



## Capítulo 2.

### **Análisis abductivo de las interacciones proteómicas y metabolómicas en un modelo de lactocito bovino**

*Camilo Calle*

*Zoot., MSc, Profesor de cátedra, Facultad de Ciencias Agrarias  
Universidad Nacional de Colombia, sede Medellín, Grupo de  
investigación en Ingeniería Agrícola*

*Tatiana Ruiz-Cortés*

*MV, MSc, PhD. Profesora titular, Facultad de Ciencias Agrarias,  
Universidad de Antioquia, Grupo de investigación Biogénesis*

### **Resumen**

Estudiar la regulación e interacciones de vías metabólicas de aminoácidos (AA), carbohidratos (CHO), lípidos y energía (E) es importante para comprender los mecanismos de síntesis de proteínas de leche. Se realizó un análisis de tipo abductivo para aproximarse a una descripción de la posible relación metabólica del proteoma en un modelo *in vitro* de lactocitos bovinos. Para estudiar las vías metabólicas se creó un archivo multi-fasta con todas las secuencias de

proteínas y se envió a BlastKOALA en la Kyoto Encyclopedia of Genes and Genomes (KEGG). Las vías de KEGG ( $n = 44$ ) se agruparon en metabolismo de AA, CHO, lípidos y E. Se creó luego una matriz de datos con las vías metabólicas y las proteínas identificadas. Se calculó una matriz de valores P de distancias chi-cuadrado. Para eliminar proteínas y vías metabólicas que no resultaron significativas ( $p > 0,05$ ) se utilizó el análisis iterativo de correspondencia múltiple y seleccionaron las variables significativas entre cada iteración. Finalmente, se realizó un análisis de clúster para detectar las posibles relaciones simultáneas entre las proteínas. Los grupos de rutas correspondientes al metabolismo de AA y de CHO representaron el mayor número de interacciones, 19 y 26 respectivamente. El algoritmo tomó cinco iteraciones hasta que el número de vías metabólicas no significativas fue igual a cero. El análisis multivariado y abductivo en nuestro modelo celular *in vitro* es una herramienta útil para estudiar y comprender mejor las interacciones metabólicas en la compleja vía de síntesis de proteínas de leche.

**Palabras clave:** *bioinformática, análisis multivariante*

## 1. Introducción

La glándula mamaria bovina madura está formada por una red de conductos que terminan en los alvéolos. La capa más interna de los alvéolos está rodeada de células epiteliales mamarias (MEC) que son responsables de la síntesis y secreción de leche durante la lactancia. Las MEC son las principales células precursoras responsables del potencial de lactancia. El proteoma de las MEC define su estado funcional y sugiere su papel en diversas actividades celulares como la proliferación y la lactancia (Jaswal et al., 2020).

El análisis de las MEC, como unidad fundamental dentro de la lactancia, implica comprender los mecanismos celulares que autorregulan sus funciones. Además, es bien sabido que el cultivo de células mamarias proporciona una herramienta experimental útil y económica, en comparación con los ensayos *in vivo*, y permite probar un gran número de tratamientos en células epiteliales aisladas (Clark & Pateman, 1978).

Ahora bien, la proteómica como área de estudio plantea muchos retos desde el punto de vista metodológico, ya que la cantidad de datos disponibles muchas veces supera la capacidad de análisis de los investigadores. Las herramientas idóneas para probar este tipo de planteamientos científicos reposan en el campo de la bioinformática, la cual constituye una de las áreas del conocimiento más promisorias. En los últimos años, este campo ha hecho importantes aportes a la proteómica, la genómica y la metabolómica, entre otras (Boehmer et al., 2008), al brindar herramientas que permiten a los científicos abordar grandes cantidades de datos e información para finalmente proponer posibles hipótesis de lo que está sucediendo en el sistema.

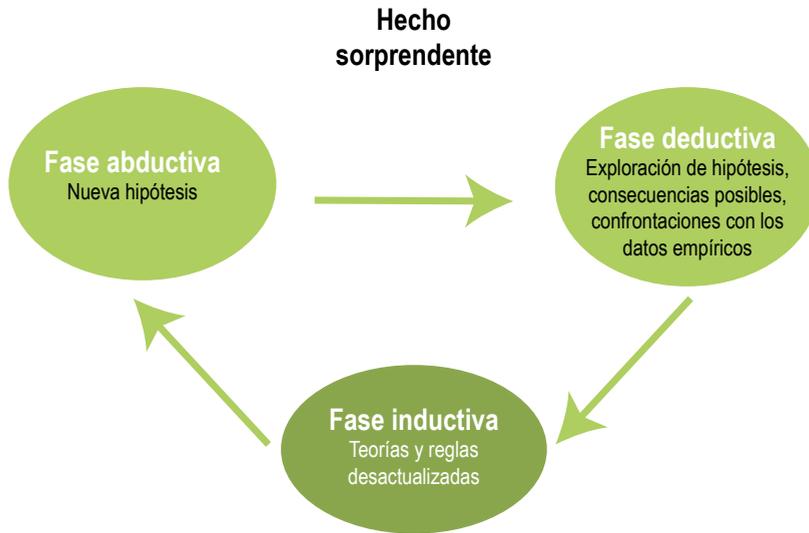
La reproducción bovina y las ciencias de la leche no son ajenas a esta tendencia científica (O'Donnell et al., 2004; Roncada et al., 2012) y, se ha publicado recientemente acerca de los perfiles metabolómicos en la glándula mamaria de yak durante el ciclo de lactancia y se reportaron metabolitos diferenciales involucrados en vías que incluyen aminoácidos, biosíntesis de otros metabolitos secundarios, carbohidratos, energía, lípidos, cofactores y vitaminas y metabolismo de nucleótidos relacionado con el desarrollo de la glándula mamaria, la lactogénesis y la lactancia (Gregorini et al., 2016). En 2020, investigadores perfilaron a profundidad el proteoma de los MEC de búfalo cultivados *in vitro* (Bu-MEC) durante la proliferación activa utilizando espectrometría de masas en tándem de alto rendimiento (Jaswal et al., 2020). Faltan estudios en



MEC sobre su estado metabólico inherente bajo desafíos de aminoácidos, como modelo *in vitro* para el suministro limitado de nutrientes, los cuales serían una pista para comprender mejor el metabolismo de la lactancia bovina y las estrategias metabólicas de MEC sometidas a estrés nutricional.

