

Víctor Rubias Rincón

Quantificació de compostos de baix pes molecular a partir de l'espectre imaginari d'espectroscòpia 1H-RMN

Treball Fi de Grau
dirigit pel Dr. Xavier Correig Blanchar

Grau d'Enginyeria en Electrònica Industrial i Automàtica



UNIVERSITAT ROVIRA I VIRGILI

Tarragona
2023

Índex

| | |
|--|----|
| 1. Introducció..... | 5 |
| 1.1 Motivacions i objectius..... | 5 |
| 1.2 Desenvolupament del projecte | 6 |
| 2. Estat de l'art..... | 7 |
| 2.1 Metabolòmica | 7 |
| 2.2 RMN (Ressonància Magnètica Nuclear)..... | 7 |
| 2.3 Compostos de baix pes molecular | 10 |
| 2.4 Polsos de RMN | 10 |
| 2.4.1 Noesy | 11 |
| 2.4.2 CPMG | 11 |
| 2.4.3 Altres polsos..... | 12 |
| 2.5 SMoIESY | 12 |
| 2.6 SMoIESY Platform..... | 15 |
| 3. Mètodes i procediments..... | 17 |
| 3.1 L'entorn de programació: MATLAB..... | 17 |
| 3.2 Estadística: coeficient de correlació de Pearson..... | 17 |
| 3.3 Conjunt de mostres utilitzat: Carratalà | 18 |
| 3.4 Preprocessat de les dades..... | 19 |
| 3.4.1 Importació de les dades a format apte per a MATLAB | 19 |
| 3.4.2 Transformació dels espectres Noesy a SMoIESY..... | 20 |
| 3.4.3 Representació de les dades | 21 |
| 3.4.4 Interpolació: correcció del problema de resolució | 21 |
| 4. Resultats..... | 24 |
| 4.1 Mètodes de quantificació..... | 24 |
| 4.1.1 Amplitud | 25 |
| 4.1.2 Àrea (integració)..... | 26 |
| 4.2 Programació de la quantificació dels compostos | 29 |
| 4.2.1 Interfície gràfica | 30 |
| 4.2.2 Transformada SMoIESY | 31 |
| 4.2.3 Interpolació..... | 31 |
| 4.2.4 Integració | 31 |
| 4.2.5 Correlació..... | 33 |
| 4.3 Correlacions (resultats)..... | 33 |
| 4.4 Discussió dels resultats | 39 |

| | | |
|-----|--|----|
| 4.5 | Estudi posterior de l'espectre: altres compostos d'interès..... | 41 |
| 5. | Conclusions..... | 46 |
| 6. | Referències | 47 |
| 7. | Annexos..... | 49 |
| | Annex 1: Algoritme d'importació de les dades a format MATLAB..... | 49 |
| | Annex 2: Algoritme per eliminar el codi identificatiu de l'arxiu de text que conté l'espectre SMoIESY..... | 50 |
| | Annex 3: Algoritme per importar els espectres transformats des de la plataforma SMoIESY en arxius independents. | 51 |
| | Annex 4: Algoritmes diversos per representar els espectres a demanda..... | 52 |
| | Annex 5: Algoritme de quantificació mitjançant l'amplitud | 53 |
| | Annex 6: Algoritme de quantificació mitjançant l'àrea (rectificant línia base) | 54 |
| | Annex 7: Algoritme de quantificació mitjançant l'àrea (calculant recta base) | 55 |
| | Annex 8: Algoritme de quantificació mitjançant l'àrea (integració directa)..... | 56 |
| | Annex 9: Algoritme de quantificació mitjançant l'àrea (referenciant punts mínims) ... | 57 |
| | Annex 10: Algoritme de la interfície gràfica | 58 |
| | Annex 11: Algoritme de la transformada de l'espectre Noesy a SMoIESY..... | 59 |
| | Annex 12: Algoritme d'interpolació | 60 |
| | Annex 13: Algoritme de quantificació..... | 61 |
| | Annex 14: Algoritme de correlació..... | 62 |

1. Introducció

Aquest Treball de Final de Grau s'ha desenvolupat en col·laboració amb *Biosfer Teslab*, una empresa *spin-off* de la Universitat Rovira i Virgili (URV) i de l'Institut de Recerca Sanitària Pere Virgili (IISPV) especialitzada en el diagnòstic in vitro. L'objectiu de la companyia és proporcionar serveis analítics per a l'estudi i seguiment de les alteracions del metabolisme amb la finalitat d'avançar en el coneixement i millorar la salut de les persones.

Encabit en el camp de la metabolòmica, que és la ciència que estudia els metabòlits, substàncies presents en les cèl·lules i els teixits de l'organisme, aquest projecte té com a finalitat avaluar l'aplicabilitat de la tècnica SMoIESY per a la quantificació de compostos de baix pes molecular en mostres biològiques. Aquesta tècnica permet utilitzar la part imaginària dels espectres Noesy per quantificar els compostos de baix pes molecular.

Aquesta quantificació es realitza mitjançant l'estudi i tractament dels espectres obtinguts amb la tècnica d'espectroscòpia de Ressonància Magnètica Nuclear de protó, una tècnica fiable i robusta que és cabdal en la recerca bioquímica i biomèdica. Els beneficis principals que presenta la tècnica SMoIESY són el d'incrementar la resolució de les molècules de baix pes i el de filtrar la contribució de les molècules més grans, facilitant la quantificació de les primeres. En aquest projecte s'han desenvolupat algorismes per transformar l'espectre Noesy en SMoIESY, quantificar-ne els compostos presents en les mostres i correlar-ho amb altres quantificacions ja existents.

Per tal de realitzar aquesta comprovació s'ha sotmès a estudi un conjunt de 151 mostres de sèrum humà (anomenat Carratalà) en el que prèviament s'han quantificat els compostos de baix pes molecular a partir dels espectres CPMG mitjançant la tècnica de l'ajust de corbes.

1.1 Motivacions i objectius

La finalitat del projecte és avaluar l'aplicabilitat de la tècnica SMoIESY per quantificar compostos de baix pes molecular i programar un algorisme per implementar-lo en Matlab.

Per tal d'avaluar l'aplicabilitat de la quantificació mitjançant espectres SMoIESY es defineixen les següents tasques:

- 1) Processament d'espectres de 1H-RMN
- 2) Quantificació de compostos de baix pes molecular a partir de l'espectre imaginari
- 3) Validació de les quantificacions en diferents conjunts de mostres de sang amb d'eines d'anàlisi multivariants i estadístiques
- 4) Integració de l'algorisme de processat i quantificació en el flux de treball en metabolòmica 1H-RMN

1.2 Desenvolupament del projecte

Per tal de complir els objectius anteriors i tenint en compte les diferents circumstàncies que s'han anat donant al llarg del procés, la realització del projecte es pot dividir en cinc fases de treball que culminen en la realització de la present memòria. Són les següents, ordenades de manera cronològica:

- Llegida de documentació i articles sobre el camp d'estudi a nivell general i sobre SMoESY en particular
- Importació dels espectres al format MATLAB i familiarització amb l'entorn de treball.
- Transformació dels espectres
- Exploració, comparació i tria de diferents mètodes de quantificació
- Millora/correcció de la resolució dels espectres mitjançant interpolació cúbica
- Càlcul de les correlacions (resultats finals) entre les quantificacions obtingudes prèviament amb els espectres CPMG per Biosfer Teslab i les obtingudes en aquest TFG amb el mètode SMoESY
- Integració/unificació final dels algorismes elaborats



Figura 1. Diagrama del flux del projecte (Font: elaboració pròpia)

2. Estat de l'art

2.1 Metabolòmica

Aquesta ciència, inclosa en el conjunt de les anomenades ciències òmiques [1] juntament amb la genòmica, transcriptòmica, epigenòmica i proteòmica, és la que s'encarrega de l'estudi del metaboloma. Aquest es defineix com el conjunt complet de les petites molècules (les de pes inferior a 1500 Da) [2], anomenades metabòlits, que es troben a qualsevol mostra biològica de l'organisme, com a resultat de la vida i els intercanvis cel·lulars. Alguns d'aquests metabòlits són els lípids, els sucres i els aminoàcids.

Gràcies a l'anàlisi de les respostes metabòliques als canvis en l'organisme, aquesta ciència és especialment important per a la detecció prematura de malalties i desenvolupament de tractaments per aquestes, així com la identificació de certs biomarcadors i l'obtenció d'informació nutricional i toxicològica [3].

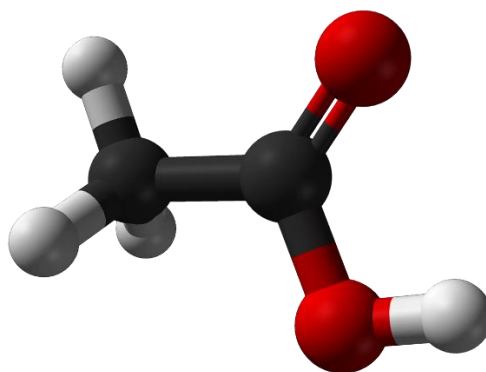


Figura 2. Estructura molecular de l'acetat (àcid acètic).

Les tècniques principals que s'utilitzen en metabolòmica són l'espectrometria de masses (MS) i la Ressonància Magnètica Nuclear (NMR o RMN), que s'utilitzen tant separatament com conjunta per tal de detectar els metabòlits presents i les seves concentracions respectives. Tot i que la MS té una major sensibilitat, la RMN es considera una tècnica més robusta, ja que requereix una mínima manipulació prèvia de les mostres i redueix el temps d'adquisició de les mateixes, tot conformant un mètode amb major reproductibilitat i fent-lo més adequat per a estudis a gran escala [4].

Les àrees dels pics dels espectres obtinguts per RMN estan directament relacionats amb les concentracions molars de certs metabòlits

2.2 RMN (Ressonància Magnètica Nuclear)

L'espectrometria en Ressonància Magnètica Nuclear és una tècnica àmpliament utilitzada per estudiar l'estructura i dinàmica de les molècules i compostos orgànics. En metabolòmica, és útil per analitzar fluids i detectar la presència dels metabòlits presents en biofluids com el sèrum, el plasma o l'orina. Ha esdevingut una eina molt potent en aquest camp, així com en moltes altres disciplines com la química i la bioquímica, des de mitjans del segle passat.



Figura 3. Espectròmetre NMR. (Font: www.bruker.com)

El procés d'obtenció d'espectres RMN consisteix a aplicar a les mostres un camp magnètic continu d'alta magnitud i sotmetre-les a una seqüència de polsos de radiofreqüència variable (irradiació). Els nuclis atòmics absorbeixen l'energia dels polsos i això modifica l'orientació dels seus *spins*. En acabar la durada del pols, els spins es relaxen fins adquirir novament la posició inicial. Després, els nuclis emeten energia en forma de senyals de radiofreqüència detectables, generant un senyal FID (Free Induction Decay).

La FID obtinguda és un senyal en el domini temporal que es transforma a l'espai de freqüència mitjançant la transformada de Fourier, obtenint com a resultat un espectre de RMN que mostra les freqüències i intensitats de les senyals que equivalen a les concentracions relatives de les espècies presents en la mostra. Aquestes freqüències es representen al llarg d'un eix que anomenem *chemical shift* o desplaçament químic [5].

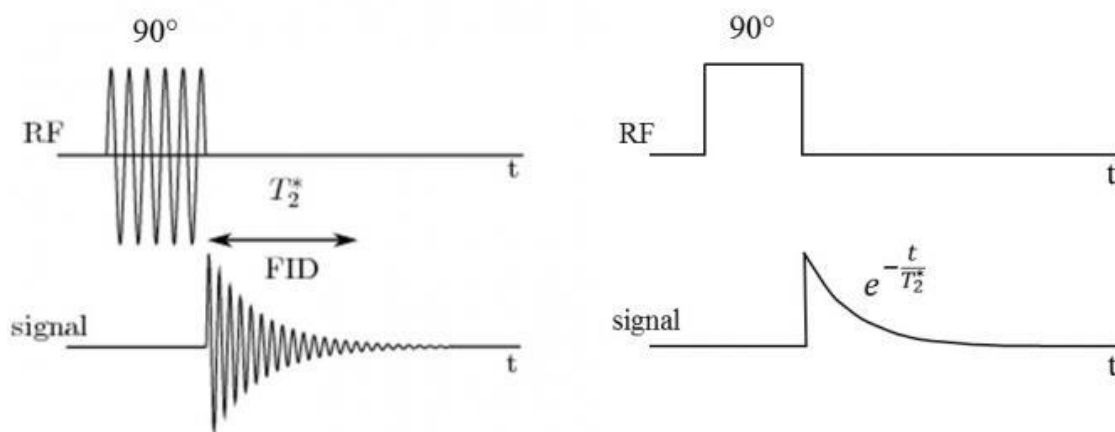


Figura 4. Senyals RF i FID (Font: <http://oftankonyv.reak.bme.hu/>)

Cada molècula vindrà caracteritzada per un nombre de pics associat als protons, mentre que en alguns punts de l'espectre freqüencial podem tenir una superposició de pics provinent de diferents molècules.

Gràcies a l'espectre obtingut es poden detectar i quantificar una àmplia varietat de metabòlits en una sola mostra, la qual cosa permet la identificació de biomarcadors específics de malalties, com ara la diabetis, l'obesitat, la malaltia cardiovascular, la malaltia renal, el càncer i les malalties neurodegeneratives.

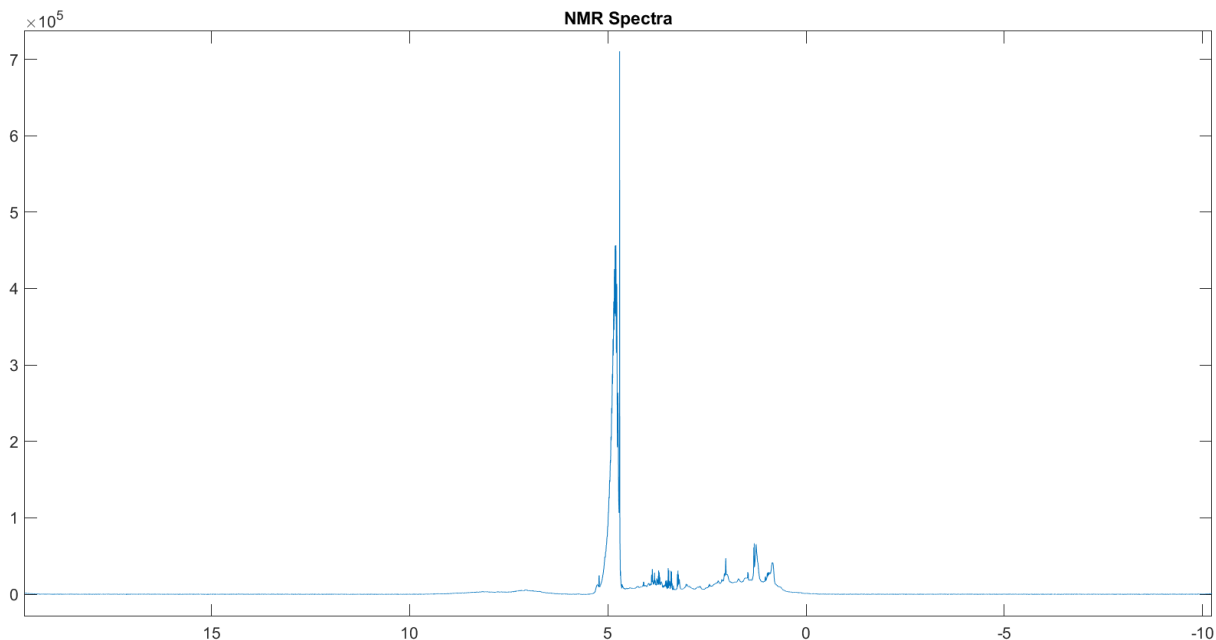


Figura 5. Exemple d'espectre RMN.

Una altra variable d'adquisició que afecta, en aquest cas, a la intensitat de l'espectre RMN és el nombre de *scans*, que és el nombre d'exploracions programades en l'adquisició de l'espectre. La quantitat de senyal s'acumula després de cada *scan*, de manera que això es tradueix sovint en una millora en la relació senyal/soroll de l'espectre, obtenint una millor resolució i sensibilitat.

Ara bé: per contra, es presenta l'inconvenient del temps d'adquisició, que també s'eleva considerablement a mesura que s'augmenta el nombre de *scans*: si dupliquem el nombre de *scans* també dupliquem aquest temps d'adquisició i, en canvi, la intensitat del senyal augmenta només en un factor $\sqrt{2}$, de tal manera que sovint és un factor decisiu a l'hora de decantar-se per un nombre més baix de *scans* en el procés d'adquisició de la mostra [6].

2.3 Compostos de baix pes molecular

Els compostos de baix pes molecular són molècules que es troben en els fluids biològics, com la sang, la saliva o l'orina, i que són identificades i quantificades per les tècniques analítiques de la metabolòmica. Aquests compostos inclouen molècules com aminoàcids, àcids orgànics, sucres, lípids i altres metabòlits de baix pes molecular.

Els compostos de baix pes molecular són d'interès en la metabolòmica perquè proporcionen informació sobre els processos metabòlics que es produeixen en l'organisme. Les alteracions en la concentració de metabòlits de baix pes molecular en una mostra biològica poden reflectir canvis en les vies metabòliques o en la resposta a una determinada malaltia o a un tractament.

Per exemple, la identificació i quantificació de metabòlits de baix pes molecular en la sang o l'orina pot proporcionar informació útil per al diagnòstic i el seguiment de malalties com la diabetis, malalties renals o malalties hepàtiques. També es poden utilitzar en la identificació de biomarcadors per a la detecció precoç de malalties, la predicció de la resposta al tractament i la monitorització de la progressió de la malaltia.

2.4 Polsos de RMN

En la RMN, es poden aplicar diferents polsos per modular el camp magnètic del nucli d'àtoms que es volen estudiar. Aquests polsos tenen diferents propietats que els diferencien entre ells i que els fan útils per a diferents aplicacions.

Els polsos de RF es caracteritzen per la seva freqüència, la seva durada i la seva intensitat. La freqüència del pols de RF ha de coincidir amb la freqüència de precessió dels nuclis que es volen excitar. La durada del pols de RF determina la quantitat d'energia que s'aplica als nuclis, mentre que la intensitat del pols de RF determina la magnitud de l'angle de gir dels nuclis excitats.

A continuació tenim exemples d'alguns d'aquests polsos, com el NOESY i el CPMG [7], que han estat els dos tipus d'espectres amb els quals s'ha treballat en aquest projecte.

2.4.1 Noesy

NOESY és l'acrònim de Nuclear Overhauser Enhancement Spectroscopy. Aquest experiment es nodreix de l'efecte Nuclear Overhauser Enhancement (NOE), que es produeix quan els nuclis de diferents àtoms estan pròxims entre ells. El NOESY mesura les interaccions dipolars entre nuclis que estan a prop, i això permet detectar la presència de connexions a través de l'espai entre diferents nuclis de la molècula. Aquest experiment és especialment útil per a l'assignació de senyals de RMN a les diferents parts de la molècula, i per a la determinació de la conformació molecular [8]

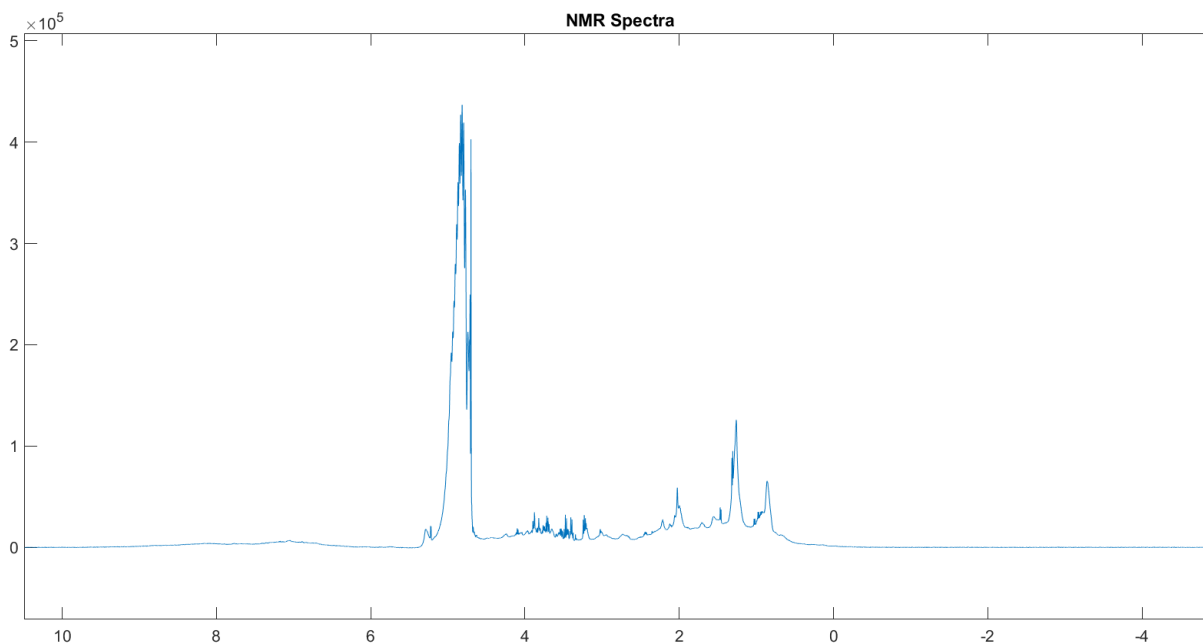


Figura 6. Espectre NOESY.

2.4.2 CPMG

CPMG és l'acrònim de Carr-Purcell-Meiboom-Gill. Aquest experiment és un tipus d'espectroscòpia d'eco de spin, que implica la generació d'un tren d'ecos de spin per mesurar la relaxació longitudinal i transversal de les espècies químiques en la mostra. En l'experiment CPMG, el tren d'ecos de spin es genera mitjançant una seqüència d'impulsos de radiofreqüència i gradients de camp magnètic, i la seva forma i intensitat són mesurades en funció del temps. Aquest experiment és útil per determinar la magnitud de les interaccions dipolars a través de l'espai i per mesurar els temps de relaxació T2 de les espècies químiques, que proporcionen informació sobre la dinàmica molecular [9]

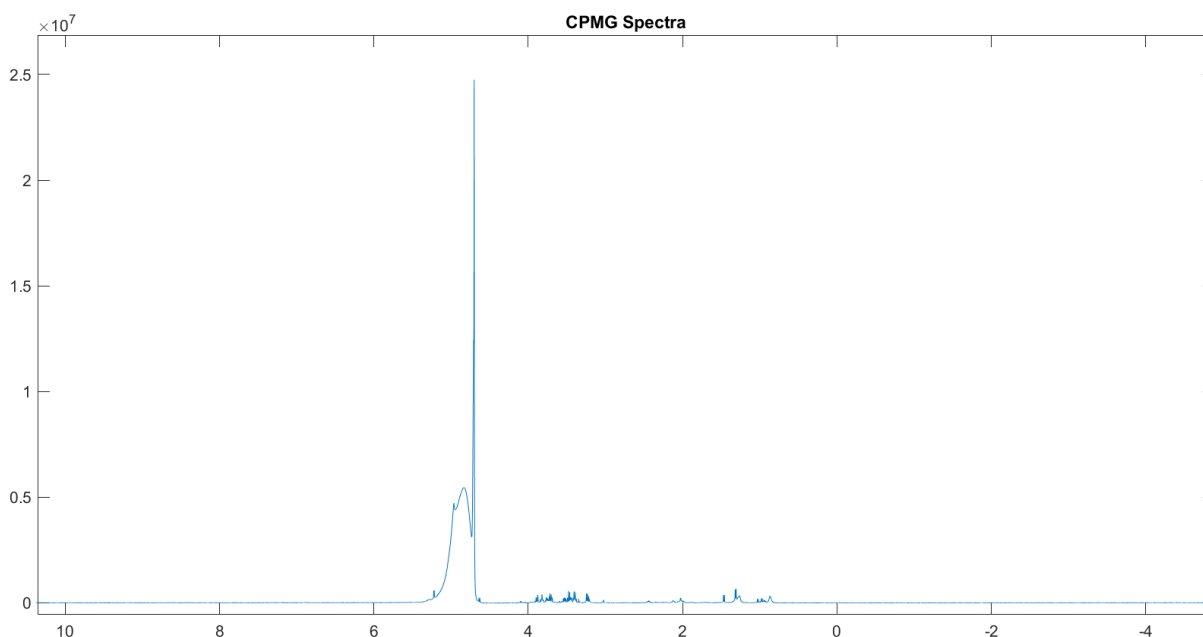


Figura 7. Espectre CPMG.

