not least, placing absorbent swabs in different areas of the mouth may influence both the amount of sample volume collected and the composition of analytes in the sample[200,201]. This variability can affect the reliability of the results, particularly for quantitative analyses where accurate measurement of saliva volume is essential.

2.3.4 Dried Saliva Spots (DSS)

Dried Saliva Spots (DSS) are an emerging technique for saliva sampling, offering several practical advantages. In DSS, a few drops of saliva are spotted onto collection card and dry at room conditions. DSS needed a low volume of saliva (50 µL) and allowed for a quantitative recovery of the analyte from a filter paper[169]. Subsequently, the DSS is extracted using a suitable solvent using a combination of vortex-assisted extraction and ultrasound-assisted extraction similar to DBS[202]. In 2016, Numako et al. utilized DSS for the determination of D- and L- lactic acid in diabetic, pre-diabetic and nominally healthy people. The study highlighted the use of DSS results high accuracy and precision and high recovery of the target molecule from the spot, while the target molecules were stable during the long storage period until analysis[203]. Indeed, the improved preservation and stability of the saliva sample is an outcome from the absorptive characteristic of the filter

paper[202]. Therefore, the advantage of DSS significantly reduces transportation costs, while DSS can be easily stored and transported without the need for refrigeration[202]. However, DSS also has some limitations. Sharing the same problems with DBS, limitations of DSS encompass variations in sample volume, sensitivity concerns, vulnerability to external factors during drying, limited sample volume, analyte stability challenges, potential risks of contamination, and difficulties in biomarker extraction[204]. Despite these challenges, DSS remains a promising tool for non-invasive monitoring in sports science and clinical research.

Moreover, recent advances in analytical techniques, such as Attenuated Total Reflection-Fourier Transform Infrared (ATR-FTIR) spectroscopy and chemometric analysis[205], have enhanced the ability to detect and interpret salivary biomarkers. These methods allow for the rapid and detailed analysis of saliva, identifying subtle changes in its composition that may not be detectable with traditional techniques. As a result, salivary analysis is becoming an increasingly important tool in the arsenal of sports scientists, offering a non-invasive, cost-effective, and reliable means of monitoring athletes.

Biofluid monitoring represents a crucial aspect of sports science, providing insights into the physiological and biochemical states of athletes during training and competition. While blood, urine, and sweat each offer unique advantages, saliva stands out for its non-invasive nature and ease of collection. As the field of sports science continues to evolve, the use of biofluid analysis is likely to expand, offering new opportunities to optimize performance, prevent injuries, and promote the health and well-being of athletes.

3. Infrared spectroscopy: Fundamentals and experimental methods

Infrared spectroscopy is certainly one of the most important analytical techniques available to today's scientists. One of the great advantages of infrared spectroscopy is that virtually any sample in virtually any state may be studied. Liquids, solutions, pastes, powders, films, fibers, gases and surfaces can all be examined with a judicious choice of sampling technique[206–211]. As a consequence of the improved instrumentation, a variety of new sensitive techniques have now been developed to examine formerly intractable samples.

Infrared spectrometers have been commercially available since the 1940s. At that time, the instruments relied on prisms to act as dispersive elements, but by the mid 1950s, diffraction gratings had been introduced into dispersive machines[212]. The most significant advances in infrared spectroscopy, however, have come about as a result of the introduction of Fourier-transform spectrometers. This type of instrument employs an interferometer and exploits the well-established mathematical process of Fourier-transformation. Fourier-transform infrared (FTIR) spectroscopy has dramatically improved the quality of infrared spectra and minimized the time required to obtain data[212,213]. In addition, with constant improvements to computers, infrared spectroscopy has made further great strides[214,215].

Infrared spectroscopy is a technique based on the vibrations of the atoms of a molecule. An infrared spectrum is commonly obtained by passing infrared radiation through a sample and determining what fraction of the incident radiation is absorbed at a particular energy. The energy at which any peak in an absorption spectrum appears corresponds to the frequency of a vibration of a part of a sample molecule.

3.1 Electromagnetic radiation

Electromagnetic radiation encompasses a broad range of energy forms, all characterized by their ability to propagate through space as oscillating electric and magnetic fields. These forms of radiation differ in their wavelengths and frequencies, giving rise to different types of interactions with matter, as depicted in the electromagnetic spectrum (Fig. 14). The visible portion of the spectrum, which can be detected by the human eye, is just a small fraction of the entire range, which also includes radiowaves, microwaves, infrared (IR), ultraviolet (UV), X-rays, and gamma rays[216,217].

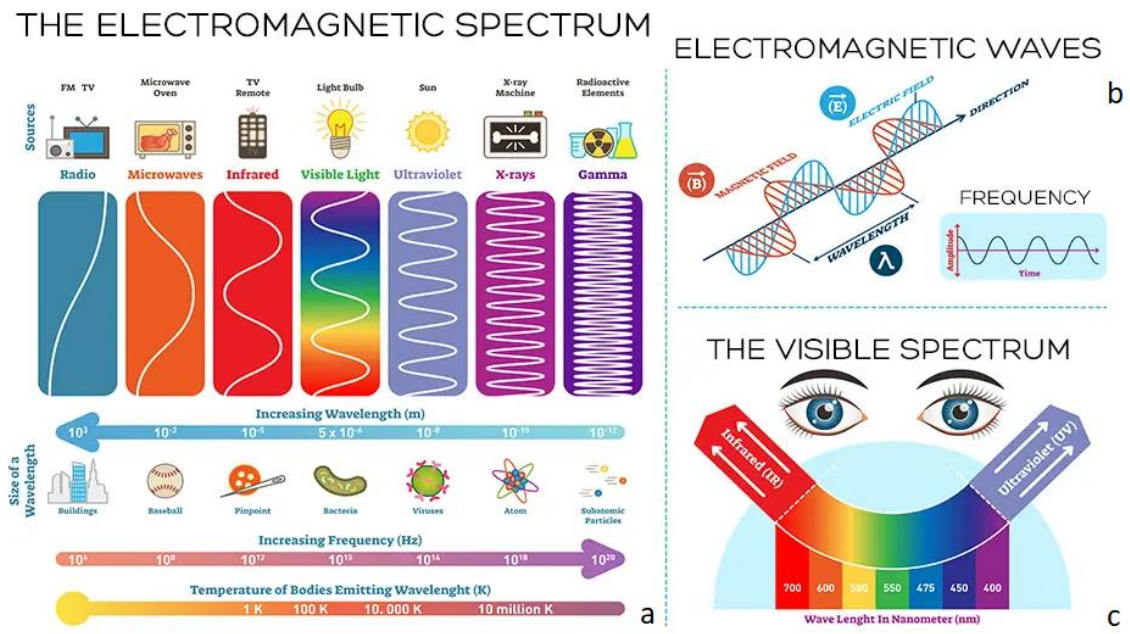


Figure 14. (a) The regions of the electromagnetic spectrum, showing various properties across the range of frequencies and wavelengths, (b) simplified picture of an electromagnetic wave (the oscillations are perpendicular to each other and to the direction of energy flow) and (c) the visible spectrum.

Spectroscopists use the interactions of radiation with matter to obtain information about a sample. The matter (analyte) is predominately in its lowest-energy or **ground**

state. The radiation then causes some of the analyte species to undergo a transition to a higher-energy or **excited state** (Fig. 15).

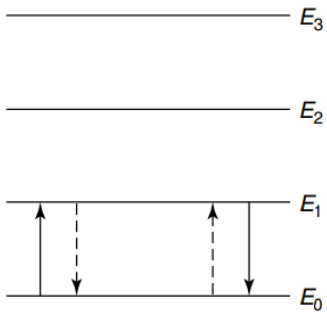


Figure 15. Illustration of quantized discrete energy levels[212] (taken from ref. 212).

The energy required for this transition must be equal to the energy of the radiation described as a stream of photons or quanta for which the energy, E, is given by the Bohr equation (Eq. 1), as follows:

$$E = h\,v \tag{1}$$

where h is the Planck constant ($h = 6.626 \times 10^{-34}$ J·s) and v is equivalent to the classical frequency of radiation[217].

The information about the analyte is acquired by measuring the electromagnetic radiation emitted as it returns to the ground state or by measuring the amount of electromagnetic radiation absorbed as a result of excitation[217].

When the sample is stimulated by applying an external electromagnetic radiation source, several processes are possible. Processes of change, including those of vibration and rotation associated with infrared spectroscopy[212], can be represented in terms of quantized discrete energy levels E_0, E_1, E_2, etc., as shown in Figure 15.

For ultraviolet and visible radiation, excitation occurs when an electron residing in a low-energy molecular or atomic orbital is promoted to a higher-energy orbital. In

addition, molecules exhibit two other types of radiation-induced transitions: vibrational transitions and rotational transitions. Vibrational transitions occur because a molecule has a multitude of quantized energy levels, or vibrational states, associated with the bonds that hold the molecule together. Figure 16 depicts the energies E_1 and E_2, two of the several electronically excited states of a molecule, relative to the energy of the ground state E_0. This phenomenon involves orbital transitions due to ultraviolet and visible radiation. In addition, the relative energies of vibrational states associated with each electronic state are indicated by the lighter horizontal lines. This kind of transitions occurs when the matter interacts with lower energy (larger wavelength) photons (infrared radiation).

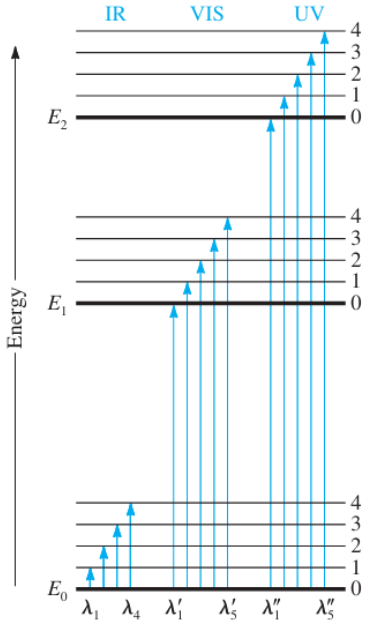


Figure 16. Energy level diagram showing some of the energy changes that occur during absorption of infrared (IR), visible (VIS), and ultraviolet (UV) radiation by a molecular species[217] (taken from ref. 217).

3.2 Infrared absorptions

Infrared radiation generally is not energetic enough to cause electronic transitions, but it can induce transitions in the vibrational and rotational states associated with the ground electronic state of the molecule (Fig. 16). With liquid or solid samples, however, rotation is often hindered or prevented, and the effects of these small energy differences are not detected[217]. In infrared spectroscopy, for a molecule to exhibit infrared absorptions, its electric dipole moment must change during the vibration, which serves as the selection rule for infrared activity[218]. Molecules that meet this criterion, such as heteronuclear diatomic molecules, are considered "infrared-active" because their dipole moment changes as the bond stretches and contracts. Conversely, homonuclear diatomic molecules are "infrared-inactive" as their dipole moment remains constant, regardless of bond length[212]. In general, the greater the polarity of the bond, the stronger its infrared absorption is. The carbonyl bond is very polar, and absorbs very strongly. The carbon-carbon triple bond in most alkynes, in contrast, is much less polar, and thus a stretching vibration does not result in a large change in the overall dipole moment of the molecule.

Infrared absorption bands are not infinitely narrow; several factors contribute to their broadening. Collisions between molecules can broaden the bands. Another factor is the finite lifetime of the states involved in the transition. In quantum mechanics, solving the Schrödinger equation for time-dependent systems reveals that energy states do not have precisely defined energies, leading to lifetime broadening[212]. According to the Heisenberg Uncertainty Principle, the shorter the lifetime of an excited state, the broader the absorption band is, reflecting a less precisely defined energy[212].

3.3 Normal modes of vibration

The interactions of infrared radiation with matter may be understood in terms of changes in molecular dipoles associated with vibrations and rotations. A basic model, a

molecule can be looked upon as a system of masses joined by bonds with spring-like properties. The atoms in the molecules can also move relative to one other, that is, bond lengths can vary or one atom can move out of its present plane. This is a description of stretching and bending movements that are collectively referred to as *vibrations*. The number of ways a molecule can vibrate is related to the number of atoms, and thus the number of bonds, it contains[217]. For a diatomic molecule, only one vibration that corresponds to the stretching and compression of the bond is possible. This accounts for one degree of vibrational freedom. Polyatomic molecules containing many (N) atoms will have 3N degrees of freedom[212] (Table 1).

Table 1. Degrees of freedom for polyatomic molecules.

Type of degrees of freedom	Linear	Non-linear
Translational	3	3
Rotational	2	3
Vibrational	3N-5	3N-6
Total	**3N**	**3N**

A molecule can only absorb radiation when the incoming infrared radiation is of the same frequency as one of the fundamental modes of vibration of the molecule. This means that the vibrational motion of a small part of the molecule is increased while the rest of the molecule is left unaffected.

Vibrations can involve either a change in bond length (stretching) or bond angle (bending) (Fig. 17). Some bonds can stretch in-Phase (symmetrical stretching) or out-of-Phase (asymmetric stretching), as shown in Figure 17a. Bending vibrations also contribute to infrared spectra and these are summarized in Figure 17b. Taking the water molecule as an example, the hydrogens can move in the same direction or in opposite directions in this plane, here the plane of the page. This results in in-plane and out-of-plane bending vibrations, as illustrated in Figure 17b. There will be many different vibrations for even

fairly simple molecules. The complexity of an infrared spectrum arises from the coupling of vibrations over a large part of or over the complete molecule. Such vibrations are called skeletal vibrations. For more complex molecules, the analysis becomes simpler since hydrogen atoms may be considered in isolation because they are usually attached to more massive, and therefore, more rigid parts of the molecule[212]. Bands associated with skeletal vibrations are likely to conform to a pattern or fingerprint of the molecule as a whole, rather than a specific group within the molecule.

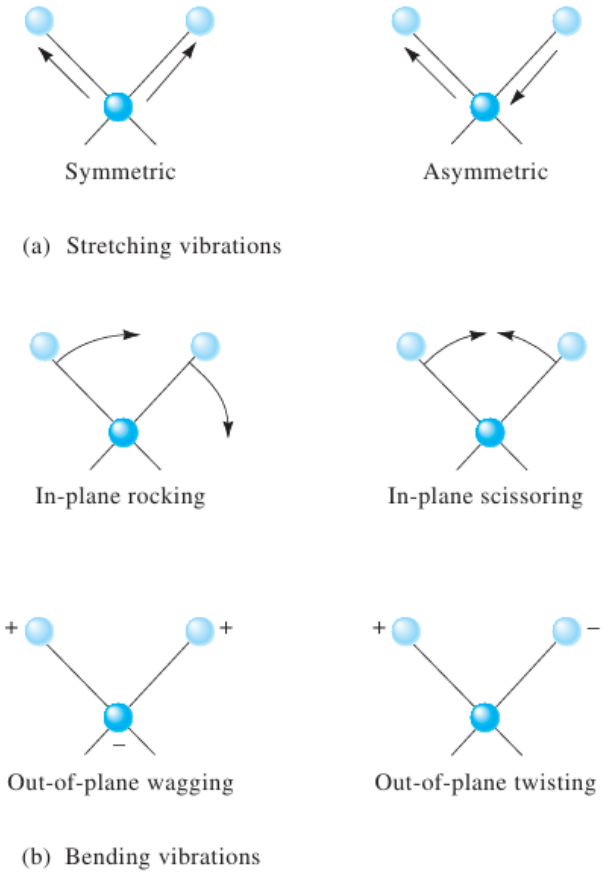


Figure 17. Types of molecular vibrations. (a) Symmetric and asymmetric stretching vibrations and (b) in-plane and out-of-plane bending vibrations[217] (taken from ref. 217).

3.4 Complicating Factors

There are a number of factors that may complicate the interpretation of infrared spectra (e.g. overtone and combination bands, Fermi resonance, coupling and vibration-rotation bands). These factors should be considered when studying spectra as they can result in important changes to the spectra and may result in the misinterpretation of bands.

3.4.1 Overtone and anharmonic bands

The sound we hear is a mixture of harmonics, that is, a fundamental frequency mixed with multiples of that frequency. Overtone bands in an infrared spectrum are analogous and are multiples of the fundamental absorption frequency. Up to now only harmonic vibrations have been discussed (Fig. 18a). If anharmonicity is present, vibration will be periodic but not a simple sine or cosine wave (Fig 18b), thus the vibrational frequency will no longer be completely independent of amplitude as it is in the harmonic case. However, any such periodic function can be resolved into simple sine or cosine components where the frequencies are integral multiples of the fundamental vibrational frequencies (Fourier analysis) (Fig. 18c). This means that if the molecular vibration is anharmonic, the dipole moment will oscillate with the fundamental frequency and integral multiples thereof[218].

These are called the fundamental, first overtone, second overtone, etc. The energy required for the first overtone is twice the fundamental, assuming evenly spaced energy levels. Since the energy is proportional to the frequency absorbed and this is proportional to the wavenumber, the first overtone will appear in the spectrum at twice the wavenumber of the fundamental[212]. The intensity of an overtone absorption is dependent on the amount of anharmonicity in the vibration. Overtones can be detected in the infrared spectrum but they are usually quite weak, which implies that although molecular

vibrations are measurably anharmonic, the anharmonicity is not great and can be ignored in a reasonably good first approximation[218].

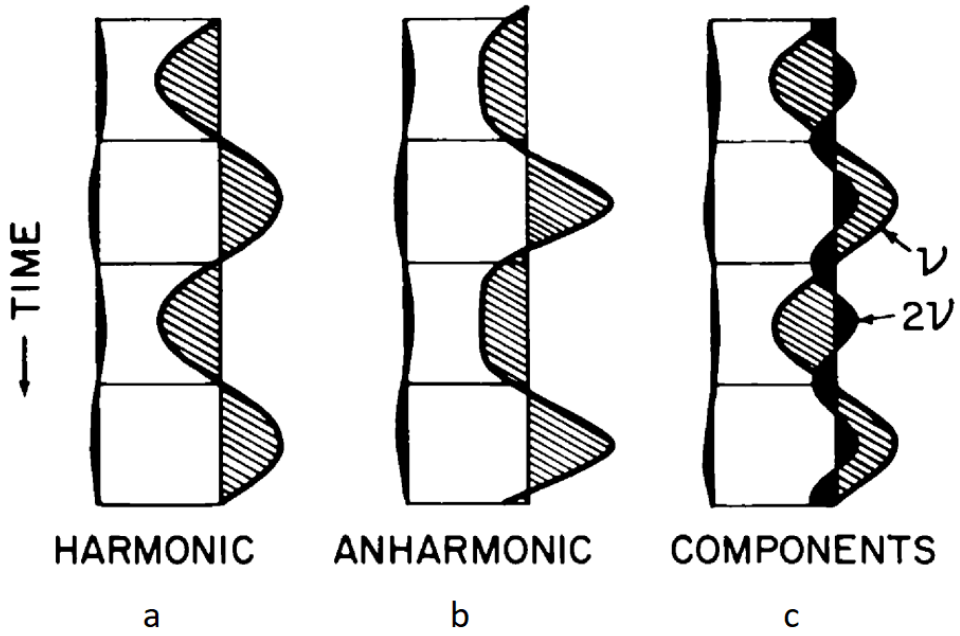


Figure 18. Plots of mass displacement versus time for (a) harmonic, (b) anharmonic vibrations and (c) the main components of the anharmonic curve in the middle[218] (taken from ref. 218).

3.4.2 Fermi resonance

The Fermi resonance effect usually leads to two bands appearing close together when only one is expected. When an overtone band has the same frequency as, or a similar frequency to, a fundamental, two bands appear, split either side of the expected value and are of about equal intensity. The effect is greatest when the frequencies match, but it is also present when there is a mismatch of a few tens of wavenumbers. The two bands are referred to as a Fermi doublet[212].

3.4.3 Coupling

Vibrations in the skeletons of molecules become coupled. Such vibrations are not restricted to one or two bonds, but may involve a large part of the carbon backbone and oxygen or nitrogen atoms if present, especially when the motions are in the same part of the molecule. The energy levels mix, hence resulting in the same number of vibrational modes, but at different frequencies, and bands can no longer be assigned to one bond. This is very common and occurs when adjacent bonds have similar frequencies. Coupling commonly occurs between C–C stretching, C–O stretching, C–N stretching, C–H rocking and C–H wagging motions[212].

3.4 Fourier-Transform infrared spectrometers

Fourier Transform Infrared (FTIR) spectrometers are powerful analytical instruments used to obtain an infrared spectrum of absorption, transmission, or emission of a solid, liquid, or gas. By measuring how a sample absorbs light across the infrared range, FTIR spectrometers provide detailed information about the molecular composition, functional groups, and chemical bonds present in a material. The technique is widely used due to its ability to analyze both organic and inorganic compounds quickly and with minimal sample preparation. Modern IR spectroscopic instruments are widely equipped with the FTIR design to accelerate the scanning and data collection process, while the fundamental unit of the FTIR spectrometer constitutes the Michelson interferometer[217]. The detailed optical layout of the ATR-FTIR spectrometer is depicted in Figure 19 along with the path ways of generating the IR spectrum. The next lines lay the description of each component.

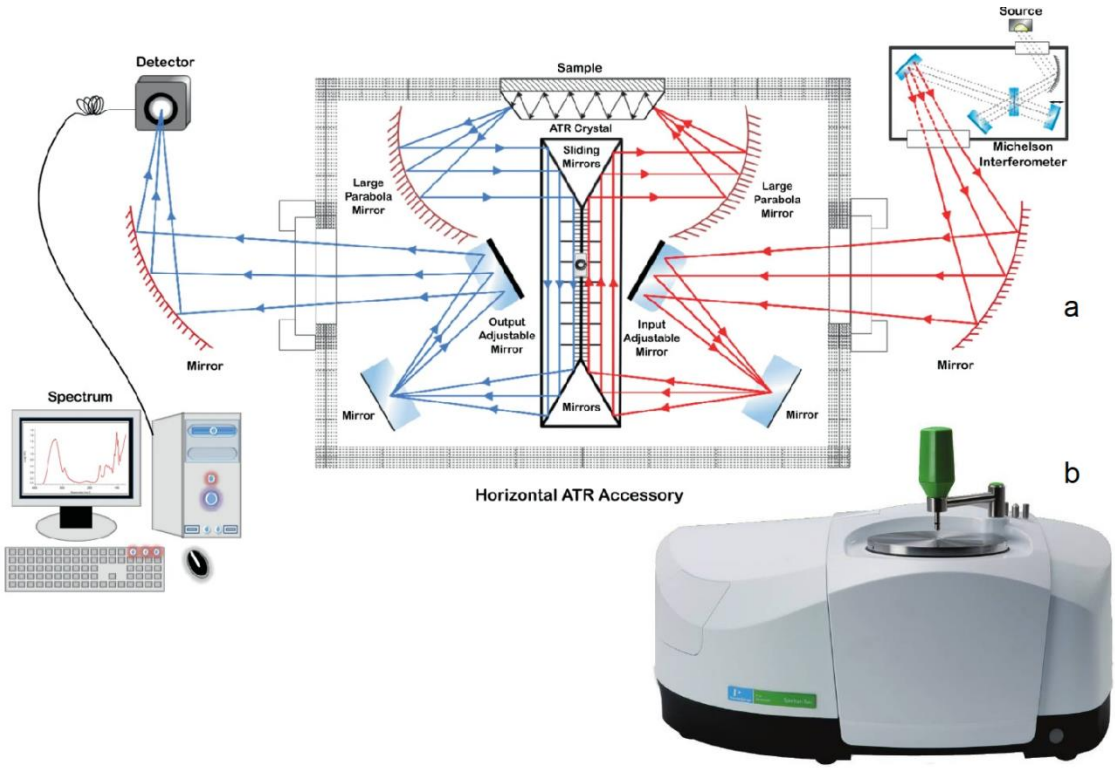


Figure 19. (a) Schematic optical layout of the FTIR spectrometer mounted with a trapezoidal crystal ATR accessory and (b) photograph of a typical ATR-FTIR instrumentation.

3.4.1 Michelson interferometers

The most common interferometer used in FTIR spectrometry is a Michelson interferometer (invented by the American physicist Albert A. Michelson), is a precision instrument that produces interference fringes by splitting a light beam into two parts and then recombining them after they have traveled different optical paths[212]. Michelson interferometer consists of two perpendicularly plane mirrors, one of which can travel in a direction perpendicular to the plane. A semi-reflecting film (half-silvered mirror), the beamsplitter, bisects the planes of these two mirrors. Figure 20 depicts the interferometer and the path of a light beam from a single point on the extended source S, which is a ground-glass plate that diffuses the light from a monochromatic lamp. The

beam strikes the beamsplitter M, where half of it is reflected to the side and half passes through it. The reflected light travels to the movable plane mirror M_1, where it is reflected back through beamsplitter M to the detector (the observer acts as a simplified version of the detector). The transmitted half of the original beam is reflected back by the stationary mirror M_2 and then toward the observer by beamsplitter.

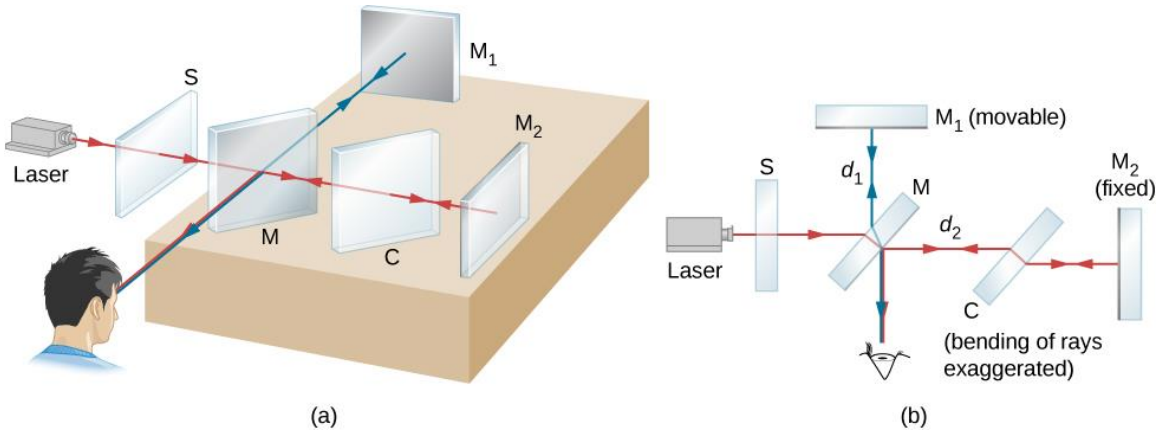


Figure 20. (a) The Michelson interferometer and (b) a planar view of the interferometer.

Notice from the figure that one beam passes through M three times and the other only once, since the reflecting surface of the beam splitter is the surface on the lower right (more clearly in Fig. 20b). To ensure that both beams traverse the same thickness of glass, a compensator plate C of transparent glass is placed in the arm containing M_2. This plate is a duplicate of M (without the silvering reflective surface) and is usually cut from the same piece of glass used to produce M. With the compensator in place, any Phase difference between the two beams is due solely to the difference in the distances they travel.

If two waves (beams) simultaneously propagate through the same region of space, the resultant field at any point in that region is the **vector sum** of the field of each wave. This is the principle of superposition. If two beams emanate from a common source, but travel over two different paths to a detector (one path is constant and the second one changes due to moving mirror), the field at the detector will be determined by the optical

path difference. This phenomenon concludes to **Phase difference** between the transmitted and reflected beam interfere destructively or constructively[212].

3.4.2 Fourier-Transformation in IR spectroscopy

Fourier-transform infrared (FTIR) spectroscopy is based on the idea of the interference of radiation between two beams to yield an interferogram. The latter is a signal produced as a function of the change of pathlength between the two beams.

An interferogram has a 'center-burst', also called the 'zero-path-difference' point, or 'ZPD'. This corresponds to the place where **maximum interference** is produced by the moving mirror in the instrument (Fig. 21)[219]. This center-burst can be placed in the middle or near to the beginning of the interferogram. The placement of the center-burst is determined by when the instrument starts data collection during the mirror travel. When the center-burst is in the middle, the interferogram is called 'double-sided' or 'symmetric'. When it is placed at the beginning of the interferogram, it is called 'single-sided' or 'asymmetric'.

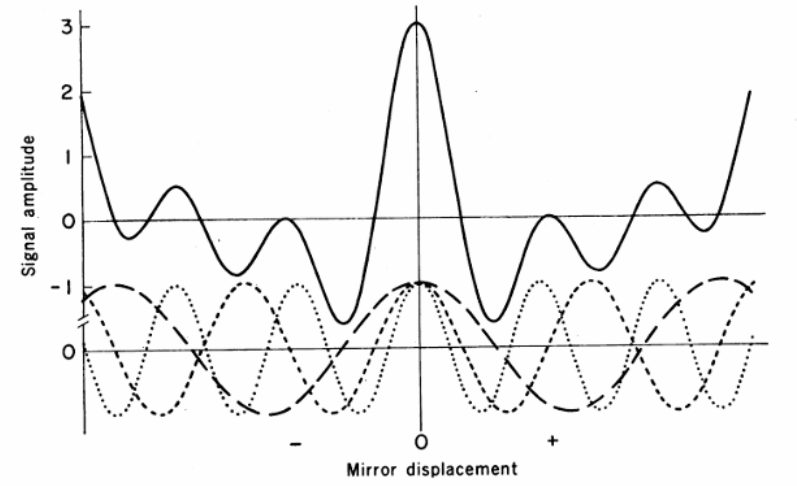


Figure 21. Observed signal from a Michelson interferometer as a function of mirror displacement for an incident wave consisting of three discrete frequencies. This signal is

the sum of the three cosine wave signals that would arise from each frequency separately, as indicated[219] (taken from ref. 219).

The two domains of distance and frequency are interconvertible by the mathematical method of ***Fourier-transformation***. In Figure 22 is summarized how the polystyrene film interferogram is transformed into a transmittance spectrum *via* Fourier analysis. A detailed documentation about Fourier transformation is described in *Section 5*.

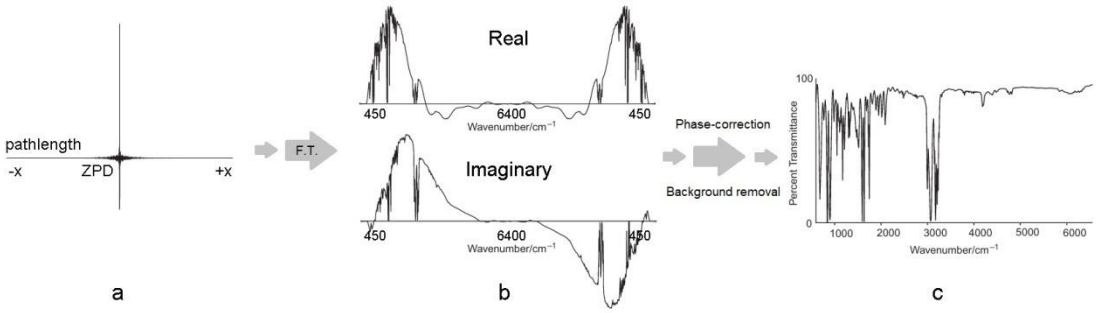


Figure 22. Pictorial essay of transformation and Phase correction by the Mertz method: (a) double-sided interferogram of a polystyrene film, (b) real and imaginary portions of the complex FT, (c) transmittance spectrum after the single-beam spectrum of sample is ratioed against a single-beam reference spectrum (note that only one-half of each curve of (c) has been retained since each half is a mirror image of the other and no information is lost when one half is discarded)[220] (taken from ref. 220).

3.5 Attenuated Total Reflectance spectroscopy

Transmission spectroscopy is the oldest and most straightforward infrared method. This technique is based upon the absorption of infrared radiation at specific wavelengths as it passes through a sample. It is possible to analyze samples in the liquid, solid or gaseous forms when using this approach. Reflectance techniques may be used for samples that are difficult to analyze by the conventional transmittance methods.

Reflectance methods can be divided into two categories. Internal reflectance measurements can be made by using an attenuated total reflectance cell in contact with the sample. There is also a variety of external reflectance measurements which involve an infrared beam reflected directly from the sample surface.

When a propagating wave (e.g. infrared light beam) hits the surface between two optical media which are characterized by two different refractive indices at a certain angle of incidence, the light is totally reflected. This angle is called the **critical angle**. Attenuated Total Reflectance (ATR) spectroscopy utilizes the phenomenon of total internal reflection. A beam of radiation entering a crystal will undergo total internal reflection when the angle of incidence at the interface between the sample and crystal is greater than the critical angle, where the latter is a function of the refractive indices of the two surfaces (Fig. 23)[212]. The IR radiation travels through the crystal and interacts with the sample on the surface in contact with the ATR crystal. The beam penetrates a fraction of a wavelength beyond the reflecting surface and when a material that selectively absorbs radiation is in close contact with the reflecting surface, the beam loses energy at the wavelength where the material absorbs. The resultant attenuated radiation is measured and plotted as a function of wavelength by the spectrometer and gives rise to the absorption spectral characteristics of the sample.

3.5.1 ATR cells

Many different ATR accessories are available for FTIR spectrometers. They can be divided into ATR cells with a single reflection (one bounce) and cells with multiple reflections (multiple bounce, 25 or more) (Fig. 23)[212]. Depending on the application and the measured samples, different materials are used as the ATR crystal. The crystals used in ATR cells are made from materials that have low solubility in water and are of a very high refractive index. Such materials include zinc selenide (ZnSe), germanium (Ge), diamond (Fig. 23c,d,e,f) and thallium–iodide (KRS-5)[221].

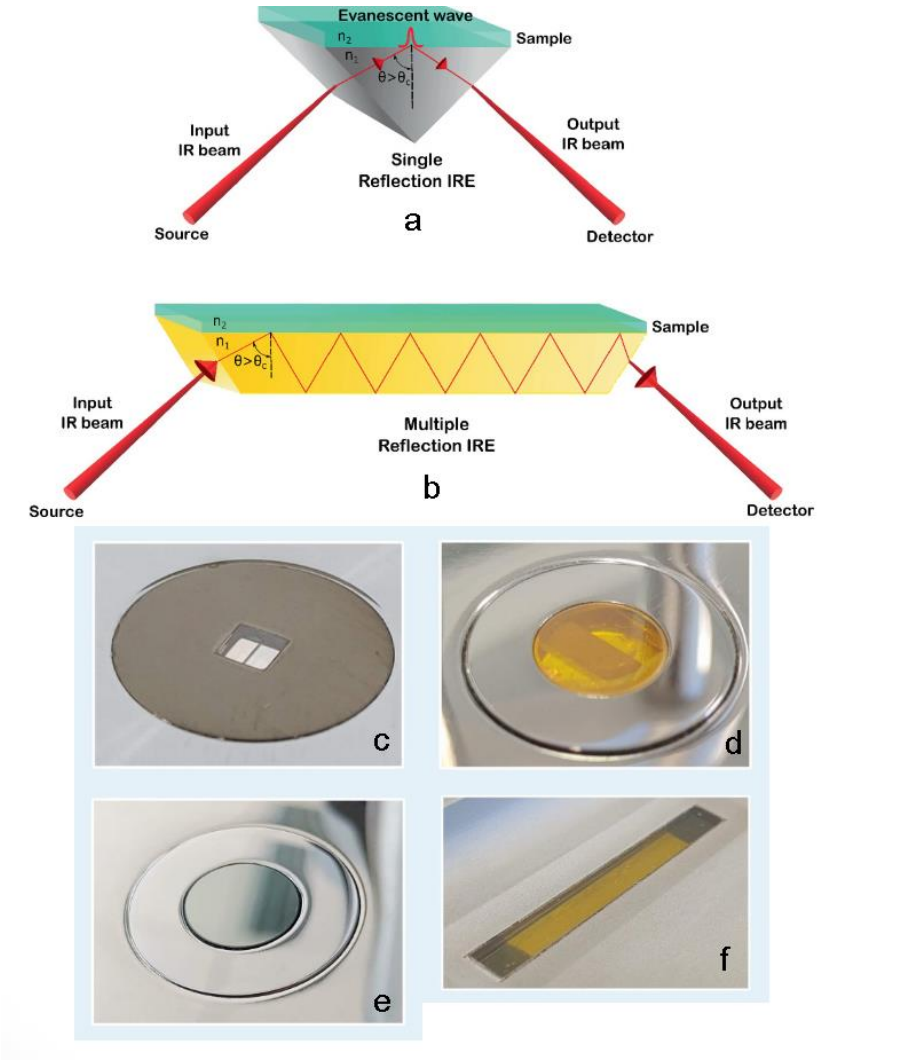


Figure 23. (a) Optical ray diagram of a single reflection and (b) multiple reflection internal reflection element geometries of ATR-FTIR spectrometer. Close-up of a (c) diamond, (d) zinc selenide (ZnSe), (e) germanium and (f) multi-reflection (ZnSe) ATR crystals[222] (taken from ref. 222).

In single-bounce ATR cell the IR beam and the sample interact only once at a single point of reflection. Modern FTIR spectrometers with sufficiently large signal-to-noise ratio (which is a measure of signal quality) allow for reasonable spectra even with only one bounce. The clear advantage of single-bounce ATR cells is the minimal amount of sample needed for a measurement. Single-bounce attenuated total reflectance cells are commonly used for solid samples or powders or whenever only small sample volumes of

liquids are available. Common single-bounce ATR accessories consist of an ATR crystal with an interface surface area of about 2 mm^2 and a clamp which is used to uniformly press solid or powder samples onto the ATR crystal's surface (Fig. 19b)[223].

In multiple-bounce ATR cells the IR beam is reflected multiple times. Each reflection on the ATR crystal's surface exhibits an evanescent wave which interacts with the sample[221]. Since the interactions are independent of one another the absorptions are additive and the sensitivity of the recorded spectrum can be increased due to significantly higher signal-to-noise ratios.

Different designs of ATR cells allow both liquid and solid samples to be examined. It is also possible to set up a flow-through ATR cell by including an inlet and outlet in the apparatus[212]. This allows for the continuous flow of solutions through the cell and permits spectral changes to be monitored with time.

3.5.2 Evanescent wave

Total internal reflection does not explain the interaction of the IR beam and sample because the IR beam never leaves the ATR crystal. Instead, the interaction of the IR beam and sample occurs through an evanescent field, often called *evanescent wave* (Fig. 24a). Upon total reflection of the incident light at the interface where the crystal touches the sample a small fraction of the light extends into the sample as an evanescent wave. When a wave cannot propagate regularly into the sample, it is concentrated in proximity to the point of total reflection and starts decaying exponentially[224]. The evanescent wave sticks out into the sample. In areas where the sample is in contact with the evanescent wave specific parts of the IR beam are absorbed based on the sample's composition. The totally reflected IR light lacks the absorbed parts and thus is attenuated.

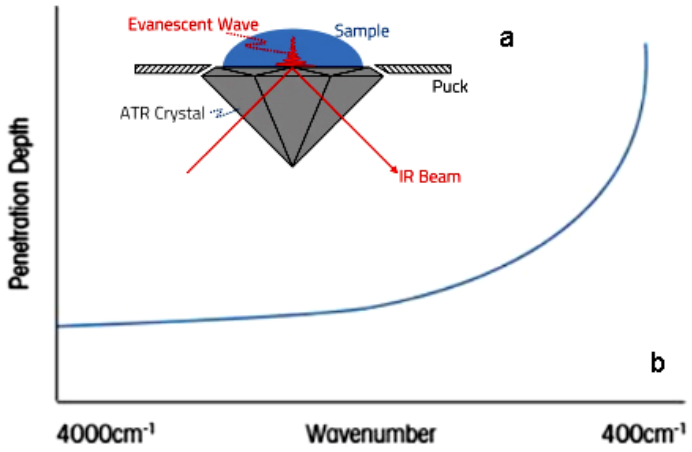


Figure 24. (a) Graphical representation of a single bounce ATR and (b) depth penetration of the evanescent wave depending on the specific wavenumber energy (wavelength, λ).

The depth of penetration is defined as the distance to the point at which the evanescent wave's amplitude has decreased to 1/e (i.e. about 37%) of its maximum value[224,225]. The depth of penetration in ATR spectroscopy is a function of the wavelength, λ (Fig. 24b); the refractive index of the crystal, n_2; and the angle of incident radiation, θ. The depth of penetration, dp, for a non-absorbing medium is given by Eq. 2:

$$d_p = \frac{\frac{\lambda}{n_1}}{2\pi\sqrt{\sin\theta - (\frac{n_1}{n_2})^2}} \quad (2)$$

where n_1 is the refractive index of the sample[212].

3.6 Qualitative applications of ATR-FTIR spectroscopy

An infrared absorption spectrum, even one for a relatively simple compound, often contains a bewildering array of sharp peaks. Peaks useful for the identification of functional groups are located in the shorter-wavelength region of the infrared (fingerprint region), where the peaks positions are only slightly affected by the carbon skeleton of the molecule[217]. This region of the spectrum thus abounds with information regarding the

overall constitution of the molecule under investigation. Table 2 gives the positions of characteristic peaks for some common functional groups.

Table 2. Some characteristic infrared absorption peaks[217] (taken from ref. 217).

		Absorption Peaks	
Vibration	**Functional Group**	**Wavenumber, cm^{-1}**	**Wavelength, µm**
O—H	Aliphatic and aromatic	3600-3000	2.8-3.3
NH_2	Also secondary and tertiary	3600-3100	2.8-3.2
C—H	Aromatic	3150-3000	3.2-3.3
C—H	Aliphatic	3000-2850	3.3-3.5
C≡N	Nitrile	2400-2200	4.2-4.6
C≡C—	Alkyne	2260-2100	4.4-4.8
COOR	Ester	1750-1700	5.7-5.9
COOH	Carboxylic acid	1740-1670	5.7-6.0
C═O	Aldehydes and ketones	1740-1660	5.7-6.0
$CONH_2$	Amides	1720-1640	5.8-6.1
C═C—	Alkene	1670-1610	6.0-6.2
Ǿ—O—R	Aromatic	1300-1180	7.7-8.5
R—O—R	Aliphatic	1160-1060	8.6-9.4

The ATR-FTIR spectroscopic technique has undergone significant advancements, enabling detailed analyses of molecular bonding, surface adsorption, interactions, molecular orientation, kinetics, and structural parameters of samples[226]. ATR-FTIR spectroscopy is a simple, non-invasive tool across a wide range of scientific disciplines. In

the healthcare sector, it has been integrated with chemometrics and multivariate analysis to enhance point-of-care diagnostics, offering highly sensitive detection capabilities for various bio-analytes and disease biomarkers[227–229]. Due to the control of the penetration depth, ATR-FTIR spectroscopy can selectively probe components within layered surfaces, such as thin biofilms, peptide layers, or electrolyte interfaces[221,230]. The potential of ATR-FTIR spectrometry has also been explored in a variety of infrared imaging applications, including histopathology, live cell and tissue analysis, and the identification of material surface properties[231–233].

4. Nuclear Magnetic Resonance (NMR) spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is a highly versatile and widely used analytical technique that exploits the magnetic properties of specific atomic nuclei. The fundamental principle of NMR is based on the fact that certain nuclei, when placed in an external magnetic field, can occupy distinct nuclear spin states. NMR detects transitions between these spin states, which are characteristic of the nucleus being studied and provide insight into its chemical environment. However, it is important to note that NMR is only applicable to nuclei with a non-zero spin quantum number ($I \neq 0$); nuclei with $I=0$ are "invisible" to NMR spectroscopy because they lack the magnetic properties required for signal generation.

Due to these unique properties, NMR has become an invaluable tool for determining molecular structures, monitoring chemical reactions, and investigating metabolic processes in living cells. Its applications extend across a broad range of fields, including medicine, biochemistry, physics, and industry, making it a cornerstone of research and development in nearly every scientific discipline.

4.1 Magnetic resonance

4.1.1 Nuclear spins

Nuclei possess a positive charge, and many behave as though they are spinning. When a charged particle is in motion, it generates a magnetic moment and produces a corresponding magnetic field. Therefore, a spinning nucleus acts like a tiny bar magnet, with its magnetic moment[234] aligned along the axis of rotation (Fig. 25a). This property is commonly referred to as nuclear spin.

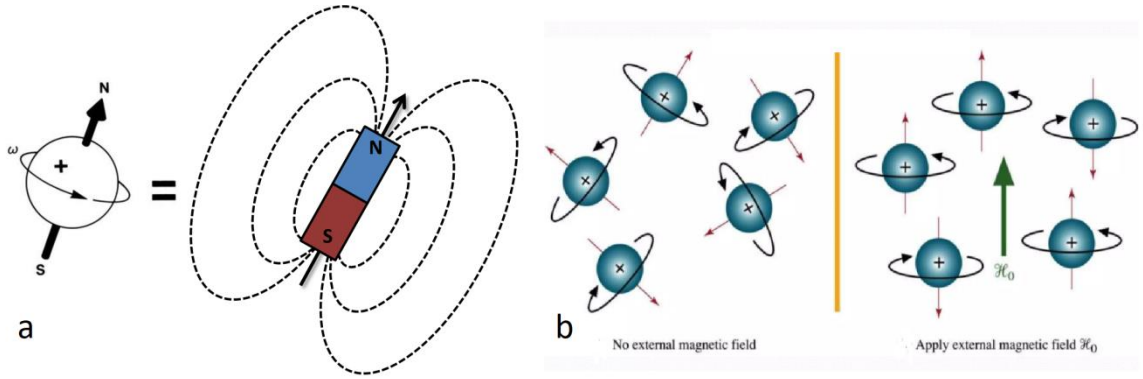


Figure 25. (a) A charged nucleus rotating with angular frequency ω creates a magnetic field B and is equivalent to a small bar magnet whose axis is coincident with the spin rotation axis. (b) Orientation of spinning nuclei in absence and presence of external magnetic field.

When such a "tiny magnet" is placed within the field of a much larger external magnet, its orientation is no longer random. Instead, there will be a preferred orientation—the most probable parallel orientation[235]. However, it is also possible for the nucleus to align itself precisely 180° opposite to this preferred direction (Fig. 25b). In

scientific terms, the more favorable orientation corresponds to the lower-energy state, while the opposite orientation represents the higher-energy state.

This two-state model is applicable to most nuclei of biological importance, such as ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P, all of which have a nuclear spin quantum number (I) of ½[235]. According to quantum mechanics, nuclei with I=½ can only occupy one of these two distinct energy states when exposed to an external magnetic field—no intermediate states are possible.

4.1.2 The resonance phenomenon

The small nuclear magnet can spontaneously "flip" between its two possible orientations (or energy states) while in the presence of a large external magnetic field, although this flipping occurs infrequently (Fig. 26a)[236]. However, when energy equivalent to the difference between the two nuclear spin energy levels (ΔE) is applied to the nucleus—or, more typically, to a group of nuclei—this flipping between energy states is greatly enhanced (Fig. 26a). This energy is delivered through a short pulse of radiofrequency (RF) irradiation, typically lasting several microseconds.

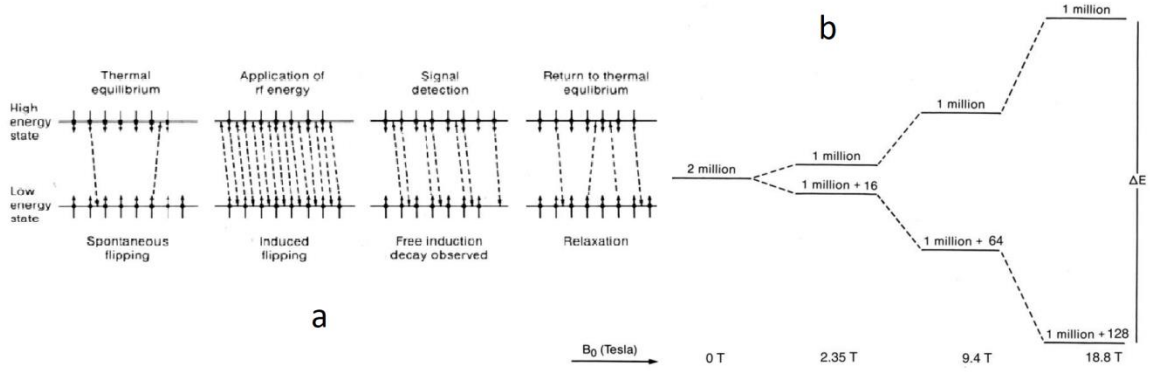


Figure 26. (a) The nuclei resonance process at thermal equilibrium and after the RF pulse. (b) Dependence on magnetic field strength B_0 of the separation of nuclear energy levels (ΔE) for spin I=½ and the relative populations of the energy levels assuming one has approximately two million protons in the sample[237] (taken from ref. 237).

The absorption of RF energy by the nuclear spins induces transitions between the two energy levels, leading to flipping both from the lower energy state to the higher energy state and vice versa. This back-and-forth flipping is a fundamental feature of the resonance process in NMR. As the nuclear spins absorb energy, they generate a voltage that can be detected by a coil of wire tuned to the appropriate frequency. This signal is then amplified and recorded as a free induction decay (FID). Eventually, relaxation processes return the nuclear spin system to its thermal equilibrium, provided no additional RF pulses are applied[236]. The energy required to induce nuclear spin flipping and generate an NMR signal corresponds exactly to the energy difference between the two spin orientations. This energy is dependent on the strength of the external magnetic field (B_o) in which the nucleus is placed, as described by Eq. 3:

$$\Delta E = \frac{\gamma h B_0}{2\pi} \tag{3}$$

where *h* is Planck's constant (6.63 x 10^{-27} erg·sec). This equation, also known as the Bohr condition, allows the frequency (ν_o) of the nuclear transition to be written as Eq. 4:

$$\nu_0 = \frac{\gamma B_0}{2\pi} \tag{4}$$

This equation is often referred to as the Larmor equation, with $\omega_o = 2\pi\nu_o$ being the Larmor resonance angular frequency[235]. The gyromagnetic ratio (γ) is a constant that is unique to each type of nucleus and directly relates to the strength of the nucleus's magnetic moment. In the magnetic fields typically used in NMR experiments, the resonance frequencies required to fulfill the conditions of the Larmor equation fall within

the RF range. For example, in a magnetic field strength of 14.1 T, the resonance frequency for ^{1}H nuclei is 600 MHz, for ^{15}N it is 60.8 MHz, and for ^{13}C it is 151 MHz[237].

4.1.3 Sensitivity and the Boltzmann equation

As previously mentioned, the nuclear spin, which behaves like a small bar magnet, can align in one of two possible orientations within an external magnetic field. The extent to which one orientation (or energy state) is favored over the other depends on both the strength of the nuclear magnetic moment (proportional to the gyromagnetic ratio) and the strength of the external magnetic field (B_o) in which the nucleus is placed[235]. The distribution of nuclei between these two energy states, in the absence of any RF perturbation, is described by the Boltzmann equation (Eq. 5):

$$\frac{N_{upper}}{N_{lower}} = e^{-\frac{h\,v}{k\,T}} \tag{5}$$

where N_{upper} and N_{lower} represent the populations of nuclei in the upper and lower energy states, respectively, k is the Boltzmann constant, and T is the absolute temperature in Kelvin.

To illustrate the impact of the magnetic field strength on the population of nuclear spin states, the distribution of about two million hydrogen nuclei, calculated using Eq. 5, is shown in Figure 26b. In a magnetic field of 18.8 T, which corresponds to a resonance frequency of 800 MHz for protons, and at thermal equilibrium at room temperature, the population ratio is approximately 0.999872[237]. This means that for every 1,000,000 nuclei in the upper energy state, there are 1,000,128 nuclei in the lower energy state. Although this is a very small population difference, it is crucial for NMR. Without this slight excess of nuclei in the lower energy state, NMR signals would not be detectable. However, this small population difference also presents a significant sensitivity challenge for NMR. Since

only the population difference—128 out of 2,000,128 nuclei in this example—contributes to the NMR signal, the vast majority of nuclei cancel each other out, leading to inherently low sensitivity. The low sensitivity of NMR, rooted in this small population difference, is one of its main limitations, particularly in applications to biological systems. However, as seen in Eq. 5, increasing the strength of the magnetic field enhances the population ratio, thereby improving the sensitivity of NMR measurements[234,238].

4.1.4 Magnetization

In a classical mechanical perspective, for a nucleus with a spin quantum number of I=½ placed in a magnetic field of strength B_o, the nucleus's magnetic moment undergoes precession around the z-axis, which is defined by the direction of the applied magnetic field. This precessional motion is depicted in Figure 27a, albeit with a simplified number of nuclear spins (represented by arrows). As discussed earlier, nuclei can align either parallel or antiparallel to the direction of the external magnetic field. Consequently, some nuclear spins precess about the positive z-axis, while others precess about the negative z-axis. The overall magnetization resulting from a real sample is the sum of all these individual nuclear magnetic moments (or spins)[234]. Given that there is a slight excess of nuclei aligned with the magnetic field—i.e., in the lower energy state—the cumulative magnetization, denoted as M_o (Fig. 27a), will be oriented along the positive z-axis. It is this total magnetization, rather than the magnetic moment of an individual nucleus, that generates the measurable NMR signal[236].

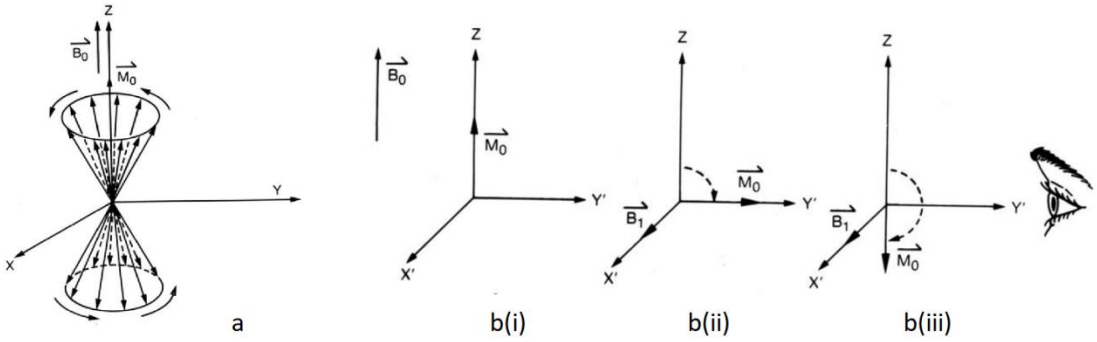


Figure 27. (a) Orientation and precession of nuclear spins (I=½) at thermal equilibrium in a stationary magnetic field B_0 that defines the z-axis. (b) Rotation of the magnetization M_0 in the rotating coordinate system that rotates about the z-axis at the NMR instrument's operating frequency. i) spin system at equilibrium in magnetic field B_0; ii) application of a 90° B_1 pulse; and iii) a 180° pulse.

4.2 The Nuclear Magnetic Resonance experiment

A proton NMR spectrum can be obtained by gradually sweeping either the magnetic field or the frequency to satisfy the resonance condition described by Eq. 3. As this condition is met for the protons in the sample, signals are generated based on their distinct chemical environments. The simplest type of NMR experiment is the continuous wave (CW) experiment, where a constant frequency is applied while the magnetic field is varied[239]. This method probes the energy levels and detects resonance as the field changes. Alternatively, the CW experiment can be conducted by maintaining a constant magnetic field while varying the frequency, though this approach is generally reserved for specific applications. However, most modern NMR spectrometers utilize pulse-based techniques. In these experiments, short bursts of radiation are applied that encompass the entire frequency range needed to excite all nuclei of a particular type, such as ^{13}C. This

pulsed method is more efficient and versatile than the traditional CW technique, and it has become the standard in NMR spectroscopy.

4.2.1 Pulsed NMR

The signal-to-noise ratio (S/N) in NMR can be significantly improved through the technique of signal averaging. In this process, the S/N ratio increases proportionally to the square root of the number of signals averaged ($\sqrt{n}$)[234]. A common and efficient way to acquire an NMR signal is by detecting the free induction decay (FID) after applying a strong RF pulse at the resonance frequency of the nuclei. This FID signal can be immediately followed by another RF pulse to generate a new FID. With the aid of a computer, the subsequent FID signals are captured and averaged with the initial one, resulting in an enhanced S/N ratio.

At equilibrium, the net magnetization vector (M_0) aligns along the direction of the external magnetic field (B_0). This state is referred to as the equilibrium magnetization, M_o. In this orientation, the component of magnetization along the z-axis (M_z) is equal to M_o and is termed longitudinal magnetization. Importantly, in this equilibrium condition, there is no transverse magnetization (M_x or M_y)[236]. To better understand the behavior of the magnetization, we use a rotating reference frame with axes x', y', and z (Fig. 27b). In this frame, the x' and y' axes rotate around the z-axis at the operating frequency (ν_o) of the NMR instrument. By adopting this rotating frame, we can more easily analyze the effect of applying an RF pulse, denoted as B_1, along the x' axis. This pulse results in a measurable signal along the y' axis.

As illustrated in Figure 27b, at thermal equilibrium, the magnetization (M_o) aligns with the external magnetic field (B_o). When an RF pulse (B_1) is applied along the x' axis, it rotates the magnetization in a plane perpendicular to B_1, typically the y'z plane[235,236,239]. The pulse must be applied at the appropriate frequency (ν_0), as described by Eq. 3. The angle of rotation (θ) depends on several factors: the gyromagnetic ratio (γ) of the nucleus,

the amplitude of the RF pulse (B_1), and the duration time (t_w) for which the pulse is applied.

Figure 27b shows the effect of a 90° (π/2) RF pulse, which rotates the magnetization (M_o) by 90°, moving it from the z-axis into the y'-axis of the rotating frame. Such a pulse is referred to as a 90° or π/2 pulse. If the B_1 field is applied for twice the duration, it results in a 180° (π) rotation, inverting the magnetization (M_o).

These rotations also have a quantum mechanical interpretation. For example, in the case of two million protons subjected to a 14.1 T magnetic field, the application of a 90° pulse equalizes the populations of the nuclear spins in the two energy states. A 180° pulse, however, inverts the population, leading to a greater number of spins in the higher-energy state, as demonstrated in Figure 28.

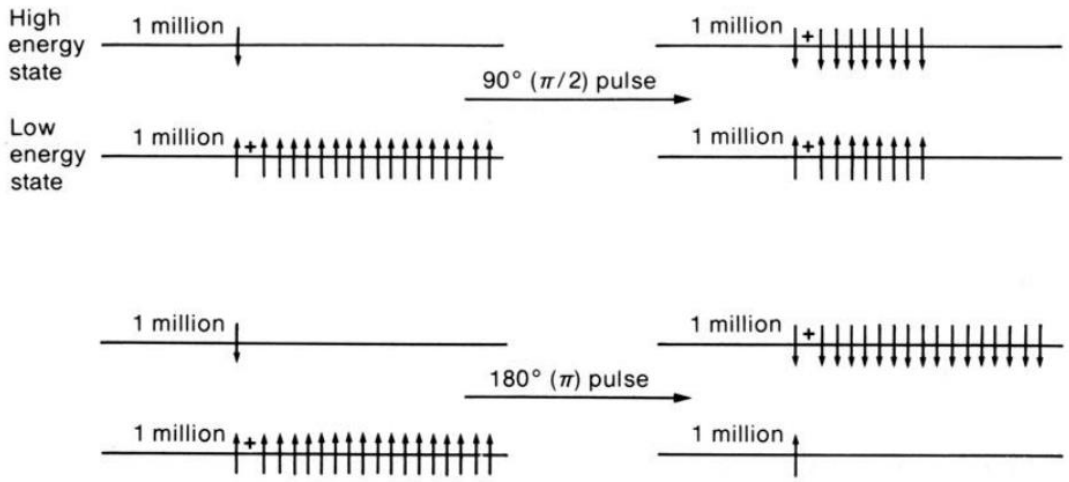


Figure 28. Effect of 90° and 180° RF pulses on the population of nuclear spins in a sample of about two million protons in a magnetic field[237] (taken from ref. 237).

4.2.2 Free Induction Decay (FID)

As long as the bulk magnetization M_0 lies along the z-axis (parallel to the applied magnetic field B_0), no NMR signal can be detected. However, when a RF pulse is applied along the x'-axis, the magnetization vector M is tipped away from the z-axis, creating a

component along the transverse (y'-axis) plane. This transverse component of magnetization is what generates the observable NMR signal.

Since the receiver detects signals along the y'-axis, signal intensity is initially at its maximum immediately after a 90° pulse. As precession continues, the magnetization vector moves around the transverse plane, resulting in sinusoidal variations in the detected signal. When the vector points directly toward the -y' axis, a negative signal is detected, reaching maximum negative amplitude at this orientation (Fig. 29)[235]. This oscillating, decaying signal is called Free Induction Decay (FID) because it decays freely after the RF pulse is turned off.

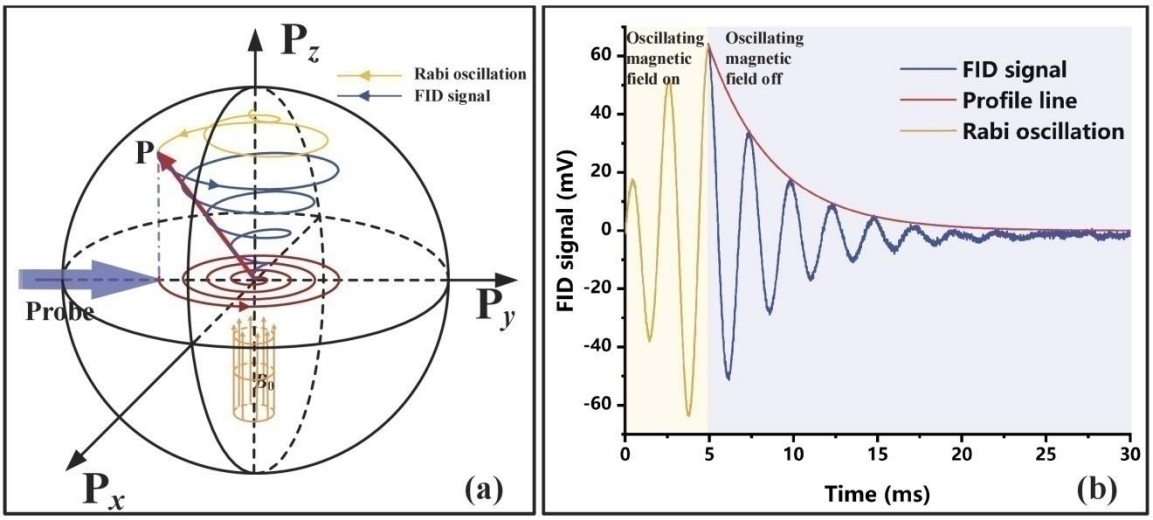


Figure 29. Dynamic evolution of the FID signal with a limited duration oscillating field B_1 and a static magnetic field B_0. (a) The spiral curve shows a trajectory of **P**, where the yellow part represents Rabi oscillation with an **oscillating field turning on**, and the blue part represents FID signal with an **oscillating field turning off**. The red vortex line represents the projection of the polarization trajectory on the x'-y' plane during the FID process. (b) The blue damped oscillation curve depicts the FID signal of y'[240] (taken from ref. 240).

4.2.3 Fourier Transform in NMR

The previously discussed principles of pulsed NMR are straightforward when only one resonance frequency needs to be monitored, such as the ^{19}F signal from fluorouracil bound to thymidylate synthase or the proton signal from water (H_2O) in biological tissues, where the water proton signal is dominant over other observable protons[236]. In these cases, a single frequency dominates the NMR signal.

However, in many practical applications, the free induction decay (FID) is a time-domain signal that often contains contributions from multiple nuclei resonating at different frequencies. For example, in a protein sample, different types of nitrogen nuclei (^{15}N) may produce several distinct signals[236]. To obtain a more interpretable frequency-domain spectrum from such a complex signal, the Fourier transform is applied to the signal-averaged FID.

By applying a Fourier transform, we convert the time-domain FID data into a frequency-domain spectrum, where individual resonances are separated and displayed based on their frequencies[234]. This process is illustrated in Figure 30, where the signal-averaged FIDs (shown on the left) are transformed into frequency spectra (on the right), making it easier to distinguish and analyze the contributions from different nuclei. A detailed documentation about Fourier transformation is described in *Section 5*.

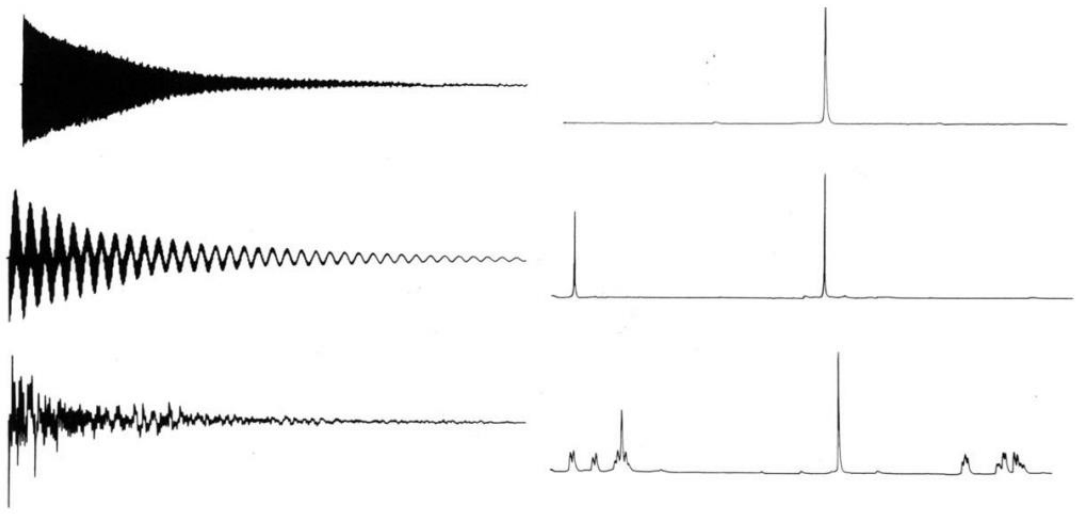


Figure 30. The free induction decay (FID) is on the left and its Fourier transform (usual frequency spectrum) is on the right.

4.3 Core components of an NMR spectrometer

Nuclear Magnetic Resonance (NMR) spectroscopy is an advanced analytical technique used for the structural and dynamic study of molecules. The performance and accuracy of an NMR experiment depend significantly on the instrument's components. This *Section* provides a detailed overview of NMR instrumentation (Fig. 31).

Figure 31. A 500 MHz NMR spectrometer (Bruker Ascend 500).

4.3.1 Magnet

The magnet is the core component of an NMR spectrometer, generating the strong magnetic field essential for nuclear resonance. This field aligns nuclear spins, making them responsive to radiofrequency pulses. NMR magnets are typically superconducting, providing both strength and stability. A superconducting magnet operates at low temperatures (around 4K)[241], using liquid helium to achieve superconductivity, which results in zero electrical resistance (Fig. 32). Liquid helium is typically surrounded by a liquid nitrogen (77.4K) container, which acts as a thermal buffer between the room temperature air (293K) and the liquid helium[239]. The resulted lack of resistance ensures a

highly stable and uniform magnetic field, crucial for obtaining reproducible and high-resolution NMR spectra. The field strength ranges today from 4.7 to 23.5 T[239].