Los resultados previos del experimento (Ruiz-Cortes et al., 2022) arrojaron resultados que sugieren que las MEC sometidas a una restricción de aminoácidos esenciales (EAA) tienen la capacidad de mantener la producción de proteínas de la leche, como la caseína, lo que implicaría algún mecanismo de homeorresis en dicho proceso. En este sentido, cobra especial importancia identificar las proteínas y rutas metabólicas que se activaron en dichos momentos de restricción nutricional.

Para ello, es clave recordar que en el campo de la proteómica y la metabolómica es de suma importancia el uso combinado de herramientas bioinformáticas y del análisis abductivo para reducir el tiempo de investigación y permitir al investigador enfocarse en las vías clave implicadas, como ya ha sido expresado por autores como Rodríguez (2005), Le, Deeth y Larsen (2017) o Cárdenas (2019). Cárdenas (2019) plantea que existe un proceso (Figura 1) que involucra estos tipos de inferencia. Así, la deducción permite generar las consecuencias (C), la inducción, establecer reglas generales (A), y la abducción, construir la hipótesis (B). La mayoría de los razonamientos y, en particular, el razonamiento científico, combina estas tres formas de razonamiento (Moscoso, 2019). Desde este punto de vista, Moscoso propone entender todo el proceso abductivo como un ciclo recursivo de abducción / deducción / inducción. Las tres etapas del ciclo de abducción / deducción / inducción, que Moscoso (2019) llamó "ciclo adi", son relativamente autónomas pues cada una implica un procedimiento específico, tienen requisitos formales que deben cumplirse y pueden ejecutarse en diferentes momentos y proyectos de investigación, como el que se expone en este capítulo.



**Figura 1.** El ciclo de la abducción/deducción/inducción

Fuente: adaptado de Moscoso (2019)

En el estudio presentado en este capítulo se diseñó una metodología que permitió realizar un estudio abductivo previo a la investigación, lo cual le permitió al investigador identificar posibles rutas o proteínas que tenían una mayor probabilidad de interactuar en las distintas situaciones de estudio. Este estudio abductivo permitió identificar la hipótesis de investigación con las que se podrán proponer nuevas investigaciones. Se aplicó dicha metodología a un conjunto de datos de aproximadamente 1.183 proteínas obtenidos de un experimento realizado en lactocitos de glándula mamaria bovina.

## 2. Metodología

Se diseñó una herramienta para el análisis abductivo proteómico y de rutas metabólicas en las MEC bajo restricciones de aminoácidos esenciales (AAE) en rutas relacionadas con el metabolismo de aminoácidos.



Para detectar las rutas metabólicas y proteínas más relevantes en las MEC bajo los diferentes tratamientos, se diseñó el flujo de trabajo que se describe a continuación.

## *2.1. Cultivo celular y tratamientos*

Se usaron células primarias bovinas aisladas de glándulas mamarias de vacas lecheras Holstein chinas (Hu et al., 2009). Para confirmar que las células eran de origen bovino y epitelial se realizó la identificación de marcadores epiteliales mediante proteómica (es decir, citoesqueleto 8 (UniProt ID = F1MU12) y citoqueratina 18 (UniPro ID = F6S1QO)) (datos no mostrados) (Tyanova et al., 2016; Yoder et al., 2019). La capacidad de síntesis de caseína se confirmó en un experimento anterior (Ruiz-Cortes et al., 2017).

Se siguió el protocolo estandarizado anteriormente (Ruiz-Cortes et al., 2022). Las células se sembraron en placas de plástico de seis pocillos (Corning®, Corning, NY) y se cultivaron en medio de Dulbecco (DMEM / F-12; ThermoScientific® Gibco®, Waltham, MO). Se añadió glucosa a 3,15 g/l y bicarbonato de sodio a 1,18 g/l. El medio se complementó con 50 mg/ml de transferrina, 2 mg/ml de insulina, 1 mg/ml de prolactina, 25 mg/ml de progesterona, 100 mg/ml de hidrocortisona (Sigma®, St Louis, MO), 0,25 mg/ml de EGF, 1% glutamina y mezcla de antibiótico-antimicótico al 1% (Sigma®) y FBS al 10% (Thermo Scientific® Gibco®). Las células se cultivaron durante 6 días después de alcanzar la confluencia (90-100%).

Las células fueron sometidas a estrés mediante la implementación de un diseño experimental factorial 2x2x2 en la que el primer factor correspondió a la variación de las concentraciones de AAE en dos niveles (2% y 100%) de la concentración completa en DMEM; el segundo factor es el tiempo de restricción (R) con dos niveles (8 horas y 24 horas); el tercer

factor es el tiempo de realimentación (RF) (8 horas y 24 horas), los dos últimos factores se relacionaron entre sí por una restricción + tiempo de realimentación (R + RF) como se describió anteriormente (Ruiz-Cortes et al., 2017). Esto dio como resultado un total de siete tratamientos que se replicaron cuatro veces por duplicado: Control-Tratamiento 1 (T1): 100% EAA; Tratamiento 2 (T2): 2% de EAA durante 8 h R; Tratamiento 3 (T3): 2% de EAA durante 24 h R; Tratamiento 4 (T4): 2% EAA durante 8 h R + 100% EAA durante 8 h RF (R + RF); Tratamiento 5 (T5): 2% de EAA durante 8 h R + 100% de EAA durante 24 h RF (R + RF); Tratamiento 6 (T6): 2% EAA durante 24 h R + 100% EAA durante 8 h RF (R + RF); Tratamiento 7 (T7): 2% EAA durante 24 h R + 100% EAA durante 24 h RF (R + RF).

