

## Estudio de la variación genética de la ruta isomerasa del D-galacturonato y análisis estructural de la enzima UxaA (D-altronato deshidratasa) para la transformación de residuos agroindustriales ricos en pectinas a etanol<sup>1</sup>

Simón David Mejía<sup>2</sup>, Julián A. Sánchez<sup>3</sup>, Alejandro Soto-Ospina<sup>4</sup>, Winston Rojas<sup>5</sup>, Leonardo Alberto Ríos-Osorio<sup>6</sup>

### Resumen

Los residuos agroindustriales son una problemática a nivel mundial ya que pueden producir contaminación de los cuerpos de agua o lixiviados al ser desechados en los rellenos sanitarios. Estos residuos, como las cáscaras de algunos frutos, son muy ricos en pectinas que no pueden ser fácilmente fermentables por levaduras como la *Saccharomyces cerevisiae*. Por lo que se hizo una búsqueda de bacterias y hongos que posean esa maquinaria metabólica, en una base de datos construida

a partir de sus secuencias nucleotídicas en un Megablastn, con los *softwares* DNAsp, Mega y Jalview. Se seleccionó la ruta isomerasa de cuatro microorganismos como *Escherichia coli*, *Lactococcus lactis*, *Erwinia carotova*, *Dickeya dadantii*, con el fin de entender la variabilidad genética y se obtuvo que la UxaA presentaba mayor variabilidad en la ruta isomerasa. Cada enzima UxaA fue modelada por predictores de estructura y se encontró una relación de alineamiento global proteico respecto a la distancia filogenética, por lo que se proponen las rutas isomerasas de dos microorganismos

- 1 Artículo original derivado del proyecto de investigación *Capital semilla pectinas de Colombia (Pepticol)*, ejecutado entre 2014 y 2015. Financiado por la Alcaldía de Medellín.
- 2 Estudiante de Microbiología Industrial y Ambiental, tecnólogo químico del SENA. Escuela de Microbiología, Universidad de Antioquia, Medellín, Antioquia, Colombia. Correo: simonmejia2009@gmail.com.
- 3 Químico, doctorando en Ciencias Químicas. Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellín, Antioquia, Colombia. Correo: julian.sanchezr@udea.edu.co.
- 4 Doctor en Ciencias Químicas de Universidad de Antioquia. Químico Universidad de Antioquia, Líder del Grupo de Investigación en Alimentos (GRIAL), Facultad de Ingenierías, Unilasallista Corporación Universitaria, Caldas, Antioquia, Colombia. Correo: jsoto72@unilasallista.edu.co. Orcid: 0000-0003-2382-7038.
- 5 Ph. D. en Biología, Universidad de Antioquia. MSc. Ciencias Básicas Biomédicas, Universidad de Antioquia. Biólogo, Universidad de Antioquia. Profesor del Instituto de Biología. Coordinador del Grupo de Investigación Genética Molecular (Genmol), Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia. Correo: winston.rojas@udea.edu.co. Orcid: 0000-0002-7654-3998.
- 6 Doctor en Sostenibilidad, Universidad de Antioquia. Bacteriólogo. Profesor titular de la Escuela de Microbiología. Grupo de Investigación Salud y Sostenibilidad, jefe del Centro de Investigación de la Escuela de Microbiología, Universidad de Antioquia. Correo: leonardo.rios@udea.edu.co. Orcid: https://orcid.org/0000-0002-7127-4629.

**Autor para Correspondencia:** leonardo.rios@udea.edu.co

Recibido: 24/05/2023      Aceptado: 10/10/2023

\*Los autores declaran que no tienen conflicto de interés

saprophytes: *Erwinia carotova* y *Dickeya dadantii* como posibles soluciones para transfectar a *Saccharomyces cerevisiae* como una alternativa fermentativa de estos residuos agroindustriales.

**Palabras clave:** pectina; diversidad nucleotídica; biología estructural; D-galacturonato; alineamiento; refinamiento.

---

**Study of the genetic variation of the D-galacturonate isomerase pathway and structural analysis of the enzyme UxaA (D-galacturonate dehydratase) for the transformation of pectin-rich agro-industrial residues to ethanol.**

**Abstract**

Agroindustrial waste is a global problem. It can cause contamination in body of water or leachate when disposed of in landfills. These residues, such as the shells of some fruits, are very rich in pectin, which cannot be easily fermented by yeasts such as *Saccharomyces cerevisiae*. Therefore, a search is made for bacteria and fungi that possess this metabolic machinery, in a database built from their nucleotide sequences in a Megablastn, with the DNAsp, Mega, and

Jalview software. The isomerase pathway of 4 microorganisms such as *Escherichia coli*, *Lactococcus lactis*, *Erwinia carotova*, *Dickeya dadantii*, was selected in order to understand the genetic variability. It was obtained that UxaA presented more significant variability in the isomerase pathway. Each UxaA enzyme was modeled by structure predictors. The global protein alignment relationship was found respect to the phylogenetic distance. The isomerase routes of two saprophytic microorganisms *Erwinia carotova* and *Dickeya dadantii* are proposed as possible solutions to transfect *Saccharomyces cerevisiae* as a fermentative alternative to these agro-industrial residues.