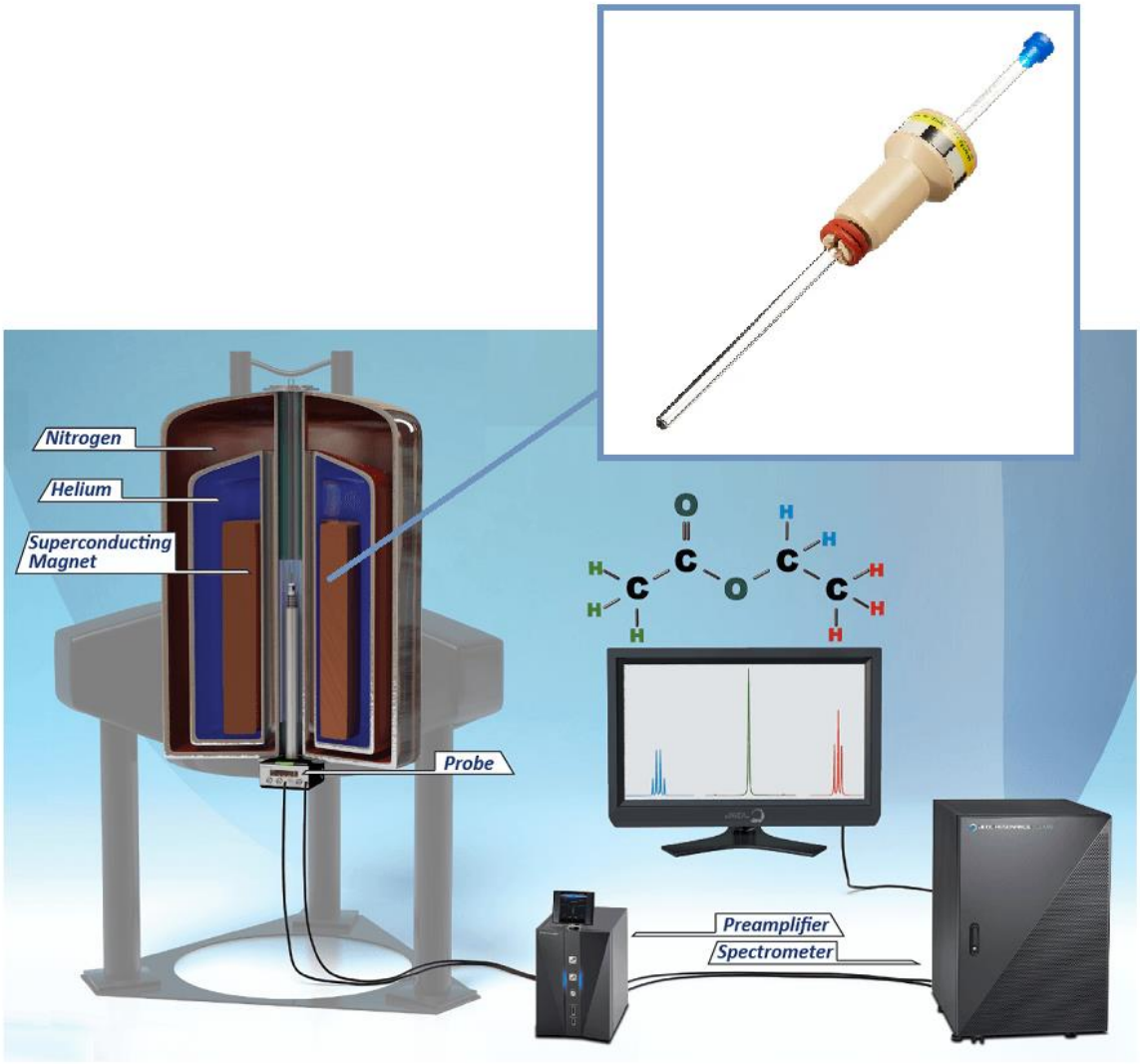


Figure 32. Schematic diagram of an NMR instrument highlighting the liquid helium and liquid nitrogen baths, outer vacuum chamber, superconducting solenoidal magnet and NMR probe with a sample spinner and sample tube.

4.3.2 Probe

The probe is an integral part of the NMR spectrometer, housing the sample spinner (thus the sample) and facilitating RF transmission and detection[242]. It is positioned precisely in the strong magnetic field that is generated by the superconducting magnet (Fig. 32). Serving as the interface between the spectrometer and the sample, NMR probes have three primary functions: holding the sample, exciting nuclear spins with radio frequency (RF) energy, and detecting the NMR signal that emanates from the sample[243]. NMR probes are intricate devices composed of various components, each contributing to the probe's overall functionality in NMR spectroscopy. Understanding these components is essential for comprehending how NMR probes operate and their role in the broader context of NMR spectroscopy.

Probe Body: The probe body, typically cylindrical, houses the internal components of the NMR probe, providing mechanical support and precise alignment for the RF coils and other elements (Fig. 32). It is specially designed to maintain stability and integrity under high magnetic fields and varying experimental temperatures[244]. The material of the probe body is a non-magnetic and resistant to RF interference to ensure optimal probe performance[244].

RF Coils: Radiofrequency coils are the "antennae" of the NMR system, broadcasting the RF signal to the sample and/or receiving the return signal; transmit and receive (transceiver). Saddle-shaped is the most frequently used type of coil in NMR instrumentations. However, surface, Helmholtz pair and bird cage coils (Fig. 33) are presented in other applications of magnetic resonance phenomenon, like MRI[245]. By generating an RF magnetic field when an alternating current flows through them, these coils excite the sample’s nuclear spins. As these spins return to equilibrium, they emit an NMR signal, which the RF coils detect. Careful design and tuning of these coils are crucial, as they need to resonate at frequencies specific to the nuclei under study. This specificity

ensures efficient excitation of the nuclei and precise signal detection, both fundamental for molecular analysis.

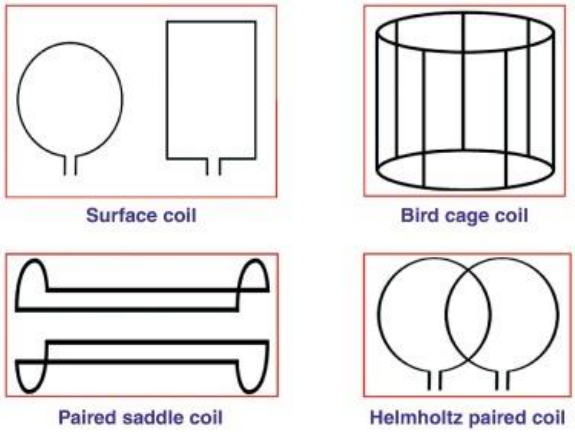


Figure 33. Schematic representation of commonly used radiofrequency coils in the NMR and MRI system[245] (taken from ref. 245).

Sample Tube Holder: The sample tube holder (spinner) secures the NMR sample tube within the probe, ensuring correct positioning in the magnetic field and relative to the RF coils (Fig. 32). Precise placement is crucial for uniform excitation and optimal signal detection. The holder is adaptable to different tube sizes based on probe type, and it facilitates easy insertion and removal of the sample.

Amplifiers: Amplifiers in NMR probes boost the weak NMR signal (from milli Watts to tens or hundreds of Watts) detected by the RF coils before it reaches the spectrometer's main console[239]. Since the raw NMR signal is typically faint, amplification is needed for accurate signal analysis. These preamplifiers are designed to be highly sensitive and minimize noise to preserve signal integrity and optimize the signal-to-noise ratio.

Shimming Coils: Shimming coils are specialized coils that fine-tune the magnetic field within the NMR probe, ensuring it is homogeneous. Any minor inhomogeneities in the magnetic field can distort the NMR spectrum[234]. These inhomogeneities could be caused by the magnet design, materials in the probe, variations in the thickness of the sample tube, sample permeability, and ferromagnetic materials around the magnet[239]. By adjusting the magnetic field at different points within the probe, shimming coils help achieve uniformity, which is essential for obtaining high-resolution spectra.

Temperature Control System: In many NMR experiments, precise temperature control of the sample is necessary, sometimes involving extreme temperatures[238,244]. The temperature control system in an NMR probe regulates sample temperature using heaters, coolers, and sensors, maintaining it at the desired level throughout the experiment. Accurate temperature control is critical for studying temperature-dependent molecular properties and in experiments monitoring reaction kinetics[246,247].

Probe Types: Liquid-state NMR probes are primarily used for samples in a liquid state. These probes are known for their high sensitivity and are optimized for analyzing homogeneous solutions[244]. They offer significant advantages in high-resolution spectroscopy, making them ideal for studying a wide range of organic and biochemical compounds. However, their use is limited to liquid samples, which can be a constraint when dealing with solid or gaseous substances. These probes are commonly employed in organic chemistry, biochemistry, and pharmaceutical research, particularly for examining molecular structures and interactions in solution. Solid-state, cryogenic, HR-MAS and benchtop probes are included in the family of NMR probes[243].

4.4 Industrial applications

Advances in NMR instrumentation continue to expand the reach and utility of NMR in diverse fields of research. Moreover, NMR plays a vital role across numerous industries,

providing precise molecular insights that enhance product development, quality control, and research. In the pharmaceutical sector, NMR is essential for drug discovery, structural verification, and purity assessment, ensuring drug safety and efficacy[248]. The chemical industry benefits from NMR's detailed analysis of polymers and petrochemicals, supporting innovation in materials science[249,250]. In food and beverage, NMR aids in quality control, nutritional analysis, and preservation research[251,252]. Environmental analysis leverages NMR for pollutant detection and monitoring[253]. These applications underscore NMR's critical contribution to scientific advancement and industry standards.

5. Fourier Transformation

The Fourier Transform (FT) is introduced as a powerful mathematical method that allows scientists to convert signals between two domains: **serial** (time or space) and **spectral** (frequency)[254]. This conversion is essential for understanding the hidden structure of signals, whether they come from sound waves, electromagnetic radiation, or molecular data in chemistry. Although originally a complex and tedious procedure, the development of computing power and the Fast Fourier Transform (FFT) algorithm has made it more accessible to scientists and engineers, improving their ability to analyze signals with greater sensitivity, speed, and resolution.

Fourier Transform is not just a mathematical tool but an approach that mimics real-world physical processes. It provides a way to understand and process signals that would otherwise be difficult or impossible to interpret directly. Assume a periodic function f(x) defined on the line of real numbers, or $x \in R$. In general cases or real-life applications, a function like f(x) can have a very complicated behavior which makes the finding of its closed form very difficult or in most cases impossible. This means that there is no closed mathematical expression that describes the complete behavior of the function exactly.

Jean Baptiste Joseph Fourier, a French mathematician and physicist discovered that any complex function (signal), whether periodic or not, can be represented as a sum of simple sine and cosine waves of different frequencies[254]. Using the sine and cosine, the properties of the very complicated function is reduced to the characteristics of the aforementioned trigonometric functions which are very well-known. This decomposition allows scientists to isolate individual components of a signal and understand its underlying structure. The FT is particularly useful when physical processes, like scattering in X-ray crystallography, don't offer direct observations. The FT allows scientists to reconstruct images of atomic structures from diffraction patterns—a task that no optical lens could achieve.

5.1 Serial-Spectral domains

The central idea behind the Fourier Transform is that many physical processes can be described in two equivalent ways: in the serial domain (over time or space) or in the spectral domain (frequency). In the serial domain, data is represented as a sequence of events occur one after another, either over time (e.g., the changing pressure amplitude of a sound wave at a given point) or across space (e.g., a densitometer measuring optical density across a plate). The spectral domain, on the other hand, describes how much of each frequency (or spatial frequency) is present in the signal[254].

For example, in time domain, we perceive the flashing of a lighthouse light as a periodic event in time, a serial behavior. However, the light's color, which is determined by its wavelength, is a spectral characteristic. In the spatial domain, we recognize an image by its arrangement of pixels (serial), but describe a fabric by the number of threads per unit length, a spatial frequency.

As mentioned earlier, any physically meaningful function can be constructed by adding together simple periodic sine or cosine waves. This idea, though surprising at first, makes sense when we consider something like the sound produced by a wind instrument. When a wind instrument is played, the instrument produces a sound with multiple pitch components, each corresponding to a periodic signal in the spectral domain, which together create the overall sound we hear in the serial domain (in this case, time). Fourier's insight bridged the gap between serial and spectral descriptions, making it possible to analyze complex phenomena in both domains.

5.2 Periodic an aperiodic signals

Fourier theory becomes most straightforward when applied to periodic signals. In this case, the periodic function can be broken down into a sum of sine waves: a fundamental wave that shares the same period as the signal, plus an infinite series of higher **harmonics** of the fundamental frequency[218]. Each of these harmonics has its own amplitude and Phase, and their combination recreates the original signal. The relation between the two domains, serial and spectral, is shown rather vividly, in Figure 34 as a three-dimensional representation of amplitude against either time or frequency, for a periodic signal. This illustrates how the original signal can be regarded as constructed by the summation of its spectral components. The Fourier series is given by Eq. 6:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) \qquad (6)$$

where a_0, a_n and b_n are the Fourier coefficients.

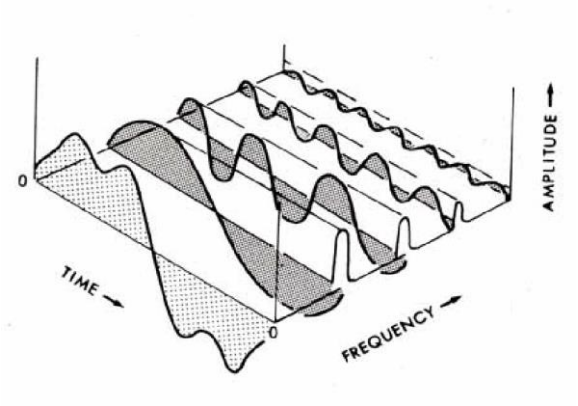


Figure 34. Decomposition of a waveform in both time (serial domain) and frequency (spectral domain)[254] (taken from ref. 254).

For aperiodic signals, the Fourier series approach doesn't work directly. Instead, the Fourier Transform generalizes the concept by considering an infinite range of frequencies. If the signal (function) is aperiodic, it can be considered as a periodic function with its period to lengthen indefinitely, when the fundamental frequency will decrease correspondingly to zero[254,255]. For example a guitar that grows ever longer, with an accompanying decrease in the fundamental frequency of the vibrations of its strings. The result of a Fourier transform is a **continuous function** (Fig. 35c,d) that represents the signal's frequency components over an infinite frequency range. Here, the infinite sum of the series synthesis of f(x) (the case of a periodic signal) will then become an integral Eq. 7:

$$F(s) = \int_{-\infty}^{+\infty} f(x) e^{-i2\pi s x} dx \qquad (7)$$

Here, F(s) gives the amplitude and Phase of each frequency s, producing a continuous frequency spectrum (Fig. 35c,d) and resulting the transform between the serial function, f(x), and the spectral function, F(s).

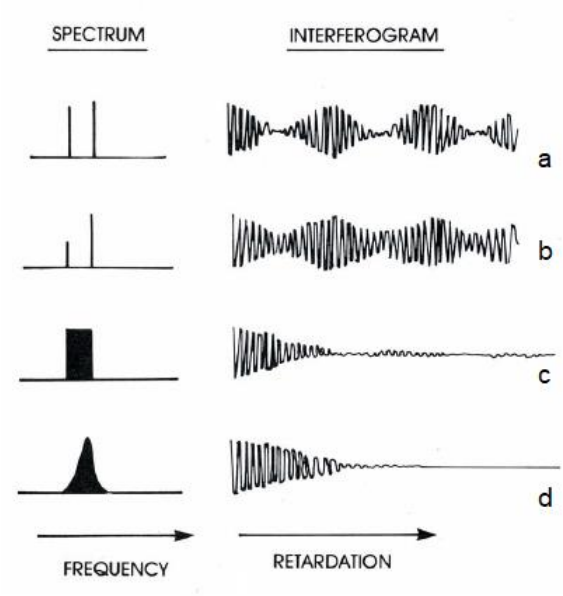


Figure 35. Fourier transform spectra and the corresponding interferograms in the case of periodic (a and b) and aperiodic (c and d) signal[254] (taken from ref. 254).

This allows scientists to analyze signals with continuous variations, such as the random noise in a chemical experiment or the gradual change in light intensity over space. In both periodic and aperiodic cases the central problem of Fourier analysis is the determination of the various coefficients.

5.3 Explaining the domain transformation

The mathematical form of FT does not demonstrate directly its capability of domain transformation. In this section a demonstration of how the complex exponential "picks out" the components of f(x) at each frequency, s, to yield the spectrum, F(s).

The complex exponential factor *exp*(-i2πsx) describes a vector that rotates with a frequency, s, (a phasor) as a function of the serial parameter, x (e.g. time or space). Thus,

the product can be illustrated as a helicoid (or coil) (Fig. 36)[254]. The sign of the phasor exponent determines the sence of rotation, clockwise or counterclockwise. The helicoid can be resolved into real and imaginary components. The FT, which is the integral of this function at the given value of s, has a real component which is the sum of these real projections, and a corresponding summed imaginary component.

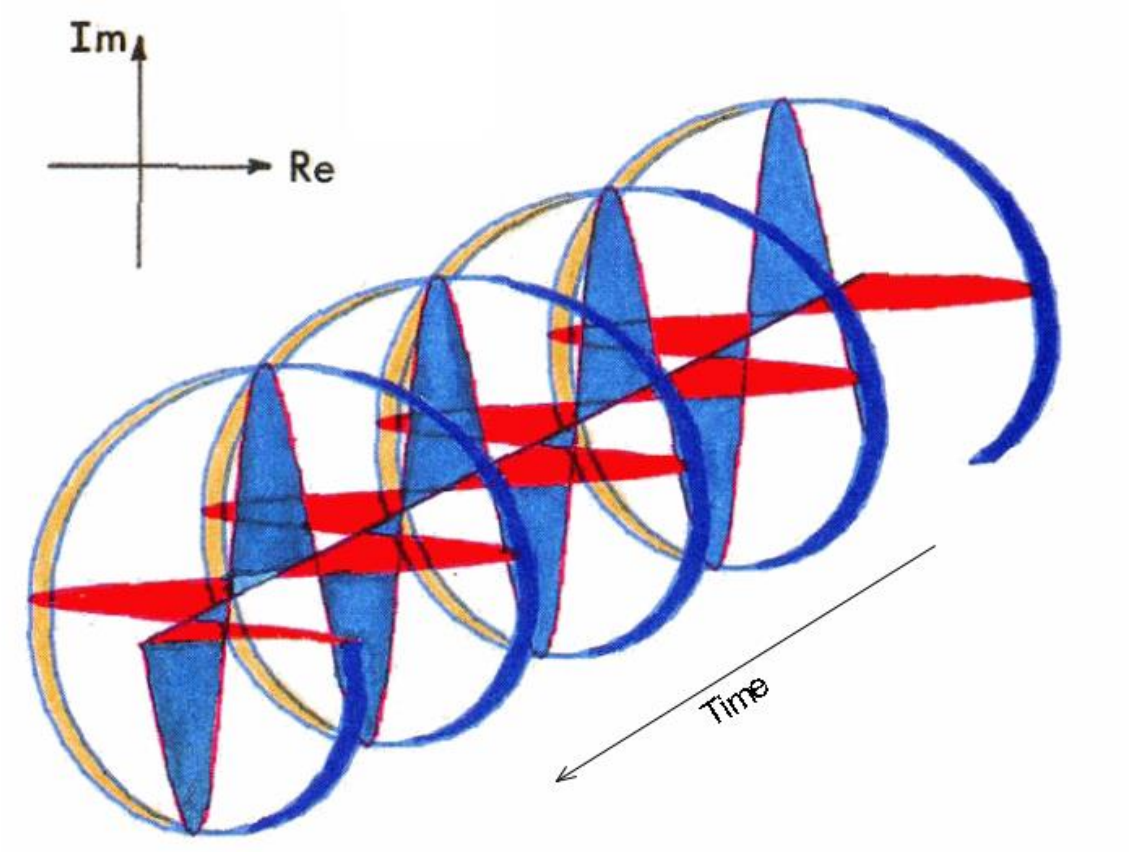


Figure 36. Resolution of the components of the helicoid into the real and imaginary planes[254] (taken from ref. 254).

In the context of FT analysis, consider a sine wave (coil) with a single, fixed frequency. The wave is coiled with a phasor of same frequency but undergoes a sign change that coincides with the reversal of the real part of the wave (the coil is coiled by the phasor) (Fig. 37a). This ends up with the phasor cancels out any contribution on the

real plane, the resulting transform yields a contribution solely to the imaginary component, with the real component being zero (Fig. 37b,c). Thus, a phasor with the same frequency is able to influence the coil only in the way of amplitude differentiation (nullification or addition)[256].

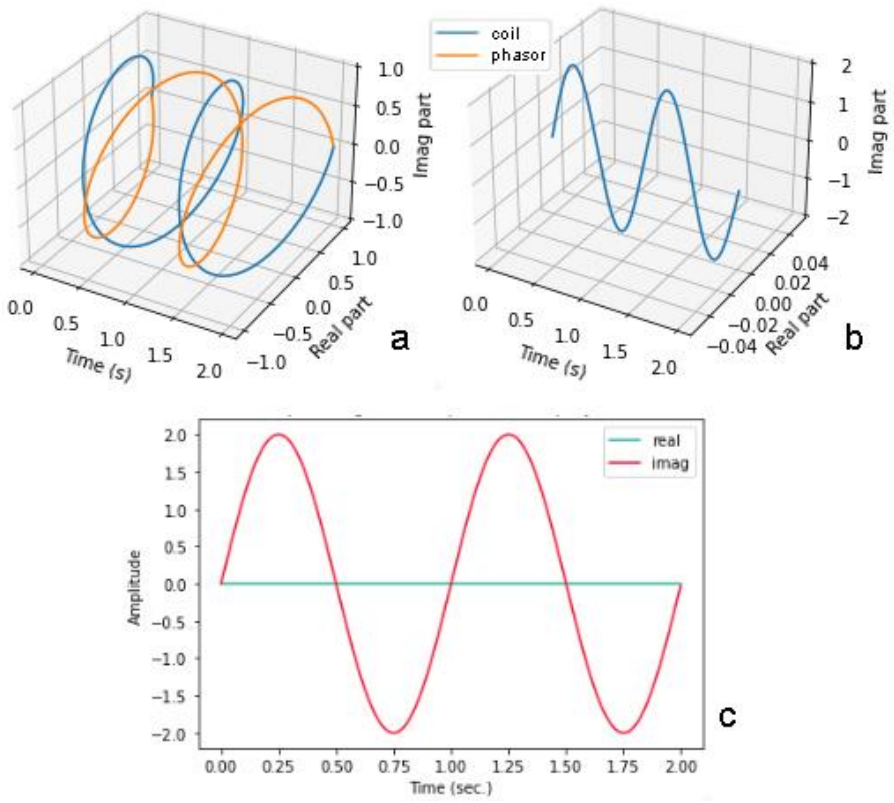


Figure 37. (a) The sine wave (coil) spread over time with the imaginary and real components follow a periodic behavior, plus the contribution of the phasor. The resulted wave after the addition of the phasor (b) in 3D and (c) in 2D.

Conversely, if the wave and the phasor are not perfectly aligned in frequency, the transform exhibits a coiling behavior (the coil then, coils about itself) (Fig. 38)[254]. Consequently, the Fourier Transform is highly effective in isolating the single frequency of the sine wave, which appears as a distinct contribution to the overall frequency spectrum

of the signal. To simplify this, the FT extracts or subtracts every single phasor that changes the original form of the coil. This process underlines the FT's ability to accurately decompose periodic functions into their constituent frequencies.

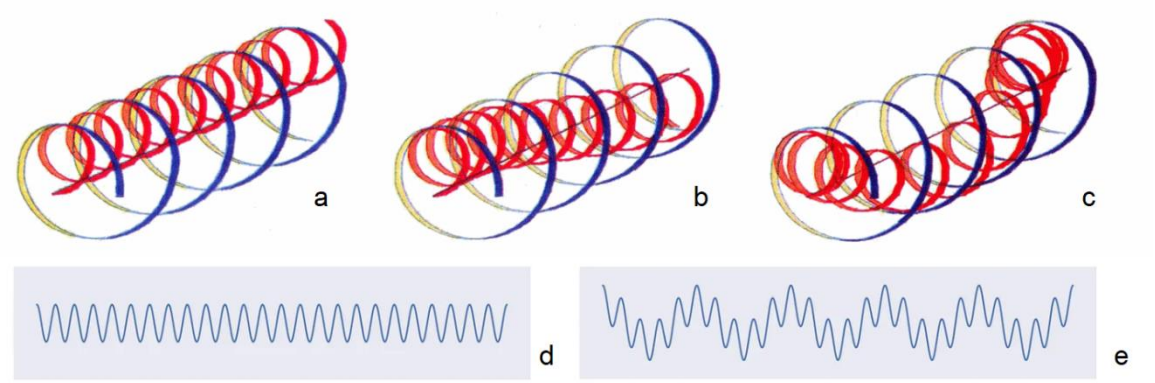


Figure 38. Coiling of the coil. The blue-yellow coil represents the original wave, while the red coil is the resulted wave after the contribution of the phasor. (a) The phasor has the same frequency with the wave (only amplitude change is observed). (b) and (c) The phasor has different frequency than the wave and different frequencies in each case[254]. (d) 2D representation (imaginary or real plane) of the resulted wave after the contribution of a same frequency phasor and (e) different frequency phasor.

In the case of an aperioodic function then, the coiling process will still produce a similar cancellation of all frequency contribution except that one corresponding to the present frequency of the helicoid (initial coil). Hence, by choosing a succession of different frequencies and amplitudes for the coiling, results to yield F(s), the spectrum of f(x).

5.4 The versatility of FT

In the study of molecular species, chemists view these entities as dynamic, exhibiting behaviors such as rotation, vibration, and flexing[257,258]. These properties are typically explored through spectroscopic techniques, where molecular systems are excited, and the emitted or transmitted electromagnetic radiation is analyzed. The radiation comprises contributions from numerous oscillators within the system, resulting in a complex time-dependent pattern. Since fluctuations in the UV-visible region can reach frequencies as high as 10^{13} Hz, no detector can accurately track such rapid changes. To manage this, the time-dependent radiation pattern is usually disregarded, and the radiation is separated into its frequency components via a monochromator (such as a prism or grating), producing an amplitude spectrum[259].

However, this approach is inherently inefficient. Single-detector systems discard much of the incoming energy when frequency bands are selected sequentially, while multi-detector systems can be costly—though alternatives like photographic films offer high sensitivity and resolution but introduce processing delays. Fourier Transform techniques provide a solution to these limitations by simultaneously capturing all frequency data and using the FT to resolve the serial-domain data into its frequency spectrum. This approach enhances both sensitivity and speed by eliminating the need for spectral separation. The versatility of FT methods extends beyond spectroscopy, offering significant advantages across various fields.

In a conventional scanning spectrometer, the light source illuminates the sample, and the transmitted radiation is dispersed and detected one sample at a time through a scanning mechanism. However, during the majority of the time, the information at different dispersion angles is discarded, leading to inefficiency. Although a single, broadband detector or multiple detectors covering the entire spectrum can resolve this issue, there are limitations that may render such a setup impractical. From an energy and information perspective, using many detectors is beneficial, but it comes with the

challenge of cost and size. If only a few spectral lines are important, the problem is minimal, but if many spectral lines need to be analyzed, then the detectors must be small and inexpensive. The accompanying table shows the minimum number of detection channels required to achieve standard resolution in various spectroscopy techniques (Table 3)[259]. To address this, the goal is to replicate the advantages of a multichannel system using only one detector. Achieving this requires removing the dispersive element, which results in a scrambled signal due to interference. This scrambled signal can then be decoded using FT methods, providing a multiplex advantage without the need for multiple detectors.

Table 3. Minimum number of channels required for various types of multidetector spectrometers[259] (taken from ref. 259).

Type of spectroscopy	Largest usual frequency (Hz)	Typical spectral frequency (Hz) range	Width of one line (Hz)	Approximate minimum number of channels
Mössbauer	6×10^{18}	10^{8}	10^{7}	10
ESCA	3.5×10^{17}	10^{17}	10^{14}	1,000
Photoelectron	5×10^{15}	3×10^{15}	10^{12}	3,000
Electronic	1.5×10^{15}	1.2×10^{15}	10^{9}	1,250,000
Vibrational	2×10^{14}	1.5×10^{14}	3×10^{9}	50,000
Rotational	4×10^{10}	3×10^{10}	10^{5}	300,000
^{13}C NMR	8×10^{7}	2×10^{4}	4×10^{-1}	50,000
ICR	2×10^{6}	2×10^{6}	10^{2}	20,000

5.4.1 FT spectroscopy

Over the past year, significant advancements in molecular structure analysis have been achieved due to the 10- to 100-fold increase in the sensitivity of infrared (IR) and nuclear magnetic resonance (NMR) spectrometers[217]. This progress is largely due to the development of FT spectroscopy, also referred to as interferometry or time-domain spectroscopy.

The two most common spectroscopic techniques that are done in an FT mode are IR and NMR spectroscopy[260,261]. Infrared spectroscopy uses the two-beam Michelson interferometer where all the beams enter and leave the interferometer in parallel streams. Initially, the beam is generated by starting with a broadband light source—containing the full spectrum of wavelengths to be measured. As the mirror in interferometer moves, each wavelength of light in the beam is periodically blocked, transmitted, blocked, transmitted, etc. by the interferometer, due to wave interference. This begins from the reference position, where the path lengths of both beams are equal, and all frequencies are in Phase simultaneously. Different wavelengths are modulated at different rates, so that at each moment or mirror position the beam coming out of the interferometer has a different spectrum[259]. The raw spectrum collected by the detector is called "interferogram" (Fig. 39a). The Fourier transform converts one domain (in this case displacement of the mirror in cm) into its inverse domain (wavenumbers in cm^{-1}) (Fig. 39b).

FTIR spectroscopy has several advantages over conventional IR spectrophotometers. It allows rapid acquisition of spectra, uses no slits (maximizing light throughput), and improves sensitivity, especially for low-concentration samples. Multiple scans can be averaged to enhance the signal-to-noise ratio, although gains diminish with increasing scans. Additionally, FTIR instruments offer superior resolution due to the highly reproducible movement of the mirrors, enabling them to distinguish nearby spectral

peaks more effectively than monochromator-based systems. This makes FTIR highly suitable for sensitive and high-resolution measurements.

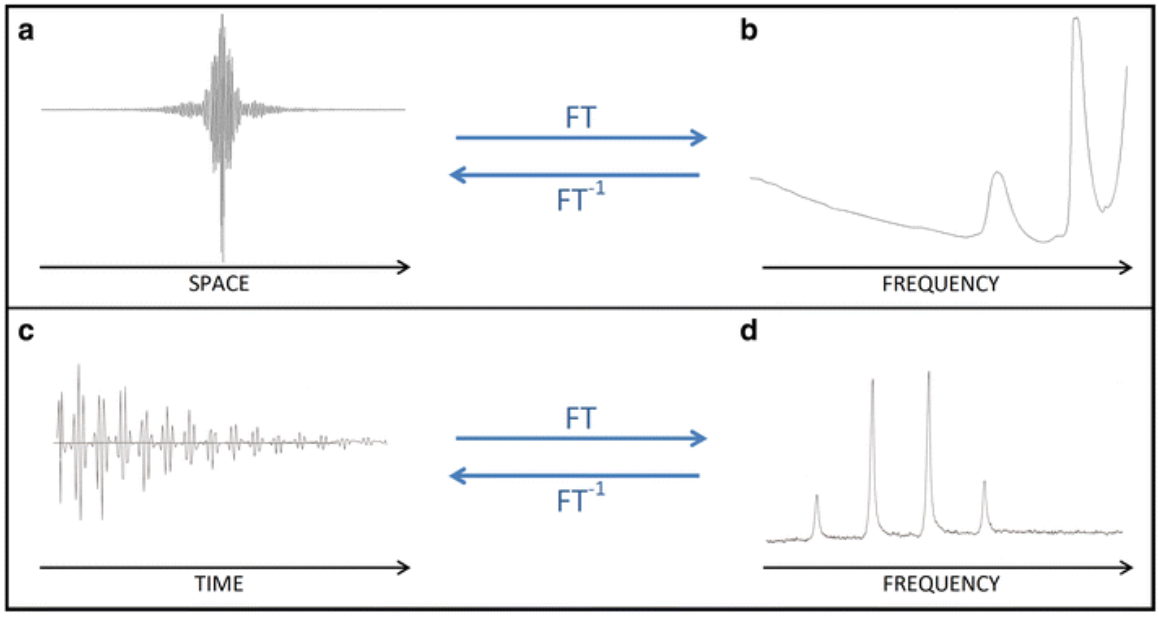


Figure 39. (a) Near infrared interferogram and (b) corresponding spectrum; (c) NMR free induction decay (FID) signal and (d) corresponding spectrum[262] (taken from ref. 262).

The key advantage of FT spectroscopy is its speed compared to conventional frequency domain methods. For example, in ^{13}C NMR, where a high-resolution spectrum would take 5000 seconds using conventional methods, FT can achieve the same result in just 1 second, with superior spectral quality. FT methods also improve the signal-to-noise ratio (S/N) through coherent signal addition, or "time averaging", which would be impractically time-consuming in conventional spectroscopy. For instance, achieving the same S/N in a ^{13}C NMR spectrum using conventional methods could take 60 days, while FT spectroscopy can do it in just 15 minutes[259].

In NMR spectroscopy, each distinct set of hydrogens in a molecule resonates at a specific frequency in a magnetic field, much like individual chimes have distinct sounds. A

traditional continuous-wave (CW) NMR spectrometer identifies each frequency by irradiating nuclei one at a time, a slow process that can take minutes for a full spectrum. In contrast, Fourier transform NMR excites all proton nuclei simultaneously using a short, strong radiofrequency pulse, creating a complex signal as the nuclei relax back to equilibrium. This complex signal (Fig. 39c), called a free induction decay (FID), is collected and analyzed using a Fourier transform to convert the time-domain data into the familiar frequency-domain spectrum (Fig. 39d). This process allows a complete spectrum to be acquired in just a few seconds. By averaging multiple FID signals from repeated pulses, FT-NMR enhances the signal strength, enabling the detection of spectra from low-abundance isotopes like ^{13}C.

5.4.2 FT advantages

One of the traditional applications of Fourier Transform (FT) is in X-ray crystallography, where it has been used since the early days of crystal structure analysis. FT helps convert the diffraction patterns obtained from X-ray scattering into real-space electron density maps, enabling the determination of the atomic arrangement within crystals. This method remains fundamental in structural biology and materials science for analyzing crystal structures. The Fellgett and Jacquinot advantages are key benefits of using Fourier Transform (FT) spectrometers[217,259]. The Fellgett advantage (also known as the multiplex advantage) refers to the ability of FT spectrometers to measure all wavelengths simultaneously, improving signal-to-noise ratio (SNR) by averaging the signal across multiple measurements, unlike dispersive instruments that measure one wavelength at a time. This advantage occurs when the noise strength is constant, independent of the signal strength; such noise originates in the detector. Fortunately, the multiplex advantage does operate for the weak sources characteristic of infrared and NMR spectroscopies. The Jacquinot advantage (or throughput advantage) arises from the fact that FT spectrometers use fewer optical elements, like slits, allowing more light to

reach the detector. This higher throughput increases the overall sensitivity, especially useful in low-light conditions or for weak signals such as in stellar spectroscopy or ^{13}C NMR. Together, these advantages make FT spectrometers faster, more sensitive, and better suited for detecting weak signals.

5.4.3 Applications in chemistry

Fourier Transforms have been successfully applied across various fields, particularly where pulsed signals generate interference (or "beat") responses. Specific examples include spectroscopic techniques like ion cyclotron resonance (a form of mass spectrometry), orbitrap mass analyzer, nuclear quadrupole resonance, dielectric and microwave responses, electron spin resonance, and muon spin rotation[263–269]. Beyond chemical applications, FTs are widely used in physics (e.g., diffusion studies), electrical engineering (e.g., antenna analysis), statistics, image processing and enhancement[270–273], demonstrating their broad utility across numerous disciplines.

6. Multivariate analysis

"We are drowning in information and starved for knowledge"

Tom Peters, Thriving on Chaos[274]

In today's data-rich world, businesses and researchers alike face the challenge of transforming vast amounts of information into valuable knowledge. With data collection and storage capabilities advancing rapidly, organizations are amassing extensive information in data warehouses, making it essential to "mine" this data effectively for strategic decision-making[275]. Simple statistics often fall short for these tasks, and sophisticated multivariate methods are now crucial in navigating and extracting insights from complex data. Advances in computing power, along with user-friendly software, have further simplified access to these powerful techniques, allowing even non-specialists to analyze intricate data efficiently.

Historically, multivariate analysis found its footing in the behavioral and biological sciences, but its utility has expanded across fields such as business, education, engineering, and beyond. Multivariate methods allow researchers and practitioners to make use of multiple measurements per unit or observation, yielding richer, multidimensional insights. This interdisciplinary relevance, coupled with modern computing, has cemented multivariate analysis as an indispensable tool for both academic and applied research across sectors, fostering knowledge-driven innovation and improvement in decision-making[276].

6.1 Measurement scales

In multivariate analysis, accurate measurement is essential for identifying variations in variables and selecting suitable analytical methods. Variables can be classified as either

non-metric (qualitative) or **metric** (quantitative), impacting how they can be analyzed[276]. Non-metric data, such as nominal and ordinal scales, describe differences by type or order without indicating actual amounts; they allow categorization and ranking but restrict mathematical operations. For example, gender is a nominal scale that simply categorizes, while satisfaction levels might use an ordinal scale, ranking items without specifying the extent of differences. Metric scales, on the other hand, such as interval and ratio scales, reflect measurable amounts and support most mathematical operations. Interval scales (like temperature) lack a true zero, while ratio scales (like weight) have an absolute zero, enabling comparisons in terms of multiples. These distinctions guide the researcher in applying the correct multivariate techniques, ensuring that non-metric or metric data are used appropriately to obtain meaningful results.

6.2 A Classification of multivariate techniques

Multivariate techniques are classified based on three key research considerations: whether variables can be defined as dependent or independent, the number of dependent variables in the analysis, and the type of measurement for both dependent and independent variables[276].

If variables can be classified as dependent and independent, a **dependence technique** is used, where the goal is to predict or explain dependent variables using independent ones. If this classification isn't possible, an **interdependence technique** is applied, which involves the simultaneous analysis of all variables without distinguishing between dependent and independent roles[276]. The difference between dependence and interdependence is illustrated in the contrast between the two situations shown in Figure 40—one in which a train car is completely dependent on the engine to pull it, and the other in which two friends provide mutual support of a helpful nature that is optional and opportunistic rather than strictly required[277].

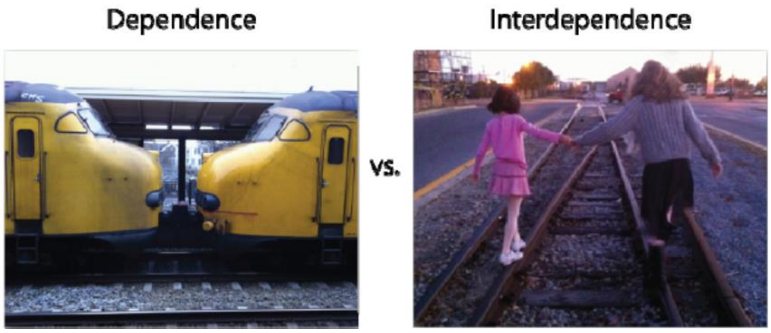


Figure 40. Dependence *vs.* Interdependence[277] (taken from ref. 277).

The selection of an appropriate multivariate technique thus depends on these research-based judgments. Figure 41 assist in becoming familiar with the specific multivariate techniques, helping to choose the most appropriate technique in each scenario.

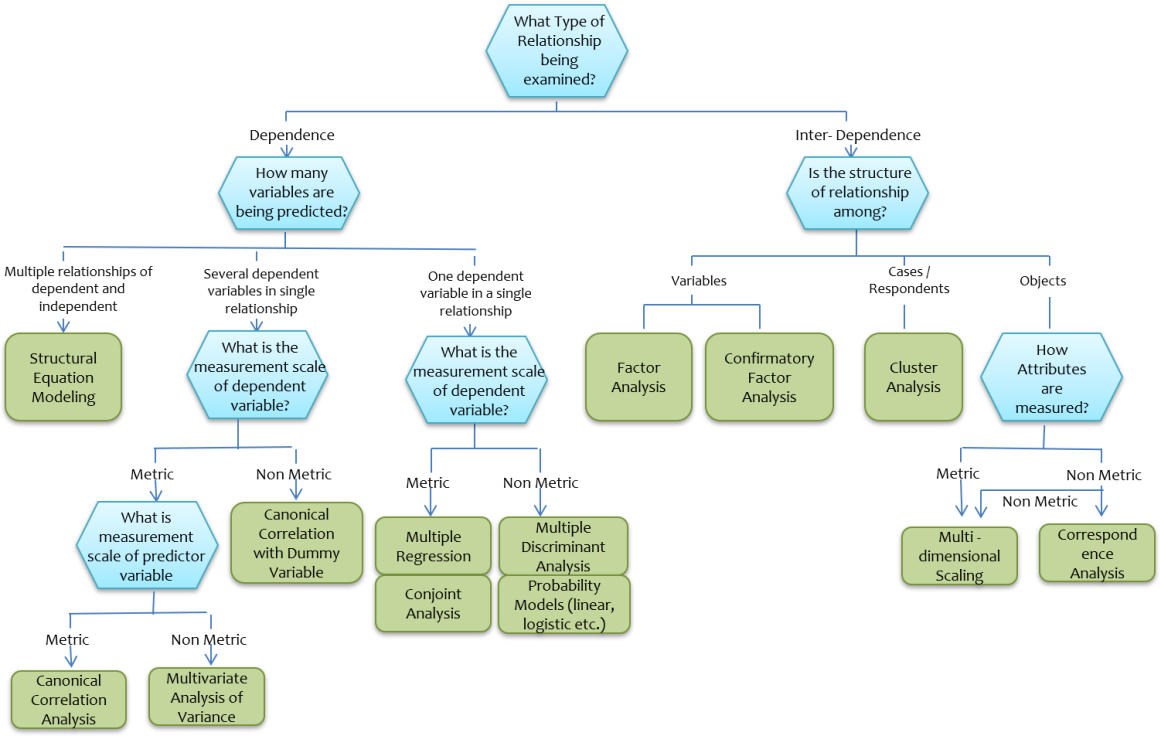


Figure 41. Selecting a multivariate technique[276] (taken from ref. 276).

6.2.1 Supervised learning - Dependence techniques

Supervised learning methods involve training a model on a dataset that includes both input features (predictor variables) and a known target variable (response)[278]. The goal of supervised learning is to predict or classify the target variable based on the input features, making it highly suitable for dependence techniques. Dependence techniques, which focus on finding relationships between dependent and independent variables, are often used in supervised learning. For instance, regression methods (such as multiple regression for metric dependent variables or logistic regression for nonmetric dependent variables) can be applied to predict an outcome based on multiple predictors[276]. Supervised learning techniques utilize dependence relationships, aiming to quantify and model the influence of independent variables on the dependent variable.

6.2.2 Usupervised learning - Interdependence techniques

Unsupervised learning, on the other hand, involves analyzing data without a predefined target variable, seeking patterns or clusters within the data[278]. This aligns well with interdependence techniques, where the focus is on analyzing relationships among all variables without classifying them as dependent or independent[276]. Techniques like factor analysis or cluster analysis are examples of interdependence methods used in unsupervised learning, as they explore the underlying structure of the data. For example, cluster analysis can group observations with similar characteristics, while factor analysis uncovers latent factors driving correlations among variables[276]. Thus, while supervised learning typically relies on dependence relationships to make predictions, unsupervised learning emphasizes interdependence to uncover patterns and groupings within the dataset[278].

6.3 Types of multivariate techniques

Multivariate analysis is an *ever-expanding* set of techniques for data analysis that encompasses a wide range of possible research situations as evidenced by Figure 41. The more established include the following[276]:

Interdependence techniques:

- Exploratory Factor Analysis (EFA): Includes principal components and common factor analyses to condense multiple variables into a smaller set of factors that represent underlying dimensions, such as combining customer satisfaction indicators into generic factors like "food quality" and "service quality."
- Cluster Analysis: Groups entities into mutually exclusive clusters based on similarities without predefined groups, e.g., categorizing restaurant customers by motivations like price or quality.

Dependence techniques:

- Multiple Regression: Predicts a metric dependent variable using multiple metric independent variables, e.g., predicting dining expenditures from income and family size.
- Multivariate Analysis of Variance (MANOVA) and Covariance (MANCOVA): Examines relationships between categorical independent variables and multiple dependent variables, adjusting for covariates if necessary.
- Multiple Discriminant Analysis (MDA): Differentiates between predefined groups based on independent variables, such as distinguishing between national-brand buyers from private-label buyers.

- Logistic Regression: Similar to multiple regression, but for binary dependent variables, useful for classifications like determining business success based on financial data.
- Structural Equation Modeling (SEM): Models complex relationships among dependent and independent variables with a focus on measurement and structural models, such as analyzing factors affecting worker satisfaction.
- Partial Least Squares SEM (PLS-SEM): Emphasizes prediction using total variance, suitable when the research goal is less about confirmation and more about prediction.
- Canonical Correlation: Extends multiple regression by correlating multiple dependent and independent variables simultaneously.
- Conjoint Analysis: Used primarily in product design to assess consumer preferences and the importance of product attributes.
- Perceptual Mapping: Maps consumer preferences or similarities between brands, aiding in competitive analysis.
- Correspondence Analysis: Suitable for nonmetric data, it maps associations in contingency tables to create perceptual maps, showing brand preferences linked to demographic variables.

Each technique is defined with its purpose and example applications, providing a toolkit for analyzing complex, multivariable datasets in research.

6.4 Principal Component Analysis

Principal Components Analysis (PCA), examines the total variance of a dataset and derives factors (components) that may include small amounts of unique and error variance, preventing any significant distortion of the factor structure[276].

PCA is particularly applied when:

- The main objective is data reduction, aiming to capture the **majority of the variance with the fewest factors possible** from the original variables.
- There is prior knowledge that specific and error variance make up a small proportion of the total variance.
- PCA serves as a preliminary step in the scale development process.

The primary objectives of Principal Component Analysis (PCA) are to reduce the dimensionality of a dataset while retaining as much variance as possible, simplify the data structure, and reveal the underlying patterns[279]. By transforming the original variables into a smaller set of uncorrelated principal components, PCA helps identify directions in the data with the most significant variance[276]. These components allow easier visualization, interpretation, and analysis of complex data, enabling efficient information summarization and facilitating subsequent analyses by reducing the complexity of the dataset.

6.4.1 Step-by-step explanation of PCA

When performing PCA for dimensionality reduction, the goal is typically to reduce a high-dimensional dataset to a lower-dimensional one, capturing most of the original variability in fewer dimensions[279,280]. In cases with only two variables, applying PCA to create two principal components might seem redundant because we're not actually reducing dimensionality. However, the following example can still be useful to illustrate how PCA works: the two new components represent the directions of maximum variance in the data, with the first principal component capturing the most variance, and the second being orthogonal to it. Even though we don't reduce the number of dimensions, this process shows how PCA reorients data in terms of variance, which is fundamental to understanding PCA in more complex cases. Let's assume that the scatter plot of a data set is as shown below (Fig. 42a).

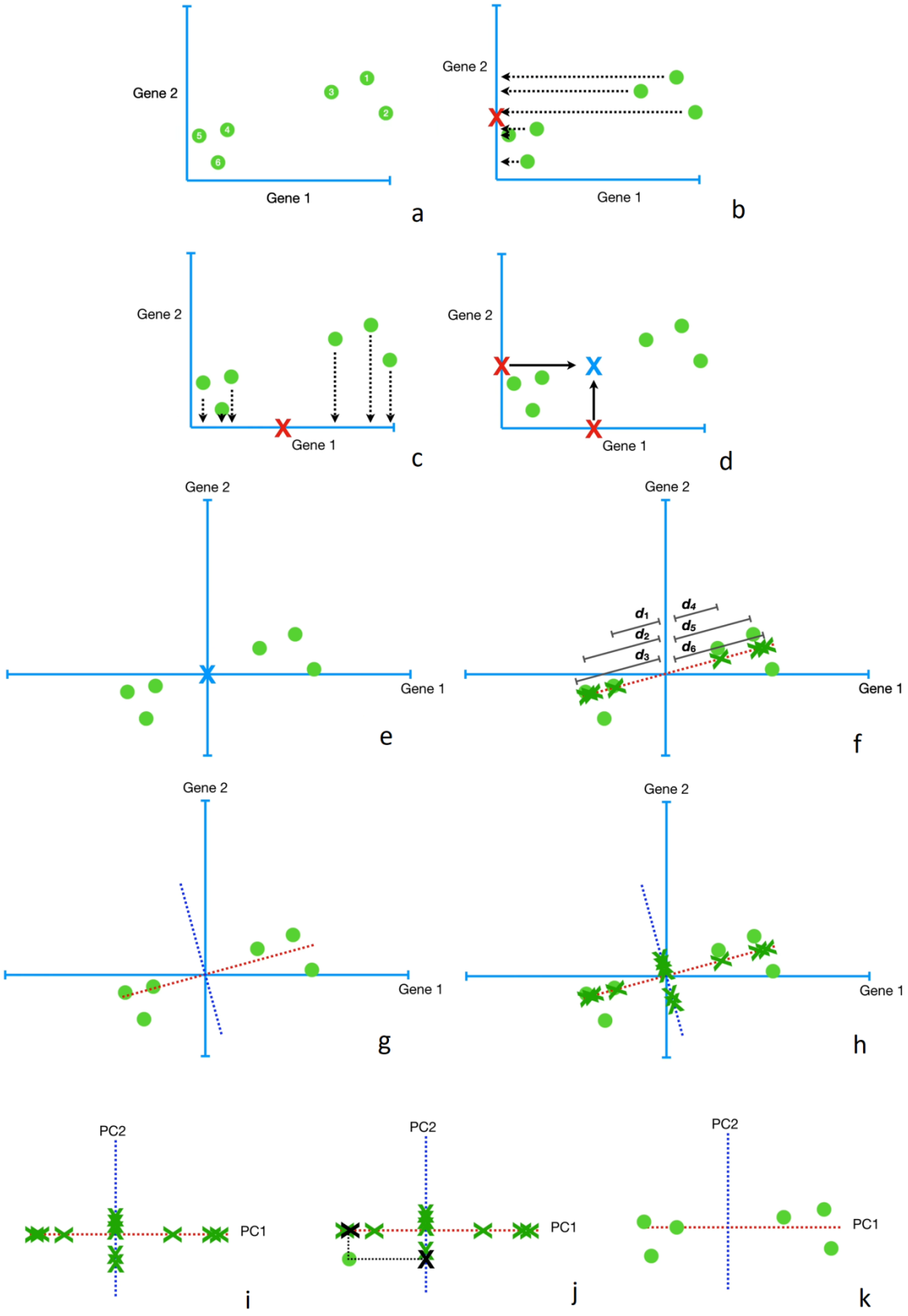

Gene 2
Gene 1
1
2
3
4
5
6
a
Gene 2
Gene 1
b
Gene 2
Gene 1
c
Gene 2
Gene 1
d
Gene 2
Gene 1
e
Gene 2
d_1
d_2
d_3
d_4
d_5
d_6
Gene 1
f
Gene 2
Gene 1
g
Gene 2
Gene 1
h
PC2
PC1
i
PC2
PC1
j
PC2
PC1
k

Figure 42. Example of a PCA application in a dataset of the measured transcription of two genes in six different mice. (a) Plotting the samples according to measured transcription. (b) Indicating the average measurement for gene 2 among samples. (c) Indicating the average measurement for gene 1 among samples. (d) Indicating the center of the data. (e) Centering the data. (f) The best fitted line due to the largest $SS_{distances}$, the construction of PC1. (g) The construction of PC2 perpendicular to PC1. (h) Projecting the data point on PC1 and PC2. (i) Plot rotation so PC1 in horizontal. (j) The use of the projected points to indicate the samples in the PCA plot. (k) The final PCA plot[281] (adjusted from ref. 281).

Step 1: Standardization

Before performing PCA, it's essential to standardize continuous variables so that they contribute equally to the analysis. PCA is sensitive to the variances of the input variables; if some variables have larger ranges than others, they will dominate the analysis and lead to skewed results. Standardization solves this by scaling the variables to a comparable range, usually by subtracting the mean and dividing by the standard deviation for each variable[282]. This ensures that all variables have similar influence on the analysis, making the PCA more balanced and effective.

When data is standardized, it is effectively shifted so that its mean is zero, which centers the data around the origin of the graph (Fig. 42e). This process involves adjusting each variable by subtracting the mean, so all values are re-centered around zero. Thus, the standardized data clusters around the origin, with each variable now having a comparable range and zero mean.

Note: Shifting the data do not change how the data points are positioned relative to each other.

Step 2: Creating the PC1

Principal components are constructed in such a manner that the first principal component (PC1) accounts for the largest possible variance in the data set. The first principal component is the line that goes through the origin and in which the projections of the points are the most spread out. Or mathematically speaking, it's the line that minimizes the distances between the data point and their projections on it, or (according to pythagoreon phenomenon) is the line that maximizes the distances between the projected points and the origin (Fig. 42f)[278,282]. The last are squared and summed and the largest resulted value ($SS_{distances}$) indicates the orientation of PC1.

Step 3: Creating PC2, PC3, etc.

Ideally, the second principal component is a unit vector that does not contain information that is already contained in the first component. Or in geometric terms the second component belongs to the subspace orthogonal to PC1 but other than that, it should maximize the same quantity as before and following reasoning similar to PC1 (Fig. 42g)[278,282]. The third component is constructed similarly, and so on.

Step 4: The explained variation

In PCA, eigenvectors and eigenvalues are key concepts used to determine the principal components, which are the directions that capture the most variation in the data[279].

Eigenvectors represent the directions of maximum variance in the data; they are the axes along which the data varies the most, and these directions/vectors/axes become the principal components[282]. Eigenvalues, on the other hand, measure the amount of variance associated with each eigenvector[282]. Each eigenvector has a corresponding eigenvalue, and together they capture the structure of the data: the eigenvector with the highest eigenvalue indicates the direction of the most significant variance, forming the first principal component, while the next highest eigenvalue's eigenvector forms the second principal component, and so on.

The concept of explained variance ties directly to eigenvalues. Explained variance quantifies how much of the total variation in the dataset is captured by each principal component. By ranking the eigenvectors based on their eigenvalues (from highest to lowest), we order the principal components by their importance. This ordering allows us to focus on the components that capture the most variance, simplifying the data analysis and interpretation process by prioritizing the most informative components[279].

Step 5: Recast the data along the principal components axes

As described previously, the first PC has the highest importance explaining the total variance better. Typically, only few of the very first components are considered, achieving dimension reduction on the dataset without losing major information[280]. To draw the final PCA plot of two dimensions (taking into consideration only the PC1 and PC2) the principal component vectors are rotated in a way that PC1 is horizontal and PC2 is vertical (Fig. 41i). Now the lines named principal components, are the axes of the plot[282]. The projected points on the components axes are used to find where the data go in the PCA plot (Fig. 42j).

Note: If the components account for a substantial amount of variation, then just using the first PCs would not create a very accurate representation of the data. However, even in this case a PCA plot of two dimensions (using the first two PCs) can be used to identify clusters of data.

PCA is a complex analytical technique that relies on sophisticated mathematical computations, making it challenging to perform manually[280]. Although the foundational principles and steps of PCA were carefully explained above, the actual calculations are typically executed by a computer due to their complexity and the large volume of data often involved[279]. This computational assistance enables the precise extraction of principal components and ensures efficient handling of the underlying matrix operations required in PCA.

6.4.2 Stopping rules: criteria for the number of components to extract

When deciding on the number of components to extract in factor analysis, researchers aim to capture the most variance in the data through a set of linear combinations of variables. The process extracts components in sequence, with each new component accounting for remaining unexplained variance, until all variance is covered[276]. The goal is to retain a manageable number of components that adequately summarize the variance across all variables. Several methods are used to determine how many factors to retain[276,282]:

1. A Priori Criterion: The number of factors is predetermined based on theory or prior research.
2. Latent Root Criterion (Kaiser Rule): Only components with eigenvalues greater than 1 are retained, meaning each component should explain as much variance as an individual variable. This method is most effective with 20-50 variables and is often a starting point, complemented by other criteria.
3. Percentage of Variance Criterion: Components are retained until they account for a specific percentage of total variance (typically 95% in natural sciences, 60% or less in social sciences).
4. Scree Test: Factors are retained until an “elbow” in a plot of eigenvalues, where the curve levels off, indicating factors with less common variance. This approach is subjective but often retains one or two more components than the Kaiser Rule.
5. Parallel Analysis: Simulated datasets are generated to compare eigenvalues, retaining components with eigenvalues above those in the random data. This is often more precise than the Kaiser Rule.

6. Heterogeneity of Respondents: When the sample includes subgroups with unique variance patterns, extra components may be retained to capture differences between groups.

Researchers generally use a combination of these criteria to ensure both parsimony and interpretability in their component (factor) solution, balancing the number of components with the goal of accurately representing the data structure[280]. Multiple solutions are typically examined to refine the final component structure.

6.4.3 Interpreting the factors – Evaluating the significance of factor loadings

In multivariate analysis, effective interpretation relies on a strong conceptual foundation, which can stem from previous research, theory, or accepted principles. Researchers must make subjective judgments on factors to extract, variable groupings, and factor solution appropriateness. Interpretation requires assessing factor-loading significance, and factor interpretation. Each process involves essential considerations, guiding the researcher to a final, conceptually sound factor structure. In interpreting factors, researchers must decide which factor loadings merit consideration and attention[276,282]. This discussion addresses issues of practical and statistical significance, as well as the number of variables, which collectively influence the interpretation of factor loadings.

Practical significance

The first guideline focuses on practical rather than statistical significance, requiring an initial examination of the factor matrix with respect to factor loadings. A factor loading represents the correlation between a variable and a factor, with the squared loading indicating the proportion of variance in the variable that the factor explains. For example, a factor loading of 0.30 accounts for approximately 10% of the variable's variance, while a

0.50 loading explains about 25%, and a loading above 0.70 accounts for 50%[276]. Therefore, larger absolute values of factor loadings generally signify greater importance in interpreting the factor matrix. In terms of practical significance, loadings can be assessed as follows[276]:

- Loadings below 0.10 can be treated as equivalent to zero, suggesting no meaningful contribution.
- Loadings from 0.30 to 0.40 meet the minimal threshold for interpreting structure.
- Loadings of 0.50 or higher are considered practically significant.
- Loadings exceeding 0.70 indicate a well-defined structure and are desirable in factor analysis.

Researchers should also recognize that extremely high loadings (0.90 or above) are uncommon and that practical significance remains an important criterion, especially for sample sizes of 100 or more where the focus is on practical interpretation rather than statistical testing.

Statistical significance

Factor loadings, which reflect the correlation between a variable and its corresponding factor, can also be evaluated statistically. While the statistical significance of correlation coefficients could theoretically apply, research has shown that factor loadings tend to exhibit larger standard errors than typical correlations[276]. Consequently, factor loadings should be assessed with stricter thresholds. To achieve a statistical power of 80% and a significance level of 0.05, researchers may refer to sample size requirements for specific loading values, as indicated in Table 4. For example, in a sample of 100, a loading of 0.55 or higher is significant, whereas a sample size of 50 requires a minimum loading of 0.75 for significance. Notably, these thresholds are conservative relative to the previously mentioned practical guidelines and the statistical standards associated with

correlation coefficients, making them useful starting points. Lower loadings can still be included in the interpretation if other contextual factors support their significance[276].

Table 4. Guidelines for identifying significant factor loadings based on sample size[276] (taken from ref. 276).

Factor Loading	Sample size needed for significance [a]
0.30	350
0.35	250
0.40	200
0.45	150
0.50	120
0.55	100
0.60	85
0.65	70
0.70	60
0.75	50

[a] Significance is based on a 0.05 significance level (α), a power level of 80%, and standard errors assumed to be twice those of conventional correlation coefficients.

Adjustments based on the number of variables

A limitation of both practical and statistical significance guidelines is the lack of consideration for the number of variables and the specific factor being analyzed. Research suggests that, as the analysis moves from the first factor to subsequent ones, the

threshold for significant loadings should increase to account for unique and error variances that emerge in later factors[276]. Similarly, as the number of variables increases, the acceptable level for defining a loading as significant decreases, making adjustments based on variable count increasingly important for later-extracted factors.

6.5 Partial Least Squares method

Partial least squares (PLS) technique was developed by Herman Wold in the 1970s by extending the multiple linear regression model[283–285]. It takes a latent variable approach to model the covariance structures in two spaces (i.e., the X and Y spaces) so that both variables X and Y are projected to a new space, which is called projection to latent (hidden) structures. Thus, PLS is alternatively called Projection to Latent Structures[286].

Note: A number of N observations (e.g samples) described by J dependent variables (e.g. concentration of each sample) are stored in a N×J matrix denoted **Y**, the values of K predictors (e.g. absorbance measurement of each sample) collected on these N observations are collected in the N×K matrix **X** (Fig. 43).

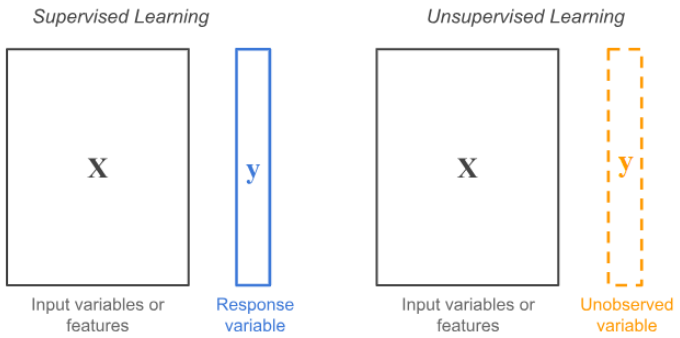


Figure 43. Inputs and outputs in supervised and unsupervised learning[278] (taken from ref. 278).

Unsupervised methods (like PCA) work solely with X. There is no Y matrix because these methods do not require predefined labels or response variables (Fig. 43). Instead, they focus on finding patterns, clusters, or the underlying structure within the X matrix alone. In contrast, in supervised methods, such as PLS, the data is divided into two matrices: X and Y. Here, X represents predictor variables, which are always available, while Y contains response variables that are aimed to be predicted or understood[278]. The supervised approach relies on learning a direct relationship from X to Y, enabling predictions for Y using new X data[287].

6.5.1 Advantages of the PLS method

Projection to Latent Structures (PLS) and Principal Component Regression (PCR) are both methods used for modeling relationships between predictor variables (X) and response variables (Y), with PLS building on PCR's foundation and offering enhanced capabilities[280]. Before going deeper, some basic principles about PCR and MLR are explained in the following lines.

Multiple Linear Regression (MLR) is a traditional linear regression method that directly models Y as a function of X. MLR, however, requires that the number of observations be greater than the number of predictor variables, and it is sensitive to noise and multicollinearity in X, which can skew results[287].

Principal Component Regression (PCR) improves on MLR by first applying PCA to X, transforming the data into orthogonal (uncorrelated) components (Fig. 44). This removes the need for variable selection and reduces noise, as these PCs are less noisy than the original X data. PCR can work with missing data and relaxes MLR's requirement due to the dimension reduction (the number of PCA components are less than initial variables)[287].

However, PCR may require extracting several components, some of which may not correlate strongly with Y, which can increase model complexity unnecessarily.

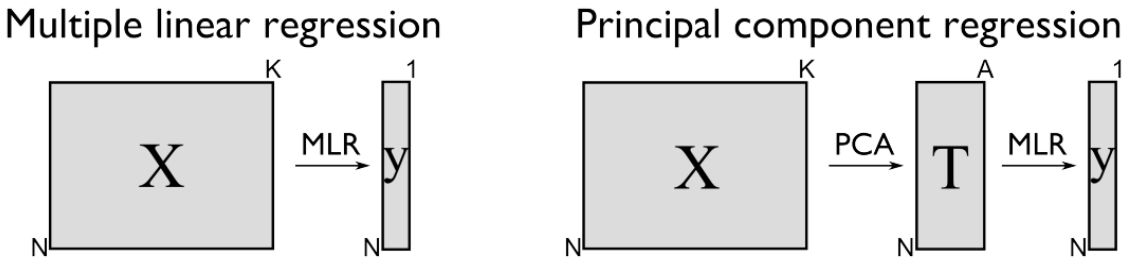


Figure 44. Schematic representation of PCR individual steps[287] (taken from ref. 287).

PLS further extends PCR by using both X and Y data in a single modeling step, making it more efficient and compact. Unlike PCR, which requires a separate model for each Y variable[280], PLS can handle multiple correlated Y variables in a single model, saving time and improving interpretability[287]. PLS also directly assumes there is error in both X and Y, providing a more realistic model of data variability. Key advantages of PLS over PCR include:

- Efficiency: PLS avoids calculating redundant scores that do not contribute to the prediction of Y, resulting in fewer components and a more streamlined model.
- Unified Model Structure: PLS maximizes the covariance between X and Y, meaning that the components are optimized for prediction in a single step, unlike PCR's two-step process.
- Holistic Use of System Variables: Since PLS uses both X and Y data simultaneously to extract components, it can reveal shared latent structures within the same system.

6.5.2 A conceptual explanation of PLS

PLS can be understood as a latent variable model with a distinct objective. Unlike PCA, which finds factors (components) to best explain variance in X, PLS seeks factors (latent variables) that simultaneously optimize three objectives[278,280,287]:

1. Best explanation of variance in the X-space,
2. Best explanation of variance in the Y-space, and
3. Maximizing the relationship between X and Y spaces.

This approach enables PLS to capture both the variability within each data block and the relationship between them, resulting in a model that effectively links X and Y for predictive or analytical purposes.

6.5.3 A geometric interpretation of PLS

In PLS, the mathematical approach builds on PCA by extending the objective to optimize variance explanation in both X and Y spaces[280]. In PCA, scores and loadings are calculated to ensure each component maximizes variance in X. PLS modifies this to simultaneously explain variance in both X and Y. Scores and loadings in PLS are calculated to capture variance in X while also explaining variance in Y[278], thereby maximizing the relationship between X and Y. This results in latent variables that reveal structural relationships across both data blocks, optimizing both predictive accuracy and interpretability.

In PLS, we can visualize the model geometrically with both X and Y data points centered and scaled to the origin, represented by corresponding points in X- and Y-space[278]. Each observation in X has a counterpart in Y, and scores are obtained by

projecting each data point onto direction vectors w_1 and c_1 (Fig. 45). These direction vectors are chosen to maximize the covariance between the X-space scores and Y-space scores (like PCA does), aligning the latent variable directions to best explain both X and Y while strengthening their relationship[287].

The second component in X-space is calculated orthogonally to the first, though it may not be strictly orthogonal in Y-space (but often nearly is). This approach ensures that each component captures distinct, meaningful variation in X that correlates with Y, enhancing predictive power.