## 2.2. Inmunodetección de caseína

Las proteínas (40 µg) se separaron electroforéticamente mediante SDS-PAGE (12%) y se transfirieron a membranas de PVDF (Millipore®, Burlington, MA). Las transferencias se bloquearon durante 1 hora con tampón de bloqueo Odyssey (LI-COR Biosciences®, Lincoln, NE) y se incubaron durante la noche a 4°C en anticuerpos primarios de conejo y ratón (1: 1000). Los anticuerpos utilizaron cadenas alfa y beta de  $\alpha$ -tubulina (Cell Signaling Technology®) y caseína reconocidas (ABCAM®, Cambridge, Reino Unido). Después de 5 lavados de 5 min con PBS que contenía Tween 20 al 0,01% (Bio-Rad Life Science®, Hercules, CA), las transferencias se incubaron durante 1 h TA en anticuerpos secundarios de cabra anti-conejo y anti-ratón (1: 10000) ligado con tintes fluorescentes (IRDye 800 CW e IRDye 680, respectivamente; LI-COR Biosciences®). Las transferencias se lavaron cinco veces más como se describió anteriormente y se escanearon usando un sistema de imágenes infrarrojas Odyssey (LI-COR Biosciences®). La intensidad de la señal de las proteínas diana se cuantificó con el software de aplicación Odyssey (versión 3.0). Los resultados



se estandarizaron como una proporción de expresión de caseína con relación a  $\alpha$ -tubulina (Ruiz-Cortes et al., 2022).

### 2.3. *Proteómica cuantitativa*

Se realizó lisis celular para liberar la proteína total intracelular y se midió la concentración de esta mediante BCA (Pierce®). Los extractos de proteína total se sometieron a proteómica cuantitativa sin marcaje basado en la movilidad iónica mejorada. El protocolo incluye análisis y cuantificación de cromatografía líquida-integración de movilidad iónica-espectrometría de masas (LC-IMS-MS) (software ISOQuant®, Mainz, Alemania) (Distler et al., 2016; Kuharev et al., 2015). Las proteínas presentes en el 90% o más de las muestras fueron 1.183; estas proteínas se utilizaron para los análisis posteriores. Se utilizaron los resultados cuantificados del TOP3 ISOQuant®.

### 2.4. *Bioinformática*

Se utilizó la base de datos KEGG para identificar las rutas metabólicas con la herramienta BlastKOALA, en el grupo taxonómico “animales”, con la base de datos de la familia “eucariotes” (Kanehisa, Sato & Morishima, 2016). De las 1.183 secuencias de proteínas se seleccionaron las proteínas incluidas en vías metabólicas de AA, CHO, y E. Se analizaron otras vías de señalización metabólica (como las de PI3K-Akt, cGMP-PKG, MAPK, insulina, calcio y ErbB) que el mismo sistema KEGG arroja (llamadas ST) y que se relacionan con la síntesis de proteínas. El número de proteínas por vía metabólica varió según el tratamiento.

### 2.5. *Creación de multifasta*

Se realizó un *pipeline* usando librerías de BioPerl para automatizar la recuperación de las secuencias de proteínas de interés de la base de datos



UniProt, con el fin de generar el archivo multifasta necesario para los análisis posteriores.

## 2.6. Identificación de las rutas metabólicas

Luego de crear el archivo multifasta de las rutas metabólicas previamente seleccionadas se procedió a montar dicho archivo en la plataforma KEGG, de allí se identificaron y ubicaron las proteínas que estuvieran presentes en las rutas metabólicas analizadas.

## 2.7. Preparación de la matriz de datos

El análisis abductivo requiere analizar las posibles relaciones entre las rutas metabólicas responsables de la síntesis láctea. No existe una hipótesis definida *a priori* por el investigador en cuanto a dichas relaciones.

Con los resultados del KEGG se creó un matriz de datos en donde se disponen las rutas metabólicas en columnas, las proteínas identificadas dentro de estas rutas en filas y los valores de las proteínas, ausente o presente (0 o 1), constituyen los valores de la base de datos.

## 2.8. Exploración de datos

Con el fin de brindar al investigador una herramienta para la exploración de datos que le permita rutas, o grupos de rutas metabólicas, se diseñó un algoritmo que crea una matriz de valores de P de distancias de chi cuadrado, de acuerdo con el nivel de significancia que determine el investigador (para este caso 5%), con la cual se crea un pictograma en donde los valores iguales o inferiores al nivel seleccionado se muestran de color rojo. De esta forma, el investigador puede identificar las rutas metabólicas que tienen probabilidad de interactuar (ya que comparten como mínimo una misma proteína).

## 2.9. Análisis de los datos

Dado que la interacción entre proteínas no es un proceso lineal y actual en conjunto, y con el fin de detectar las proteínas y rutas metabólicas más significativas que en conjunto puedan interactuar, se diseñó un algoritmo que implementa un análisis de correspondencias múltiples (ACM). Este análisis detecta las relaciones simultáneas entre las rutas metabólicas mediante el cálculo del R2 de cada modalidad (valor de presente o ausente), de cada ruta metabólica y su respectivo valor de P con respecto a la primera dimensión o al primer componente que representa un grupo de rutas metabólicas que tienen algo en común (que para el presente caso fue la síntesis de proteína). El algoritmo elimina las rutas metabólicas con un valor de P superior a 0,05, y corre de nuevo el análisis de correspondencias múltiples. Este proceso se itera hasta que la cantidad de variables con un valor de P inferior a 0,05 es igual al número inicial de variables, es decir que se eliminan variables hasta que todas sean significativas.