2.4.3 Altres polsos

A més dels dos polsos anomenats, n'existeixen molts altres: COSY, JRES, HMQC..., cadascun dels quals presenta les seves propietats i s'utilitza per a una aplicació concreta.

Cada pols té una freqüència, una durada i una intensitat específiques que es trien en funció dels nuclis que es volen estudiar i de l'objectiu de l'estudi.

Aquests altres polsos no han estat objecte d'estudi en aquest projecte, però cal fer menció al LED (Longitudinal Eddy-current delay) degut a la seva importància destacada en la detecció, precisament, de lípids i lipoproteïnes.

La seqüència LED es basa en la transferència de magnetització de nuclis d'hidrogen a nuclis de carboni mitjançant l'ús de polsos d'irradiació selectiva. Això permet obtenir espectres de RMN amb senyal elevat per a nuclis de carboni de baixa abundància en mostres biològiques, cosa que pot ser difícil amb altres seqüències de RMN convencionals.

2.5 SMoIESY

"SMoIESY" és una abreviatura de "Small Molecule Enhancement Spectroscopy". Es tracta d'una tècnica derivada dels espectres NOESY en la qual es mesura la derivada de la part imaginària de l'espectre NOESY a baixes freqüències.

La tècnica SMoIESY s'ha desenvolupat per resoldre un problema que es produeix en l'anàlisi de l'espectre Noesy: la superposició de pics amples corresponents a la ressonància de macromolècules que dificulten la identificació de les molècules més petites.

Aquest problema es pot millorar eliminant les macromolècules físicament, per exemple, mitjançant la ultracentrifugació, però això pot alterar la integritat de la mostra inicial, tot

perdent un dels avantatges que representa la tècnica NMR, que és que no precisa el processament de la mostra, a més de suposar un cost inassumible tant a nivell econòmic com de temps.

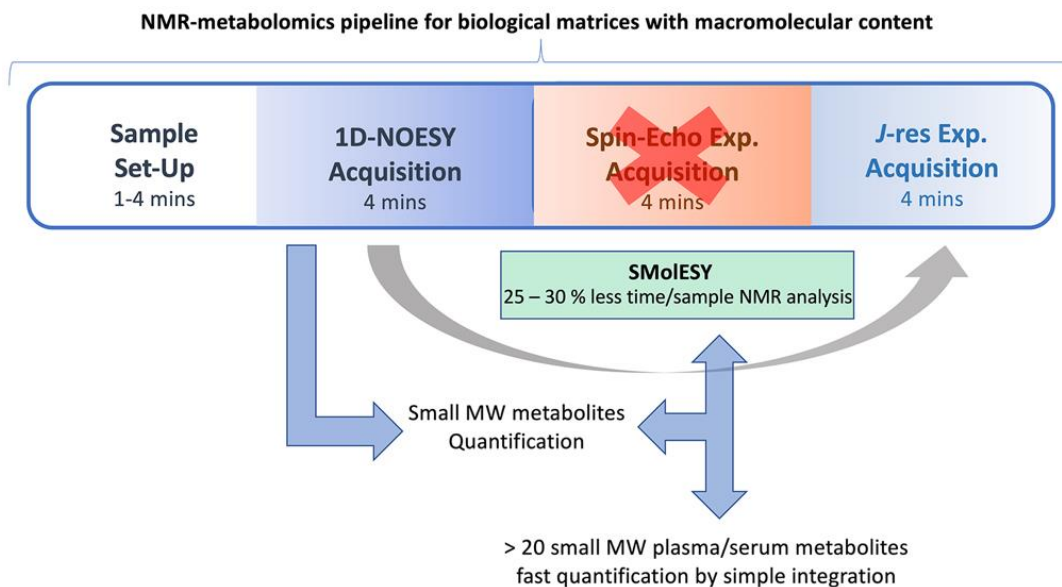


Figura 8. Utilitat de l'alternativa SMoIESY en el procés d'anàlisi NMR. (Font: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8159292/>)

També hi ha l'opció d'utilitzar el pols CPMG amb filtres per aquestes macromolècules, però aquest és també un procés altament costós pel que fa a temps i poc precís a l'hora de fer-ne la quantificació degut a l'alta variabilitat del temps de relaxació T2 de les molècules petites.

És per aquest motiu que s'obre l'alternativa del SMoIESY com una tècnica molt més eficient per aquest objectiu i que permet incrementar considerablement la resolució i filtrar la contribució de les molècules més grans directament des de l'espectre [10].

Aquesta aproximació es realitza calculant la primera derivada parcial de la part imaginària de l'espectre Noesy (això fa que no depengui de T2). De forma experimental, l'estudi esmentat demostra un decreixement mitjà d'un 30% de la relació senyal-soroll i permet detectar un 34% més de senyals respecte el CPMG.

El fonament matemàtic d'aquesta derivada és el següent: entenent la derivada parcial com el pendent de la recta tangent de l'espectre en cada punt, es demostra que la màxima intensitat de senyal de l'espectre derivat és inversament proporcional al seu ample a la meitat de l'altura. Modelitzant un pic com una funció de tipus Lorentziana, tindriem el següent:

$$f(x) = \frac{I_{\delta}}{1 + \left(\frac{x - \delta}{\Delta v_{1/2}}\right)^2} \tag{1}$$

On:

I_δ és la màxima intensitat de senyal a un punt específic de l'eix del desplaçament químic (δ).

$\Delta\nu_{1/2}$ és l'ample de pic a mitja altura.

La derivada és la següent:

$$f'(x) = \frac{-2I_\delta \cdot (x - \delta)}{\Delta\nu_{1/2} \cdot \left(1 + \frac{(x - \delta)^2}{\Delta\nu_{1/2}^2}\right)^2} \quad (2)$$

D'aquesta manera, els senyals amples ($\Delta\nu_{1/2}$ elevat) es veuen suprimitats, mentre que els senyals més estrets encara s'hi fan més.

Aquest mètode és útil perquè pot millorar la resolució d'un senyal i els senyals més amples de les macromolècules poden ser completament atenuats, tenint en compte, a més, que l'amplitud del senyal resultant continua sent directament proporcional a l'original ja que la diferenciació és una tècnica lineal. Això facilita molt la quantificació de les molècules més petites.

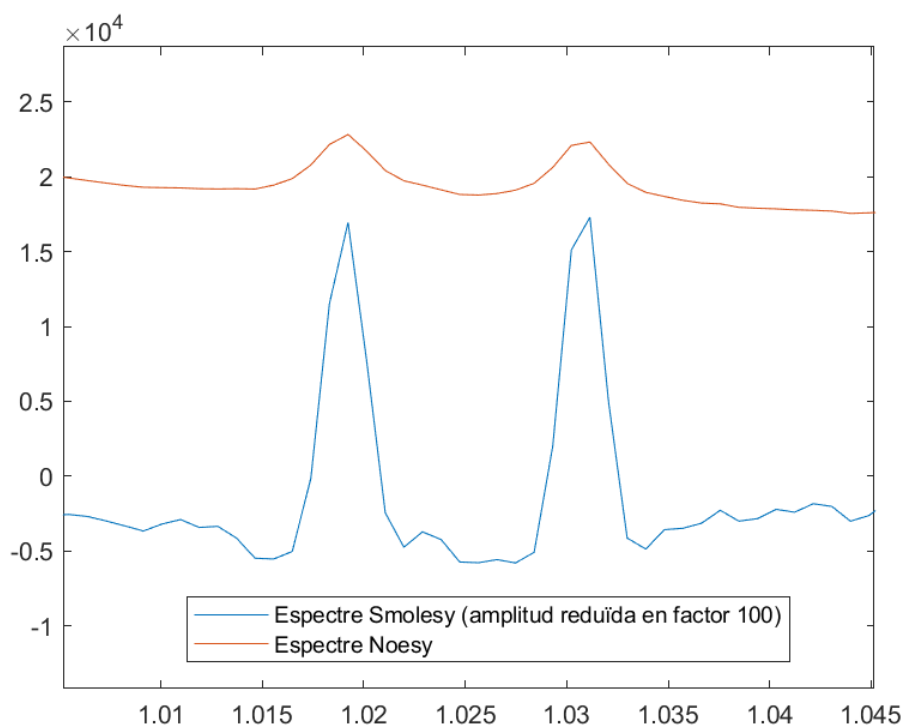


Figura 9. Comparació de les amplituds de pics en espectres Smoiesy (atenuat) respecte el Noesy corresponent a la mateixa mostra. (Font: elaboració pròpia)

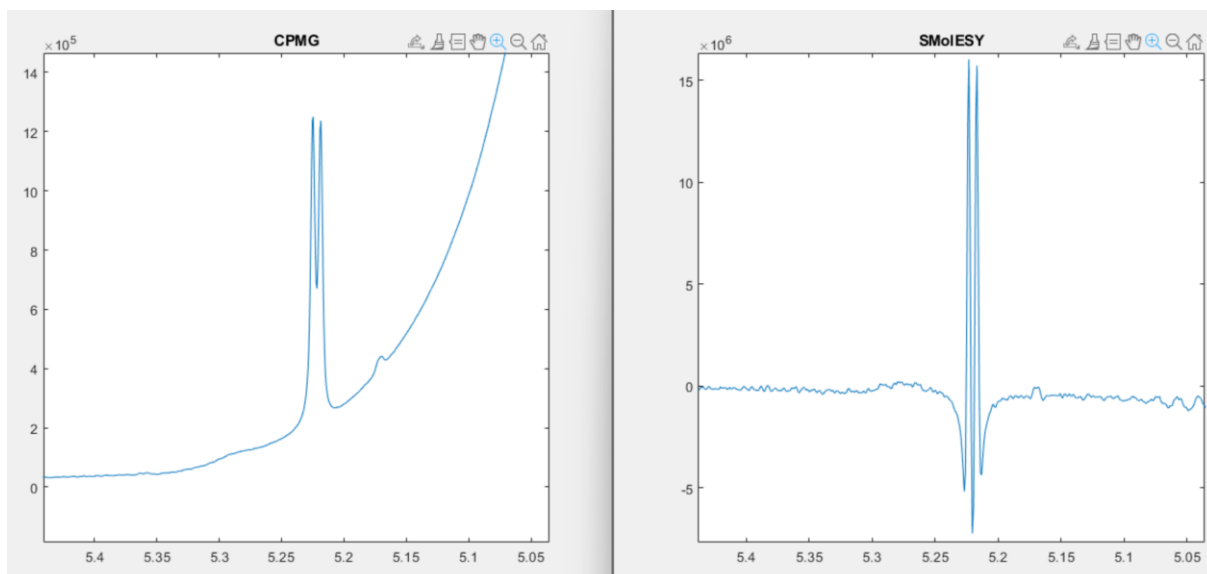


Figura 10. Comparació del pic de la glucosa en espectre Noesy respecte SMoIESY (Font: elaboració pròpia).

2.6 SMoIESY Platform

"SMoIESY_platform" és una eina quimiinformàtica per a produir i processar dades SMoIESY a partir d'espectres 1H-NMR. És una eina de lliure descàrrega disponible a *GitHub* [11].

Es tracta d'una interfície gràfica integrada com a *toolbox* de *MATLAB* que permet la transformació dels espectres, representació, calibració, alineació de pics per a la quantificació, quantificació absoluta dels pics i exportació de les quantificacions.

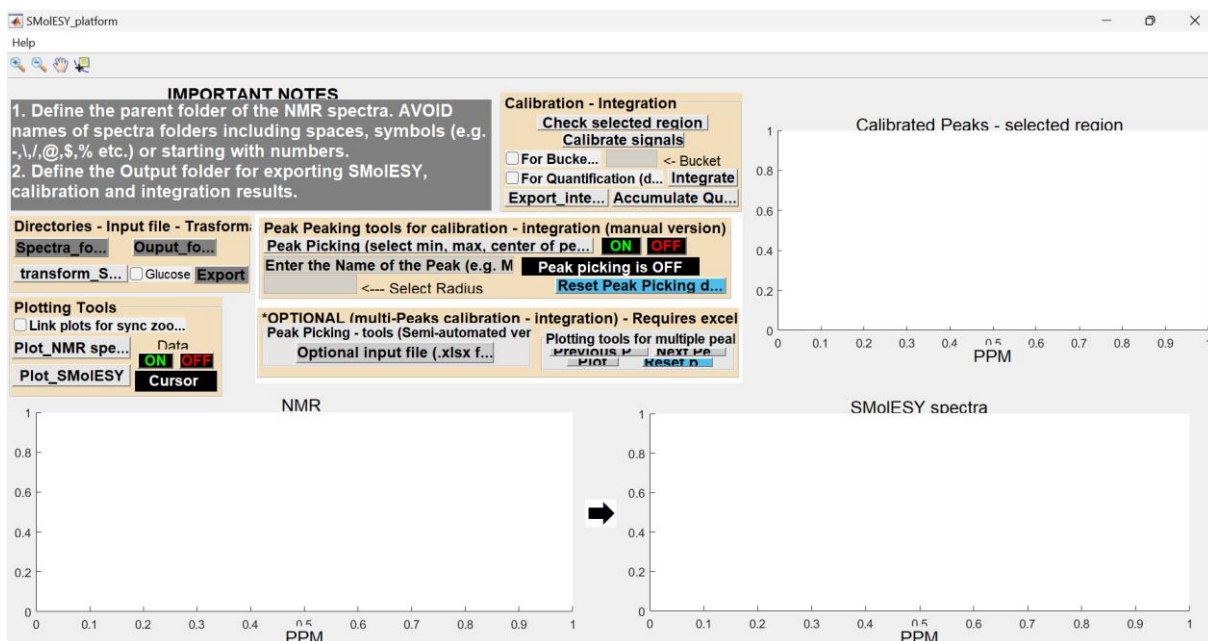


Figura 11. Interfície gràfica de la toolbox "SMoIESY Platform".

Aquesta *toolbox* és molt completa donat que permet carregar diversos espectres Noesy al mateix temps per a transformar-los, tot i que el format de sortida és un arxiu de text que conté tots els punts de l'espectre *SMoIESY* i, per tant, requereix un post-processament per obtenir els mateixos de forma independent.

Tot seguit, la plataforma permet fer una selecció de pics (peak Picking) per a calibrar (alinear el seu màxim a un punt de l'eix del desplaçament químic prèviament seleccionat per l'usuari) i integrar els mateixos. Aquesta selecció de pics es pot realitzar de forma manual clicant a la gràfica sobre els extrems del pic que es vol quantificar i sobre el seu màxim, creant així una finestra (*bucket*) d'integració, o bé es pot automatitzar generant un full de càlcul on s'introdueix el nom del metabòlit corresponent a cada pic i els tres valors necessaris per a crear aquesta finestra i integrar el pic.

L'opció manual és poc precisa i no és útil per a realitzar les quantificacions a gran escala previstes per al treball. L'opció automàtica pot ser més útil, però es considera interessant, com a objectiu final del present projecte, adaptar tot el procés de transformació, calibració, quantificació i correlació al format de dades del qual es disposa i integrar-ho tot en un aplicatiu únic i independent de la plataforma i de les limitacions o problemes que aquesta pugui presentar.

D'altra banda, la quantificació absoluta (integració directa) proposada per la plataforma genera dubtes pel que fa al tractament dels possibles valors negatius de les àrees. És per això que més endavant s'estudia la manera més acurada d'integrar els pics per obtenir unes millors correlacions amb les quantificacions originals del CPMG.

3. Mètodes i procediments

3.1 L'entorn de programació: MATLAB

El programa MATLAB és un software d'ús molt estès en el camp de la ciència, les matemàtiques i l'enginyeria i, en particular, de la metabolòmica, donada la seva capacitat i facilitat per analitzar i organitzar dades de tipus matricial de forma molt eficient i visual. També és un programa que permet presentar els resultats gràfics de manera personalitzada i fàcilment editable.

Pel que fa al llenguatge de programació, es tracta d'un llenguatge propi d'alt nivell que presenta una sintaxi molt semblant al C, però que permet dissenyar algorismes de forma més senzilla. Presenta funcions integrades que faciliten càlculs matemàtics i d'enginyeria d'ús freqüent.

Permet realitzar tant operacions ràpides mitjançant la línia de comandes com desenvolupar *scripts* per realitzar aplicacions a gran escala.

En el camp de la metabolòmica existeixen també funcions predissenyades d'ús públic, com és el cas de la funció *RBNMR* que s'ha utilitzat en aquest projecte, que permet llegir i tractar de forma senzilla dades espectrals de RMN al *Workspace* de MATLAB. Aquesta funció importa les dades en un format d'estructura a partir dels fitxers obtinguts del procés [12].

3.2 Estadística: coeficient de correlació de Pearson

Per tal de relacionar estadísticament les quantificacions obtingudes experimentalment amb les que es disposa a partir dels espectres CPMG de les mateixes mostres, s'ha utilitzat el coeficient de correlació de Pearson, que és una mesura de dependència lineal entre dues variables quantitatives contínues [13] [14].

Per a dades mostrals, es calcula mitjançant la següent fórmula:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

On:

n és la mida de la mostra

x_i i y_i són dades mostrals individuals

\bar{x} i \bar{y} són les mitjanes de les mostres

El programa MATLAB ja disposa d'una funció *fitlm* que retorna un model de regressió lineal de dues variables disposades en format taula. Aquesta funció mostra per pantalla la fórmula del model, els coeficients estimats i un resum estadístic del model (desviacions estàndard, t-

statistic i p-value, així com nombre d'observacions, grau de llibertat, error quadrat mitjà, coeficient de determinació R^2 , etc.) [15].

El coeficient de determinació R^2 , per a models lineals, és simplement el quadrat del coeficient de correlació de Pearson.

3.3 Conjunt de mostres utilitzat: Carratalà

S'ha disposat dels espectres i les dades d'un conjunt de 151 mostres d'individus proporcionat per *Biosfer Teslab*. D'aquests individus es disposa de:

- Els espectres NOESY
- Els espectres CPMG
- Els arxius de l'espectre LED
- Les quantificacions dels diferents metabòlits realitzades en base a l'espectre CPMG, en format full de càlcul.

Els metabòlits que conté aquest arxiu, altrament dit, que són quantificats a *Biosfer Teslab*, són els del llistat següent:

- 3-hidroxiabutirat
- Acetat
- Acetona
- Alanina
- Creatinina
- Creatina
- Glucosa
- Glutamat
- Glutamina
- Glicerol
- Glicina
- Histidina
- Lactat
- Mannitol
- Treonina
- Tirosina
- Valina
- Isoleucina
- Leucina

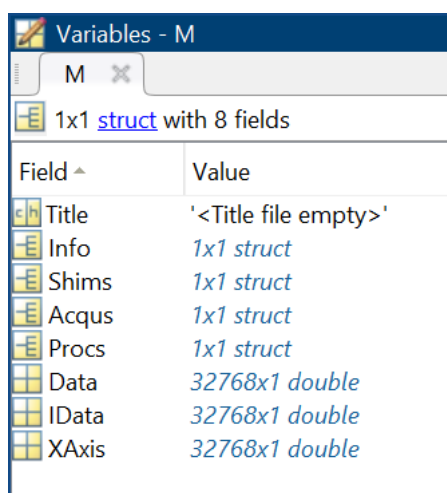
3.4 Preprocessat de les dades

3.4.1 Importació de les dades a format apte per a MATLAB

Aquesta ha estat la primera fase purament pràctica del projecte. La base de dades (veure 3.3.) està dins d'una carpeta anomenada *nmr* que conté els 151 espectres en una estructura de carpetes anomenades *ID_COVID_CARRATALA_n*, on *n* és el nombre identificatiu de la mostra comprès entre el 462 i el 612.

Cadascuna d'aquestes carpetes conté 4 subcarpetes, de les quals ens interessen la 11 i la 13, ja que contenen els arxius de l'espectre CPMG i Noesy, respectivament. Les altres corresponen al pols LED i a arxius de comprovació.

El primer pas per obtenir les dades en un format matricial apte per a MATLAB consisteix a crear dos scripts d'un sol ús: un per a obtenir els espectres Noesy que posteriorment es transformaran, i un altre per a obtenir els CPMG per a poder-los representar i comparar amb el *SMoIESY*. Aquest codi cridarà la funció *RBNMR* esmentada a l'apartat 3.1. per a transformar les dades i guardar-les en una estructura *M*.



| Field ^ | Value |
|---------|----------------------|
| Title | '<Title file empty>' |
| Info | 1x1 struct |
| Shims | 1x1 struct |
| Acqus | 1x1 struct |
| Procs | 1x1 struct |
| Data | 32768x1 double |
| IData | 32768x1 double |
| XAxis | 32768x1 double |

Figura 12. Contingut de l'estructura dels espectres de la base de dades

S'observa que aquests espectres tenen una resolució de 32.768 punts.

En aquest algoritme s'utilitzen les funcions de MATLAB:

int2str, que converteix un enter en un string per poder interpretar-lo com a text;

strcat, que permet concatenar els camins on s'allotjaran les matriu;

save, que guardarà l'estructura *M* a la ruta especificada amb el nom d'arxiu corresponent al nombre identificatiu entre el 462 i el 612. Aquestes variables s'allotjaran a la carpeta especificada (*cpmg* i *noesy*, respectivament).

També s'hi inclou una *waitbar* per informar a l'usuari del progrés.

El codi generat es pot consultar a l'Annex 1.

Finalment, les dades d'aquestes estructures es redimensionen en format matriu per tal d'optimitzar el temps d'accés i facilitar-ne i agilitzar-ne l'ús.

3.4.2 Transformació dels espectres Noesy a SMoIESY

Inicialment es decideix realitzar aquesta transformació mitjançant la plataforma, tot i que amb la vista posada en fer-la de forma independent en un futur.

En el primer intent es creu erròniament que la càrrega, transformació i exportació dels espectres està limitada a una sola mostra, de forma que es comença a fer el procés de forma manual.

```

PPM-20      19.6996      19.6993709748269  19.6991419496538  19.6989129244806  19.6986838993075  19.6984548741344
19.6982258489613  19.6979968237882  19.6977677986151  19.6975387734419  19.6973097482688  19.6970807230957
19.6968516979226  19.6966226727495  19.6963936475763  19.6961646224032  19.6959355972301  19.695706572057
19.6954775468839  19.6952485217108  19.6950194965376  19.6947904713645  19.6945614461914  19.6943324210183
19.6941033958452  19.693874370672  19.6936453454989  19.6934163203258  19.6931872951527  19.6929582699796
19.6927292448065  19.6925002196333  19.6922711944602  19.6920421692871  19.691813144114  19.6915841189409
19.6913550937677  19.6911260685946  19.6908970434215  19.6906680182484  19.6904389930753  19.6902099679022
19.689980942729  19.6897519175559  19.6895228923828  19.6892938672097  19.6890648420366  19.6888358168634

```

Figura 13. Contingut de l'arxiu de text que s'obté en transformar un espectre Noesy a SMoIESY mitjançant la plataforma.

Com s'observa a la figura anterior, l'output resultant és un fitxer en format *.txt* que presenta a l'inici un codi identificatiu que s'ha d'eliminar (veure algoritme a l'Annex 2).

Cal destacar l'ús de la funció *readmatrix* que permet llegir un arxiu de text i que el programa l'interpreti com a matriu.

Posteriorment, s'estudia fer-ho amb diversos espectres i s'observa que, reorganitzant l'estructura de carpetes, es poden transformar diverses mostres al mateix temps, amb la qual cosa l'algoritme anterior queda obsolet.