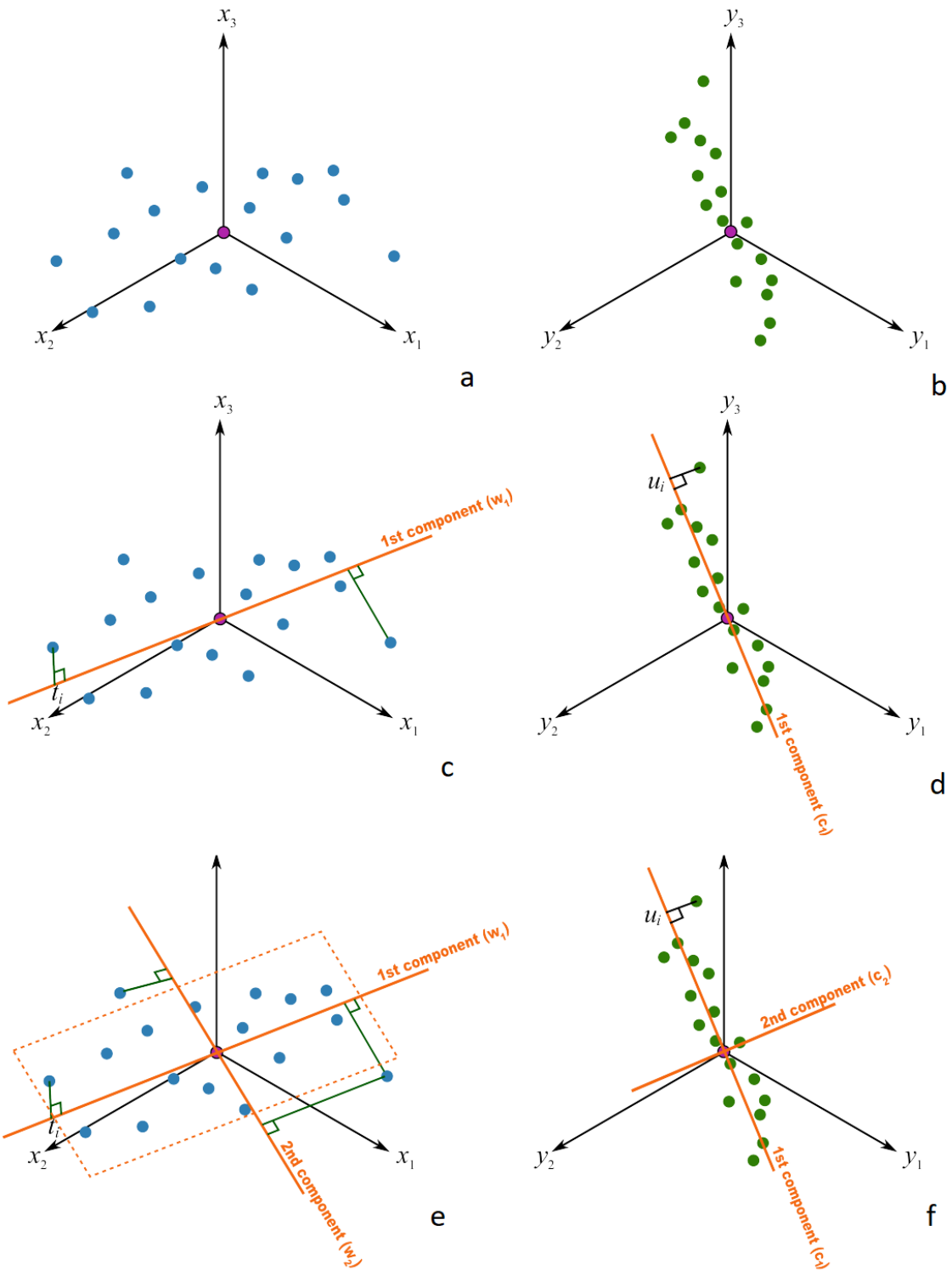


Figure 45. Interpretation of a dataset via PLS method. (a) Plotting the X matrix, (b) plotting the Y matrix. The first component explaining the variation (c) in X and (d) in Y. Construction of the second component (e) in X and (f) in Y matrix[287] (taken from ref. 287).

6.5.4 Interpreting the scores and loadings in PLS

In PLS, scores summarize the data in X and Y blocks, and have maximal covariance. Interpreting PLS scores is similar to PCA: looking for clusters, outliers, and patterns in score plots[280]. However, unlike PCA scores, which only explain variance in X, PLS scores capture variance in both X and Y while maximizing the relationship between them. This results in similar but not identical orientations between PCA and PLS scores (Fig. 46a)[287].

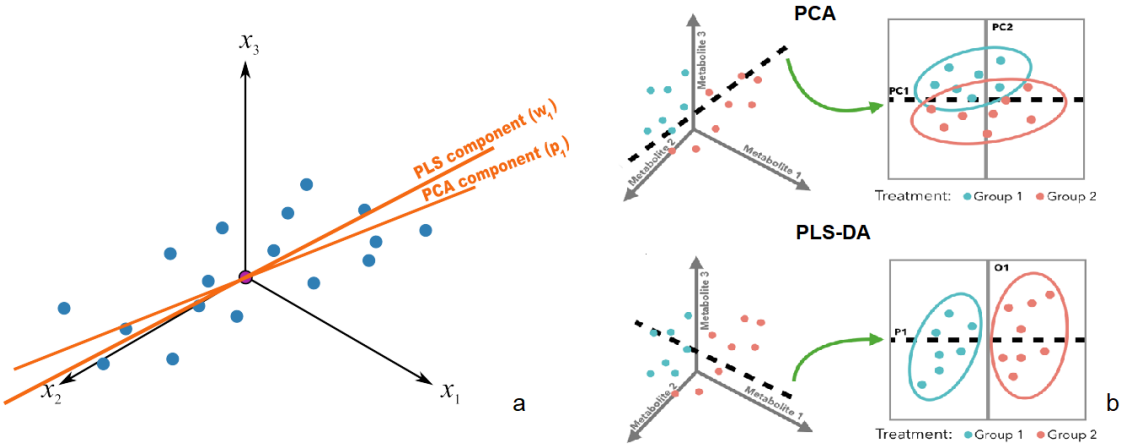


Figure 46. Examples of two datasets. (a) First components' orientations of both PCA and PLS methods applied in the same dataset and (b) recasting the data on PCA and PLS components plane.

In PLS, interpreting the loadings, also called weights[287], follows similar principles to PCA loadings. Highly correlated variables have similar weights and appear close together in loading plots. A unique feature in PLS is that we often plot the loadings for X and Y simultaneously[287]. This combined view highlights relationships not only among X variables and Y variables but also between all variables, reflecting that X and Y come from the same system.

6.5.5 Validation of model results

The process of validation is essential to avoid the pitfalls of overfitting, which occurs when model parameters are too finely tuned to the specific characteristics of the sample rather than to the general population[276]. Overfitting leads to models that perform well on the sample data but fail to generalize, meaning they perform poorly on new or independent data. Overfitting is especially common when the sample size is small, or the model includes a large number of parameters.

Split-Sample Validation

The simplest method for validation is the split-sample approach, in which the data is divided into two subsets[276]. One subset, the estimation or training sample, is used to develop the model, while the other, the holdout or validation or test sample, serves as an independent dataset to test the model's performance. Since the holdout sample does not contribute to model estimation, it provides a separate and unbiased measure of the model's validity and accuracy[288].

Cross-Validation

When a sample is too small for effective split-sample validation, cross-validation techniques are more suitable. Cross-validation divides the data into multiple smaller subsets and averages the model performance across all subsets[280,288]. Three commonly used cross-validation methods are: (a) *K-Fold Cross-Validation*: The data is divided into K subsets, with each subset being used once as the validation sample while the other K–1 subsets serve as the estimation sample[276]. This process repeats K times, each time using a different subset for validation. This technique is effective even with small datasets as it allows for a smaller validation sample. (b) *Repeated Random Sampling (Resampling):* In

this approach, multiple random samples are drawn as validation samples. The size of these samples is flexible, independent of the number of subsets, making it useful when specific sample size requirements are needed[276]. (c) *Leave-One-Out Cross-Validation (Jackknife):* This approach is a variation of K-fold validation where each fold contains only a single observation, meaning one observation is left out each time as the validation sample[276]. This is repeated until all observations have been used as validation samples.

In multivariate analysis, the goal is not solely to achieve a model that fits the sample but to build one that accurately reflects the population. Validation provides the means to achieve this by minimizing overfitting and confirming that the model is both statistically significant and representative of broader data. Thus, validation is critical in creating models that not only capture the specifics of the sample data but also hold predictive power for the population at large.

Determining the Number of Components with Cross-Validation

Cross-validation is a versatile tool that helps prevent overfitting and can be applied to various models, not just latent variable models. When additional components are added to a model, the model's fit improves as it explains more of the data's variance, but this also risks capturing noise instead of meaningful patterns[287]. For latent variable models like PLS, cross-validation helps determine the appropriate number of components by dividing the data into groups and calculating the residual variance in the cross-validated error matrix[278,287].

In each cross-validation fold, a PLS model is built on a subset of the data X. The cumulative residual errors across all folds yield a cross-validated error matrix, from which a measure (e.g. mean square error, MSE) is computed[278,287]. MSE eventually decreases after adding non-informative components, signaling overfitting. For model selection, practitioners examine plots of MSE to judge component relevance[280].

Although cross-validation provides guidance, there is no exact answer for the "best" number of components; it should be determined by the model's purpose, considering

both predictive accuracy and interpretability[287]. Cross-validation is particularly valuable for predictive models like PLS but may require adjustment for exploratory or process-optimization models where component relevance may vary.

6.5.6 PLS in continuous and categorical data

Partial Least Squares (PLS) is a family of models. Partial Least Squares Regression (PLS-R) and Partial Least Squares Discriminant Analysis (PLS-DA) are both based on PLS but differ in their objectives and applications[289,290]. PLSR is used for regression tasks, where the goal is to predict **continuous outcomes** by modeling relationships between predictor variables and a continuous response variable. In contrast, PLS-DA is a classification method; it adapts PLS for discriminant analysis by converting class labels (e.g., categories) into a **binary or a dummy variable** (coding 0–1) rather than a block of continuous variables, aiming to maximize the separation between predefined classes[276,280]. While PLS-R seeks to minimize prediction error for continuous data, PLS-DA focuses on maximizing class discrimination, making it useful in cases like biomarker identification or diagnostic classification.

6.6 Chemometrics

Having explored the fundamentals of multivariate analysis, the term of chemometrics — a field that applies these techniques to solve complex chemical problems — can be introduced. Chemometrics leverages multivariate tools, such as PCA and regression, to analyze chemical data, helping to identify patterns, optimize processes, and make predictive models. This approach has become essential in handling the intricate, high-dimensional data common in modern chemical research, transforming raw data into actionable chemical insights[280].

An actual definition of chemometrics is:

the chemical discipline that uses mathematical and statistical methods, (a) to design or select optimal measurement procedures and experiments, and (b) to provide maximum chemical information by analyzing chemical data[291].

Aim and Objectives

The overarching aim of this research is to develop and evaluate protocols that leverage spectroscopic techniques and chemometric analysis for effective monitoring and characterization of physiological responses in the context of physical exercise. This work seeks to advance our understanding of physiological responses across different exercise intensities through reliable, non-invasive methods, with a focus on establishing robust, reproducible protocols using saliva and dried blood spot (DBS) analysis.

The 1st objective is to create and validate protocols that characterize the distinct salivary profiles of athletes by employing ATR-FTIR spectroscopy alongside chemometric techniques. This objective involves defining and standardizing spectral markers in saliva that reliably differentiate athletes based on factors like training, fitness levels, and physiological adaptations to exercise (**Phase 1**).

Furthermore, the 2nd objective is to refine methods for distinguishing exercise intensities through the chemometric analysis of oral biofluid markers, utilizing ATR-FTIR spectroscopy to correlate biochemical and spectral variations in saliva with physical activity levels. This aspect of the study prioritizes the development of a rapid and accessible protocol for exercise intensity monitoring, potentially providing insights for both training and recovery (**Phase 2**).

Additionally, a 3rd objective is to quantitatively measure biomarkers in saliva that are associated with different exercise intensities. This Phase focuses on developing and validating a precise analytical technique to assess thiocyanate ions in saliva, enabling the evaluation of their variations in response to physical activity. The goal is to establish a reliable and practical method for monitoring exercise-induced biochemical changes, with

potential applications in athletic performance assessment and recovery management **(Phase 3)**.

The final objective is to explore protocols for monitoring biomarkers in exercise contexts *via* DBS analysis, employing NMR spectroscopy and multivariate analysis. This objective seeks to establish a minimally invasive, standardized protocol for DBS that reliably tracks biochemical changes in response to exercise, enhancing the practical utility of DBS in exercise biomarker monitoring (**Phase 4**).

Collectively, these objectives aim to deliver a rigorous, multivariate approach to biomarker protocol development in exercise physiology. By integrating ATR-FTIR, NMR, and advanced chemometric analysis, this research aspires to support personalized and accessible methods for assessing training responses and physiological adaptation at the molecular level.

EXPERIMENTAL PART

Experimental Part

7. Bioethical considerations

All procedures conducted in this study adhered to ethical guidelines and were approved by the Independent Personal Data Protection Department of the University of Ioannina, Greece. The approval was obtained under protocol number 10253/18-1-2022. Each participant provided informed consent before their involvement in the study, with assurances of confidentiality and proper management of personal data. Participants were thoroughly informed about the study objectives, procedures, and their right to withdraw at any stage. Measures were taken to ensure anonymity through the coding of collected samples.

8. Participants

The research group maintained active communication with sports centers and academies in the city of Ioannina for a period of four years. This engagement aimed to inform stakeholders about the study's objectives and recruit suitable athlete participants. By collaborating with local sports institutions and personal coaches, the team ensured the selection of a diverse and representative cohort, which contributed to the reliable development of experiments and the generation of robust, meaningful results.

This study included 260 athletes recruited from various sports disciplines, each meeting specific eligibility criteria to ensure the validity of the findings. A detailed

questionnaire was administered to all participants to collect information about their training habits, lifestyle, and general health (Appendix 1). The questionnaire included questions about training frequency, intensity, and duration, allowing for the categorization of athletes into low-level and high-level groups. Additional sections addressed dietary habits, and the use of supplements or medications, ensuring that participants met the study's eligibility criteria. This structured approach provided critical baseline data to contextualize biomarker analysis and support the investigation of differences between recreational and competitive athletes.

8.1 Athletic cohort and management

The cohort management process was tailored to support the study's objectives and ensure the reliability of results across different sports disciplines and training intensities.

In the preliminary study of the first Phase, a single national-level long-distance runner was monitored over a period of 12 consecutive days. This initial investigation aimed to assess salivary biochemical changes before and after a standardized training session. The findings from this pilot study provided a foundational understanding of the methodology and helped refine the experimental design for subsequent Phases.

Following the preliminary study, 57 male athletes participated in the main study of the first Phase. These athletes represented seven distinct sports: football (13 athletes), basketball (13 athletes), tennis (7 athletes), muay thai (9 athletes), karate (5 athletes), boxing (9 athletes), and long-distance running (1 athlete). These sports were chosen for their varying demands on aerobic and anaerobic fitness, as well as the distinct metabolic adaptations they induce, thus providing a diverse sample for biochemical analysis. Participants were divided into low-level and high-level groups based on their training frequency and intensity, according to the questionnaire. The athletes' ages and physical characteristics reflected the general profiles of amateur and professional competitors in

these sports, capturing a broad spectrum of training regimens and fitness levels. This Phase specifically aimed to investigate how salivary profile vary between these two groups, offering a unique perspective on the biochemical distinctions associated with recreational and competitive athletic participation.

The second Phase of the study focused on a cohort of 32 high-level male athletes. This cohort included 22 short-distance runners and 10 triathletes actively engaged in swimming, cycling, and running disciplines. These athletes were selected due to their rigorous training schedules and the physiological demands of their respective sports, which offered a valuable framework for studying salivary biomarkers under varying exercise intensities. By including high-level athletes from disciplines with demanding training regimens, this study captured a nuanced perspective on exercise-induced biochemical variations, contributing to the broader understanding of athletic performance and metabolic monitoring.

In Phase 3, participant recruitment was carried out through strategic collaborations with sports centers, academies, and local athletic organizations. Coaches and institutional representatives played an active role in identifying and engaging athletes from a broad range of sporting disciplines, including football, basketball, pole dancing, aerial hoops, tennis, volleyball, middle-distance running, and aerobic gymnastics. In total, 162 athletes (88 males and 74 females) were enrolled in this Phase, and personal data were collected by an additional questionnaire (Appendix 2). The experimental protocol was divided into two distinct parts. The first part involved a subset of 21 non-smoking athletes who underwent controlled treadmill exercise trials. The second part was designed to validate the initial findings under more naturalistic training conditions. For this purpose, a larger and more heterogeneous cohort of 141 athletes (both smokers and non-smokers) was recruited. A detailed presentation of Phase 3 participants' management is presented in Figure 47. This approach allowed for the assessment of salivary thiocyanate fluctuations under real-world training conditions.

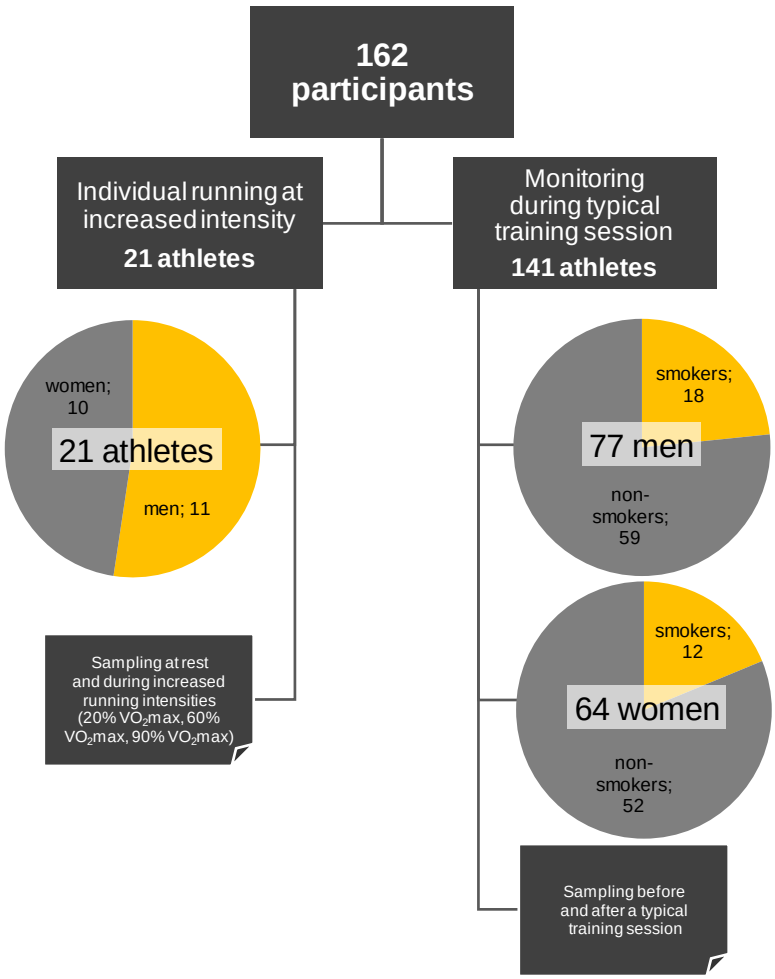


Figure 47. Cohort management and demographic statistics of Phase 3 participants.

In Phase 4, nine male middle-distance (800 m) runners participated, focusing on blood metabolome alterations associated with exercise intensity. The initial part involved two individuals: one professional and one non-professional athlete, selected to represent contrasting training statuses and physiological baselines. These participants were monitored over multiple sessions (4 distinct days) to assess intra-individual variation and the potential influence of training background on metabolic responses.

Subsequently, the cohort was expanded to include seven additional professional athletes, bringing the total number of participants to nine. All professional athletes were actively engaged in structured training programs and competitive sports at the national or

international level. The inclusion of a larger number of athletes enabled the investigation of inter-individual consistency in metabolic alterations associated with exercise and facilitated the validation of preliminary findings from the initial comparative part. They participated in a carefully designed experimental protocol at specific speeds, with blood samples collected. The inclusion of this group allowed for an in-depth exploration of blood-based biomarkers linked to exercise-induced metabolic changes, offering a complementary perspective to the salivary analyses conducted in earlier Phases.

A detailed representation of cohort management in presented in Figure 48. This meticulous approach to cohort management was crucial in maintaining the integrity of the data and achieving the research objectives. Together, these cohorts provided a comprehensive framework for examining the biochemical and metabolic impacts of varying athletic conditions and exercise intensities, advancing the understanding of athletic performance and physiological adaptations.

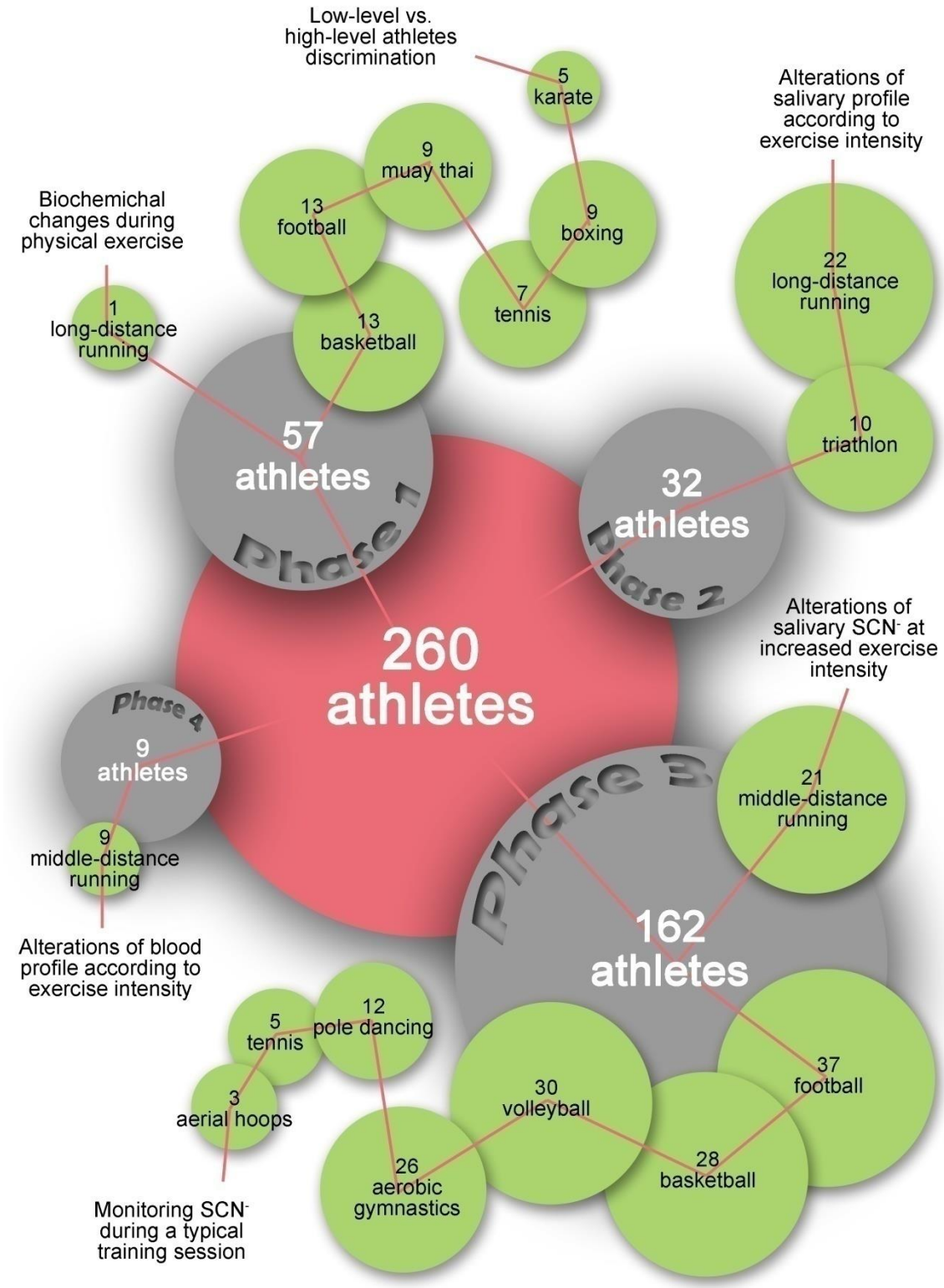

Low-level vs. high-level athletes discrimination
5 karate
9 muay thai
13 football
9 boxing
7 tennis
13 basketball
Biochemichal changes during physical exercise
1 long-distance running
57 athletes
Phase 1
Alterations of salivary profile according to exercise intensity
22 long-distance running
10 triathlon
32 athletes
Phase 2
260 athletes
Alterations of salivary SCN- at increased exercise intensity
Phase 4
9 athletes
9 middle-distance running
Alterations of blood profile according to exercise intensity
Phase 3
21 middle-distance running
162 athletes
5 tennis
12 pole dancing
3 aerial hoops
30 volleyball
37 football
26 aerobic gymnastics
28 basketball
Monitoring SCN- during a typical training session

Figure 48. Athletes' cohort management. Each circle indicates the number of participants and the athletic expertise of each. The paths indicate the recruited athletes in every Phase.

8.2 Anthropometric characteristics

The athletes participating represented a range of sports disciplines, providing a diverse sample for analysis. Participants in the first Phase were divided into low-level and high-level training groups, reflecting differences in training frequency and intensity. The rest Phases focused mainly on high-level athletes. Table 5 summarizes the key anthropometric characteristics of the participants.

Table 5. Anthropometric characteristics of participated athletes.

Phase	Group (no. of participants)	Age (years)	Height (m)	BMI (kg/m²)
1	Low-level athletes (39)	28.8 ± 4.2	1.76 ± 0.05	24.78 ± 0.87
	High-level athletes (18)	25.6 ± 3.0	1.79 ± 0.04	23.09 ± 0.42
2	Short-distance runners and triathletes (32)	26.3 ± 3.0	1.82 ± 0.02	24.10 ± 0.40
3	Men (88)	24.6 ± 6.8	1.82 ± 0.06	22.14 ± 1.41
	Women (74)	24.1 ± 4.9	1.76 ± 0.05	21.02 ± 1.19
4	Middle-distance runners (9)	26.5 ± 3.2	1.84 ± 0.04	21.81 ± 0.90

The anthropometric characteristics of the athletes who participated reflect their training habits and the physical demands of their respective sports disciplines. In the first Phase low-level athletes, primarily recreational participants (hobbyists), demonstrated

slightly higher body mass indices (BMIs) compared to their high-level counterparts. This trend aligns with their less frequent training routines, typically one to two sessions per week at moderate intensity. Conversely, high-level athletes exhibited lower BMIs, indicative of their rigorous training schedules and higher physical conditioning. These participants engaged in structured, high-intensity training regimens at least four times per week, with their sports requiring specialized endurance, strength, or speed. Such consistent physical activity not only enhances fitness levels but also promotes lean body composition.

The athletes in Phases 2, 3 and 4 were all young individuals, with ages ranging from 19 to 29 years, reflecting the prime years of physical performance and athletic development. Their anthropometric data—characterized by taller heights and lean BMIs, compared to lower level athletes of Phase 1—are indicative of highly trained and physically optimized athletes. These characteristics align with the demands of competitive sports such as short/middle-distance running and marathon, emphasizing speed, power, and endurance[292]. The data underscores their status and provides a consistent basis for exploring advanced metabolic and biochemical adaptations to exercise.

8.3 Exclusion criteria

To ensure the integrity and reliability of the findings, strict exclusion criteria were applied. These criteria were designed to minimize potential confounding variables and standardize the participant pool, ensuring that observed variations in salivary and blood biomarkers and other measurements were directly attributable to the study's focus on physical activity and exercise intensity.

8.3.1 General health and lifestyle

Participants were excluded if they engaged in behaviors or had conditions that could significantly alter their physiological or metabolic states. Smoking was prohibited, as tobacco use can influence cardiovascular health, salivary biochemistry, and overall metabolism. Smoking is known to alter biomarkers such as lactate and thiocyanate[293,294], which could confound results. However, in Phase 3 smoker athletes were included but treated separately. In addition, athletes undergoing any medication or treatment that could affect their physical performance or biochemical markers, such as anti-inflammatory drugs, hormonal treatments, or metabolic supplements, were excluded across every Phase. The use of stimulants, energy-boosting supplements, or other performance-enhancing substances was also not permitted, as these could distort baseline profiles and comparisons between participants.

8.3.2 Sex-based differences

Participants were selected based on predefined inclusion and exclusion criteria to ensure the reliability and consistency of the study. In Phases 1, 2, and 4, female participants were excluded due to the small number of female volunteers and the known variability in metabolomics between sexes. Hormonal fluctuations, such as those associated with the menstrual cycle, can significantly affect biomarker levels, introducing additional complexity to data interpretation. However, in Phase 3, both male and female athletes participated, for more robust results.

8.3.3 Nutritional and pre-sampling guidelines

To maintain standard sampling conditions, participants were required to follow strict pre-sampling guidelines. These included avoiding heavy meals for at least three hours before sampling (when sampling during day) or 10-12 hours –overnight fasting– (when sampling in the morning) to prevent the influence of recent food intake on salivary and blood glucose, phosphate, and other metabolites[295,296]. Participants were also instructed to refrain from consuming caffeinated or alcoholic beverages within the same timeframe, as these substances can alter metabolic and hormonal activity[297].

8.3.4 Medical history and conditions

Finally, participants with known medical conditions that could influence metabolic biomarkers were excluded. Conditions such as diabetes, thyroid disorders, and chronic inflammatory or autoimmune diseases often result in altered baseline metabolic states[298–300], which could interfere with the study's ability to detect exercise-induced changes.

By implementing these exclusion criteria, the studies ensured a homogenous participant group that was representative of the targeted populations. This approach minimized external variables and enabled a more accurate assessment of the biochemical and metabolic changes associated with physical exercise and training intensity.

8.4 COVID-19 considerations

Given that the study was conducted during the COVID-19 pandemic, extensive measures were implemented to ensure the safety of participants and researchers, as well

as the reliability of the data. All athletes were required to undergo SARS-CoV-2 testing immediately prior to participation, ensuring that only those with a confirmed negative result were allowed to proceed with the study. This precaution was essential to prevent any potential transmission of the virus during sample collection and to maintain a controlled and safe research environment.

Athlete selection and recruitment processes were adapted to comply with public health guidelines. The research team maintained close communication with sports centers and academies to minimize large gatherings and conducted much of the initial recruitment virtually or in small, staggered sessions to adhere to social distancing protocols. On-site sample collection was carefully organized, with participants arriving individually at scheduled times to avoid overlap and ensure proper distancing. Additional hygiene measures were enforced throughout the study. Researchers wore appropriate personal protective equipment (PPE), including masks and gloves, always during interactions with the participants. Hand hygiene was prioritized, and sanitization of surfaces and equipment was carried out regularly between sampling sessions.

These considerations not only safeguarded the health of everyone involved but also ensured the continuity of the research under challenging circumstances. The measures demonstrated the study's commitment to ethical research practices and public health compliance, reflecting the adaptability of the methodology in response to unprecedented global challenges.

9. Experimental design

The experimental design across all four Phases, was carefully structured to evaluate the physiological responses to a variety exercise intensities. In Phase 1, a preliminary study was conducted with a single, national-level, long-distance runner, monitored over 12 consecutive days to examine the impact of regular training sessions. Thus, saliva

samples were collected before and directly after his training session. In the main work of the Phase 1, 57 athletes from diverse sports disciplines completed a standardized post-exercise assessment, so samples were collected after their workout.

The experimental protocol for Phases 2, 3 and 4 focused on assessing physiological responses at specific exercise intensities. In Phase 2, the athletes came to the outdoor running track stadium for sampling at 09:00 a.m. to 10:00 a.m. Athletes performed four distinct exercise states in sequence: rest (0 km/h), walking (4–5 km/h), jogging (9–10 km/h), and running (14–15 km/h). Sampling was conducted at each of these time points to capture the effects of progressive intensity levels. The total distance covered by athletes during the protocol was 2-3 km on a standard running track, ensuring consistency across participants.

On the other hand, Phases 3 and 4 were conducted on a running treadmill to standardize further the athletic conditions during sampling. Herein, the exercise intensity was evaluated according to each athlete's VO_2max value for more representative workout. More specifically, the VO_2max value of each athlete was measured *via* ergometric test in a third-party laboratory using the Bruce treadmill protocol, a widely accepted graded exercise test designed to progressively increase workload until volitional exhaustion. Each test was performed using the same metabolic cart and treadmill model to maintain consistency across measurements. Prior to each test, equipment calibration and participant familiarization were conducted in accordance with the laboratory's standard operating procedures. Based on the globally recognized American College of Sports Medicine® (ACSM) Metabolic Equations[301], the targeted exercise effort (i.e., 20% VO_2max, 60% VO_2max and 90% VO_2max) converted to the set speed on treadmill. For instance, Table 6 presents the VO_2max value of each participant in Phase 4 and the corresponding speed-state who run. The presentation of exercise intensity as a percentage of VO_2max in Phases 3 and 4 aligns well with the less mature approach used in Phase 2, where intensity was expressed more directly through running speeds (km/h). Although expressing intensity in km/h represents a more straightforward and less

physiologically tailored method, using % VO_2max provides a more individualized and standardized measure of exertion across athletes with varying fitness levels. By linking both methods, the study maintains continuity between Phases while enhancing precision in exercise prescription and physiological interpretation in these Phases.

Table 6. Running speeds corresponding to 20%, 60%, and 90% of VO_2max for each athlete participated in Phase 4.

Athlete	VO_2max (mL/kg/min)	Set speed (km/h)		
		20% VO_2max	60% VO_2max	90% VO_2max
A	51.6	4.1	8.2	12.9
B	57.4	4.8	9.3	14.4
C	63.0	5.5	10.3	16.0
D	52.3	4.2	8.4	13.1
E	63.7	5.5	10.4	16.2
F	62.5	5.4	10.2	15.8
G	56.7	4.7	9.1	14.2
H	59.5	5.0	9.7	15.0
I	65.0	5.7	10.7	16.5
Average	**59.1**	**5.0**	**9.6**	**14.9**
SD	**4.9**	**0.6**	**0.9**	**1.3**

The athletes, in both Phases 3 and 4, covered a distance of 1 km in each exercise state, exercising in total of 3 kilometers. It should be mentioned that at the end of Phase

3, an additional sampling was conducted prior and after a typical training routine of a large cohort of athletes (see section 8.1) to establish the Phase's findings. This structured design allowed for a precise evaluation of how different exercise intensities influence athletic performance and physiological responses. Figure 49 depicts the experimental design of the whole study.

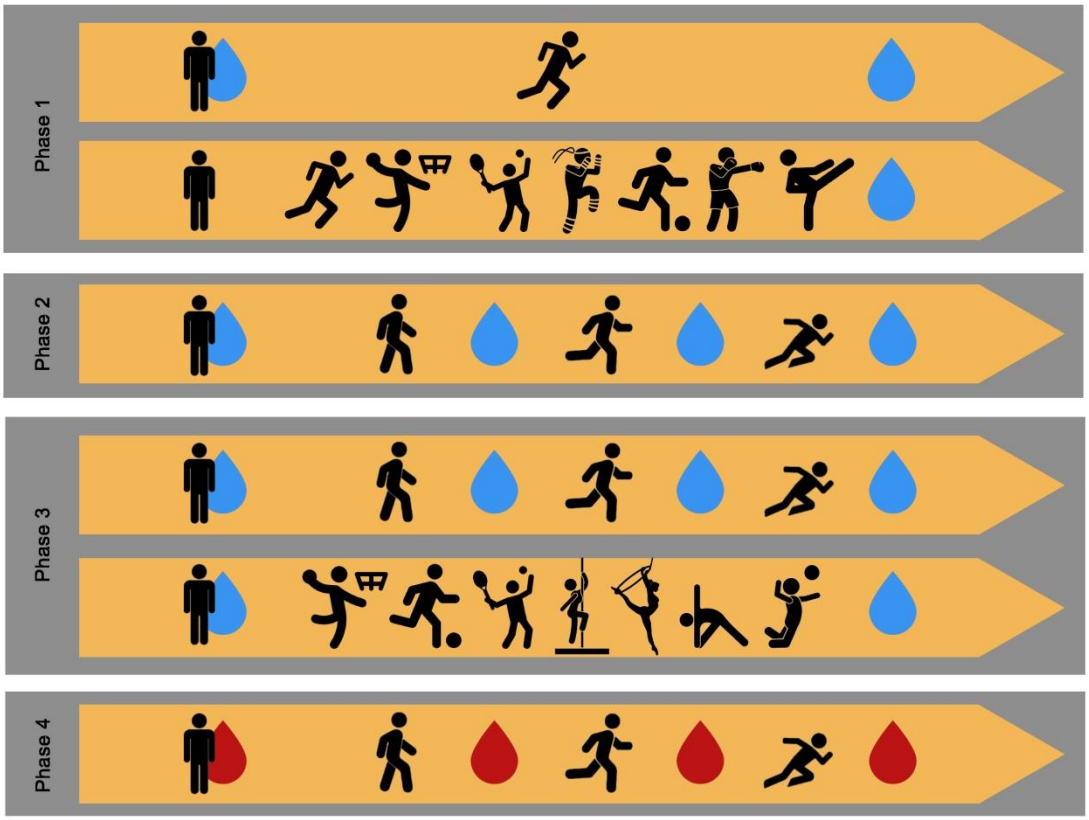


Figure 49. Sampling time-points in each Phase. Phase 1: saliva sampling before and after running session, and after training session of various sports. Phase 2: saliva sampling at rest and after running session of different intensities. Phase 3: saliva sampling at rest and after running session of different intensities, and before and after training sessions of various sports. Phase 4: blood sampling at rest and after running session of different intensities.

10. Sample collection

The sample collection process was meticulously designed to ensure consistency, reliability, and minimal disruption to the athletes' training routines. Specific protocols were followed for each type of sample, with saliva sampling conducted in Phases 1, 2 and 3 and blood sampling conducted in Phase 4.

10.1 Saliva sampling

Saliva samples were collected before and immediately post-exercise as described above. Athletes were instructed to avoid food, caffeine, and alcohol for at least three hours before sampling and to refrain from oral hygiene practices, such as brushing teeth, flossing or using mouthwash, for 30 minutes before the session to avoid contamination of saliva with dental products that might interfere with the analysis[302].

In Phase 1, on the day of sampling, participants rinsed their mouths with water approximately 10 minutes before the session to remove any residual contaminants. Saliva was collected directly into sterile 1.5 mL Eppendorf tubes by passive drooling, avoiding stimulation techniques.

In both Phases 2 and 3, participants used the Salivette® (SARSTEDT, Numbrecht, Germany) collection technique for sample collection, which involved chewing on a sterile swab for a standardized period to stimulate saliva production. The swab was then placed into a sterile Salivette® tube designed to collect and filter the saliva efficiently. This method ensured the consistent volume and quality of saliva collected across all participants.

Each sample was labeled with a unique code to maintain anonymity and facilitate precise tracking. After sampling the containers with the saliva samples were analyzed

immediately or were stored in refrigerator at 4°C until analysis (the analysis done in a 4 h window), to prevent degradation of some components and bacterial growth[303,304].

10.2 Blood enzymatic assay

In Phase 2 of the study, alongside saliva sampling, blood glucose and lactate levels were measured using commercially available portable assays. Blood samples were obtained via finger prick at rest and immediately after each exercise intensity (walking, jogging, and running). The measurements were conducted using validated handheld devices, ensuring rapid and accurate quantification of glucose and lactate concentrations in the field. Lactate analyzer Accutrend® Plus (Roche®, Germany) and reagent strips (BM-Lactate, Roche®) were utilized to determinate the concentration level of lactate in blood, while Contour® analyzer (Bayer®, Germany) and Contour® blood glucose test strips (Bayer®, Germany) were used for glucose measurements. These biomarkers provided additional insights into the physiological responses to incremental exercise intensities, complementing the salivary analysis, comparing the results with responses in saliva, and offering a more comprehensive understanding of the athletes' metabolic adaptations.

10.3 Blood sampling

In Phase 4, blood samples were obtained using the minimally invasive dried blood spot (DBS) technique. In the field of blood microsampling, commercially available materials such as the Whatman 403 Protein Saver Cards are widely used due to their reliability and standardization. However, the development and application of synthesized materials for blood sampling present significant advantages, including potential improvements in analyte adsorption, selective extraction and extraction efficiency, and cost-effectiveness. By exploring alternative materials inspired by the FPSE extraction

technique, it becomes possible to optimize sampling performance, enhance selectivity for specific biomarkers, and tailor the properties of the substrate to suit diverse analytical needs[305]. This approach not only expands the range of available microsampling media but also opens new avenues for method customization and innovation in bioanalytical research.

10.3.1 Synthesis of adsorptive-FPSE based materials

In this study, four synthesized materials were prepared for blood microsampling, inspired by the Fabric Phase Sorptive Extraction (FPSE) technique. Two fabric substrates – Whatman Cellulose filter of 125 mm and Whatman Microfiber Glass filter of 110 mm (GE Healthcare Bio-sciences Corp, Piscataway, NJ, USA) – and two different polymers – polyethylene glycol (PEG 300) and poly(ethylene glycol)-block-poly(propylene glycol)-block-poly(ethylene glycol) (PEG-PPG-PEG 5.800) (Sigma-Aldrich, Burlington, Massachusetts, U.S.) – were tested. In total, four different combinations arose for the blood microsampling proposes.

The synthesis process involved two main steps: pretreatment of fabric substrates and the application of sol–gel coatings to create the desired sorbent materials.

Pretreatment of Fabric Substrates:

Both fabrics/substrates chosen to produce sol–gel covered sorbents were initially soaked in deionized water under sonication to ensure thorough wetting and removal of impurities. The fabrics were then subjected to a cleaning/activation process by treating them with 1 M sodium hydroxide (NaOH) under sonication for an hour. Following this, the fabrics were washed extensively with deionized water to remove residual NaOH. Subsequently, the fabrics were treated with 0.1 M hydrochloric acid (HCl) under sonication for an hour, washed again with deionized water, and finally dried overnight in

an inert atmosphere. The dried fabric substrates were stored in clean, airtight glass containers until further use.

Preparation of sol–gel solution:

The pretreated fabric substrates were coated with sol–gel derived sorbents to create four distinct extraction media. The sol–gel synthesis and coating processes is summarized in the following lines: Each sol–gel solution was prepared by dissolving 5 g of the corresponding polymer in 10 mL mixture of acetone and dichloromethane (50/50% v/v) and vortexing for 1 min. Then in a separated container, 5 mL of the generic precursor molecule methyltrimethoxysilane (MTMS) was added, followed by the addition of 2 mL trifluoroacetic acid (TFA) catalyst to hydrolyze MTMS, initiating polycondensation. The two mixtures were combined (5 min vortex) forming a growing sol–gel network incorporating the polymer.

Coating of the media:

The pretreated fabric media were immersed in this solution for 4 h. The coated fabric was removed from the sol–gel solution and left in a desiccator overnight for solvent evaporation. It was then rinsed with acetone/dichloromethane (50/50% v/v) under sonication for 30 min to remove unreacted residues. Finally, the fabric was cut into 1 cm diameter circles and stored in a sealed container to prevent contamination.

The resulting sol–gel coated fabrics served as the synthesized materials for blood microsampling, each possessing unique sorptive properties tailored through the sol–gel chemistry employed during their preparation.

10.3.2 Blood microsampling with different materials

Athletes performed four exercise states—rest, 20% VO_2max, 60% VO_2max, 90% VO_2max, —and blood samples were collected immediately after each state. Capillary blood samples (approx. 45 µL) were taken via finger prick from the third (middle) or fourth (ring) fingertip, which had been thoroughly cleaned with water to avoid interference from alcohol or soap residues.

Using a sterile lancet, a drop of blood was allowed to form and was carefully applied to the center of a Whatman 903 Protein Saver Card (GE Healthcare Bio-sciences Corp, Piscataway, NJ, USA) (Fig. 50a) or on the precut synthesized materials. Each medium was left to dry at room temperature for two hours to ensure stability before storage at -80°C until analysis. The DBS technique allowed for convenient field sampling, reducing logistical challenges while maintaining the reliability of the collected samples.

These standardized protocols ensured the collection of high-quality saliva and blood samples, facilitating robust analyses and reliable comparisons across the different Phases of the study

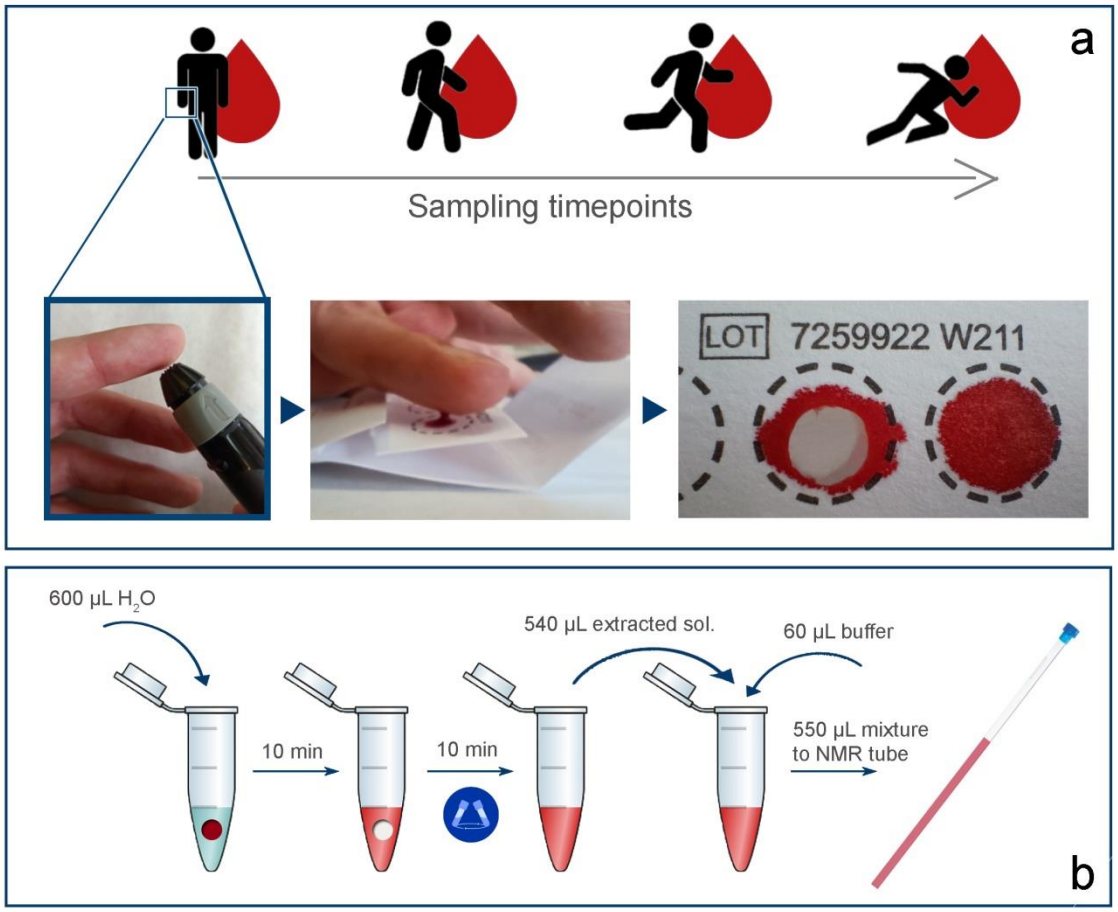


Figure 50. (a) Collection of blood microvolume from the fingertip on the Whatman 903 DBS card and (b) blood extraction process.

.

11. Sample pre-processing

The sample pre-processing protocols were designed to ensure the optimal preparation of saliva and blood samples for subsequent analyses. Specific procedures were implemented for each biofluid, reflecting the objectives and methodologies of the respective study Phases.

11.1 Saliva handling

In Phases 1 and 2, prior analysis, all saliva samples were dried to eliminate residual humidity, which can interfere with ATR spectra by masking key molecular vibrations and reducing spectral clarity. Removing moisture ensures that the detected signals represent the chemical composition of the sample rather than water-related artifacts. Optimizing the drying step is therefore critical to achieving consistent and high-quality spectra, enabling reliable interpretation and reproducibility in ATR-based saliva analysis.

Therefore, to optimize the drying process of saliva for ATR-FTIR analysis, a pooled saliva sample was prepared and evaluated using two drying methods. In the first method, 1 mL of the pooled saliva sample was placed on a watch glass and dried at 37°C. After the overnight drying process the dried saliva was manually scraped and transferred to the ATR crystal for analysis. This procedure was repeated ten times to assess the precision of the technique. For the second method, 10 µL of the pooled saliva sample was deposited onto a cover glass and dried at the same temperature. Once the initial layer of saliva was dried, an additional 10 µL was applied directly on top of the dried layer to form a second stacked layer. This process was repeated iteratively to create a total of 100 stacked layers. The cover glass was then inverted to place the dried layers in contact with the ATR crystal for measurement. This method was also repeated ten times to evaluate its precision. Both methods were compared to determine which produced higher-quality ATR spectra.

Thus, in Phase 1, saliva samples were centrifuged by Centurion Scientific K241 (Centurion Scientific Ltd, Chichester, WS, UK) at 10,000 rpm for 10 minutes to remove large molecular debris and obtain a clear supernatant. The supernatant was then dried, with the optimum drying method, to prepare it for ATR-FTIR spectroscopy analysis.

In Phase 2, saliva samples collected using the Salivette® technique were first centrifuged at 10,000 rpm for 5 minutes to extract absorbed saliva from the cotton

swabs. The collected saliva was then subjected to further processing, similar to Phase 1. The biofluid was subjected again to centrifugation at 10,000 rpm for 10 min to separate and precipitate larger molecules. Aliquots of the centrifuged saliva were dried under controlled conditions to ensure consistent preparation for analysis. This drying process concentrated the salivary components, allowing for more precise spectral readings and biomarker identification. Finally, dried saliva samples were analyzed with ATR-FTIR instrumentation and the collected spectra were further processed.

In Phase 3, saliva samples, also collected using the Salivette® technique, were centrifuged at 10,000 rpm for 5 minutes to extract absorbed saliva from the cotton swabs, and the supernatant was collected for thiocyanate determination.

11.2 Blood handling

In Phase 4, blood samples were processed using the dried blood spot (DBS) technique. In the case of the synthesized materials the media were already precut, so blood was deposited directly on the top on the media, where they were then placed in an Eppendorf tube. On the other hand, discs (6 mm in diameter) were punched from the dried blood spots of the Whatman 903 DBS cards using a sterile puncher. Each disc, equivalent to approximately 2.2 mg of dried blood, was placed in an Eppendorf tube. In both cases, a volume of 600 µL of ultrapure water (produced by a Milli-Q system, Evoqua, Pittsburg, USA) was added to the tube, and the mixture was gently vortexed for 10 minutes to rehydrate the dried blood and extract metabolites (Fig. 50b). Using water as the sole extraction solvent was sufficient to elute blood cells and other endogenous components from the synthesized materials and DBS cards. The volume of water used was kept to a minimum to ensure the extracted solution remained concentrated while still being adequate for NMR analysis. The solution was then centrifuged at 12,000 rpm for 10 minutes to remove the paper punch and solid materials from the card, if appeared.

Each NMR sample was prepared by combining 10% (60 µL) of a deuterated buffer solution with 90% (540 µL) of the blood extract (Fig. 50b). The deuterated buffer, widely employed in NMR-based metabolomics of urine, consisted of 1.5 M potassium dihydrogen phosphate (KH_2PO_4) dissolved in 99.9% deuterium oxide (D_2O). The buffer was adjusted to a pH of 7.4 and contained additional components, including 2 mM sodium azide to prevent microbial growth and 0.1% sodium 3-(trimethylsilyl)propionate-d_4 (TSP) as a chemical shift reference standard. The resulting mixture was thoroughly homogenized to ensure consistency before transferring a final volume of 550 µL into a 5 mm NMR tube. This volume and composition were optimized to provide sufficient sample integrity and signal quality for subsequent NMR analysis.

These standardized protocols for saliva and blood handling ensured high-quality samples were prepared for advanced biochemical and metabolomic analyses, enabling robust and reproducible findings across the study Phases.

12. Thiocyanate assay in saliva

In Phase 3, thiocyanate ions were studied in detail. The determination of thiocyanate in saliva was performed according to the well-established method of thiocyanatoiron (III) ion[306].

12.1 Calibration curve *via* artificial saliva

To develop and validate a method for thiocyanate determination in saliva, a series of thiocyanate standards were prepared at concentrations ranging from 0.01 to 1.5 mM. Standard thiocyanate solutions were prepared using potassium thiocyanate (KSCN) dissolved in artificial saliva to create a stock solution of 1.5 M, which was further diluted

as needed. The calibration curve was constructed using artificial saliva instead of deionized water to better mimic the sample matrix and account for potential matrix effects. Artificial saliva was prepared following a standardized composition[307], consisting of 125.6 mg/L sodium chloride (NaCl), 963.9 mg/L potassium chloride (KCl), 227.9 mg/L calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$), 178 mg/L ammonium chloride (NH_4Cl), 336.5 mg/L sodium sulphate (Na_2SO_4), 200 mg/L urea (CH_4N_2O), 630.8 mg/L sodium bicarbonate ($NaHCO_3$), and 654.5 mg/L potassium dihydrogen phosphate (KH_2PO_4), all dissolved in deionized water. Thus, standard thiocyanate solutions were prepared in this artificial saliva medium.

12.2 Colored complex formation

The reagent of iron(III) nitrate solution was prepared at a concentration of 0.2 M by dissolving iron(III) nitrate $Fe(NO_3)_3$ nonahydrate in 1 M HNO_3. For analysis, 1.0 mL of the standard solution or real saliva sample was mixed with 1.0 mL of the iron nitrate reagent. The mixture was allowed to react (vortex), and the complex was formed immediately. The resulting solution was then transferred to a cuvette, and the absorbance was measured photometrically at 458 nm.

13. Instrumentation

13.1 ATR-FTIR spectra acquisition

The acquisition of ATR-FTIR (Attenuated Total Reflectance-Fourier Transform Infrared) spectra was performed in Phases 1 and 2 to analyze the biochemical composition of saliva samples. The same instrument and operational conditions were applied in both Phases to ensure consistency and comparability of results.

A Spectrum Two FT-IR spectrometer equipped with a UATR Two Accessory (Perkin Elmer, Waltham, MA, USA) and Spectrum 10 Spectroscopy Software v. 10.5.4 (Perkin Elmer, Waltham, MA, USA), was used for spectral acquisition. The system operated in transmittance mode in the mid-infrared range (4000–450 cm^{-1}). Background spectrum was obtained before measurements, to subtract it from each sample spectrum, providing higher quality spectra. Each dried saliva sample was placed directly on the ATR crystal, ensuring full contact for optimal signal acquisition. Approximately 1.5 mg of dried saliva was applied to the crystal surface (2.0 × 2.0 mm) for analysis. The spectra were recorded with a resolution of 4 cm^{-1}, and 32 scans were averaged per sample to enhance the signal-to-noise ratio. Prior to each measurement, the ATR crystal was cleaned thoroughly with isopropanol to prevent cross-contamination between samples.

The recorded spectra provided detailed molecular fingerprints of the saliva samples, enabling the identification and quantification of specific biochemical changes associated with exercise intensity and athlete training levels.

13.2 Photometric acquisition for thiocyanate assay

The quantification of thiocyanate in saliva was performed using a spectrophotometric assay based on the formation of an iron(III) thiocyanate complex. When thiocyanate reacts with iron(III) nitrate, a red-colored complex forms, exhibiting a characteristic absorbance at 458 nm. The intensity of the color is directly proportional to the thiocyanate concentration, allowing for its quantification. Absorbance measurements were conducted using a UV-Vis spectrophotometer (UV-1800 Spectrtophotometer, Shimadzu, Kyoto, Japan), and sample concentrations were determined by interpolation from a calibration curve prepared using artificial saliva standards.

13.3 NMR spectra acquisition

One-dimensional (1D) ^{1}H-NMR spectra were acquired using a Bruker AV500 spectrometer operating at a frequency of 500.13 MHz (Bruker Biospin, Rheinstetten, Germany). The temperature was maintained at a constant 310 K throughout the measurements, which were conducted using the TopSpin v4 software suite. The spectra were recorded using the 1D ^{1}H zgpr pulse sequence, with 98,000 data points collected across a spectral width of 14,705 Hz. A total of 512 scans were performed, preceded by 8 dummy scans, and a relaxation delay of 4 seconds was applied between scans.

The acquired spectra were zero-filled to 132K data points and Fourier-transformed following the application of an exponential apodization function with a 0.3 Hz line broadening factor. Baseline and Phase corrections were initially performed using Bruker's automated software, IconNMR, integrated with TopSpin 4.06. When necessary, manual corrections were applied, including zero-order and first-order Phase adjustments, followed by baseline correction. Baseline adjustments were performed by fitting a third-degree polynomial function to spectral regions devoid of peaks. These manual corrections were carried out using the "apk" (automatic Phase correction) and "abs" (automatic baseline correction) functions in TopSpin.

The TSP signal was calibrated to 0.0 ppm for all spectra to ensure consistency. Additionally, two reference samples (SRs) were prepared by pooling aliquots from all study samples. These pooled reference samples were analyzed periodically throughout the study's Phase to serve as quality control (QC) standards, ensuring reproducibility and reliability of the NMR measurements The ^{1}H-NMR spectra of all samples were acquired using identical experimental settings (PQN normalization).

14. Data analysis

Working on Phase 1 of the study, ATR-FTIR spectra were first preprocessed by converting transmittance to absorbance, followed by baseline correction using the Sensitive Nonlinear Iterative Peak (SNIP) algorithm[308] and normalization based on the broad band at 3200–3300 cm^{-1} corresponding to water[309]. Principal Component Analysis (PCA) was applied to explore variations in salivary spectral profiles during physical exercise, while both PCA and Partial Least Squares Discriminant Analysis (PLS-DA) was used to classify athletes based on their fitness levels. Model optimization and validation were conducted using a 3-fold repeated stratified cross-validation (30 repeats), determining the optimal number of components based on the Root Mean Squared Error of Cross-Validation (RMSECV) and Prediction (RMSEP). All analyses were performed using Python (v3.8) with scikit-learn[310], pandas[311], and numpy[312] libraries.

In Phase 2 spectra were baseline corrected using the Spectragryph licensed application software version 1.2.15, and multiplicative scatter correction (MSC) was applied to remove light scattering effects caused by surface irregularities in dried oral fluid. A second derivative transformation was performed to enhance the separation of overlapping peaks[313]. PCA was initially used to explore patterns in the dataset, while PCA-LDA and PLS-DA was applied for classification of saliva spectra based on exercise intensity. The total spectral range (4000-450 cm^{-1}) was utilized for analysis, with the dataset split into 70% for training and 30% for testing. A 10-fold repeated stratified cross-validation (100 repeats) was conducted to determine the optimal number of components while minimizing prediction error. All multivariate procedures and analyses were performed using MATLAB (R2019a, The Mathworks, Natick, MA, USA) with in-house scripts based on “Partial Least Squares Regression and Principal Components Regression - MATLAB & Simulink Example.”, in Mathworks website (Appendix 3). To assess statistical significance in specified metabolites variations across exercise intensities, Analysis of Variance (ANOVA) was performed. However, as ANOVA does not indicate where the differences occur, post hoc tests were conducted to compare group means pairwise. The Bonferroni

correction was applied to adjust for multiple comparisons, reducing the risk of false positives and ensuring the reliability of the findings. Statistical analyses were conducted using Microsoft Excel 2007 (Microsoft Corporation, Redmond, WA, USA).

Similarly, statistical analysis in Phase 3 was done with Microsoft Excel 2007 (Microsoft Corporation, Redmond, WA, USA).

The following data analysis pertains to Phase 4 of the study. The processed 1D ^{1}H NMR spectra were imported into the SMolESY platform[314] to remove macromolecular baseline contributions while preserving the quantitative integrity of the spectra[315]. This preprocessing step enabled accurate normalization of spectral data by taking into consideration of dilution effects on total signal intensity[316]. Spectral datasets after SMolESY normalization, ranging from 0.50 to 10.00 ppm, while the suppressed H_2O region from 4.3 to 5.2 ppm was removed, were binned using a 0.02 ppm bin width, resulting in a dataset of 425 variables. This dataset was used for the untargeted analysis. In addition, a targeted dataset was created by semi-automated peak assignment and quantification for metabolites identification, similarly with previous study[317]. Both datasets were normalized using Probabilistic Quotient Normalization (PQN)[318] to account for dilution variability introduced during the extraction of blood from DBS.

All normalized datasets were imported into MATLAB software (v. 2021b, MathWorks) and analyzed using Principal Component Analysis (PCA) and Partial Least Squares - Discriminant Analysis (PLS-DA), through the PLS_Toolbox (v. 8.7.1, Eigenvector Research, Inc., Manson, WA, USA). Prior to the statistical processing, the SMolESY profiling data were mean-centered. All statistical models and corresponding performance metrics were generated following a cross-validation procedure. To identify important spectral features, variables loadings and Variable Importance in Projection (VIP) scores from the PLS-DA models were examined, with VIP values equal to or higher than 1 considered indicative of high influence. Additionally, the statistical significance of metabolite concentrations was evaluated through univariate analysis using one-way ANOVA build-in functions available in MATLAB

(https://uk.mathworks.com/help/stats/one-way-anova.html). To correct for multiple testing and control the false discovery rate, the Benjamini-Hochberg correction was applied using the "fdr_bh" script (https://www.mathworks.com/matlabcentral/fileexchange/27418-fdr_bh).

RESULTS AND DISCUSSION

Results and Discussion

15. Saliva drying optimization

To optimize the drying process of saliva for ATR-FTIR analysis, two drying methods were evaluated. In the first method, 1 mL of pooled saliva was dried at 37°C for approximately 20 hours (overnight) (Fig. 51c), manually scraped, and transferred to the ATR crystal (Fig. 51d). In the second method, 10 µL aliquots of saliva were sequentially layered and dried on a cover glass to create progressively 100 stacked layers (Fig. 51b), which were then inverted onto the ATR crystal (Fig. 51e).

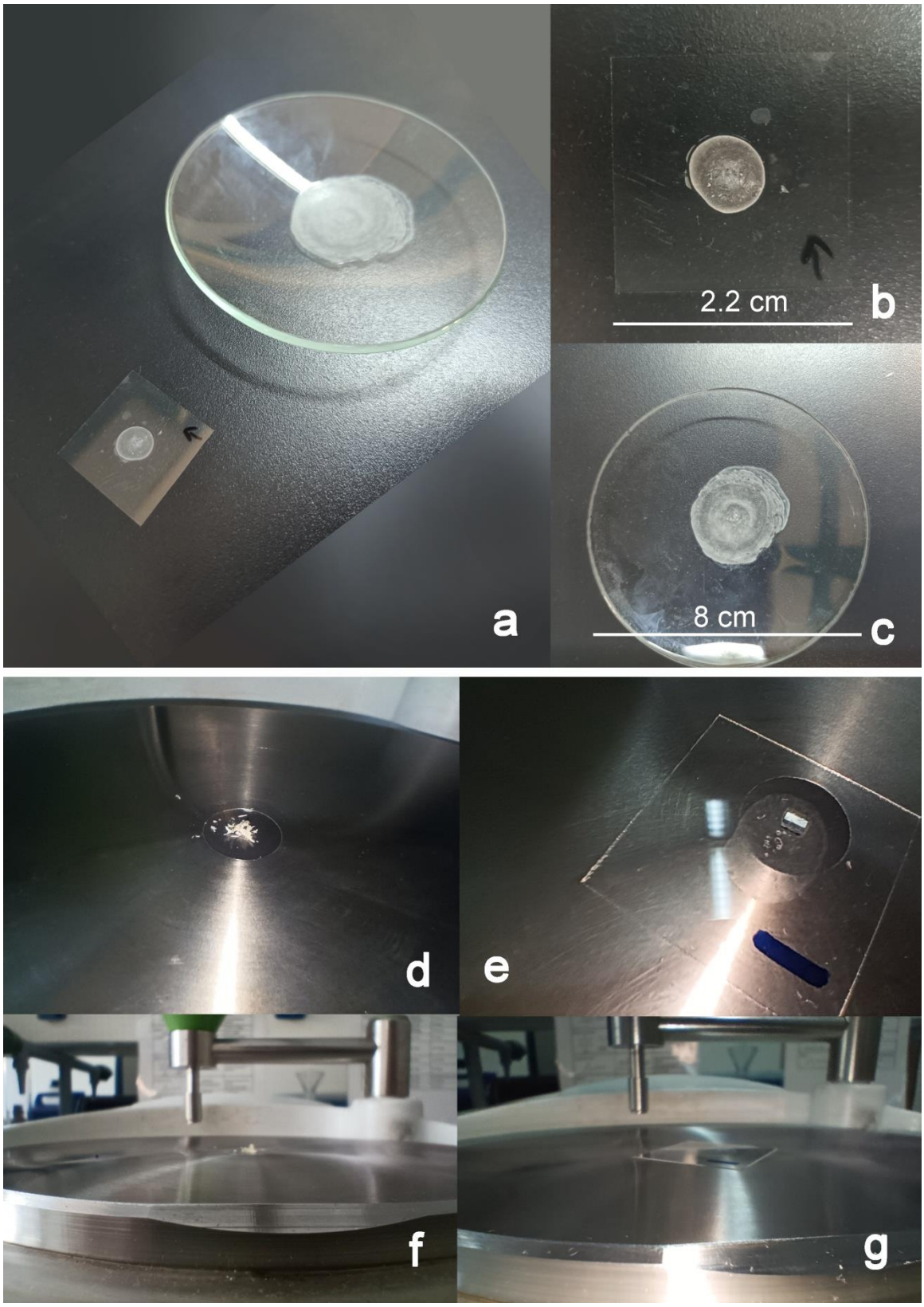

a
2.2 cm
b
8 cm
c
d
e
f
g

Figure 51. Images of the drying process of saliva with the two methods: drying on watch glass and manual scrapping and drying on cover glass and direct placement on ATR crystal. (a) Side by side viewpoint, (b) top view of dried saliva with cover glass method and (c) top view of dried saliva with watch glass method. (d) Top view of manually scrapped dried saliva placed on ATR crystal, (e) top view of inverted cover glass placed on ATR crystal, (f) side view of manually scrapped dried saliva placed on ATR crystal and (g) side view of inverted cover glass placed on ATR crystal.

The evaluation of the two drying methods revealed notable differences in the quality and intensity of the ATR-FTIR spectra obtained (Fig. 52). The method involving the manual scraping of dried saliva demonstrated superior performance in terms of spectral intensity and sensitivity. The spectra acquired using this method consistently exhibited sharper peaks and higher signal intensities, indicating improved interaction between the solid saliva sample and the ATR crystal surface. This enhanced contact is attributed to the direct application of the scraped saliva, which likely allowed for more uniform and intimate contact with the crystal, minimizing gaps or inconsistencies.

In contrast, the stacked-layer method, while innovative, resulted in relatively lower spectral intensity. This reduction in sensitivity may be due to the inherent structural inconsistencies introduced during the iterative layering process. Although the stacked layers formed a solid structure upon drying, their contact with the ATR crystal was likely less uniform due to potential air gaps or surface irregularities caused by the stacking technique. Furthermore, the inversion of the cover glass to place the dried layers in contact with the crystal could have introduced additional challenges, such as uneven pressure or incomplete contact over the crystal's surface.

Reproducibility tests for both methods showed consistent results across the ten repetitions, confirming that the drying processes were stable and repeatable. However, the manual scraping method not only demonstrated better spectral quality but also

proved to be more practical for ensuring optimal sample-crystal interaction in ATR-FTIR analysis.