De esta forma se detecta el conjunto de rutas metabólicas que están relacionadas entre ellas de forma significativa. Finalmente, se realiza un análisis de clúster en el cual se detectan las posibles relaciones simultáneas entre las proteínas.

## 2.10. Análisis estadístico

En cada tratamiento se realizó un análisis abductivo para proporcionar una aproximación de la posible relación metabólica del proteoma en nuestro modelo *in vitro* de MEC. Usando los resultados del software KEGG se creó una matriz de datos con los valores de presente o ausente de las proteínas. Se calculó una matriz de p-valores de distancias chi-cuadrado y se creó un pictograma para eliminar proteínas y vías metabólicas que no eran significativas ( $p > 0,05$ ). Se utilizó un análisis de

correspondencia múltiple (MCA) y se calculó el coeficiente de determinación ( $R^2$ ) de cada vía metabólica para determinar su relevancia en el análisis, eliminando las rutas que no fueron significativas para la primera dimensión o componente (síntesis proteica). Este procedimiento se realizó hasta que no hubo vías metabólicas que eliminar.

Las proteínas detectadas en la primera iteración se utilizaron para analizar los posibles grados de interacción entre grupos de proteínas. De esto se trata la técnica multivariante, cuyo principal propósito es agrupar objetos formando conglomerados (clústeres) de objetos con un alto grado de homogeneidad interna y heterogeneidad externa y que, además, en nuestro caso interactúan al hacer parte de una misma vía metabólica.

Luego se compara la igualdad de medias entre las proteínas de los clústeres por medio de un análisis ANOVA cuantitativo para detectar diferencias entre la expresión de dichas proteínas. Las proteínas significativas ( $p < 0,05$ ) se seleccionaron por homogeneidad en sus variaciones. Un estudio *post hoc* determinó la diferencia entre la expresión de proteínas significativa en comparación con el grupo de control (T1). Para los análisis estadísticos se utilizó el FactoMineR del programa R (RFoundation, 2017) y el programa IBM SPSS Statistics 25 (Armonk, NY).

### **3. Resultados y discusión: aplicación de la herramienta**

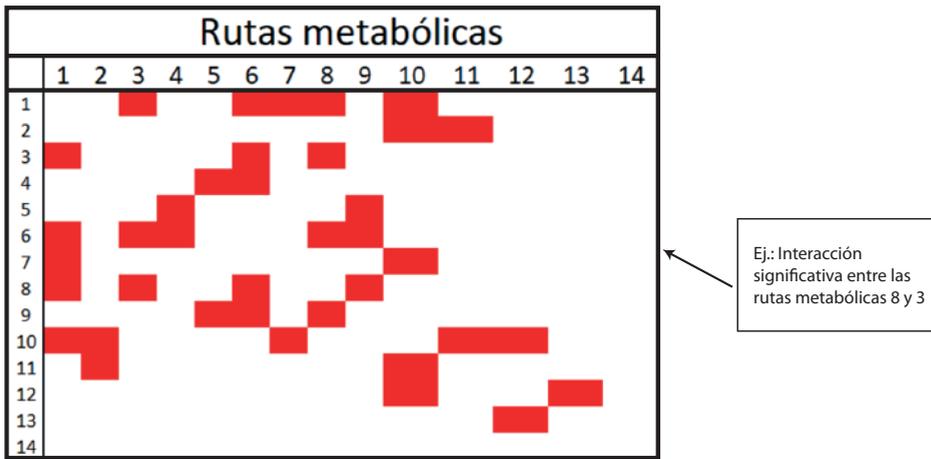
Los resultados que se muestran en este artículo provienen de muestras control como modelo para describir el procedimiento multivariado y su análisis.

Se seleccionaron 44 rutas metabólicas agrupadas en metabolismo de carbohidratos, metabolismo de lípidos, metabolismo de aminoácidos y



metabolismo energético, ya que el estado del arte de dicho fenómeno implica el estudio de todos estos grupos de rutas metabólicas.

El Pictograma 1 muestra una matriz de valores significativos de P ( $P < 0,05$ ) en la que se pueden evidenciar las relaciones entre las rutas metabólicas seleccionadas.



**Pictograma 1.** Interacción entre rutas metabólicas en MEC cultivadas *in vitro* durante 6 días después de la confluencia. Se agregaron aminoácidos esenciales (EAA) al 100% y al 2% (restricción) de concentraciones completas de DMEM durante 8 h y 24 h, así como realimentación con 100% EAA durante 8 h y 24 h (restricción + realimentación).

Fuente: elaboración propia

En la Tabla 1 se puede ver la cantidad de interacciones de las rutas metabólicas que se extrajeron inicialmente del KEGG. Los resultados muestran una gran cantidad de rutas metabólicas presentes y de interacciones que se dan entre estas. Esta situación evidencia la necesidad de reducir dicha complejidad con el fin de formular hipótesis más efectivas en los procesos de investigación.

**Tabla 1.** Proteínas y rutas metabólicas identificadas en el KEGG e interacciones

Ruta	Número de rutas con las que interactúan	Proteínas identificadas
Glycolysis / Gluconeogenesis	9	22
Citrate cycle TCA cycle	3	14
Pentose phosphate pathway	4	12
Pentose and glucuronate interconversions	3	6
Fructose and mannose metabolism	2	9
Galactose metabolism	5	9
Ascorbate and aldarate metabolism	9	3
Starch and sucrose metabolism	4	6
Amino sugar and nucleotide sugar metabolism	3	14
Pyruvate metabolism	9	15
Glyoxylate and dicarboxylate metabolism	3	6
Propanoate metabolism	7	8
Butanoate metabolism	7	5
Inositol phosphate metabolism	0	3
Oxidative phosphorylation	1	21
Carbon fixation in photosynthetic organisms	5	11
Carbon fixation pathways in prokaryotes	2	6
Methane metabolism	2	9
Nitrogen metabolism	0	2
Sulfur metabolism	0	4
Fatty acid biosynthesis	0	1
Fatty acid elongation	7	6
Fatty acid degradation	10	11
Synthesis and degradation of ketone bodies	2	2
Steroid hormone biosynthesis	0	3
Glycerolipid metabolism	10	4
Glycerophospholipid metabolism	1	3
Ether lipid metabolism	1	3
Sphingolipid metabolism	0	3
Arachidonic acid metabolism	0	3
alpha-Linolenic acid metabolism	1	1