Per a fer-ho, cal eliminar les subcarpetes de cada mostra fins a deixar-ne una de sola (la número 98888, la 12 i la 11 o 13, segons interressi). S'accedeix des de línia de comandes de Windows (cmd) aplicant el següent codi, seleccionant el nom de la carpeta que es desitja eliminar en cada cas [16]:

```

FOR /d /r . %d IN ("11") DO @IF EXIST "%d" rd /s /q "%d"
FOR /d /r . %d IN ("12") DO @IF EXIST "%d" rd /s /q "%d"
FOR /d /r . %d IN ("13") DO @IF EXIST "%d" rd /s /q "%d"
FOR /d /r . %d IN ("98888") DO @IF EXIST "%d" rd /s /q "%d"

```

Com s'explica a l'apartat 2.6, la plataforma retorna tots els espectres en un fitxer *.txt* únic que es separa mostra a mostra i es guarda en format matricial (consultar algoritme a l'Annex 3).

A l'apartat 4 (resultats) s'explica com, finalment, aquesta transformació s'inclou en l'algoritme definitiu, sense necessitat de la *SMoIESY Platform*.

3.4.3 Representació de les dades

A continuació, s'elaboren algoritmes que permeten representar les dades CPMG i SMoIESY, tant de forma separada com conjunta, i amb objectius diferents: veure l'alineació dels pics, comparar ambdós espectres per veure diferències en l'amplitud, representar un fitxer concret, etc.

Aquí veiem totes les representacions que el projecte ha anat requerint:

- Representar tots els espectres CPMG
- Representar tots els espectres SMoIESY
- Representar qualsevol espectre triant l'arxiu concret
- Representar un sol espectre SMoIESY
- Comparar conjunts CPMG i SMoIESY
- Comparar CPMG i SMoIESY mostra a mostra

Els algoritmes corresponents a aquestes representacions s'inclouen a l'Annex 4.

3.4.4 Interpolació: correcció del problema de resolució

En importar i representar els espectres anteriors, s'observa en els espectres SMoIESY el que es creu que és una pèrdua de resolució respecte l'equivalent CPMG degut a la transformació, malgrat que els espectres mostren el mateix nombre de punts: 32768.

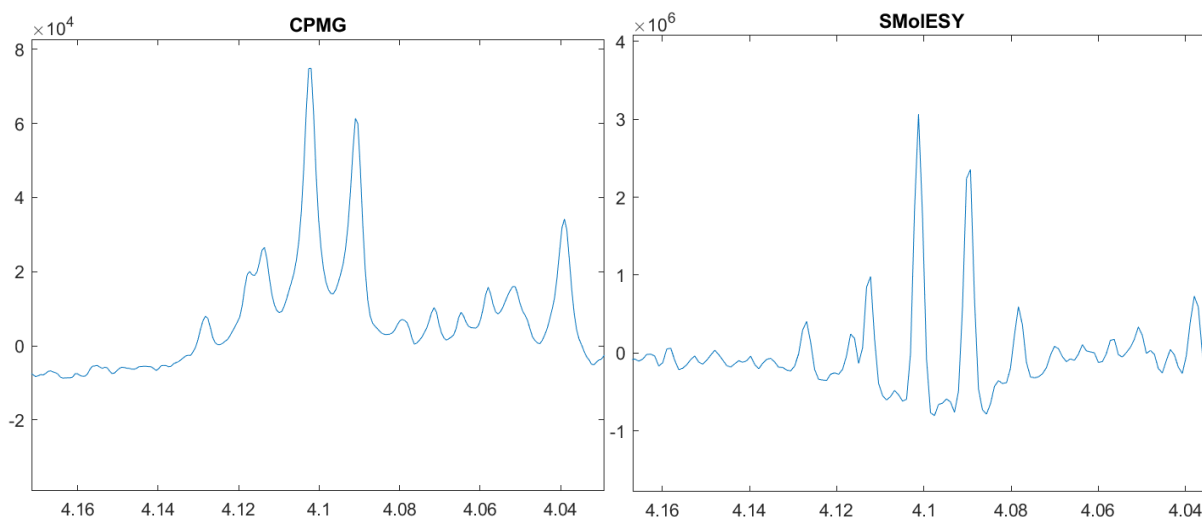


Figura 14. Comparació de l'espectre CPMG amb el seu corresponent SMOESY. El primer presenta una major resolució.

Posteriorment s'observa que no és una pèrdua en la resolució, sinó que, en obtenir les dades de l'espectròmetre, la finestra de dispersió (domini de l'eix del desplaçament químic) és diferent entre l'espectre CPMG, que va des de les 14 ppm fins a -5 ppm, i el Noesy original, que va de 20 ppm a -10 ppm.

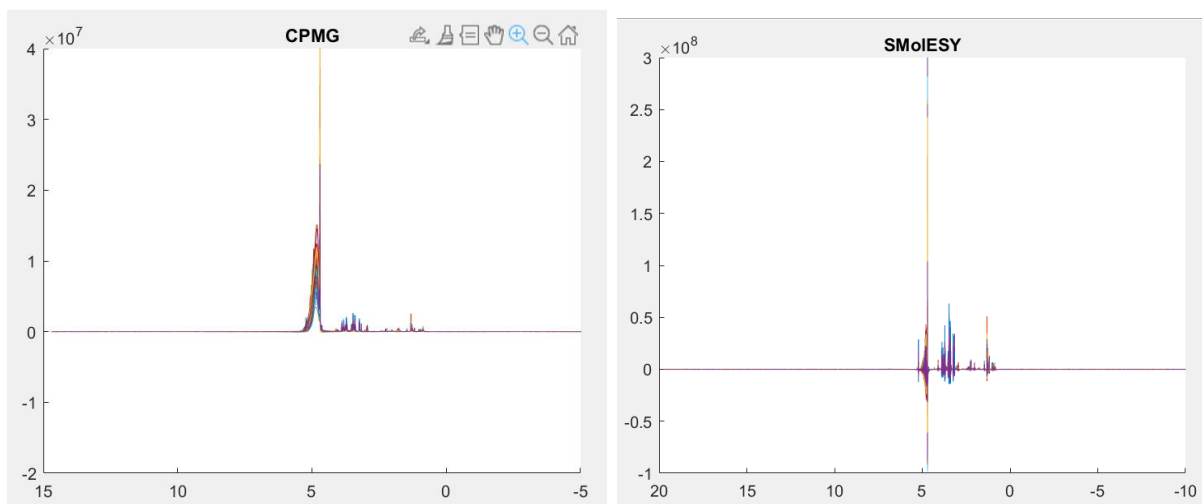


Figura 15. Comparació d'ambdós espectres complets. Els dominis varien.

Per resoldre aquesta qüestió, es decideix recórrer a la interpolació. La funció de MATLAB *interp1* permet interpolat funcions en una sola dimensió. La sintaxi és la següent:

$$interp1(x,v,xq,method)$$

On els paràmetres d'entrada són:

Un vector x que conté els punts de la mostra.

Un vector v que conté els respectius valors.

Un vector xq que conté les coordenades dels punts desitjats.

El paràmetre *method* que especifica el mètode d'interpolació.

El vector dels punts desitjats s'obté agafant com a referència el conjunt de punts d'uns espectres d'exemple existents en descarregar la *SMolESY Platform*. Són espectres que contenen 131072 punts (és a dir, multiplica per 4 la resolució original). Paral·lelament, també es pot agafar com a referència uns espectres d'una base de dades diferent proporcionats per l'empresa *Biosfer Teslab* que presenten aquesta mateixa resolució.

Pel que fa al darrer paràmetre, la interpolació que permet obtenir pics amb una morfologia similar als desitjats és la de tipus *spline*, on els valors interpolats en els nous punts es basen en una interpolació cúbica dels valors als punts veïns de cada dimensió respectiva.

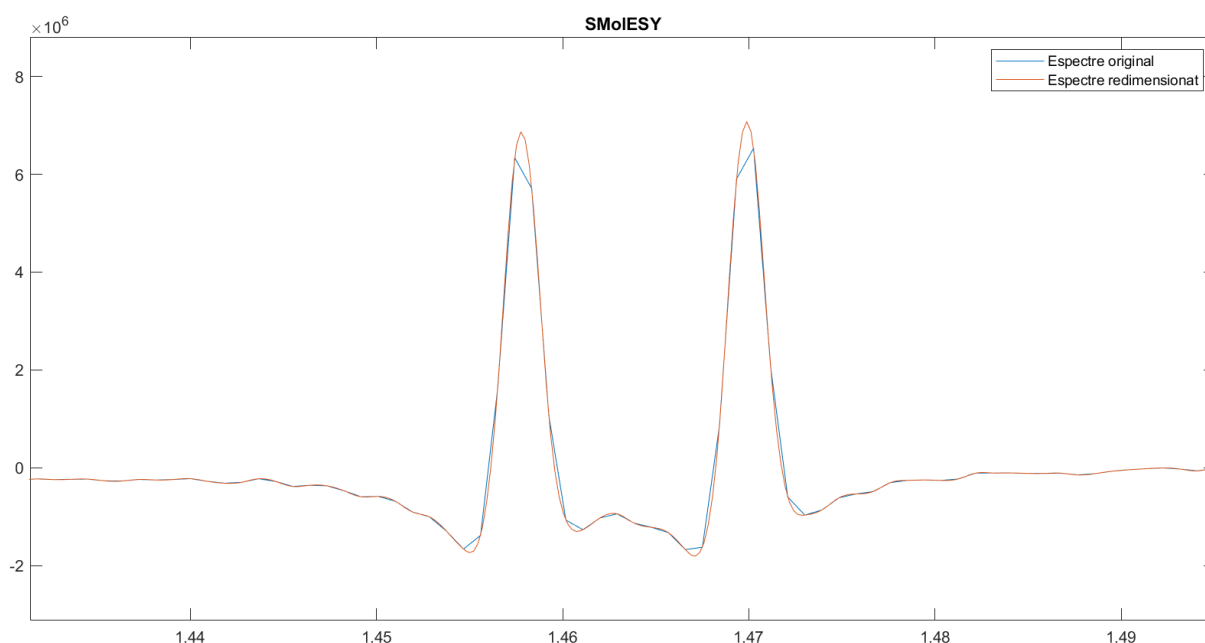


Figura 16. Comparació de l'espectre SMolESY original amb l'interpolat.

Comparant-ho amb els resultats parcials que s'havien realitzat fins al moment, les correlacions estadístiques obtingudes després d'interpolar aquests espectres milloren de forma considerable i, per tant, es valida aquest mètode i es treballa amb els espectres redimensionats a partir d'aquest punt.

A l'apartat 4.2.3 s'aprofundeix més en aquesta qüestió a nivell de codi.

4. Resultats

4.1 Mètodes de quantificació

Per tal de quantificar els metabòlits es pensa en diverses opcions: en primer lloc basant-se en l'amplitud dels pics i en segon lloc, calculant l'àrea d'aquests. Un altre repte que apareix de forma simultània és la manera referenciar els pics i/o els espectres per tal d'unificar els càlculs, ja que la línia base varia en les diferents mostres.

Finalment, cal tenir en compte que hi ha diversos tipus de metabòlits que es volen quantificar, classificats segons el nombre de pics que s'observen a l'espectre:

- Singlets (pic únic)
- Doblets (dos pics)
- Triplets (tres pics)
- Quadruplets (quatre pics)
- Multiplets (altres formes: dobles dobles, dobles triplets, etc.)

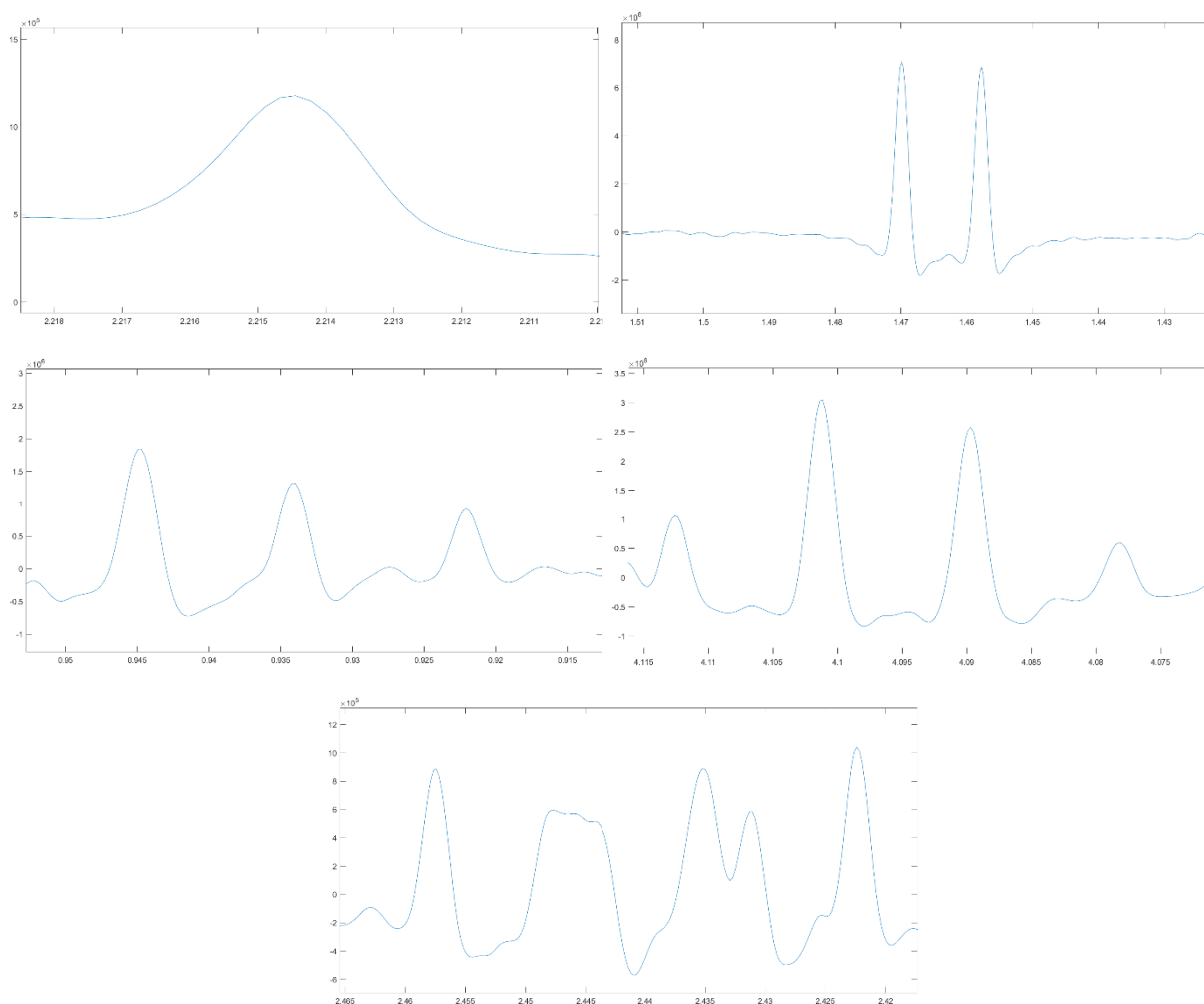


Figura 17. Imatge de diferents morfologies dels metabòlits (singlet, doblet, triplet, quadruplet i multiplet, respectivament).

Com a estratègia inicial, es decideix començar per la integració més senzilla, la dels singlets, per continuar més endavant amb la resta de tipus.

A continuació es detallen les diverses opcions per a la quantificació que s'han considerat i avaluat fins a la tria de la definitiva.

4.1.1 Amplitud

La primera opció, i més senzilla, consisteix a calcular l'amplitud de cada pic. Per a referenciar-los s'opta per calcular una mitjana entre els dos mínims de la funció dins d'una subfinestra determinada i rectificar aquesta mitjana per a fer-ho coincidir amb l'eix horitzontal:

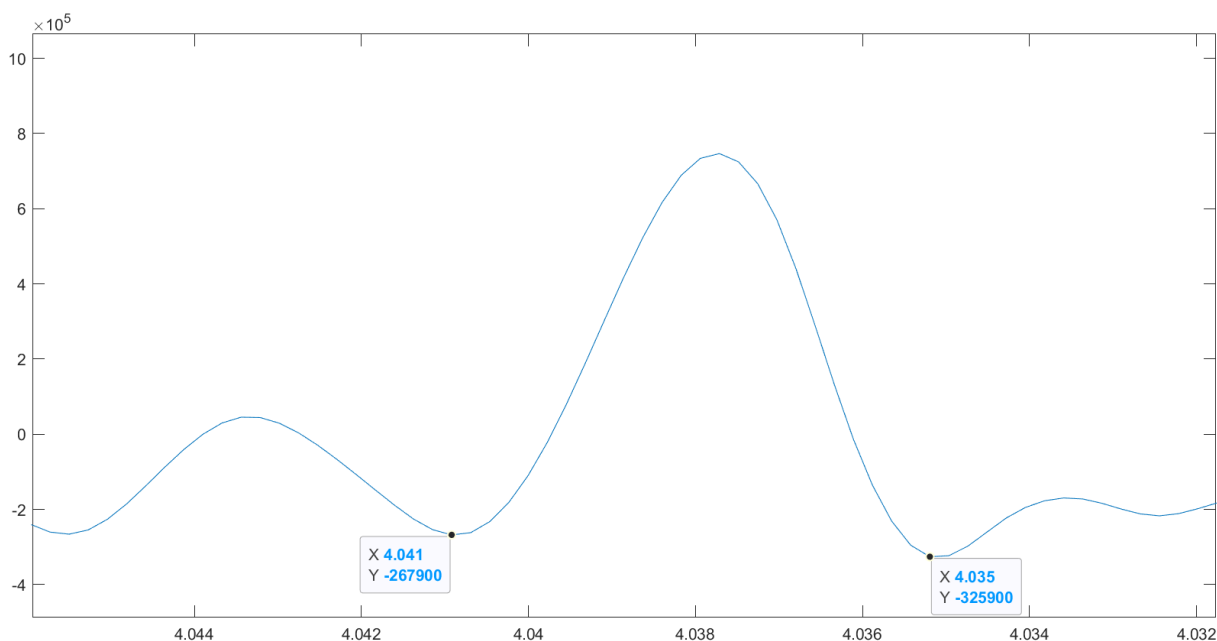


Figura 18. Pic de la creatinina. Quantificació per amplitud.

No obstant, optant per aquest model les correlacions que s'observen són molt pobres, amb valors de R^2 per sota de 0.8, per la qual cosa es descarta aquesta via.

Cal especificar que les correlacions es fa servir la funció *fitlm* de MATLAB tal com s'explica a l'apartat 3.2.

L'algoritme dissenyat ha de contemplar alguns altres problemes i necessitats:

Cal carregar primer les dades del full de càlcul on es troben les quantificacions del CPMG que es volen comparar. Per això, s'utilitzen les funcions de MATLAB *readtable*, que crea una taula a partir d'un fitxer en el format .xls o similar; i *table2array*, que converteix la taula anterior en un vector.

Cal fer una detecció prèvia de possibles dades en format no numèric (NaN). Per això s'empra la funció de MATLAB *isnan*. Aquestes dades s'han d'eliminar tant del vector creat a partir de la taula anterior com de les quantificacions calculades (vector *a*), per tal que els valors que es correlen corresponguin a les mateixes mostres.

Els detalls d'aquest algoritme es troben a l'Annex 5.

4.1.2 Àrea (integració)

A continuació es demana assessorament amb *Biosfer Teslab* per aquesta qüestió i es constata que les quantificacions són molt més fidels si es realitzen via integració dels pics enlloc de l'amplitud.

Per tal d'agilitzar-ne el càlcul, es consulta la metodologia de treball utilitzada més habitualment en aquests casos i es decideix treballar amb una opció equivalent a la que presenta el seu software de quantificació, que consisteix a agafar una finestra més àmplia dels pics, de manera que s'hi inclogui la línia base de l'espectre. S'agafen els extrems d'aquesta finestra en diferents graus (es fan proves amb un 5%, 10% i 20% de punts de la finestra) per a cada espectre i se'n fa una mitjana. El valor obtingut és el que s'utilitza per corregir la diferència respecte zero, unificant força les línies base.

Paral·lelament, també es comenta la necessitat d'alinejar els pics. En la següent figura es mostra la millora en els dos aspectes anteriors.

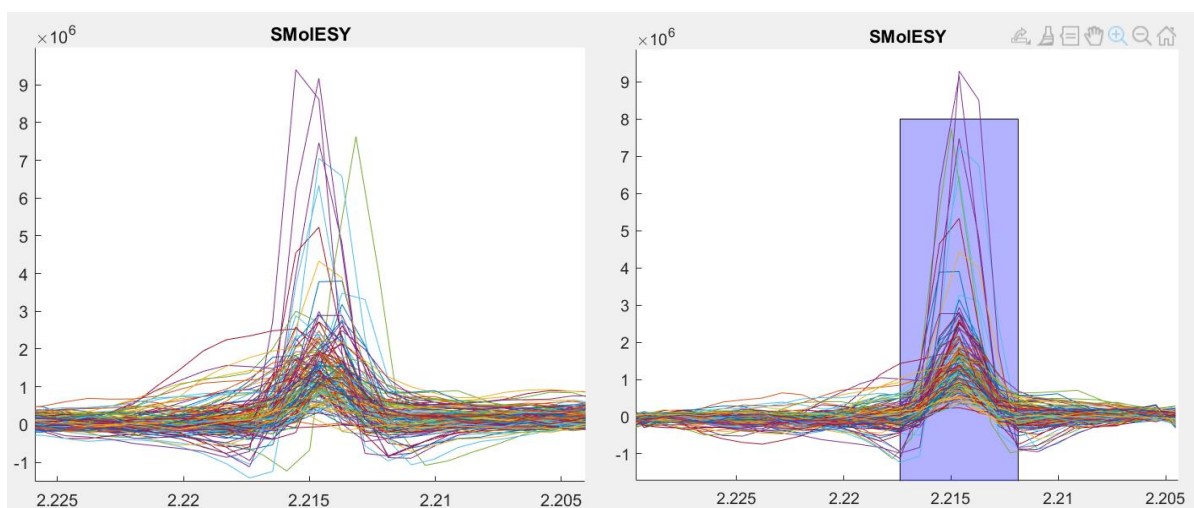


Figura 19. Comparació entre els espectres originals sense aplicar correcció (esquerra) i aplicant-hi la correcció de la línia base i l'alineació del pic (dreta).

Per a l'alineació de pics es calcula el punt més habitual on es troba el màxim, i es fa un desplaçament de tota aquesta finestra amb la funció *circshift*.

Finalment, s'integren els diferents pics utilitzant la funció de MATLAB *trapz(x,y)*. Aquesta funció realitza una integració trapezoidal numèrica de la funció *y* respecte les coordenades de finestra *x* especificades.

Cal puntualitzar que, a l'algoritme, s'inclou un canvi de signe als valors de les integrals ja que, en tenir els eixos del desplaçament químic orientats de forma descendent, aquestes surten amb valors negatius.

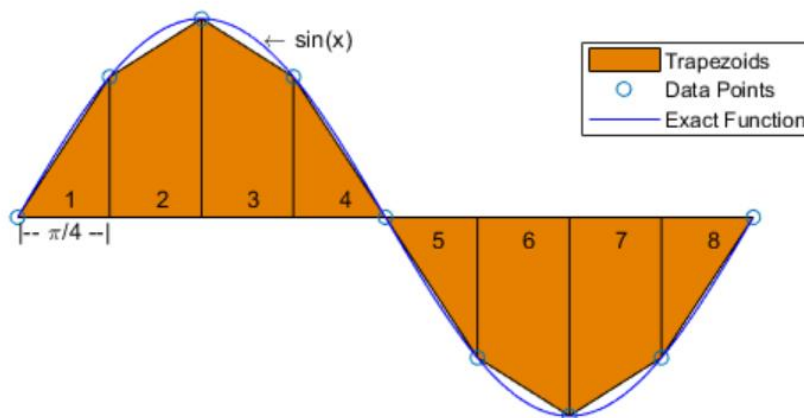


Figura 20. Integració trapezoidal (Font: MATLAB Help)

El codi emprat es pot consultar a l'annex 6

A les proves realitzades després d'aquestes modificacions s'observa una millora substancial en les correlacions, tot i que no definitives, donats que en alguns pics puntuals amb morfologies particulars aquesta correlació és, fins i tot, més baixa que quantificats amb l'amplitud.