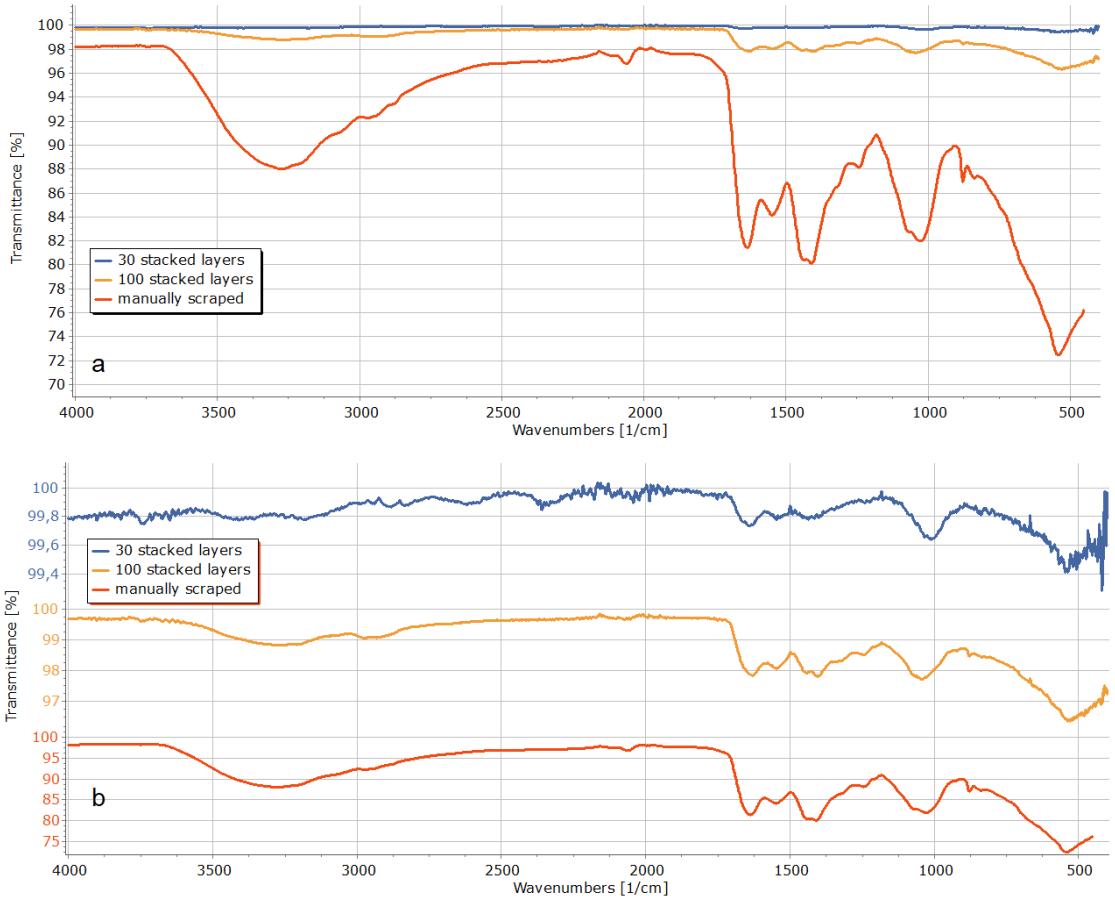


Figure 52. Averaged spectra (n=10) of dried pooled saliva with the two drying methods: drying on watch glass and manual scrapping and drying on cover glass and direct placement on ATR crystal (30 layers and 100 layers are presented). (a) Overlay presentation, (b) stacked presentation.

These findings underscore the critical role of sample preparation in ATR-FTIR saliva analysis. The manual scraping method's ability to produce high-quality spectra with greater sensitivity highlights its suitability on later applications requiring detailed and

reliable chemical characterization. In contrast, while the stacked-layer method offers an alternative approach, its limitations in spectral intensity suggest that further refinement is necessary to enhance its efficacy for ATR-FTIR analysis.

16. Phase 1 – Spectroscopic and chemometric characterization of saliva for athlete profiling: A study using ATR-FTIR

Driven by the scientific community's increasing interest in non-invasive biomarker monitoring, this Phase focused on utilizing a robust approach for tracking biomolecular changes in saliva during physical exercise through ATR-FTIR spectroscopy. Saliva was chosen as the biofluid of interest due to its ease of collection, participant-friendly nature, and rich biochemical composition reflective of systemic physiological changes[319]. ATR-FTIR spectroscopy was preferred in this study due to its minimal sample preparation requirements and ability to analyze complex biological matrices, such as saliva, directly. The technique's sensitivity to molecular vibrations allows for the detection of subtle biochemical changes, making it ideal for monitoring minimal biomarker variations in minimal sample volume during physical exercise[303].

16.1 Salivary biochemical changes before and after physical exercise

A preliminary study was performed to investigate the utility of our methodology to monitor the salivary profile during exercise. The preliminary study conducted in Phase 1 aimed to reveal distinct biochemical changes in saliva before and after physical exercise, using ATR-FTIR spectroscopy.

A single national-level long-distance athlete was monitored over 12 consecutive days to evaluate the feasibility of using ATR-FTIR spectroscopy for salivary biomarker analysis. Saliva samples were collected both before and after the athlete's daily training sessions, capturing the biochemical changes associated with exercise. This longitudinal approach provided valuable insights into the day-to-day metabolic fluctuations induced

by variables such as sleep patterns, physical activity outside training, or hydration status. By collecting saliva samples consistently before and after each workout over consecutive days, the study reduced the influence of transient factors, allowing a clearer focus on exercise-induced biochemical changes. This approach ensured more reliable and representative data.

Thus, PCA was employed to evaluate these spectral differences and to identify the key biochemical features driving discrimination. The scree plot displaying the explained variance of up to 10 principal components (PCs) showed a steep decline in variance explained by the first few PCs, followed by a gradual leveling off (Fig. 53). PC1 accounted for a significant portion of the variance (83.27%) reflecting the dominant variation in the data. PC2 added another substantial contribution (6.03%) bringing the cumulative explained variance to 89.30%. Subsequent PCs, such as PC3 and PC4, explained progressively smaller amounts of variance 3.53% and 2.76%, respectively. Beyond PC5, the variance contributions diminished further, with PC6 to PC10 each explaining less than 1%, reflecting noise or minor variations unrelated to the primary patterns in the data.

The "elbow" of the scree plot clearly observed on PC2, indicating that the majority of the variance can be captured with the first two components, making them the most significant for further analysis. According to the Percentage of Variance Criterion, which recommends retaining components until a specific percentage of the total variance is explained, the first three PCs are more than sufficient to capture the majority of the data’s variability (>90%). This approach balances data simplification with information retention. Following this criterion, the score plots of these three principal components were examined, providing a detailed visualization of the clustering and separation between pre- and post-exercise samples, highlighting the key biochemical differences induced by physical activity.

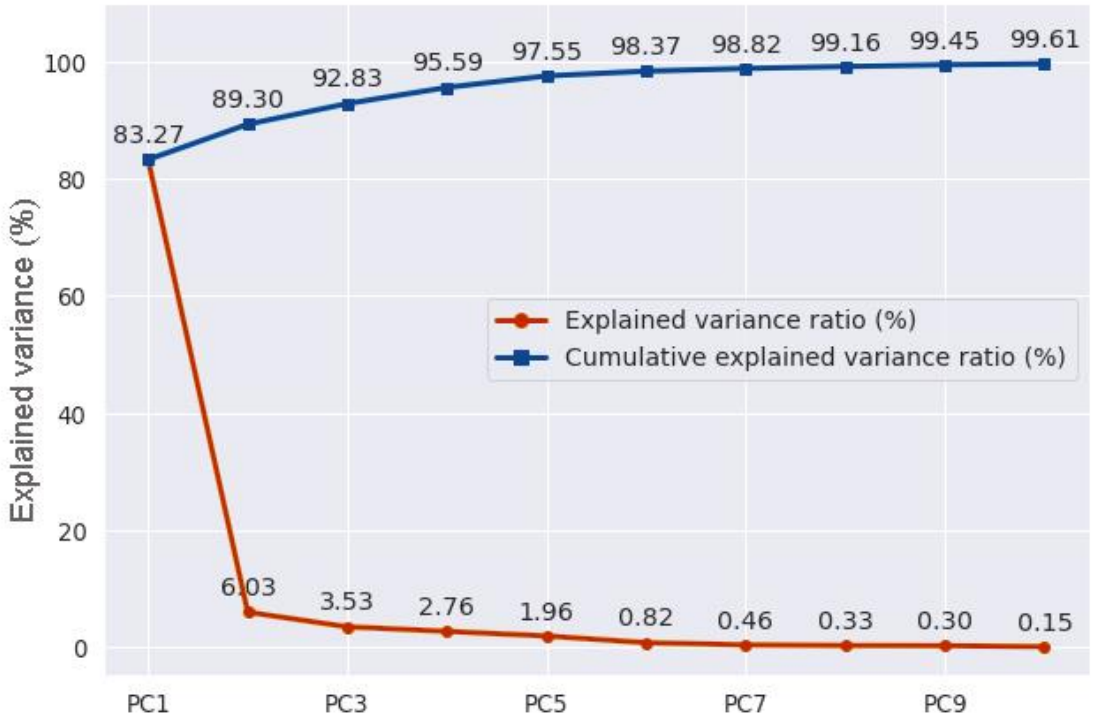


Figure 53. Scree plot of PCA. The explained variation in each PC and the cumulative explained variance (before vs. after physical exercise dataset).

The score scatter plots among PC1, PC2, and PC3 reveal clear patterns of discrimination between the two groups (before and after exercise) (Fig. 54). It is also depicted the data distribution within each principal component, providing a more comprehensive view of the underlying variability. The score plots indicate that the groups are well-separated, with minimal overlap indicating robust discrimination, suggesting distinct biochemical profiles associated with the pre- and post-exercise states. Notably, PC2 appears to be the primary contributor to this discrimination, as it shows the most significant separation between the groups compared to PC1 and PC3. This finding underscores the importance of the biochemical features represented by PC2 in capturing the metabolic changes induced by physical activity. The inclusion of data distributions within the plots enhances the clarity of these results, visually confirming the distinct clustering of samples on PC2 (Fig. 54) and supporting the robustness of the PCA model in differentiating the two groups. These results demonstrate the efficacy of ATR-FTIR

spectroscopy combined with multivariate analysis for detecting and characterizing exercise-induced biochemical changes in saliva.

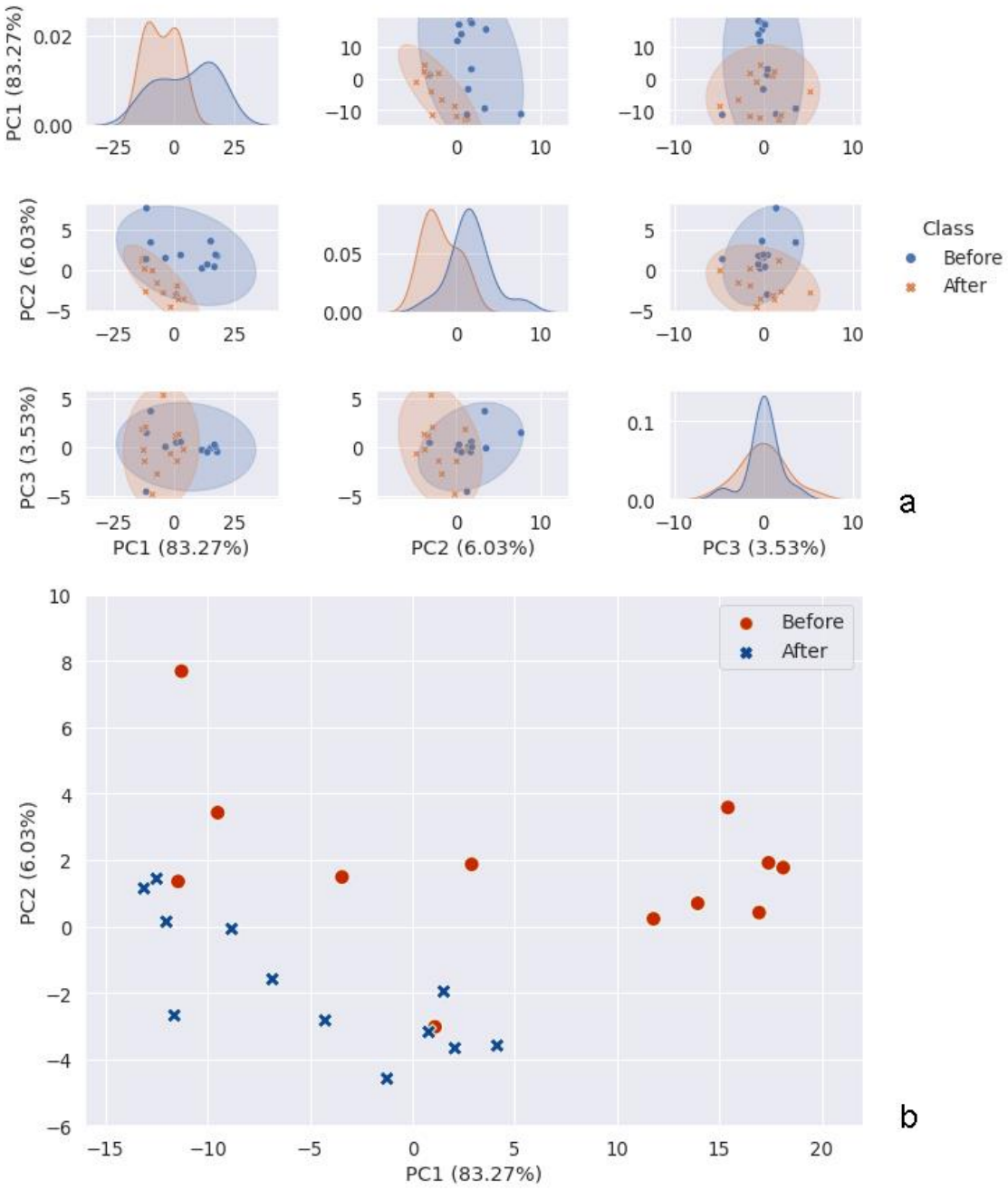


Figure 54. (a) Score scatter plots of PCA among the first three PCs and distribution graphs of data on each component. (b) Score plot of PC1-PC2 highlighting the groups' discrimination (before vs. after physical exercise dataset).

The loading plots for PC1, PC2, and PC3 provide insights into the specific spectral features driving the discrimination between pre- and post-exercise salivary samples (Fig. 55). These plots highlight the wavenumbers most strongly contributing to the variance captured by each principal component, offering a deeper understanding of the biochemical changes associated with physical exercise.

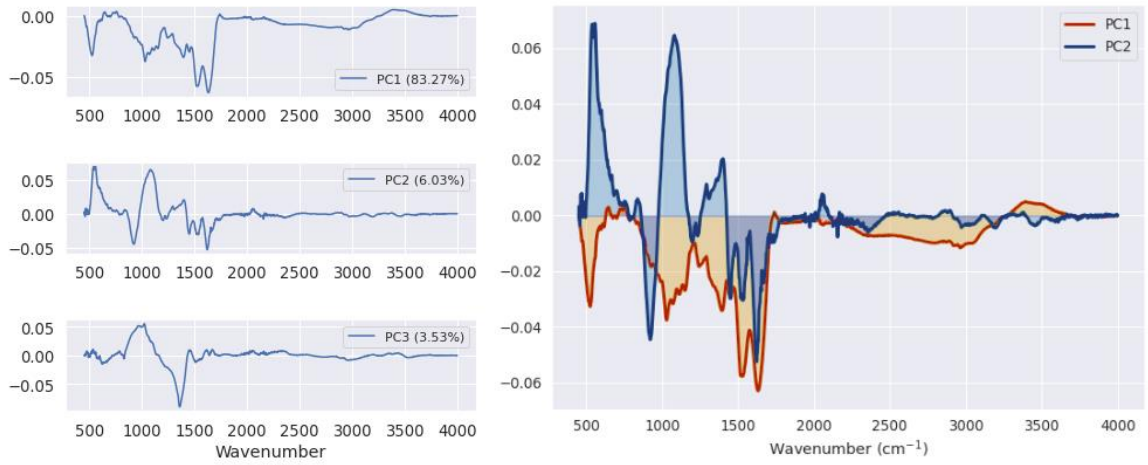


Figure 55. Loading plots of the first three principal components (before vs. after physical exercise dataset).

PC1, which explains the majority of the variance, is primarily influenced by wavenumbers associated with general biochemical content in saliva. Key contributors include the amide I (1631 cm^{-1}, C=O stretching vibration) and amide II (1527 cm^{-1}, N–H bending vibration) regions, indicative of proteins and their secondary structures[303,320]. The high loadings in these regions suggest significant protein-related changes, reflecting increased protein turnover or secretion post-exercise. Medium-intensity bands in the infrared spectra also play a significant role, contributing to the overall variance captured by this principal component. Specifically, the broad band at 982–1207 cm^{-1}, associated with glucose, glycogen, and sugar moieties (C–C and C–O stretching vibrations), highlights the involvement of carbohydrate-related biochemical changes[320,321]. These peaks indicate

shifts in energy metabolism, such as glycogen breakdown and glucose utilization, which are essential metabolic responses to physical activity. Additionally, the band at 1395 cm^{-1}, attributed to carboxylate groups (COO^- stretching vibrations), is indicative of molecules like lactic acid or carboxyl side chains of salivary proteins[321]. This peak suggests an alteration in lactic acid levels, a well-known byproduct of anaerobic metabolism during exercise[58].

For PC2, the primary classification between pre- and post-exercise samples is driven by specific spectral bands that highlight key biochemical changes. Notably, the band at 622 cm^{-1}, associated with the amide I region, reflects alterations in protein secondary structure[322], potentially indicative of changes in salivary protein dynamics due to exercise-induced stress or metabolic activity. The band at 921 cm^{-1}, linked to membrane lipids/phospholipids and carbohydrates[320], suggests a metabolic shift involving lipid mobilization or structural changes in carbohydrate-linked components[323]. These changes may correspond to enhanced energy demands and cellular membrane activity during physical exertion. A particularly strong and wide band at 1080 cm^{-1}, attributed to sugar moieties[320,321], shows the most significant contribution to the discrimination along PC2. This peak highlights changes in glycosylation patterns or variations in glucose-related metabolites, underscoring the critical role of carbohydrate metabolism in separating pre- and post-exercise states. The prominent influence of these bands on PC2 reinforces its role as a key driver of group discrimination, reflecting targeted biochemical adaptations to physical activity. These findings emphasize the sensitivity of ATR-FTIR spectroscopy in detecting nuanced molecular responses to exercise.

The strong peak at 1080 cm^{-1}, associated with sugar moieties, provides direct evidence of the critical role that carbohydrates play in energy metabolism during physical exercise[324]. Simple carbohydrates, such as glucose, are rapidly absorbed and utilized by muscles for immediate energy, while more complex carbohydrates are gradually broken down into glucose, fructose, and galactose, offering a sustained energy release[325]. Excess glucose is stored as glycogen in the liver and muscles, serving as a readily available energy

reserve, particularly during short, high-intensity activities like sprinting[326]. During exercise, glycogen stores are converted back into glucose to meet the body's heightened energy demands[326]. Post-exercise, the depletion of blood glucose levels often leads to fatigue[326], further emphasizing the central role of carbohydrate metabolism in sustaining physical activity. The prominent spectral band at 1080 cm^{-1} highlights the metabolic activity involving sugar moieties and glycogen mobilization, confirming these physiological processes.

In general, the analysis of the principal components suggests distinct biochemical roles for PC1 and PC2 in differentiating before and after exercise salivary samples. PC1 appears to be primarily associated with variables related to proteins, as evidenced by the strong contributions from the amide I and II regions. In contrast, PC2 is more strongly correlated with sugar moieties and carbohydrates, highlighted by prominent peaks such as the band at 1080 cm^{-1}. Together, these components provide complementary insights into the molecular adaptations underpinning physical exertion.

Overall, the use of ATR-FTIR provided a rapid and reliable analytical approach to capture molecular fingerprints, enabling the identification of exercise-induced metabolic alterations. This innovative approach sought to establish saliva as a practical and effective medium for real-time biomarker assessment in sports science and exercise physiology.

16.2 Biochemical adaptations aid by training habits: high- *vs.* low-level athletes discrimination

The findings from the before and after exercise discrimination highlighted the significant biochemical changes induced by physical activity, demonstrating the utility of ATR-FTIR spectroscopy for tracking metabolic responses. Building on this foundation, the study progressed to examine how these biochemical markers vary between athletes of different training levels.

While the pre- and post- exercise analysis focused on acute metabolic changes due to physical exertion, the comparison between high-level and low-level athletes aimed to explore long-term adaptations driven by training intensity and frequency. The intense and frequent physical exercise is a defining characteristic of high-level athletes, as it drives the physiological and metabolic adaptations required for enhanced performance. Consistent, structured training leads to significant improvements in cardiovascular and metabolic health by enhancing cardiac output, improving insulin sensitivity, and promoting favorable changes in blood lipid profiles[327]. Additionally, increasing training frequency, even with the same total load, can enhance muscle and bone adaptations by maintaining a more consistent positive protein balance and more frequent training stimuli[328]. Higher exercise frequency may improve recovery and reduce fatigue, and potentially optimize motor skill learning[328]. These adaptations highlights a broader perspective on the importance of sustained physical activity in shaping metabolic profiles, providing insights into the physiological differences that distinguish highly trained professionals from recreational or low-level athletes. Understanding these distinctions not only enhances our comprehension of training adaptations but also underscores the societal and scientific relevance of promoting structured physical activity.

16.2.1 Questionnaire results: distinguishing low- and high-level athletes

Although a definition about novice, intermediate, advanced and elite athletes has been defined[329], it is challenging to practically differentiate among them. The questionnaire results revealed distinct differences between low-level and high-level athletes, emphasizing the contrasting approaches to training and athletic development. Figure 56 presents the results of a questionnaire designed to differentiate athletes into low- and high-level cohorts based on their training habits, affiliations, and motivations.

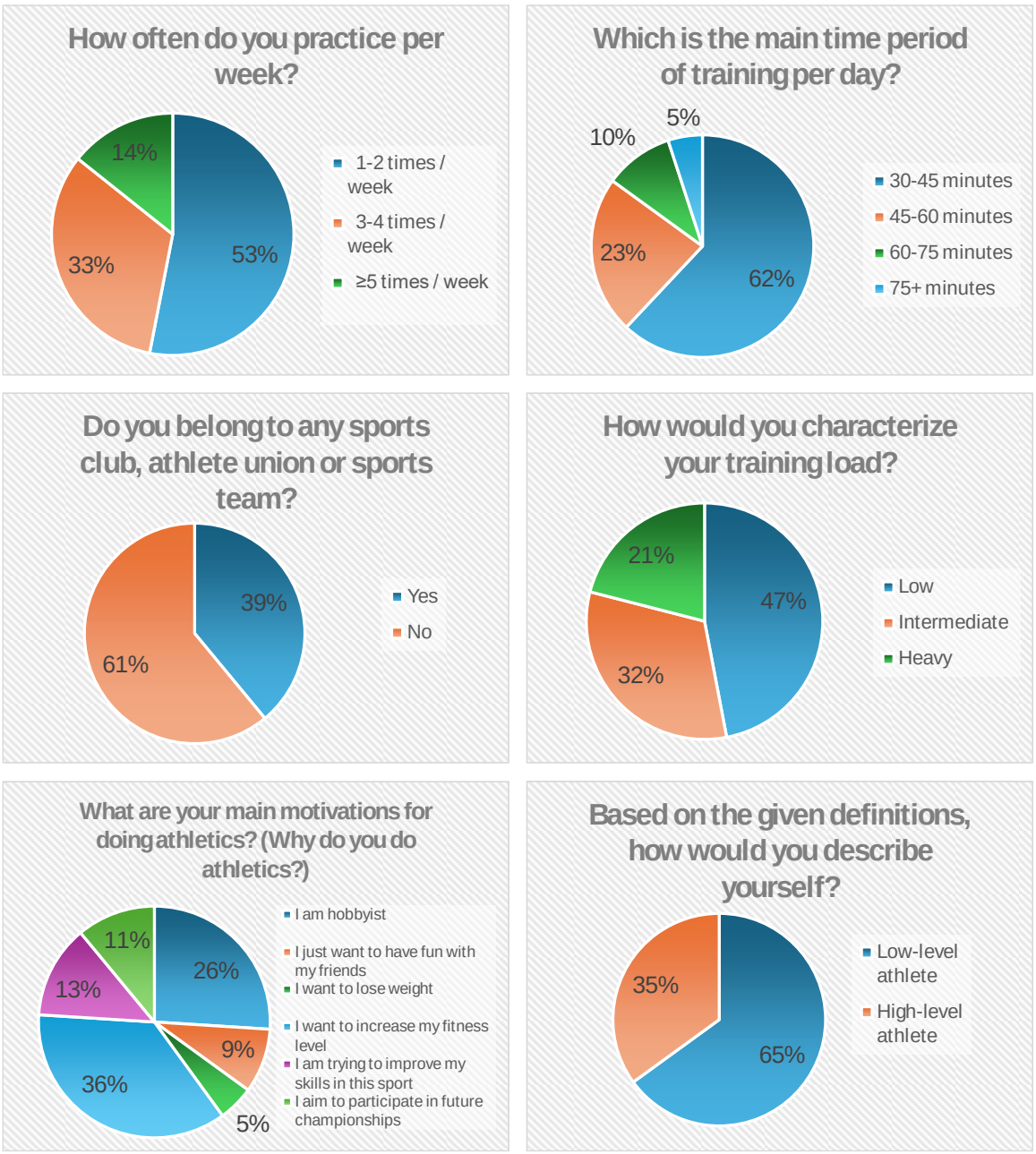


Figure 56. Questionnaire outcomes.

The majority of respondents practice 1–2 times per week, with fewer training 3–4 times or more than five times weekly, suggesting a predominance of moderate engagement in athletic activities. Most participants also report training durations of 30–

45 minutes per session, while fewer exceed 45 minutes. This highlights a tendency toward shorter training sessions, typical of recreational or fitness-focused athletes.

When asked about affiliation with sports organizations, 61% indicated they are not members of a club, team, or union. This lack of formal affiliation supports the idea that most respondents might be participating at a recreational rather than competitive level. Similarly, nearly half of the respondents characterize their training load as low, with fewer describing it as intermediate or heavy, further indicating a lighter approach to athletic participation. The motivations for engaging in athletics reveal that most participants aim to increase fitness levels or engage in the activity as a hobby. Performance-driven motivations, such as skill improvement or preparing for championships, were less common, suggesting that competitive goals are not a primary factor for most respondents.

Finally, when asked to self-identify based on predefined criteria, 65% described themselves as low-level athletes, with only 35% identifying as high-level athletes. This self-assessment aligns with the observed training patterns and motivations, reinforcing the predominance of recreational athletes in the sample. The inclusion of professional coaching guidance as a criterion further validated the distinction between the groups, ensuring a reliable categorization based on both objective training metrics and expert input.

Overall, the data provides clear distinctions between low- and high-level athletes. The low-level group is characterized by lighter training loads, shorter sessions, and fitness or recreational motivations, while the high-level cohort exhibits more intense training regimens, competitive aspirations, and greater organizational affiliation. Thus, based on the questionnaire results and the recommendations from professional coaches, a total of 17 athletes were classified as high-level, while 39 athletes were categorized as low-level. This classification was based on factors such as training frequency, intensity, coaching involvement, and the athletes' overall commitment to their training regimens, ensuring a

clear distinction between the two groups for further analysis. This analysis supports the classification criteria for cohort discrimination.

16.2.2 Discrimination between groups – PCA

Having classified the athletes into high-level and low-level groups based on the questionnaire results and coach recommendations, the next step was to explore the biochemical differences between these groups. To achieve this, we applied multivariate analysis techniques, specifically PCA and PLS-DA, to identify and visualize the key metabolic variations between the two groups. By examining the spectral data from both groups, we aimed to uncover underlying patterns and biochemical markers that distinguish high-level athletes from their low-level counterparts.

The mean absorption spectra for low-level and high-level athletes, along with their standard deviations, provide an overview of the average biochemical profiles for the two groups (Fig. 57). This visualization illustrates the general consistency of the spectral data within each group while highlighting overall trends in absorption intensities across the measured wavenumber range. The spectral regions, including those associated with proteins (amide I and II regions), carbohydrates (sugar moieties), lipids, and lactate-related components, display consistent patterns within each group. The standard deviation bands further emphasize the reliability of the measurements, showing minimal variation around the mean, which supports the robustness of the data. These spectra serve as a foundational comparison for the subsequent PCA analysis, which delves deeper into the specific wavenumbers contributing to the observed differences between low-level and high-level athletes.

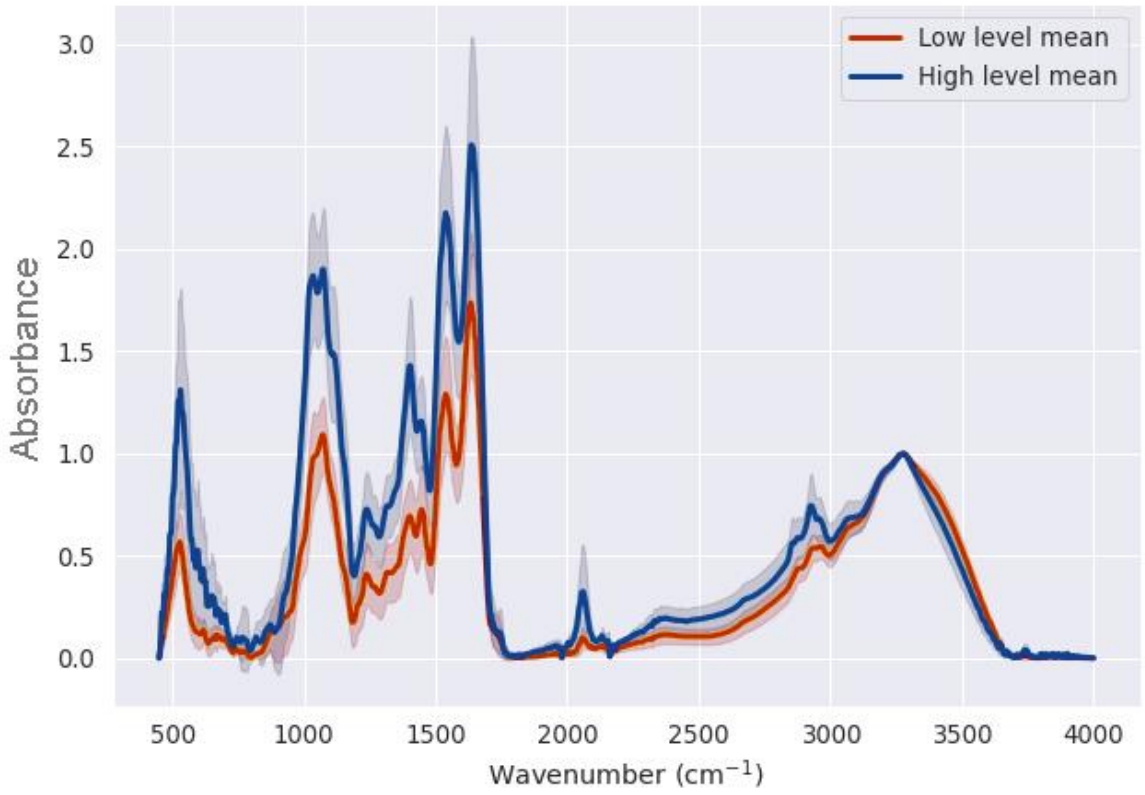


Figure 57. Mean FTIR spectra with SDs of high and low fitness level athletes.

In this Phase of the study, a total of 56 athletes participated, comprising 39 low-level and 17 high-level athletes. The dataset was divided into two subsets: a training set containing data from 42 athletes (29 low-level and 13 high-level) and a test set comprising 14 athletes (10 low-level and 4 high-level). For the purposes of PCA and later PLS-DA model construction, only the training dataset was utilized to ensure the development of a robust and unbiased predictive model. The remaining test dataset served as a validation set, acting as "unknown" data since it was not presented during the PLS model training. This separation ensures that the model's performance is evaluated on unseen data, reflecting its ability to generalize and accurately classify new samples.

The scree plot for the PCA of the low-level and high-level athletes dataset reveals the proportion of variance explained by each principal component (up to ten PCs). A

significant amount of variance is captured within the first few PCs, emphasizing their importance in distinguishing between the two groups. PC1 accounts for a dominant portion of the variance, explaining 75.13%, reflecting the most significant differences in the dataset related to athletic level of participants and associated metabolic adaptations (Fig. 58). PC2 contributes an additional 8.29%, increasing the cumulative explained variance to 83.41%. This indicates the supplemental role of the second principal component in capturing further variability between low-level and high-level athletes. PC3, PC4 and PC5 explain 5.75%, 3.56% and 2.21%, respectively, bringing the total variance explained by the first three PCs to 94.93%. The rest of components explain less than 1% each.

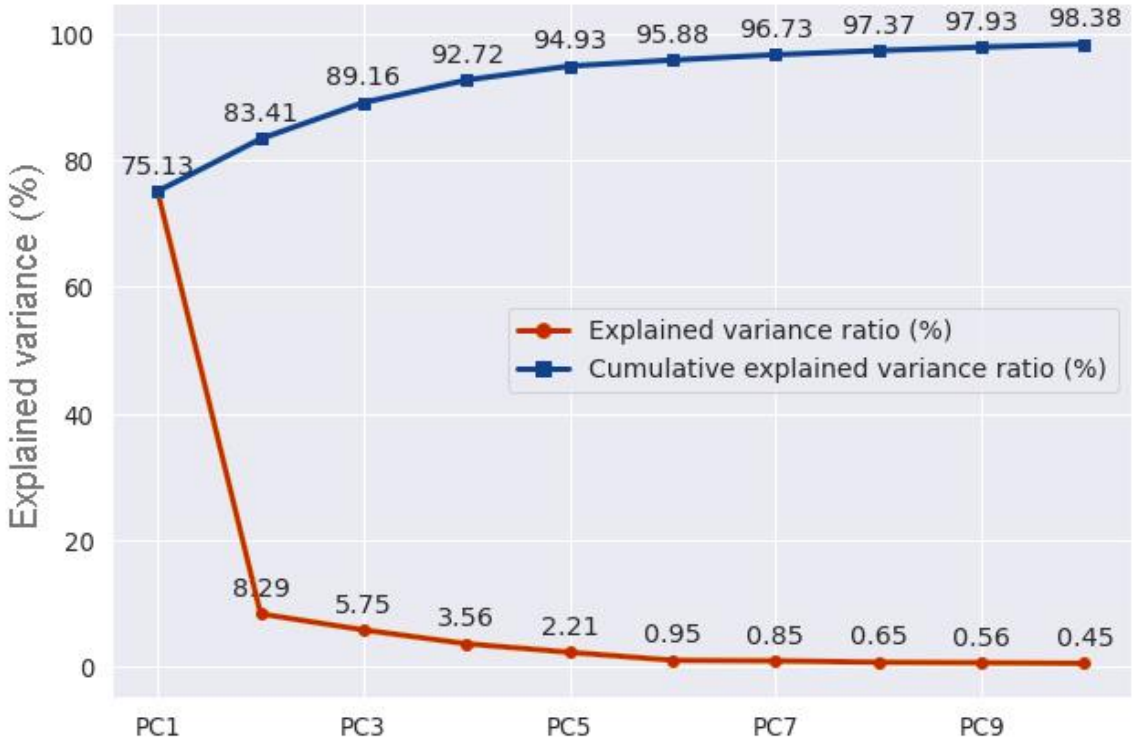


Figure 58. Scree plot of PCA. The explained variation in each PC and the cumulative explained variance (low-level vs. high-level athletes dataset).

The scree plot displays an “elbow” at PC2, indicating a point where the explained variance levels off, and subsequent PCs contribute only marginally to the dataset. These results suggest that PC1 and PC2 effectively capture the critical biochemical and metabolic distinctions between the two groups, while PC3 provides additional nuance. This forms the foundation for subsequent score and loading plot analyses, which elucidate the specific features driving these separations.

The score plots among PC1, PC2, and PC3, alongside the corresponding distribution plots, provide insights into the biochemical differences between low-level and high-level athletes (Fig. 59). The analysis highlights PC1 as the dominant separator, while separation along PC2 and PC3 is not clearly visible. PC1, explaining 65% of the total variance, is the primary contributor to group separation. The score plot along PC1 shows distinct clustering of high-level and low-level athletes, with minimal overlap between the groups. This indicates that PC1 effectively captures the key biochemical differences associated with athletic level and long-term physical adaptations. The distribution plot for PC1 further reinforces this separation, with two well-defined and minimally intersecting peaks representing each group. PC2 and PC3, explaining 20% and 6% of the variance, respectively, do not show clear separation between the groups. The score plots along PC2 and PC3 display significant overlap, suggesting that these components primarily capture individual variability rather than features distinguishing the two groups. While PC2 and PC3 add some nuance to the overall dataset, they do not contribute significantly to the discrimination between high-level and low-level athletes, emphasizing the importance of PC1 in distinguishing the metabolic profiles of the two groups.

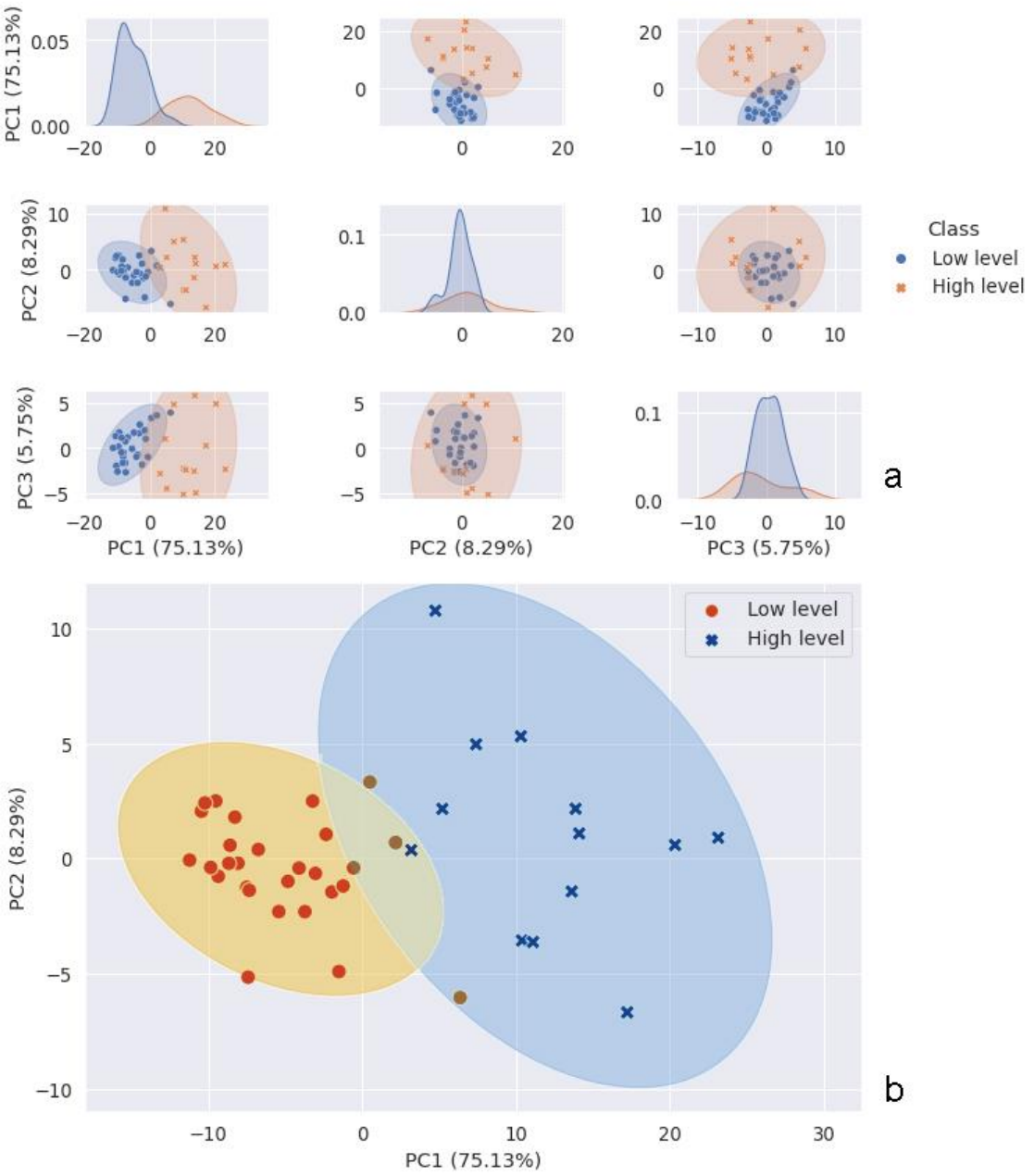


Figure 59. (a) Score scatter plots of PCA among the first three PCs and distribution graphs of data on each component. (b) Score plot of PC1-PC2 highlighting the groups' discrimination (low-level vs. high-level athletes dataset).

The loading plots provide crucial insights into the spectral features driving the discrimination between low-level and high-level athletes (Fig. 60). These plots highlight

the specific wavenumbers and their associated biochemical components that contribute most significantly to the observed differences in metabolic profiles. While PC1 serves as the primary axis of separation, PC2 provides complementary information, though it plays a secondary role in the discrimination. PC3, which explains a minor portion of the total variance, does not contribute significantly to the separation.

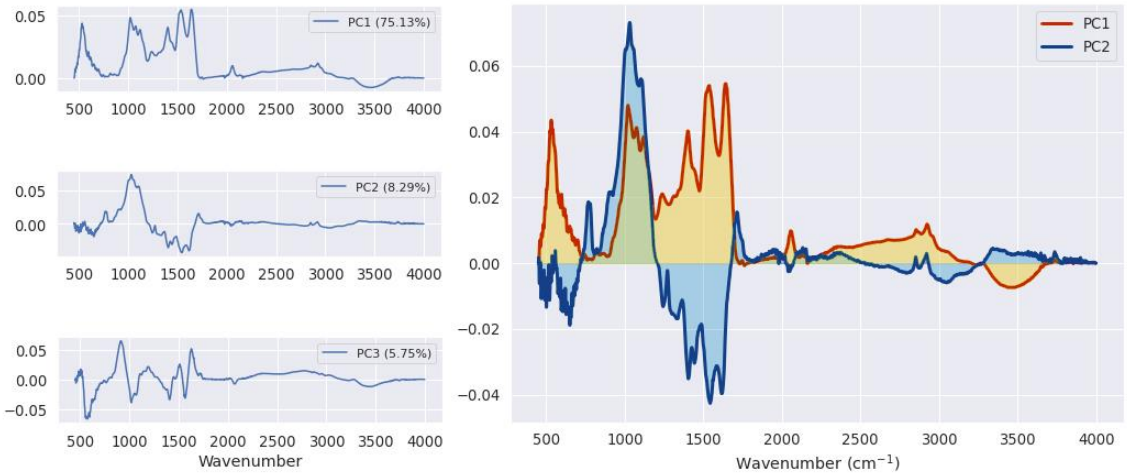


Figure 60. Loading plots of the first three principal components (low-level vs. high-level athletes dataset).

PC1, explaining 75.13% of the variance, reveals the dominant biochemical variables responsible for distinguishing the two groups. Strong spectral bands identified in the loading plot include:

- 992–1196 cm^{-1}: This broad band is associated with glucose, glycogen, and sugar moieties, representing C–C and C–O stretching vibrations[320,321]. These features indicate differences in carbohydrate metabolism between the two groups.
- 1529 cm^{-1} (amide II): Attributed to N–H bending vibrations[303,320], this band reflects changes in salivary protein content, likely due to differences in protein turnover or secretion influenced by training intensity.

- 1642 cm^{-1} (amide I): Corresponding to C=O stretching vibrations[303,320], this band further underscores the role of proteins in differentiating the groups.

Additionally, the peak at 1392 cm^{-1}, attributed to carboxylate groups (COO^- stretching vibrations), represents molecules such as lactic acid[321]. The presence of this band highlights the role of lactate metabolism in high-level athletes, who often engage in lactate threshold training. This training promotes the production of proteins that aid in the absorption and conversion of lactic acid into energy[330], emphasizing the metabolic adaptations driven by higher exercise intensity and frequency.

The loading plot of PC1 suggests that the primary separation between low-level and high-level athletes arises from proteins (amide I and II regions) and carbohydrate-related peaks, reflecting energy utilization and recovery processes linked to structured training. Proteins (the main discrimination variable on PC1 axis) can also be broken down and used as a last resort of energy, although this stresses the kidneys and restricts the body's ability to build and maintain muscle tissue[331,332].

PC2, explaining an additional portion of the variance, complements PC1 by highlighting secondary biochemical features. Bands at 1039 cm^{-1} (sugar moiety-related band) reinforcing the role of carbohydrate metabolism in group differentiation and 1543 cm^{-1} (amide II) and 1619 cm^{-1} (amide I) emphasizing structural differences in protein metabolism between the groups are shared in both PCs. The 1401 cm^{-1} band, attributed to carboxylate groups, is common to both PCs, further supporting the role of lactic acid and related metabolites in distinguishing between the two groups. Lastly, a medium-intensity band at 1446 cm^{-1}, associated with CH_2 bending vibrations in lipids or amines, appears only in PC2, suggesting its relevance to lipid dynamics in the metabolic profiles.

The loading plots confirm that the discrimination between low-level and high-level athletes is driven primarily by metabolic markers of protein turnover, carbohydrate metabolism, and lactate dynamics. PC1 emphasizes proteins and sugar moieties as the dominant features, reflecting energy utilization and recovery processes, while PC2 adds

depth by highlighting lipid-related components and additional protein features. Together, these components illustrate the biochemical adaptations associated with different levels of training intensity and frequency, providing a detailed molecular fingerprint of athletic performance.

16.2.3 Discrimination between groups – PLS-DA

PLS-DA was performed subsequent to PCA to enhance the discrimination between low-level and high-level athletes by focusing directly on the separation of predefined groups. While PCA is an unsupervised method that identifies patterns and reduces dimensionality based solely on variance, it does not explicitly consider group labels. In contrast, PLS-DA is a supervised method that incorporates group membership into the analysis, maximizing the variance related to class separation. This makes PLS-DA particularly well-suited for identifying and visualizing the biochemical features most strongly associated with the differences between low-level and high-level athletes.

From the 56 collected spectra (56 participated athletes), 42 were randomly selected in a stratified manner and used for PCA. Subsequently, PLS-DA was performed on the same set of samples to create a predictive model capable of assessing the performance level of “unknown” samples. For this predictive model, the PCA-tested samples served as the training dataset, while the remaining 14 “unknown” spectra (10 belonging to low-level athletes and 4 belonging to high-level athletes) were used as the test dataset.

Cross-validation was performed as part of the PLS-DA to determine the optimal number of latent variables (LVs) required for effective discrimination between low-level and high-level athletes. This step is crucial to balance model complexity and performance, ensuring that the PLS-DA model captures meaningful patterns without overfitting. A 3-fold repeated stratified cross-validation (30 repeats) was carried out on the training dataset to identify the optimal number of latent variables (LVs). The root mean squared

error of cross-validation (RMSECV) and prediction (RMSEP) were plotted against the number of components, as shown in Figure 61. This analysis revealed that the lowest RMSE values were observed at two components, indicating that a model with two latent variables provides the best predictive performance while minimizing the risk of overfitting. The RMSECV curve showed stabilization and then error increment beyond two components, suggesting that increasing the number of components would not significantly enhance prediction accuracy but might lead to model overfitting. Thus, the optimal PLS-DA model was constructed using two components, balancing prediction accuracy and robustness.

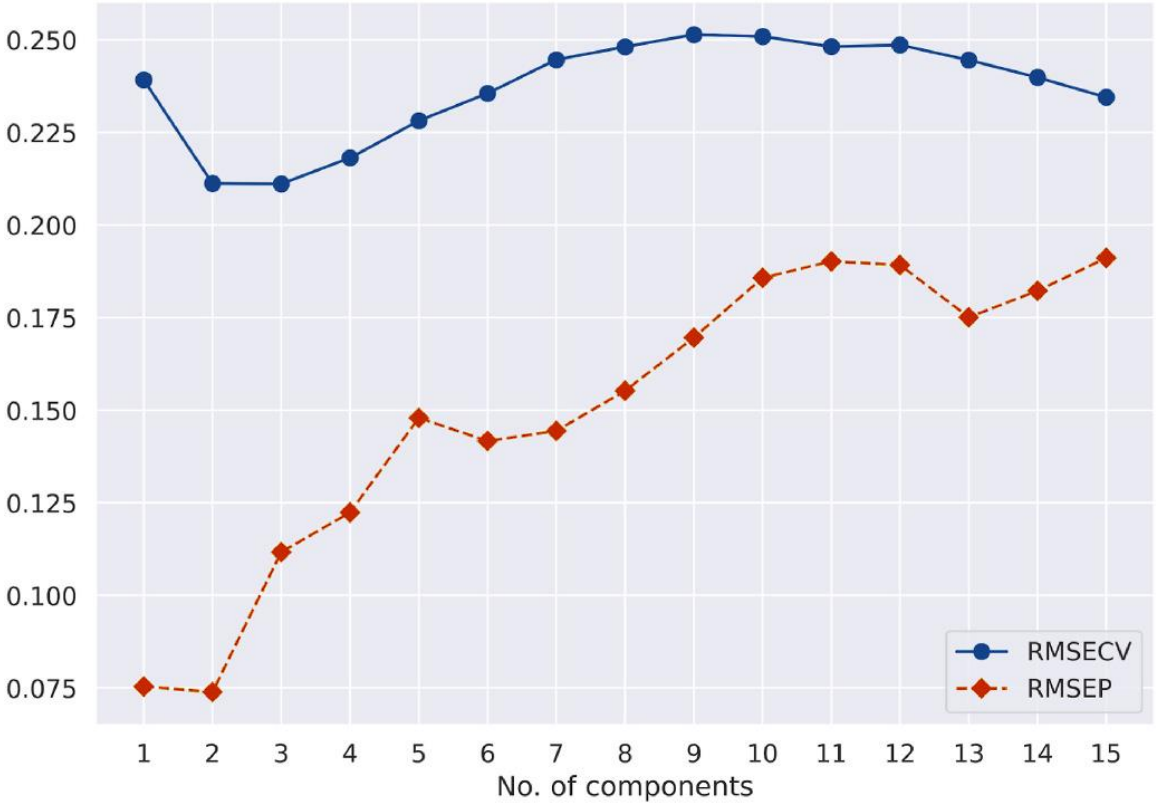


Figure 61. RMSE of cross-validation (RMSECV) and prediction (RMSEP) indicating the number of components that should be used for the predictive PLS model.

The cross-validation results emphasize the robustness of the PLS-DA model and its ability to generalize well to unseen data. By selecting the optimal number of LVs, the analysis ensured that the model was neither underfitted (missing key patterns) nor overfitted (capturing noise), laying a strong foundation for subsequent interpretation of the PLS-DA loadings and score plots.

The PLS-DA score plot demonstrated an excellent separation between the low-level and high-level athlete groups (Fig. 62). The two clusters formed by their respective spectra were distinctly separated along the first latent variable (LV1), which accounted for the majority of the variance linked to the performance level differentiation. This clear segregation indicates that the model effectively captured the underlying biochemical differences between the two groups. Such results highlight the effectiveness of ATR-FTIR spectroscopy in combination with PLS-DA to differentiate between athlete performance levels based on their biochemical salivary profiles.

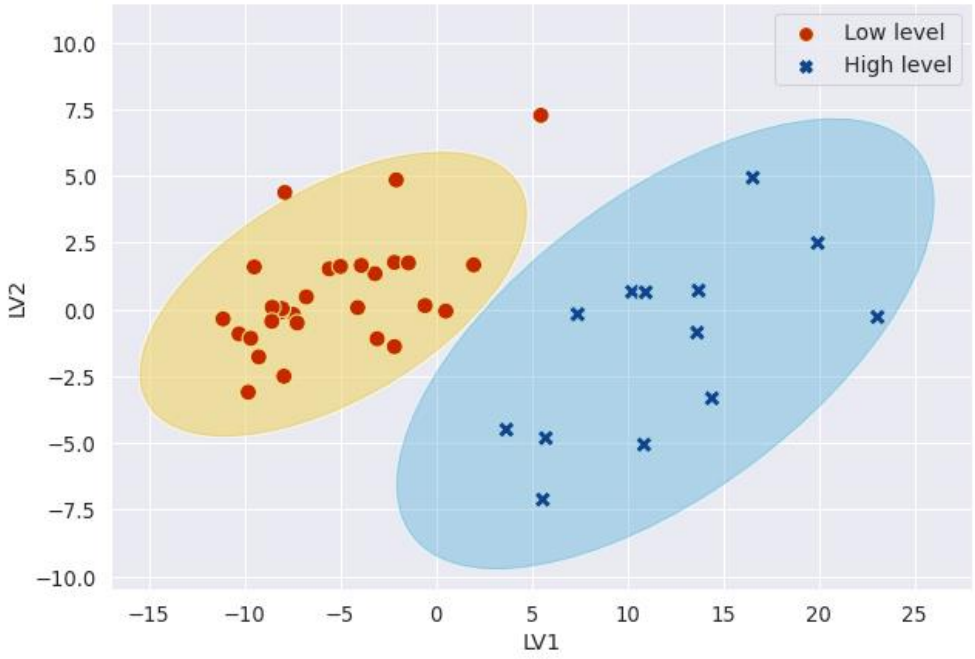


Figure 62. Scores of athletes plotted in LV1 against LV2 and highlighting the groups' discrimination in PLS analysis.

The performance of the PLS-DA model was assessed through multiple metrics, including the confusion matrix, ROC curve, and classification performance indicators. The confusion matrix, shown in Figure 63a, reveals the classification accuracy of the PLS-DA model. As mentioned previously, the remain test dataset was utilized to evaluate the model's discrimination ability. This dataset, having 10 low-level and 4 high-level athletes' spectral data, acts as an "unknown" because it was not included in the building of the model. Thus, the matrix shows that the model successfully predicted 9 low-level athletes and 4 high-level athletes correctly, while there was only 1 misclassification of a low-level athlete as a high-level athlete. This indicates that the model demonstrated high classification accuracy (93%), especially for distinguishing between low-level and high-level athletes. The ROC curve, shown in Figure 63b, further confirms the excellent predictive performance of the PLS-DA model. The decided threshold indicates a near-perfect separation between the two athlete groups, with a true positive rate (sensitivity, 1.0) of 1.0 and a false positive rate (1-specificity, 1-0.9) of 0.1. The Area Under the Curve (AUC) value of 0.95 indicates a strong ability of the model to correctly classify the athletes. AUC values near 1 signify an outstanding predictive model, reinforcing the conclusion from the confusion matrix that the PLS-DA model performed with a high degree of accuracy and robustness. Together, the confusion matrix and ROC curve illustrate the exceptional classification ability of the PLS-DA model in distinguishing between low-level and high-level athletes, providing strong evidence for the efficacy of this approach in athlete performance level classification based on biochemical markers measured in saliva.

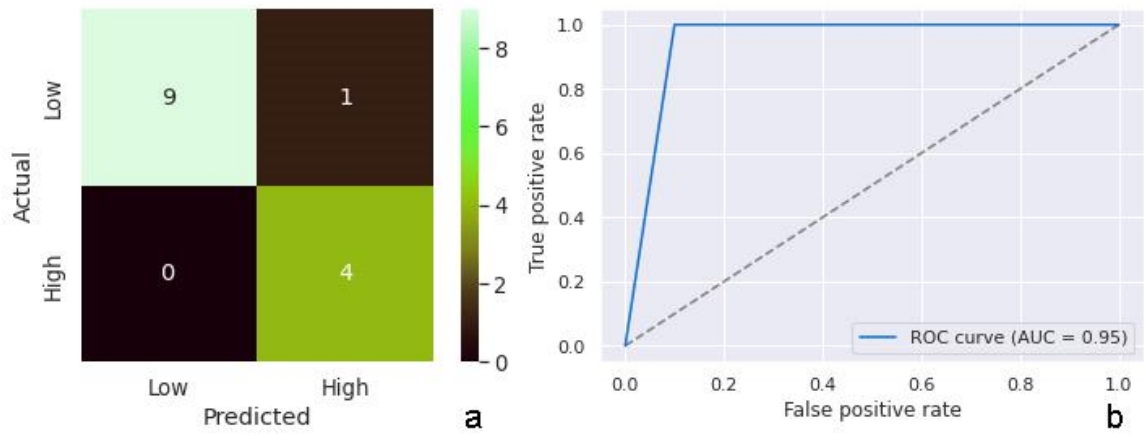


Figure 63. (a) Confusion matrix of the predictive PLS model and (b) ROC curve indicating sensitivity and specificity of the model.

Following this evaluation, PLS-DA was performed on the entire dataset (56 athletes) using the described model. When the predictive model was applied to the full set of collected data, the performance metrics improved, with a sensitivity of 1.0, a specificity of 0.94, and an overall accuracy of 98%. These results highlight the reliability and robustness of the constructed PLS-DA model in discriminating between low-level and high-level athletes.

The PLS-DA loading plot reveals key spectral features that contribute to the discrimination between low-level and high-level athletes (Fig. 64). As seen in the loading plot, the peaks associated with the discrimination are largely consistent with those identified in the PCA analysis, including bands related to glucose, glycogen, and sugar moieties, as well as protein-related bands (amide I and II). These features continue to play a major role in differentiating the two groups, consistent with the previous PCA results.

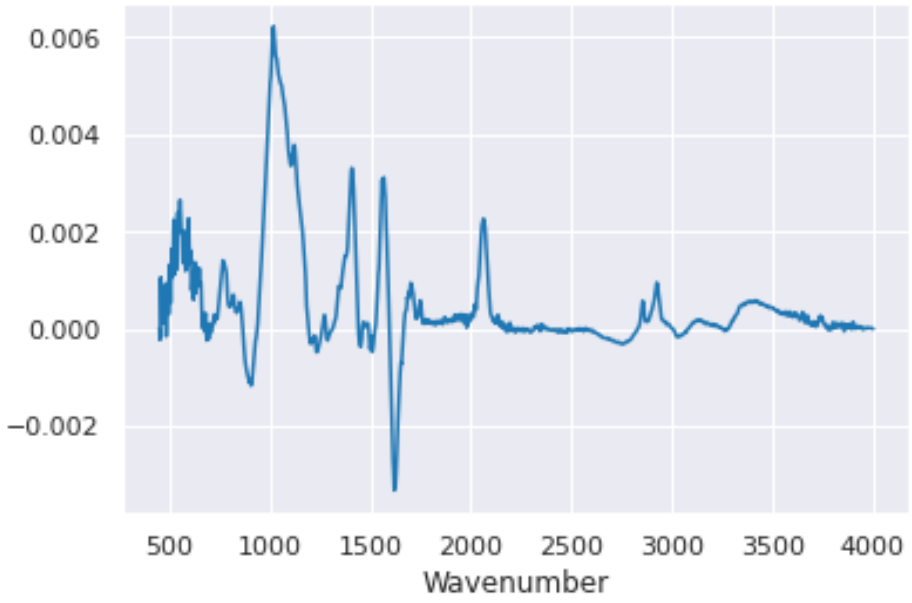


Figure 64. PLS coefficients plot of the first latent variable (LV1).

However, in the PLS-DA analysis, a new spectral feature emerges. A peak at 2060 cm^{-1}, attributed to thiocyanates[320,321,333–335], is observed, although it is less intense compared to the other spectral bands. Thiocyanates, which are known to be associated with various metabolic processes, appear to contribute to the discrimination between low-level and high-level athletes, providing additional information that enhances the classification model.

Thiocyanate (SCN^-) is a significant ion present in human saliva, arising from both endogenous and exogenous sources. Endogenously, it is a product of the metabolism of sulfur-containing compounds, such as glucosinolates found in cruciferous vegetables[336,337]. Exogenously, thiocyanate is notably characteristic of smokers' saliva, as it results from the metabolic conversion of cyanide compounds present in tobacco smoke[338]. While its elevated concentration in smokers makes it a valuable biomarker for smoking-related studies, thiocyanate is also naturally present in non-smokers, albeit at lower levels[293]. In saliva, typical concentrations range from 0.5 to 3 mM, with higher levels

observed in individuals exposed to tobacco smoke[339,340]. Thiocyanate plays a vital role in the antimicrobial defense system of saliva through its involvement in the lactoperoxidase system[341]. However, excessive thiocyanate levels, particularly in smokers, may interfere with iodine uptake by the thyroid gland, highlighting the dual significance of this molecule in both health and disease contexts[342].

This new feature suggests that thiocyanate levels may serve as a supplementary biomarker in distinguishing athletes based on their training intensity or performance levels (more details are presented in Phase 3, *Section 18*). This addition to the discrimination model underscores the utility of PLS-DA in identifying subtle, yet informative, spectral features that may not have been apparent in the simpler PCA model.

Phase 1 represents the first study to employ ATR-FTIR spectroscopy combined with chemometric analysis (PCA and PLS-DA) to discriminate salivary biochemical profiles between low- and high-level athletes after physical exercise. While previous studies had demonstrated that FTIR can capture metabolic shifts in saliva due to exertion, none had specifically applied this approach to stratify athletes by fitness level using a single post-exercise saliva sample. This Phase introduced a viable novelty: Salivary infrared spectra were linked not only to rest/exercise states but also to training status, revealing that athletes with different fitness levels exhibit distinguishable salivary biochemical fingerprints after exertion.

In resume, Phase 1 successfully demonstrated the potential of ATR-FTIR spectroscopy coupled with multivariate analysis to distinguish biochemical changes in saliva related to physical exercise and athlete performance levels. Through PCA and PLS-DA, it was shown that saliva can act as a valuable non-invasive biomonitoring tool, reflecting significant differences in the molecular composition before and after physical exercise, as well as between low-level and high-level athletes. The findings revealed that proteins and sugar moieties were the major contributors to these discriminations, with additional insights from thiocyanates during PLS-DA. Moreover, the high sensitivity, specificity, and accuracy of the constructed models underscore the reliability of this

approach. These results lay a solid foundation for further investigations into using saliva for athletic performance monitoring in sports science.

These findings suggest that salivary composition may reflect long-term physiological adaptations associated with training status. However, given that sample collection occurred post-exercise, it remains unclear to what extent these differences were influenced by the athletes' chronic fitness level versus the acute intensity of exertion at the time of sampling. This ambiguity highlighted the need for a more controlled approach to disentangle training-related biochemical signatures from those arising dynamically during exercise.

17. Phase 2 – Utilizing ATR-FTIR spectroscopy of saliva for monitoring and differentiating exercise intensity levels

In Phase 1, we observed distinct salivary profiles between low- and high-level athletes, likely reflecting the physiological adaptations associated with regular high-intensity training. However, because high-level athletes are habitually exposed to more intense physical workloads, it remained unclear whether the observed differences were solely due to training status or were also influenced by the intensity of effort at the time of sampling. To address this, we designed Phase 2 to investigate the acute effects of increasing exercise intensity on salivary composition in a controlled setting. This allowed us to distinguish between baseline training-related adaptations and the dynamic response of saliva to varying levels of physical exertion, thereby strengthening the interpretation of our findings from both phases.

In Phase 2, the experimental design was structured to monitor salivary biochemical changes under varying running intensities walking (4–5 km/h), jogging (9–10 km/h) and running (14–15 km/h). The primary aim was to assess whether ATR-FTIR spectroscopy could capture dynamic physiological adaptations during exercise, offering a non-invasive tool to evaluate athlete performance and recovery. This Phase involved collecting saliva samples at different stages of exercise, representing distinct running states, including low-intensity aerobic running, moderate-intensity steady-state running, and high-intensity anaerobic sprints.

Running intensity was chosen as the variable of interest because it plays a critical role in influencing biochemical processes within the body. Low-intensity aerobic running primarily utilizes fat metabolism, while moderate-intensity states rely on a mix of carbohydrate and fat metabolism[343]. High-intensity exercise shifts energy production toward anaerobic pathways, leading to increased lactic acid production and significant

biochemical changes, such as glycogen breakdown and elevated stress markers[58,343]. Monitoring these physiological changes offers valuable insights into energy utilization, metabolic efficiency, and recovery strategies.

Saliva was chosen as the biofluid of interest due to its non-invasive collection method, making it suitable for frequent sampling without disrupting exercise performance. Additionally, saliva contains biomarkers reflective of systemic metabolic states, such as glucose, proteins, and lactate, which are highly relevant to exercise physiology. This approach aimed to correlate salivary biochemical profiles with running intensities, enabling real-time monitoring of athlete performance and recovery during training sessions.

By integrating these running states into the experimental design, Phase 2 sought to validate ATR-FTIR spectroscopy as a versatile and practical tool for assessing athlete status across a wide spectrum of exercise conditions.

In Phase 2, saliva samples were collected during different running intensities from all participating athletes. A total of 128 spectra were initially recorded. However, as part of data quality control, spectra with excessive noise, poor signal-to-noise ratio, or technical issues during acquisition were excluded to ensure accurate analysis. After this preprocessing step, 125 spectra remained for further analysis, ensuring high-quality data was used (Fig. 65).

For multivariate analysis, the spectra were divided into two main groups: training dataset, used for model construction, accounting for 70% of the total spectra, and test dataset, consisting of the remaining 30%, was reserved for model validation to assess the predictive capability of the developed models. This separation ensures that the predictive models are trained on one dataset and validated on unseen data, simulating real-world scenarios. The division was performed in a stratified manner, maintaining an even distribution of samples across the various running states to avoid bias in the analysis. Such careful preprocessing and data splitting procedures were essential to ensure reliable and robust results in the subsequent multivariate analyses.

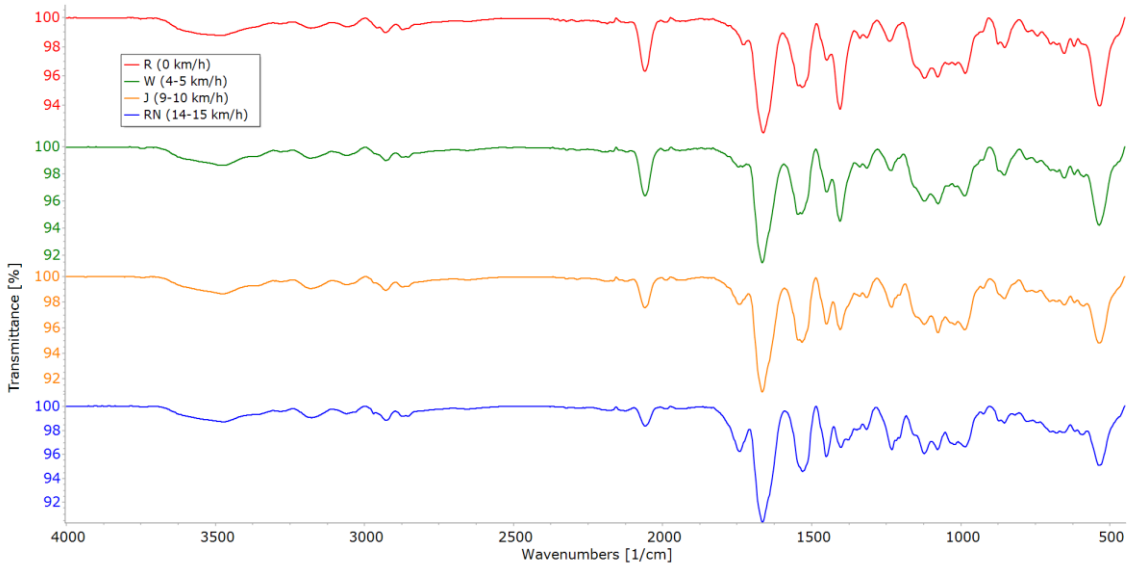


Figure 65. Mean FTIR spectra of saliva samples at rest and three different physical exercise intensities.