Ruta	Número de rutas con las que interactúan	Proteínas identificadas
Biosynthesis of unsaturated fatty acids	4	4
Alanine, aspartate and glutamate metabolism	4	9
Glycine, serine and threonine metabolism	2	6
Cysteine and methionine metabolism	5	12
Valine, leucine and isoleucine degradation	11	16
Lysine degradation	9	7
Arginine biosynthesis	5	6
Arginine and proline metabolism	9	10
Histidine metabolism	9	2
Tyrosine metabolism	7	4
Phenylalanine metabolism	7	3
Tryptophan metabolism	11	4
Phenylalanine, tyrosine and tryptophan biosynthesis	7	2

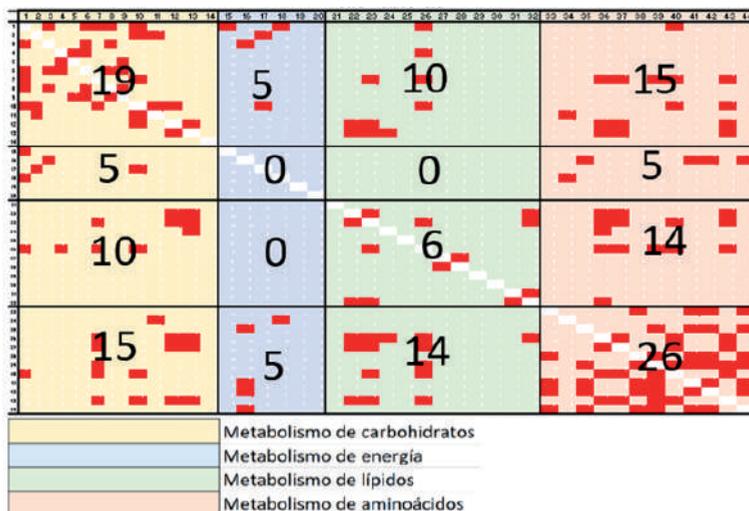
Fuente: elaboración propia

### *3.1. Cálculo e interpretación del pictograma de interacciones*

En el Pictograma 2 se puede ver cómo los grupos de rutas correspondiente al metabolismo de aminoácidos y el metabolismo de los carbohidratos representan la mayor cantidad de interacciones, 19 y 26 respectivamente. Pese a ello, la mayor cantidad de rutas no poseen ningún grado de interacción con las demás. Esta situación genera una dificultad desde el punto de vista estadístico, ya que involucra ruido en los análisis conjuntos. Por otro lado, la eliminación de proteínas y rutas metabólicas de forma subjetiva puede introducir un sesgo al análisis.

### *3.2. Determinación de las proteínas y rutas metabólicas clave en el estudio*

Bajo el supuesto de que existe una combinación lineal que representa las relaciones entre las rutas metabólicas, basado en las proteínas que comparten, se diseñó un algoritmo que itera análisis de corresponden-



**Pictograma 2.** Matriz de interacciones individuales entre las rutas metabólicas analizadas

Fuente: elaboración propia

cias múltiples y selecciona las variables significativas entre cada iteración, eliminando las rutas que no son significativas ( $P < 0,05$ ) con la primera dimensión o componente que fue síntesis de proteínas.

Para el presente estudio, el algoritmo tardó cinco iteraciones hasta que la cantidad de rutas metabólicas no significativas fuese igual a cero.

En la Tabla 2 se puede ver la relación de las rutas metabólicas (R2) con el primer componente, se puede ver cómo el metabolismo del triptófano es la ruta que más explica el fenómeno (70%), mientras que la biosíntesis de arginina fue la que menos explicó el primer componente (20%).

Aunque el objetivo de este artículo no es el análisis de los resultados obtenidos utilizando la herramienta que se describe, es importante resaltar a manera de ejemplo, cómo esta permite llegar a resultados ya discutidos por otros autores (Ruiz-Cortes et al., 2022). Uno de los ha-

**Tabla 2.** Relaciones entre las rutas metabólicas

COD	RUTA	R2	Valor de P	GRUPO	PROTEÍNAS IDENTIFICADAS
rm43	Tryptophan metabolism	0,7068472	0,00000	metabolismo de aminoácidos	4
rm40	Histidine metabolism	0,6962692	0,00000	metabolismo de aminoácidos	2
rm7	Ascorbate and aldarate metabolism	0,5130543	0,00000	metabolismo carbohidratos	3
rm37	Lysine degradation	0,4504774	0,00001	metabolismo de aminoácidos	7
rm26	Glycerolipid metabolism	0,3893448	0,00006	metabolismo de lípidos	4
rm23	Fatty acid degradation	0,3638985	0,00013	metabolismo de lípidos	11
rm33	Alanine, aspartate and glutamate metabolism	0,2887553	0,00087	metabolismo de aminoácidos	9
rm36	Valine, leucine and isoleucine degradation	0,2530229	0,00207	metabolismo de aminoácidos	16
rm38	Arginine biosynthesis	0,206432	0,00611	metabolismo de aminoácidos	6