Tot seguit es contempla una opció que consisteix a crear una recta base per a cada pic, traçada entre els dos punts mínims dins d'una petita finestra, tal com es mostra a la imatge:

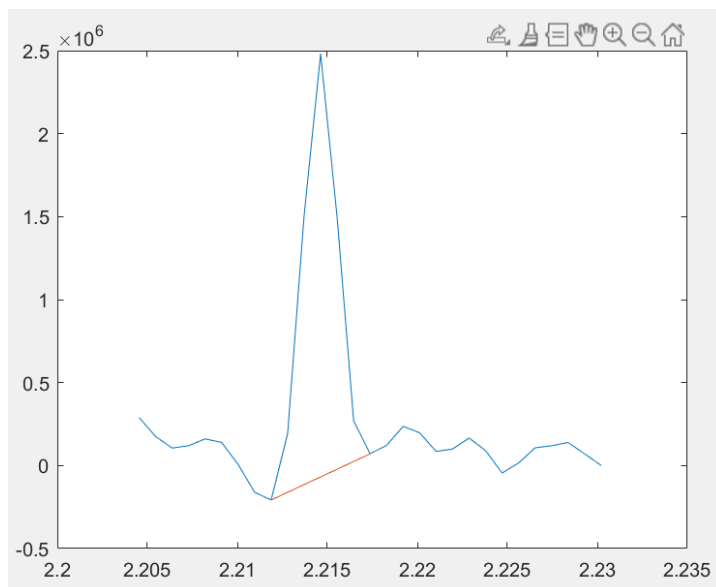


Figura 21. Integració via càlcul de la recta base.

En primer lloc s'han calculat els mínims d'una finestra prèviament estudiada i la seva abscissa. A continuació es registren les coordenades d'aquests punts a les variables p i q per, finalment, obtenir la fórmula de la recta amb la funció *polyfit*, que retorna els coeficients d'una funció de grau n a partir dels paràmetres anteriors.

Per últim, es calculen els punts d'aquesta recta amb la mateixa resolució de la finestra, de manera que quedin units els dos punts desitjats, i s'integra l'àrea compresa entre ambdues funcions tot restant els punts de l'espectre menys la recta calculada.

Es poden consultar els detalls del codi a l'Annex 7.

S'observa que, per pics molt estrets, les correlacions milloren considerablement respecte tots els casos anteriors. Ara bé, per pics més amples o amb possible superposició d'altres metabòlits, algunes d'aquestes correlacions encara són força baixes.

A continuació, es fa un estudi minuciós del codi font de la *SMoIESY Platform* i s'acaba determinant que la integració es fa de forma directa: es fixa una certa finestra de punts del pic en qüestió i se'n calcula la integral. No obstant, es desconeix els possibles pretractaments o referenciats que es poden haver fet en aquest cas, així com el tractament que es fa dels valors negatius, que a priori sumen el seu valor a la integral que es calcula, tal com s'observa a la imatge.

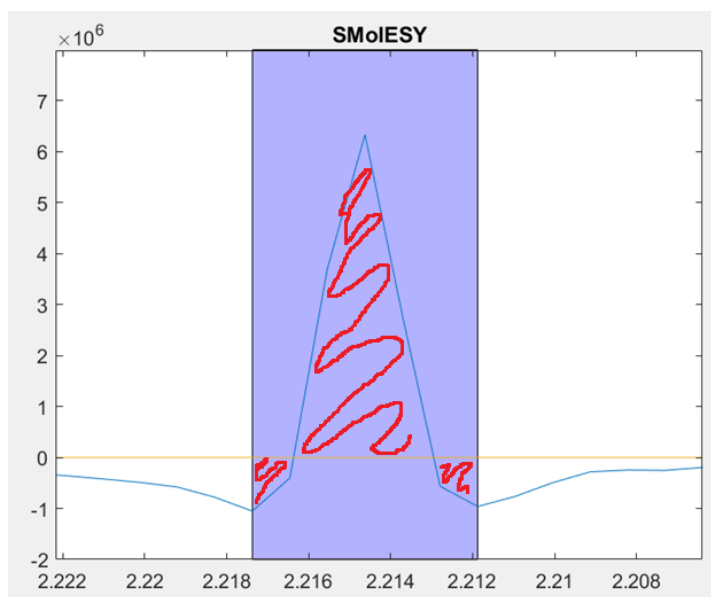


Figura 22. Càlcul de l'àrea per quantificació directa.

Aquest codi d'integració s'adapta directament del codi font de la plataforma, estudiant-ne el funcionament i modificant els noms de les variables corresponents. Per abastar totes les possibilitats, es quantifiquen les àrees amb les diverses opcions: modificant el codi per a que els trams d'àrea positiva i negativa es sumin o es restin, respectivament, i calculant un rectificat de línia base, però les correlacions no són, en cap cas, millors que les del mètode anterior.

Finalment, es troba la solució definitiva que consisteix a referenciar a zero el punt mínim d'una subfinestra més reduïda dins de la finestra on es troba ubicat el pic, de manera que l'àrea calculada sempre presenti un valor positiu.

A la figura següent es poden observar bé alguns dels casos: les figures blava i groga presenten el punt mínim a l'extrem dret de la subfinestra seleccionada, mentre que la figura de color vermell fixa aquest "zero" en un punt molt proper a l'extrem esquerre de la subfinestra.

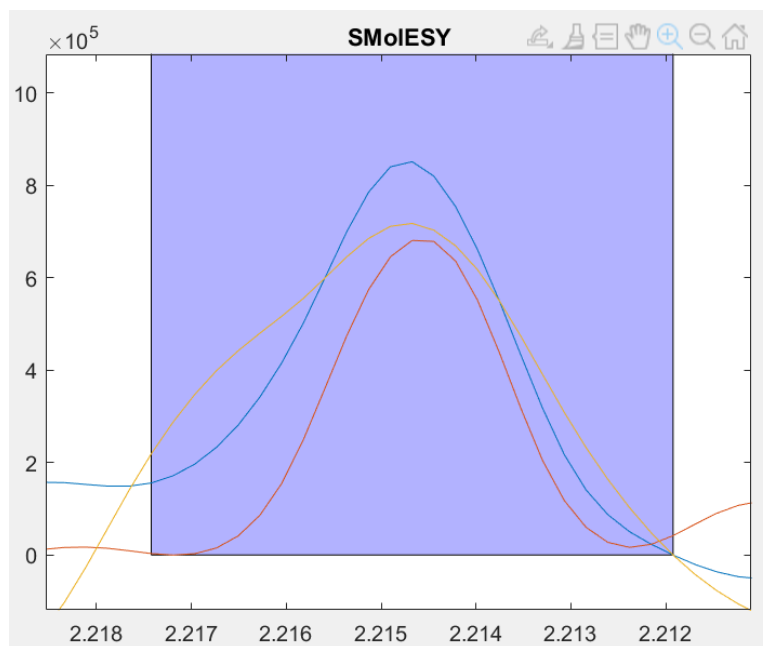


Figura 23. Quantificació de l'àrea mitjançant rectificació del mínim punt de la subfinestra.

El codi es troba a l'annex 9.

S'observen molt bones correlacions, en molts casos, amb valors de R^2 al voltant de 0.9 i, en qualsevol dels casos, superiors notablement a les altres estratègies estudiades. De fet, es considera que, gràcies a aquest pretractament, les quantificacions milloren respecte les de la SM Platform.

4.2 Programació de la quantificació dels compostos

Finalment es procedeix a generar un programa que englobi tots els processos i algorismes programats. Es generarà també una interfície gràfica senzilla i clarificadora, que inclogui totes les funcionalitats.

Per començar, es dissenya el següent diagrama de flux pensant en els possibles usos que necessitaria l'usuari o usuària:



Figura 24. Diagrama de flux del programa.

4.2.1 Interfície gràfica

La interfície creada presenta sis botons: tres que executen cadascun una funció de l'algoritme (transformar, interpolar o quantificar) i tres que serveixen com a instruccions, tot obrint una finestra emergent per tal que l'usuari conegui com ha d'utilitzar les diferents funcionalitats d'aquesta interfície. Aquesta finestra emergent es genera amb la funció *msgbox*.

Per generar i dotar de funcionalitat als botons s'utilitza la funció *uicontrol* que contindrà tots els paràmetres de disseny, així com la funció a la qual cridarà en cada cas.

Es pot consultar el codi a l'annex 10.

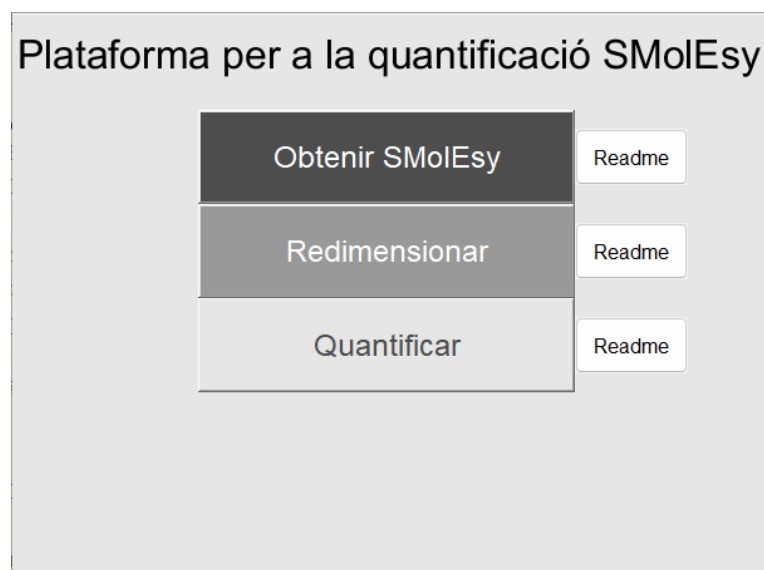


Figura 25. Interfície gràfica de la plataforma.

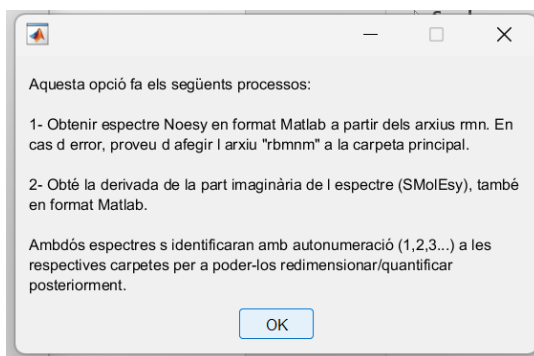


Figura 26. Missatges "readme" d'ajuda a l'usuari/ària.

4.2.2 Transformada SMoIESY

Per aquest apartat es crea una interacció amb l'usuari que podrà seleccionar la carpeta que conté el conjunt d'arxius, així com les carpetes on s'hi allotjaran els arxius Noesy i SMoIESY sortints.

S'hi afegeix també un codi de detecció de les carpetes existents al directori seleccionat per l'usuari.

La resta de detalls de la transformació s'expliquen prèviament als apartats 3.4.1 i 3.4.2., mentre que el codi es pot trobar a l'annex 11.

4.2.3 Interpolació

Igual que a l'apartat anterior, s'afegeix la part d'interacció amb l'usuari, ja que es requereix prèviament, a més de les carpetes d'entrada i sortida, un espectre amb la resolució desitjada per a que l'algoritme tingui la referència per a interpolar.

A l'annex 12 se'n poden consultar els detalls.

4.2.4 Integració

Aquesta és la funció més extensa de tot l'algoritme. Inicialment, s'inicialitza una variable de tipus taula que genera tota la informació dels metabòlits necessària per l'algoritme: nom del metabòlit, tipus de pic, finestres i subfinestres dels pics, punts on es troba el màxim (per alinear els espectres) i la referència de la columna del full de càlcul per tal de fer la posterior correlació.

Es sol·licita a l'usuari que triï el conjunt d'espectres que vol quantificar.

Es fa aparèixer un menú d'opcions, ja que els metabòlits es quantifiquen un a un.

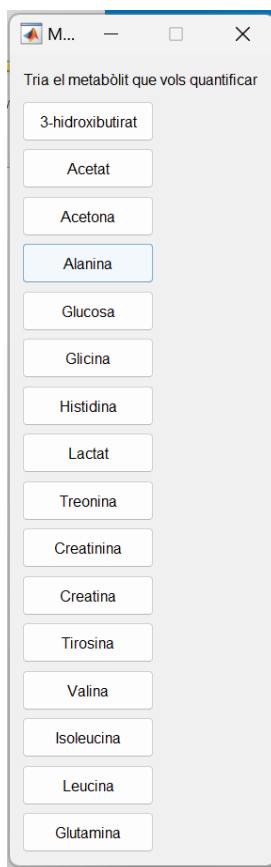


Figura 27. Interfície de l'opció "quantificar".


A continuació, el programa va a cercar les dades del metabòlit prèviament inicialitzades.

A partir d'aquest punt, el programa diferencia entre els diversos tipus de pics que es poden quantificar mitjançant un *switch*: singlets, dobles, triplets, quadruplets i casos especials.

En funció del cas, el programa ha d'integrar un sol pic o bé diversos i sumar-ne les àrees calculades. D'altra banda, el cas especial és el de la creatina, ja que el pic on es troba està superposat amb la creatinina, i el que es fa es quantificar-los per separat i restar la integral de la creatinina (obtingut a un altre punt de l'espectre) de l'àrea total.

Finalment, s'exporten les quantificacions obtingudes a un full de càlcul, en format columna, que es guarda amb el nom del metabòlit que s'està integrant, a diferència del cas de la Plataforma SMolESY que donava aquest *output* com a fitxer *.txt*. Això està pensat per a facilitar-ne el tractament a posteriori.

Aquest algorisme es troba complet a l'annex número 13.



| | A |
|----|---------|
| 1 | 1375,13 |
| 2 | 1029,36 |
| 3 | 1429,83 |
| 4 | 763,462 |
| 5 | 1680,73 |
| 6 | 1113,24 |
| 7 | 1710,89 |
| 8 | 1090,37 |
| 9 | 1679,37 |
| 10 | 1226,51 |
| 11 | 1080,15 |
| 12 | 1311,99 |
| 13 | 1466 |
| 14 | 1551,37 |
| 15 | 1653,51 |
| 16 | 1413,64 |
| 17 | 790,306 |
| 18 | 1437,16 |
| 19 | 853,937 |
| 20 | 1018,86 |
| 21 | 1005,32 |
| 22 | 1645,48 |

Figura 28. Format de sortida de les dades quantificades.

4.2.5 Correlació

Finalment, dins de la funció anterior, es correlen les quantificacions mitjançant el coeficient de Pearson amb les del full de càlcul. Aquest pas només té sentit per a la base de dades proporcionada per *Biosfer Teslab*, però s'haurà d'adaptar el codi en cas que es treballi amb una base de dades diferent amb una altra estructura.

Al 14è i últim annex es pot consultar aquest algoritme.

4.3 Correlacions (resultats)

Com a resultat final de l'estudi, s'aplica l'algoritme a tot el conjunt de dades i es quantifiquen i correlen tots els metabòlits amb les quantificacions prèvies disponibles a partir del CPMG.

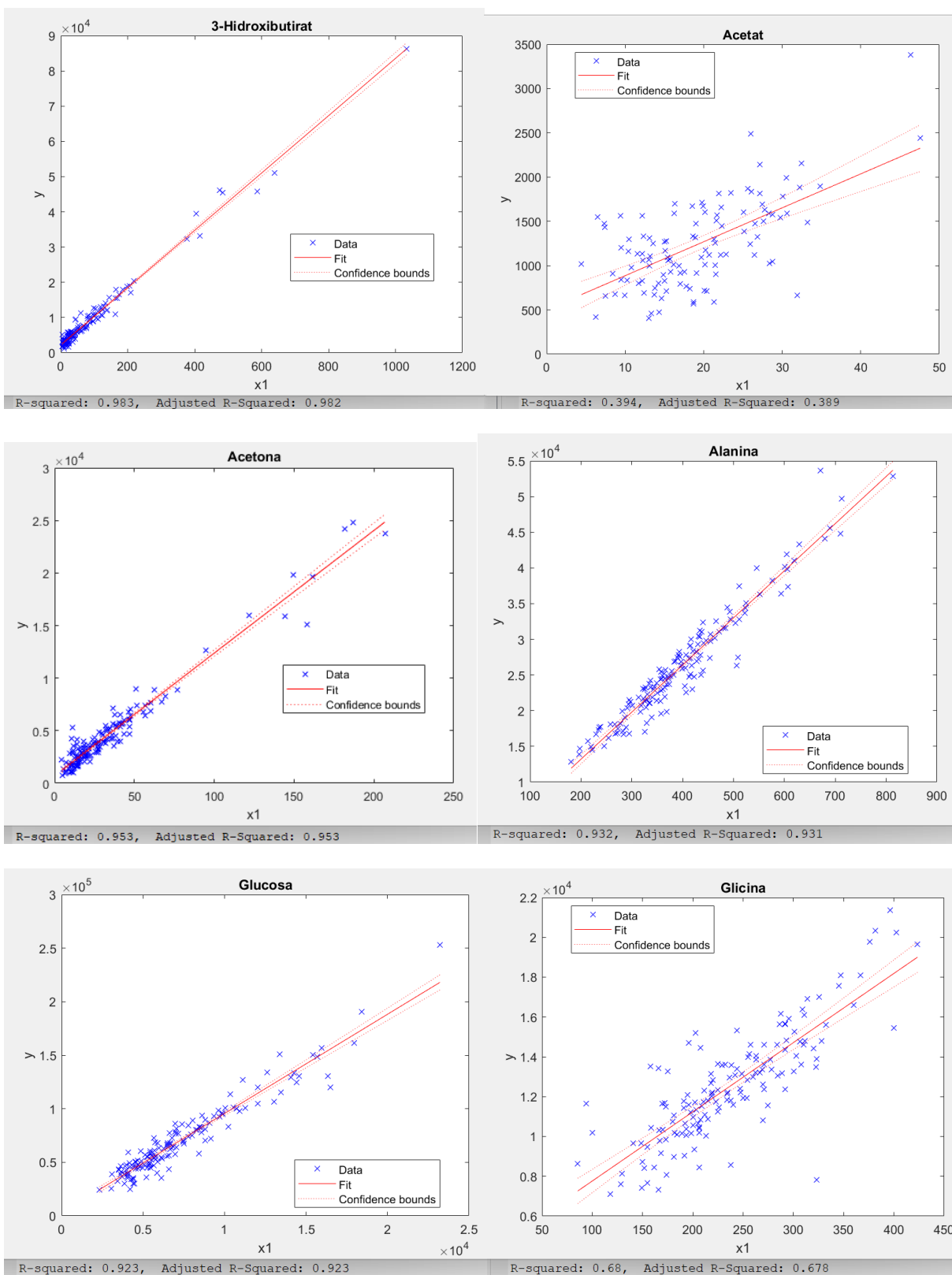
Biosfer Teslab proporciona en la següent taula un llistat dels metabòlits més habituals que es quantifiquen.

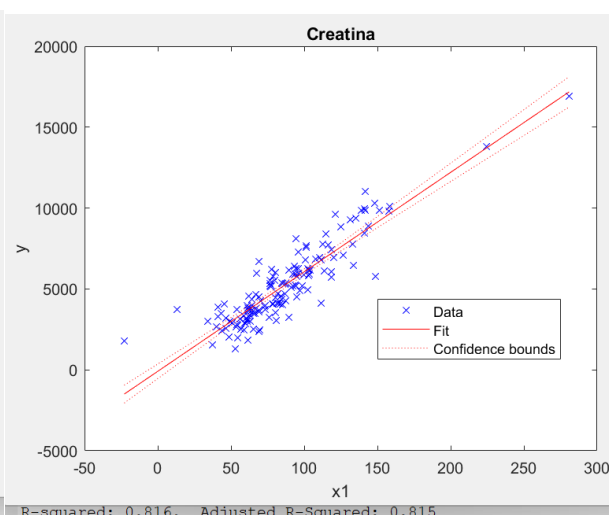
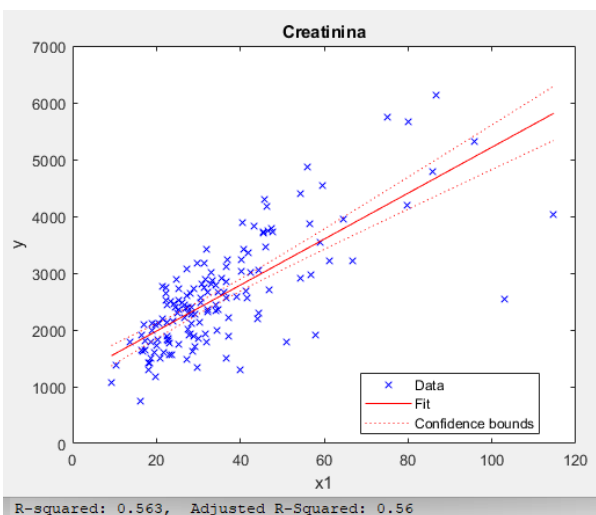
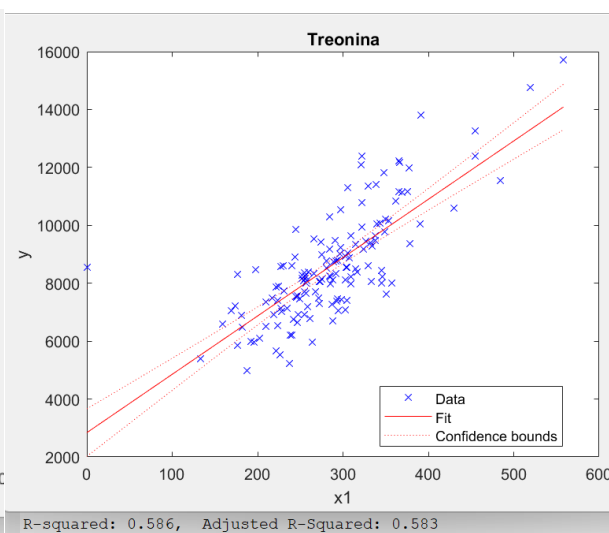
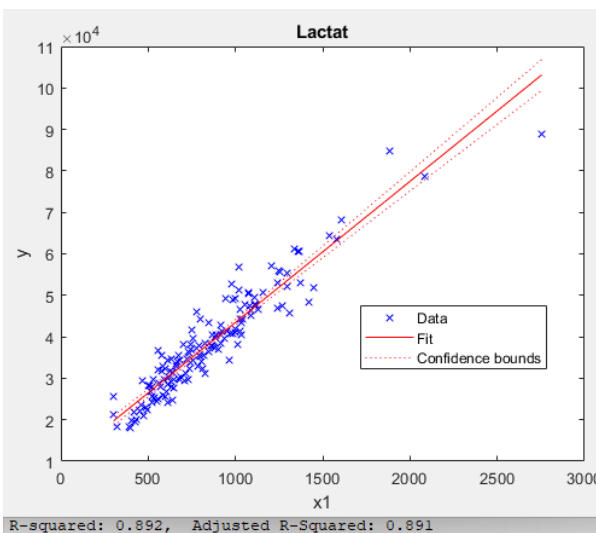
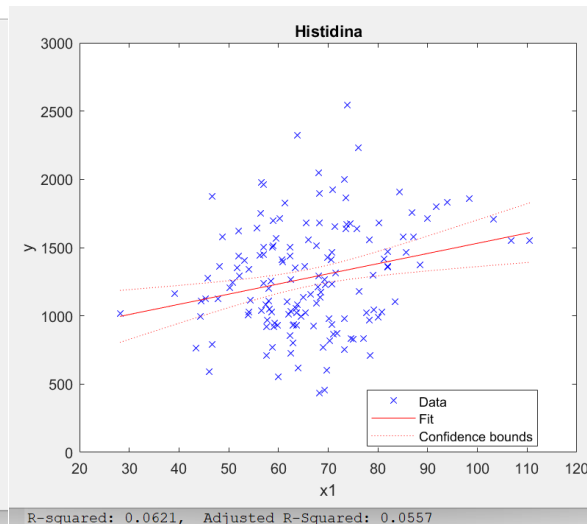
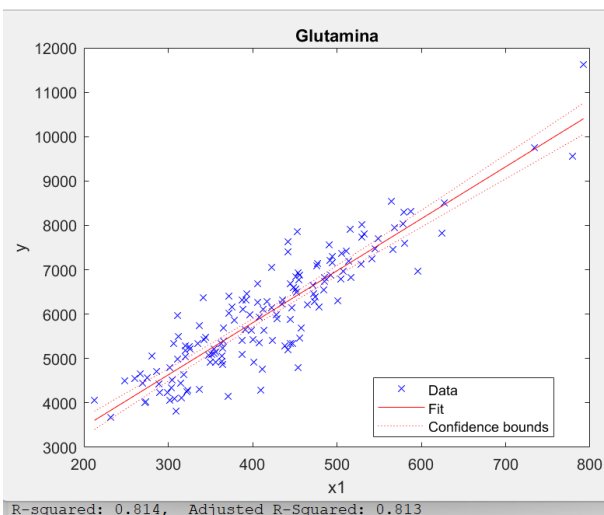
| Metabòlit | Regió (ppm) | Forma | Link |
|-------------------|----------------------|-------------------|---|
| 3-Hydroxybutyrate | 1.22-1.16 | Doblet | https://hmdb.ca/metabolites/HMDB0000011 |
| Acetate | 1.91-1.895 | Singlet | https://hmdb.ca/metabolites/HMDB0000042 |
| Acetone | 2.23-2.205 | Singlet | https://hmdb.ca/metabolites/HMDB0001659 |
| Alanine | 1.48-1.45 | Doblet | https://hmdb.ca/metabolites/HMDB0000161 |
| Glucose | 5.235-5.21 | Doblet | https://hmdb.ca/metabolites/HMDB0000122 |
| Glutamate | 2.36-2.32 | Multiplet | https://hmdb.ca/metabolites/HMDB0000148 |
| Glutamine | 2.47-2.41 | Multiplet | https://hmdb.ca/metabolites/HMDB0000641 |
| Glycerol | 3.7-3.62 | Doble doblet | https://hmdb.ca/metabolites/HMDB00131 |
| Mannitol | 3.7-3.62 | Multiplet | https://hmdb.ca/metabolites/HMDB0000765 |
| Glycine | 3.56-3.5425 | Singlet | https://hmdb.ca/metabolites/HMDB0000123 |
| Histidine | 7.06-7.035 | Singlet | https://hmdb.ca/metabolites/HMDB0000177 |
| Lactate | 1.33-1.29/4.12-4.08 | Doblet/Quadruplet | https://hmdb.ca/metabolites/HMDB0000190 |
| Threonine | 3.61-3.585 | Doblet | https://hmdb.ca/metabolites/HMDB0000167 |
| Creatinine | 4.05-4.03 | Singlet | https://hmdb.ca/metabolites/HMDB0000562 |
| Cretaine | 3.035-3.02 | Singlet | https://hmdb.ca/metabolites/HMDB0000064 |
| Tyrosine | 6.92-6.86 | Doblet | https://hmdb.ca/metabolites/HMDB0000158 |
| Valine | 1.045-1.015 | Doblet | https://hmdb.ca/metabolites/HMDB0000883 |
| Isoleucine | 1.01-0.995/0.93-0.91 | Doblet/Triplet | https://hmdb.ca/metabolites/HMDB0000172 |
| Leucine | 0.97-0.93 | Triplet | https://hmdb.ca/metabolites/HMDB0000687 |

Figura 29. Llistat detallat dels metabòlits més habituals quantificats a Biosfer Teslab.