17.1 Chemometric discrimination of different physical exercise intensities

In Phase 2, PCA was initially performed to explore the feasibility of discriminating among the four exercise intensity groups—rest (0 km/h), walking (4–5 km/h), jogging (9–10 km/h), and running (14–15 km/h)—based on the salivary ATR-FTIR spectra. PCA, as an unsupervised dimensionality reduction technique, was applied to identify patterns, clusters, and relationships in the spectral dataset without incorporating prior class information. The primary goal was to determine whether the spectral profiles of saliva samples clustered according to exercise intensity.

By projecting the high-dimensional spectral data onto principal components (PCs), the analysis revealed that samples from different classes overlapped significantly (Fig. 66),

suggesting that individual variability was the primary factor driving sample positioning rather than the exercise intensity class. Consequently, PCA alone was insufficient to achieve the desired separation between groups. These results highlighted the limitations of unsupervised techniques in capturing the subtle biochemical differences associated with exercise intensity. This outcome necessitated the application of supervised classification methods, which are specifically designed to optimize discrimination based on predefined classes.

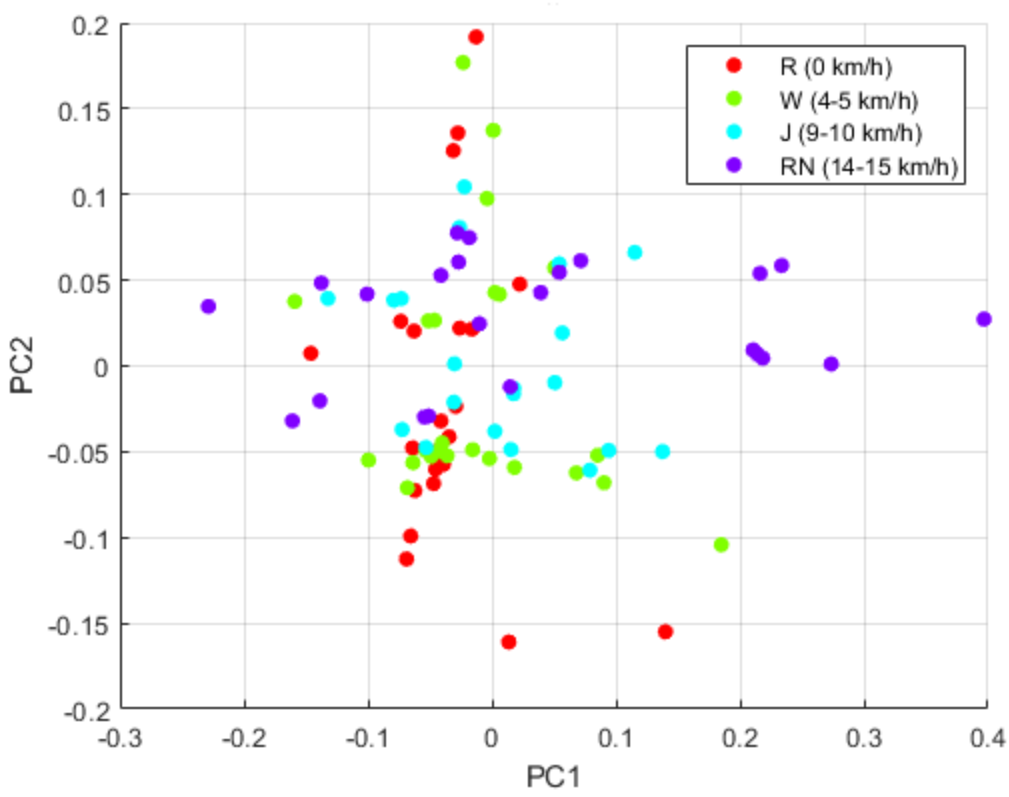


Figure 66. PCA scatter plot (PC1 and PC2) of saliva samples at different exercise intensities.

17.1.1 Model selection

To address the inefficiencies observed in the unsupervised PCA analysis, the study investigated whether pre-processing the ATR-FTIR spectra using Multiplicative Scatter Correction (MSC) would enhance the classification accuracy. MSC is a widely adopted pre-processing technique that minimizes spectral scattering effects caused by sample surface irregularities and instrumental artifacts, thereby improving data quality and interpretability[344]. Both PCA-Discriminant Analysis (PCA-DA) and Partial Least Squares-Discriminant Analysis (PLS-DA) were evaluated under two conditions: using MSC pre-processed spectra and non-pre-processed spectra.

The score scatter plot of PCA-DA without MSC (Fig. 67a) revealed poor clustering of samples from the four exercise intensity groups, indicating a lack of distinct separation among classes. The plot exhibited significant overlap among the groups, with no discernible patterns. This suggests that the variability in the spectral data was driven more by noise and scattering artifacts than by meaningful biochemical differences. Without MSC pre-processing, the spectral data retained inconsistencies caused by surface irregularities and instrumental effects, masking subtle distinctions necessary for effective classification. The inability of PCA-DA to differentiate between groups in this condition underscores the necessity of applying MSC to improve the interpretability and quality of spectroscopic data before analysis. The score scatter plot of PLS-DA without MSC (Fig. 67b) showed slightly better performance compared to PCA-DA, with some small clustering observed for the exercise intensity groups. However, the clustering was very poor, with significant overlap among samples from different groups (rest, walking, jogging, and running). The limited separation indicates that the model struggled to capture the subtle biochemical differences necessary for accurate classification. The lack of MSC pre-processing left scattering effects and instrumental artifacts uncorrected, which likely contributed to the poor clustering. Although PLS-DA performed marginally better than PCA-DA, the results highlight the insufficient quality of the raw spectral data

for reliable discrimination without pre-treatment. This reinforces the critical role of MSC in enhancing the robustness of supervised classification methods.

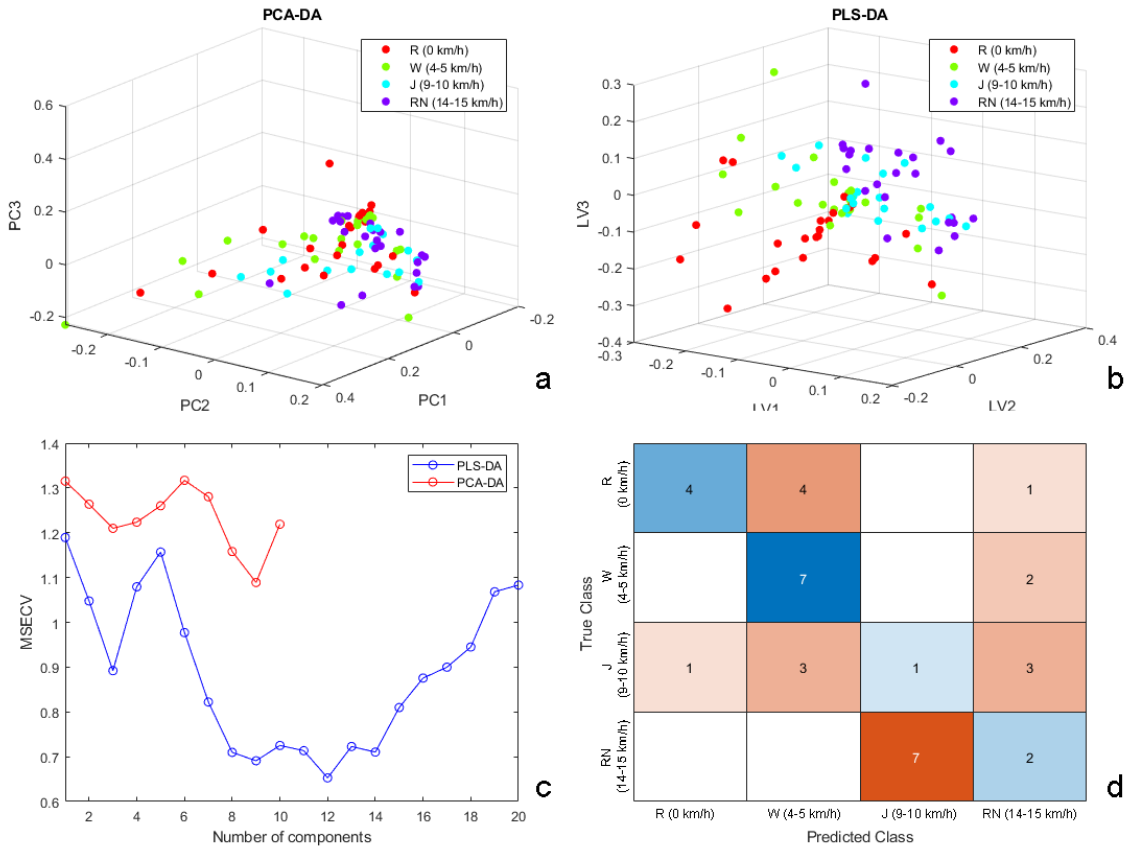


Figure 67. Multivariate analysis without MSC pre-processed data: (a) score scatter plot of PCA-DA model, (b) score scatter plot of PLS-DA model, (c) mean square error of cross-validation against the number of components for the two models, (d) confusion matrix of 12-LVs PLS-DA model.

The models were cross-validated and the process indicated that certain factors, likely stemming from uncorrected scattering effects and instrumental noise, contributed to an increased error amount in both models. In both cases, these factors were particularly evident in the higher mean square error of cross-validation (MSECV),

reflecting the model's reduced ability to generalize and accurately classify samples (Fig. 67c).

The confusion matrix for the PLS-DA model (without MSC) reveals notable misclassifications across the four exercise intensity groups: rest (R), walking (W), jogging (J), and running (RN) (Fig. 67d). For the rest group, only a portion of the samples were correctly classified, with others misclassified predominantly as walking or running. The walking group showed relatively better performance, with most samples correctly identified, though some were misclassified as running. The jogging group exhibited the poorest classification, with only a single sample correctly identified, and most being misclassified. For the running group, a majority of samples were incorrectly predicted. These misclassifications highlight the difficulty of distinguishing among spectral features of different exercise intensities. This demonstrates the limitations of the model when applied to raw spectral data, further supporting the need for pre-processing to enhance class separation and prediction accuracy. Following this, the overall model accuracy of 40% (with 12 latent variables) underscores these challenges.

Following the application of MSC, the next step was to determine which supervised model, PCA-DA or PLS-DA, would be most appropriate for discriminating between the exercise intensity groups. The score scatter plots of both models were analyzed to assess their ability to separate the groups effectively.

In the PCA-DA score scatter plot, the application of MSC resulted in a slight improvement compared to the unprocessed data, but the separation of the groups remained limited (Fig. 68a). There was still big overlap among the different exercise intensity groups, indicating that PCA-DA might not be able to capture the full complexity of the data. While the PCA-DA plot showed a minimal tendency to cluster samples, the separation was not distinct enough to reliably classify the groups. On the other hand, the PLS-DA score scatter plot exhibited much clearer group separation after MSC pre-processing (Fig. 68b). PLS-DA, being a supervised technique, showed much better discrimination between the four exercise intensity classes. Samples from the rest,

walking, jogging, and running groups were more distinctly separated, indicating that PLS-DA was better suited to handling the complexity of the data and highlighting the importance of supervised methods in such analyses. This improved clustering in the PLS-DA plot suggested that this model could more effectively capture the biochemical differences between the groups, making it a more appropriate choice for the analysis.

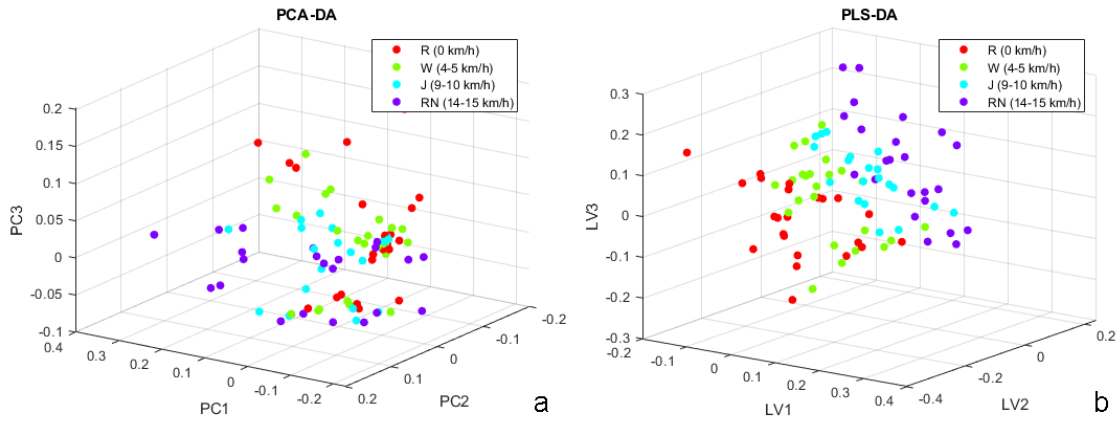


Figure 68. Scatter plot of saliva second derivative spectra at different exercise intensities of (a) PCA-DA model and (b) PLS-DA model.

In general, the score scatter plots of both models demonstrated that MSC pre-processing significantly enhanced the ability of the models to discriminate between the groups. While PCA-DA provided some clustering, it lacked the clear separation observed in the PLS-DA plot, confirming PLS-DA as the more appropriate supervised method for this study.

To minimize the expected error and optimize the performance of both models (PCA-DA and PLS-DA), cross-validation was performed to evaluate how well each model generalizes to new data. Cross-validation is an essential step in assessing the robustness and predictive power of multivariate models, ensuring that the chosen model does not overfit to the training data but can accurately predict unseen samples. Thus, the number

of components in both PCA-DA and PLS-DA models was carefully investigated. The goal was to find the optimal number of components that provided the best trade-off between model complexity and prediction accuracy, without overfitting the data.

Initially, both models were tested with different numbers of components, with up to 10 components for PCA-DA and 20 components for PLS-DA. Twenty components may lead overfitting the data, but diagnostics were used to choose a model with minimum necessary components. While a higher number of components can lead to a better fit with the training data, it can also cause overfitting, where the model captures noise or random fluctuations rather than meaningful patterns. Overfitting can result in poor performance when the model is applied to new, unseen data (test data).

In the coefficient of determination (R^2) analysis, representing the fitting ability of the two models, PLS-DA showed a much better fitting ability compared to PCA-DA (Fig. 69a). For both models, the fitting ability improved with an increasing number of components, but the R^2 value plateaued after 10 components for PLS-DA (R^2 index of 0.9927), indicating no significant improvement in fitting with additional components. In contrast, PCA-DA's R^2 value was significantly lower, even with 10 components (R^2 index of 0.4421), demonstrating that PCA-DA was not as effective in capturing the data's underlying structure for exercise intensity discrimination.

To further investigate the optimal number of components, 10-fold repeated cross-validation (100 repeats) was performed. Cross-validation is a more statistically robust method for determining the ideal number of components, as it evaluates the model's performance on unseen data and helps prevent overfitting. The cross-validation results showed that PCA-DA had higher mean squared error of cross-validation (MSECV) compared to PLS-DA, indicating a less accurate prediction model (Fig. 69b).

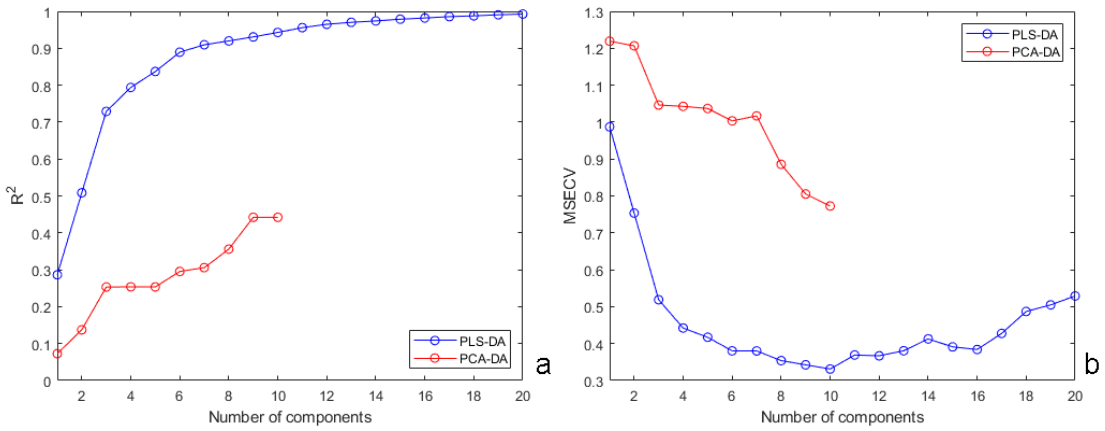


Figure 69. Diagnostics of PCA-DA vs. PLS-DA models: (a) Fitting ability of the models, (b) mean squared error of 10-fold cross validation (MSECV) of the models in association with the number of components.

Specifically, the fourth and seventh components in the PCA-DA model were found to increase the MSECV, suggesting that the combination of predictor variables in those components was not strongly related to the response variable. This happens because the PCA constructs components to explain variation in predictor variables (spectral intensities), not response variable (exercise intensity). This highlighted that the model was capturing unnecessary variation, leading to an increased error and further reinforcing the limitations of PCA-DA for this task. In contrast, the PLS-DA model showed minimum error with 10 components (Fig. 69b), with the MSECV stabilizing at this point. This result indicated that 10 components were sufficient to capture the relevant information in the data without overfitting, as adding more components did not significantly improve the model's performance. The PLS-DA model with 10 components reached a fitting of 0.9927 R^2 and an MSECV of 0.3362, demonstrating a high-quality prediction model.

The results of the component optimization and cross-validation analyses confirmed that PLS-DA with 10 components was the optimal model for discriminating exercise intensities using ATR-FTIR salivary spectra. The PLS-DA model achieved the best balance

between fitting the data and avoiding overfitting, making it suitable for distinguishing among saliva spectra at different exercise intensities. The careful choice of 10 components based on diagnostic measures ensured that the model was both parsimonious and robust, providing accurate predictions without capturing irrelevant variations or noise.

17.1.2 Model evaluation

The performance of the PLS-DA model with 10 components was assessed using the confusion matrix of the test dataset, as shown in (Fig. 70). The confusion matrix provides an evaluation of how well the model predicts the classes of the "unknown" test samples by comparing the predicted labels with the actual. The confusion matrix evaluates the performance of the selected PLS-DA model with 10 LVs (latent variables) for classifying different exercise intensities (R: 0 km/h, W: 4-5 km/h, J: 9-10 km/h, RN: 14-15 km/h) based on salivary ATR-FTIR spectra. The matrix shows how the predicted classes correspond to the true classes, with the diagonal elements representing correctly classified samples and the off-diagonal elements indicating misclassifications.

From the confusion matrix (Fig. 70), we can make the following observations:

- Class R (0 km/h): The model performed very well in predicting class R, with 8 true positives (correctly predicted as R) and only 1 misclassification (predicted as W). The model demonstrated high sensitivity, correctly identifying 89% of the samples with this exercise intensity. This means the model is quite effective in predicting the rest condition (R), with very few misclassifications. Additionally, the specificity for class R is perfect at 1, indicating that the model correctly identified all non-R samples as not being class R, with no false positives.

- Class W (4-5 km/h): Class W had 7 true positives, but also 1 misclassification as class J and 1 as class RN. Thus, the sensitivity is good at 0.78, meaning the model correctly identified 78% of the samples in this class. While this is fairly good, there were some misclassifications, suggesting some overlap with J and RN classes. The specificity for class W is strong at 0.92, indicating that the model was able to correctly identify most samples that were not class W. There was a small amount of misclassification as class R, but overall, class W was well discriminated from other exercise intensities.
- Class J (9-10 km/h): Class J showed 5 misclassifications as class RN and 1 misclassification as class W. This suggests that class J shares some similarities with other intensities (especially RN and W), leading to some errors in classification. The low sensitivity suggests that class J is difficult to distinguish from other classes, and the model struggles to correctly classify these samples. However, the specificity for class J is high at 0.96, meaning that when the model predicted class J, it was rarely wrong.
- Class RN (14-15 km/h): The model performed well in predicting RN, with 9 true positives and zero misclassifications. The sensitivity is perfect at 1, indicating that the model correctly identified all the samples in this class. This suggests that class RN is well discriminated from other intensities, and the model is highly accurate in detecting the high-intensity condition. The specificity for class RN is 77%, which is lower than that of the other classes. This means that some samples from other classes, particularly J, were misclassified as RN. Despite this, the model still maintains a good level of specificity in most cases.

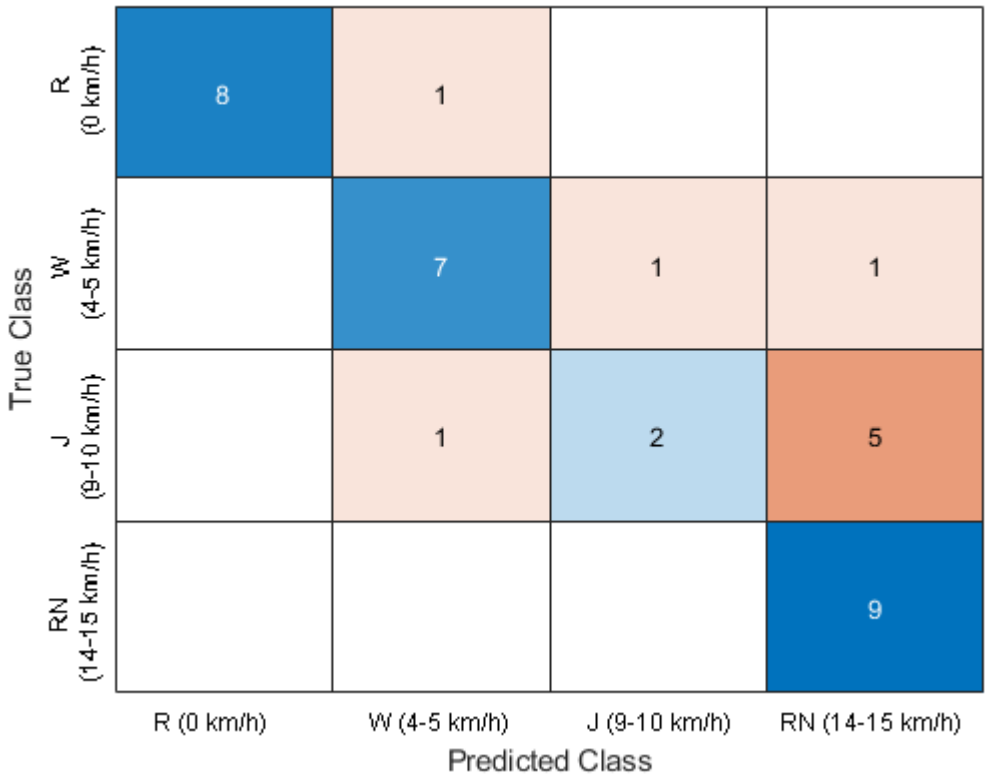


Figure 70. Confusion matrix of predictive 10-components PLS-DA model.

For the selected PLS-DA model with 10 components, class RN (14-15 km/h) shows the highest sensitivity with 9 true positive test samples, indicating it is very well discriminated by the model. However, class W is moderately well predicted, while class J has the most difficulty, exhibiting low sensitivity but high specificity. In total of 35 samples, and out of those, only 9 were misclassified, achieving a total accuracy of 74.3%. The sensitivity and specificity of each class were extracted from the confusion matrix and are summarized in Table 7.

Table 7. Sensitivity and specificity of different exercise intensities.

Exercise intensity	Sensitivity	Specificity
R (0 km/h)	0.89	1
W (4-5 km/h)	0.78	0.92
J (9-10 km/h)	0.25	0.96
RN (14-15 km/h)	1	0.77

17.1.3 Factors interpretation

To uncover the biochemical features underlying the discrimination of exercise intensities, the loading plot of the first two latent variables (LVs) in the PLS-DA model was examined. The loading plot is a critical analytical tool that highlights the spectral wavenumbers most influential in classifying the saliva samples according to exercise intensity. By focusing on these features, meaningful biochemical insights can be derived. The analysis of the loading plot (Fig. 71) revealed that a broad spectral range contributed to the differentiation of exercise intensity classes.

However, several specific peaks stood out as particularly significant:

- 2056 cm^{-1}: This peak corresponds to the vibrations of thiocyanate ions[320,321,333–335].
- 1724 cm^{-1}: Assigned to the carbonyl group (C=O) vibrations of esters found in phospholipids[320,334], this peak reflects membrane dynamics and lipid metabolism during exercise.
- 1410 cm^{-1}: Linked to lactate vibrations[321,345], this feature is crucial for monitoring exercise-induced metabolic shifts, particularly during high-intensity anaerobic activity where lactate production increases significantly.

- 1234 cm^{-1}: Associated with the asymmetric stretch of phosphate (PO_2^-) mainly in nucleic acids[303,320,334,345–347], this peak suggests changes in cellular activity and nucleic acid dynamics in response to metabolic stress.
- 1052 cm^{-1}: Corresponding to glucose and sugar moieties[320,321,345], this feature underscores the role of carbohydrate metabolism in energy production across different exercise intensities.
- 912 cm^{-1}: Assigned to the additional vibration of lactic acid[348].

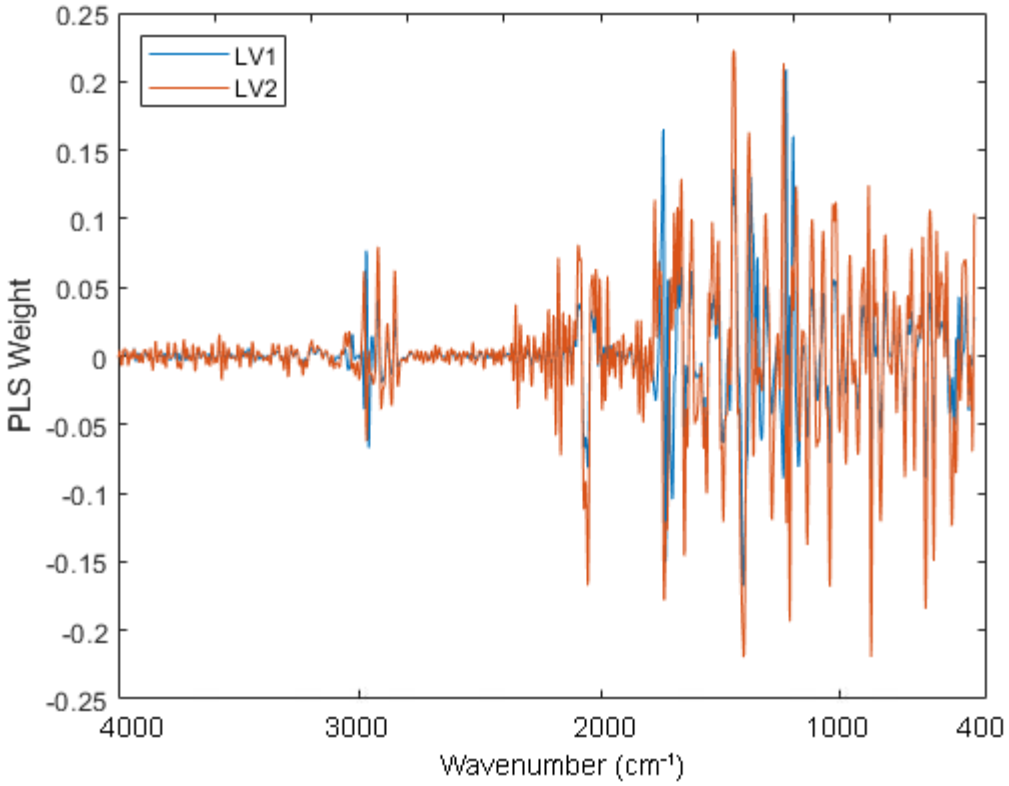


Figure 71. Loading plot of the first two latent variables LVs of PLS-DA model.

The loading plot analysis provides a detailed biochemical perspective on how saliva spectra vary with exercise intensity. These findings support the physiological relevance of

the model's predictions and emphasize the potential of ATR-FTIR spectroscopy for non-invasive metabolic monitoring, offering a nuanced understanding of the physiological adaptations to exercise. By linking spectral features to specific biochemical processes, this analysis provides both a mechanistic and practical foundation for saliva-based diagnostics.

17.2 Trends of salivary *vs.* blood biomarkers: Lactate and Glucose

After interpreting the factors obtained from the multivariate analysis, we focused on two widely recognized exercise induced biomarkers—lactate and glucose—selected from the five chemical species identified above (section 17.1.3). Thus, the focus was on evaluating the agreement between salivary and blood biomarkers, lactate and glucose, as indicators of exercise. The use of ATR-FTIR spectroscopy for saliva analysis was compared with enzymatic analyzers used for blood samples. The aim was to determine whether non-invasive saliva analysis could accurately track biochemical trends during varying physical activity levels. Thus, the selected salivary biomarkers qualitatively tracked *via* infrared spectroscopy and the findings compared with the corresponding blood samples *via* reference methods.

Firstly, PLS-Regression (PLS-R) was employed to evaluate the correlation between the full range salivary ATR-FTIR spectra and blood concentrations of glucose and lactate. A training set of 90 samples was employed, covering glucose concentrations between 84–149 mg/dL and lactate concentrations between 0.8–9.9 mmol/L. An independent test set of 35 samples was used for validation, with glucose and lactate ranges of 85–142 mg/dL and 0.8–9.8 mmol/L, respectively. This analysis aimed to determine the extent to which information being partially present in other salivary spectral regions could predict blood biomarker levels, thereby validating the use of saliva as a non-invasive matrix for biochemical monitoring during exercise.

The glucose PLS-R model was constructed using second derivative ATR-FTIR spectra and optimized with four latent variables (LVs), determined *via* 10-fold cross-validation (Fig. 72a). The model demonstrated a moderate fit for the training set, with an R^2 value of 0.7008 (Fig. 72b). However, the test set performance revealed significant scattering for samples with higher glucose concentrations, resulting in limited predictive accuracy (Fig. 72c). Regression coefficients identified key vibrational bands at 1740 cm^{-1}, 1408 cm^{-1}, and 1244 cm^{-1} as being additionally associated with blood glucose levels (Fig. 72d), while the characteristic vibration of glucose at 1050 cm^{-1} is weak. These bands are likely linked to the molecular vibrations of glucose-related functional groups. Despite these associations, the model's weaker predictive power for the test set suggests variability in glucose trends between saliva and blood, possibly due to physiological differences in glucose transport and metabolism in these two matrices[349]. Factors such as oral enzymatic activity and the complexity of glucose regulation may contribute to this variability. Additionally, inter-individual variability—including differences in hydration status and circadian rhythms—may significantly influence salivary glucose profiles[349].

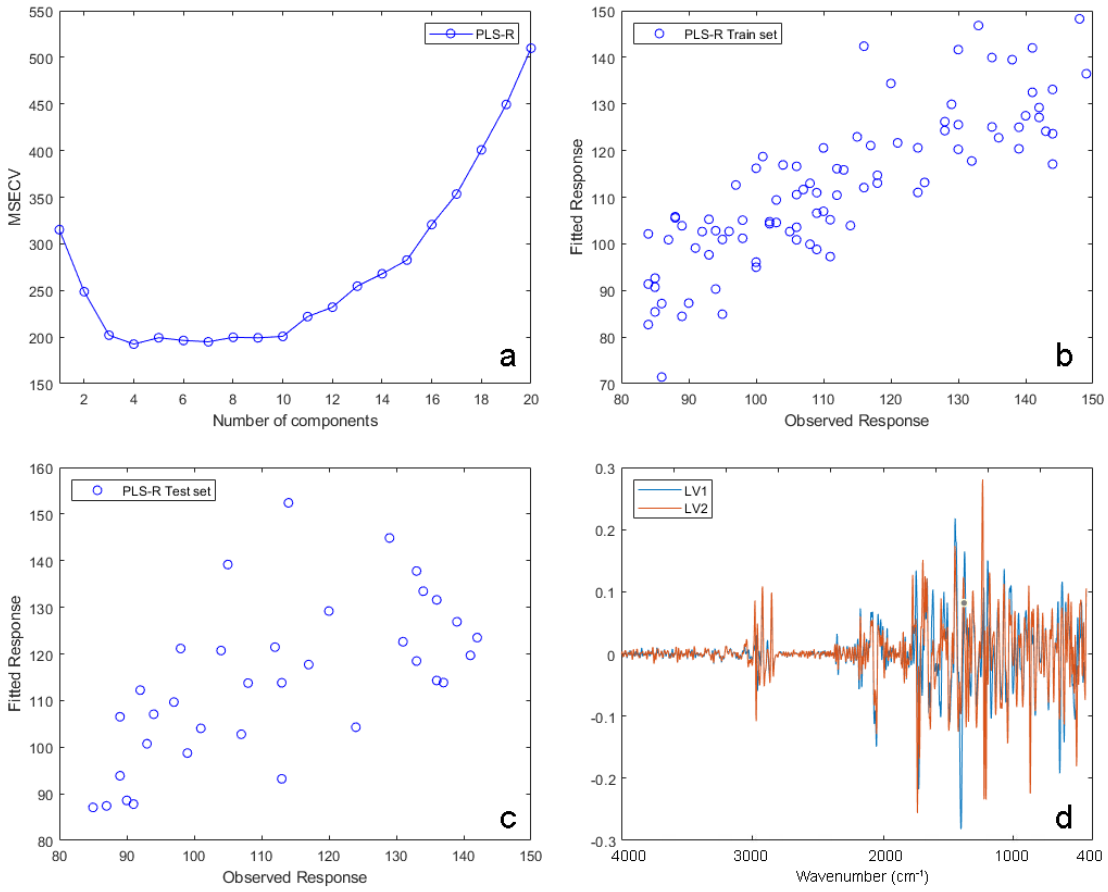


Figure 72. 4-LVs PLS-R model for glucose: (a) 10-fold cross-validation, (b) predicted vs. measured glucose in training set, (c) predicted vs. measured glucose in test set, (d) regression coefficients from PLS-R.

For lactate, the PLS-R model utilized nine latent variables, achieving a much stronger correlation with an R^2 value of 0.9276 for the training set (Fig. 73b) and 0.7295 for the test set (Fig. 73c). This robust performance underscores the reliability of salivary lactate as a proxy for blood lactate levels. Regression coefficients highlighted key spectral bands at 1744 cm^{-1}, 1418 cm^{-1}, 1246 cm^{-1} and 910 cm^{-1} as indicative of lactate concentration in blood (Fig. 73d). These vibrations correspond to functional groups linked to lactate molecules, reaffirming their suitability as markers for metabolic changes of salivary lactate during exercise.

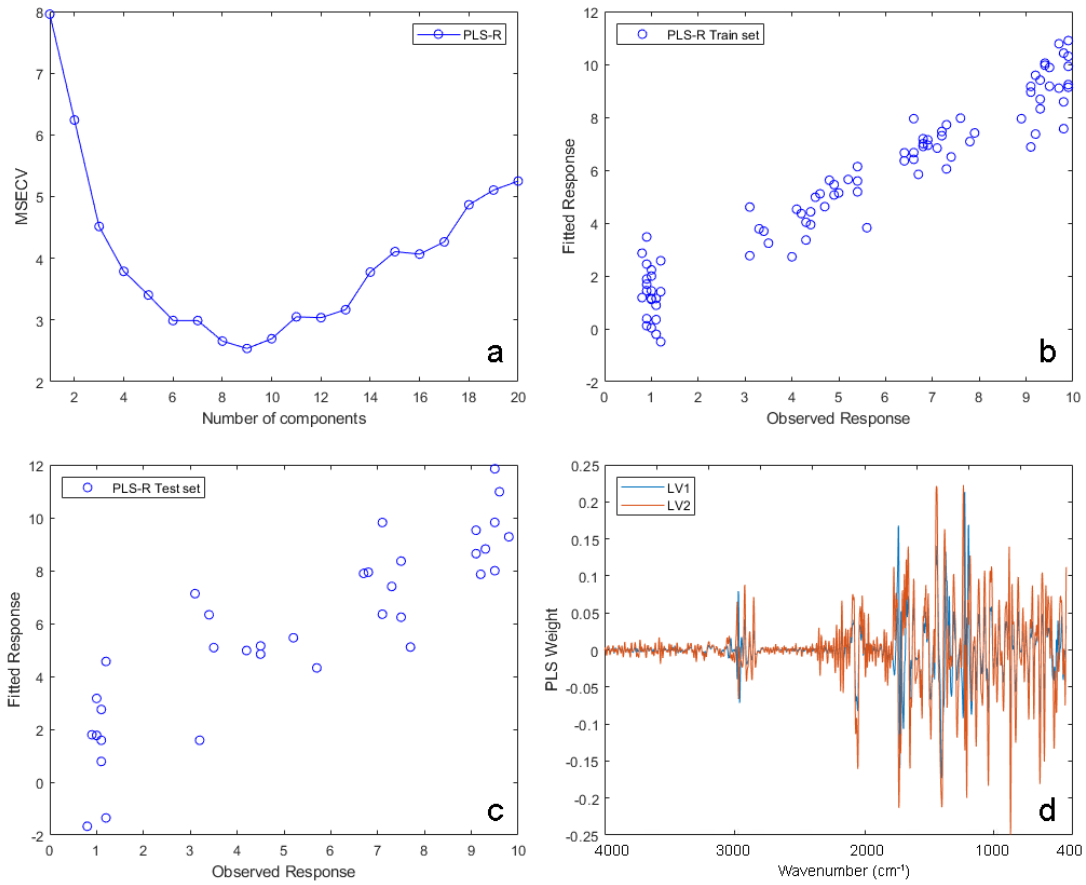


Figure 73. 9-LVs PLS-R model for lactate: (a) 10-fold cross-validation, (b) predicted *vs.* measured lactate in training set, (c) predicted vs. measured lactate in test set, (d) regression coefficients from PLS-R.

Overall, the results from PLS-R analysis reinforce the potential of ATR-FTIR spectroscopy for non-invasive biomarker monitoring, at least for the case of lactate. The strong correlation and predictive accuracy validate salivary lactate as a reliable marker for tracking exercise intensity. The identification of specific vibrational bands (e.i. 1418 cm^{-1}, 910 cm^{-1} [321,345,348]) strengthens the biochemical basis for this correlation.

Following the PLS-R analysis, we proceeded to semi-quantitatively monitor the trends of lactate and glucose by integrating their corresponding IR spectral bands and

plotting their alterations across the increasing exercise intensities. A similar approach was applied to the blood measured values of these biomarkers. This parallel comparison helped clarify the outcomes of the PLS model, reinforcing the link between salivary spectral features and systemic metabolic changes.

Blood lactate levels (measured using Contour® analyzer) exhibited a clear and expected exponential increase as exercise intensity rose, starting from a baseline of approximately 1.0 mmol/L at rest to a peak of 9.5 mmol/L during high-intensity running (14–15 km/h). This aligns with well-established physiological responses, where lactate production intensifies with increased reliance on anaerobic metabolism under higher physical exertion. Salivary lactate levels, as measured by ATR-FTIR spectroscopy, mirrored this trend but displayed a more gradual rise. To obtain quantitative information from the spectra, the area of the band corresponding to a particular analyte was measured. The spectral bands associated with lactate (1428–1391 cm^{-1})[345] showed a consistent increase across exercise intensities. This correlation suggests that salivary lactate could serve as a non-invasive proxy for blood lactate in monitoring exercise intensity. The biochemical similarity between trends in saliva and blood supports the validity of saliva as an alternative matrix for lactate analysis, reinforcing its potential utility in sports science and athlete monitoring.

In contrast to lactate, glucose trends revealed notable differences between blood and saliva. Blood glucose levels (measured using Accutrend® Plus analyzer) exhibited a steady increase with exercise intensity, reaching a maximum of 127 mg/dL at the highest intensity. This increase corresponds to the physiological mobilization of glucose to meet energy demands during sustained physical activity[58]. Salivary glucose, however, showed a less pronounced pattern. While there was a slight overall increase in the glucose-related spectral band (1094–1042 cm^{-1})[345,350], the trend was less clear, with a minor drop observed at the jogging intensity (9–10 km/h). The small agreement between salivary and blood glucose levels in multivariate regression above, is now more clear. The mentioned discrepancy could be attributed to various factors, including the differential transport and

regulation of glucose in saliva compared to blood, the individual variability, even to the simulated collection method of saliva with cotton swab[349]. These findings in salivary glucose trend align with previous research[345] that noted similar patterns, with salivary glucose levels fluctuating depending on the specific exercise intensity. Vieira et al. reported the same tendency: as salivary glucose increases from rest up to 14 km/h, it then drops at 15 km/h and remains unchanged until 17 km/h, a point where sharply increases[345].

The box plots (Fig. 74a–d) visualize the variations of salivary (ATR-FTIR measurements) and blood (specific enzymatic measurements) lactate and glucose. Direct comparison is difficult to follow attributed to the different units expressing the variations. Thus, in Figure 74e,f the co-linearity was examined.

These findings emphasize the importance of refining analytical techniques, highlight the promise of ATR-FTIR spectroscopy for real-time, non-invasive biomarker monitoring in exercise physiology and open pathways for personalized athletic performance assessments.

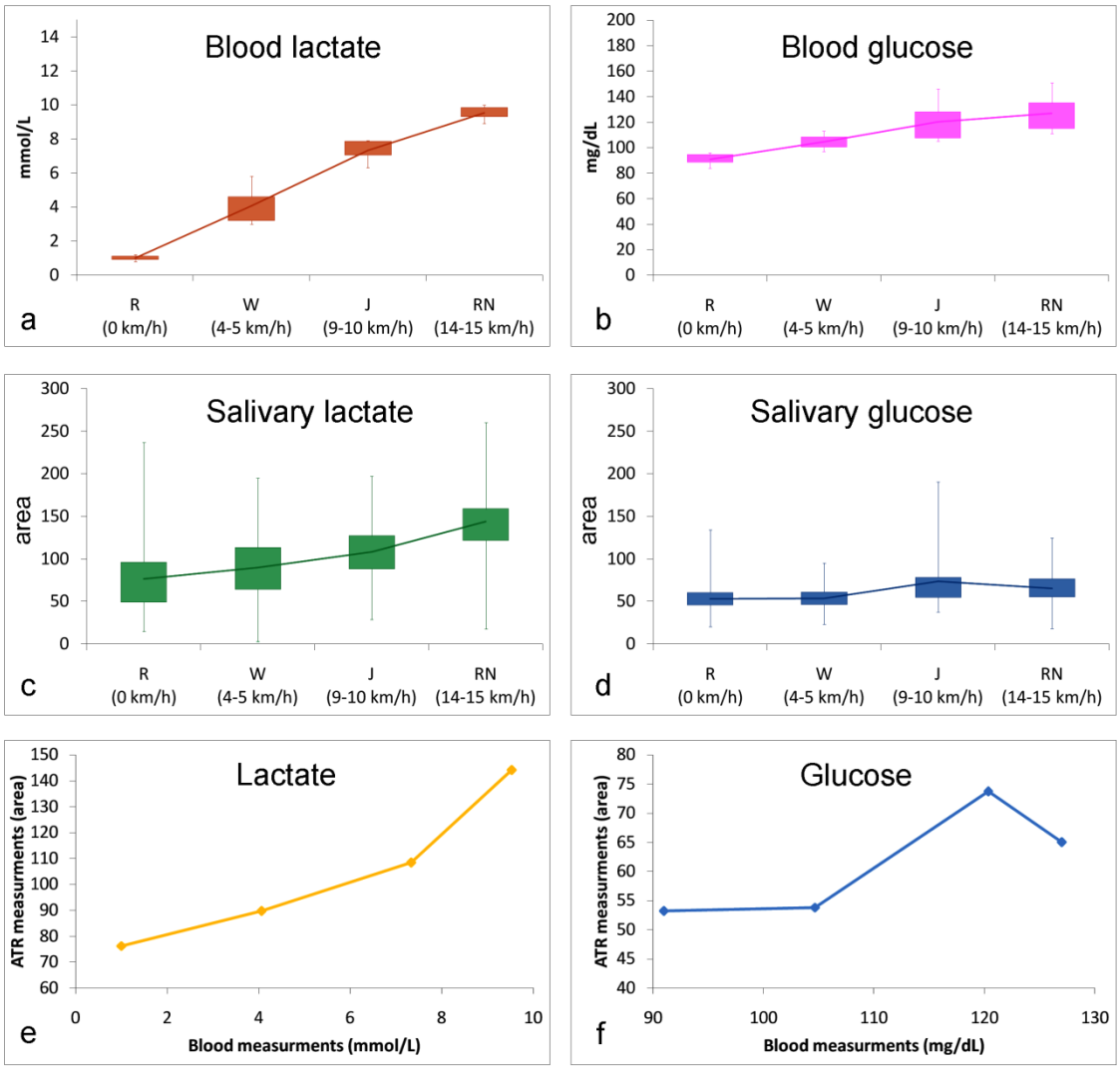


Figure 74. Variations and mean values of (a) area of lactate's second derivative spectral band in saliva (ATR-FTIR), (b) lactate in blood (enzymatic analyzer), (c) area of glucose's second derivative spectral band in saliva (ATR-FTIR), and (d) glucose in blood (enzymatic analyzer), during increment of exercise intensity. Plot of salivary (ATR-FTIR) *vs.* blood (enzymatic analyzers) measurements for (e) lactate and (f) glucose.

17.3 Monitoring biomarkers during increased physical effort

Having established the efficacy of PLS-DA in distinguishing saliva samples based on exercise intensity, the focus now shifts to monitoring specific biochemical markers identified as key contributors to this discrimination. Through PLS-DA analysis, five biomarkers –lactate, glucose, thiocyanate, phosphate, lipids (phospholipids)– were pinpointed as critical for capturing metabolic changes associated with physical activity. These biomarkers are directly linked to physiological adaptations during exercise, providing valuable insights into energy metabolism, oxidative stress, and cellular activity[58]. These observations underscore the dynamic biochemical responses of the body during physical activity and highlight the potential of saliva as a non-invasive medium for monitoring exercise-induced physiological alterations. By focusing on these biomarkers, the subsequent analysis delves deeper into their individual trajectories during exercise. Thus, the lactate, glucose and the rest chemical species were monitored and statistically examined, similarly to previous semi-quantitatively evaluation of lactate and glucose during the increased exercise intensity.

For phosphates, the spectral intensity showed a slow but consistent linear increase across all exercise intensities, reflecting gradual metabolic changes (Fig. 75a). In contrast, phospholipids exhibited a faster and more pronounced increase, particularly at the highest exercise intensity (RN, 14–15 km/h) (Fig. 75b). This substantial rise, despite variability in data, points to the metabolic role of phospholipids in energy-demanding activities[351–353], with their higher mean levels aligning with heightened physical exertion. Furthermore, the intensity of glucose bands displayed only a slight increase, suggesting limited salivary glucose variation during exercise (Fig. 75c). This minor change aligns with its relatively weaker influence in the PLS-DA loadings plot (region of 1052 cm^{-1}) compared to other biomarkers, such as lactate. The minimal glucose response emphasizes the need to consider alternative or complementary biomarkers for better monitoring glucose-related metabolic activity in saliva. Lactate, on the other hand, exhibited a clear upward trend in spectral intensity, strongly correlating with increased exercise intensity (Fig. 75d).

This trend aligns with previous findings[345] and blood lactate data from the current Phase, reinforcing lactate's significance as an indicator of skeletal muscle metabolism[125]. As lactate levels are closely tied to energy production and overtraining syndrome[125], its robust and consistent increase highlights its potential as a reliable biomarker for tracking exercise physiology. Only thiocyanates showed a decrease in alteration during the prolonged exercise intensity (Fig. 75e).

These patterns were further evaluated using an Analysis of Variance (ANOVA). According to the analysis, responses of all five examined metabolites alter statistically significantly (p <0.05) during the prolonged exercise. ANOVA performed for glucose indicated the highest p-value (although <0.05), fact that confirms the visually inspected inconsistent trend. The performed ANOVA with five metabolites and four exercise intensity group, conclude that there is a statistically significant difference somewhere among the four group means. However, ANOVA doesn't specify where those differences are. For instance, ANOVA might indicate that not all the groups are the same, but it won't tell whether the difference is between group A and group B, or between A and C, or some other combination.

To pinpoint which groups are significantly different, post hoc test was conducted. This family of tests compares the means of the groups pair by pair. This level of detail is critical for understanding the resulted trends and drawing meaningful conclusions. Additionally, when comparing multiple groups, there's a higher chance of detecting a difference just by chance. Post hoc tests adjust for this increased risk, ensuring that any significant results are less likely to be false positives. Thus, the Bonferroni correction is a statistical method of the post hoc family tests, used to reduce the likelihood of false positives errors when conducting multiple comparisons. It adjusts the significance level (p) by dividing it by the number of comparisons being made. This makes post hoc tests essential for maintaining the reliability of the findings when dealing with multiple comparisons.

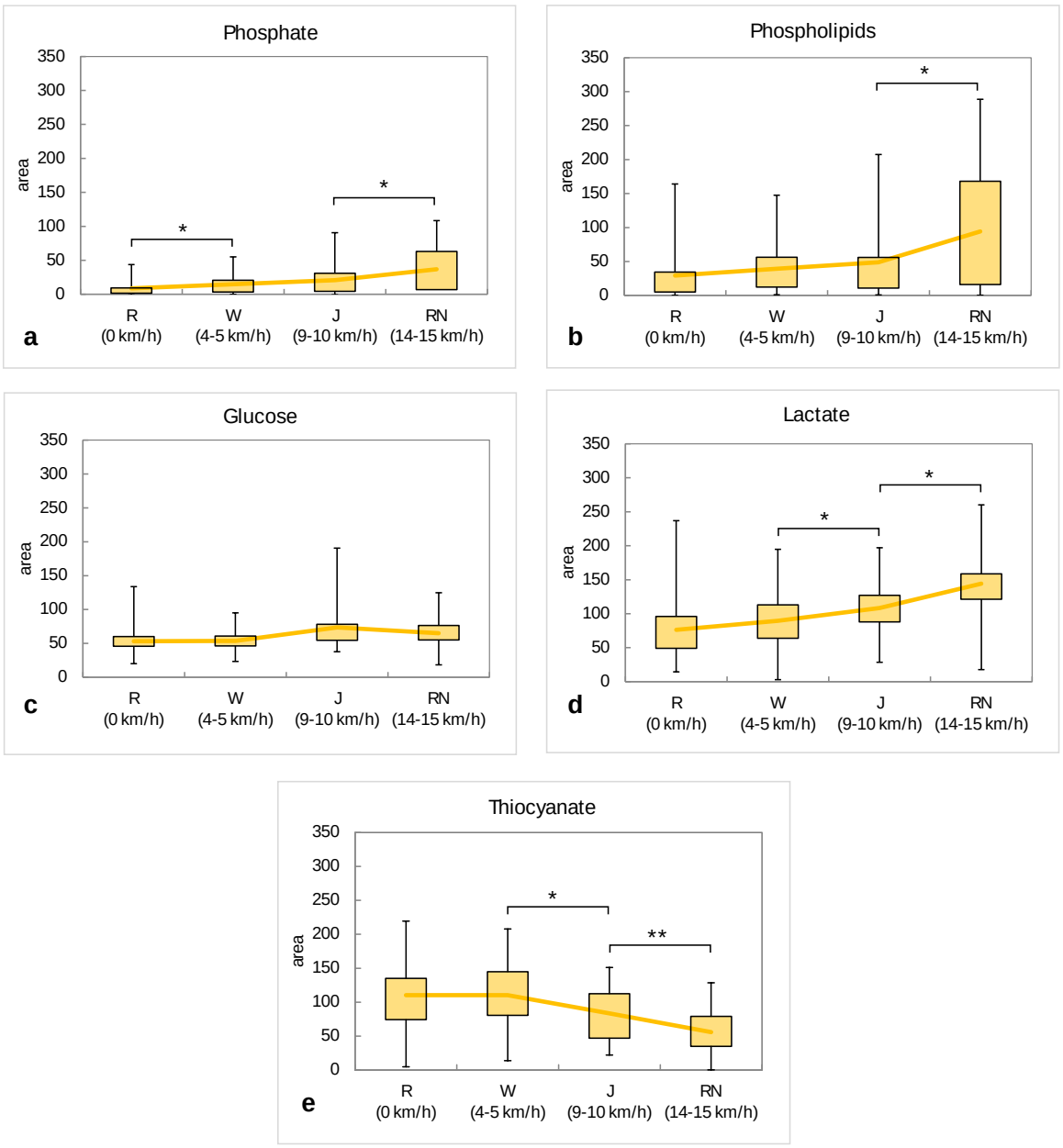


Figure 75. Area variations and mean values of second derivative spectral band in saliva, during increase of exercise intensity: (a) phosphate, (b) phospholipids, (c) glucose, (d) lactate, (e) thiocyanate. *p-adjusted <0.05, **p-adjusted <0.01

Bonferroni analysis was performed on the five metabolites across R-W, W-J and J-RN running states and the adjusted p-values are presented in Table 8. For glucose, the

statistical output indicates that no significant differences were found between any of the exercise intensity groups after applying the Bonferroni correction (*p*-adjusted >0.05). This suggests that the observed variations in salivary glucose levels across intensities were not strong enough to reach statistical significance under this stringent multiple comparisons adjustment. The absence of statistical significance in glucose trends was expected, as previously indicated (section 17.2) by the saliva-blood study of agreement. In that section, the weak similarity between salivary and blood glucose levels was already evident. This prior observation aligns with the current results and reinforces the physiological disconnect between the two matrices.

On the other hand, for thiocyanate, the *p*-adjusted value being <0.01 after Bonferroni correction demonstrates highly significant differences between J and RN groups (Fig. 75e). This indicates that thiocyanate levels changed consistently and significantly, surpassing the threshold for statistical significance even under the strict correction for multiple comparisons.

Table 8. ANOVA (*p*-values) and Post Hoc - Bonferroni (adjusted *p*-values, <0.05 are in bold) results across exercise intensity groups.

	ANOVA	Post hoc test (Bonferroni)		
		R - W	W - J	J - RN
Phosphate	2.26E-06	**0.007274**	0.108541	**0.003733**
Phospholipids	4.49E-05	0.173148	0.175853	**0.003991**
Glucose	0.004475	0.934226	0.028657	0.217306
Lactate	4.34E-07	0.100275	**0.011024**	**0.004729**
Thiocyanates	3.13E-07	0.997643	**0.016418**	**0.000198**

Most of the above biomarkers have already been mentioned that are correlated with physical exercise[303,319,345,354]. However, this study marks the first comprehensive analysis of thiocyanate ions (SCN^-) in saliva as a biomarker for distinguishing exercise intensities. While most other biomarkers identified in this research have already been associated with physical exercise, the discovery of SCN^- as a discriminatory marker is novel and sheds light on its potential role in monitoring training load and metabolic responses.

Thiocyanate, an acidic pseudohalide thiolate, is found in high concentrations in saliva, ranging from 0.5 to 3 mM, making saliva the richest body fluid in terms of SCN^- content[339,340]. Its role extends beyond being a simple metabolite; thiocyanate acts as a precursor for antimicrobial agents in the presence of salivary peroxidase enzymes such as lactoperoxidase. During physical exercise, biochemical interactions involving SCN^-, lactate, and peroxidase activity underscore its dynamic behavior as a biomarker. The reduction in salivary SCN^- levels during increasing exercise intensity (Fig. 75e) can be attributed to its oxidative conversion. Hydrogen peroxide, generated by the action of lactoperoxidases[355], reacts with SCN^- to produce antimicrobial products such as $OSCN^-$, HOSCN, O_2SCN^-, and O_3SCN^-[356]. This oxidative pathway is stimulated by lactate[356], which is known to increase significantly during higher exercise intensities. The interplay between lactate metabolism and salivary peroxidase activity suggests a direct biochemical link influencing SCN^- levels. The increase in lactate during exercise likely amplifies salivary peroxidase activity, boosting hydrogen peroxide production[356,357]. This enhanced peroxidase activity accelerates the oxidation of SCN^-, reducing its concentration in saliva. These findings suggest that SCN^- depletion is not merely a passive outcome but is actively mediated by exercise-induced biochemical processes involving lactate and salivary enzymatic activity[125].

The strong correlation between lactate levels and SCN^- absorption bands implies a functional relationship between lactate metabolism, oxidative stress, and thiocyanate concentration[355]. As exercise load increases, the simultaneous rise in lactate and salivary

peroxidase activity results in SCN^- oxidation, making it a reliable marker for tracking exercise intensity. Thus, the observed decrease in SCN^- reflects a complex interplay of metabolic and enzymatic processes unique to saliva. These interconnected biochemical pathways position SCN^- as a dual indicator: it signals both the metabolic demands of exercise and the systemic oxidative responses.

Overall, thiocyanate's distinctive response to exercise intensity, coupled with its strong correlation with lactate, highlights its utility as a salivary biomarker for monitoring training load and metabolic activity. By capturing the interplay between lactate metabolism and salivary peroxidase activity, SCN^- provides a unique perspective on the physiological adaptations to exercise (more details are presented in Phase 3, *Section 18*). This finding not only expands the understanding of salivary biochemistry during physical activity but also paves the way for its application in non-invasive, real-time exercise monitoring and personalized training strategies.

In summary, Phase 2 demonstrated the feasibility of using saliva as a non-invasive tool to monitor biochemical changes associated with physical exercise through ATR-FTIR second derivative spectroscopy. Saliva samples collected at varying exercise intensities—rest, walking, jogging, and running—were analyzed, with PLS-DA achieving a classification accuracy of 74.3%. Five key biomarkers—thiocyanate, phospholipids, lactate, phosphate, and glucose—were identified as significant drivers of discrimination. Among these, thiocyanate (SCN^-) exhibited a novel and significant decrease with increasing exercise intensity, marking it as a potential biomarker for tracking oxidative stress and metabolic responses. This study highlights saliva's potential for real-time, personalized monitoring of exercise intensity, providing valuable insights for sports science and training optimization.

Phase 2 established, for the first time, a non-invasive and data-driven framework capable of classifying graded physical exercise intensities based on salivary biochemical profiles. By integrating second-derivative ATR-FTIR spectroscopy with advanced chemometric modeling (PCA-LDA and PLS-DA), this Phase expanded the analytical

resolution beyond the binary design of Phase 1, enabling the detection of subtle, intensity-dependent metabolic changes. The predictive models developed demonstrated high classification accuracy, underscoring the robustness of saliva-based spectroscopic fingerprinting (highlighting phosphate, phospholipids, glucose, lactate, thiocyanate) for physiological load assessment. Notably, thiocyanate (SCN^-) emerged as a novel, exercise-responsive spectral feature, as mentioned previously. While this Phase demonstrated the feasibility of real-time, non-invasive exertion monitoring, the findings also pointed to the need for systemic metabolic insights—beyond the oral cavity. This realization set the foundation for Phase 3 and Phase 4, where quantitative and multi-biofluid strategies were introduced to validate and expand the physiological interpretations of these initial steps of the study.

18. Phase 3 – Exploring salivary thiocyanate as a novel biomarker of physical activity response

The outcomes of Phase 2 confirmed that salivary biochemical composition is sensitive to graded exercise intensities and can be effectively monitored using infrared spectroscopy coupled with chemometric modeling. However, while the spectroscopic approach provided valuable qualitative insights and revealed promising candidate biomarkers such as lactate, glucose, and thiocyanate (SCN^-), it lacked the ability to deliver quantitative measurements necessary for clinical validation and broader physiological interpretation. Among the spectral markers identified, SCN^- emerged as a particularly responsive feature, yet its exact concentration dynamics during exercise remained unexplored. To bridge this gap, Phase 3 focused on the development of a robust, specific, and scalable analytical method for the quantitative determination of salivary thiocyanate.

This transition marked a deliberate methodological shift—from holistic, multivariate spectral fingerprinting to a targeted, molecular-specific photometric approach. The goal of Phase 3 was twofold: first, to validate thiocyanate as a physiologically meaningful, exercise-responsive salivary biomarker, and second, to examine its concentration trends across multiple exercise intensities in a broader athlete population. This phase thus aimed to transform a statistically identified spectral feature into a quantifiable biological marker, advancing the biomarker discovery pipeline from exploratory detection to functional characterization.

The determination of thiocyanate concentration in saliva is based on the reaction between iron(III) ions (Fe^{3+}) and thiocyanate ions (SCN^-) to form the thiocyanatoiron(III) complex, $FeSCN^{2+}$. This reaction follows the equilibrium (Eq. 8):

$$Fe^{3+}_{(aq)} + SCN^-_{(aq)} \rightleftharpoons FeSCN^{2+}_{(aq)} \quad (8)$$

When the concentration of iron(III) ions is significantly higher than that of thiocyanate, the formation of higher-order complexes such as $Fe(SCN)_2^+$ and $Fe(SCN)_3$ can be excluded. The $FeSCN^{2+}$ complex exhibits a deep orange-red color, making it ideal for spectrophotometric quantification.

To construct the calibration curve, standard solutions were prepared where $[Fe^{3+}] >> [SCN^-]$, ensuring that all thiocyanate ions were converted to $FeSCN^{2+}$. This approach allows for accurate determination of thiocyanate concentrations in saliva samples by comparing their absorbance values to the calibration curve. The high sensitivity and selectivity of this method make it a reliable tool for analyzing thiocyanate variations during physical exercise.

18.1 Method development

To ensure accurate quantification of thiocyanate in saliva, the method was first developed by analyzing the visible absorption spectrum of standard $FeSCN^{2+}$ solutions. Figure 76 presents the absorbance spectra of these standard solutions, showing a characteristic peak at approximately 458 nm. This peak corresponds to the maximum absorption of the $FeSCN^{2+}$ complex, confirming its formation and suitability for spectrophotometric analysis.

The observed absorption peak at 458 nm is in close agreement with the literature value of 447 nm[306], with slight variations possibly attributed to differences in experimental conditions such as solvent composition. The intensity of this peak increases proportionally with thiocyanate concentration, demonstrating a direct relationship between absorbance and analyte concentration. This characteristic enables the construction of a calibration curve, which forms the basis for thiocyanate determination.

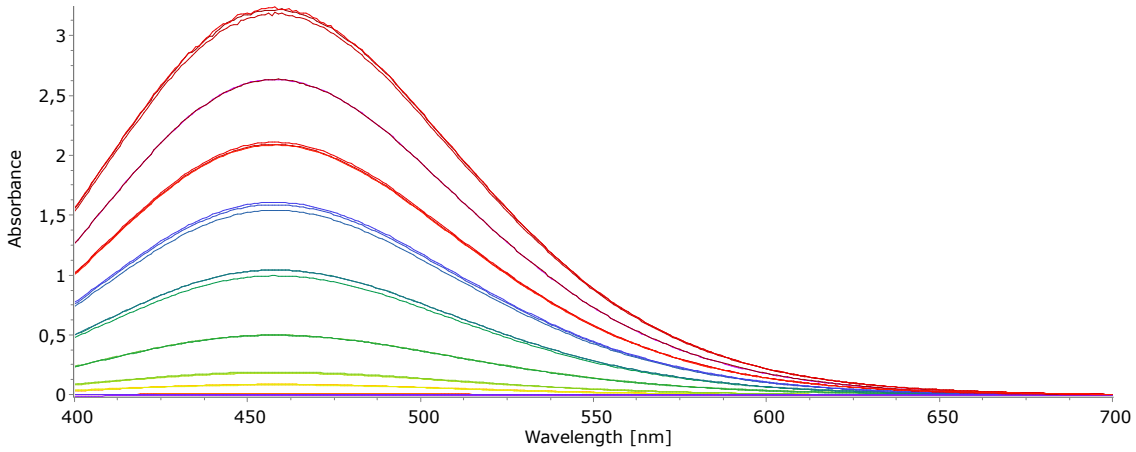


Figure 76. Vis spectra of thiocyanate standards ($FeSCN^{2+}$ complex) at various concentration levels.

The calibration curve for thiocyanate determination was constructed in the form of y = b ($\pm S_b$) x + α ($\pm S_\alpha$), where slope is indicated as 'b' (S_b: random error of slope) and intercept indicated as 'α' (S_α: random error of intercept). Using standard SCN^- solutions prepared in artificial saliva, a concentration range from 0.01 to 1.5 mM (ten concentration levels, three replicates each) was covered. The calibration curve, along with confidence and prediction intervals, is presented in Figure 77a. The residuals plot in Figure 77b confirms the absence of a funneling/trumpet-shaped pattern at higher concentration levels, indicating that the homoscedasticity assumption was met. Furthermore, the horizontal lines ($\pm\ t(0.05, [\text{conc. levels}]\text{-}2) \times S_{y/x}$) in the residuals plot define the deviation limits for each individual data point, confirming that no outliers were detected in the measurements. This ensures the reliability and robustness of the calibration model for thiocyanate quantification in saliva samples.

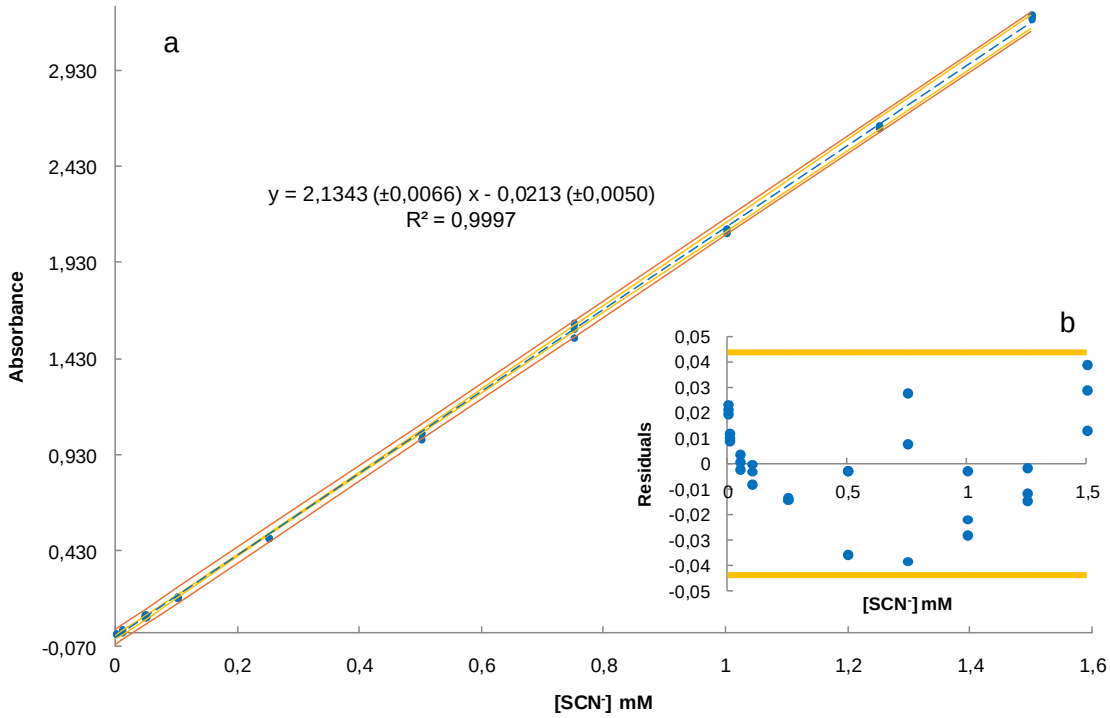


Figure 77. (a) Calibration curve of thiocyanate with confidence and prediction 95% intervals, (b) regression residuals plot with deviation limits.

The calibration curve for thiocyanate quantification was evaluated using linear regression analysis, and the statistical outputs are summarized in Table 9. The results confirm a strong linear relationship between absorbance and thiocyanate concentration, with a coefficient of determination (R^2) of 0.9997. This indicates that 99.97% of the variability in absorbance is explained by the model, demonstrating an excellent fit.