**Keywords:** diversity; pectins; enzymes; structural biology; yeast; isomerase; D-galacturonate acid.

---

**Estudo da variação genética da via da D-galacturonato isomerase e análise estrutural da enzima UxaA (D-galacturonato desidratase) para a transformação de resíduos agro-industriais ricos em pectina em etanol.**

**Resumo**

Os resíduos agro-industriais são um problema global. Pode causar contaminação

no corpo de água ou lixiviados quando depositados em aterros sanitários. Estes resíduos, tais como as cascas de algumas frutas, são muito ricos em pectina, que não podem ser facilmente fermentados por leveduras como a *Saccharomyces cerevisiae*. Portanto, é feita uma pesquisa de bactérias e fungos que possuem esta maquinaria metabólica, numa base de dados construída a partir das suas sequências nucleotídicas numa Megablastn, com

o software DNAsp, Mega e Jalview. O percurso da isomerase de 4 microrganismos como *Escherichia coli*, *Lactococcus lactis*, *Erwinia carotova*, *Dickeya dadantii*, foi seleccionado de modo a compreender a variabilidade genética. Obteve-se que UxaA apresentava uma variabilidade mais significativa na via da isomerase. Cada enzima UxaA foi modelada por preditores de estrutura. A relação de alinhamento global das proteínas foi encontrada em relação à

distância filogenética. As vias de isomerase de dois microrganismos saprófitos *Erwinia carotova* e *Dickeya dadantii* são propostas como possíveis soluções para transfectar *Saccharomyces cerevisiae* como alternativa fermentativa a estes resíduos agro-industriais.

**Palavras-chave:** diversidade, pectinas, enzimas, biologia estrutural, levedura, isomerase, ácido D-galacturonado.

---

## Introducción

La producción de compuestos químicos, combustibles y energía mediante tecnologías amigables con el medio ambiente, suponen un desafío global en la lucha contra el cambio climático. La biotecnología ha demostrado ser una alternativa para la producción verde empleando la biomasa de segunda y tercera generación como materia prima de bajo costo y disponible para la biotransformación a biocombustibles, biomoléculas y otros productos de interés (Benz *et al.*, 2014; Kuivanen *et al.*, 2012; Nielsen *et al.*, 2013; Protzko *et al.*, 2018). Residuos agroindustriales ricos en pectinas, generados del procesamiento industrial de frutas y vegetales como la remolacha, la manzana, las frutas cítricas, el mango, el cacao, el banano etc., están siendo subvalorados en toda la cadena de valor, son más de 500 millones de toneladas anuales de residuos aprovechables para la obtención de compuestos, metabolitos y bienes de alto valor agregado como el etanol (Choi *et al.*, 2013; Huisjes, Luttik *et al.*, 2012;

Kuivanen *et al.*, 2012; Maury *et al.*, 2008; United Nations, 2022). En los diversos tipos de frutas y verduras, las pectinas están contenidas en la cáscara, la pulpa, el bagazo y la corteza, aproximadamente del 10 % al 40 % en peso seco. Esta biomasa rica en azúcares y polisacáridos no ha sido aprovechada integralmente en sistemas fermentativos, debido a que la levadura *Saccharomyces cerevisiae*, microorganismo popular de ingeniería metabólica y biotecnología microbiana para la producción de etanol y otros metabolitos, no posee naturalmente la maquinaria proteica necesaria para transformar los residuos del ácido D-galacturónico y el D-galacturonato a etanol, por lo que hasta ahora la fermentación alcohólica de este tipo de sustratos empleando levaduras transformadas genéticamente no ha sido reportada aún en la literatura (Huisjes, 2013; Huisjes, Luttik *et al.*, 2012; Putnik *et al.*, 2017; Van Maris *et al.*, 2006). La pectina es un heteropolisacárido estructural de la pared celular primaria y lámina media de las células vegetales formado principalmente

por enlaces  $\alpha$ -(1→4) de residuos del ácido D-galacturónico (Alimardani-Theuil *et al.*, 2011; Edwards y Doran-Peterson, 2012; Gwanpua *et al.*, 2014; Mojzita *et al.*, 2010; Protzko *et al.*, 2018; Shpigelman *et al.*, 2015; Souffriau *et al.*, 2012; Zhang *et al.*, 2011).