Posteriorment es decideix prescindir de la quantificació d'alguns d'ells per diferents motius:

- Mannitol: a la base de dades donada amb les quantificacions del CPMG només està quantificat en un petit grup de mostres, de manera que, amb tan poques dades, no es considera que el resultat pugui ser concloent.
- Glicerol: per la forma del pic (doble doblet) de moment es descarta el seu estudi. A més, és difícil de quantificar donada la seva superposició amb el mannitol (mateixa franja de l'espectre).
- Glutamat: per la forma dels pics (multiplet) i la seva baixa amplitud en la gran majoria de mostres també es descarta la seva quantificació.





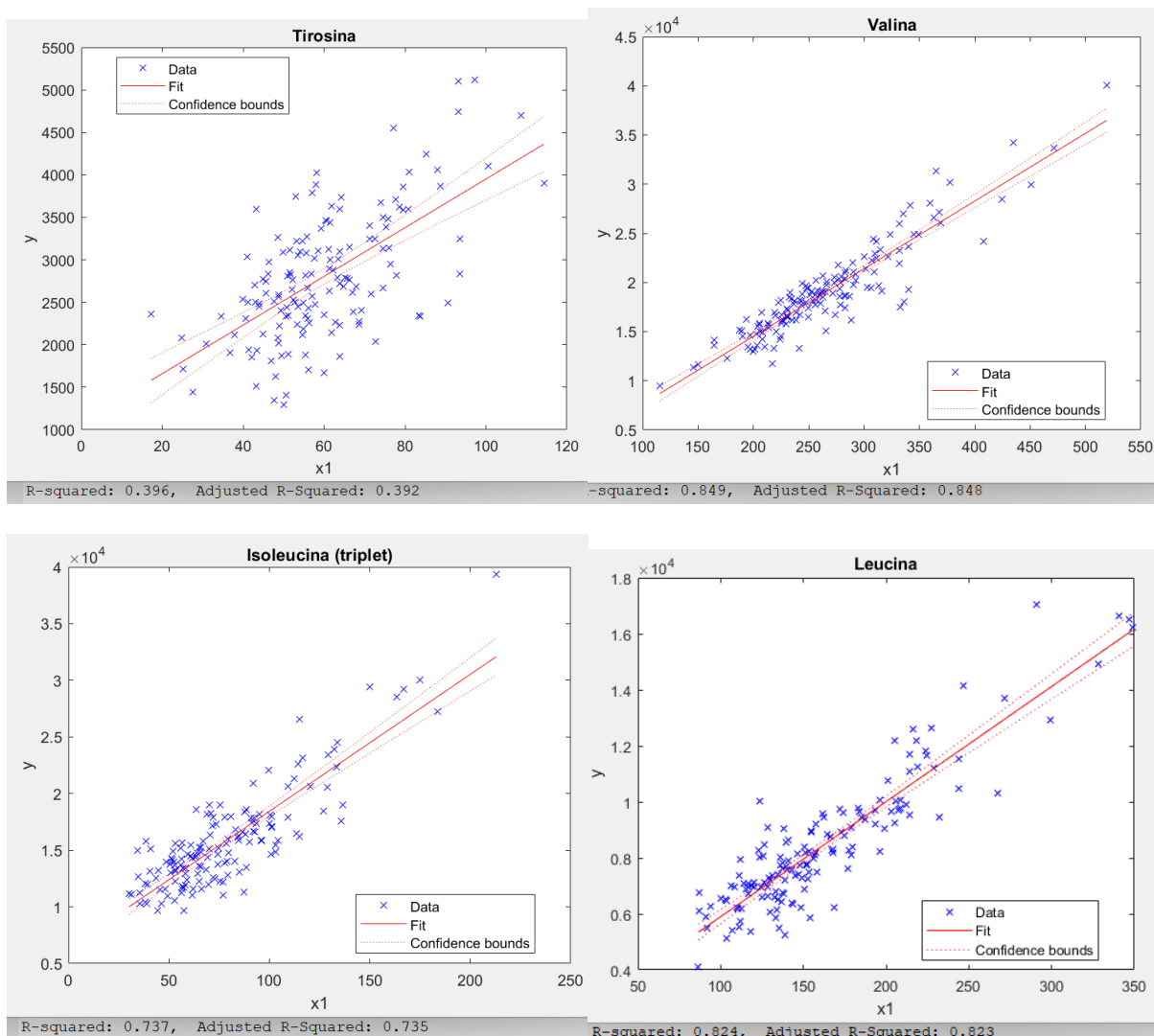


Figura 30. Correlacions entre les dades originals (quantificacions de Biosfer amb CPMG) i les quantificacions del projecte (SMoESY).

Les correlacions obtingudes es recullen a la següent taula:

Taula 1. Correlacions entre CPMG i SMoESY

| Metabòlit | R² |
|------------------|----------------------|
| 3-hidroxibutirat | 0.983 |
| Acetat | 0.389 |
| Acetona | 0.953 |
| Alanina | 0.932 |
| Glucosa | 0.923 |
| Glicina | 0.68 |
| Glutamina | 0.814 |
| Histidina | 0.06 |
| Lactat | 0.892 |
| Treonina | 0.586 |
| Creatinina | 0.563 |
| Creatina | 0.816 |
| Tirosina | 0.396 |
| Valina | 0.849 |
| Isoleucina | 0.737 |
| Leucina | 0.824 |

4.4 Discussió dels resultats

Com es pot observar a les figures anteriors, hi ha una quantitat considerable de metabòlits que presenten correlacions molt elevades entre les quantificacions realitzades i les originals (CPMG). El 3-hidroxiacetat, l'acetona, l'alanina i la glucosa presenten valors de R^2 que superen el 0.9; seguidament, la glutamina, el lactat, la creatina, la valina i la leucina presenten correlacions de valor força acceptable de $R^2 > 0.8$, i fins i tot amb la isoleucina ($R^2 > 0.737$). S'observa que aquests pics són els que tenen les amplituds més elevades:

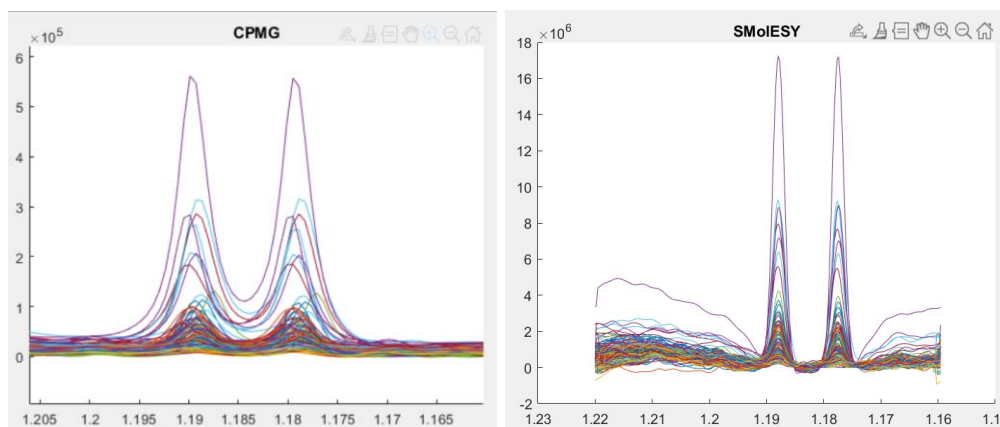


Figura 31. Comparativa CPMG vs SMOIESY del metabòlit 3-hidroxiacetat

D'altra banda, s'observen alguns pics amb unes correlacions força lluny de les esperades: és el cas de l'acetat, la glicina, la histidina, la treonina, la creatinina i la tirosina. Aquests resultats es posen en coneixement de *Biosfer Teslab* per tal de realitzar-ne una valoració més acurada basada en veus expertes. Es conclou que aquests metabòlits presenten també algunes característiques comunes:

L'acetat i la creatinina són pics que presenten força dificultats a l'hora de ser quantificats també per la via del CPMG. Això és degut a que es troben superposats amb altres metabòlits propers i això obliga a utilitzar una tècnica molt recurrent que és la deconvolució.

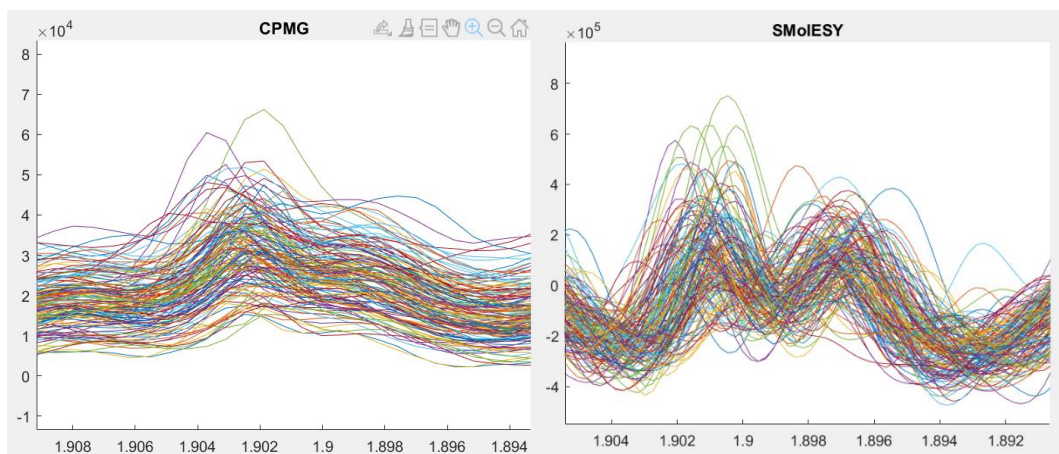


Figura 32. Exemple de l'acetat (pic de l'esquerra en ambdues imatges). Malgrat es troba més aïllat al SMOESY, es superposa amb un altre metabòlit (pic de la dreta), fent necessària la seva deconvolució.

La histidina i la tirosina són pics amb molt poca amplitud. La relació senyal/soroll és molt baixa tant al CPMG com a l'SMOESY, de forma que és molt complicada la seva quantificació en ambdós casos.

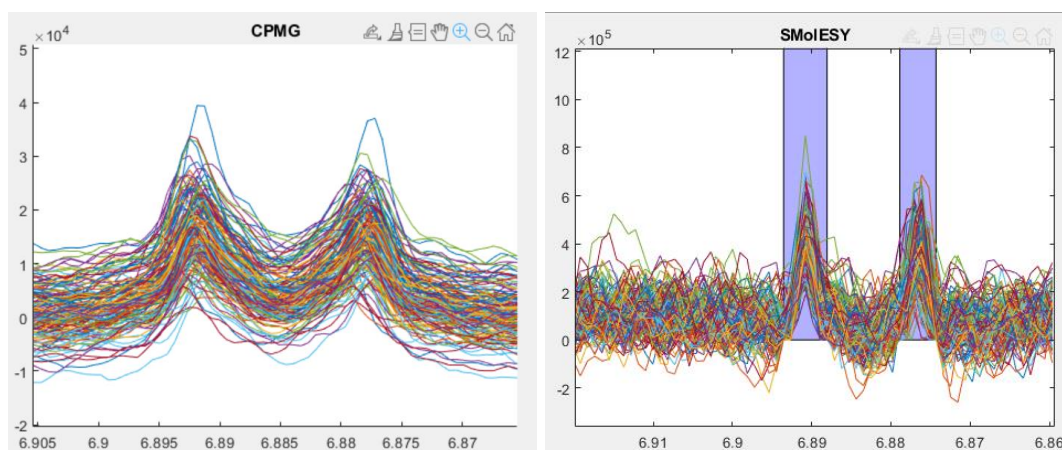


Figura 33. Comparativa CPMG vs SMOESY del metabòlit tirosina.

Quelcom similar succeeix amb a treonina i la glicina, tot i que en aquests casos les amplituds són majors als casos anteriors, fet que explica que, dins d'aquest grup de metabòlits amb correlacions més baixes, és en els quals s'obtenen millors resultats ($R^2=0.58$ i 0.68 , respectivament).

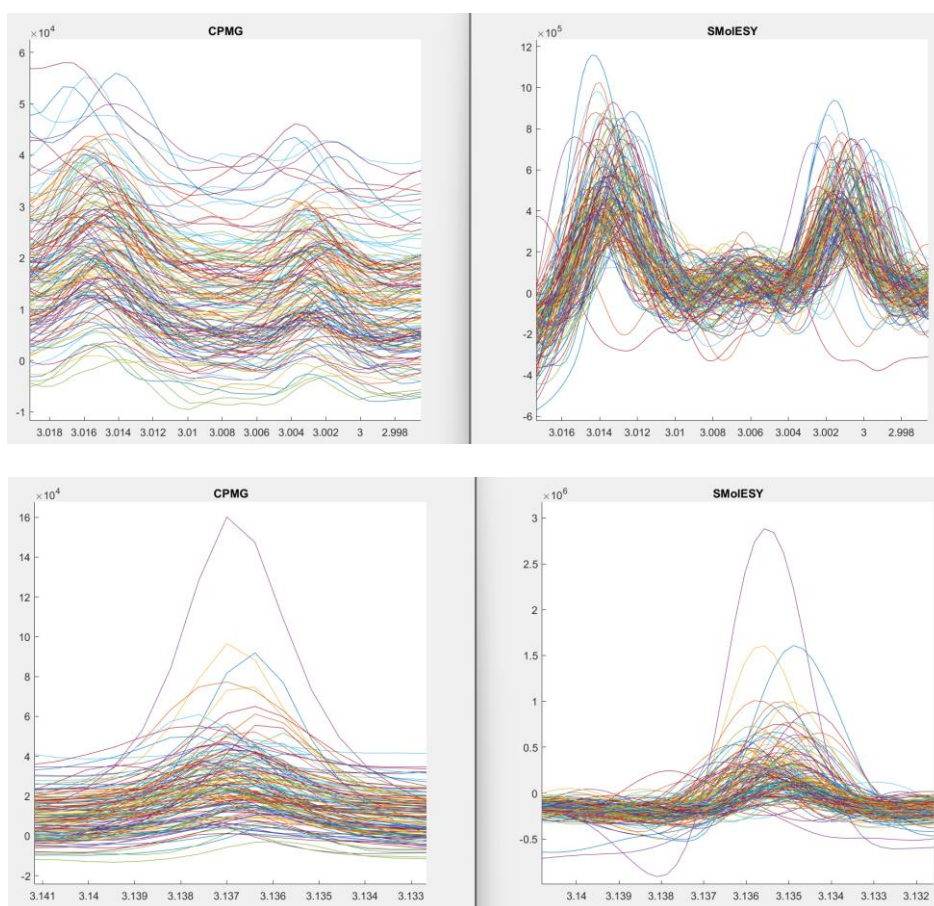
Això porta a concloure que, donada la comparació d'amplituds entre uns i altres espectres (CPMG vs SMOESY), aquesta tècnica de quantificació combinada en determinats casos amb la tècnica de deconvolució de pics pot esdevenir més precisa.

4.5 Estudi posterior de l'espectre: altres compostos d'interès

Com a part final del projecte, i davant dels resultats positius obtinguts en les correlacions, es planteja la possibilitat que altres metabòlits més enllà dels quantificats a partir dels espectres CPMG puguin ser quantificats amb la tècnica SMoIESY.

Inicialment, s'observa una sèrie de metabòlits que apareixen citats a l'article *SMoIESY: an efficient and quantitative alternative to on-instrument macromolecular 1H-NMR signal suppression* [8], però que no són objecte de quantificació amb el software de *Biosfer Teslab*: es tracta de la lisina, el dimetilsulfòxid, l'àcid cítric o citrat i la colina.

En comparació, s'observa com aquests presenten una millor amplitud a l'espectre transformat respecte al CPMG, esdevenint així més fiable la seva quantificació.



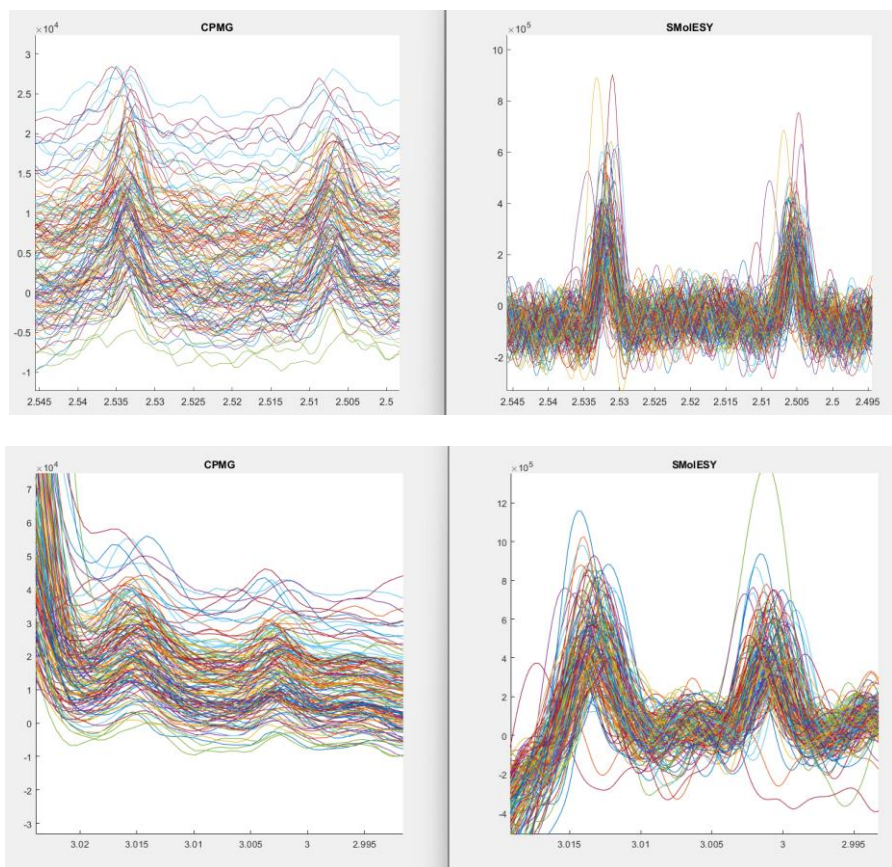


Figura 34. Comparació dels pics de la lisina, dimetilsulfòxid, citrat i colina (de dalt a baix) en espectre CPMG (esquerra) vs espectre SMoIESY (dreta).

Per ampliar encara aquesta recerca, es fa un estudi minuciós al llarg de l'eix del desplaçament químic per observar les ressonàncies a cada punt. Com a referència, es parteix de l'article *750 MHz 1H and 1H-13C NMR Spectroscopy of Human Blood Plasma* [17].