The ANOVA (Analysis of Variance) results validate the statistical significance of the regression model. The calculated F-value of 105079.18 is substantially greater than the critical F-value (4.3512434) at the given degrees of freedom, with an extremely low p-value of 1.35×10^{-51}, confirming that the linear regression model is highly significant. The small residual sum of squares (SS_{RES} = 0.0101) and standard error (0.01899) further indicate a precise fit of the experimental data to the regression model.

Table 9. Regression output of thiocyanate calibration curve.

Regression Statistics	
Multiple R	0.999866794
R Square	0.999733605
Adjusted R Square	0.999724091
Standard Error	0.018990959[a]
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	37.89749192	37.89749192	105079.1775	1.35487E-51
Residual	28	0.010098383	0.000360657		
Total	29	37.9075903			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.021333701	0.004970841[b]	-4.291768989	0.000191323
X Variable 1	2.134258228	0.006583982[c]	324.1591855	1.35487E-51
	Lower 95%	*Upper 95%*		
Intercept	-0.031516007	-0.011151395		
X Variable 1	2.120771552	2.147744903		

[a] $S_{y/x}$ = Residual standard deviation, [b] S_α = Intercept standard deviation, [c] S_b = Slope standard deviation

The regression coefficients provide additional insights into the model's performance. The intercept (-0.0213 ± 0.0050) is close to zero, suggesting minimal systematic error in the calibration. The slope of 2.1343 ± 0.0066 confirms the strong proportional relationship between thiocyanate concentration and absorbance. The confidence intervals (95%) for both the intercept and slope further demonstrate the reliability and precision of the estimated parameters.

Overall, the results confirm that the calibration curve follows a highly linear trend, fulfilling the assumptions for reliable thiocyanate quantification. The high *F*-value, low residual error, and strong R^2 support the model's reliability, making it well-suited for the analysis of thiocyanate in saliva samples.

18.2 Method validation

To ensure the reliability and robustness of the developed method, further validation was conducted by determining the limits of detection (LOD) and quantification (LOQ). These values were calculated using the standard formulas (Eq. 9):

$$LOD = 3 \times \frac{\sigma}{b}, \quad LOQ = 10 \times \frac{\sigma}{b} \tag{9}$$

where b is the slope of the calibration curve (2.1343) and σ represents the standard deviation of the response. The standard deviation was estimated using three different approaches: (i) the residual standard deviation of the regression ($S_{y/x}$), (ii) the standard deviation of the intercept (S_{α}), and (iii) the standard deviation of blank measurements. The lowest values obtained were considered as the theoretical LOD and LOQ, which were then experimentally confirmed.

The analytical parameters of the developed thiocyanate quantification method are summarized in Table 10. The method exhibits excellent sensitivity, with a limit of detection (LOD) of 0.004 mM and a limit of quantification (LOQ) of 0.01 mM. This low LOQ allows for the reliable detection of thiocyanate even at very low concentrations, making the method suitable for physiological studies where small variations in thiocyanate levels are of interest. The working range of the method spans from 0.01 to 1.5 mM, covering the expected concentration range in saliva samples collected before and after exercise.

The accuracy of the method was evaluated at three concentration levels (0.1, 0.7, and 1.25 mM). The accuracy at the lowest concentration (0.1 mM) was slightly overestimated (110.22%), which could be attributed to matrix effects or a higher relative impact of small instrumental variations at low concentrations. However, at higher

concentrations, the accuracy improved significantly, with values of 98.36% (0.7 mM) and 97.57% (1.25 mM), indicating that the method provides reliable quantification across most of its working range.

Table 10. Analytical parameters of thiocyanate quantification method (n=20).

	Analytical parameters	
LOD (mM)	0.004	
LOQ (mM)	0.01	
Working range (mM)	0.01 – 1.5	
Uncertainty (%)	4.51	
Accuracy (%)	0.1 mM	110.22
	0.7 mM	98.36
	1.25 mM	97.57
Intra-day repeatability (% RSD)	0.1 mM	3.70
	0.7 mM	0.57
	1.25 mM	0.54
Inter-day repeatability / reproducibility (% RSD)	0.1 mM	3.13
	0.7 mM	0.83
	1.25 mM	0.56

Precision was assessed through intra-day and inter-day repeatability studies, expressed as the relative standard deviation (% RSD). The intra-day repeatability was highly satisfactory, with RSD values below 4% for all tested concentrations (3.70% at 0.1

mM, 0.57% at 0.7 mM, and 0.54% at 1.25 mM), demonstrating excellent consistency within a single day of analysis. The inter-day reproducibility was also very good, with RSD values of 3.13% at 0.1 mM, 0.83% at 0.7 mM, and 0.56% at 1.25 mM, indicating that the method remains robust over multiple days of analysis.

Overall, these results confirm that the developed method is highly sensitive, accurate, and precise for thiocyanate quantification in saliva. The minor variability observed at the lowest concentration does not significantly impact its reliability, and the method's strong reproducibility makes it a suitable tool for monitoring thiocyanate variations during physical exercise.

18.3 Thiocyanate determination during increased exercise intensity

Aiming to examine the effect of increasing exercise intensity on salivary thiocyanate levels in male and female athletes, a controlled treadmill exercise protocol was conducted with systematic saliva sampling. The goal was to establish how thiocyanate concentrations respond to progressive physical exertion. This investigation builds upon previous findings and provides deeper insight into the kinetics of this biomarker under controlled conditions.

Saliva samples were collected from 11 male and 10 female athletes at four time points: at rest, after running 1 km at 20% VO_2max, 1 km at 60% VO_2max, and 1 km at 90% VO_2max. The violin plots in Figure 78 illustrate the distribution of thiocyanate concentrations in both groups at the different intensities.

For both male and female athletes, a progressive decrease in thiocyanate concentration was observed with increasing exercise intensity. At rest, thiocyanate levels were highest, with a noticeable decline after the first stage of exercise (20% VO_2max). This trend continued as the exercise intensity increased, reaching the lowest thiocyanate concentrations at 90% VO_2max. However, variability among individuals was evident, as

indicated by the spread of values in the box plots, highlighting inter-individual differences in thiocyanate response to exercise.

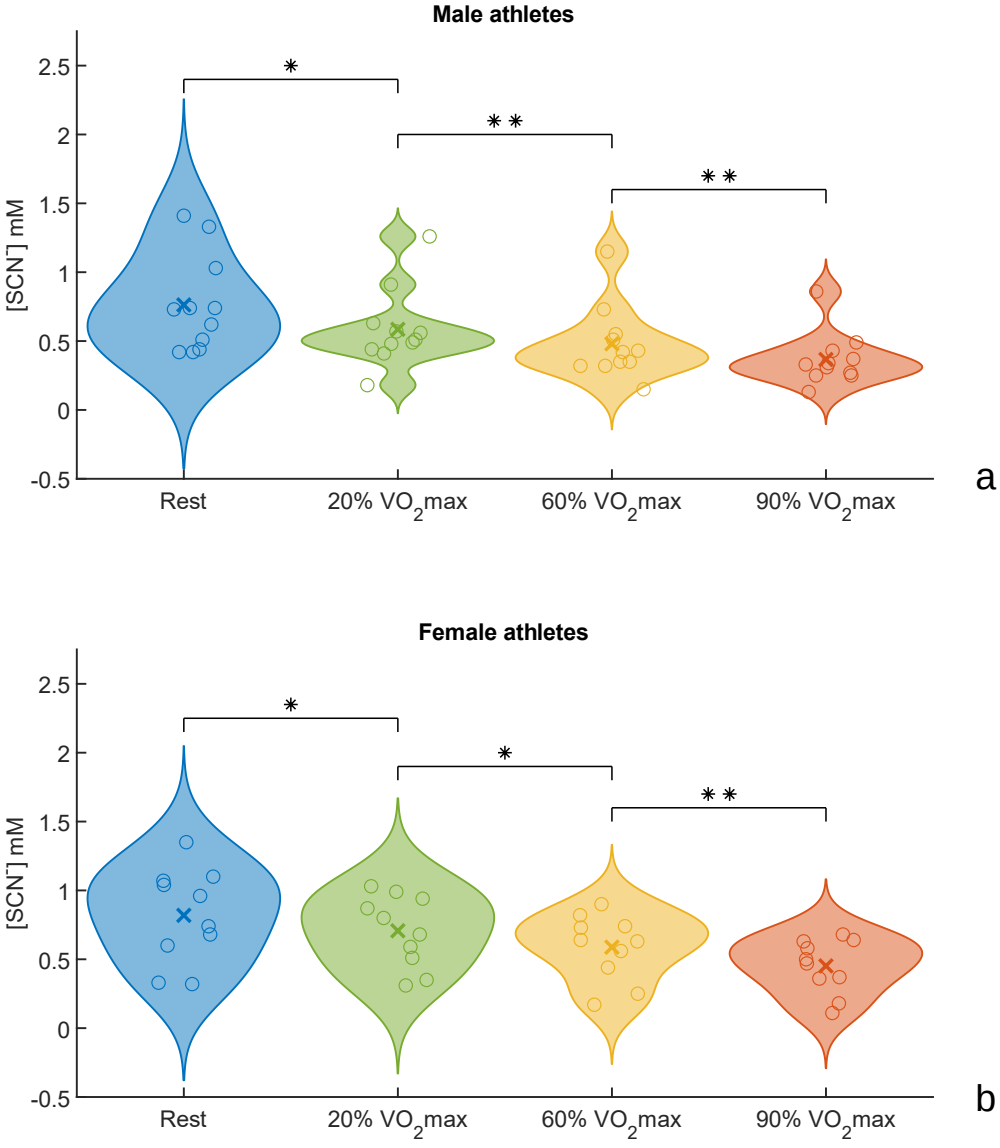


Figure 78. Thiocyanate variations in saliva, during increase of exercise intensity: (a) male athletes and (b) female athletes. Circle points represent the measured values and X represents the mean of each group. **p*-adjusted <0.05, ***p*-adjusted <0.01.

To assess the significance of thiocyanate alterations during exercise, a repeated-measures ANOVA was performed separately for male and female athletes (Table 11). The analysis revealed a statistically significant effect of exercise intensity on salivary thiocyanate levels in both groups ($p < 0.05$), confirming that thiocyanate concentration is affected by increasing exercise intensity.

Table 11. ANOVA (*p*-values) and Post Hoc - Bonferroni (adjusted *p*-values <0.05 are underlined and <0.01 are double-underlined) results across exercise intensity groups.

	ANOVA	Post hoc test (Bonferroni)		
		Rest – 20% VO_2max	20% - 60% VO_2max	60% - 90% VO_2max
Men	0.013949	0.016197	0.001770	0.001243
Women	0.023837	0.009487	0.015651	0.001174

Further pairwise comparisons were conducted using the Bonferroni post-hoc test to identify which specific exercise intensities led to significant changes in thiocyanate concentrations, providing deeper insight into between SCN^- and physical exercise.

As a result of the post-hoc Bonferroni test, in the male athlete group, the comparison between the resting state and the 20% VO_2max condition yielded a statistically significant reduction in thiocyanate concentration with a *p*-adjusted value <0.05 (Table 11). In contrast, the subsequent intensity transitions—20% to 60% VO_2max and 60% to 90% VO_2max—demonstrated even more pronounced reductions, both achieving statistical significance with *p*-adjusted values <0.01.

On the other hand, in the female athlete group, a slightly different pattern was observed. The first two comparisons—rest to 20% VO_2max and 20% to 60% VO_2max—

both showed statistically significant changes with *p*-adjusted values <0.05. However, only the final transition from 60% to 90% VO_2max resulted in a thiocyanate concentration difference that met the stricter significance threshold of *p*-adjusted <0.01 (Table 11).

These results demonstrate that thiocyanate concentration in saliva decreases progressively as exercise intensity increases, supporting the hypothesis that this biomarker is sensitive to physical exertion.

While both male and female athletes exhibited a significant reduction in salivary thiocyanate levels, differences in the magnitude of decline were noted. The violin plots suggest a more pronounced decrease in male athletes compared to females. This could be attributed to physiological differences such as higher metabolic demands[358], or respiratory adjustments in males[359]. Additionally, differences in salivary flow rate and gland size between genders[360,361] could also influence the extent of thiocyanate depletion.

This observed decreasing trend in salivary thiocyanate concentration with increasing exercise intensity aligns with the findings of Phase 2, where ATR-FTIR spectroscopy revealed a progressive reduction in the intensity of the IR band associated with the thiocyanate group (approximately 2050-2060 cm^{-1}) as exercise intensity increased. The coherence between the spectroscopic fingerprint and the quantitative photometric data supports the hypothesis that exercise induces a measurable depletion of thiocyanates in saliva. This concordance not only validates the utility of the developed photometric method but also reinforces the role of salivary thiocyanate as a potential non-invasive biomarker for exercise-induced oxidative stress and related immune changes.

The findings from this controlled treadmill experiment provide robust evidence that thiocyanate is a biomarker responsive to exercise intensity. The significant reductions observed at increasing VO_2max levels suggest that thiocyanate is systematically affected by physical activity, reinforcing its potential as a non-invasive biomarker for monitoring physiological stress during exercise. However, while these results are highly promising, a

larger cohort is necessary to validate these findings and account for potential inter-individual variability more comprehensively.

Building on this controlled laboratory study, the next step was to examine thiocyanate fluctuations in a real-world training environment, assessing athletes from various sports before and after their typical training sessions. This approach allowed for a broader evaluation of thiocyanate as an exercise-induced biomarker across different types of physical activity.

18.4 Insights from a large athletic cohort

This part of the study aimed to monitor thiocyanate as a salivary biomarker in a large number of participants, ensuring more reliable and robust results. A high sample size is particularly valuable in biomarker studies, as it reduces the impact of individual variability and enhances statistical power. In smaller sample sizes, interindividual differences—such as metabolism, hydration status, or lifestyle habits—may overshadow true biological trends. However, with a substantial number of participants, these variations become less influential, allowing for more generalizable conclusions regarding the responsiveness of thiocyanate to exercise. By investigating a diverse cohort of athletes, this study provides stronger evidence for the potential of thiocyanate as an exercise-sensitive biomarker.

To assess the impact of physical activity on salivary thiocyanate concentration, athletes from various sports disciplines, including football, basketball, tennis, pole dancing, aerial hoops, and aerobic/fitness programs, were recruited. Each participant provided a saliva sample before and immediately after their typical training session. A total of 141 athletes participated in the study, comprising 77 men and 64 women. To further investigate individual differences, participants were categorized into four

subgroups based on gender and smoking habits: male smokers, male non-smokers, female smokers, and female non-smokers.

A paired *t*-test was applied within each group to compare thiocyanate levels before and after exercise. The paired approach was chosen because each participant served as their own control, allowing for a direct comparison of the same individual's salivary thiocyanate concentration before and after exercise. Moreover, the saliva samples collected before and after exercise are inherently linked, representing measurements from the same individuals at two different time points. This approach accounts for intra-individual variability, enhancing the sensitivity of the analysis by focusing on changes within each participant rather than differences between independent groups. The results revealed a highly significant decrease in thiocyanate levels among male smokers (p <0.0001), male non-smokers (p <0.0001), and female non-smokers (p <0.001), highlighting a consistent trend of thiocyanate depletion due to physical exertion.

Interestingly, the female smokers group did not exhibit a statistically significant change in thiocyanate levels following exercise. This could be attributed to the lower number of individuals in this group, which may have limited statistical power. Additionally, the higher baseline thiocyanate levels in smokers—due to the well-documented thiocyanate accumulation from tobacco smoke[338]—might have masked any potential exercise-induced depletion.

Figure 79 below illustrates individual changes in thiocyanate concentration for each group, with the mean values before and after exercise represented by the yellow lines. Among male smokers, the mean thiocyanate concentration decreased from 0.98 mM to 0.70 mM, while male non-smokers showed a reduction from 0.86 mM to 0.62 mM. Similarly, female non-smokers experienced a decline from 0.77 mM to 0.66 mM. As mentioned above, female smokers exhibited no significant alteration, with a mean value remaining approximately constant at 1.03–1.00 mM.

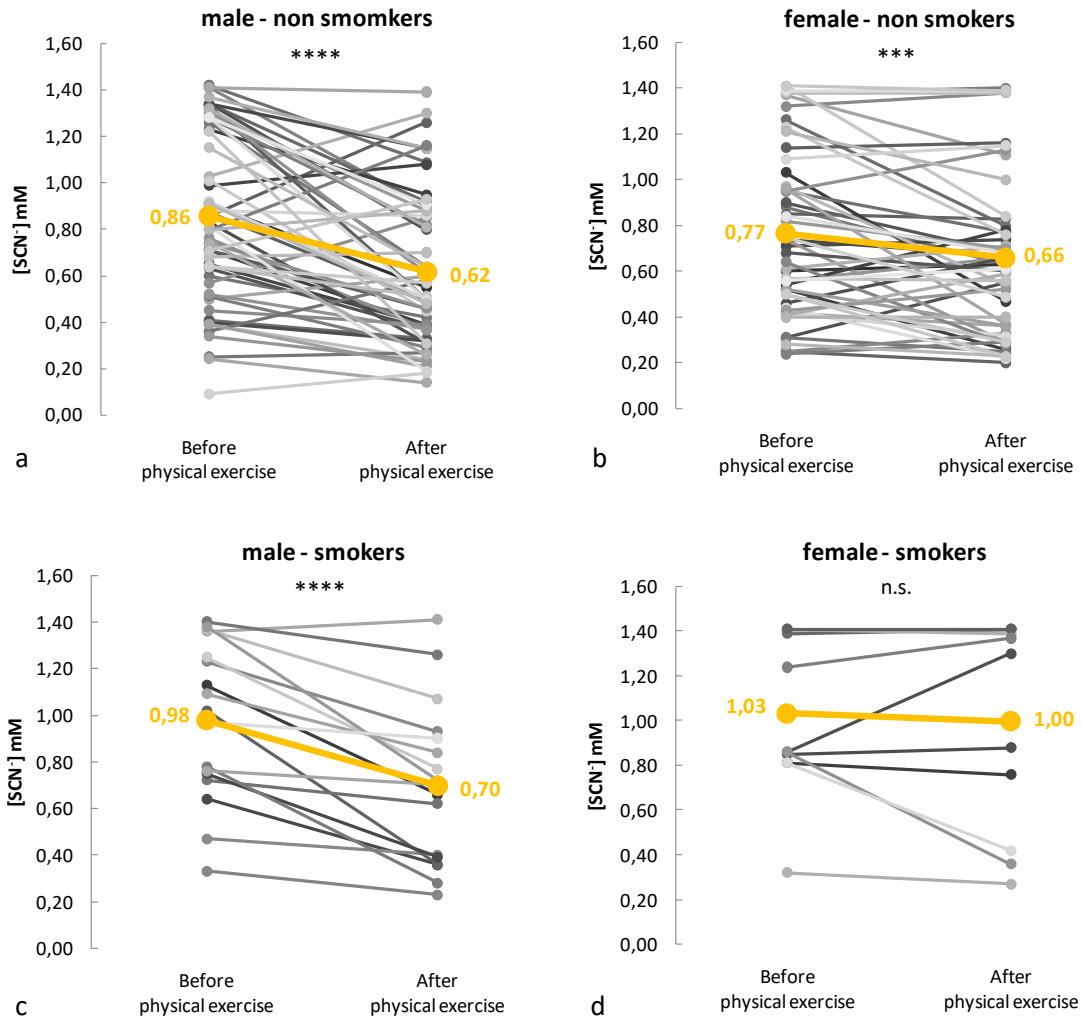


Figure 79. Individual variations in salivary thiocyanate concentration before and after physical exercise in (a) male non-smoker (n=59), (b) female non-smoker (n=52), (c) male smoker (n=18) and (d) female smoker athletes (n=12). The yellow highlighted lines represent the mean values of each group. *p*-values for pre/post paired comparisons are indicated as: n.s. (non significant), **p* <0.05, ***p* <0.01, ****p* <0.001, *****p* <0.0001.

The analysis revealed that both male and female smokers exhibited higher initial salivary thiocyanate concentrations compared to their non-smoking counterparts. This finding aligns with the well-documented association between tobacco exposure and elevated thiocyanate levels, as thiocyanate is a major detoxification product of cyanide

found in cigarette smoke[338,340,362]. The consistently higher baseline concentrations across both sexes suggest that smoking significantly influences thiocyanate homeostasis.

A notable observation in this study is the greater decrease in thiocyanate levels among male compared to female following exercise. Specifically, thiocyanate concentrations in male non-smokers decreased 0.24 mM, whereas in female non-smokers, the reduction was more moderate, 0.11 mM. This difference suggests potential physiological and biochemical distinctions between males and females in response to exercise-induced oxidative stress[363]. This observation aligns with the previously mentioned findings in paragraph 18.3.

While the exact mechanisms underlying this disparity remain to be fully elucidated, the findings suggest that sex-specific physiological factors should be considered when using thiocyanate as a biomarker for exercise-induced stress. Future studies could further investigate these differences by incorporating additional biochemical markers and exploring the effects of varying exercise intensities on thiocyanate dynamics in both males and females.

Overall, these findings suggest that thiocyanate responds dynamically to physical exercise, particularly in non-smokers, where the reduction is more pronounced. The differences between smokers and non-smokers emphasize the role of lifestyle factors in modulating salivary biomarkers, potentially influencing their sensitivity and reliability as indicators of physiological stress. The results further support the potential application of thiocyanate as a non-invasive biomarker for exercise-induced oxidative stress.

These findings provide strong evidence that salivary thiocyanate levels consistently decrease following physical exercise across a large and diverse athletic population. By incorporating a substantial number of participants and accounting for key lifestyle factors such as smoking, this study minimizes the impact of individual variability and enhances the reliability of thiocyanate as a novel biomarker. The observed significant reductions reinforce the association between exercise-induced oxidative stress and thiocyanate metabolism.

18.5 Biochemical pathway of exercise-induced thiocyanate reduction

Thiocyanate (SCN^-), an acidic pseudohalide thiolate, is found in high concentrations in saliva, ranging from 0.5 to 3 mM, making saliva the richest body fluid in terms of SCN^- content[339]. In non-smokers, levels typically range from 0.5 to 2 mM, but in heavy smokers, they can reach up to 6 mM due to dietary and environmental exposure[340]. Thiocyanate acts as a precursor for antimicrobial agents in the presence of salivary peroxidase enzymes such as lactoperoxidase[341]. The lactoperoxidase (LPO) system, which consists of SCN^-, hydrogen peroxide (H_2O_2), and the LPO enzyme, plays a critical role in the innate immune defense of saliva, producing antimicrobial molecules such as hypothiocyanite ($OSCN^-$)[364].

During physical exercise, biochemical interactions involving SCN^- and peroxidase activity[365] underscore its dynamic behavior as a biomarker. Recent studies observed that salivary peroxidase activity in athletes increased immediately following an intense treadmill run but gradually returned to baseline within an hour post-exercise[366,367]. Thus, the reduction in salivary SCN^- levels during increasing exercise intensity can be attributed to its oxidative conversion, aid by lactoperoxidase. Hydrogen peroxide, generated by the action of lactoperoxidase, reacts with SCN^- to produce antimicrobial products such as $OSCN^-$, O_3SCN^-, O_2SCN^-, and HOSCN[356].

This oxidative pathway is stimulated by salivary lactate, which is known to increase significantly during higher exercise intensities[125]. Lactate accumulation during anaerobic metabolism enhances the availability of H_2O_2[368,369], via a flavin-dependent lactate oxidase pathway[370], which H_2O_2 is required for the oxidation of SCN^- into antimicrobial products[356]. The interplay between lactate metabolism and salivary peroxidase activity[356,357] suggests a biochemical link influencing SCN^- levels. As lactate increases with exercise intensity, it likely amplifies salivary peroxidase activity, boosting H_2O_2 production.

This enhanced peroxidase activity accelerates the oxidation of SCN^-, reducing its concentration in saliva. The above aligns with a recent study where a strong correlation was found between salivary lactate concentration and thiocyanate IR band[355].

These findings suggest that SCN^- reduction is not merely a passive outcome but is actively mediated by exercise-induced biochemical processes involving lactate and salivary enzymatic activity. The observed decrease in SCN^- concentration during physical activity could serve as a functional biomarker of exercise intensity.

To conclude, this Phase establishes, for the first time according to our knowledge, salivary thiocyanate as a biomarker responsive to physical exercise, demonstrating its potential application in exercise physiology and sports science. Future research should further explore the underlying mechanisms and assess the biomarker's utility in different athletic and clinical settings.

Despite the strong evidence supporting thiocyanate as a non-invasive marker of physiological stress, the findings of Phase 3 also revealed considerable interindividual variability, likely influenced by external factors such as dietary intake and overall lifestyle habits. Since thiocyanate levels are known to be affected by exogenous sources (e.g., cruciferous vegetables and almonds) this biochemical variability introduces challenges for its use as a standalone marker in all contexts. These limitations highlighted the need to broaden the metabolic scope of our investigation by including additional endogenous biomarkers that may offer more stable or complementary physiological insights. This motivation led to Phase 4, where systemic metabolic profiling was performed using blood (reflects the systemic circulation with rich metabolic profile – the gold standard matrix in metabolomics), allowing for a more comprehensive and unbiased exploration of exercise-related biochemical changes across multiple metabolites.

19. Phase 4 – Metabolomic profiling of exercise intensity *via* a novel approach of DBS microsampling and proton NMR analysis

To complement the salivary results and capture systemic metabolism, Phase 4 expands the investigation to a novel blood metabolomics approach. While previous Phases focused exclusively on saliva, the need for a complementary biofluid—capable of capturing broader physiological processes—led to the incorporation of capillary blood sampling. To address the challenges of invasiveness, logistics, and sample stability often associated with traditional blood collection, Phase 4 adopted a minimally invasive strategy based on dried blood spot (DBS) sampling. Utilizing DBS coupled with NMR spectroscopy, this Phase focuses on exploring systemic metabolic responses to exercise intensity. This transition, from saliva to blood, allows for a more comprehensive assessment of exercise-induced changes by leveraging the broader metabolic coverage offered by blood as a biofluid.

Importantly, this Phase was also designed to validate and expand upon previously monitored metabolites. By shifting to blood and applying a different analytical platform, we aimed to cross-confirm key exercise-induced metabolic alterations observed in prior studies, and assess whether similar physiological signatures—now measured in blood—could be reliably detected using DBS–NMR. This provided both a methodological advancement and a biological continuity across phases of the research.

The objective of Phase 4 was to explore, for the first time, the use of dried blood spot (DBS) sampling combined with nuclear magnetic resonance (NMR) spectroscopy for monitoring metabolic alterations induced by physical exercise. By combining the practicality of DBS sampling with the analytical precision of NMR spectroscopy, this Phase introduces a minimally invasive and field-friendly method for monitoring exercise metabolism. For the first time, different adsorptive materials evaluated for capillary blood

collection. This Phase aimed to capture systemic metabolic shifts across different exercise intensities and evaluate the method's robustness, repeatability, and capacity to reflect physiological differences between athletes of varying training status.

To address these objectives comprehensively, the study was structured in two parts:

1. Part I focused on repeatability and individual comparison. Two athletes—one professional and one non-professional—were repeatedly sampled across four non-consecutive days at four defined exercise intensities. This design allowed for the assessment of intra-individual consistency in metabolic responses and highlighted potential differences attributable to training level.
2. Part II expanded the analysis to a broader athletic population. Seven additional professional athletes were included, enabling the validation of observed metabolic trends in a larger cohort. This second part aimed to determine whether consistent and generalizable metabolic patterns could be established among trained individuals using the DBS–NMR approach.

The results of this Phase are anticipated to provide deeper insights into athlete-specific adaptations, optimize performance strategies, and further establish the potential of DBS-NMR in sports science and personalized training applications.

19.1 Selection of blood microsampling adsorptive material

To optimize the efficiency and reliability of blood microsampling for downstream NMR analysis, the selection and evaluation of suitable adsorptive materials was a critical initial step in Phase 4. In addition to using a commercial DBS card (Whatman 903), novel sorptive materials inspired by fabric-phase sorptive extraction (FPSE) were synthesized and tested as alternative microsampling substrates. This allowed for a comparative evaluation of material performance in sample handling and compatibility with

downstream NMR analysis, while also exploring the potential for customized, low-cost alternatives to commercial DBS solutions. In this study, four synthesized materials inspired by the FPSE technique and one commercially available material (DBS card, Whatman 903 Protein Saver Card) were evaluated for their potential use in blood sampling.

19.1.1 Characterization of the synthesized materials with SEM analysis

Characterization of the synthesized materials was performed to assess their morphological properties. While all four materials were successfully synthesized, a detailed surface morphology analysis using Scanning Electron Microscopy (SEM) was conducted on the fiberglass filter coated with PEG polymer. SEM imaging was performed on both the uncoated and coated materials to evaluate the impact of the sol–gel modification on the substrate's structure.

SEM analysis provided insights into the surface morphology of the fiberglass filter before and after sol–gel coating. Figures 80a and 80b correspond to the uncoated fiberglass filter at different magnifications. The uncoated fiberglass filter exhibited a porous, fibrous network, characteristic of its structure, which facilitates absorption and diffusion of liquids, making it a suitable substrate for microsampling applications. In contrast, Figures 80c and 80d, which depict the fiberglass filter after sol–gel coating with PEG polymer, show significant morphological changes. The coated material displayed a smoother surface with a thin polymeric layer covering the fibers, indicating successful deposition of the sol–gel network. It is visible that the sol-gel network diffuses deep into the substrate, not just on the surface. It could be likened to the sauce among the spaghetti. This modification is expected to influence the material’s sorptive properties by enhancing adsorption capabilities and selectivity for targeted metabolites.

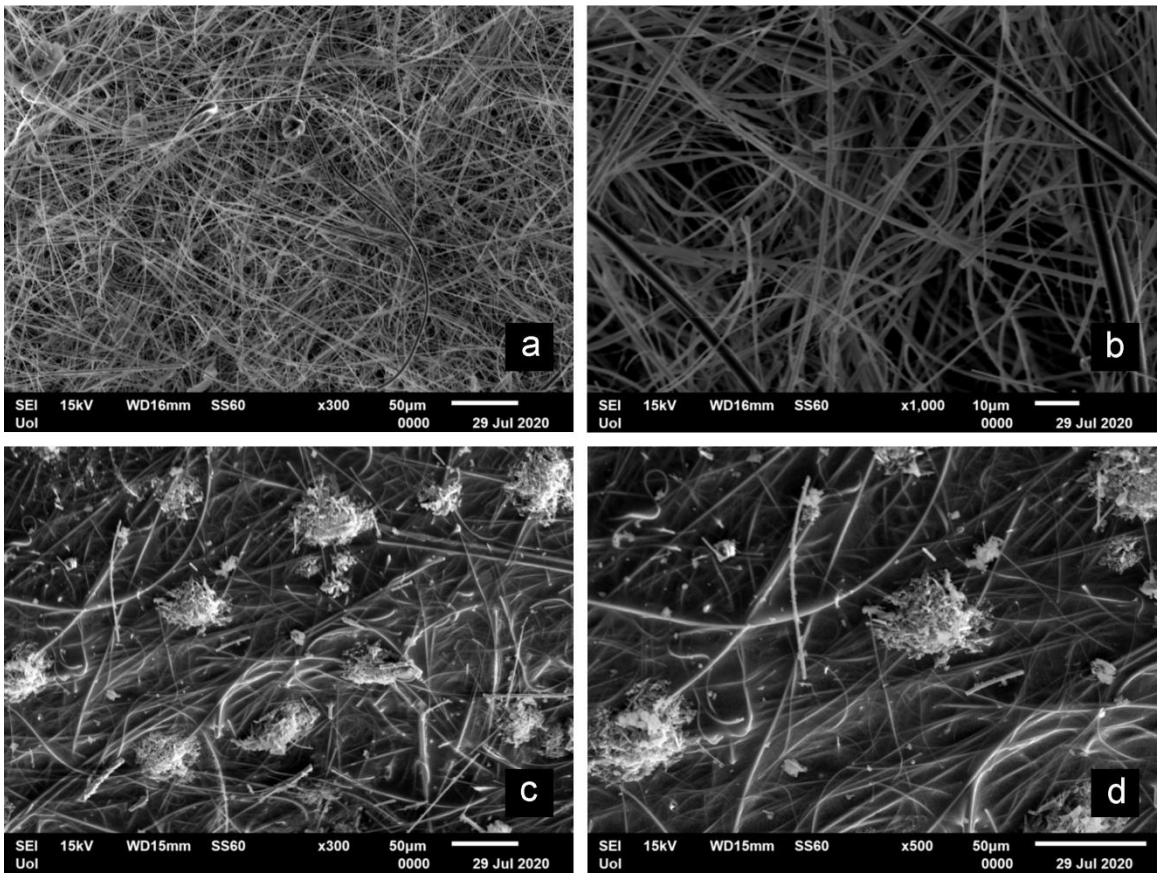


Figure 80. SEM images of the surface of fiberglass fabric before treatment at a) x300 and b) x1000 times magnification; and the surface of fiberglass fabric modified with a sol-gel solution coating of PEG polymer at a) x300 and x500 times magnification.

The comparison between the uncoated and coated materials confirms the successful immobilization of the sol–gel layer while preserving the inherent porosity of the fiberglass substrate. These structural changes suggest that the modification process effectively integrates the polymer and sol–gel network within the substrate, potentially improving the material's performance in blood microsampling applications.

19.1.2 Comparison of adsorptive materials and NMR analysis of blank extracts

The suitability of the synthesized and commercial blood microsampling materials was evaluated by analyzing potential leachable impurities. This assessment is critical, as any compounds released from the sampling material may interfere with downstream metabolomic analysis, masking relevant biological signals. To investigate this, an extraction was performed on blank (unexposed) materials, and the resulting extracts were analyzed *via* NMR spectroscopy.

The NMR spectra of the blank extracts (Fig. 81) revealed significant differences in the chemical profiles of the five tested materials. The four synthesized materials exhibited prominent signals in the 3.6-3.8 ppm region, indicating the presence of impurities leaching from the materials. Furthermore, the two sol–gel coated materials incorporating PEG-PPG-PEG polymer showed additional impurity signals around 1.2 ppm and 3.5 ppm, further suggesting unwanted compound release. These findings raise concerns about the potential impact of such leachables on blood sample integrity, as they may obscure key metabolic signals in upcoming analyses.

In contrast, the commercially available Whatman 903 Protein Saver Card demonstrated a much cleaner spectral profile, with minimal detectable impurities. The absence of significant interfering signals suggests that this material is more suitable for blood microsampling applications, as it reduces the risk of contamination and unwanted spectral overlap in metabolomic studies.

The presence of extractable impurities in the synthesized materials poses a considerable limitation for their application in metabolomic blood analysis. The overlapping signals within critical spectral regions could compromise the detection of endogenous metabolites, thereby reducing the reliability of the analytical results. Given these concerns, the Whatman 903 Protein Saver Card was selected as the preferred material for blood microsampling in subsequent metabolomic investigations. Its low

background interference ensures a more accurate representation of the metabolomic profile, improving the reliability of downstream data interpretation.

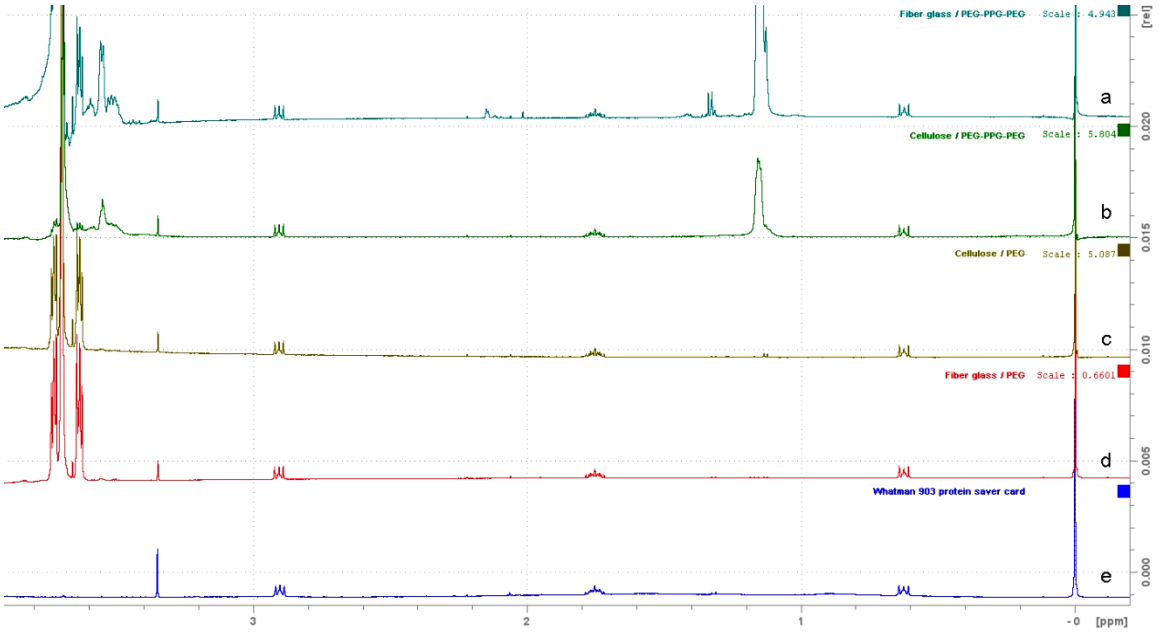


Figure 81. NMR spectra of blank extracts from the five tested blood microsampling materials. (a) Fiber glass filter coated with PEG-PPG-PEG sol-gel, (b) cellulose filter coated with PEG-PPG-PEG sol-gel, (c) cellulose filter coated with PEG sol-gel, (d) fiber glass filter coated with PEG sol-gel and (e) commercially available Whatman 903 Protein Saver Card.

19.2 Quality Control (QC) assessment

To ensure the reliability and reproducibility of the results in Phase 4, two quality control (QC) samples were analyzed throughout the sequence. These QC samples were included to monitor the stability of the NMR spectrometer, assess potential batch effects, and verify the consistency of the metabolic profiling process[371]. By periodically analyzing the same QC samples alongside the experimental samples, any instrumental drift or variability in sample preparation could be identified and accounted for, ensuring the

robustness and accuracy of the generated data. This step was essential for maintaining high-quality standards and ensuring that the metabolic trends observed were truly reflective of exercise-induced changes.

Thus, two additional reference samples (SRs) were prepared by pooling portions of all study samples, serving as part of the quality control (QC) process. Principal Component Analysis (PCA) was applied to evaluate the quality of the data. Specifically, the PCA scores plot included all DBS samples (SS) and the two SRs, with the first two principal components (PC1 and PC2) accounting for approximately 85% of the total variability in the dataset (Fig. 82). The use of PQN (probabilistic quotient normalization) during preprocessing ensured that the data were normalized and comparable across all samples, minimizing the impact of technical variation.

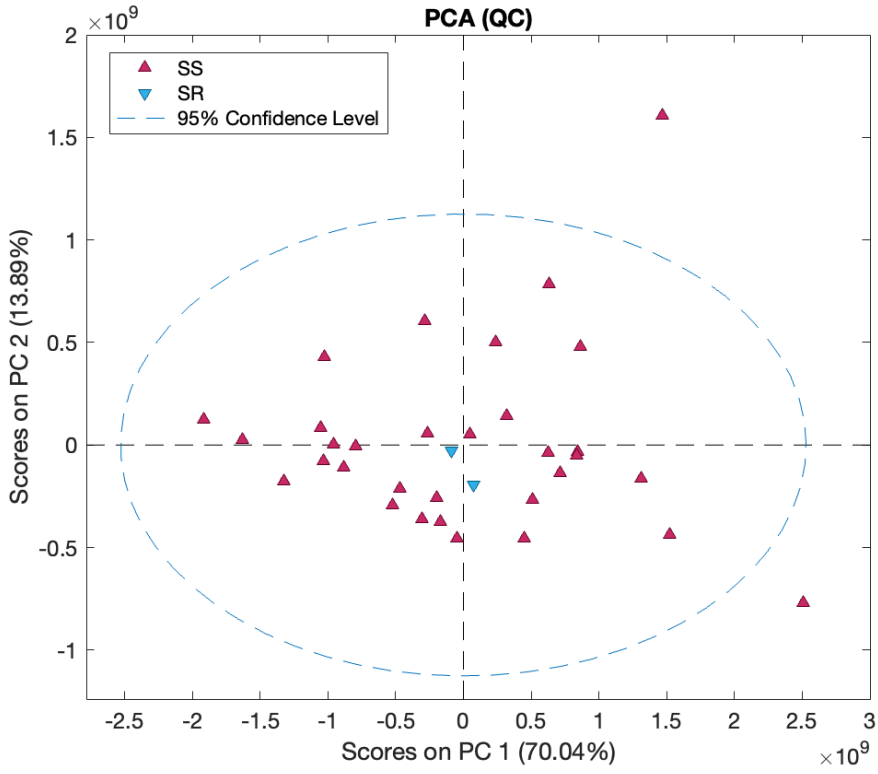


Figure 82. PCA scores plot demonstrating the consistency of reference samples (SRs) in quality control analyses.

The results of the PCA analysis demonstrated the reliability of the QC process. The SR samples clustered tightly (near-overlap) near the center of the scores plot, with minimal variability between them. This tight clustering is a strong indication that the NMR data acquisition and sample preparation steps were consistent and reproducible. The close proximity of the SRs on the PCA plot underscores the high quality of the experimental workflow, ensuring that the datasets produced are reliable for downstream metabolomics analysis in the context of Phase 4.

19.3 Intra-individual repeatability

The PCA quality control assessment confirmed the analytical robustness and biological relevance of the DBS–NMR dataset, providing a solid foundation for more targeted investigations. With confidence in the system's technical reproducibility, we proceeded to a two-part analysis aimed at exploring both intra-individual metabolic consistency and broader population-level trends.

19.3.1 Untargeted metabolomics

In the first part of the Phase, we evaluated the repeatability of exercise-induced metabolic responses across four sessions in two individuals—a professional and a non-professional athlete. This design enabled us to examine whether metabolic alterations were consistent over time and to explore differences in exercise metabolism between differently trained individuals.

This strategy recognizes the unique physiological conditions and responses of each athlete, ensuring that intra-individual variability does not mask the metabolic trends associated with different running speeds. By focusing on individual-specific analyses, the pilot study aimed to improve the reliability of identifying metabolic changes and their correlation with exercise intensity. This approach is particularly advantageous when using DBS microsampling, as it evaluates whether small but significant metabolic shifts can be effectively detected and whether these changes are robustly captured through ^{1}H NMR analysis. The individualized analysis gives an additional layer of precision, laying the groundwork for more targeted investigations in larger cohorts.

Partial Least Squares Discriminant Analysis (PLS-DA), a supervised multivariate method, allows for the classification of NMR profiles according to predefined group labels, such as running speeds. Thus, PLS-DA was performed separately for each athlete to evaluate whether DBS–NMR profiles could reliably distinguish between the different exercise intensities.

The PLS-DA models were constructed using latent variables (LVs) to maximize discrimination between the groups, with the primary focus on LV1 and LV2 to represent the key variance captured in the data. The PLS-DA scores plots (Fig. 83a,b) reveal clear separation of the metabolic profiles corresponding to the four running intensities for both non-professional and professional athletes.

for each running speed (0 km/h [DBS-A0], 5 km/h [DBS-A5], 10 km/h [DBS-A10], 15 km/h [DBS-A15]). LV1 captures 59.45% of the variance, while LV2 captures an additional 29.93%, indicating that the first two components account for nearly 90% of the variability in the dataset. The diagonal progression of clusters from 0 km/h to 15 km/h along both LV1 and LV2 reflects a consistent metabolic shift corresponding to increasing running intensity. The clusters are relatively tight, emphasizing the reproducibility of the metabolic profiles for each condition, with minimal overlap between adjacent running intensities, signifying robust discrimination by the PLS-DA model.

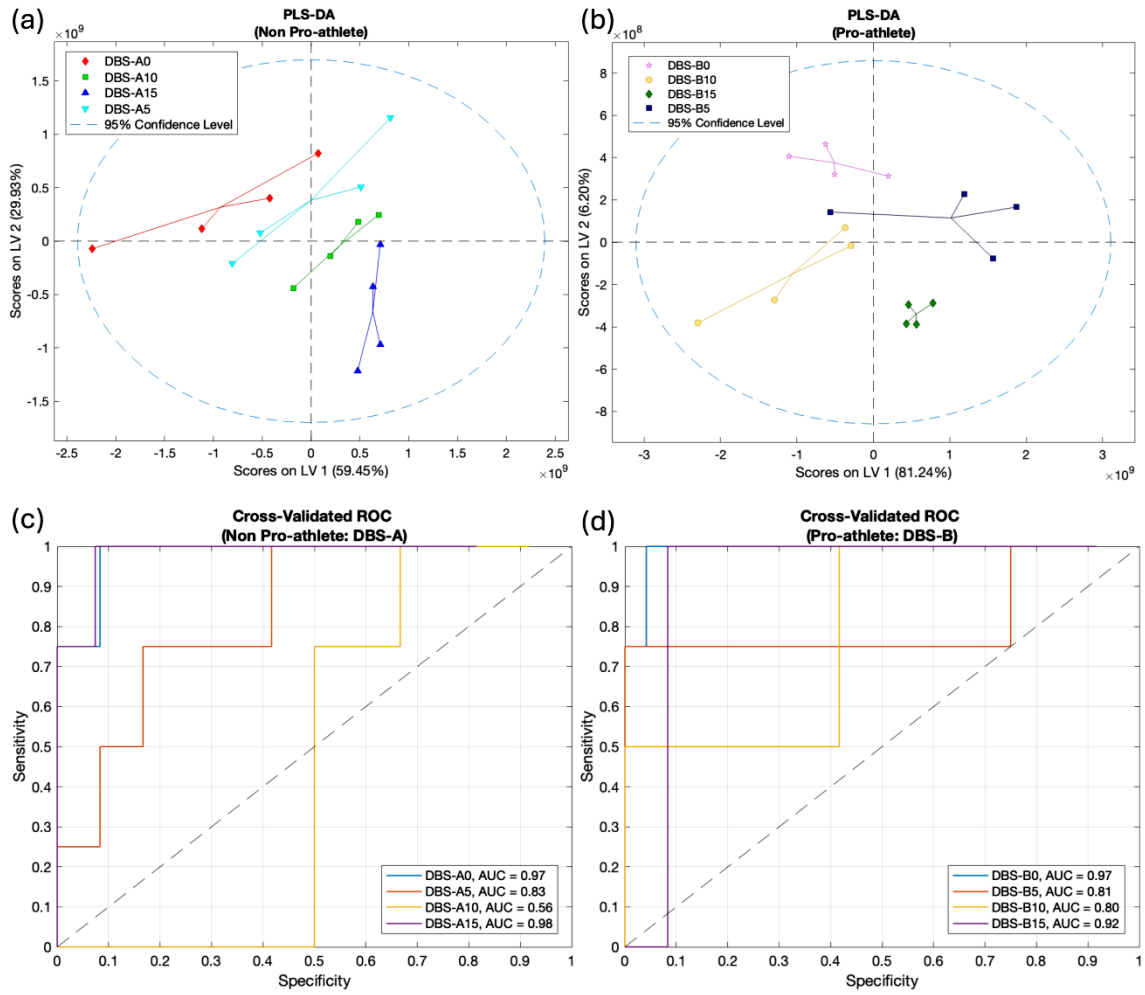


Figure 83. PLS-DA analysis on ^{1}H NMR profiles of DBS extracts. The score plots depict the first two latent variables for (a) non-professional (DBS-A) and (b) professional (DBS-B) athletes across four running speeds: 0 km/h (DBS-A0, DBS-B0), 5 km/h (DBS-A5, DBS-B5), 10 km/h (DBS-A10, DBS-B10), and 15 km/h (DBS-A15, DBS-B15). ROC curves indicating sensitivity and specificity of the model for both (c) non-professional and (d) professional athlete profiles.

For the non-professional athlete, the scores plot (Fig. 83a) shows distinct clustering

For the professional athlete, the scores plot (Fig. 83b) demonstrates a similar pattern of distinct clustering for each running speed (0 km/h, 5 km/h, 10 km/h, 15 km/h).

However, the explained variance in LV1 (81.24%) and LV2 (26.20%) is even higher than that observed for the non-professional athlete, underscoring the greater contribution of these two components to the model's overall discrimination capability. The clusters are well-separated, particularly for 0 km/h and 15 km/h, indicating pronounced metabolic differences at rest and at the highest running intensity. Some kind of overlap particularly in LV2 is observed between intermediate running speeds (5 km/h and 10 km/h), discontinuing the diagonal trend of discrimination, and suggesting metabolic similarities at these intensities or a more gradual physiological adaptation in the professional athlete.

The scores plots for both athletes highlight the effectiveness of PLS-DA in capturing the metabolic distinctions associated with varying running speeds. The separation achieved in the models suggests that the metabolic profiles undergo significant changes in response to physical exertion, with clear trends correlating to intensity. This separation was consistent across all four days, suggesting that the physiological response to exercise, as captured in the DBS metabolome, is robust and repeatable over time.

These findings serve as a robust foundation for further validation, which is explored in the following analysis using cross-validated Receiver Operating Characteristic (ROC) curves to quantitatively assess the classification performance of the PLS-DA models for each running speed in both non-professional (Figure 83c) and professional athletes (Figure 83d). The Area Under the Curve (AUC) values are used as performance metrics to assess the model's ability to differentiate between metabolic profiles corresponding to different running intensities.

The cross-validated ROC curves for the non-professional athlete demonstrate strong classification performance for most running speeds. The AUC values for 0 km/h (resting state) and 15 km/h (highest running intensity) are 0.97 and 0.98, respectively, indicating near-perfect discrimination of these metabolic states from the others. These results highlight the significant metabolic differences between rest and maximal exertion, which align with the distinct clustering observed in the PLS-DA scores plot. Intermediate running speeds, 5 km/h and 10 km/h, show slightly lower AUC values of 0.83 and 0.56,

respectively. The reduced performance for 10 km/h suggests some overlap in the metabolic profiles at moderate intensities, possibly reflecting transitional physiological states. Nevertheless, the overall classification performance remains robust, emphasizing the ability of the model to capture key metabolic variations across different running intensities.

For the professional athlete, the cross-validated ROC curves reveal similarly strong classification performance. The AUC value for 0 km/h (resting state) is 0.97, consistent with the significant metabolic differences between rest and active states observed in the scores plot. The highest running intensity, 15 km/h, also demonstrates excellent discrimination, with an AUC of 0.92. The AUC values for intermediate running speeds, 5 km/h (0.81) and 10 km/h (0.80), indicate good classification performance, albeit slightly lower than the resting and maximal intensity. This result aligns with the partial phenomenon observed in the scores plot for these intensities (described above), suggesting that the metabolic adaptations during moderate exercise intensities in professional athletes are less pronounced compared to the extremes of rest and maximal effort.

Overall, the ROC curve analysis confirms the reliability of PLS-DA models in distinguishing metabolic profiles across running speeds, with high AUC values highlighting significant metabolic shifts, especially at rest and maximal intensity. The slightly lower classification performance at intermediate intensities suggests variability in physiological responses: for non-professionals, this may stem from lower fitness or individual exertion differences, while for professionals, it likely reflects gradual metabolic adaptation due to their advanced training and energy efficiency.

These findings demonstrate the utility of ^{1}H NMR spectroscopy and PLS-DA modeling in capturing exercise-induced metabolic changes. The ROC curve analysis provides additional confidence in the discriminatory power of the models and supports the use of DBS microsampling as a reliable method for studying metabolic adaptations during physical activity. The results set the stage for further exploration of specific

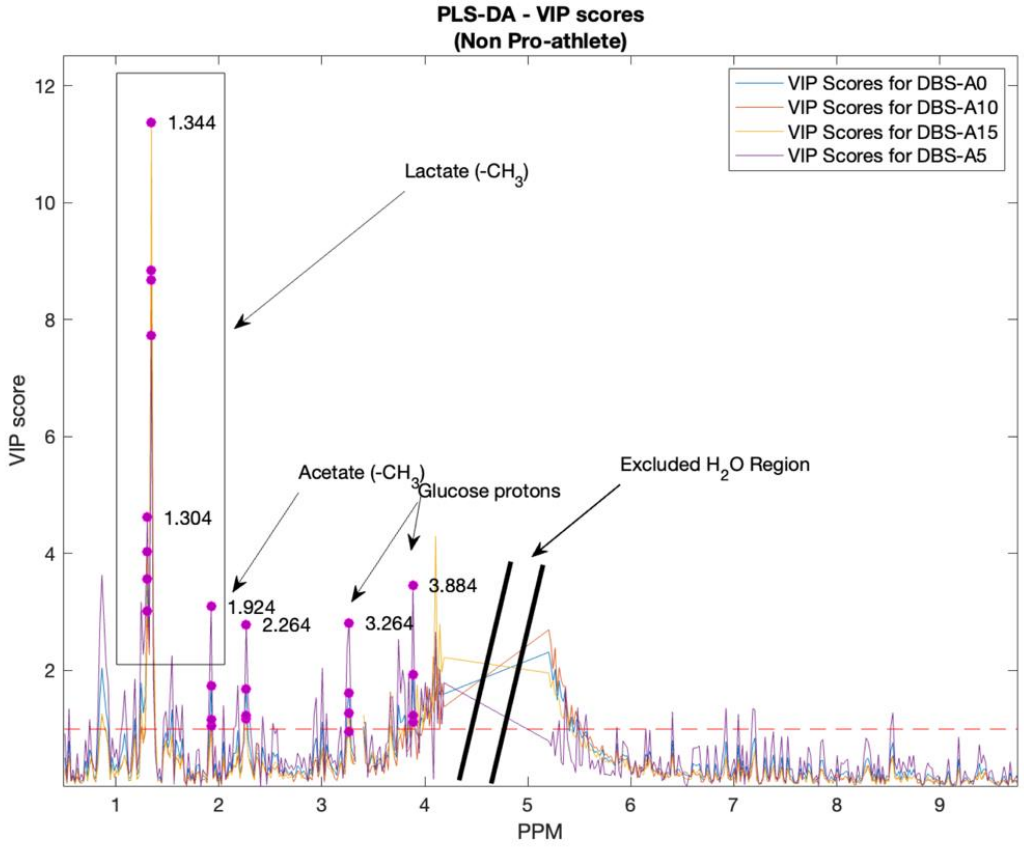
metabolites contributing to these observed trends, which will be discussed in the subsequent sections.

Subsequently, a comprehensive analysis of the Variable Importance in Projection (VIP) provided crucial insights into the discriminative metabolites contributing to the separation of metabolic profiles at varying running speeds for professional and non-professional athletes. In both cases, the spectral bin corresponding to the methyl protons of lactate at approximately 1.3 ppm (Fig. 84) consistently showed the highest VIP scores, confirming its central role in differentiating between exercise intensities. This is consistent with its established role as a key biomolecule of physical activity intensity, aligning with existing literature[372–374].

For non-professional athlete (Fig. 84a), additional metabolic markers, such as acetate (-CH_3) at 1.9 ppm and glucose protons at 3.26 and 3.88 ppm, also displayed elevated VIP scores. These metabolites indicate a broader range of metabolic adaptations in response to exercise for less-trained individuals. In contrast, professional athlete (Fig. 84b) demonstrated a more streamlined metabolic response, with lactate dominating the VIP scores and few additional metabolites contributing significantly to the discrimination. This streamlined response suggests a more efficient metabolic adaptation in professional athletes, potentially due to their enhanced physiological conditioning.

The exclusion of the water region around 4.7 ppm further ensured that the VIP scores were not influenced by artifacts, improving the reliability of the analysis. The application of NMR combined with VIP analysis highlights the potential of this approach in identifying exercise-induced metabolic shifts, with DBS microsampling proving to be an effective method for real-time analysis.

(a)


PLS-DA - VIP scores
(Non Pro-athlete)
VIP Scores for DBS-A0
VIP Scores for DBS-A10
VIP Scores for DBS-A15
VIP Scores for DBS-A5
Lactate (-CH3)
Acetate (-CH3)
Glucose protons
Excluded H2O Region
1.344
1.304
1.924
2.264
3.264
3.884
VIP score
PPM

(b)

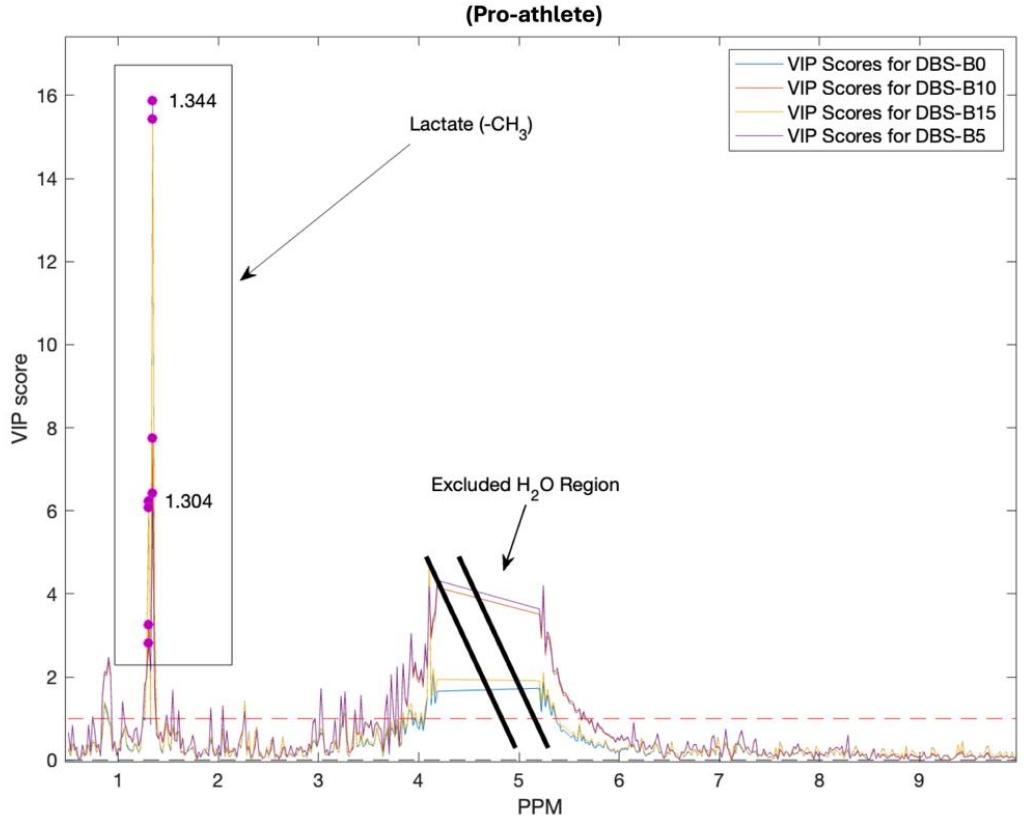

(Pro-athlete)
VIP Scores for DBS-B0
VIP Scores for DBS-B10
VIP Scores for DBS-B15
VIP Scores for DBS-B5
Lactate (-CH3)
Excluded H2O Region
1.344
1.304
VIP score
PPM

Figure 84. Plots of the VIP scores from the two PLS-DA analyses for the (a) non-professional and (b) professional athletes, respectively for across all running speeds—0 km/h (DBS-A0, DBS-B0), 5 km/h (DBS-A5, DBS-B5), 10 km/h (DBS-A10, DBS-B10), and 15 km/h (DBS-A15, DBS-B15). VIP scores above the read dashed threshold reveal the most significant variables (i.e., NMR spectral bins).

19.3.2 Targeted metabolomics – metabolic changes

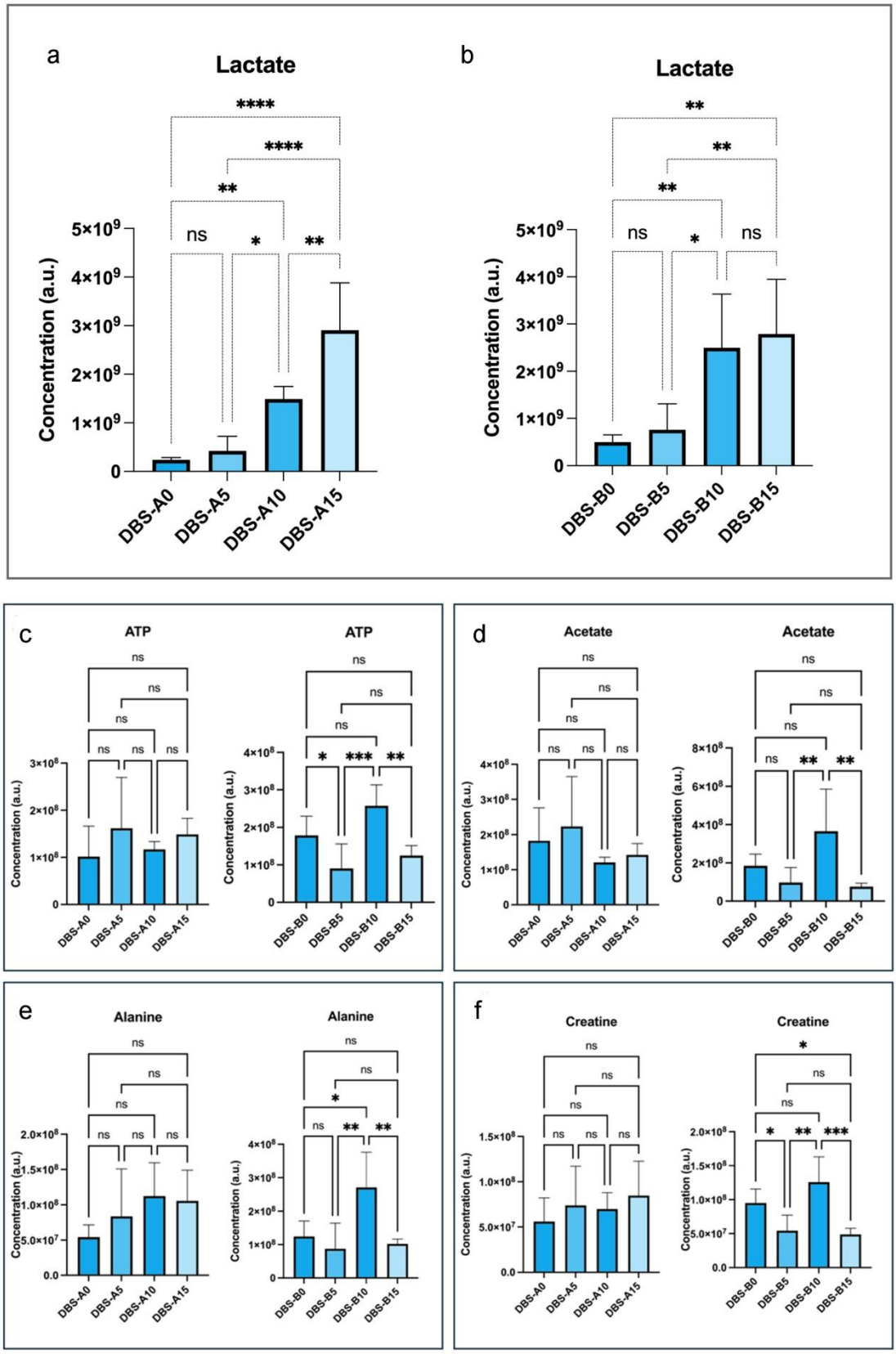
The targeted metabolomics approach conducted in this study enabled a quantitative evaluation of metabolic changes induced by varying exercise intensities, using DBS samples analyzed via ^{1}H NMR spectroscopy. In total, 11 metabolites were identified: ATP (Adenosine-5'-triphosphate), acetate, alanine, creatine, creatinine, formate, glucose, lactate, leucine, pyruvate, and valine. Out of the 11 metabolites reliably detected and quantified from the NMR spectra, lactate emerged as the most significant metabolite showing distinct concentration trends across different running speeds for both the non-professional and professional athletes. The results reinforce the robustness of NMR-based targeted metabolomics in investigating exercise-induced metabolic changes, while also shedding light on the differential physiological adaptations between the two athletic profiles.

The univariate statistical analysis applied in this Phase utilized one-way ANOVA with multiple comparisons to discern the statistical significance of changes in metabolite concentrations across four running speeds (rest, walking, jogging, running). The datasets were PQN-normalized to account for inherent variability in the data, ensuring accurate and reliable comparison of metabolite levels. Lactate demonstrated a statistically significant progressive increase in concentration with increasing exercise intensity, aligning with established metabolic responses to physical exertion.

In Figure 85, the lactate concentrations (presented in arbitrary units, a.u.) for both professional and non-professional athlete at the four running speeds exhibit a steep upward trend, with significant differences noted between successive speeds. The data reveal a statistically significant difference in lactate levels between baseline (0 km/h) and moderate to high-intensity exercise (10 km/h and 15 km/h). The lack of significance (ns) between rest and walking (5 km/h) indicates that at lower exercise intensities, the metabolic shift toward anaerobic glycolysis is not yet pronounced[375]. However, the sharp increase of lactate, in both participants, at 10 km/h and 15 km/h highlights the onset of anaerobic metabolism due to the increased energy demand surpassing aerobic capacity.

In Figure 85b, the professional athlete's lactate concentrations show a more progressive rate of increase with exercise intensity compared to the non-professional athlete. The professional athlete exhibits significant differences in lactate levels across most running speeds, with a more prominent increase evident between baseline and the highest exercise intensities.