Fuente: elaboración propia

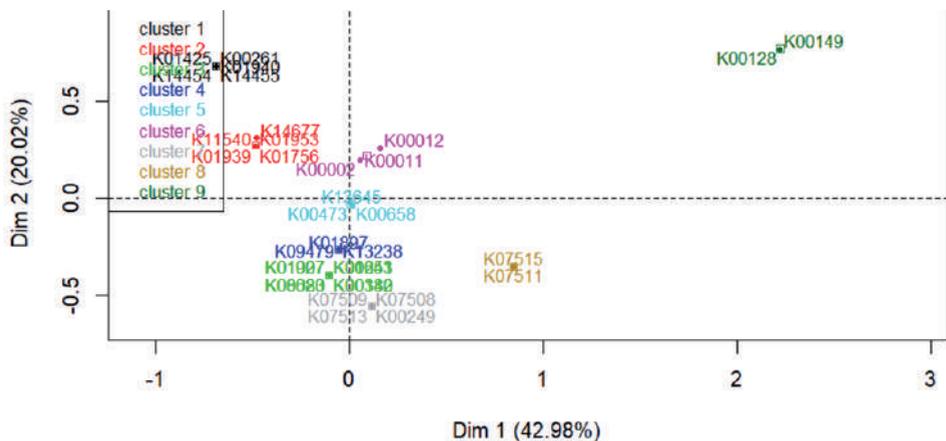
llazgos del presente estudio abductivo, ya reportado previamente, fue que el metabolismo del triptófano es una de las rutas metabólicas más significativas ya que este aminoácido es el precursor de la serotonina. La serotonina es un aminoácido modificado: el L-triptófano que se convierte en 5-hidroxitriptófano (5-HTP) utilizando la enzima triptófano hidroxilasa (TPH) que limita la velocidad de síntesis de la serotonina.

Existen 2 enfermedades potencialmente devastadoras para las vacas que se encuentran lactando: la cetosis y la hipocalcemia. Una terapia común podría representar una solución poderosa dentro de la industria láctea. De hecho, la serotonina tiene el potencial de servir como una herramienta esencial en la prevención de enfermedades de las vacas lecheras relacionadas con la transición. Es un potente regulador tanto

de la homeóstasis del calcio como de la homeóstasis energética durante la lactancia en modelos de roedores y en vacas lecheras. En 2014 se reportó que la serotonina es necesaria para la estructura y función adecuadas de la glándula mamaria, pues regula el transporte de calcio de la sangre y las células epiteliales mamarias durante la lactancia (Laporta et al., 2014). En otro estudio se determinó que las infusiones preparto de 5-hidroxitriptófano (5-HTP) podrían modular la homeostasis energética a nivel del hepatocito en vacas lecheras Holstein y Jersey posparto al estimular una adaptación autocrino-paracrina a la lactancia en el hígado de la vaca (Weaver et al., 2017).

El potencial de esta herramienta en el estudio y diseño de proyectos de investigación radica en que permite identificar hipótesis certeras y eficaces. Para este ejemplo se podría plantear la siguiente hipótesis: el metabolismo del triptófano regula los estados de estrés por diferentes vías. En la Gráfica 1 se puede ver cómo se crearon nueve grupos de proteínas, resultado que le puede dar mejores indicios al investigador sobre las proteínas hacia las que se puede dirigir la investigación.

**Gráfica 1.** Análisis de clúster de las proteínas seleccionadas.



Fuente: elaboración propia

Los resultados de la Tabla 3 muestran las posibles interacciones que se pueden dar entre las proteínas detectadas en el experimento.

**Tabla 3.** Posible interacción entre proteínas

PROTEÍNA	Ascorbate and aldarate metabolism	Fatty acid degradation	Glycerolipid metabolism	Alanine, aspartate and glutamate metabolism	Valine, leucine and isoleucine degradation	Lysine degradation	Arginine biosynthesis	Histidine metabolism	Tryptophan metabolism	CLUSTER
K00261	0	0	0	1	0	0	1	0	0	1
K01425	0	0	0	1	0	0	1	0	0	1
K01940	0	0	0	1	0	0	1	0	0	1
K14454	0	0	0	1	0	0	1	0	0	1
K14455	0	0	0	1	0	0	1	0	0	1
K01756	0	0	0	1	0	0	0	0	0	2
K01939	0	0	0	1	0	0	0	0	0	2
K01953	0	0	0	1	0	0	0	0	0	2
K11540	0	0	0	1	0	0	0	0	0	2
K14677	0	0	0	0	0	0	1	0	0	2
K00020	0	0	0	0	1	0	0	0	0	3
K00140	0	0	0	0	1	0	0	0	0	3
K00253	0	0	0	0	1	0	0	0	0	3
K00382	0	0	0	0	1	0	0	0	0	3
K01027	0	0	0	0	1	0	0	0	0	3
K01641	0	0	0	0	1	0	0	0	0	3
K01907	0	0	0	0	1	0	0	0	0	3
K08683	0	0	0	0	1	0	0	0	0	3
K01897	0	1	0	0	0	0	0	0	0	4
K09479	0	1	0	0	0	0	0	0	0	4
K13238	0	1	0	0	0	0	0	0	0	4
K00473	0	0	0	0	0	1	0	0	0	5

PROTEÍNA	Ascorbate and aldarate metabolism	Fatty acid degradation	Glycerolipid metabolism	Alanine, aspartate and glutamate metabolism	Valine, leucine and isoleucine degradation	Lysine degradation	Arginine biosynthesis	Histidine metabolism	Tryptophan metabolism	CLÚSTER
K00658	0	0	0	0	0	1	0	0	0	5
K13645	0	0	0	0	0	1	0	0	0	5
K00002	0	0	1	0	0	0	0	0	0	6
K00011	0	0	1	0	0	0	0	0	0	6
K00012	1	0	0	0	0	0	0	0	0	6
K00249	0	1	0	0	1	0	0	0	0	7
K07508	0	1	0	0	1	0	0	0	0	7
K07509	0	1	0	0	1	0	0	0	0	7
K07513	0	1	0	0	1	0	0	0	0	7
K07511	0	1	0	0	1	1	0	0	1	8
K07515	0	1	0	0	1	1	0	0	1	8
K00128	1	1	1	0	1	1	0	1	1	9
K00149	1	1	1	0	1	1	0	1	1	9

Fuente: elaboración propia

### 3.3. Resultados del análisis de varianza

Al realizar el análisis de ACM y de niveles de caseína y estudiar la posible interacción entre proteínas se hallaron 32 proteínas pertenecientes a ocho vías metabólicas; se desarrolló una base de datos en la que se incluyeron los resultados de cuantificación de cada proteína para cada tratamiento. A partir de estos resultados se realizó el ANAVA y el análisis *post hoc*, con lo que se identificaron 20 proteínas con expresión significativa y solo 13 diferentes ( $P < 0,05$ ) con respecto al control (Tabla 4).