S'estableix una comparativa entre tots aquests metabòlits i se'n registren algunes conclusions, recollides en la següent taula comparativa:

LLEGENDA

- Ja han estat quantificats.
- Millora la seva amplitud respecte el CPMG.
- Resultats similars.
- Millora els resultats del CPMG però cal identificar els pics de forma aïllada
- No millora resultats del CPMG

| Molècula/es | Comparativa, observacions i conclusions |
|---|--|
| Formate | Millor al CPMG |
| 1-metilhistidina | Millor al CPMG |
| Histidina | Cap dels dos |
| 3-metilhistidina | Cap dels dos |
| Fenilalanina | Baixa intensitat, però millor a SMoIESY |
| Tirosina | Pics independents |
| 1-metilhistidina | Cap dels dos |
| Histidina | JA QUANTIFICAT AMB SMOLESY |
| 3-metilhistidina | Cap dels dos |
| Tirosina | JA QUANTIFICAT AMB SMOLESY |
| "unsaturated lipid" | Cap dels dos |
| Glucosa | JA QUANTIFICAT AMB SMOLESY |
| Glycerol of lipids | Cap dels dos |
| B-glucosa | Millor al CPMG |
| B-galactosa | Cap dels dos |
| Colina | Cap dels dos |
| Glycerol of lipids | Cap dels dos |
| Treonina | Molt poca amplitud, millor al CPMG |
| 3-hidroxibutirat | Cap dels dos |
| Prolina | Cap dels dos |
| Lactat | JA QUANTIFICAT AMB SMOLESY |
| Myo-inositol | Cap dels dos |
| Glycerol of lipids | Cap dels dos |
| Creatinina | JA QUANTIFICAT AMB SMOLESY |
| Histidina/fenilalanina | Més aïllats al SMoIESY |
| Tirosina | Doble doblet. Té millor amplitud |
| Creatina | Millor amplitud al SMoIESY |
| B-glucosa | Pics independents |
| Glicerol | Cap dels dos |
| A-glucosa | Doble doblet doblet. Pics més independents |
| Alanina A-glucosa Citrulina Leucina Glutamina | Cal identificar-los per separat. Millor al SMoIESY |
| Colina | Poca amplitud en tots dos |
| Glicerol/myo-inositol | Només es veu un doblet. Lleugerament millor al SMoIESY |
| Treonina | JA QUANTIFICAT AMB SMOLESY |
| Valina/glicerol/myo-inositol | Pics independents |
| Glicina | Més definit al SMoIESY |
| A-glucosa | Més definit al SMoIESY |

| | |
|---|---|
| B-glucosa, treonina, prolina | Triplet glucosa i doble doblet treonina? Més definits SMoESY |
| A,B-glucosa i taurina | Molt millor a SMoESY (triplets) |
| Prolina | Lleugerament millor SMoESY (multiplet). Es pot quantificar en qualsevol cas. |
| Myo-inositol | Menys superposició (triplet), però poca amplitud |
| Fenilalanina TMAO (trimetilamine oxine) Taurina Histidina B-glucosa Arginina Colina | Molt similar quant a amplitud |
| Tirosina, Citrulina, histidina, fenilalanina | Haurien de ser dobles dobles i triplet. Hi ha un pic resolt que es desconeix a què correspon. |
| Creatina, creatinina | JA QUANTIFICAT AMB SMOESY |
| Tirosina | Doble doblet. Millor al SMoESY els dos pics de la dreta. |
| Albumin lysyl | Cap dels dos |
| Asparagine | Cap dels dos |
| Trimetilamina | Cap dels dos (molt poca amplitud) |
| Aspartate | Cap dels dos |
| Lipid | Cap dels dos |
| Dimetilamina | Millor al CPMG |
| Citrat | Cap dels dos |
| Aspartate | Cap dels dos |
| Metilamina | Molt millor a SMoESY |
| Citrat | Cap dels dos |
| 2-oxoglutarat | Cap dels dos |
| Glutamina | JA QUANTIFICAT AMB SMOESY |
| 3-hidroxi-butirat, prolina, piruvat | Més amplitud però més soroll al SMoESY. |
| Glutamat | JA QUANTIFICAT AMB SMOESY |
| 3-hidroxi-butirat | Molt sorollós. Millor al CPMG |
| Valina + lipid | Més estret al SMoESY |
| Acetoacetat (acetona) | JA QUANTIFICAT AMB SMOESY |
| Glutamat, methioninne | Pics més definits |
| Glutamina | Millor al SMoESY |
| Glicoproteïna/prolina | Millor al SMoESY. Es pot deconvolucionar, tot i que glicoproteïnes es mesuren millor amb LED. |
| Glutamat, prolina, isoleucina + lipids | Més amplitud però més soroll al SMoESY |
| Acetat | JA QUANTIFICAT AMB SMOESY |
| Citrulina | Cap dels dos |
| Leucina, lípid, lisina, arginina | Més amplitud però més soroll al SMoESY |
| Citrulina, lípid (VLDL) | Cap dels dos |
| Lisina, isoleucina | Cap dels dos |
| Alanina | JA QUANTIFICAT AMB SMOESY |
| Lactat | JA QUANTIFICAT AMB SMOESY |

| | |
|--------------------------------------|---|
| Lípids, treonina, focuse, isoleucina | SMoIESY hi ha pic, al CPMG no |
| Lípids (LDL) | Més definit al SMoIESY. Millor amb el LED. |
| 3-hidroxitirac | JA QUANTIFICAT AMB SMOLESY |
| Isobutirat | Més amplitud però més soroll al SMoIESY |
| Valina | JA QUANTIFICAT AMB SMOLESY |
| Isoleucina | JA QUANTIFICAT AMB SMOLESY |
| Valina | JA QUANTIFICAT AMB SMOLESY |
| Leucina | JA QUANTIFICAT AMB SMOLESY |
| Isoleucina | JA QUANTIFICAT AMB SMOLESY |
| Colesterol | Més amplitud però més soroll al SMoIESY. Millor LED |
| VLDL, LDL, lípid, colesterol | Pics més definits (no es superposen). Millor LED |
| Colesterol | Cap dels dos |

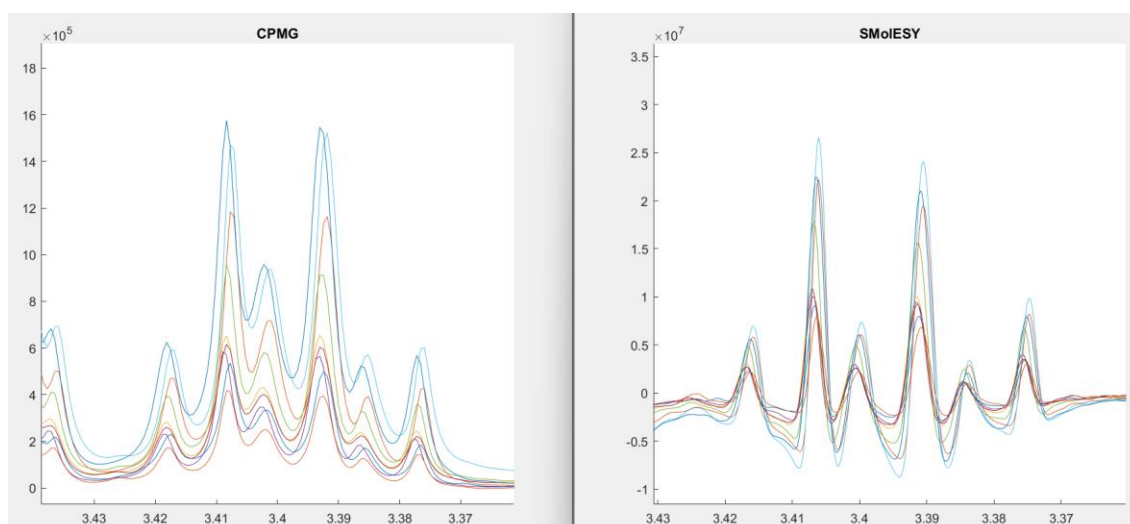


Figura 35. Comparativa de l'aïllament dels pics en CPMG (esquerra) vs SMoIESY (dreta). Es facilita la seva identificació en el segon cas.

La resta de comparatives es poden visualitzar a l'annex 15.

5. Conclusions

Com a tancament d'aquest treball, podem concloure que l'estudi que s'ha realitzat demostra de forma satisfactòria les hipòtesis generals plantejades inicialment:

- El mètode proposat permet quantificar de forma fiable un gran nombre de compostos de baix pes molecular a partir de la transformació dels espectres Noesy amb el mètode SMoIESY, de tal manera que es fa innecessària l'adquisició dels espectres CPMG, fet que suposa un important estalvi de temps en la seva mesura.

- La correlació de les quantificacions de l'SMoIESY amb les del CPMG són molt bones per la majoria de compostos, especialment amb metabòlits que tenen més amplitud. Això permet també obrir la possibilitat de quantificar compostos que amb altres tècniques eren complicats d'obtenir.

- S'ha dissenyat la plataforma per a la quantificació SMoIESY que integra totes les funcions necessàries per la quantificació dels compostos.

- S'han comparat diverses estratègies d'integració i quantificació, resultant que el millor mètode és el de referenciar els mínims i alinear els pics. En els metabòlits amb més d'un pic (doblets, triplet, quadruplet, doble doblet...) les quantificacions més precises s'obtenen integrant cada pic de forma independent i sumant-ne les àrees.

6. Referències

1. Frigolet, Maria E. y Gutiérrez-Aguilar, Ruth (2017). "Ciencias "ómicas", ¿cómo ayudan a las ciencias de la salud?", en *Revista Digital Universitaria (RDU)*, vol. 18, núm. 7, septiembre-octubre. DOI: <http://doi.org/10.22201/codeic.16076079e.2017.v18n7.a3> (Últim accés: 24 d'abril de 2023).
2. "Metabolomics: Eavesdropping on silent conversations between hosts and their unwelcome guests". PubMed Central (PMC). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5886577/> (Últim accés: 24 d'abril de 2023).
3. Koulman, A., & Volmer, D. A. (2008). Perspectives for Metabolomics in Human Nutrition: An Overview. *Nutrition bulletin*, 33(4), 324–330. <https://doi.org/10.1111/j.1467-3010.2008.00733.x> (Últim accés: 24 de maig de 2023).
4. "Comparison of NMR and MS | Metabolomics". EMBL-EBI homepage | EMBL-EBI. <https://www.ebi.ac.uk/training/online/courses/metabolomics-introduction/designing-a-metabolomics-study/comparison-of-nmr-and-ms/> (Últim accés: 24 de maig de 2023).
5. "Basic NMR Concepts: A Guide for the Modern Laboratory". (Últim accés: 24 de maig de 2023). [En línia]. Disponible: https://sites.bu.edu/cheminst/files/2021/06/CICNMR_basicconcepts.pdf
6. NMR-CTP Tutorial/Record Spectrum B12. http://www-usr.rider.edu/~grushow/nmr/NMR_tutor/pages/record/record_b12.html (Últim accés: 24 de maig de 2023).
7. Bliziotis, N.G., Engelke, U.F.H., Aspers, R.L.E.G. et al. A comparison of high-throughput plasma NMR protocols for comparative untargeted metabolomics. *Metabolomics* 16, 64 (2020). <https://doi.org/10.1007/s11306-020-01686-y> (Últim accés: 24 de maig de 2023).
8. Forster, M., Jones, C., & Mulloy, B. (1989). NOEMOL: integrated molecular graphics and the simulation of Nuclear Overhauser effects in NMR spectroscopy. *Journal of molecular graphics*, 7(4), 196–217. [https://doi.org/10.1016/0263-7855\(89\)80002-5](https://doi.org/10.1016/0263-7855(89)80002-5)
9. de Andrade, F. D., Netto, A. M., & Colnago, L. A. (2011). Qualitative analysis by online nuclear magnetic resonance using Carr-Purcell-Meiboom-Gill sequence with low refocusing flip angles. *Talanta*, 84(1), 84–88. <https://doi.org/10.1016/j.talanta.2010.12.033>
10. Takis, P. G.; Jiménez, B.; Sands, C. J.; Chekmeneva, E.; Lewis, M. R. SMoESY: An Efficient and Quantitative Alternative to on-Instrument Macromolecular 1H-NMR Signal Suppression. *Chem. Sci.* 2020, 11 (23), 6000– 6011. <https://doi.org/10.1039/D0SC01421D>
11. Takis, P. G., Jiménez, B., Al-Saffar, N. M. S., Harvey, N., Chekmeneva, E., Misra, S., & Lewis, M. R. (n.d.). SUPPORTING INFORMATION A computationally lightweight algorithm for deriving reliable metabolite panel measurements from 1D 1 H NMR. https://github.com/pantakis/SMoESY_platform (Últim accés: 24 de maig de 2023).

12. Nils Nyberg (2023) "RBNMR". MathWorks - Makers of MATLAB and Simulink - MATLAB & Simulink. <https://www.mathworks.com/matlabcentral/fileexchange/40332-rbnmr> (Últim accés: 24 de maig de 2023).
13. Islam, M. A., & Al-Shiha, A. (2018). Foundations of biostatistics. In Foundations of Biostatistics. Springer Singapore. <https://doi.org/10.1007/978-981-10-8627-4> (Últim accés: 24 de maig de 2023).
14. Si_Afifi, Abdelmonem A., et al., Practical Multivariate Analysis Sixth Edition. (2020).
15. "Documentation- MATLAB & Simulink- MathWorks España". MathWorks - Creadores de MATLAB y Simulink - MATLAB y Simulink - MATLAB & Simulink. <https://es.mathworks.com/help/> (Últim accés: 24 de maig de 2023).
16. "MS Windows batch file: need for-loop over multiple lines". Stack Overflow. <https://stackoverflow.com/questions/56694464/ms-windows-batch-file-need-for-loop-over-multiple-lines> (Últim accés: 24 de maig de 2023).
17. Nicholson, J. K., D Foxall, P. J., Spraul, M., Duncan Farrant, R., Lindon, J. C., & Clin, D. W. (1996). 750 MHz 1H and 1H-13C NMR Spectroscopy of Human Blood Plasma. In Sonksen, P. H. Biochem. J (Vol. 67, Issue 2). London.

7. Annexos

Annex 1: Algoritme d'importació de les dades a format MATLAB.

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 2: Algoritme per eliminar el codi identificatiu de l'arxiu de text que conté l'espectre SMoESY.

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 3: Algoritme per importar els espectres transformats des de la plataforma SMoIESY en arxius independents.

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 4: Algoritmes diversos per representar els espectres a demanda

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 5: Algoritme de quantificació mitjançant l'amplitud

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 6: Algoritme de quantificació mitjançant l'àrea (rectificant línia base)

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 7: Algoritme de quantificació mitjançant l'àrea (calculant recta base)

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 8: Algoritme de quantificació mitjançant l'àrea (integració directa)

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 9: Algoritme de quantificació mitjançant l'àrea (referenciant punts mínims)

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 10: Algoritme de la interfície gràfica

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 11: Algoritme de la transformada de l'espectre Noesy a SMoIESY

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 12: Algoritme d'interpolació

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 13: Algoritme de quantificació

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 14: Algoritme de correlació

Aquest apartat conté part de l'algoritme dissenyat, informació que es considera confidencial, i no s'inclou en aquesta versió de la memòria.

Annex 15: Estudi comparatiu d'altres compostos d'interès: figures

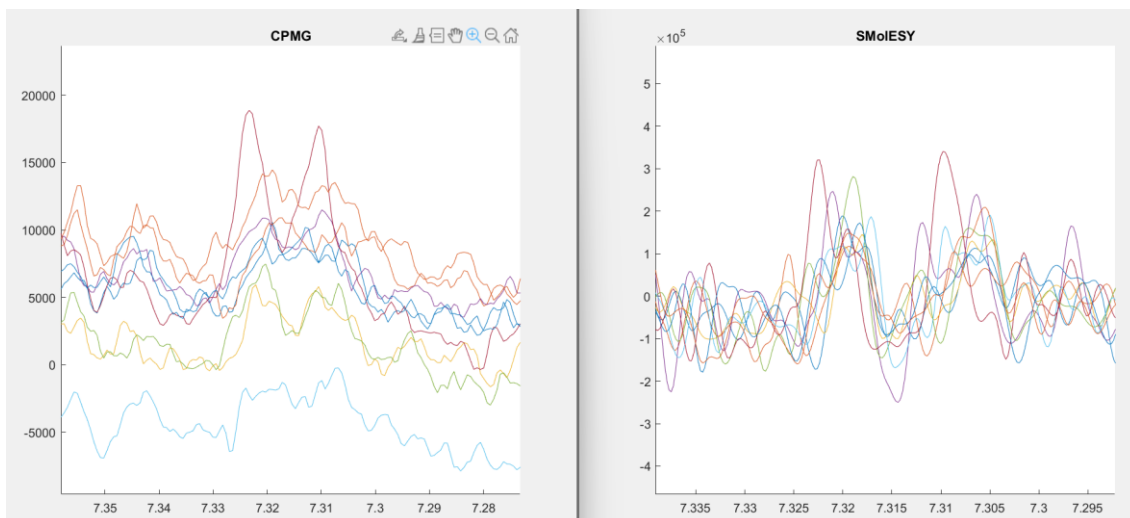


Figura 36. Fenilalanina

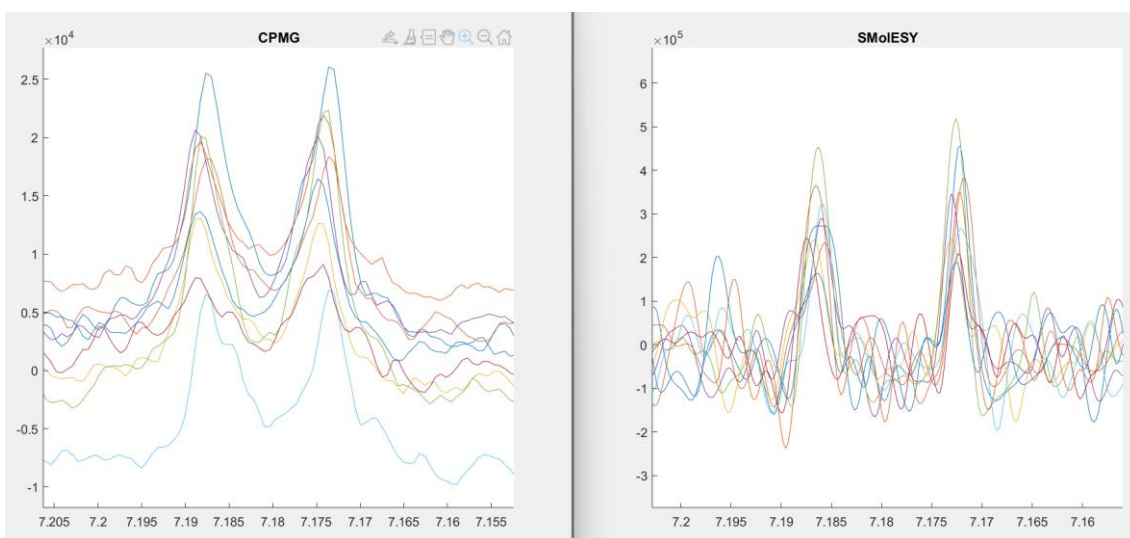


Figura 37. Tirosina

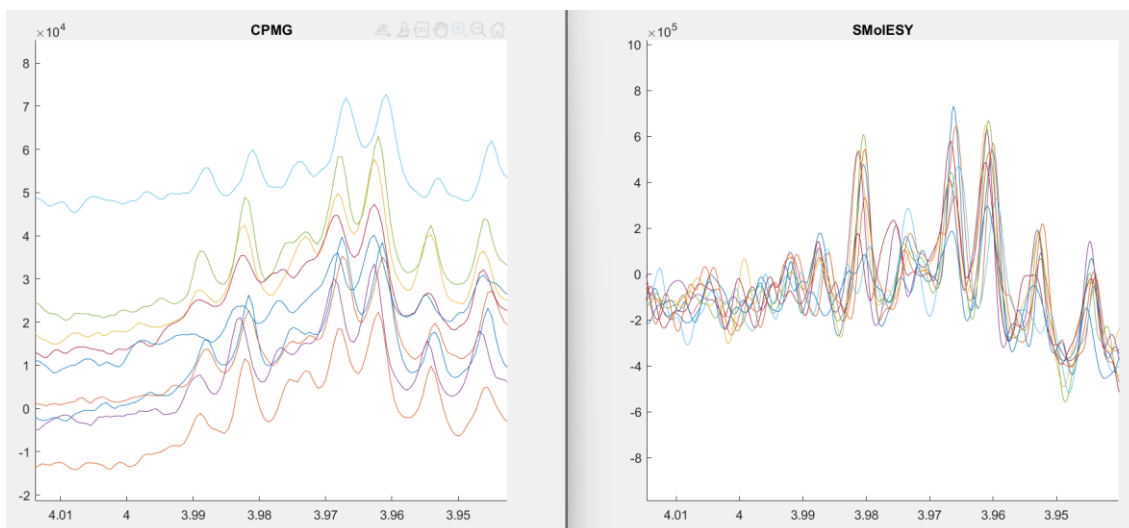


Figura 38. Histidina/fenilalanina

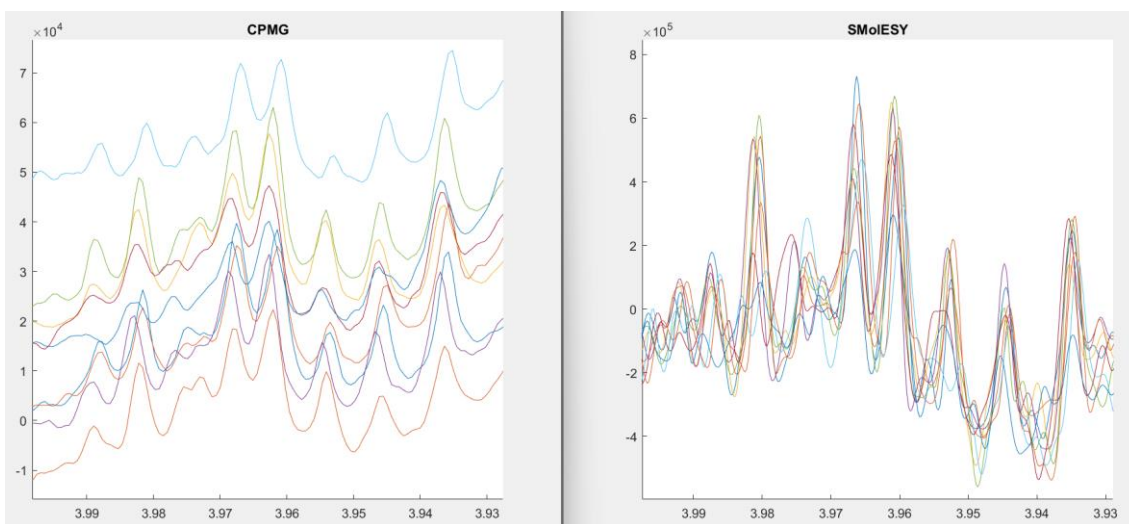


Figura 39. Tirosina

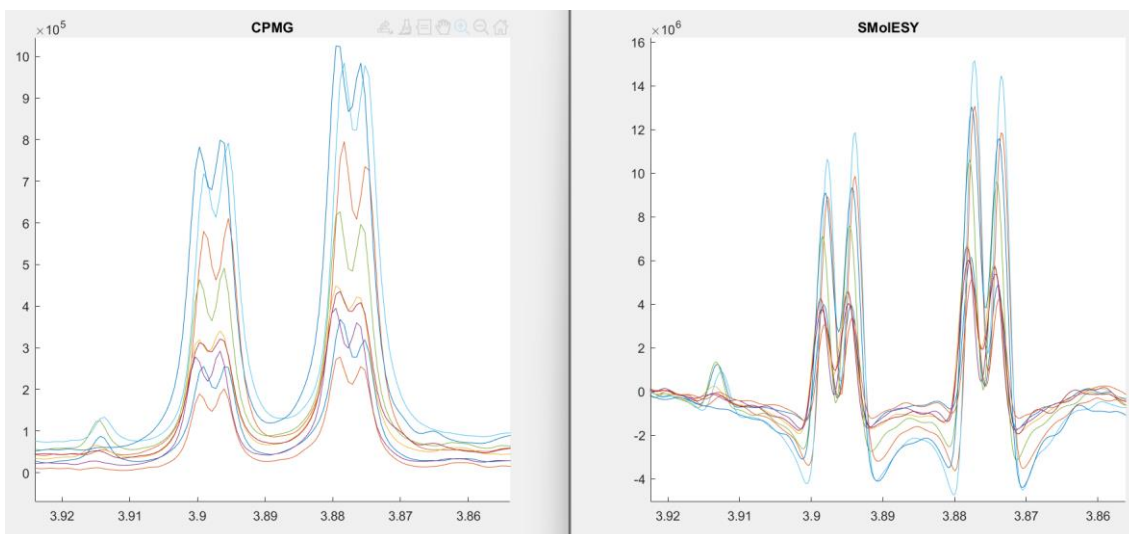


Figura 40. Creatina (3.93 ppm) i B-glucosa (doble doblet al 3.90ppm)