Estructuralmente la pectina se compone de complejas regiones de polisacáridos y azúcares siendo el dominio homogalacturonano la principal región lineal la cual se compone de monómeros D-galacturonato de enlaces  $\alpha$ -(1→4) y otros polímeros no lineales como ramnogalacturonano I, ramnogalacturonano II y xilogalacturonano (Huisjes, 2013; Richard and Hilditch, 2009; Shpigelman *et al.*, 2015). Es sabido que naturalmente las levaduras no poseen la maquinaria metabólica necesaria para catabolizar el monómero de pectina (D-galacturonato), sin embargo, en bacterias y eukaryas se ha documentado la existencia de tres rutas catabólicas para el metabolismo del D-galacturonato: la ruta isomerasa estudiada extensivamente en *Escherichia coli*, la segunda una ruta oxidativa identificada en *Agrobacterium tumefaciens* y *Pseudomonas*, y la tercera una ruta eucariota recién descubierta en *Trichoderma reesei* denominada reductiva (Gullón *et al.*, 2013; Richard and Hilditch, 2009; Rorick *et al.*, 2011; Wilkins, 2009). Al revisar las tres rutas, diferentes autores afirman que la ruta oxidativa presenta inconvenientes debido al exceso de producción de NAD(H) y su posterior efecto en el metabolismo central vía 2-ceto-glutarato (Huisjes, 2013). Al igual que la ruta eucariota, debido a la participación de los cofactores redox NAD(H), NADP(H) y el glicerol como intermediario, representaría

un problema en los procesos anaerobios en *S. cerevisiae* para la producción de etanol (Huisjes, De Hulster *et al.*, 2012; Van Maris *et al.*, 2006), caso contrario ocurre con la ruta isomerasa que produce migración de fragmentos estructurales o reacomodos de grupos funcionales con gastos energéticos más bajos para obtener los isómeros y productos deseados. De igual forma, Huisjes evaluó estas rutas y determinó que la conversión del ácido D-galacturónico a etanol requiere la entrada de dos equivalentes redox, debido al mayor grado de oxidación del ácido D-galacturónico en comparación con otro sustrato como la D-glucosa, lo que conlleva a la oxidación de dos moléculas de NADH o NADPH por una molécula de ácido D-galacturónico cuando se emplea la ruta isomerasa o reductiva (Huisjes, 2013; Kuivanen *et al.*, 2019).

Profundizando a nivel bioquímico en el entendimiento del sistema fermentativo, una vez se identificaron las rutas metabólicas que podrían ayudar en la degradación de las pectinas, cuya mejor ruta podría ser introducida en *S. cerevisiae*; la ruta metabólica isomerasa es la más promisoría para la implementación de ingeniería metabólica en *S. cerevisiae*, si se tiene en cuenta que, durante la fermentación de hidrolizados en condiciones anaeróbicas, parte de los azúcares son convertidos a glicerol para ser reoxidados a NADH citológico en reacciones de biosíntesis (31). Así que, el metabolismo del ácido D-galacturónico podría funcionar como un disipador redox del exceso de NADH, evitando de este modo la producción de glicerol y aumentando el rendimiento global de etanol, caso contrario de la ruta redox que

presenta un desbalance estequiométrico en relación 2:1 para la transformación de ácido D-galacturónico (Huisjes, 2013; Huisjes, Luttik *et al.*, 2012).

La vía metabólica bacteriana isomerasa del D-galacturonato implica cinco reacciones catalizadas por enzimas, y es iniciada por la isomerización del D-galacturonato a D-tagaturonato mediada por la enzima UxaC (D-galacturonato isomerasa), después una reducción del D-tagaturonato a D-altronato es mediada por la enzima dependiente de NADH UxaB (D-tagaturonato reductasa), luego el D-altronato es deshidratado a 2-deshidro-3-desoxi-D-gluconato por medio de la enzima UxaA (D-altronato deshidratasa), posteriormente hay una fosforilación y se produce el 2-ceto-3-desoxi-6-fosfogluconato vía la enzima KdgK (2-ceto-3-desoxi gluconato quinasa). Este compuesto, finalmente es dividido a piruvato y gliceraldehído-3-fosfato por la enzima KdgA (2-ceto-3-desoxi-6-fosfogluconato aldolasa) vinculando así el metabolismo del D-galacturonato con la glucólisis (Huisjes, De Hulster *et al.*, 2012; Schäfer *et al.*, 2020). Por consiguiente, la expresión de la vía para el metabolismo del D-galaturonato en *S. cerevisiae* para la generación de etanol por métodos de ingeniería metabólica, presenta una serie de retos a nivel de expresión y transformación genética, redes metabólicas y de balance de flujos. A su vez, la poca información que se tiene de su comportamiento y transformarlo por vía fermentativa a etanol, empleando como sustrato residuos agro-industriales hidrolizados ricos en pectinas se ha visto limitado debido a problemas de inhibición de

crecimiento y represión catabólica (Biz *et al.*, 2016; Huisjes, 2013; Marsiglia *et al.*, 2016; Moncada, 2012; Tsukamoto *et al.*, 2013).