Lactate's role as a key biomarker of exercise intensity is well-documented, reflecting the balance between energy demand and oxygen availability during physical activity[376]. The measurement of the blood lactate concentration is widely used for assessing the involvement of anaerobic glycolysis in providing energy for muscle work[375]. The observed differences in lactate dynamics, mostly in 10 km/h, between the two athletes underscore their contrasting metabolic profiles[372]. In general, blood lactate levels of the high-level athlete seem to be increased earlier, in agreement with the recent literature[377]. This finding aligns with the professional athlete's greater glycolytic capacity and ability to recruit fast-twitch muscle fibers, which primarily rely on anaerobic pathways for ATP production under high-intensity conditions[377]. In contrast, the non-professional athlete's slower accumulation of lactate at middle intensities, coupled with a more pronounced increase at higher speeds, indicates a limited glycolytic capacity and a reduced ability to buffer the associated acidosis in blood and muscles[378,379], resulting in earlier fatigue during intense exercise.


a
Lactate
Concentration (a.u.)
DBS-A0
DBS-A5
DBS-A10
DBS-A15
b
Lactate
Concentration (a.u.)
DBS-B0
DBS-B5
DBS-B10
DBS-B15
c
ATP
ATP
d
Acetate
Acetate
e
Alanine
Alanine
f
Creatine
Creatine

(continued)

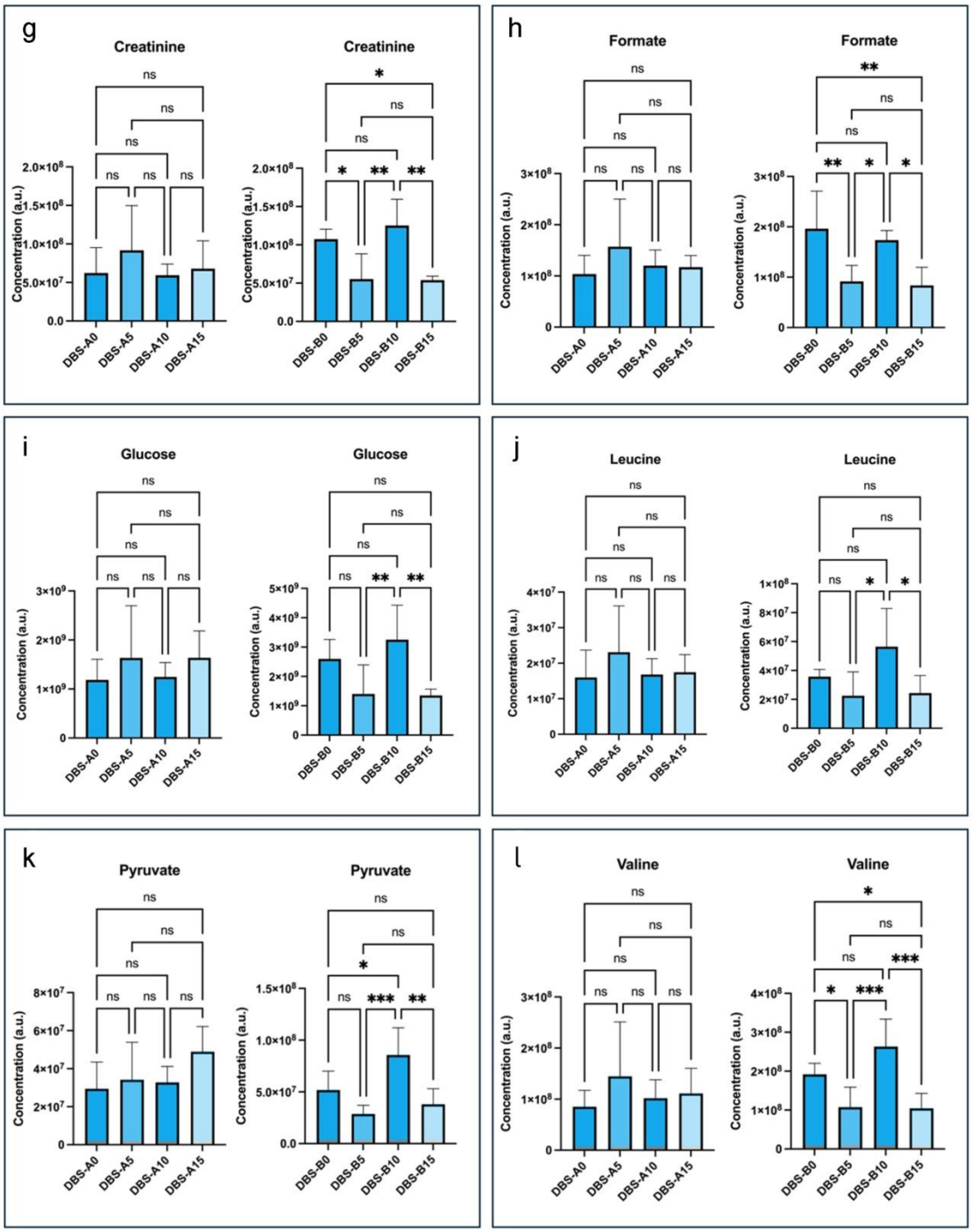


Figure 85. One-way ANOVA with post-hoc multiple comparisons was conducted to analyze metabolite concentrations: (a) Lactate variations of both non-professional (DBS-

A) and (b) professional (DBS-B) athletes. (c) ATP, (d) acetate, (e) alanine, (f) creatine, (g) creatinine, (h) formate, (i) glucose, (j) leucine, (k) pyruvate, and (l) valine variations in dried blood samples across four running speeds: 0 km/h (DBS-0), 5 km/h (DBS-5), 10 km/h (DBS-10), and 15 km/h (DBS-15). Statistical significance is denoted as follows: ns (not significant, $p>0.05$), * ($0.01<p\leq0.05$), ** ($0.001<p\leq0.01$), *** ($0.0001<p\leq0.001$), and **** ($p\leq0.0001$).

In the case of the non-professional athlete, there were no additional metabolites demonstrating statistically significant variations across different exercise intensities, suggesting a more uniform or limited metabolic response beyond lactate. This likely reflects a less complex metabolic regulation system, potentially influenced by lower levels of training adaptation[380]. Evidence is that the professional athlete exhibited a greater exercise-induced metabolic response, including elevated creatinine levels compared to the non-professional athlete (Fig. 85g), a finding consistent with previously reported literature[381].

In professional athlete, various metabolites exhibited fluctuations in concentration, but no clear or consistent trends were evident across different running speeds (Fig. 85c-l). These variable changes might be attributed to temporary factors independent of exercise intensity, such as short-term metabolic adjustments, hydration levels, or dietary intake[58,296]. Nevertheless, certain patterns emerged when comparing resting states to higher-intensity exercise. Notably, alanine concentrations showed a progressive rise with increasing exercise intensity in both athletes[101,363,382] (Fig. 85e). Furthermore, specific trends identified in previous research were confirmed, such as reductions in formate (Fig. 85h), leucine (Fig. 85j), and valine (Fig. 85l) levels, along with an increase in pyruvate levels in both individuals[101,382,383] (Fig. 85k). Additionally, slight elevations in acetate, creatine, and creatinine levels were noted at 10 km/h compared to baseline, consistent with findings from recent studies[383].

Professional athletes exhibit a remarkable ability to perform at significantly higher exercise intensities than non-professionals, a capacity rooted in their advanced physiological adaptations and metabolic efficiency[384]. As running intensities increased up to 10 km/h, trends in blood metabolite concentrations—either increasing or decreasing—were observed and were generally consistent with previously reported findings[101,363,372]. Notably, at the highest running intensity (15 km/h), many of these metabolites either returned to baseline levels or ceased to show statistically significant changes. This phenomenon reflects the superior efficiency of professional athletes' metabolic systems in maintaining homeostasis and stabilizing metabolite levels under extreme physical demand[384]. This regulatory efficiency underscores their ability to tolerate and adapt to the physiological challenges of strenuous exercise, highlighting the role of optimized metabolic pathways in sustaining high-intensity performance.

A similar pattern was evident in lactate dynamics (Fig. 85a,b). In the professional athlete, lactate accumulation began earlier and reached higher concentrations compared to the non-professional athlete, with significant increases noted at 10 km/h. However, during the highest running intensity (15 km/h), lactate levels stabilized rather than continuing to rise. This behavior can be attributed to the professional athlete's capacity to operate at greater intensities, which increases reliance on anaerobic glycolysis (as mentioned above), leading to elevated lactate production. Despite this, professional athletes exhibit highly efficient mechanisms for managing lactate, including enhanced lactate clearance, recycling, and utilization[373]. These mechanisms allow lactate levels to stabilize even under maximal effort[384], preventing excessive acidification of the blood and maintaining muscular performance.

The observed stabilization of both lactate and other metabolites at peak exercise intensities highlights the advanced metabolic flexibility and resilience of professional athletes. Such adaptations include a greater capacity for buffering hydrogen ions, improved oxidative metabolism, and an ability to sustain higher energy demands without overwhelming metabolic pathways. These findings align with the current understanding

of elite athletic performance, where metabolic optimization plays a crucial role in sustaining activity at near-maximal workloads[384]. By contrast, non-professional athletes, with less developed metabolic systems, may struggle to maintain stable metabolite levels under similar conditions, leading to earlier onset of fatigue and reduced performance capacity. This comparison emphasizes the importance of training-induced physiological adaptations in shaping metabolic responses to high-intensity exercise.

Overall, the first part of the Phase 4 demonstrated that DBS–NMR metabolomic profiling is capable of reliably capturing structured and reproducible metabolic changes induced by exercise within individuals, across multiple days. Both the professional and non-professional athlete exhibited intensity-dependent metabolic shifts, with distinct separation across exercise stages.

19.4 Expanded cohort validation

Building on the observations from the initial athlete comparison, this phase was extended to include seven additional professional athletes, forming a cohort of nine in total. The aim of this second part was to validate the previously observed metabolic trends in a larger population of trained individuals and to assess the results across increased inter-individual biological variability. To this end, DBS–NMR data from all nine athletes were analyzed using PLS-DA, to evaluate whether samples could still be clearly discriminated based on exercise intensity. This approach tested the generalizability of the identified metabolic signatures and provided insights into the consistency of physiological responses within a performance-oriented cohort.

The resulting PLS-DA score plot is shown in Figure 86. The model revealed a strong and structured separation among the four exercise intensity levels (before exercise, 5, 10, and 15 km/h). The first latent variable (LV1) accounted for 72.13% of the variance, while

the second latent variable (LV2) explained an additional 15.77%, together capturing nearly 90% of the total variation in the dataset.

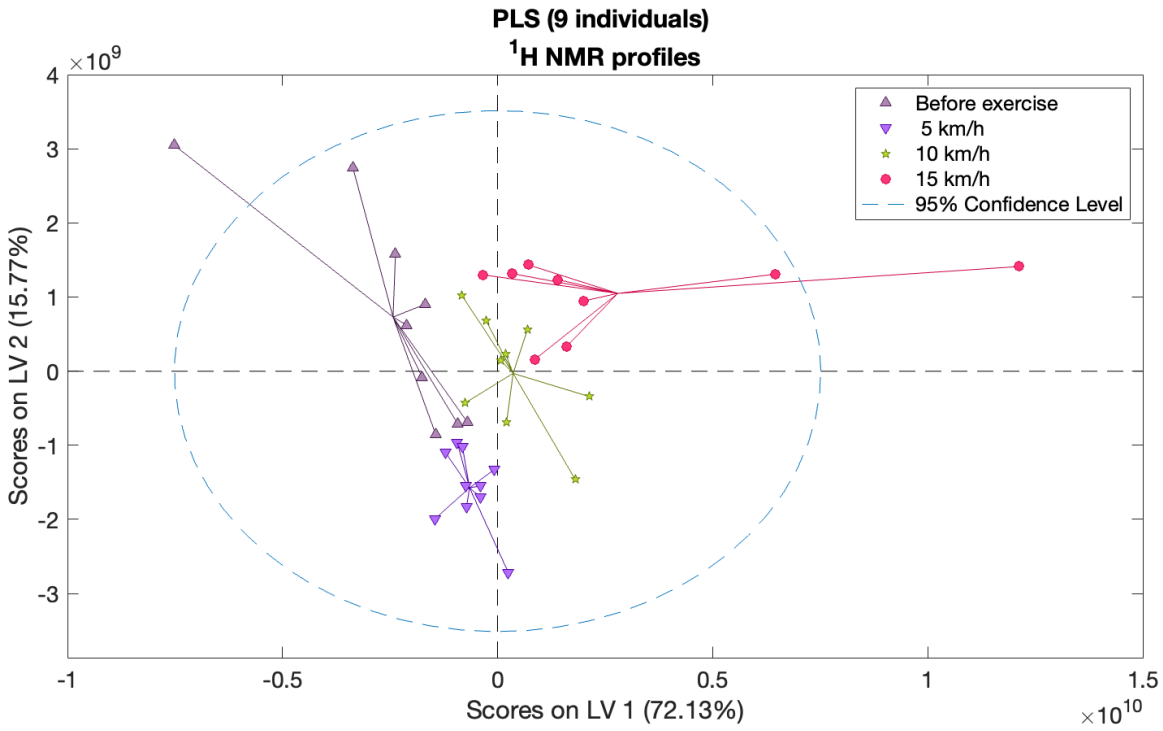


Figure 86. PLS-DA score plot (LV1 vs LV2) of DBS–NMR data from nine athletes across four exercise intensities (0, 5, 10, 15 km/h). Each point represents a DBS sample, colored by intensity.

This high cumulative variance indicates that the majority of the metabolic signal associated with exercise intensity is well-explained by the model. Samples from the resting condition clustered distinctly on the left side of the LV1 axis, progressively shifting rightward through increasing intensity levels, with high intensity exercise (15 km/h) samples forming a separate and compact cluster at the far end. This gradient distribution reflects a consistent and exercise-dependent metabolic response across athletes.

Importantly, despite individual variability in metabolic baselines, samples from the same intensity class tended to cluster closely together, suggesting a shared underlying metabolic response to physical exertion in trained individuals.

The structure of the score plot also suggests that the principal source of variance (LV1) is tightly linked to exercise intensity, while LV2 may reflect inter-individual variability or subtle differences in physiological adaptation. However, the separation remained robust, confirming that exercise load remains the dominant factor shaping the DBS metabolomic profile in this cohort.

To assess the classification accuracy and robustness of the multivariate model, cross-validated ROC analysis was performed for each exercise intensity class using the pooled dataset of nine athletes. ROC curves evaluate the sensitivity and specificity of the PLS-DA model in correctly assigning samples to their respective exercise intensity levels.

As shown in Figure 87, the model achieved excellent performance for most exercise stages. The area under the curve (AUC) was 0.8452 for samples collected at rest (0 km/h), indicating good discrimination of baseline metabolic profiles. Classification performance improved further for low and high intensities: 5 km/h samples yielded an AUC of 0.9704, and 15 km/h samples achieved an AUC of 0.9643, both reflecting high sensitivity and specificity in detecting metabolic shifts at these stages. By contrast, classification performance was less robust at 10 km/h, where the AUC dropped to 0.6230, suggesting greater overlap or heterogeneity in metabolic profiles at this intermediate workload. This finding is consistent with previous observations in part I, where moderate intensities also showed increased variability—potentially reflecting transitional physiological states or inter-athlete differences in aerobic-anaerobic thresholds.

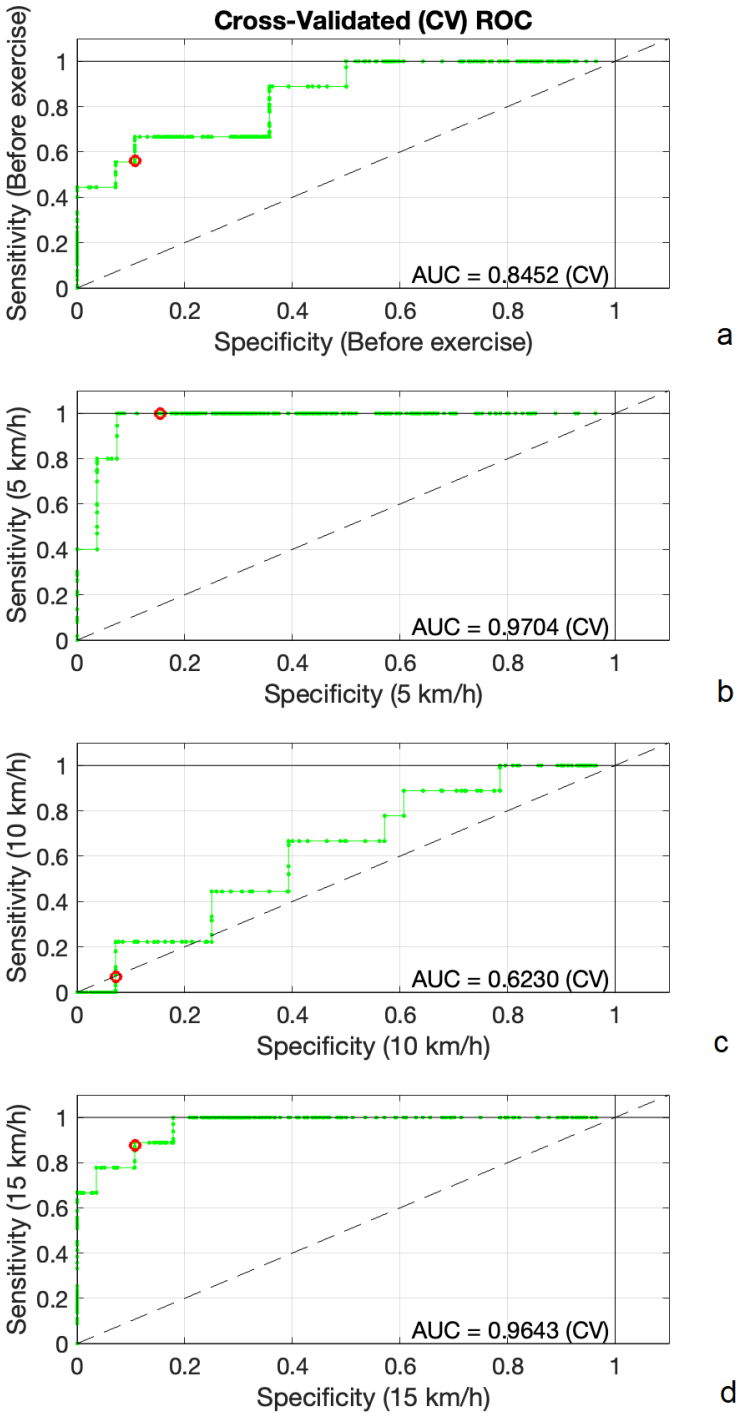


Figure 87. Cross-validated ROC curves for PLS-DA classification of DBS–NMR samples by exercise intensity. (a) Before exercise (0 km/h), (b) low intensity (5 km/h), (c) moderate intensity (10 km/h), and (d) high intensity (15 km/h) clusters.

Despite this localized reduction in discriminatory power, the overall ROC results support the model's strong ability to distinguish metabolic profiles across most exercise intensities. Particularly at rest and maximal exertion, the DBS–NMR method reliably captures systemic shifts, while partial overlap at moderate intensity may reflect meaningful biological variability rather than methodological limitations.

These findings validate the generalizability of the PLS-DA model in a trained population and affirm the diagnostic potential of DBS-based metabolomic profiling for monitoring exercise load.

To identify which metabolites contributed most to the separation across exercise intensities, Variable Importance in Projection (VIP) scores were calculated for LV1, the primary component explaining 72.13% of the total variance. The VIP analysis (Fig. 88) highlights the spectral features most influential in the model's classification performance, offering biological insight into the metabolic changes underlying physical effort.

The most prominent discriminant variable was the spectral signal at 1.33 ppm, which exhibited the highest VIP score across all features. This chemical shift corresponds to lactate, a well-established marker of anaerobic metabolism and muscular energy turnover[370]. Its dominant contribution to LV1 indicates that lactate accumulation is a central and consistent feature of the metabolic response to increasing exercise intensity across athletes. The progressive separation observed along LV1 in the PLS-DA score plot is therefore largely driven by the rising lactate signal with advancing workloads.

In addition to lactate, the region around 3.37 ppm also contributed significantly to sample discrimination. This signal is attributed to glucose, another critical metabolite in exercise metabolism. The involvement of glucose reflects its role in both immediate energy provision and longer-term substrate utilization dynamics[326]. While its contribution was secondary to lactate, the glucose signal nonetheless helped to refine the intensity-dependent clustering, particularly at lower and intermediate workloads where glucose availability and uptake may vary between individuals and over time.

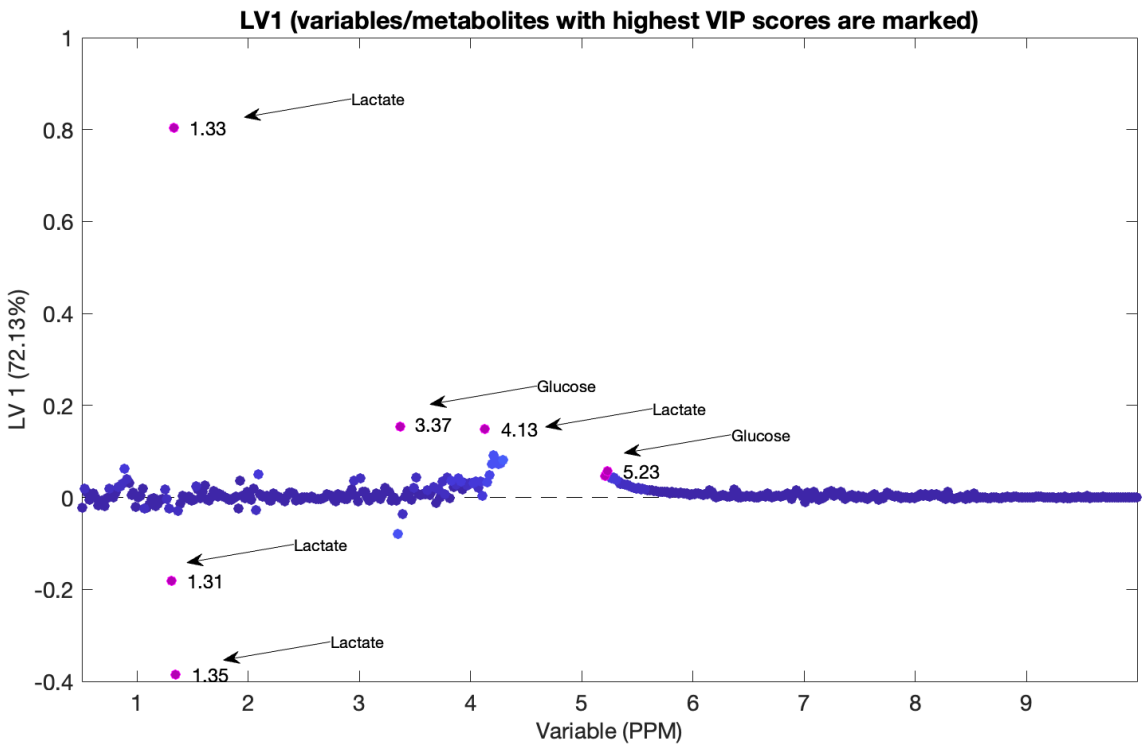


Figure 88. VIP scores from the PLS-DA model (LV1) showing the most influential NMR spectral variables for discrimination across exercise intensities.

Together, these findings confirm that the primary drivers of the DBS–NMR model are physiologically meaningful markers of energy metabolism. The results also demonstrate the ability of the method to capture both aerobic and anaerobic (lactate-associated) components of the exercise response.

The expanded analysis in nine professional athletes confirmed that DBS–NMR metabolomic profiling can sensitively and consistently discriminate between different exercise intensities, even in the presence of inter-individual biological variability. Multivariate modeling revealed a clear intensity-dependent structure, with lactate and glucose emerging as the principal metabolites driving this separation. These results

validate the repeatability and discriminatory power observed in previous part of this Phase, demonstrating that the exercise-related metabolic shifts captured *via* DBS sampling are robust, reproducible, and generalizable across trained populations.

Overall, Phase 4 introduced a novel and powerful analytical approach by combining dried blood spot (DBS) microsampling with nuclear magnetic resonance (NMR) spectroscopy for the metabolomic profiling of exercise responses. This combination proved to be highly effective, offering a minimally invasive, practical, and reproducible method for capturing systemic metabolic changes associated with physical effort. The strength of this method lies in its ability to provide a comprehensive metabolic snapshot from a single, easily collected blood drop, making it uniquely suited for repeated sampling in real-world athletic settings.

However, the Phase is not without its limitations. The small sample size (n=9) restricts the ability to generalize the results to broader populations, and the focus on a limited subset of metabolites may overlook more extensive metabolic networks and interactions. Expanding future research efforts to include larger cohorts and a more extensive range of metabolites, utilizing complementary high-sensitivity techniques such as LC-MS, will provide a more holistic understanding of metabolic shifts during exercise. Additionally, controlling for confounding variables such as diet, hydration, and individual metabolic variability will further enhance the robustness of these findings.

Despite these constraints, the study provides compelling preliminary evidence for the efficacy of this approach in identifying subtle metabolic differences between athlete types and exercise intensities. What sets this Phase apart is not only the robustness of the results, but also the scalability and translational potential of the DBS–NMR technique. This novel combination in athletic context opens new avenues for high-throughput, field-deployable athlete monitoring with the precision of metabolomic resolution—demonstrating the analytical viability of microvolume blood sampling for metabolic profiling during physical exertion—bridging, this way, the gap between laboratory capabilities and practical athletic application. In this context, Phase 4 represents a

significant methodological and conceptual advancement within the study, aiding the salivary analysis by ATR-FTIR spectroscopy and expanding both the analytical framework and the biological understanding of how the human body responds to exercise.

This methodology sets the stage for broader applications in sports science and clinical research, offering a pathway for uncovering intricate metabolic dynamics and advancing our understanding of exercise physiology.

Conclusions

This doctoral thesis presents a comprehensive, multi-phase research effort aimed at developing and validating minimaly-invasive analytical methodologies for monitoring biochemical and metabolic responses to physical exercise. By integrating saliva and dried blood spot (DBS) microsampling with advanced spectroscopic techniques—namely attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy, visible-range photometry, and proton nuclear magnetic resonance (^{1}H NMR) spectroscopy—this work responds to a growing scientific and practical demand for real-time, field-deployable, and physiologically meaningful diagnostic tools. Each of the four experimental Phases targeted a specific aspect of this framework, and collectively, they form a novel and robust approach to the analysis of exercise-induced biochemical changes.

The overarching goal of this research was to explore how easily accessible biofluids, such as saliva and capillary blood, can serve as reliable sources of metabolic information during and after physical exertion. A total of 260 athletes from diverse disciplines participated in the study, including endurance runners, team sport players, combat athletes, and artistic athletes. This heterogeneous cohort ensured a broad representation of physiological responses and training modalities. The wide demographic and sport-specific diversity allowed for greater generalizability of findings and increased the translational value of the proposed methodologies.

Phase 1 established the foundation of this thesis by applying ATR-FTIR spectroscopy to the analysis of post-exercise saliva samples from low- and high-level athletes. The primary aim was to evaluate the feasibility of using salivary spectral profiles to distinguish between levels of physical conditioning. Multivariate statistical analysis, particularly principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-

DA) successfully classified the two groups based on distinctive biochemical patterns. This was the first demonstration of the use of ATR-FTIR saliva fingerprinting as a viable method for assessing physiological adaptations to fitness level. The results highlighted specific spectral regions—associated with proteins, carbohydrates, and carboxylic acids—as differentiating variables, thereby confirming, for the first time, the sensitivity of saliva to athletic level expressed by the metabolic profile. Phase 1 thus provided a proof-of-concept that salivary infrared spectroscopy could be used as a non-invasive, rapid screening tool for athlete profiling, setting the stage for more nuanced investigations.

Phase 2 expanded on the proof-of-concept established in Phase 1 by investigating the acute biochemical response to graded physical activity. Using a controlled treadmill protocol (0, 5, 10, 15 km/h), saliva samples were collected at each stage and analyzed using second-derivative ATR-FTIR spectroscopy. This Phase introduced, for the first time, advanced chemometric tools—specifically PCA-LDA and PLS-DA—to achieve accurate classification of samples based on exercise intensity.

Thiocyanate (SCN^-) emerged as a particularly responsive spectral feature, alongside other well-characterized exercise-related metabolites such as lactate. This was the first time that thiocyanate had been proposed as a candidate salivary biomarker for physical exertion, based on its distinct spectral behavior. The classification models demonstrated high predictive accuracy, reinforcing the potential of second-derivative ATR-FTIR spectroscopy as a rapid method for monitoring real-time physiological stress. Phase 2 thus represented a methodological and analytical advancement, bridging the gap between non-invasive sampling and real-time metabolic tracking.

Phase 3 addressed a key limitation of the previous phases—namely, the lack of absolute concentration data—by developing and validating a photometric method for the quantitative determination of salivary thiocyanate. A large-scale study involving 161 athletes was conducted to evaluate the consistency, reproducibility, and physiological relevance of thiocyanate concentration in response to exercise. The method was based

on the formation of a thiocyanato-iron complex and allowed for high-throughput, low-cost analysis under field-relevant conditions.

The results confirmed that thiocyanate levels consistently decreased with increasing exercise intensity, providing quantitative support to the spectral observations made in Phase 2. Furthermore, the study revealed statistically significant effects of gender and smoking status on thiocyanate concentrations, underscoring the importance of interindividual variability in biomarker expression. This was the first comprehensive demonstration of thiocyanate as a practical, exercise-responsive salivary biomarker validated in a real-world athletic setting. The large cohort size and detailed statistical treatment of the data enhanced the robustness and generalizability of the findings, making a strong case for the biomarker's inclusion in broader physiological monitoring frameworks.

Phase 4 broadened the scope of the thesis by moving from localized (salivary) analysis to systemic metabolic profiling using DBS samples analyzed with ^{1}H NMR spectroscopy. This approach addressed several limitations of traditional venipuncture, including invasiveness and logistical constraints, by leveraging minimally invasive blood microsampling techniques. Both commercial DBS cards (i.e., Whatman 903) and, for the first time, custom sorptive substrates developed in-house were used for sample collection. This methodological innovation allowed for the evaluation of different matrix materials in terms of compatibility with NMR analysis.

In addition to using commercial DBS cards (i.e., Whatman 903), this Phase also explored the application of FPSE-inspired adsorptive materials as alternative media for dried blood microsampling. Although these experimental substrates showed limited success in terms of analyte recovery and compatibility with NMR analysis, their evaluation provided valuable insights into the material-specific constraints of microsampling workflows. Future work may focus on optimizing surface chemistries, polymer coatings, and sample elution strategies to enhance their performance. With further refinement,

such materials could offer low-cost, customizable alternatives for decentralized biomonitoring in sports and clinical settings.

Despite the inherent challenges of low sample volume and matrix effects, NMR spectroscopy successfully differentiated metabolic profiles associated with different exercise intensities. Key metabolites such as lactate, alanine, and valine were identified as responsive to physical exertion, validating earlier findings from salivary analyses and extending them into the domain of systemic metabolism. Phase 4 thus marked the first time that DBS-NMR was applied in a sports science context to explore exercise-related metabolic changes, highlighting the technique's utility for minimally invasive, high-resolution biomonitoring.

Together, the four phases of this thesis form a coherent and innovative body of work that advances the scientific understanding of exercise-induced biochemical changes. By employing both untargeted (ATR-FTIR and NMR spectroscopy) and targeted (Vis-photometry) analytical techniques on non-invasively/minimal-invasively collected samples, this research bridges a crucial gap between laboratory capability and field applicability.

From a methodological perspective, the thesis showcases the power of chemometric modeling in extracting meaningful physiological information from complex spectral data. The integration of second-derivative spectral processing, predictive multivariate models, and rigorous statistical validation contributes to the growing field of applied chemometrics in biomedical sciences. The exploration of novel sampling materials and the validation of DBS protocols for NMR further extend the analytical toolkit available to exercise scientists and physiologists.

Thus, the work introduces thiocyanate as a novel, exercise-responsive biomarker and validates it through both qualitative and quantitative means. Moreover, it demonstrates the feasibility of using both saliva and DBS as reliable matrices for metabolic monitoring.

The practical implications are equally significant. The non-invasive, rapid, and scalable nature of the proposed protocols makes them ideal for real-time monitoring in athletic settings, training personalization, and potentially even clinical applications in stress and recovery evaluation. By combining scientific rigor with translational relevance, this work lays the groundwork for future applications in digital health, wearable diagnostics, and personalized performance management.

In conclusion, this thesis delivers a comprehensive, analytically validated, and field-applicable framework for monitoring physiological responses to exercise using saliva and dried blood spots. It contributes novel biomarkers, methods, and materials to scientific literature and opens new avenues for minimally invasive biomonitoring in both sports and health sciences.

Bibliography

1. Biróné NE. *Sportpedagógia*. (Biróné NE, ed.). Pécs: Dialóg Campus Publishing House – Nordex Ltd.; 2011.

2. Ács P. *Research Methodology in Sport Sciences*. (Ács P, ed.). Pécs: University of Pécs, Faculty of Health Sciences; 2015.

3. Heggie V. Sport (and exercise) medicine in Britain: healthy citizens and abnormal athletes. *Canadian bulletin of medical history*. 2011;28(2):249-269. doi:10.3138/cbmh.28.2.249

4. Bernardes-Ribeiro M, Patrone LGA, Cristina-Silva C, Bícego KC, Gargaglioni LH. Exercise derived myokine irisin as mediator of cardiorespiratory, metabolic and thermal adjustments during central and peripheral chemoreflex activation. *Scientific Reports*. 2024;14(1):1-17. doi:10.1038/s41598-024-62650-7

5. Huschtscha Z, Silver J, Gerhardy M, et al. The Effect of Palmitoylethanolamide (PEA) on Skeletal Muscle Hypertrophy, Strength, and Power in Response to Resistance Training in Healthy Active Adults: A Double-Blind Randomized Control Trial. *Sports Medicine - Open*. 2024;10(1):1-16. doi:10.1186/S40798-024-00732-6

6. Baxter BA, Baross AW, Ryan DJ, Kay AD. Effects of detraining on neuromuscular function and structural adaptations following once- or twice-weekly eccentric resistance training in older adults. *Aging Clinical and Experimental Research*. 2024;36(1):1-11. doi:10.1007/S40520-024-02828-1

7. Powers SK, Radak Z, Ji LL, Jackson M. Reactive oxygen species promote endurance exercise-induced adaptations in skeletal muscles. *Journal of sport and health science*. 2024;13(6). doi:10.1016/J.JSHS.2024.05.001

8. Ammar A, Boujelbane MA, Simak ML, et al. Unveiling the acute neurophysiological responses to strength training: An exploratory study on novices performing weightlifting bouts with different motor learning models. *Biology of sport*. 2024;41(2):249-274. doi:10.5114/BIOLSPORT.2024.133481

9. Morgan JA, Corrigan F, Baune BT. Effects of physical exercise on central nervous system functions: a review of brain region specific adaptations. *Journal of Molecular Psychiatry*. 2015;3(1):1-13. doi:10.1186/S40303-015-0010-8

10. Ament W, Verkerke G. Exercise and fatigue. *Sports medicine*. 2009;39(5):389-422. doi:10.2165/00007256-200939050-00005

11. Schiphof-Godart L, Roelands B, Hettinga FJ. Drive in sports: How mental fatigue affects endurance performance. *Frontiers in Psychology*. 2018;9(AUG):360941. doi:10.3389/FPSYG.2018.01383

12. Wender CLA, Manninen M, O'Connor PJ. The Effect of Chronic Exercise on Energy and Fatigue States: A Systematic Review and Meta-Analysis of Randomized Trials. *Frontiers in Psychology*. 2022;13:907637. doi:10.3389/FPSYG.2022.907637

13. Puetz TW. Physical activity and feelings of energy and fatigue: epidemiological evidence. *Sports medicine*. 2006;36(9):767-780. doi:10.2165/00007256-200636090-00004

14. Yánez CA, Cohen D, Serrato M, Ospina AL. Acute fatigue in endurance athletes: The association between countermovement jump variables and creatine kinase response. *European Journal of Human Movement*. 2023;51:115-125. doi:10.21134/eurjhm.2023.51.10

15. Stephenson SD, Kocan JW, Vinod A V., Kluczynski MA, Bisson LJ. A Comprehensive Summary of Systematic Reviews on Sports Injury Prevention Strategies. *Orthopaedic Journal of Sports Medicine*. 2021;9(10). doi:10.1177/23259671211035776

16. Vergeer I, Johansson M, Cagas JY. Holistic movement practices – An emerging category of physical activity for exercise psychology. *Psychology of Sport and Exercise*. 2021;53:101870. doi:10.1016/J.PSYCHSPORT.2020.101870

17. Biddle SJH. Physical activity research in Australia: A view from exercise psychology and behavioural medicine. *Asian Journal of Sport and Exercise Psychology*. 2021;1(1):12-20. doi:10.1016/J.AJSEP.2021.03.006

18. Close GL, Hamilton DL, Philp A, Burke LM, Morton JP. New strategies in sport nutrition to increase exercise performance. *Free radical biology & medicine*. 2016;98:144-158. doi:10.1016/J.FREERADBIOMED.2016.01.016

19. Larrosa M, Gil-Izquierdo A, González-Rodríguez LG, et al. Nutritional Strategies for Optimizing Health, Sports Performance, and Recovery for Female Athletes and Other Physically Active Women: A Systematic Review. *Nutrition Reviews*. July 2024. doi:10.1093/NUTRIT/NUAE082

20. Maughan RJ, Shirreffs SM. Nutrition for sports performance: issues and opportunities. *Proceedings of the Nutrition Society*. 2012;71(1):112-119. doi:10.1017/S0029665111003211

21. Seçkin AÇ, Ateş B, Seçkin M. Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities. *Applied Sciences*. 2023;13(18):10399. doi:10.3390/APP131810399

22. Li RT, Kling SR, Salata MJ, Cupp SA, Sheehan J, Voos JE. Wearable Performance Devices in Sports Medicine. *Sports Health: A Multidisciplinary Approach*. 2015;8(1):74-78. doi:10.1177/1941738115616917

23. Rana M, Mittal V. Wearable Sensors for Real-Time Kinematics Analysis in Sports: A Review. *IEEE Sensors Journal*. 2021;21(2):1187-1207. doi:10.1109/JSEN.2020.3019016

24. Physical K, Sanchez-Oliver J, García-Fernández J, de Dieu Habyarimana J,

Tugirumukiza E, Zhou K. Physical Education and Sports: A Backbone of the Entire Community in the Twenty-First Century. *International Journal of Environmental Research and Public Health*. 2022;19(12):7296. doi:10.3390/IJERPH19127296

25. Kjønniksen L, Fjørtoft I, Wold B. Attitude to physical education and participation in organized youth sports during adolescence related to physical activity in young adulthood: A 10-year longitudinal study. *European Physical Education Review*. 2010;15(2):139-154. doi:10.1177/1356336X09345231

26. Ahmetov II, Fedotovskaya ON. Current Progress in Sports Genomics. *Advances in Clinical Chemistry*. 2015;70:247-314. doi:10.1016/BS.ACC.2015.03.003

27. Ahmetov II, Fedotovskaya ON. Sports genomics: Current state of knowledge and future directions. *Cellular and Molecular Exercise Physiology*. 2012;1(1). doi:10.7457/CMEP.V1I1.E1

28. Kanope T, Santos CGM dos, Veneroso CE, et al. Polygenic Profile and Magnitude Impact of Inflammatory Response in Soccer Athletes. *Open Journal of Genetics*. 2024;14(3):47-76. doi:10.4236/OJGEN.2024.143005

29. Chávez-Guevara IA, Hernández-Torres RP, González-Rodríguez E, Ramos-Jiménez A, Amaro-Gahete FJ. Biomarkers and genetic polymorphisms associated with maximal fat oxidation during physical exercise: implications for metabolic health and sports performance. *European Journal of Applied Physiology*. 2022;122(8):1773-1795. doi:10.1007/S00421-022-04936-0

30. Bouchard C, Rankinen T, Chagnon YC, et al. Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE family study. *Journal of Applied Physiology*. 2000;88(2):551-559. doi:10.1152/JAPPL.2000.88.2.551

31. Hall ECR, Semenova EA, Bondareva EA, et al. Association of muscle fiber composition with health and exercise-related traits in athletes and untrained subjects. *Biology of Sport*. 2021;38(4):659-666.

doi:10.5114/BIOLSPORT.2021.102923

32. Ahmetov II, John G, Semenova EA, Hall ECR. Genomic predictors of physical activity and athletic performance. *Advances in Genetics*. 2024;111:311-408. doi:10.1016/BS.ADGEN.2024.01.001

33. Nishikawa K, Biewener AA, Aerts P, et al. Neuromechanics: an integrative approach for understanding motor control. *Integrative and Comparative Biology*. 2007;47(1):16-54. doi:10.1093/ICB/ICM024

34. Taiar R, Bernardo-Filho M, Sañudo B, Ivanenko Y. Editorial: The Relationship Between Neural Circuitry and Biomechanical Action. *Frontiers in Human Neuroscience*. 2022;16:838028. doi:10.3389/FNHUM.2022.838028

35. Gardner A, Kay-Lambkin F, Stanwell P, et al. A systematic review of diffusion tensor imaging findings in sports-related concussion. *Journal of neurotrauma*. 2012;29(16):2521-2538. doi:10.1089/NEU.2012.2628

36. Cantu RC. Role of Diffusion Tensor Imaging MRI in Detecting Brain Injury in Asymptomatic Contact Athletes. *World Neurosurgery*. 2013;80(6):792-793. doi:10.1016/J.WNEU.2013.10.011

37. Churchill NW, Hutchison MG, Richards D, Leung G, Graham SJ, Schweizer TA. Neuroimaging of sport concussion: persistent alterations in brain structure and function at medical clearance. *Scientific Reports*. 2017;7(1):1-9. doi:10.1038/s41598-017-07742-3

38. Nadikattu RR. Implementation of New Ways of Artificial Intelligence in Sports. *Journal of Xidian University*. June 2020. doi:10.2139/SSRN.3620017

39. Talha M. Research on the use of 3D modeling and motion capture technologies for making sports training easier. *Revista de Psicologia del Deporte*. 31:1-10.

40. Venzke J, Hohmann R, Krombholz A, Platen P, Reichert M. Enhancing Learning

Experiences in Sports Science through Video and AI-generated Feedback. *Learning Analytics und Künstliche Intelligenz in Studium und Lehre*. 2024:79-95. doi:10.1007/978-3-658-42993-5_5

41. Foina AG, Badia RM, EI-Deeb A, Ramirez-Fernandez FJ. Player tracker - A tool to analyze sport players using RFID. In: *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2010*. ; 2010:772-775. doi:10.1109/PERCOMW.2010.5470538

42. Claudino JG, Capanema D de O, de Souza TV, Serrão JC, Machado Pereira AC, Nassis GP. Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. *Sports Medicine - Open*. 2019;5(1):1-12. doi:10.1186/S40798-019-0202-3

43. Tan L, Ran N. Applying Artificial Intelligence Technology to Analyze the Athletes' Training Under Sports Training Monitoring System. *International Journal of Humanoid Robotics*. 2022;20(6). doi:10.1142/S0219843622500177

44. Beck KL, Thomson JS, Swift RJ, Hurst PR von. Role of nutrition in performance enhancement and postexercise recovery. *Open Access Journal of Sports Medicine*. 2015;6:259. doi:10.2147/OAJSM.S33605

45. Dziewiecka H, Buttar HS, Kasperska A, et al. Physical activity induced alterations of gut microbiota in humans: a systematic review. *BMC Sports Science, Medicine and Rehabilitation 2022 14:1*. 2022;14(1):1-22. doi:10.1186/S13102-022-00513-2

46. Ivy JL, Ferguson-Stegall LM. Nutrient Timing: The Means to Improved Exercise Performance, Recovery, and Training Adaptation. *American Journal of Lifestyle Medicine*. 2013;8(4):246-259. doi:10.1177/1559827613502444

47. Ghazzawi HA, Hussain MA, Raziq KM, et al. Exploring the Relationship between Micronutrients and Athletic Performance: A Comprehensive Scientific Systematic Review of the Literature in Sports Medicine. *Sports*. 2023;11(6):109.

doi:10.3390/SPORTS11060109

48. Marinho AH, Gonçalves JS, Araújo PK, Lima-Silva AE, Ataide-Silva T, de Araujo GG. Effects of creatine and caffeine ingestion in combination on exercise performance: A systematic review. *Critical Reviews in Food Science and Nutrition*. 2023;63(20):4785-4798. doi:10.1080/10408398.2021.2007470

49. Spradley BD, Crowley KR, Tai CY, et al. Ingesting a pre-workout supplement containing caffeine, B-vitamins, amino acids, creatine, and beta-alanine before exercise delays fatigue while improving reaction time and muscular endurance. *Nutrition & Metabolism*. 2012;9:28. doi:10.1186/1743-7075-9-28

50. Hill J, Howatson G, van Someren K, Leeder J, Pedlar C. Compression garments and recovery from exercise-induced muscle damage: a meta-analysis. *British Journal of Sports Medicine*. 2014;48(18):1340-1346. doi:10.1136/BJSPORTS-2013-092456

51. Tejero-Fernández V, Membrilla-Mesa M, Galiano-Castillo N, Arroyo-Morales M. Immunological effects of massage after exercise: A systematic review. *Physical Therapy in Sport*. 2015;16(2):187-192. doi:10.1016/J.PTSP.2014.07.001

52. Rose C, Edwards KM, Siegler J, Graham K, Caillaud C. Whole-body Cryotherapy as a Recovery Technique after Exercise: A Review of the Literature. *International Journal of Sports Medicine*. 2017;38(14):1049-1060. doi:10.1055/S-0043-114861

53. Vitale KC, Owens R, Hopkins SR, Malhotra A. Sleep Hygiene for Optimizing Recovery in Athletes: Review and Recommendations. *International Journal of Sports Medicine*. 2019;40(8):535-543. doi:10.1055/A-0905-3103

54. Johnston LH, Carroll D. The psychological impact of injury: effects of prior sport and exercise involvement. *British Journal of Sports Medicine*. 2000;34(6):436-439. doi:10.1136/BJSM.34.6.436

55. Li C, Wang CKJ, Pyun DY, Kee YH. Burnout and its relations with basic psychological needs and motivation among athletes: A systematic review and meta-analysis.

Psychology of Sport and Exercise. 2013;14(5):692-700. doi:10.1016/J.PSYCHSPORT.2013.04.009

56. Brajendra B, Rajesh T, Bhadauriya B. Stress management technique for athletes during sports: A critical review. *Journal of Drug Delivery and Therapeutics*. 2018;8(5-s):67-72. doi:10.22270/JDDT.V8I5-S.1956

57. Birrer D, Morgan G. Psychological skills training as a way to enhance an athlete's performance in high-intensity sports. *Scandinavian Journal of Medicine & Science in Sports*. 2010;20(SUPPL. 2):78-87. doi:10.1111/J.1600-0838.2010.01188.X

58. Hargreaves M, Spriet LL. Skeletal muscle energy metabolism during exercise. *Nature Metabolism*. 2020;2(9):817-828. doi:10.1038/s42255-020-0251-4

59. Jürimäe J, Mäestu J, Jürimäe T, Mangus B, Von Duvillard SP. Peripheral signals of energy homeostasis as possible markers of training stress in athletes: a review. *Metabolism*. 2011;60(3):335-350. doi:10.1016/J.METABOL.2010.02.009

60. Chen J, Zhou R, Feng Y, Cheng L. Molecular mechanisms of exercise contributing to tissue regeneration. *Signal Transduction and Targeted Therapy 2022 7:1*. 2022;7(1):1-24. doi:10.1038/s41392-022-01233-2

61. Rusdiawan A, Sholikhah AM atus, Prihatiningsih S. The Changes in pH Levels, Blood Lactic Acid and Fatigue Index to Anaerobic Exercise on Athlete after NaHCO3 Administration. *Malaysian Journal of Medicine and Health Sciences*. 2020;16:50-56. https://scholar.unair.ac.id/en/publications/the-changes-in-ph-levels-blood-lactic-acid-and-fatigue-index-to-a. Accessed September 8, 2024.

62. Wells GD, Selvadurai H, Tein I. Bioenergetic provision of energy for muscular activity. *Paediatric Respiratory Reviews*. 2009;10(3):83-90. doi:10.1016/J.PRRV.2009.04.005

63. Mac Laren D, Morton J. *Biochemistry for Sport and Exercise Metabolism*. 2nd ed. Wiley; 2024.

64. Hultman E, Bergström J, Anderson NM. Breakdown and Resynthesis of Phosphorylcreatine and Adenosine Triphosphate in Connection with Muscular Work in Man. *Scandinavian Journal of Clinical and Laboratory Investigation*. 1967;19(1):56-66. doi:10.3109/00365516709093481

65. Nieman DC. Multiomics Approach to Precision Sports Nutrition: Limits, Challenges, and Possibilities. *Frontiers in Nutrition*. 2021;8:796360. doi:10.3389/FNUT.2021.796360

66. Dahl MA, Areta JL, Jeppesen PB, et al. Coingestion of protein and carbohydrate in the early recovery phase, compared with carbohydrate only, improves endurance performance despite similar glycogen degradation and AMPK phosphorylation. *Journal of Applied Physiology*. 2020;129(2):297-310. doi:10.1152/JAPPLPHYSIOL.00817.2019

67. Markus I, Constantini K, Hoffman JR, Bartolomei S, Gepner Y. Exercise-induced muscle damage: mechanism, assessment and nutritional factors to accelerate recovery. *European Journal of Applied Physiology*. 2021;121(4):969-992. doi:10.1007/S00421-020-04566-4

68. Millet GY, Tomazin K, Verges S, et al. Neuromuscular Consequences of an Extreme Mountain Ultra-Marathon. *PLOS ONE*. 2011;6(2):e17059. doi:10.1371/JOURNAL.PONE.0017059

69. Sudo M, Ando S, Poole DC, Kano Y. Blood flow restriction prevents muscle damage but not protein synthesis signaling following eccentric contractions. *Physiological Reports*. 2015;3(7). doi:10.14814/PHY2.12449

70. Kraemer WJ, Ratamess NA, Hymer WC, Nindl BC, Fragala MS. Growth Hormone(s), Testosterone, Insulin-Like Growth Factors, and Cortisol: Roles and Integration for Cellular Development and Growth With Exercise. *Frontiers in Endocrinology*. 2020;11:33. doi:10.3389/FENDO.2020.00033

71. Hill EE, Zack E, Battaglini C, Viru M, Viru A, Hackney AC. Exercise and circulating cortisol levels: the intensity threshold effect. *Journal of endocrinological investigation*. 2008;31(7):587-591. doi:10.1007/BF03345606

72. Review S, Torres R, Koutakis P, Forsse JS. The Effects of Different Exercise Intensities and Modalities on Cortisol Production in Healthy Individuals: A Review. *Journal of Exercise and Nutrition*. 2021;4(4):19. doi:10.53520/JEN2021.103108

73. Hackney AC, Walz EA. Hormonal adaptation and the stress of exercise training: the role of glucocorticoids. *Trends in sport sciences*. 2013;20(4):165. /pmc/articles/PMC5988244/. Accessed September 8, 2024.

74. Viru A, Viru M. Cortisol - Essential adaptation hormone in exercise. *International Journal of Sports Medicine*. 2004;25(6):461-464. doi:10.1055/S-2004-821068

75. Gharahdaghi N, Phillips BE, Szewczyk NJ, Smith K, Wilkinson DJ, Atherton PJ. Links Between Testosterone, Oestrogen, and the Growth Hormone/Insulin-Like Growth Factor Axis and Resistance Exercise Muscle Adaptations. *Frontiers in Physiology*. 2021;11:621226. doi:10.3389/FPHYS.2020.621226

76. Romagnoli C, Zonefrati R, Sharma P, Innocenti M, Cianferotti L, Brandi ML. Characterization of Skeletal Muscle Endocrine Control in an In Vitro Model of Myogenesis. *Calcified tissue international*. 2020;107(1):18-30. doi:10.1007/S00223-020-00678-3

77. Yoshida T, Delafontaine P. Mechanisms of IGF-1-Mediated Regulation of Skeletal Muscle Hypertrophy and Atrophy. *Cells*. 2020;9(9):1970. doi:10.3390/CELLS9091970

78. Kruse R, Petersson SJ, Christensen LL, et al. Effect of long-term testosterone therapy on molecular regulators of skeletal muscle mass and fibre-type distribution in aging men with subnormal testosterone. *Metabolism*. 2020;112:154347. doi:10.1016/J.METABOL.2020.154347

79. Takarada Y. Evaluation of muscle damage after a rugby match with special reference to tackle plays. *British Journal of Sports Medicine*. 2003;37(5):416. doi:10.1136/BJSM.37.5.416

80. Gleeson M. Biochemical and Immunological Markers of Over-Training. *Journal of Sports Science & Medicine*. 2002;1(2):31. /pmc/articles/PMC3963240/. Accessed September 8, 2024.

81. Nix C, Hemmati M, Cobraiville G, Servais AC, Fillet M. Blood Microsampling to Monitor Metabolic Profiles During Physical Exercise. *Frontiers in Molecular Biosciences*. 2021;8:681400. doi:10.3389/FMOLB.2021.681400

82. Soligard T, Schwellnus M, Alonso JM, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British Journal of Sports Medicine*. 2016;50(17):1030-1041. doi:10.1136/bjsports-2016-096581

83. Al-Khelaifi F, Abraham D, Diboun I, Elrayess MA. Proteomics and metabolomics research in exercise and sport. In: *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press; 2019:539-566. doi:10.1016/B978-0-12-816193-7.00023-3

84. Prado E, Souza GHMF, Pegurier M, et al. Non-targeted sportomics analyses by mass spectrometry to understand exercise-induced metabolic stress in soccer players. *International Journal of Mass Spectrometry*. 2017;418:1-5. doi:10.1016/J.IJMS.2017.02.002

85. Muniz-Santos R, Magno-França A, Jurisica I, Cameron LC. From Microcosm to Macrocosm: The -Omics, Multiomics, and Sportomics Approaches in Exercise and Sports. *OMICS A Journal of Integrative Biology*. 2023;27(11):499-518. doi:10.1089/OMI.2023.0169

86. Khoramipour K, Sandbakk Ø, Keshteli AH, Gaeini AA, Wishart DS, Chamari K.

Metabolomics in Exercise and Sports: A Systematic Review. *Sports Medicine*. 2022;52(3):547-583. doi:10.1007/S40279-021-01582-Y

87. Steckl AJ, Ray P. Stress Biomarkers in Biological Fluids and Their Point-of-Use Detection. *ACS Sensors*. 2018;3(10):2025-2044. doi:10.1021/ACSSENSORS.8B00726

88. Stojiljković S, Gavrilović L, Pejić S, et al. Effects of Endurance Training on Antioxidant and Hormonal Status in Peripheral Blood of Young Healthy Men. *Life*. 2024;14(8):921. doi:10.3390/LIFE14080921

89. Barton W, Penney NC, Cronin O, et al. The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. *Gut*. 2018;67(4):625-633. doi:10.1136/GUTJNL-2016-313627

90. Klous L, de Ruiter CJ, Scherrer S, Gerrett N, Daanen HAM. The (in)dependency of blood and sweat sodium, chloride, potassium, ammonia, lactate and glucose concentrations during submaximal exercise. *European Journal of Applied Physiology*. 2021;121(3):803-816. doi:10.1007/S00421-020-04562-8

91. Antunes BM, Rosa-Neto JC, Batatinha HAP, Franchini E, Teixeira AM, Lira FS. Physical fitness status modulates the inflammatory proteins in peripheral blood and circulating monocytes: role of PPAR-gamma. *Scientific Reports*. 2020;10(1):1-12. doi:10.1038/s41598-020-70731-6

92. Lavoie JM. Blood metabolites during prolonged exercise in swimming and leg cycling. *European Journal of Applied Physiology and Occupational Physiology*. 1982;48(1):127-133. doi:10.1007/BF00421172/METRICS

93. Abdelkrim N Ben, Castagna C, Fazaa S El, Tabka Z, Ati J El. Blood metabolites during basketball competitions. *Journal of Strength and Conditioning Research*. 2009;23(3):765-773. doi:10.1519/JSC.0B013E3181A2D8FC

94. Lewis GD, Farrell L, Wood MJ, et al. Metabolic signatures of exercise in human

plasma. *Science Translational Medicine*. 2010;2(33). doi:10.1126/SCITRANSLMED.3001006

95. Stander Z, Luies L, Mienie LJ, et al. The altered human serum metabolome induced by a marathon. *Metabolomics*. 2018;14(11):1-11. doi:10.1007/S11306-018-1447-4

96. Lei BUW, Prow TW. A review of microsampling techniques and their social impact. *Biomedical Microdevices*. 2019;21(4):1-30. doi:10.1007/S10544-019-0412-Y

97. Jesuthasan A;, Ali A;, Lee JKW;, et al. Assessment of Changes in Physiological Markers in Different Body Fluids at Rest and after Exercise. *Nutrients*. 2022;14(21):4685. doi:10.3390/NU14214685

98. Armstrong LE. Hydration Assessment Techniques. *Nutrition Reviews*. 2005;63(suppl_1):S40-S54. doi:10.1111/J.1753-4887.2005.TB00153.X

99. Chlíbková D, Rosemann T, Posch L, Matoušek R, Knechtle B. Pre- and Post-Race Hydration Status in Hyponatremic and Non-Hyponatremic Ultra-Endurance Athletes. *The Chinese journal of physiology*. 2016;59(3):173-183. doi:10.4077/CJP.2016.BAE391

100. Pechlivanis A, Kostidis S, Saraslanidis P, et al. 1H NMR study on the short-and long-term impact of two training programs of sprint running on the metabolic fingerprint of human serum. *Journal of Proteome Research*. 2013;12(1):470-480. doi:10.1021/PR300846X

101. Pechlivanis A, Kostidis S, Saraslanidis P, et al. 1H NMR-based metabonomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine. *Journal of Proteome Research*. 2010;9(12):6405-6416. doi:10.1021/PR100684T

102. Pechlivanis A, Papaioannou KG, Tsalis G, Saraslanidis P, Mougios V, Theodoridis GA. Monitoring the response of the human urinary metabolome to brief maximal exercise by a combination of RP-UPLC-MS and 1H NMR spectroscopy. *Journal of*

Proteome Research. 2015;14(11):4610-4622. doi:10.1021/ACS.JPROTEOME.5B00470

103. Arnaoutis G, Kavouras SA, Angelopoulou A, et al. Fluid Balance during Training in Elite Young Athletes of Different Sports. *Journal of Strength and Conditioning Research*. 2015;29(12):3447-3452. doi:10.1519/JSC.0000000000000400

104. Shirreffs SM, Sawka MN. Fluid and electrolyte needs for training, competition, and recovery. *Journal of sports sciences*. 2011;29 (Suppl((Suppl. 1)). doi:10.1080/02640414.2011.614269

105. Baker LB, Model JB, Barnes KA, et al. Skin-interfaced microfluidic system with personalized sweating rate and sweat chloride analytics for sports science applications. *Science Advances*. 2020;6(50):3929-3940. doi:10.1126/SCIADV.ABE3929

106. Baker LB, De Chavez PJD, Ungaro CT, et al. Exercise intensity effects on total sweat electrolyte losses and regional vs. whole-body sweat [Na+], [Cl−], and [K+]. *European Journal of Applied Physiology*. 2019;119(2):361-375. doi:10.1007/S00421-018-4048-Z

107. Vinoth R, Nakagawa T, Mathiyarasu J, Mohan AMV. Fully Printed Wearable Microfluidic Devices for High-Throughput Sweat Sampling and Multiplexed Electrochemical Analysis. *ACS Sensors*. 2021;6(3):1174-1186. doi:10.1021/ACSSENSORS.0C02446

108. Xuan X, Pérez-Ràfols C, Chen C, Cuartero M, Crespo GA. Lactate Biosensing for Reliable On-Body Sweat Analysis. *ACS Sensors*. 2021;6(7):2763-2771. doi:10.1021/ACSSENSORS.1C01009

109. Mena-Bravo A, Luque de Castro MD. Sweat: A sample with limited present applications and promising future in metabolomics. *Journal of Pharmaceutical and Biomedical Analysis*. 2014;90:139-147. doi:10.1016/J.JPBA.2013.10.048

110. Parrilla M, Cánovas R, Jeerapan I, et al. A Textile-Based Stretchable Multi-Ion Potentiometric Sensor. *Advanced Healthcare Materials*. 2016;5(9):996-1001. doi:10.1002/ADHM.201600092

111. McCaul M, Porter A, Barrett R, et al. Wearable Platform for Real-time Monitoring of Sodium in Sweat. *ChemPhysChem*. 2018;19(12):1531-1536. doi:10.1002/CPHC.201701312

112. Gao W, Emaminejad S, Nyein HYY, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature*. 2016;529(7587):509-514. doi:10.1038/nature16521

113. Patel V, Mardolkar A, Shelar A, Tiwari R, Srivastava R. Wearable sweat chloride sensors: materials, fabrication and their applications. *Analytical Methods*. 2024;16(10):1439-1453. doi:10.1039/D3AY01979A

114. Pham S, Yeap D, Escalera G, et al. Wearable Sensor System to Monitor Physical Activity and the Physiological Effects of Heat Exposure. *Sensors*. 2020;20(3):855. doi:10.3390/S20030855

115. Hussain JN, Mantri N, Cohen MM. Working Up a Good Sweat – The Challenges of Standardising Sweat Collection for Metabolomics Analysis. *The Clinical Biochemist Reviews*. 2017;38(1):13. /pmc/articles/PMC5548369/. Accessed September 10, 2024.

116. DeCaro JA. Methodological considerations in the use of salivary α-amylase as a stress marker in field research. *American Journal of Human Biology*. 2008;20(5):617-619. doi:10.1002/AJHB.20795

117. Pathiyil V, Udayasankar R, Pathiyil V, Udayasankar R. Salivary Diagnostics. *Saliva and Salivary Diagnostics*. March 2019. doi:10.5772/INTECHOPEN.84722

118. Melguizo-Rodríguez L, Costela-Ruiz VJ, Manzano-Moreno FJ, Ruiz C, Illescas-Montes R. Salivary Biomarkers and Their Application in the Diagnosis and Monitoring of the

Most Common Oral Pathologies. *International Journal of Molecular Sciences 2020, Vol 21, Page 5173*. 2020;21(14):5173. doi:10.3390/IJMS21145173

119. Hough J, Leal D, Scott G, Taylor L, Townsend D, Gleeson M. Reliability of salivary cortisol and testosterone to a high-intensity cycling protocol to highlight overtraining. *Journal of Sports Sciences*. 2021;39(18):2080-2086. doi:10.1080/02640414.2021.1918362

120. Vaamonde D, García-Manso JM, Algar-Santacruz C, Abbasi A, Sarmiento S, Valverde-Esteve T. Behaviour of salivary testosterone and cortisol in men during an Ironman Triathlon. *European Journal of Sport Science*. 2022;22(9):1335-1342. doi:10.1080/17461391.2021.1955011

121. Sellami M, Bragazzi NL, Aboghaba B, Elrayess MA. The Impact of Acute and Chronic Exercise on Immunoglobulins and Cytokines in Elderly: Insights From a Critical Review of the Literature. *Frontiers in Immunology*. 2021;12:631873. doi:10.3389/FIMMU.2021.631873

122. Brandtzaeg P. Salivary Immunoglobulins. In: *Human Saliva*. 1st ed. Boca Raton: CRC Press; 2021:1-54. doi:10.1201/9781003210405-1

123. Bel'skaya L V., Sarf EA, Kosenok VK, Gundyrev IA. Biochemical Markers of Saliva in Lung Cancer: Diagnostic and Prognostic Perspectives. *Diagnostics*. 2020;10(4):186. doi:10.3390/DIAGNOSTICS10040186

124. Bazzano M, Laghi L, Zhu C, et al. Exercise Induced Changes in Salivary and Serum Metabolome in Trained Standardbred, Assessed by 1H-NMR. *Metabolites*. 2020;10(7):298. doi:10.3390/METABO10070298

125. Ntovas P, Loumprinis N, Maniatakos P, Margaritidi L, Rahiotis C. The Effects of Physical Exercise on Saliva Composition: A Comprehensive Review. *Dentistry Journal 2022, Vol 10, Page 7*. 2022;10(1):7. doi:10.3390/DJ10010007

126. McBride EM, Lawrence RJ, McGee K, et al. Rapid liquid chromatography tandem

mass spectrometry method for targeted quantitation of human performance metabolites in saliva. *Journal of Chromatography A*. 2019;1601:205-213. doi:10.1016/J.CHROMA.2019.04.071

127. Serra R, Ielapi N, Barbetta A, et al. Adverse complications of venipuncture: A systematic review. *Acta Phlebologica*. 2018;19(1):11-15. doi:10.23736/S1593-232X.18.00408-3

128. World Health Organization. *WHO Guidelines on Drawing Blood*. World Health Organization; 2010. https://www.ncbi.nlm.nih.gov/books/NBK138650/. Accessed September 13, 2024.