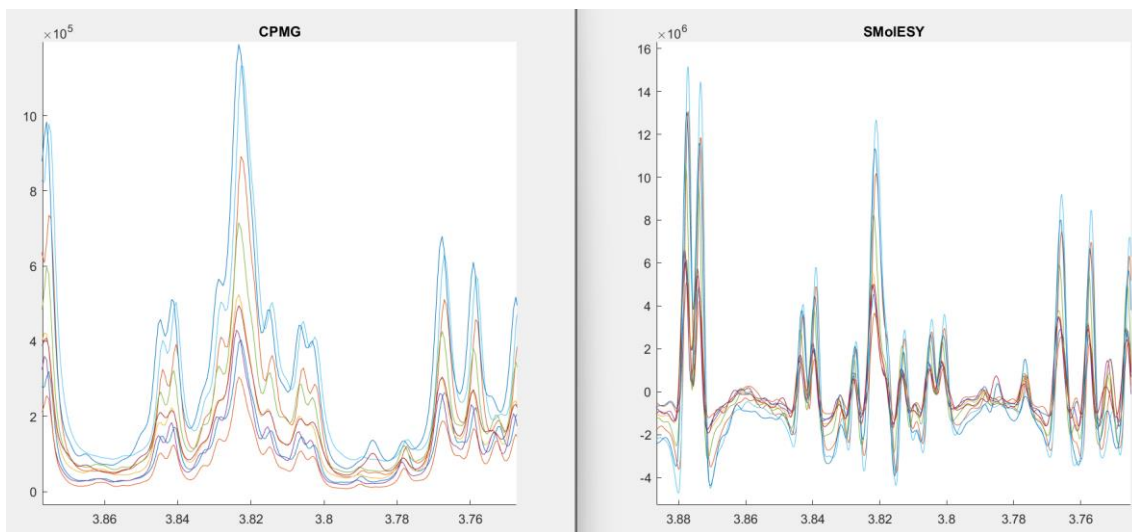


Figura 41. A-glucosa

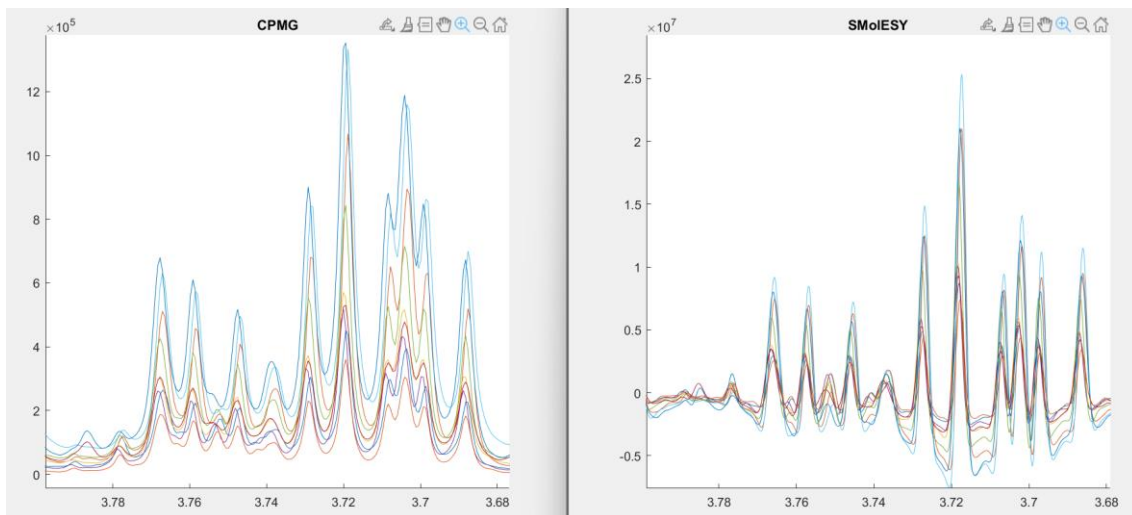


Figura 42. Pics entre 3.76 i 3.68ppm (alanina, A-glucosa, citrulina, leucina, glutamina)

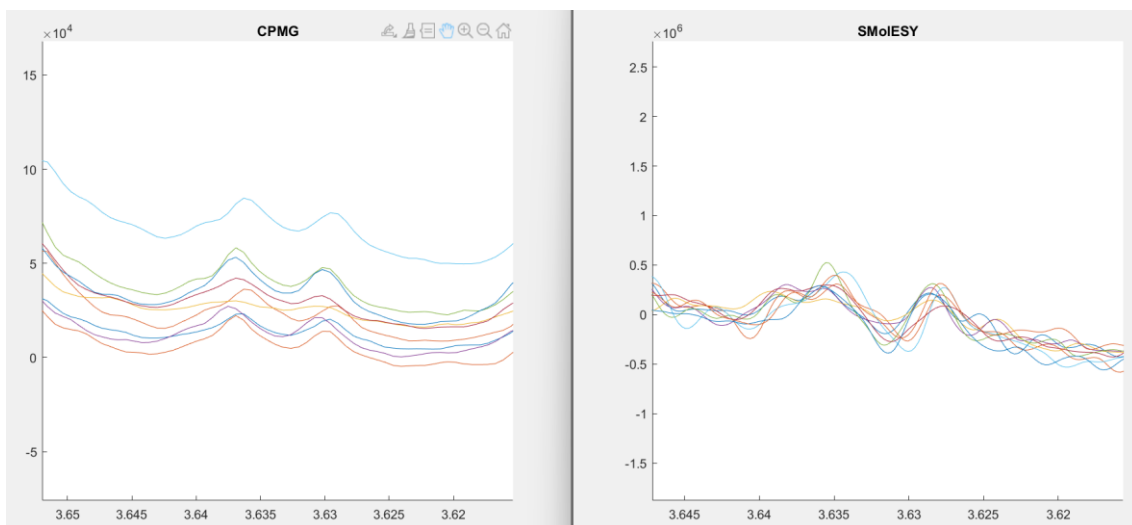


Figura 43. Glicerol/myo-inositol

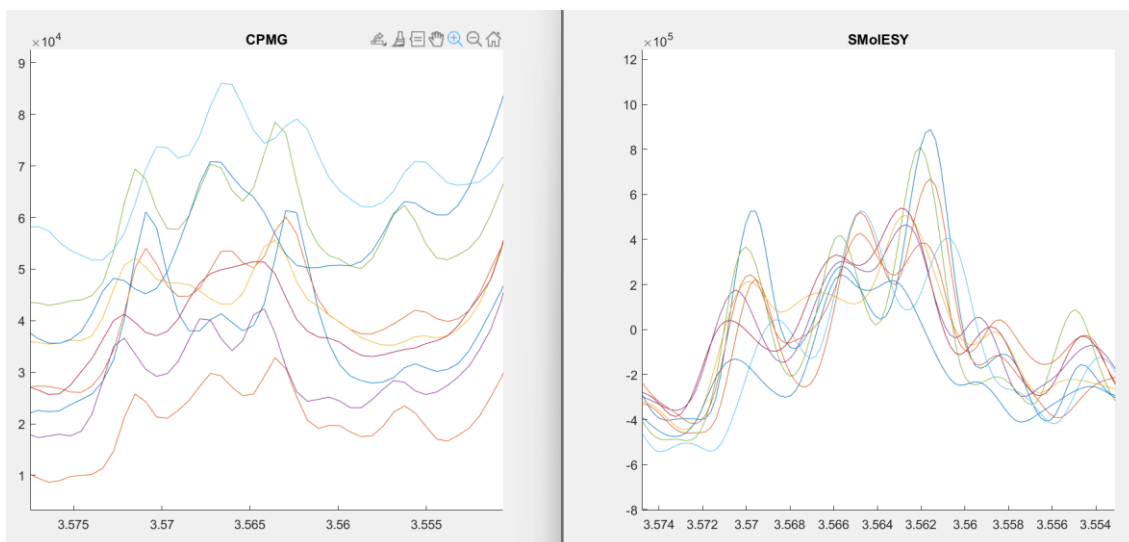


Figura 44. Valina(d)/glicerol(dd)/myo-inositol(dd)

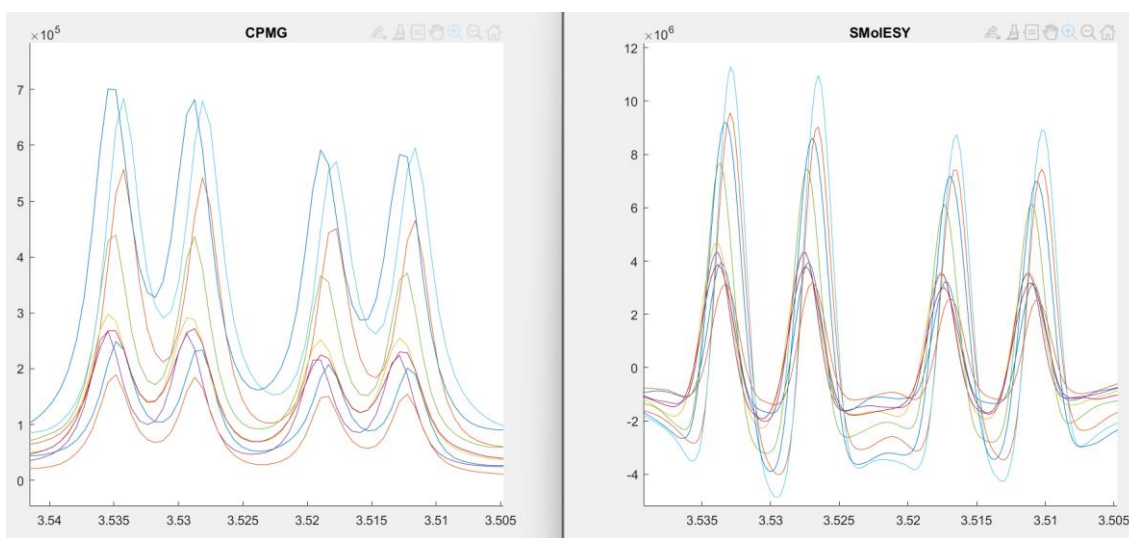


Figura 45. A-glucosa (3.54ppm)

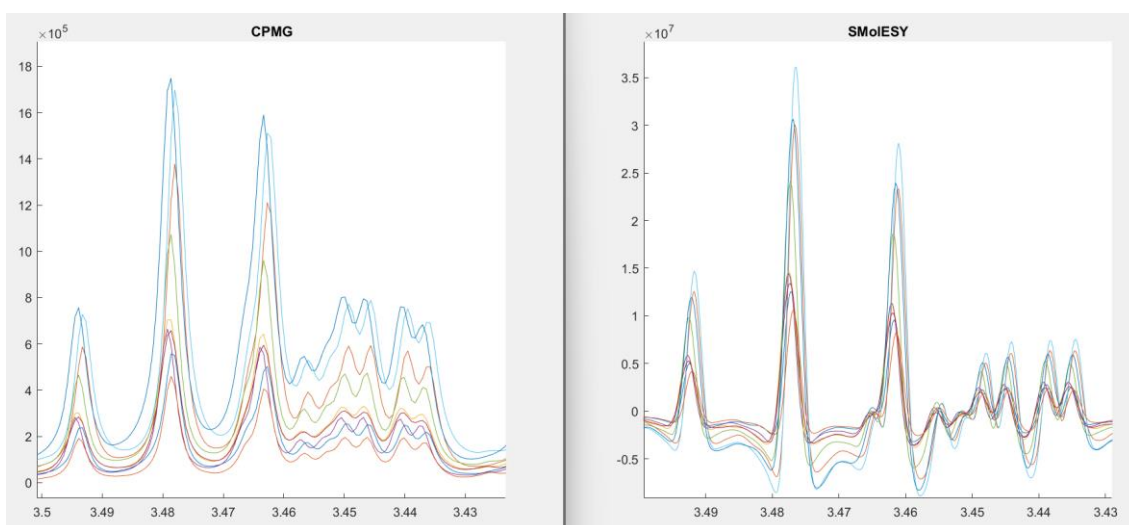


Figura 46. B-glucosa, treonina, prolina (3.48 a 3.45 ppm)

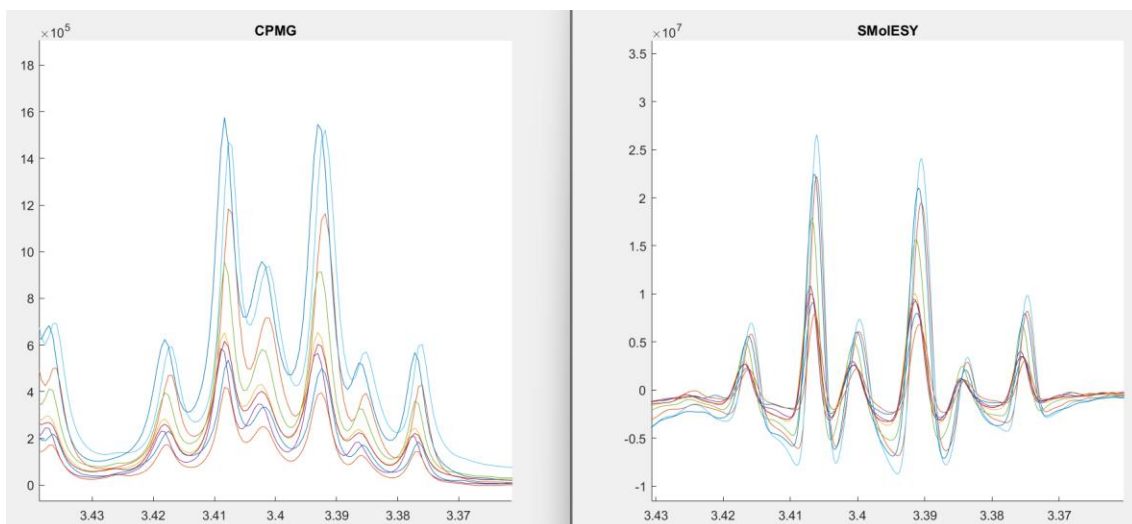


Figura 47. A,B-glucosa i taurina (3.42-3.40ppm)

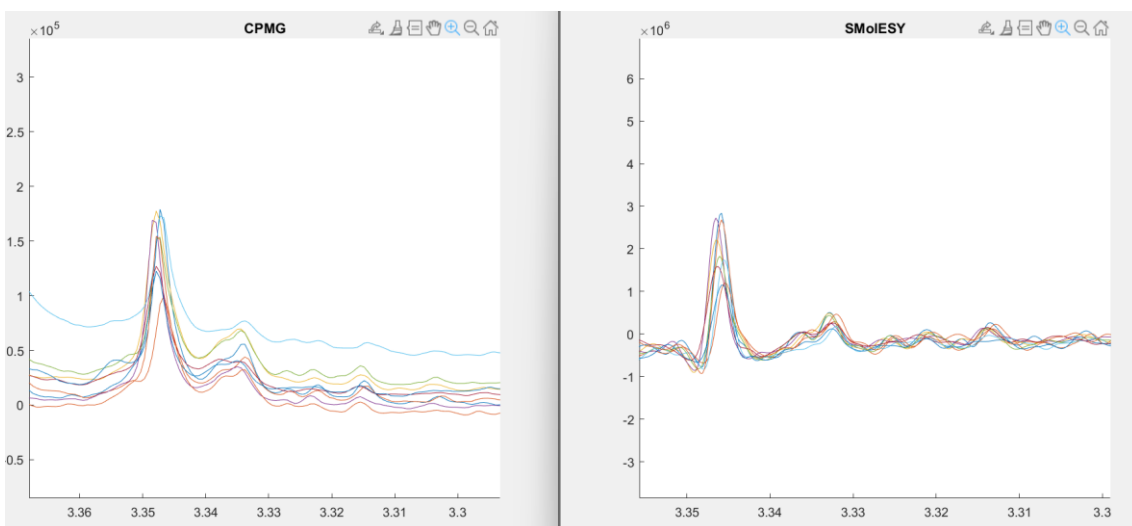


Figura 48. Prolina (multiplet 3.34ppm)

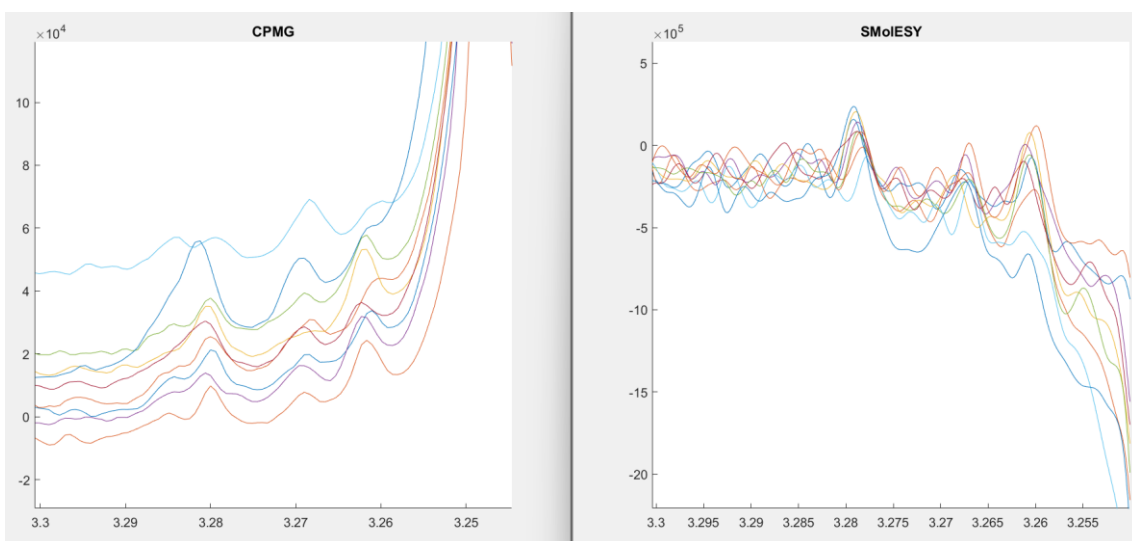


Figura 49. Myo-inositol (3.28 ppm)

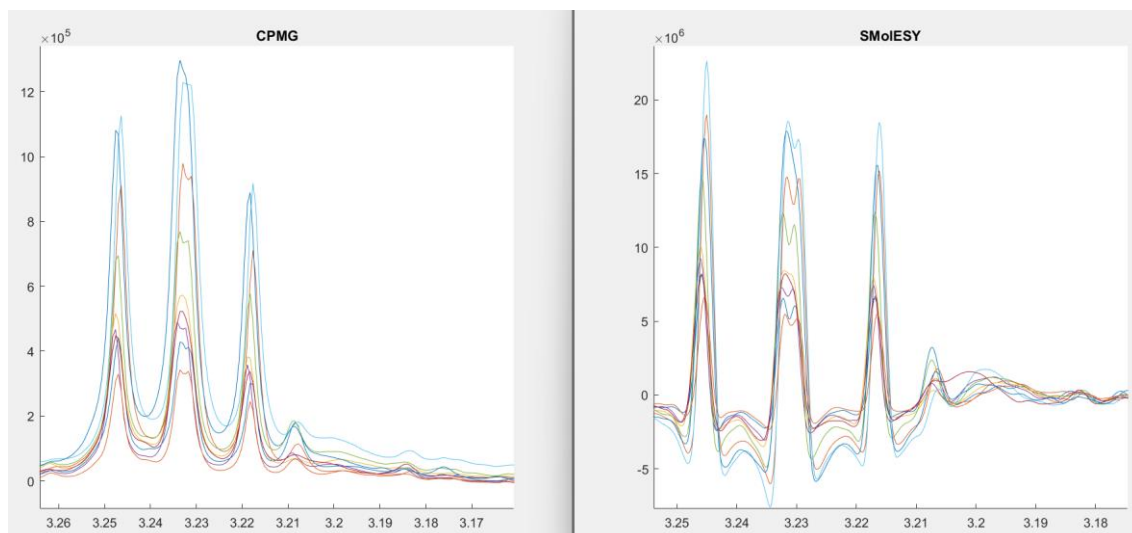


Figura 50. Franja 3.26-3.21ppm: Fenilalanina, TMAO, Taurina, Histidina, B-glucosa, Arginina, Colina

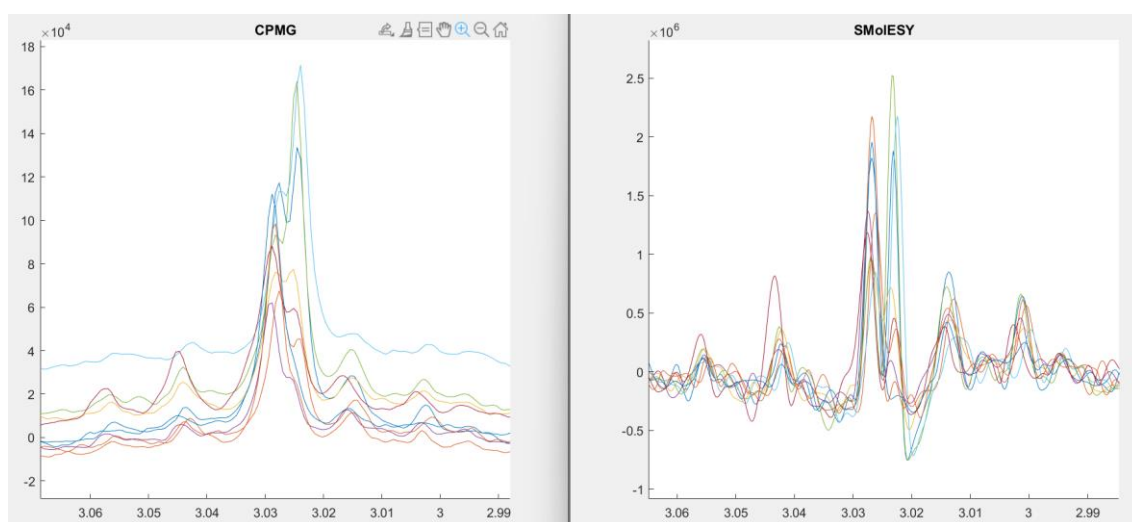


Figura 51. Tirosina (a banda i banda, des del 3.04ppm)

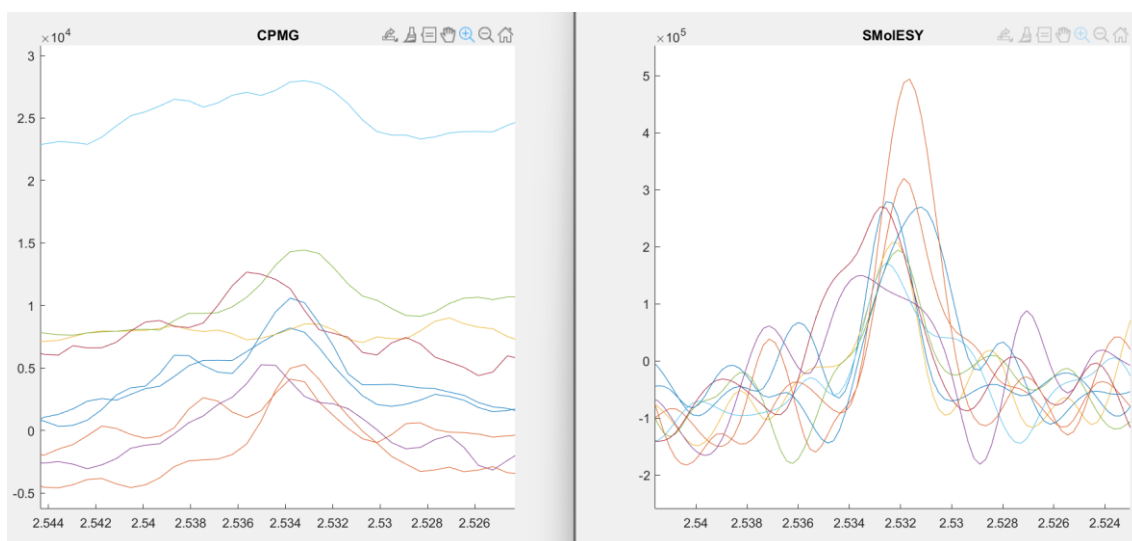


Figura 52. Metilamina (2.54ppm)

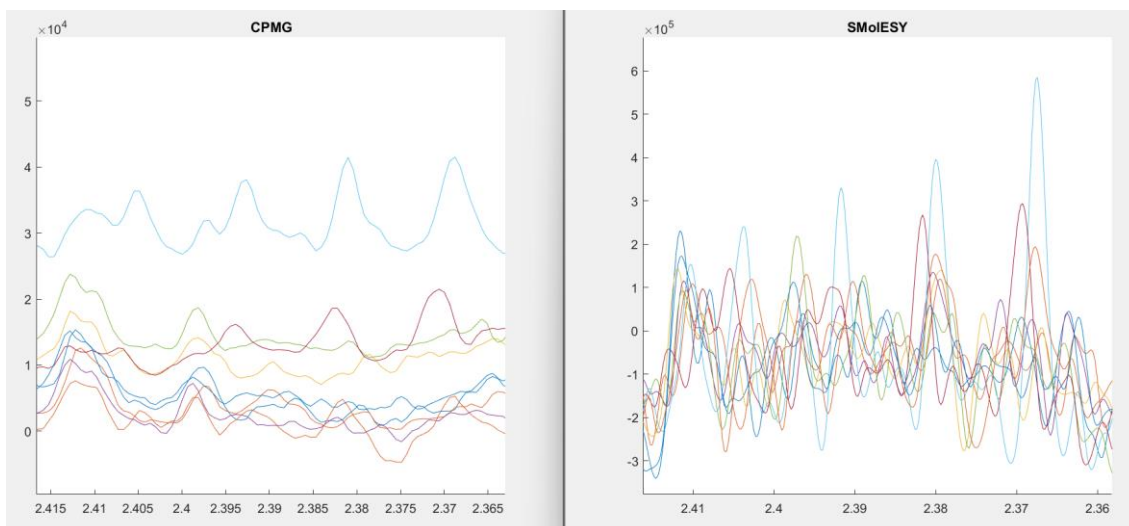


Figura 53. 3-hidroxibutirat, prolina, piruvat (3.38-3.36ppm)