**Tabla 4.** Variación de la expresión de proteína (aumenta-disminuye, flechas) con diferencia significativa ( $p \leq 0,05$ ) comparado con el grupo control (Control-T1: 100% AAE)

Cód. KEGG	Proteína <sup>1</sup>	Tratamientos					
		T2	T3	T4	T5	T6	T7
		2% EAA 8 h R <sup>2</sup>	2% EAA 24 h R	2% AAE 8 h R+ 100% AAE 8 h RA <sup>3</sup>	2% AAE 8 h R+ 100% AAE 24 h RA	2% AAE 24 h R+ 100% AAE 8 h RA	2% AAE 24 h R + 100% AAE 24 h RA
K00016	LDHA	↓	↓	–	–	↓	–
K00026	MDH2	↑	↑	–	↑	↑	↑
K00789	MAT2A	–	–	–	–	–	↑
K00802	SMS	↓	–	–	–	–	–
K01425	GLS	–	–	–	–	↓	–
K01940	ASS1	↓	↓	↓	↓	↓	↓
K01953	ASNS	–	↑	–	–	↑	–
K11540	CAD	–	↓	–	↓	↓	–
K14455	GOT2	↓	–	–	↓	↓	–
K14677	ACY1	↓	↓	–	–	↓	–
K14085	ALDH7A1	–	↓	–	–	–	–
K00121	ADH5	↓	↓	–	–	↓	–
K01758	CTH	–	–	–	↓	–	–

Fuente: elaboración propia

1 L-lactate dehydrogenase (LDHA); mitochondrial malate dehydrogenase (MDH2); S-adenosylmethionine synthetase (MAT2A); Spermine synthase (SMS); Glutaminase (GLS); Argininosuccinate synthase (ASS1); Asparagine synthase (ASNS); Carbamoyl-phosphate synthase (CAD); Mitochondrial aspartate aminotransferase (GOT2); Aminoacylase (ACY1); Aldehyde dehydrogenase family 7 member A1 (ALDH7A1); S-(hydroxymethyl) glutathione dehydrogenase (ADH5); Cystathionine gamma-lyase (CTH).

2 Restricción (R).

3 Restricción + re-alimentación (R+RA)



Los resultados que se muestran en este artículo provienen de muestras control como se indicó anteriormente para describir el procedimiento multivariado y su análisis. Con cada tratamiento se siguieron las mismas etapas, lo que permitió finalmente comparar y estudiar qué proteínas tuvieron una expresión diferencial frente al grupo control (sin restricción de EAA) y en qué vías metabólicas estuvieron implicadas (López-Diez et al., 2021).

Se han desarrollado muchos repositorios de datos y recursos disponibles públicamente para respaldar la gestión de información relacionada con las proteínas, la generación de hipótesis impulsada por datos y el descubrimiento de conocimientos biológicos. Existen revisiones completas (con categorización y descripción) de las principales bases de datos de bioinformática de proteínas para ayudar a los investigadores a encontrar rápidamente los recursos informáticos apropiados relacionados con las proteínas. Estas revisiones incluyen la discusión de los desafíos y oportunidades para desarrollar bases de datos y recursos bioinformáticos de proteínas de próxima generación para respaldar la integración de datos y el análisis de datos en la era de Big Data (Chen et al., 2020; Chen et al., 2017).

Nuestra metodología está acorde a recientes publicaciones que afirman que los avances recientes en la proteómica basada en espectrometría de masas han permitido un enorme progreso en la comprensión de los mecanismos celulares, la progresión de la enfermedad y la relación entre genotipo y fenotipo (Cifani & Kentsis, 2017).

Aunque muchos métodos bioinformáticos populares en proteómica se derivan de otros estudios ómicos, se requieren estrategias de análisis novedosas para tratar las características únicas de los datos proteómicos (Chen et al., 2020). Consideramos que la nuestra hace parte de estas novedosas técnicas de análisis, pues se usaron herramientas bioinformáticas diseñadas para la identificación y cuantificación de proteínas



basadas en espectrometría de masas y luego se apropiaron los diferentes métodos estadísticos y de aprendizaje automático que se han desarrollado para realizar análisis integrales en estudios proteómicos y metabolómicos.

Se utilizaron los datos cuantitativos de proteínas para reconstruir interacciones de proteínas y redes de señalización activadas frente a retos nutricionales de células epiteliales de glándula mamaria bovina, como ha sido propuesto recientemente por otros autores (Chen et al., 2020).

## 4. Conclusiones

Nuestro estudio de MEC *in vitro* utilizó una herramienta de análisis multivariado que integró la bioinformática, la proteómica y la metabolómica, con lo que se logró utilizar datos cuantitativos de proteínas para contribuir a una mejor comprensión fisiológica del metabolismo de MEC bovino *in vitro* (López-Diez, et al., 2021).

Mostramos cómo a partir de una gran cantidad de datos se puede abducir y enfocar el análisis hacia los puntos de interés, así como contribuir con la generación de nuevas hipótesis (Cárdenas, 2019). Para el caso específico de estudio, se aporta información en torno a las posibles vías alternas que usan las MEC para garantizar/mantener la producción de proteínas de leche frente a situaciones de estrés nutricional como parte de un interesante proceso metabólico homeorrético. Esto podría ayudar a dilucidar en el futuro estrategias metabólicas reguladas para la nutrición de vacas lactantes.