129. Thangavelu MU, Wouters B, Kindt A, Reiss IKM, Hankemeier T. Blood microsampling technologies: Innovations and applications in 2022. *Analytical Science Advances*. 2023;4(5-6):154-180. doi:10.1002/ANSA.202300011

130. Bodley T, Chan M, Levi O, et al. Patient harm associated with serial phlebotomy and blood waste in the intensive care unit: A retrospective cohort study. *PLOS ONE*. 2021;16(1):e0243782. doi:10.1371/JOURNAL.PONE.0243782

131. Sikorova L, Hrazdilova P. The effect of psychological intervention on perceived pain in children undergoing venipuncture. *Biomedical Papers*. 2011;155(2):149-154. doi:10.5507/BP.2011.024

132. Buowari OY, Buowari OY. Complications of venepuncture. *Advances in Bioscience and Biotechnology*. 2013;4(1):126-128. doi:10.4236/ABB.2013.41A018

133. Ialongo C, Bernardini S. Phlebotomy, a bridge between laboratory and patient. *Biochemia Medica*. 2016;26(1):17-33. doi:10.11613/BM.2016.002

134. Bevan E, Chee L, McGhee S, McInnes G. Patients' attitudes to participation in clinical trials. *British Journal of Clinical Pharmacology*. 1993;35(2):204. doi:10.1111/j.1365-2125.1993.tb05687.x

135. Trolio WM, Coville WE. Pre-Analytical Specimen Preparation. *Journal of the Association for Laboratory Automation*. 2000;5(1):72-78. doi:10.1016/S1535-5535-04-00055-3

136. Roberts JL, Whiley L, Gray N, Gay M, Lawler NG. Advanced Microsamples: Current Applications and Considerations for Mass Spectrometry-Based Metabolic Phenotyping Pipelines. *Separations*. 2022;9(7):175. doi:10.3390/SEPARATIONS9070175

137. Bossi E, Limo E, Pagani L, et al. Revolutionizing Blood Collection: Innovations, Applications, and the Potential of Microsampling Technologies for Monitoring Metabolites and Lipids. *Metabolites*. 2024;14(1):46. doi:10.3390/METABO14010046

138. Guthrie R. Blood Screening for Phenylketonuria. *JAMA*. 1961;178(8):863-863. doi:10.1001/JAMA.1961.03040470079019

139. Aktuğlu Zeybek AÇ. Newborn Screening: From the Past to the Future. *Turkish Archives of Pediatrics*. 2022;57(5):473-475. doi:10.5152/TURKARCHPEDIATR.2022.16082022

140. Couacault P, Avella D, Londoño-Osorio S, et al. Targeted and untargeted metabolomics and lipidomics in dried blood microsampling: Recent applications and perspectives. *Analytical Science Advances*. 2024;5(5-6):e2400002. doi:10.1002/ANSA.202400002

141. Ottosson F, Russo F, Abrahamsson A, et al. Effects of Long-Term Storage on the Biobanked Neonatal Dried Blood Spot Metabolome. *Journal of the American Society for Mass Spectrometry*. 2023;34(4):685-694. doi:10.1021/JASMS.2C00358

142. Baillargeon KR, Brooks JC, Miljanic PR, Mace CR. Patterned Dried Blood Spot Cards for the Improved Sampling of Whole Blood. *ACS measurement science au*. 2022;2(1):31-38. doi:10.1021/ACSMEASURESCIAU.1C00031

143. De Kesel PMM, Sadones N, Capiau S, Lambert WE, Stove CP. Hemato-critical issues in quantitative analysis of dried blood spots: challenges and solutions. *Bioanalysis*. 2013;5(16):2023-2041. doi:10.4155/BIO.13.156

144. Hall EM, Flores SR, De Jesús VR. Influence of Hematocrit and Total-Spot Volume on Performance Characteristics of Dried Blood Spots for Newborn Screening. *International Journal of Neonatal Screening*. 2015;1(2):69-78. doi:10.3390/IJNS1020069

145. O'Mara M, Hudson-Curtis B, Olson K, Yueh Y, Dunn J, Spooner N. The Effect of Hematocrit and Punch Location on Assay Bias During Quantitative Bioanalysis of Dried Blood Spot Samples. *Bioanalysis*. 2011;3(20):2335-2347. doi:10.4155/BIO.11.220

146. Sinha S, Basu A, Shukla J, Dasgupta S, Dutta G, Das S. A smartphone-integrated low-cost, reagent-free, non-destructive dried blood spot-based paper sensor for hematocrit measurement. *Analytical Methods*. 2023;15(29):3532-3542. doi:10.1039/D3AY00688C

147. Denniff P, Spooner N. Volumetric absorptive microsampling: a dried sample collection technique for quantitative bioanalysis. *Analytical chemistry*. 2014;86(16):8489-8495. doi:10.1021/AC5022562

148. Lenk G, Hansson J, Beck O, Roxhed N. The Effect of Drying on The Homogeneity of DBS. *Bioanalysis*. 2015;7(16):1977-1985. doi:10.4155/BIO.15.135

149. Kok MGM, Nix C, Nys G, Fillet M. Targeted metabolomics of whole blood using volumetric absorptive microsampling. *Talanta*. 2019;197:49-58. doi:10.1016/J.TALANTA.2019.01.014

150. Carling RS, Emmett EC, Moat SJ. Evaluation of volumetric blood collection devices for the measurement of phenylalanine and tyrosine to monitor patients with phenylketonuria. *Clinica Chimica Acta*. 2022;535:157-166.

doi:10.1016/J.CCA.2022.08.005

151. Londhe V, Rajadhyaksha M. Opportunities and obstacles for microsampling techniques in bioanalysis: Special focus on DBS and VAMS. *Journal of Pharmaceutical and Biomedical Analysis*. 2020;182:113102. doi:10.1016/J.JPBA.2020.113102

152. Baillargeon KR, Mace CR. Microsampling tools for collecting, processing, and storing blood at the point-of-care. *Bioengineering & Translational Medicine*. 2023;8(2):e10476. doi:10.1002/BTM2.10476

153. Coleman D, Smith G, Lawrence R, McManus D, Diaram S, Edwards J. Capillary microsampling in nonclinical safety assessment: practical sampling and bioanalysis from a CRO perspective. *Bioanalysis*. 2017;9(10):787-798. doi:10.4155/BIO-2017-0032

154. Luo Y, Korfmacher W, Ho S, et al. Evaluation of two blood microsampling approaches for drug discovery PK studies in rats. *Bioanalysis*. 2015;7(18):2345-2359. doi:10.4155/BIO.15.126

155. Liu Y, Yu Q, Luo X, Yang L, Cui Y. Continuous monitoring of diabetes with an integrated microneedle biosensing device through 3D printing. *Microsystems & Nanoengineering*. 2021;7(1):1-12. doi:10.1038/s41378-021-00302-w

156. Kim YC, Park JH, Prausnitz MR. Microneedles for drug and vaccine delivery. *Advanced Drug Delivery Reviews*. 2012;64(14):1547-1568. doi:10.1016/J.ADDR.2012.04.005

157. Prow TW, Lin LL, Raphael AP, et al. Microbiopsy engineered for minimally invasive and suture-free sub-millimetre skin sampling. *F1000Research*. 2013;2. doi:10.12688/F1000RESEARCH.2-120.V2

158. Samant PP, Prausnitz MR. Mechanisms of sampling interstitial fluid from skin using a microneedle patch. *Proceedings of the National Academy of Sciences of the*

United States of America. 2018;115(18):4583-4588. doi:10.1073/PNAS.1716772115

159. Mukerjee E V., Collins SD, Isseroff RR, Smith RL. Microneedle array for transdermal biological fluid extraction and in situ analysis. *Sensors and Actuators A: Physical*. 2004;114(2-3):267-275. doi:10.1016/J.SNA.2003.11.008

160. Gardeniers HJGE, Luttge R, Berenschot EJW, et al. Silicon micromachined hollow microneedles for transdermal liquid transport. *Journal of Microelectromechanical Systems*. 2003;12(6):855-862. doi:10.1109/JMEMS.2003.820293

161. Li T, Barnett A, Rogers KL, Gianchandani YB. A blood sampling microsystem for pharmacokinetic applications: Design, fabrication, and initial results. *Lab on a Chip*. 2009;9(24):3495-3503. doi:10.1039/B910508E

162. Khumpuang S, Horade M, Fujioka K, Sugiyama S. Geometrical strengthening and tip-sharpening of a microneedle array fabricated by X-ray lithography. *Microsystem Technologies*. 2007;13(3-4):209-214. doi:10.1007/S00542-006-0173-4

163. Li CG, Joung HA, Noh H, Song MB, Kim MG, Jung H. One-touch-activated blood multidiagnostic system using a minimally invasive hollow microneedle integrated with a paper-based sensor. *Lab on a Chip*. 2015;15(16):3286-3292. doi:10.1039/C5LC00669D

164. Mohan AMV, Windmiller JR, Mishra RK, Wang J. Continuous minimally-invasive alcohol monitoring using microneedle sensor arrays. *Biosensors and Bioelectronics*. 2017;91:574-579. doi:10.1016/J.BIOS.2017.01.016

165. Miller CS, Foley JD, Bailey AL, et al. Current developments in salivary diagnostics. *Biomarkers in medicine*. 2010;4(1):171. doi:10.2217/BMM.09.68

166. Verkman AS, Yang B, Song Y, Manley GT, Ma T. Role of water channels in fluid transport studied by phenotype analysis of aquaporin knockout mice. *Experimental physiology*. 2000;85 Spec No(SPEC. ISS.). doi:10.1111/J.1469-445X.2000.TB00028.X

167. Haeckel R, Hänecke P. Application of saliva for drug monitoring. An in vivo model for transmembrane transport. *European journal of clinical chemistry and clinical biochemistry : journal of the Forum of European Clinical Chemistry Societies*. 1996;34(3):171-192. doi:10.1515/cclm.1996.34.3.171

168. Malamud D. Saliva as a diagnostic fluid. *BMJ (Clinical research ed)*. 1992;305(6847):207-208. doi:10.1136/BMJ.305.6847.207

169. Bellagambi FG, Lomonaco T, Salvo P, et al. Saliva sampling: Methods and devices. An overview. *TrAC Trends in Analytical Chemistry*. 2020;124:115781. doi:10.1016/J.TRAC.2019.115781

170. Priya Y, Prathibha K M. Methods of collection of saliva-A Review. *International Journal of Oral Health Dentistry*. 2007;3(3):149-153. doi:10.18231/2395-499X.2017.0032

171. Granger DA, Kivlighan KT, Fortunato C, et al. Integration of salivary biomarkers into developmental and behaviorally-oriented research: Problems and solutions for collecting specimens. *Physiology & Behavior*. 2007;92(4):583-590. doi:10.1016/J.PHYSBEH.2007.05.004

172. Gug IT, Tertis M, Hosu O, Cristea C. Salivary biomarkers detection: Analytical and immunological methods overview. *TrAC Trends in Analytical Chemistry*. 2019;113:301-316. doi:10.1016/J.TRAC.2019.02.020

173. Al Habobe H, Haverkort EB, Nazmi K, Van Splunter AP, Pieters RHH, Bikker FJ. The impact of saliva collection methods on measured salivary biomarker levels. *Clinica Chimica Acta*. 2024;552:117628. doi:10.1016/J.CCA.2023.117628

174. Vasquez-Salgado Y, Thwaits A, Serrano JP, Guan S-SA. Protocol for Safely Collecting Saliva as a Biospecimen During and Post the COVID-19 Pandemic. August 2021. doi:10.21203/RS.3.PEX-1597/V1

175. Galar A, Catalán P, Vesperinas L, et al. Use of Saliva Swab for Detection of Influenza

Virus in Patients Admitted to an Emergency Department. *Microbiology Spectrum*. 2021;9(1). doi:10.1128/SPECTRUM.00336-21

176. Dawes C, Kubieniec K. The effects of prolonged gum chewing on salivary flow rate and composition. *Archives of Oral Biology*. 2004;49(8):665-669. doi:10.1016/J.ARCHORALBIO.2004.02.007

177. Polland KE, Higgins F, Orchardson R. Salivary flow rate and pH during prolonged gum chewing in humans. *Journal of Oral Rehabilitation*. 2003;30(9):861-865. doi:10.1046/J.1365-2842.2003.01177.X

178. Iorgulescu G. Saliva between normal and pathological. Important factors in determiningsystemic and oral health. *Journal of Medicine and Life*. 2009;2(3):303. /pmc/articles/PMC5052503/. Accessed September 18, 2024.

179. Almståhl A, Wikström M. Electrolytes in stimulated whole saliva in individuals with hyposalivation of different origins. *Archives of Oral Biology*. 2003;48(5):337-344. doi:10.1016/S0003-9969(02)00200-5

180. Dawes C. The effects of flow rate and duration of stimulation on the concentrations of protein and the main electrolytes in human submandibular saliva. *Archives of Oral Biology*. 1974;19(10):887-895. doi:10.1016/0003-9969(74)90051-X

181. Justino AB, Teixeira RR, Peixoto LG, Jaramillo OLB, Espindola FS. Effect of saliva collection methods and oral hygiene on salivary biomarkers. *Scandinavian Journal of Clinical and Laboratory Investigation*. 2017;77(6):415-422. doi:10.1080/00365513.2017.1334261

182. Dong C, Puckett AD, Dawes C. The effects of chewing frequency and duration of gum chewing on salivary flow rate and sucrose concentration. *Archives of Oral Biology*. 1995;40(7):585-588. doi:10.1016/0003-9969(95)00012-E

183. Higashi T, Hijikuro M, Yamagata K, Ogawa S. Influence of saliva flow rate stimulated by gum-chewing on salivary concentrations of catecholamine metabolites. *Clinica*

Chimica Acta. 2012;414:248-252. doi:10.1016/J.CCA.2012.09.031

184. Lomonaco T, Ghimenti S, Biagini D, et al. The effect of sampling procedures on the urate and lactate concentration in oral fluid. *Microchemical Journal*. 2018;136:255-262. doi:10.1016/J.MICROC.2017.02.032

185. Golatowski C, Gesell Salazar M, Dhople VM, et al. Comparative evaluation of saliva collection methods for proteome analysis. *Clinica Chimica Acta*. 2013;419:42-46. doi:10.1016/J.CCA.2013.01.013

186. Matos-Gomes N, Katsurayama M, Makimoto FH, et al. Psychological stress and its influence on salivary flow rate, total protein concentration and IgA, IgG and IgM titers. *Neuroimmunomodulation*. 2010;17(6):396-404. doi:10.1159/000292064

187. Cohen M, Khalaila R. Saliva pH as a biomarker of exam stress and a predictor of exam performance. *Journal of Psychosomatic Research*. 2014;77(5):420-425. doi:10.1016/J.JPSYCHORES.2014.07.003

188. Humphrey SP, Williamson RT. A review of saliva: normal composition, flow, and function. *The Journal of prosthetic dentistry*. 2001;85(2):162-169. doi:10.1067/MPR.2001.113778

189. Golden SH, Wand GS, Malhotra S, Kamel I, Horton K. Reliability of hypothalamic-pituitary-adrenal axis assessment methods for use in population-based studies. *European journal of epidemiology*. 2011;26(7):511-525. doi:10.1007/S10654-011-9585-2

190. Papacosta E, Nassis GP. Saliva as a tool for monitoring steroid, peptide and immune markers in sport and exercise science. *Journal of Science and Medicine in Sport*. 2011;14(5):424-434. doi:10.1016/J.JSAMS.2011.03.004

191. Bruijns B. What Are the Limitations and Challenges of Swab-Based DNA Sampling? *Forensic Sciences*. 2024;4(1):76-95. doi:10.3390/FORENSICSCI4010006

192. Fortes MB, Diment BC, Di Felice U, Walsh NP. Dehydration decreases saliva antimicrobial proteins important for mucosal immunity. *Applied physiology, nutrition, and metabolism*. 2012;37(5):850-859. doi:10.1139/H2012-054

193. Brajović G, Bogdanović S, Anđelski-Radičević B, Popović B, Mandinić Z, Juloski J. The effect of exercise during sport training on levels of salivary diagnostic markers. *Srpski arhiv za celokupno lekarstvo*. 2021;149(11-12):660-665. doi:10.2298/SARH210913095B

194. Navazesh M, Kumar SKS. Measuring salivary flow: Challenges and opportunities. *The Journal of the American Dental Association*. 2008;139:35S-40S. doi:10.14219/JADA.ARCHIVE.2008.0353

195. Crouch DJ. Oral fluid collection: The neglected variable in oral fluid testing. *Forensic Science International*. 2005;150(2-3):165-173. doi:10.1016/J.FORSCIINT.2005.02.028

196. Gröschl M, Rauh M. Influence of commercial collection devices for saliva on the reliability of salivary steroids analysis. *Steroids*. 2006;71(13-14):1097-1100. doi:10.1016/J.STEROIDS.2006.09.007

197. Wood I, Park S, Tooke J, Smith O, Morgan RM, Meakin GE. Efficiencies of recovery and extraction of trace DNA from non-porous surfaces. *Forensic Science International: Genetics Supplement Series*. 2017;6:e153-e155. doi:10.1016/j.fsigss.2017.09.022

198. Hedman J, Akel Y, Jansson L, et al. Enhanced forensic DNA recovery with appropriate swabs and optimized swabbing technique. *Forensic science international Genetics*. 2021;53. doi:10.1016/J.FSIGEN.2021.102491

199. Seiberle I, Währer J, Kron S, et al. Collaborative swab performance comparison and the impact of sampling solution volumes on DNA recovery. *Forensic science international Genetics*. 2022;59. doi:10.1016/J.FSIGEN.2022.102716

200. *Collection Methods: SalivaBio Oral Swab (SOS)*. Salimetrics LLC; 2024. https://salimetrics.com/wp-content/uploads/2019/03/oral-swab-saliva-collection-device-instructions.pdf. Accessed September 18, 2024.

201. *Saliva Collection Handbook*. Salimetrics LLC; 2024. https://salimetrics.com/saliva-collection-handbook/. Accessed September 18, 2024.

202. Han Y, Li XL, Zhang M, Wang J, Zeng S, Min JZ. Potential use of a dried saliva spot (DSS) in therapeutic drug monitoring and disease diagnosis. *Journal of Pharmaceutical Analysis*. 2022;12(6):815-823. doi:10.1016/J.JPHA.2021.11.001

203. Numako M, Takayama T, Noge I, et al. Dried Saliva Spot (DSS) as a Convenient and Reliable Sampling for Bioanalysis: An Application for the Diagnosis of Diabetes Mellitus. *Analytical Chemistry*. 2016;88(1):635-639. doi:10.1021/ACS.ANALCHEM.5B04059

204. Chaurasia N, Jain A, Dubey N. Dried saliva spot as an effective tool in the diagnosis and management of chronic disease. *The Thai Journal of Pharmaceutical Sciences*. 2024;48(1):6. doi:10.56808/3027-7922.2862

205. Vieira CAGA, Pupin B, Bhattacharjee TT, Sakane KK. Infrared Spectroscopy Based Study of Biochemical Changes in Saliva during Maximal Progressive Test in Athletes. *Analytical sciences*. 2021;37(8):1157-1163. doi:10.2116/ANALSCI.20P395

206. Bruni S, Guglielmi V, Della Foglia E, Castoldi M, Bagnasco Gianni G. A non-destructive spectroscopic study of the decoration of archaeological pottery: from matt-painted bichrome ceramic sherds (southern Italy, VIII-VII B.C.) to an intact Etruscan cinerary urn. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2018;191:88-97. doi:10.1016/J.SAA.2017.10.010

207. Aykas DP, Borba KR, Rodriguez-Saona LE. Non-Destructive Quality Assessment of Tomato Paste by Using Portable Mid-Infrared Spectroscopy and Multivariate Analysis. *Foods*. 2020;9(9):1300. doi:10.3390/FOODS9091300

208. Wu D, He Y, Feng S. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. *Analytica Chimica Acta*. 2008;610(2):232-242. doi:10.1016/J.ACA.2008.01.056

209. Casci Ceccacci A, Cagliani A, Marizza P, Schmid S, Boisen A. Thin Film Analysis by Nanomechanical Infrared Spectroscopy. *ACS Omega*. 2019;4(4):7628-7635. doi:10.1021/ACSOMEGA.9B00276

210. Zhou J, Yu L, Ding Q, Wang R. Textile Fiber Identification Using Near-Infrared Spectroscopy and Pattern Recognition. *Autex Research Journal*. 2019;19(2):201-209. doi:10.1515/AUT-2018-0055

211. Esler MB, Griffith DWT, Wilson SR, Steele LP. Precision trace gas analysis by FT-IR spectroscopy. 1. Simultaneous analysis of CO2, CH4, N2O, and CO in air. *Analytical Chemistry*. 2000;72(1):206-215. doi:10.1021/AC9905625

212. Stuart BH. *Infrared Spectroscopy: Fundamentals and Applications*. John Wiley & Sons, Ltd; 2005. doi:10.1002/0470011149

213. van de Voort FR. Fourier transform infrared spectroscopy applied to food analysis. *Food Research International*. 1992;25(5):397-403. doi:10.1016/0963-9969(92)90115-L

214. Yang J, Xu J, Zhang X, Wu C, Lin T, Ying Y. Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Analytica Chimica Acta*. 2019;1081:6-17. doi:10.1016/J.ACA.2019.06.012

215. Hirschfeld T. New trends in the application of Fourier transform infrared spectroscopy to analytical chemistry. *Applied Optics*. 1978;17(9):1400-1412. doi:10.1364/AO.17.001400

216. Heinrich Rolf Kuehni JG. Light, Electromagnetic Spectrum. In: *Encyclopedia of Color Science and Technology*. Springer, New York, NY; 2016:843-849. doi:10.1007/978-1-4419-8071-7_204

217. Skoog DA, West DM, Holler FJ, Crouch SR. *Fundamentals of Analytical Chemistry*. 9th ed. Cengage Learning; 2013.

218. Colthup NB, Daly LH, Wiberley SE. *Introduction to Infrared and Raman Spectroscopy*. 2nd ed. Academic Press; 1990.

219. Becker ED, Farrar TC. Fourier Transform Spectroscopy: New methods dramatically improve the sensitivity of infrared and nuclear magnetic resonance spectroscopy. *Science*. 1972;178(4059):361-368. doi:10.1126/SCIENCE.178.4059.361

220. Griffiths PR, De Haseth JA. *Fourier Transform Infrared Spectrometry*. John Wiley & Sons, Inc.; 2006. doi:10.1002/047010631X

221. Hind AR, Bhargava SK, McKinnon A. At the solid/liquid interface: FTIR/ATR — the tool of choice. *Advances in Colloid and Interface Science*. 2001;93(1-3):91-114. doi:10.1016/S0001-8686(00)00079-8

222. Attenuated Total Reflectance (ATR) | Bruker. https://www.bruker.com/en/products-and-solutions/infrared-and-raman/ft-ir-routine-spectrometer/what-is-ft-ir-spectroscopy/atr-attenuated-total-reflectance.html. Accessed October 18, 2024.

223. Attenuated total reflectance (ATR) | Anton Paar Wiki. https://wiki.anton-paar.com/en/attenuated-total-reflectance-atr/. Accessed October 18, 2024.

224. Künzelmann U, Schumacher H. Characterization of surface processes during oxide CMP by in situ FTIR spectroscopy. *Advances in Chemical Mechanical Planarization (CMP)*. January 2016:359-396. doi:10.1016/B978-0-08-100165-3.00014-0

225. Oheim M, Salomon A, Weissman A, Brunstein M, Becherer U. Calibrating Evanescent-Wave Penetration Depths for Biological TIRF Microscopy. *Biophysical Journal*. 2019;117(5):795-809. doi:10.1016/J.BPJ.2019.07.048

226. Kaur H, Rana B, Tomar D, Kaur S, Jena KC. Fundamentals of ATR-FTIR Spectroscopy

and Its Role for Probing In-Situ Molecular-Level Interactions. *Progress in Optical Science and Photonics*. 2021;13:3-37. doi:10.1007/978-981-33-6084-6_1

227. Butler HJ, Smith BR, Fritzsch R, Radhakrishnan P, Palmer DS, Baker MJ. Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy. *Analyst*. 2018;143(24):6121-6134. doi:10.1039/C8AN01384E

228. Roy S, Perez-Guaita D, Andrew DW, et al. Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide. *Analytical Chemistry*. 2017;89(10):5238-5245. doi:10.1021/ACS.ANALCHEM.6B04578

229. Perez-Guaita D, Richardson Z, Heraud P, Wood B. Quantification and Identification of Microproteinuria Using Ultrafiltration and ATR-FTIR Spectroscopy. *Analytical Chemistry*. 2020;92(3):2409-2416. doi:10.1021/ACS.ANALCHEM.9B03081

230. Shi F, Ross PN, Somorjai GA, Komvopoulos K. The Chemistry of Electrolyte Reduction on Silicon Electrodes Revealed by in Situ ATR-FTIR Spectroscopy. *Journal of Physical Chemistry C*. 2017;121(27):14476-14483. doi:10.1021/ACS.JPCC.7B04132

231. Kazarian SG, Chan AKL. Attenuated total reflection Fourier-transform infrared (ATR-FTIR) imaging of tissues and live cells. *Chemical Society Reviews*. 2016;45(7):1850-1864. doi:10.1039/C5CS00515A

232. Kazarian SG, Chan AKL, Maquet V, Boccaccini AR. Characterisation of bioactive and resorbable polylactide/Bioglass® composites by FTIR spectroscopic imaging. *Biomaterials*. 2004;25(18):3931-3938. doi:10.1016/J.BIOMATERIALS.2003.10.099

233. Kumar S, Lahlali R, Liu X, Karunakaran C. Infrared spectroscopy combined with imaging: A new developing analytical tool in health and plant science. *Applied Spectroscopy Reviews*. 2016;51(6):466-483. doi:10.1080/05704928.2016.1157808

234. Becker DE. *High Resolution NMR: Theory and Chemical Applications*. 3rd ed.

Academic Press; 1999. https://www.amazon.com/High-Resolution-NMR-Chemical-Applications/dp/0120846624. Accessed October 29, 2024.

235. Macomber SR. *A Complete Introduction to Modern NMR Spectroscopy*. John Wiley and Sons; 1992.

236. James TL. *Nuclear Magnetic Resonance in Biochemistry*. Academic Press; 1975.

237. James TL. Fundamentals of NMR. In: *Nuclear Magnetic Resonance*. Quantum Device Lab; 1998:1-31. https://qudev.phys.ethz.ch/static/content/courses/phys4/studentspresentations/nmr/James_Fundamentals_of_NMR.pdf. Accessed October 30, 2024.

238. Bai S. Nuclear Magnetic Resonance Instrumentation. *Encyclopedia of Analytical Chemistry*. June 2016:1-26. doi:10.1002/9780470027318.A6108.PUB3

239. Hornak JP. *The Basics of NMR*. Rochester Institute of Technology; 1997. https://www.cis.rit.edu/htbooks/nmr/index.html. Accessed October 30, 2024.

240. Wang Y, Tang J, Zhou B, et al. Rabi oscillation of spin-polarized rubidium in the spin-exchange relaxation-free regime. *Optics Express*. 2022;30(19):35071-35084. doi:10.1364/OE.471475

241. Laukien DD, Tschopp WH. Superconducting NMR magnet design. *Concepts in Magnetic Resonance*. 1994;6(4):255-273. doi:10.1002/CMR.1820060402

242. Friebolin H. *Basic One- and Two-Dimensional NMR Spectroscopy*. 5th ed. Wiley; 2010.

243. NMR Probes Selection Guide: Types, Features, Applications. GlobalSpec. https://www.globalspec.com/learnmore/labware_scientific_instruments/spectrometers_analytical_photometers/nmr_probes. Accessed October 30, 2024.

244. Doty FD. Probe Design and Construction. In: *Encyclopedia of Magnetic Resonance*. John Wiley & Sons, Ltd; 2007. doi:10.1002/9780470034590.emrstm0414.pub2

245. Pradeep D, Tembhre MK, Parihar AS, Rao C. Magnetic resonance imaging: Basic principles and advancement in clinical and diagnostics approaches in health care. In: *Biomedical Imaging Instrumentation: Applications in Tissue, Cellular and Molecular Diagnostics*. Academic Press; 2022:45-66. doi:10.1016/B978-0-323-85650-8.00005-X

246. Sullivan TD, Parsekian AD. The effects of temperature on shallow borehole NMR measurements in permafrost. *Cold Regions Science and Technology*. 2023;211:103850. doi:10.1016/J.COLDREGIONS.2023.103850

247. Wenter P, Fürtig B, Hainard A, et al. A Caged Uridine for the Selective Preparation of an RNA Fold and Determination of its Refolding Kinetics by Real-Time NMR. *ChemBioChem*. 2006;7(3):417-420. doi:10.1002/CBIC.200500468

248. Craik DJ, Scanlon MJ. *Pharmaceutical Applications of NMR*. Vol 42. Academic Press; 2000. doi:10.1016/S0066-4103(00)42004-1

249. Edwards JC. A Review of Applications of NMR Spectroscopy in the Petroleum Industry. *Spectroscopic Analysis of Petroleum Products and Lubricants*. October 2011:423-472. doi:10.1520/MONO10117M

250. Petrakis L, Edelheit E. The Utilization of Nuclear Magnetic Resonance Spectroscopy for Petroleum, Coal, Oil Shale, Petrochemicals, and Polymers. Phenomenology, Paradigms of Applications, and Instrumentation. *Applied Spectroscopy Reviews*. 1979;15(2):195-260. doi:10.1080/05704927908060398

251. El Nokab MEH, Alassmy YA, Abduljawad MM, et al. Solid-State NMR Spectroscopy: Towards Structural Insights into Starch-Based Materials in the Food Industry. *Polymers*. 2022;14(21):4686. doi:10.3390/POLYM14214686

252. Hatzakis E. Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*. 2019;18(1):189-220. doi:10.1111/1541-4337.12408

253. Conte P, Chillura Martino DF, Lo Meo P. *The Environment in a Magnet : Applications of NMR Techniques to Environmental Problems*. Royal Society of Chemistry; 2024.

254. Glasser L. Fourier transforms for chemists. Part 1. Introduction to the Fourier transform. *Journal of Chemical Education*. 1987;64(10). doi:10.1021/ED064PA228

255. Chong YD. *Complex Methods for the Sciences*. LibreTexts; 2024.

256. Smith BC. *Fundamentals of Fourier Transform Infrared Spectroscopy*. 2nd ed. CRC Press; 2011. doi:10.1201/B10777

257. Schmiedt H, Jensen P, Schlemmer S. Rotation-vibration motion of extremely flexible molecules – The molecular superrotor. *Chemical Physics Letters*. 2017;672:34-46. doi:10.1016/J.CPLETT.2017.01.045

258. Forrey C, Douglas JF, Gilson MK. The Fundamental Role of Flexibility on the Strength of Molecular Binding. *Soft matter*. 2012;8(23):6385. doi:10.1039/C2SM25160D

259. Glasser L. Fourier transforms for chemists. Part 2. Fourier transforms in chemistry and spectroscopy. *Journal of Chemical Education*. 1987;64(11). doi:10.1021/ED064PA260

260. Abed AQ, AbdulAemah MA, Hussein SF, Abdullkadem MH. Pharmaceutical and Medical Applications of Fourier Transform Infrared Spectroscopy (FTIR), and Nuclear Magnetic Resonance (NMR). *Journal of Current Medical Research and Opinion*. 2024;7(06):2787-2800. doi:10.52845/CMRO/2024/7-6-16

261. Delsuc MA, O'Connor P. The Fourier transform in analytical science. *Nature Reviews Methods Primers*. 2024;4(1):1-16. doi:10.1038/s43586-024-00326-2

262. Seeber R, Ulrici A. Analog and digital worlds: Part 1. Signal sampling and Fourier Transform. *ChemTexts*. 2016;2(4):1-12. doi:10.1007/S40828-016-0037-1

263. Hrlimann MD, Pennington CH, Fan NQ, Clarke J, Pines A, Hahn EL. Pulsed Fourier-transform NQR of 14N with a dc SQUID. *Physical review letters*. 1992;69(4):684-687. doi:10.1103/PHYSREVLETT.69.684

264. Sonier JE. Muon spin rotation studies of electronic excitations and magnetism in the vortex cores of superconductors. *Reports on Progress in Physics*. 2007;70(11):1717. doi:10.1088/0034-4885/70/11/R01

265. Borbat PP, Costa-Filho AJ, Earle KA, Moscicki JK, Freed JH. Electron Spin Resonance in Studies of Membranes and Proteins. *Science*. 2001;291(5502):266-269. doi:10.1126/SCIENCE.291.5502.266

266. Park GB, Field RW. Perspective: The first ten years of broadband chirped pulse Fourier transform microwave spectroscopy. *Journal of Chemical Physics*. 2016;144(20). doi:10.1063/1.4952762

267. Bizet A, Nakamura N, Teramoto Y, Hatakeyama T. Measurement of dielectric properties of polymers using Fourier analysis. *Thermochimica Acta*. 1994;237(1):147-157. doi:10.1016/0040-6031(94)85194-8

268. Lange O, Damoc E, Wieghaus A, Makarov A. Enhanced Fourier transform for Orbitrap mass spectrometry. *International Journal of Mass Spectrometry*. 2014;369:16-22. doi:10.1016/J.IJMS.2014.05.019

269. Cochran D, Powers R. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Applications for Metabolomics. *Biomedicines*. 2024;12(8):1786. doi:10.3390/BIOMEDICINES12081786/S1

270. Marino RA. Application of the Fourier transform in antenna pattern measurements. *Microwave Journal*. 1998;41(10):122-127. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=01926225&v=2.1&it=r&id=GALE%7CA21269470&sid=googleScholar&linkaccess=fulltext. Accessed October 18, 2024.

271. Tivnan M, Teneggi J, Lee T-C, et al. Fourier Diffusion Models: A Method to Control MTF and NPS in Score-Based Stochastic Image Generation. March 2023. doi:10.1109/TMI.2025.3553805

272. Dilworth SJ, Koldobsky AL. The Fourier transform of order statistics with applications to Lorentz spaces. *Israel Journal of Mathematics*. 1995;92(1-3):411-425. doi:10.1007/BF02762092/METRICS

273. El-Adawi M, Abdel-Ghany S, Shalaby S. Diffusion in Semiconductors by Using Fourier Series Expansion Technique. *Material Science Research India*. 2014;11(2):90-97. doi:10.13005/MSRI/110201

274. Peters T. *Thriving on Chaos: Handbook for a Management Revolution*. Harper Perennial; 1988.

275. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*. 2012;36(4):1165-1188. doi:10.2307/41703503

276. Hair JF, Babin BJ, Anderson RE, Black WC. *Multivariate Data Analysis*. 8th ed. Cengage; 2018.

277. Johnson M, Bradshaw JM, Feltovich PJ, Jonker CM, Van Riemsdijk B, Sierhuis M. The fundamental principle of coactive design: Interdependence must shape autonomy. In: *Coordination, Organizations, Institutions, and Norms in Agent Systems VI - COIN 2010 International Workshops*. Toronto, Canada; 2010. doi:10.1007/978-3-642-21268-0_10

278. Sanchez G, Marzban E. *All Models Are Wrong: Concepts of Statistical Learning*.; 2020. https://allmodelsarewrong.github.io/. Accessed November 13, 2024.

279. Jolliffe IT. *Principal Component Analysis*. 2nd ed. New York: Springer; 2002. doi:10.1007/B98835

280. Wehrens R. *Chemometrics with R*. Springer; 2011. doi:10.1007/978-3-642-17841-2

281. Starmer J. StatQuest: Principal Component Analysis (PCA), Step-by-Step - YouTube. https://www.youtube.com/watch?v=FgakZw6K1QQ&t=871s&ab_channel=StatQuestwithJoshStarmer. Published 2018. Accessed November 13, 2024.

282. Rencher AC, William FC. *Methods of Multivariate Analysis*. 3rd ed. Wiley; 2012. doi:10.1002/9781118391686

283. Wold H. Estimation of principal components and related models by iterative least squares. *Multivariate analysis*. 1966:391-420.

284. Wold S, Ruhe A, Wold H, Dunn WJ. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM*. 1984;5(3):735-743. doi:10.1137/0905052

285. Wold H. Partial Least Squares. *Encyclopedia of Statistical Sciences*. 2004;6. doi:10.1002/0471667196.ESS1914

286. Xia Y. Correlation and association analyses in microbiome study integrating multiomics in health and disease. In: *Progress in Molecular Biology and Translational Science*. Vol 171. Academic Press; 2020:309-491. doi:10.1016/BS.PMBTS.2020.04.003

287. Dunn K. *Process Improvement Using Data*. 10d109 ed.; 2023. https://learnche.org/pid/. Accessed November 13, 2024.

288. Hair JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S. *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. Cham: Springer; 2021. doi:10.1007/978-3-030-80519-7

289. Garson DG. *Partial Least Squares Regression and Structural Equation Models*. 2016th ed. Statistical Associates; 2014.

290. Vinzi VE, Chin WW, Henseler J, Wang H. *Handbook of Partial Least Squares*. Springer; 2010. doi:10.1007/978-3-540-32827-8

291. Otto M. *Chemometrics Statistics and Computer Application in Analytical Chemistry*. 4th ed. Wiley-VCH; 2024.

292. Sedeaud A, Marc A, Marck A, et al. BMI, a Performance Parameter for Speed Improvement. *PLOS ONE*. 2014;9(2):e90183. doi:10.1371/JOURNAL.PONE.0090183

293. Madiyal A, Ajila V, Babu SG, et al. Status of thiocyanate levels in the serum and saliva of non-smokers, ex-smokers and smokers. *African Health Sciences*. 2018;18(3):727. doi:10.4314/AHS.V18I3.31

294. Sumartiningsih S, Rahayu S, Handoyo E, et al. Systemic Lactate Elevation Induced by Tobacco Smoking during Rest and Exercise Is Not Associated with Nicotine. *International Journal of Environmental Research and Public Health*. 2022;19(5):2902. doi:10.3390/IJERPH19052902

295. Palleschi G, Faridnia MH, Lubrano GJ, Guilbault GG. Determination of lactate in human saliva with an electrochemical enzyme probe. *Analytica Chimica Acta*. 1991;245(C):151-157. doi:10.1016/S0003-2670(00)80215-9

296. Barbieri A, Fuk A, Gallo G, et al. Cardiorespiratory and metabolic consequences of detraining in endurance athletes. *Frontiers in Physiology*. 2023;14:1334766. doi:10.3389/FPHYS.2023.1334766

297. Pesta DH, Angadi SS, Burtscher M, Roberts CK. The effects of caffeine, nicotine, ethanol, and tetrahydrocannabinol on exercise performance. *Nutrition & Metabolism*. 2013;10(1):1-15. doi:10.1186/1743-7075-10-71

298. Straub RH, Cutolo M, Buttgereit F, Pongratz G. Energy regulation and neuroendocrine-immune control in chronic inflammatory diseases. *Journal of Internal Medicine*. 2010;267(6):543-560. doi:10.1111/J.1365-2796.2010.02218.X

299. Lei Y, Yang J, Li H, Zhong H, Wan Q. Changes in glucose-lipid metabolism, insulin resistance, and inflammatory factors in patients with autoimmune thyroid disease. *Journal of Clinical Laboratory Analysis*. 2019;33(7):e22929. doi:10.1002/JCLA.22929

300. Villarreal-Pérez JZ, Villarreal-Martínez JZ, Lavalle-González FJ, et al. Plasma and urine metabolic profiles are reflective of altered beta-oxidation in non-diabetic obese subjects and patients with type 2 diabetes mellitus. *Diabetology and Metabolic Syndrome*. 2014;6(1):1-8. doi:10.1186/1758-5996-6-129

301. American College of Sports Medicine. *ACSM'S Metabolic Calculations Handbook*. Lippincott Williams & Wilkins; 2007.

302. Poureslami HR, Torkzadeh M, Sefadini MR. Study of changes in phosphate, calcium and fluoride ions in plaque and saliva after the administration of a fluoride mouth rinse. *Journal of the Indian Society of Pedodontics and Preventive Dentistry*. 2007;25(3):122-125. doi:10.4103/0970-4388.36561

303. Caetano PC, Strixino JF, Raniero L. Analysis of saliva by fourier transform infrared spectroscopy for diagnosis of physiological stress in athletes. *Research on Biomedical Engineering*. 2015;31(2):116-124. doi:10.1590/2446-4740.0664

304. Chiappin S, Antonelli G, Gatti R, De Palo EF. Saliva specimen: A new laboratory tool for diagnostic and basic investigation. *Clinica Chimica Acta*. 2007;383(1-2):30-40. doi:10.1016/J.CCA.2007.04.011

305. Lastovka A V., Rogachev AD, Il'ina I V., et al. Comparison of dried matrix spots and fabric phase sorptive extraction methods for quantification of highly potent analgesic activity agent (2R,4aR,7R,8aR)-4,7-dimethyl-2-(thiophen-2-yl)octahydro-2H-chromen-4-ol in rat whole blood and plasma using LC–MS/MS. *Journal of Chromatography B*. 2019;1132:121813. doi:10.1016/J.JCHROMB.2019.121813

306. Lahti M, Vilpo J, Hovinen J. Spectrophotometric Determination of Thiocyanate in Human Saliva. *Journal of Chemical Education*. 1999;76(9):1281-1282. doi:10.1021/ED076P1281

307. Pavelka V, Lazanas AC, Sarigiannidou M, Hrbac J, Prodromidis MI. Nanostructured cobalt(II) phthalocyanine modified screen-printed electrodes for the determination

of thiocyanate in human saliva. *Electroanalysis*. 2024;36(10):e202400011. doi:10.1002/ELAN.202400011

308. Morháč M. An algorithm for determination of peak regions and baseline elimination in spectroscopic data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2009;600(2):478-487. doi:10.1016/J.NIMA.2008.11.132

309. Ryabchykov O, Guo S, Bocklitz T. Analyzing Raman spectroscopic data. *Physical Sciences Reviews*. 2019;4(2). doi:10.1515/PSR-2017-0043

310. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2012;12:2825-2830. doi:10.48550/arxiv.1201.0490

311. McKinney W. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference*. SciPy; 2010:56-61. doi:10.25080/MAJORA-92BF1922-00A

312. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362. doi:10.1038/S41586-020-2649-2

313. Rieppo L, Saarakkala S, Närhi T, Helminen HJ, Jurvelin JS, Rieppo J. Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage. *Osteoarthritis and Cartilage*. 2012;20(5):451-459. doi:10.1016/J.JOCA.2012.01.010

314. Takis PG. SMolESY_platform. https://github.com/pantakis/SMolESY_platform. Published 2020. Accessed January 4, 2025.

315. Takis PG, Jiménez B, Sands CJ, Chekmeneva E, Lewis MR. SMolESY: an efficient and quantitative alternative to on-instrument macromolecular 1H-NMR signal suppression. *Chemical Science*. 2020;11(23):6000-6011. doi:10.1039/D0SC01421D

316. Correia GDS, Takis PG, Sands CJ, et al. 1H NMR Signals from Urine Excreted Protein Are a Source of Bias in Probabilistic Quotient Normalization. *Analytical Chemistry*. 2022;94(19):6919-6923. doi:10.1021/ACS.ANALCHEM.2C00466

317. Takis PG, Jiménez B, Al-Saffar NMS, et al. A Computationally Lightweight Algorithm for Deriving Reliable Metabolite Panel Measurements from 1D ^{1}H NMR. *Analytical Chemistry*. 2021;93(12):4995-5000. doi:10.1021/ACS.ANALCHEM.1C00113

318. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical chemistry*. 2006;78(13):4281-4290. doi:10.1021/AC051632C

319. Caetano Júnior PC, Lemes LC, Aguiar JC, Strixino JF, Raniero L. Application of FT-IR spectroscopy to assess physiological stress in rugby players during fatigue test. *Revista Brasileira de Engenharia Biomedica*. 2016;32(2):123-128. doi:10.1590/2446-4740.05315

320. Naseer K, Ali S, Qazi J. ATR-FTIR spectroscopy as the future of diagnostics: a systematic review of the approach using bio-fluids. *Applied Spectroscopy Reviews*. 2021;56(2):85-97. doi:10.1080/05704928.2020.1738453

321. Scott DA, Renaud DE, Krishnasamy S, et al. Diabetes-related molecular signatures in infrared spectra of human saliva. *Diabetology and Metabolic Syndrome*. 2010;2(1). doi:10.1186/1758-5996-2-48

322. Barth A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2007;1767(9):1073-1101. doi:10.1016/J.BBABIO.2007.06.004

323. Ranallo RF, Rhodes EC. Lipid metabolism during exercise. *Sports Medicine*. 1998;26(1):29-42. doi:10.2165/00007256-199826010-00003/METRICS

324. Mul JD, Stanford KI, Hirshman MF, Goodyear LJ. Exercise and Regulation of Carbohydrate Metabolism. *Progress in molecular biology and translational science*.

2015;135:17. doi:10.1016/BS.PMBTS.2015.07.020

325. Qi X, Tester RF. Fructose, galactose and glucose – In health and disease. *Clinical Nutrition ESPEN*. 2019;33:18-28. doi:10.1016/J.CLNESP.2019.07.004

326. Murray B, Rosenbloom C. Fundamentals of glycogen metabolism for coaches and athletes. *Nutrition Reviews*. 2018;76(4):243-259. doi:10.1093/NUTRIT/NUY001

327. Nystoriak MA, Bhatnagar A. Cardiovascular Effects and Benefits of Exercise. *Frontiers in Cardiovascular Medicine*. 2018;5:408204. doi:10.3389/FCVM.2018.00135

328. Kumar G, Pandey V, Thapa RK, Weldon A, Granacher U, Ramirez-Campillo R. Effects of Exercise Frequency with Complex Contrast Training on Measures of Physical Fitness in Active Adult Males. *Sports*. 2023;11(1):11. doi:10.3390/SPORTS11010011

329. Rippetoe M, Baker A. *Practical Programming for Strength Training*. 3rd ed. The Aasgaard Company; 2013.

330. Nalbandian M, Takeda M. Lactate as a Signaling Molecule That Regulates Exercise-Induced Adaptations. *Biology*. 2016;5(4):38. doi:10.3390/BIOLOGY5040038

331. Cheng Y, Zheng G, Song Z, Zhang G, Rao X, Zeng T. Association between dietary protein intake and risk of chronic kidney disease: a systematic review and meta-analysis. *Frontiers in Nutrition*. 2024;11:1408424. doi:10.3389/FNUT.2024.1408424

332. Argilés JM, Campos N, Lopez-Pedrosa JM, Rueda R, Rodriguez-Mañas L. Skeletal Muscle Regulates Metabolism via Interorgan Crosstalk: Roles in Health and Disease. *Journal of the American Medical Directors Association*. 2016;17(9):789-796. doi:10.1016/J.JAMDA.2016.04.019

333. Bel'skaya L V., Sarf EA, Makarova NA. Use of Fourier Transform IR Spectroscopy for the Study of Saliva Composition. *Journal of Applied Spectroscopy*. 2018;85(3):445-451. doi:10.1007/S10812-018-0670-0

334. Bel'skaya L V., Sarf EA, Solomatin D V. Age and Gender Characteristics of the Infrared Spectra of Normal Human Saliva. *Applied Spectroscopy*. 2020;74(5):536-543. doi:10.1177/0003702819885958

335. Mikkonen JJW, Raittila J, Rieppo L, Lappalainen R, Kullaa AM, Myllymaa S. Fourier Transform Infrared Spectroscopy and Photoacoustic Spectroscopy for Saliva Analysis. *Applied Spectroscopy*. 2016;70(9):1502-1510. doi:10.1177/0003702816654149

336. Dewhurst-Trigg R, Yeates T, Blackwell JR, et al. Lowering of blood pressure after nitrate-rich vegetable consumption is abolished with the co-ingestion of thiocyanate-rich vegetables in healthy normotensive males. *Nitric Oxide*. 2018;74:39-46. doi:10.1016/J.NIOX.2018.01.009

337. Connolly EL, Sim M, Travica N, et al. Glucosinolates From Cruciferous Vegetables and Their Potential Role in Chronic Disease: Investigating the Preclinical and Clinical Evidence. *Frontiers in Pharmacology*. 2021;12:767975. doi:10.3389/FPHAR.2021.767975

338. Aggarwal A, Keluskar V, Goyal R, Dahiya P. Salivary thiocyanate: A biochemical indicator of cigarette smoking in adolescents. *Oral Health and Preventive Dentistry*. 2013;11(3):221-227. doi:10.3290/j.ohpd.a30169

339. Rodrigues RPCB, Aguiar EMG, Cardoso-Sousa L, et al. Differential Molecular Signature of Human Saliva Using ATR-FTIR Spectroscopy for Chronic Kidney Disease Diagnosis. *Brazilian Dental Journal*. 2019;30(5):437-445. doi:10.1590/0103-6440201902228

340. Flieger J, Kawka J, Tatarczak-Michalewska M. Levels of the Thiocyanate in the Saliva of Tobacco Smokers in Comparison to e-Cigarette Smokers and Nonsmokers Measured by HPLC on a Phosphatidylcholine Column. *Molecules*. 2019;24(20):3790. doi:10.3390/MOLECULES24203790

341. Magacz M, Kędziora K, Sapa J, Krzyściak W. The Significance of Lactoperoxidase System in Oral Health: Application and Efficacy in Oral Hygiene Products. *International Journal of Molecular Sciences*. 2019;20(6):1443. doi:10.3390/IJMS20061443

342. Pontikides N, Krassas G. Influence of cigarette smoking on thyroid function, goiter formation and autoimmune thyroid disorders. *Hormones*. 2002;1(2):91-98. doi:10.14310/HORM.2002.1156

343. Mul JD, Stanford KI, Hirshman MF, Goodyear LJ. Exercise and Regulation of Carbohydrate Metabolism. *Progress in molecular biology and translational science*. 2015;135:17. doi:10.1016/BS.PMBTS.2015.07.020

344. Rinnan Å, Berg F van den, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*. 2009;28(10):1201-1222. doi:10.1016/J.TRAC.2009.07.007

345. Vieira CAGA, Pupin B, Bhattacharjee TT, Sakane KK. Infrared Spectroscopy Based Study of Biochemical Changes in Saliva during Maximal Progressive Test in Athletes. *Analytical Sciences*. 2021;37(8):1157-1163. doi:10.2116/analsci.20P395

346. Barauna VG, Singh MN, Barbosa LL, et al. Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity. *Analytical Chemistry*. 2021;93(5):2950-2958. doi:10.1021/acs.analchem.0c04608

347. Rohman A, Windarsih A, Lukitaningsih E, Rafi M, Betania K, Fadzillah NA. The use of FTIR and Raman spectroscopy in combination with chemometrics for analysis of biomolecules in biomedical fluids: A review. *Biomedical Spectroscopy and Imaging*. 2020;8(3-4):55-71. doi:10.3233/BSI-200189

348. Campanella B, Legnaioli S, Onor M, Benedetti E, Bramanti E. The Role of the Preanalytical Step for Human Saliva Analysis via Vibrational Spectroscopy.

Metabolites. 2023;13(3):393. doi:10.3390/METABO13030393/S1

349. Cui Y, Zhang H, Zhu J, Liao Z, Wang S, Liu W. Correlations of Salivary and Blood Glucose Levels among Six Saliva Collection Methods. *International Journal of Environmental Research and Public Health*. 2022;19(7):4122. doi:10.3390/IJERPH19074122

350. Caixeta DC, Lima C, Xu Y, et al. Monitoring glucose levels in urine using FTIR spectroscopy combined with univariate and multivariate statistical methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2023;290:122259. doi:10.1016/J.SAA.2022.122259

351. Helge JW, Ben J, Willer M, Daugaard JR, Storlien LH, Kiens B. Training affects muscle phospholipid fatty acid composition in humans. *Journal of Applied Physiology*. 2001;90(2):670-677. doi:10.1152/JAPPL.2001.90.2.670

352. Andersson A, Sjödin A, Hedman A, Olsson R, Vessby B. Fatty acid profile of skeletal muscle phospholipids in trained and untrained young men. *American Journal of Physiology - Endocrinology and Metabolism*. 2000;279(4 42-4). doi:10.1152/AJPENDO.2000.279.4.E744

353. Andersson A, Sjödin A, Olsson R, Vessby B. Effects of physical exercise on phospholipid fatty acid composition skeletal muscle. *American Journal of Physiology - Endocrinology and Metabolism*. 1998;274(3 37-3). doi:10.1152/AJPENDO.1998.274.3.E432

354. Khaustova S, Shkurnikov M, Tonevitsky E, Artyushenko V, Tonevitsky A. Noninvasive biochemical monitoring of physiological stress by Fourier transform infrared saliva spectroscopy. *Analyst*. 2010;135(12):3183-3192. doi:10.1039/c0an00529k

355. Bel'skaya L V., Sarf EA. Biochemical composition and characteristics of salivary FTIR spectra: Correlation analysis. *Journal of Molecular Liquids*. 2021;341:117380.

doi:10.1016/J.MOLLIQ.2021.117380

356. Welk A, Rudolph P, Kreth J, Schwahn C, Kramer A, Below H. Microbicidal efficacy of thiocyanate hydrogen peroxide after adding lactoperoxidase under saliva loading in the quantitative suspension test. *Archives of Oral Biology*. 2011;56(12):1576-1582. doi:10.1016/J.ARCHORALBIO.2011.04.016

357. Barnard JP, Stinson MW. Influence of environmental conditions on hydrogen peroxide formation by Streptococcus gordonii. *Infection and Immunity*. 1999;67(12):6558-6564. doi:10.1128/IAI.67.12.6558-6564.1999

358. Jagim AR, Camic CL, Askow A, et al. Sex Differences in Resting Metabolic Rate Among Athletes. *Journal of strength and conditioning research*. 2019;33(11):3008-3014. doi:10.1519/JSC.0000000000002813

359. Molgat-Seon Y, Peters CM, Sheel AW. Sex-differences in the human respiratory system and their impact on resting pulmonary function and the integrative response to exercise. *Current Opinion in Physiology*. 2018;6:21-27. doi:10.1016/J.COPHYS.2018.03.007

360. Li-Hui W, Chuan-Quan L, Long Y, Ru-Liu L, Long-Hui C, Wei-Wen C. Gender differences in the saliva of young healthy subjects before and after citric acid stimulation. *Clinica Chimica Acta*. 2016;460:142-145. doi:10.1016/J.CCA.2016.06.040

361. Inoue H, Ono K, Masuda W, et al. Gender difference in unstimulated whole saliva flow rate and salivary gland sizes. *Archives of Oral Biology*. 2006;51(12):1055-1060. doi:10.1016/J.ARCHORALBIO.2006.06.010

362. Valdés MG, Díaz-García ME. Determination of Thiocyanate within Physiological Fluids and Environmental Samples: Current Practice and Future Trends. *Critical Reviews in Analytical Chemistry*. 2004;34(1):9-23. doi:10.1080/10408340490273726

363. Cendali F, D'Alessandro A, Nemkov T. Dried blood spot characterization of sex-based metabolic responses to acute running exercise. *Analytical Science Advances*. 2023;4(1-2):37-48. doi:10.1002/ANSA.202200039

364. Reiter B. The Lactoperoxidase-Thiocyanate-Hydrogen Peroxide Antibacterium System. In: Fitzsimons DW, ed. *Oxygen Free Radicals and Tissue Damage*. John Wiley & Sons, Ltd; 1978:285-294. doi:10.1002/9780470715413.CH16

365. Grzesiak-Gasek I, Kaczmarek U. Influence of Swimming Training Session on Selected Saliva Components in Youth Swimmers. *Frontiers in Physiology*. 2022;13:869903. doi:10.3389/FPHYS.2022.869903

366. Damirchi A, Saati Zareei A, Sariri R. Salivary antioxidants of male athletes after aerobic exercise and garlic supplementation on: A randomized, double blind, placebo-controlled study. *Journal of Oral Biology and Craniofacial Research*. 2015;5(3):146-152. doi:10.1016/J.JOBCR.2015.08.001

367. Sariri R, Damirchi A. Salivary antioxidant variations in athletes after intense exercise. *Medicina Sportiva*. 2013;9(1):2043-2050.

368. de Bari L, Valenti D, Atlante A, Passarella S. l-Lactate generates hydrogen peroxide in purified rat liver mitochondria due to the putative l-lactate oxidase localized in the intermembrane space. *FEBS Letters*. 2010;584(11):2285-2290. doi:10.1016/J.FEBSLET.2010.03.038

369. Passarella S, de Bari L, Valenti D, Pizzuto R, Paventi G, Atlante A. Mitochondria and l-lactate metabolism. *FEBS Letters*. 2008;582(25-26):3569-3576. doi:10.1016/J.FEBSLET.2008.09.042

370. Bartoloni B, Mannelli M, Gamberi T, Fiaschi T. The Multiple Roles of Lactate in the Skeletal Muscle. *Cells*. 2024;13(14):1177. doi:10.3390/CELLS13141177

371. Sands CJ, Wolfer AM, Correia GDS, et al. The nPYc-Toolbox, a Python module for the pre-processing, quality-control and analysis of metabolic profiling datasets.

Bioinformatics (Oxford, England). 2019;35(24):5359-5360. doi:10.1093/BIOINFORMATICS/BTZ566

372. Hinojosa JN, Hearon CM, Kowalsky RJ. Blood lactate response to active recovery in athletes vs. non-athletes. *Sport Sciences for Health*. 2021;17(3):699-705. doi:10.1007/S11332-021-00735-W

373. Lee S, Choi Y, Jeong E, et al. Physiological significance of elevated levels of lactate by exercise training in the brain and body. *Journal of Bioscience and Bioengineering*. 2023;135(3):167-175. doi:10.1016/J.JBIOSC.2022.12.001

374. Todd JJ, Stewart K. Lactate: valuable for physical performance and maintenance of brain function during exercise. *Bioscience Horizons: The International Journal of Student Research*. 2014;7:2014. doi:10.1093/BIOHORIZONS/HZU001

375. Zając B. Analysis of Course of Changes in Blood Lactate Concentration in Response to Graded Exercise Test and Modified Wingate Test in Adolescent Road Cyclists. *Journal of Clinical Medicine*. 2024;13(2):535. doi:10.3390/JCM13020535

376. San-Millán I, Stefanoni D, Martinez JL, Hansen KC, D'Alessandro A, Nemkov T. Metabolomics of Endurance Capacity in World Tour Professional Cyclists. *Frontiers in Physiology*. 2020;11:578. doi:10.3389/FPHYS.2020.00578

377. Mastalerz A, Johne M, Mróz A, et al. Changes of Anaerobic Power and Lactate Concentration following Intense Glycolytic Efforts in Elite and Sub-Elite 400-meter Sprinters. *Journal of human kinetics*. 2024;91(Spec Issue):165-174. doi:10.5114/JHK/186074

378. Hirvonen J, Nummela A, Rusko H, Rehunen S, Härkönen M. Fatigue and changes of ATP, creatine phosphate, and lactate during the 400-m sprint. *Canadian Journal of Sport Sciences*. 1992;17(2):141-144. doi:10.1080/02640419208729920

379. Ohkuwa T, Miyamura M. Peak blood lactate after 400 m sprinting in sprinters and long-distance runners. *The Japanese journal of physiology*. 1984;34(3):553-556.

doi:10.2170/JJPHYSIOL.34.553

380. Weiss A, Alack K, Klatt S, et al. Sustained Endurance Training Leads to Metabolomic Adaptation. *Metabolites*. 2022;12(7):658. doi:10.3390/METABO12070658

381. Gailiūnienė A, Stasiulis A, Michailovienė J. The Effect of Submaximal Exercise on Blood Creatinine, Urea, Total Protein and Uric Acid Levels of Trained and Untrained Subjects. *Baltic Journal of Sport and Health Sciences*. 2007;3(66):5-10. doi:10.33607/BJSHS.V3I66.542

382. Schranner D, Kastenmüller G, Schönfelder M, Römisch-Margl W, Wackerhage H. Metabolite Concentration Changes in Humans After a Bout of Exercise: a Systematic Review of Exercise Metabolomics Studies. *Sports Medicine - Open*. 2020;6(1):1-17. doi:10.1186/S40798-020-0238-4

383. Hudson JF, Phelan MM, Owens DJ, Morton JP, Close GL, Stewart CE. "Fuel for the Damage Induced": Untargeted Metabolomics in Elite Rugby Union Match Play. *Metabolites*. 2021;11(8):544. doi:10.3390/METABO11080544

384. Bloom SR, Johnson RH, Park DM, Rennie MJ, Sulaiman WR. Differences in the metabolic and hormonal response to exercise between racing cyclists and untrained individuals. *The Journal of Physiology*. 1976;258(1):1-18. doi:10.1113/JPHYSIOL.1976.SP011403

Appendix 1

Athlete Questionnaire

Lab of Analytical Chemistry
Chemistry Dpt
University of Ioannina

Sample code number _____

1. Age _____ 2. Height _____ 3. Weight _____

Basic Information

4. How often do you practice per week?

☐ 1-2 times / week ☐ 3-4 times / week

☐ ≥5 times / week

5. Which is the main time period of training per day?

☐ 30-45 minutes ☐ 45-60 minutes

☐ 60-75 minutes ☐ 75+ minutes

6. Do you belong to any sports club, athlete union or sports team? (Gym is not included as a positive answer)

☐ Yes ☐ No

7. Are you trained by a professional coach?

☐ Yes ☐ No

8. Do you push yourself to higher training loads?

☐ Yes ☐ No

9. How would you characterize your training load?

☐ Low ☐ Intermediate ☐ Heavy

Medication and Suppliments

12. Do you use Ephedrine or any other energy boosters / weight cutters?

☐ Yes ☐ No

13. Do you use Protein or Creatine or any other Weight Gainers?

☐ Yes ☐ No

14. Do you use energy drinks? ("Red Bull", "Rock Star", Etc)

☐ Yes ☐ No

15. Do you use anabolic steroids or steroids of any sort?

☐ Yes ☐ No

16. Do you use any other hormones? (HGH, Insulin, Thyroxine, Etc)

☐ Yes ☐ No

17. Do you take anything to enhance recovery from training?

☐ Yes ☐ No

18. Have you taken ANY prescription medications or other substances in past 3 months?

☐ Yes ☐ No

☐ I hereby consent to the processing of the personal data that I have provided and declare my agreement with the data protection regulations in the data privacy statement

10. What are your main motivations for doing athletics? (Why do you do athletics?) (You can fill more than one options)

- ☐ I am hobbyist
- ☐ I just want to have fun with my friends
- ☐ I want to lose weight
- ☐ I want to increase my fitness level
- ☐ I am trying to improve my skills in this sport
- ☐ I aim to participate in future championships

11. Based on the below definitions, how would you decribe yourself?

☐ Low-level athlete ☐ High-level athlete

Low-level athletes are individuals who engage in physically demanding sporting activities typically on the weekends despite minimal physical activity during the work week. These athletes perform high-intensity workout regimen without the proper preparation, probably leading to an increased risk of injury. These weekend athletes perform solely for pleasure, while on the other hand high level athletes continuously strive to meet a perfect physical standard.

High-level athletes are usually drafted in higher rounds or playing in higher divisions and are perceived as having greater performance ability than that of their peers in the same sport and play at a higher level within a sport (division I vs II, professional vs amateur).

Appendix 2

Athlete Questionnaire

Lab of Analytical Chemistry
Chemistry Dpt
University of Ioannina

Sample code number _____

1. Date _____ 2. Sport _____ 3. Sport club _____

4. Name _____

5. Gender ☐ Male ☐ Female

6. Age _____ 7. Height _____ 8. Weight _____

Basic Information

9. Are you a smoker?

☐ Yes ☐ No

10. When was the last time you smoked?

☐ less than an hour ☐ 1-2 hours before

☐ more than 2 hours

11. When was the last time you consumed any meal?

☐ less than an hour ☐ 1-2 hours before

☐ more than 2 hours

12. When was the last time you drunk coffee?

☐ less than an hour ☐ 1-2 hours before

☐ more than 2 hours

13. When was the last time you had any oral hygiene procedure?

☐ less than an hour ☐ 1-2 hours before

☐ more than 2 hours

Medication and Suppliments

14. Do you use Ephedrine or any other energy boosters / weight cutters?

☐ Yes ☐ No

15. Do you use Protein or Creatine or any other Weight Gainers?

☐ Yes ☐ No

16. Do you use anabolic steroids or steroids of any sort?

☐ Yes ☐ No

17. Do you use any other hormones? (HGH, Insulin, Thyroxine, Etc)

☐ Yes ☐ No

18. Do you take anything to enhance recovery from training?

☐ Yes ☐ No

19. Have you taken ANY prescription medications or other substances in past 3 months?

☐ Yes ☐ No

Statement of informed consent for participation

☐ I hereby consent to the processing of the personal data that I have provided and declare my agreement with the data protection regulations in the data privacy statement.

☐ I am fully aware of the implications of publication the research findings online in open access format and accept any associated risk.

Athlete signature

_____

Appendix 3

```
% --------------------------Inport the csv----------------------

Matrix = readtable('spectra.csv');

X = Matrix(2:end,3:end);
Y = Matrix(2:end,2);

% --------------------------Table to array----------------------

X = table2array(X);
Y = table2array(Y);

whos X Y
[numSamples, numWavenumbers] = size(X);

%--------------------------Loading the Spectra-------------------

[dummy,h] = sort(Y);
oldorder = get(gcf,'DefaultAxesColorOrder');
set(gcf,'DefaultAxesColorOrder',jet(numSamples));
plot3(repmat(1:numWavenumbers,numSamples,1)',...
      repmat(Y(h),1,numWavenumbers)',X(h,:)');
set(gcf,'DefaultAxesColorOrder',oldorder);
xlabel('Variable / Wavenumber'); ylabel('Samples'); axis('tight');
grid on

% --------------------------Split the data----------------------
```

```
% Find unique groups in Y (assuming Y contains categorical labels
% or integers representing groups)
uniqueGroups = unique(Y);

% Initialize cell array to store indices for each group
groups = cell(length(uniqueGroups), 1);

% Assign samples to corresponding groups based on Y labels
for i = 1:length(uniqueGroups)
    % Find indices of samples belonging to each group
    groups{i} = find(Y == uniqueGroups(i));
end

% Initialize the training and testing indices
trainIndices = [];
testIndices = [];

% Randomly split each group (70% training, 30% testing)

for i = 1:length(groups)
    groupIndices = groups{i};
    numGroupSamples = length(groupIndices);

    % Shuffle the group indices
    shuffledGroupIndices                                    =
groupIndices(randperm(numGroupSamples));

    % Split the group into training and testing sets
    % (70% training, 30% testing)
    numTrain = round(0.7 * numGroupSamples);  % 70% for training
```

```
    trainIndices                    =                    [trainIndices;
shuffledGroupIndices(1:numTrain)];
    testIndices                     =                     [testIndices;
shuffledGroupIndices(numTrain+1:end)];
end

% Create the training and testing data
X_train = X(trainIndices, :);
Y_train = Y(trainIndices);

X_test = X(testIndices, :);
Y_test = Y(testIndices);

Y_train_D = dummyvar(Y_train);
Y_test_D = dummyvar(Y_test);

% --------------------------------PLS-DA---------------------------

[n,p] = size(X_train);
[Xloadings,Yloadings,Xscores,Yscores,betaPLS,PLS_percent_of_varian
ce,...
 PLS_MSEcv,stats] = plsregress(X_train,Y_train_D,20,'CV',10);

% ---------------------------------PCA-----------------------------

[PCALoadings,PCAScores,...
 PCA_percent_of_variance] = pca(X_train,'Economy',false);

% ----------------------------Y vs comps-------------------------

figure
```

```
plot(1:20,cumsum(100*PLS_percent_of_variance(2,:)),'-bo');
xlabel('Number of Components');
ylabel('Percent Variance Explained in Y');

% ----------------------------X vs comps-------------------------

figure
plot(1:20,100*cumsum(PLS_percent_of_variance(1,:)),'b-o',...
     1:20,100*cumsum(PCA_percent_of_variance(1:20))/...
     sum(PCA_percent_of_variance(1:20)),'r-o');
xlabel('Number of Components');
ylabel('Percent Variance Explained in X');
legend({'PLS-DA', 'PCA-LDA'},'location','SE');

% -------------------------Cross Validation-----------------------

PCA_MSEcv = sum(crossval(@pcrsse,X_train,Y_train,'KFold',10),1) /
n;

figure
plot(0:20,PLS_MSEcv(2,:),'b-o', 0:10,PCA_MSEcv,'r-o');
xlim([1,20])
xlabel('Number of Components');
ylabel('MSECV');
legend({'PLS-DA','PCA-LDA'},'location','NE');

% ----------------------------Fitting --------------------------

betaPCR = regress(Y_train-mean(Y_train), PCAScores(:,1:10));
betaPCR = PCALoadings(:,1:10)*betaPCR;
betaPCR = [mean(Y_train) - mean(X_train)*betaPCR; betaPCR];
```

```
yfitPCR = [ones(n,1) X_train]*betaPCR;

TSS_nonD = sum((Y_train-mean(Y_train)).^2);

RSS_PCR = sum((Y_train-yfitPCR).^2);
RsquaredPCR = 1 - RSS_PCR./TSS_nonD


yfitPLS = [ones(n,1) X_train]*betaPLS;

TSS = sum((Y_train_D-mean(Y_train_D)).^2);

RSS_PLS = sum(stats.Yresiduals.^2);
R2PLS = 1 - RSS_PLS./TSS;
R2PLS_uniform = mean(R2PLS)


% ----------------------------Score plot-------------------------

figure
hold on
y_uni = unique(Y_train);
colors = hsv(length(y_uni));
for k = 1 : length(y_uni)

    % Get indices of this particular unique group:
    ind = Y_train==y_uni(k);
    plot3(Xscores(ind,1),Xscores(ind,2),Xscores(ind,3),...
          '.','color',colors(k,:),'markersize',20);

end
```

```
legend('R (0 km/h)','W (4-5 km/h)','J (9-10 km/h)','RN (14-15
km/h)');
title('PLS-DA');
grid on; view(-30,30);
xlabel('LV1'); ylabel('LV2'); zlabel('LV3');


figure
hold on
y_uni=unique(Y_train);
colors = hsv(length(y_uni));
for k = 1 : length(y_uni)

    % Get indices of this particular unique group:
    ind = Y_train==y_uni(k);
    plot3(PCAScores(ind,1),PCAScores(ind,2),PCAScores(ind,3),...
          '.','color',colors(k,:),'markersize',20);

end

legend('R (0 km/h)','W (4-5 km/h)','J (9-10 km/h)','RN (14-15
km/h)');
title('PCA');
grid on; view(-30,30);
xlabel('PC1'); ylabel('PC2');zlabel('PC3');

% ----------------------------Loading plot-----------------------

figure
plot(1:numWavenumbers,stats.W(:,1:2),'-');
```

```
xlabel('Variable / Wavenumber');
ylabel('PLS Weights');
legend({'LV1' 'LV2'},  ...
      'location','NW');

figure
plot(1:numWavenumbers,PCALoadings(:,1:2),'-');
xlabel('Variable / Wavenumber');
ylabel('PCA Loadings');
legend({'PC1' 'PC2'},'location','NW');


% ----------------------------Prediction-------------------------

Y_pred = [ones(size(X_test,1),1) X_test]*betaPLS;

[s1, s2] = size(Y_pred);

Y_pred_D = Y_pred;

for i = 1 : s1
    for j = 1 : s2
        if Y_pred_D(i,j) < max(Y_pred(i,:))
            Y_pred_D(i,j) = 0.0;
        end
    end
end

for i = 1 : s1
    for j = 1 : s2
        if Y_pred_D(i,j) > 0.0
```

```
            Y_pred_D(i,j) =1.0;
        end
    end
end

g1 = 0;
for i = 1 : s1
        if Y_pred_D(i,:) == Y_test_D(i,:)
            g1 = g1 + 1;
        end
end

percentage_of_success = (g1/s1)*100


[~,Y_pred_unD] = max(Y_pred_D, [], 2);


figure
confusionchart(Y_test,Y_pred_unD)
title("Confusion matrix");
```