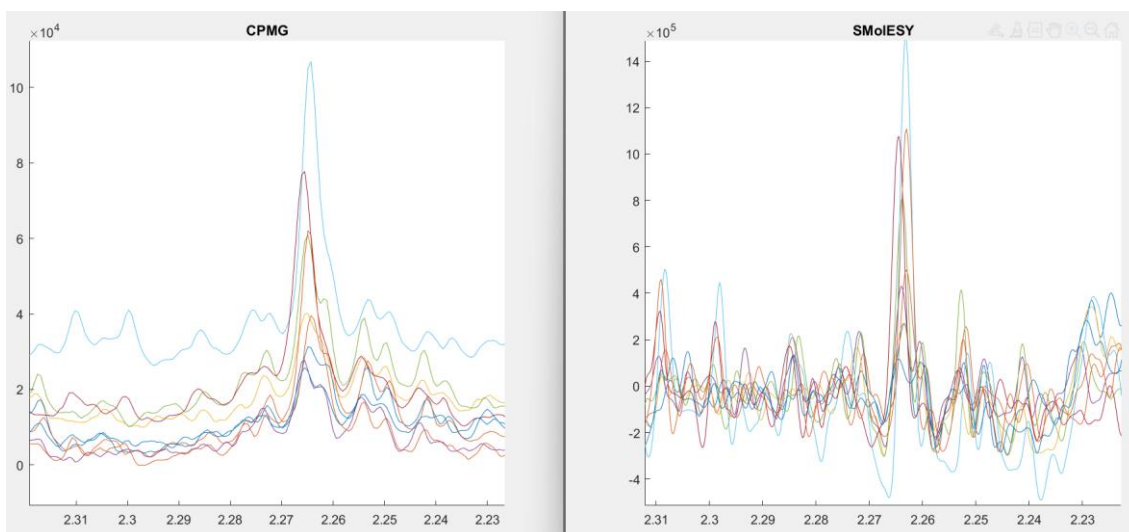


Figura 54. Valina (conjunt doblets) + lípid (2.24-2.23)

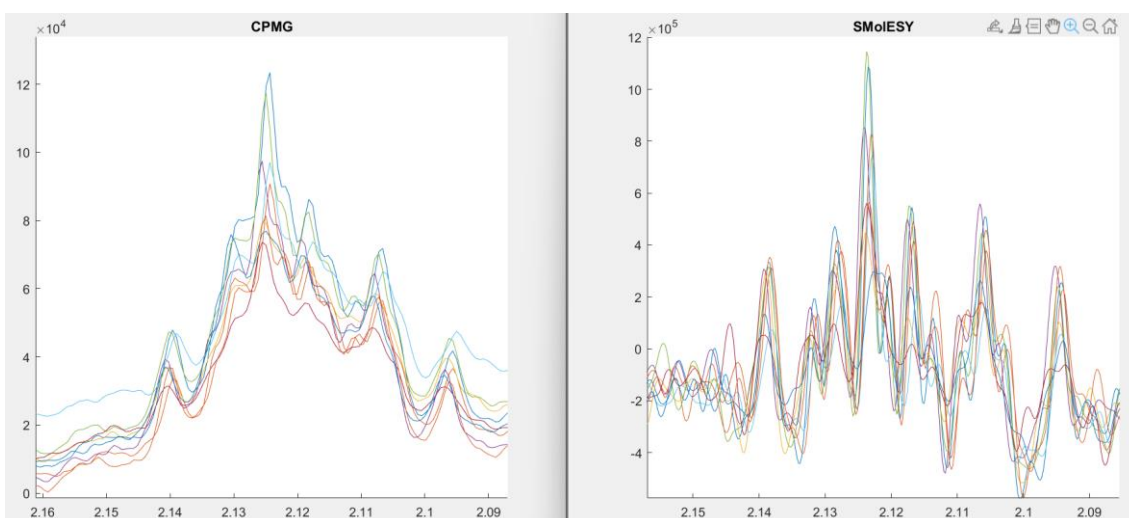


Figura 55. Glutamat, methionine (2.14-2.13ppm)

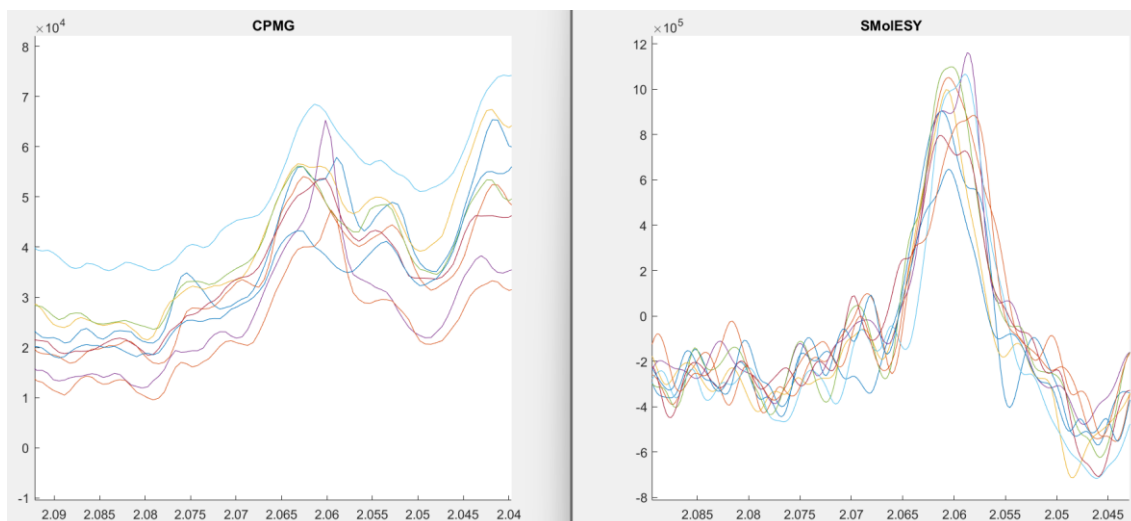


Figura 56. Glutamina (2.09-2.08ppm)

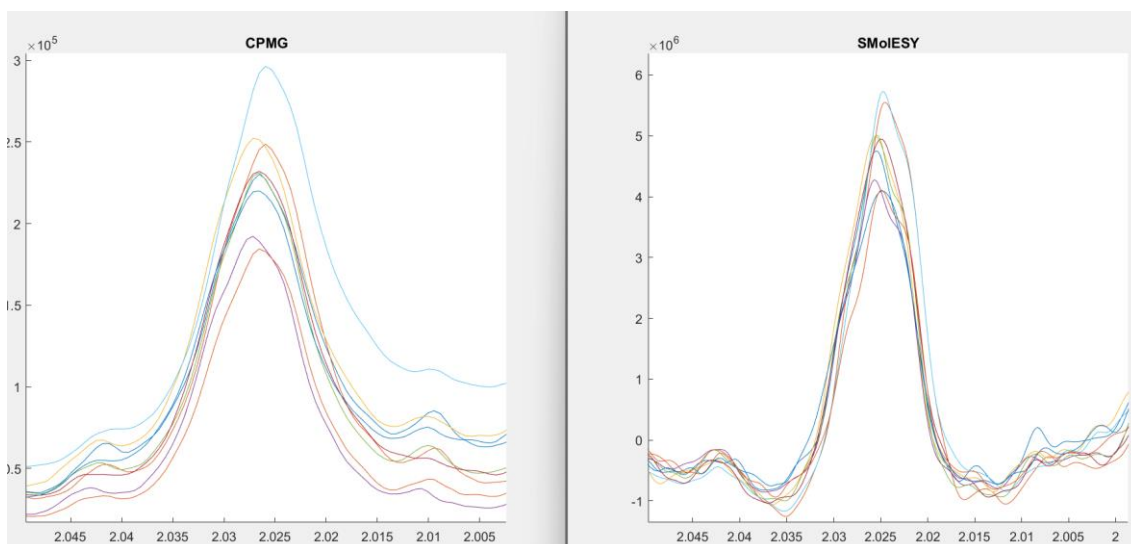


Figura 57. Glicoproteina (2.04ppm)

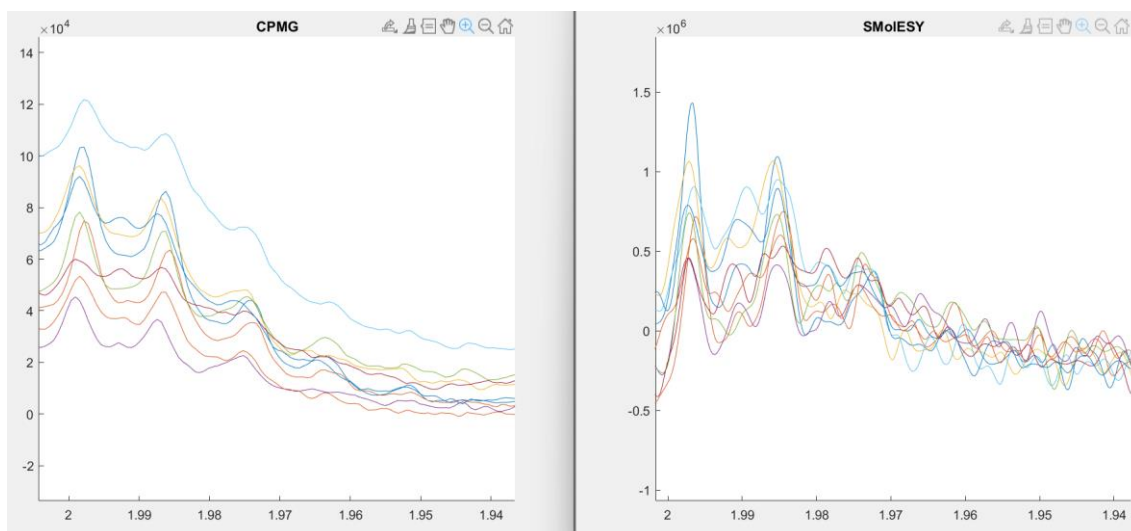


Figura 58. Glutamat, prolina, isoleucina + lipids (2.00-1.96ppm)

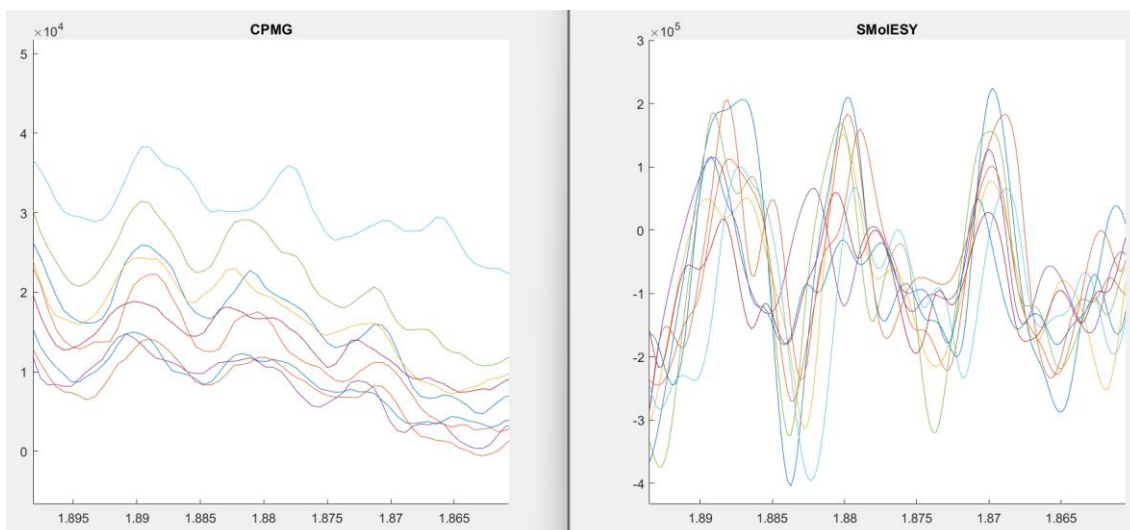


Figura 59. Arginina, lisina (1.91ppm)

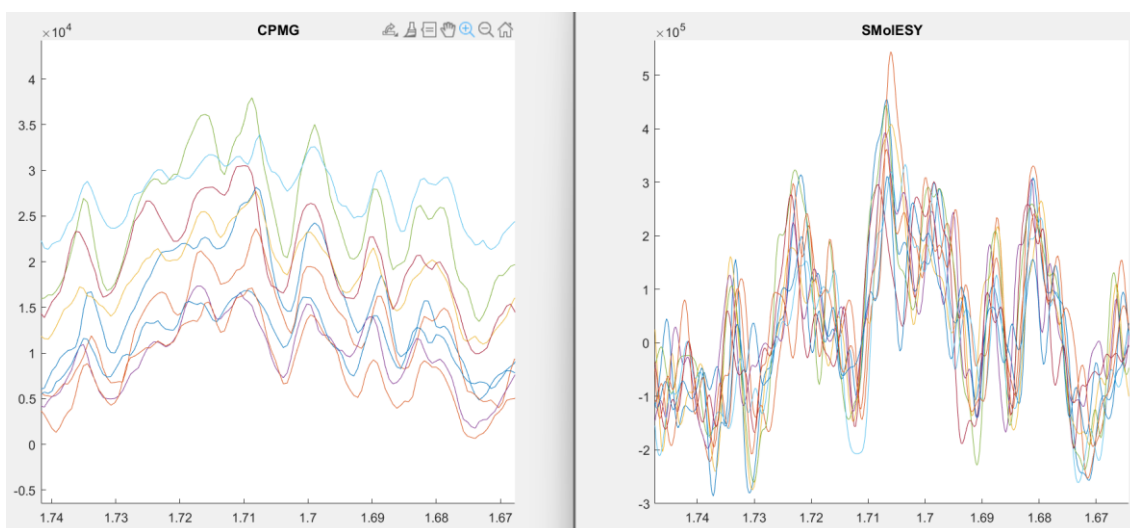


Figura 60. Leucina, lípid, lisina, arginina (1.71-1.68ppm)

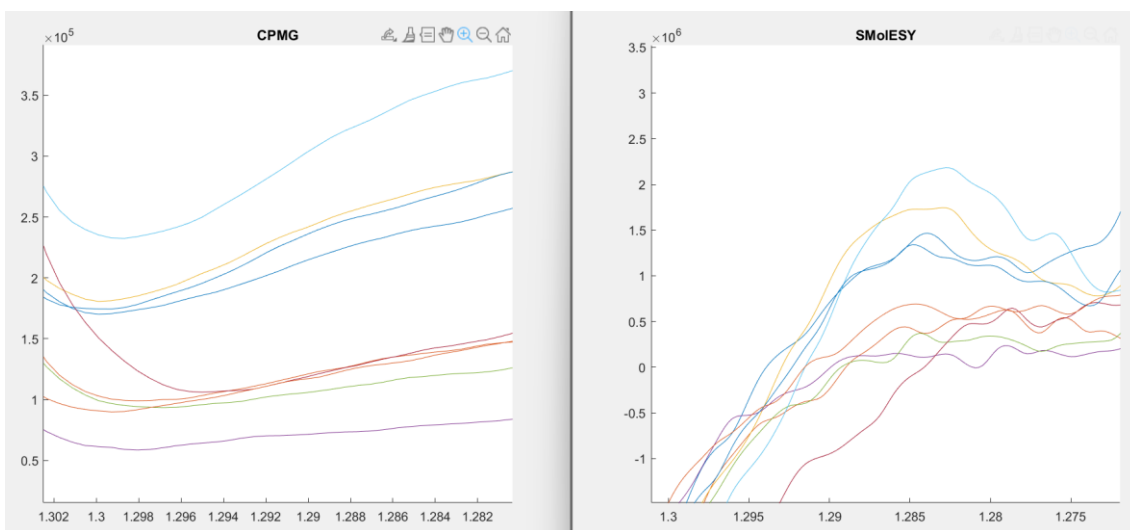


Figura 61. Lípids, treonina, fucose, isoleucina (1.32-1.28ppm)

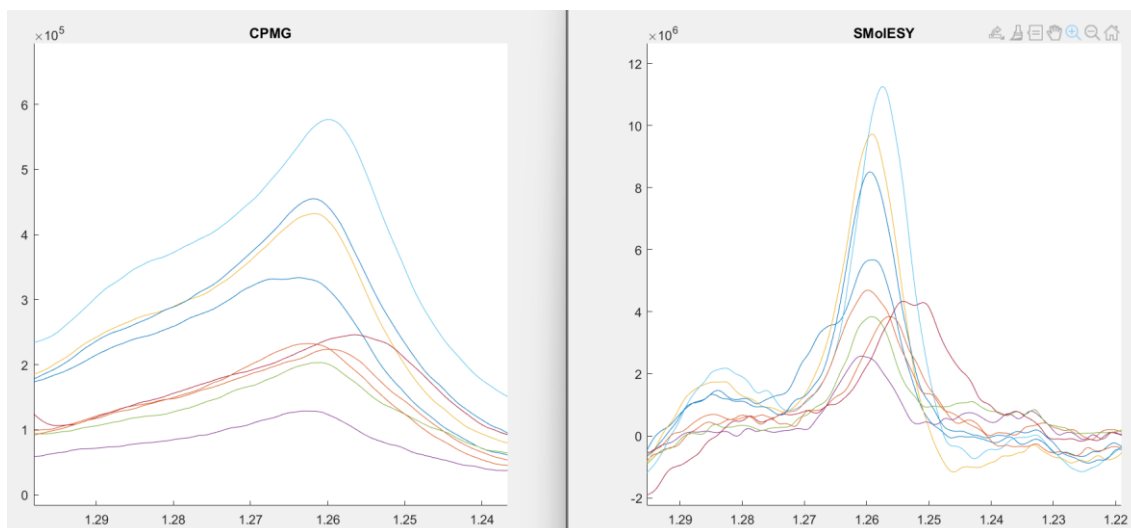


Figura 62. Lípids (LDL) (1.26-1.25ppm)

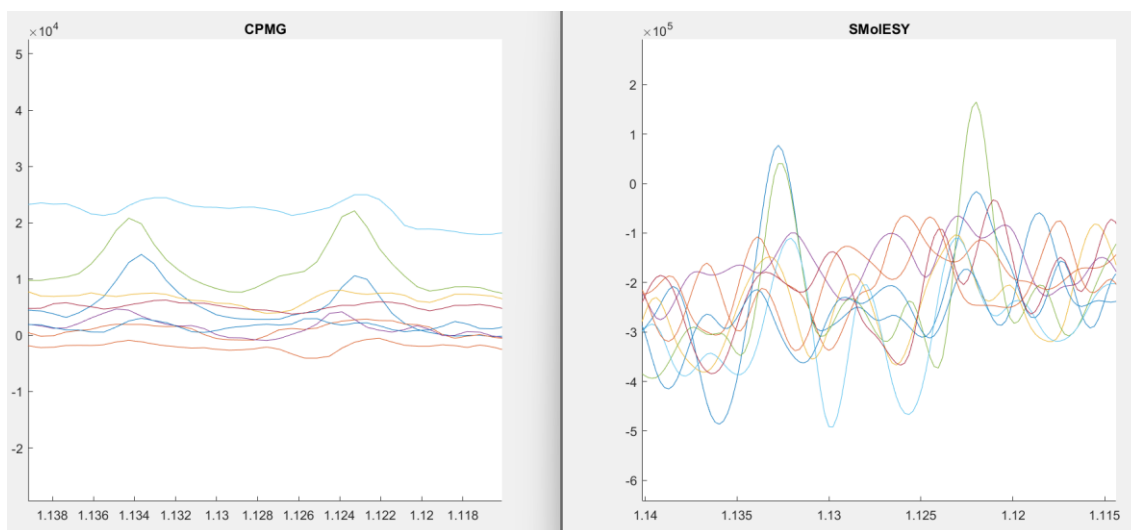


Figura 63. Isobutirat (1.13ppm)

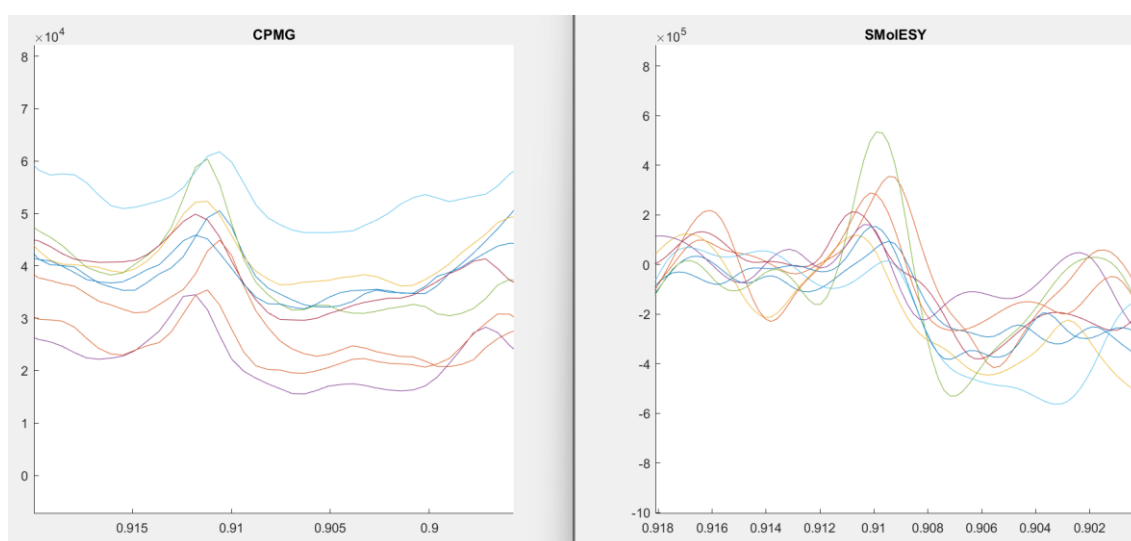


Figura 64. Colesterol (0.91ppm)

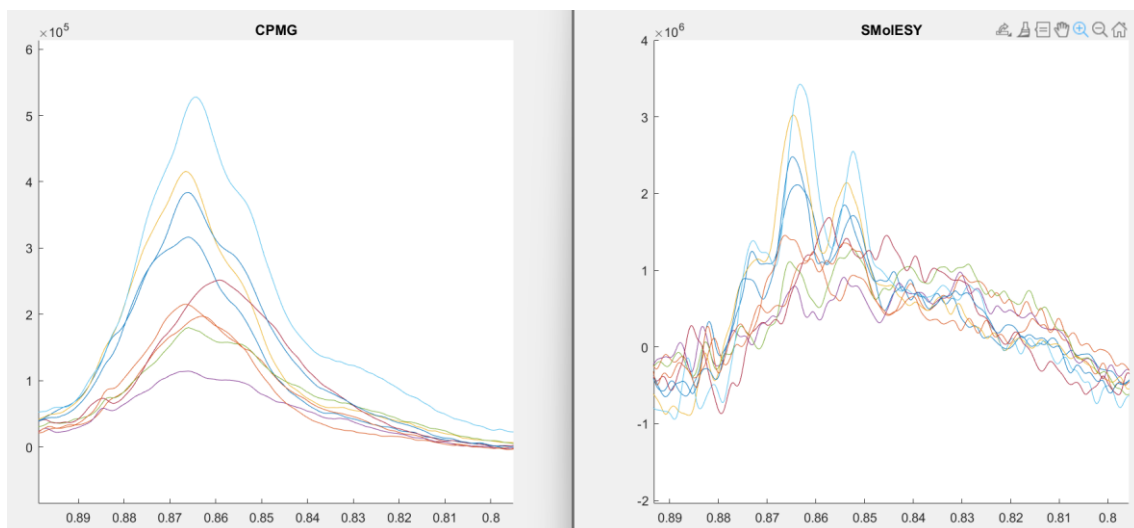


Figura 65. VLDL, LDL, colesterol