Un método abductivo eficiente y previo a la investigación puede ayudar a dirigir mejor los objetivos, especialmente en el campo de la proteómica, ya que se cuenta con una gran cantidad de datos que pueden abrumar al investigador o evitar que avance. En este sentido, la bioinformática puede aportar importantes herramientas para el acercamiento a los datos de proteómica y de otros campos.

## Referencias bibliográficas

- Boehmer, J. L., Bannerman, D. D., Shefcheck, K. & Ward, J. L. (2008). Proteomic analysis of differentially expressed proteins in bovine milk during experimentally induced *Escherichia coli* mastitis. *J Dairy Sci*, 91. doi:10.3168/jds.2008-1297
- Cárdenas, J. (2019). El método abductivo de investigación científica. sus orígenes en el romanticismo oscuro estadounidense y algunas reflexiones y ejemplos en torno a contextos multiculturales y la enseñanza de la música en la desglobalización. *Neuma (Talca)*, 12, 60-75. doi:10.4067/S0719-53892019000100060
- Cifani, P. & Kentsis, A. (2017). Towards comprehensive and quantitative proteomics for diagnosis and therapy of human disease. *Proteomics*, 17(1-2). doi:10.1002/pmic.201600079
- Clark, J. M. & Pateman, J. A. (1978). Isolation method affects transformed cell line karyotype. *Nature*, 272(5650), 262-264. doi:10.1038/272262a0
- Chen, C., Hou, J., Tanner, J. J. & Cheng, J. (2020). Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *Int J Mol Sci*, 21(8). doi:10.3390/ijms21082873
- Chen, C., Huang, H. & Wu, C. H. (2017). Protein Bioinformatics Databases and Resources. *Methods Mol Biol*, 1558, 3-39. doi:10.1007/978-1-4939-6783-4\_1
- David, A. (1999). *Logique, épistémologie et méthodologie en sciences de gestion*. Paper presented at the Conférence de l'AIMS, Ecole des Mines de Paris (CGS).
- Distler, U., Kuharev, J., Navarro, P. & Tenzer, S. (2016). Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. *Nature Protocols*, 11(4), 795-812. doi:10.1038/nprot.2016.042
- Gregorini, P., Beukes, P. C., Dalley, D. & Romera, A. J. (2016). Screening for diets that reduce urinary nitrogen excretion and methane emissions while maintaining or increasing production by dairy cows. *Sci Total Environ*, 551-552, 32-41. doi:10.1016/j.scitotenv.2016.01.203
- Jaswal, S., Anand, V., Kumar, S., Bathla, S., Dang, A. K., Kaushik, J. K. & Mohanty, A. K. (2020). In-depth proteome analysis of more than 12,500 proteins in buffalo mammary epithelial cell line identifies protein signatures for active proliferation and lactation. *Sci Rep*, 10(1), 4834. doi:10.1038/s41598-020-61521-1
- Kuharev, J., Navarro, P., Distler, U., Jahn, O. & Tenzer, S. (2015). In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics*, 15(18), 3140-3151. doi:10.1002/pmic.201400396



- Laporta, J., Keil, K. P., Vezina, C. M. & Hernandez, L. L. (2014). Peripheral serotonin regulates maternal calcium trafficking in mammary epithelial cells during lactation in mice. *PloS one*, 9(10), e110190. doi:10.1371/journal.pone.0110190
- López-Diez, L., Calle-Velásquez, C., Hanigan, M. D. & Ruiz-Cortes, Z. T. (2021). Aminoacids metabolomics profiles in bovine mammary epithelial cells under essential amino acids restriction. *Animals*, 11, 1334. <https://doi.org/10.3390/ani11051334>.
- Moscoso, J. N. (2019). Abductive reasoning: a contribution to knowledge creation in education. *Cadernos de Pesquisa*, 49, 308-329.
- RFoundation, S. c. (2017). R: A language and environment for statistical computing, version 3.4.2. <https://www.r-project.org/>
- Rodríguez, R. J. R. (2005). Abducción en el contexto del descubrimiento científico. *Rev. Filosofía Univ. Costa Rica*, XLIII Número doble (109/110), 87-97.
- Ruiz-Cortes, Z. T., Yoder, P. S. & Hanigan, M. D. (2017). *Essential amino acid deficiency in primary mammary epithelial bovine cells and effects on GCN2/IF2 expression and casein protein synthesis*. Paper presented at the Society for the Study of Reproduction 50th Annual Meeting Scientific Program, Washington, DC.
- Ruiz-Cortes, Z.T., Yoder, P. & Hanigan, M.D. (2022). Effects of Essential Amino Acid Deficiency on General Control Nonderepressible 2/Eukaryotic Initiation Factor 2 Signaling and Proteomic Changes in Primary Bovine Mammary Epithelial Cells. *Curr. Issues Mol. Biol.* 2022, 44, 1075-1086. <https://doi.org/10.3390/cimb44030071>
- Tyanova, S., Temu, T. & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*, 11(12), 2301-2319. doi:10.1038/nprot.2016.136
- Weaver, S. R., Prichard, A. S., Maerz, N. L., Prichard, A. P., Endres, E. L., Hernández-Castellano, L. E., Akins, M., Bruckmaier R. M. & Hernandez, L. L. (2017). Elevating serotonin pre-partum alters the Holstein dairy cow hepatic adaptation to lactation. *PloS one*, 12(9), e0184939-e0184939. doi:10.1371/journal.pone.0184939
- Yoder, P. S., Ruiz-Cortes, T., Castro, J. J. & Hanigan, M. D. (2019). Effects of varying extracellular amino acid profile on intracellular free amino acid concentrations and cell signaling in primary mammary epithelial cells. *J Dairy Sci*, 102(10), 8977-8985. doi:10.3168/jds.2018-16122