En el entendimiento de este sistema y en el estudio de la diferencia genética a nivel de bacterias, la bioinformática y la biología estructural son ramas de la biología que permiten conceptualizar los genes, su expresión y funcionalidad proteica. Una contribución bioinformática clásica ha sido la tasa evolutiva de un gen en cuestión, la cual es basada sobre los patrones de conservación de la secuencia genética entre genes relacionados filogenéticamente (Jacobs *et al.*, 2017). Asimismo, estudiando la diferencia genética a partir de los cambios estructurales y tridimensionales de las enzimas involucradas en las transformaciones, hace que el modelamiento molecular sea una buena alternativa en el estudio de los sistemas biológicos (Alvarez *et al.*, 2023).

Las técnicas experimentales de cristalización son difíciles de ejecutar, ya que cuentan con muchas limitaciones en la obtención de los cristales, por la ausencia de la propiedad física como la difracción de la luz por parte del cristal o por la formación de zonas de baja electrodensidad en la detección de técnicas como la microscopía criogénica de imagen, la cual es favorecida por el comportamiento flexible y desordenado de estructuras secundarias tipo bucles o *loops* (Higgins y Lea, 2017; Soni y Madhusudhan, 2017) las cuales no permiten elucidar bien la estructura tridimensional y generar un modelo estructural más completo (Al-Karadaghi, 2017). Los métodos computacionales para predecir una



estructura tridimensional son un gran aliado en la obtención de modelos para entender la funcionalidad proteica, estos utilizan metodologías que derivan de algoritmos que conciben búsquedas por homología de la secuencia primaria de aminoácidos, reconocimientos de plegamiento estructural a partir de los aminoácidos que componen la proteína y cuando no existe plantilla estructural se utilizan algoritmos *ab-initio* que usan *threading* (hilamiento) de bloques de ensamble y adoptan formas dependiendo del campo de fuerza externo que el *software* le aplica al modelo, esta es una técnica naciente e *in silico* en el análisis de datos (Cavasotto y Phatak, 2009; Dakal *et al.*, 2017; Guven *et al.*, 2019; Iyer *et al.*, 2019; Johansson-Åkhe *et al.*, 2019; Rimola *et al.*, 2011; Rodríguez *et al.*, 1998; Soni y Madhusudhan, 2017; Soto-Ospina *et al.*, 2021; Tejada Moreno *et al.*, 2022).

Finalmente, al realizar la revisión respectiva de la literatura se corroboró que en la actualidad la transformación metabólica de la pectina (ácido D-galacturonato) en sistema de levadura, no ha sido posible por medio de ingeniería genética y con el fin de aprovechar esta biomasa agroindustrial que sirva como nueva fuente de materia prima de bajo costo para la obtención de etanol, el presente trabajo propone conocer el contexto genético y evolutivo de los genes de la vía metabólica isomerasa en bacterias, mediante la aplicación *in silico* de herramientas bioinformáticas y de biología estructural con el fin de sentar las bases a futuras investigaciones para la transformación de la levadura *S. cerevisiae* por medio de ingeniería metabólica.

## Materiales y métodos

### Recopilación de datos de secuencia

Los genes que codifican la ruta isomerasa para el metabolismo del D-galacturonato se evaluaron en los siguientes microorganismos de interés: *Escherichia coli*, *Lactococcus lactis*, *Erwinia carotova*, *Dickeya dadantii*. Para cada microorganismo se obtuvo la secuencia de nucleótidos de los genes *uxaC*, *uxaB*, *uxaA*, *kdgK* y *kdgA* y su correspondiente secuencia de aminoácidos (proteína), a partir de las bases de datos del NCBI y PDB-UniProt. A su vez, se verificó la anotación de todos los genes y su respectiva proteína y la información se almacenó en formato FASTA (Apweiler *et al.*, 2014; Berman *et al.*, 2012; UniProt Consortium, 2010; 2012; Crooks *et al.*, 2004) organized by the Worldwide Protein Data Bank, was held at Cold Spring Harbor Laboratory (CSHL).

Para obtener las secuencias de alineamiento, se realizó un Basic Local Alignment Search Tool (BLASTn) a nivel de nucleótido para cada secuencia génica por microorganismo dentro de la especie. El algoritmo se ajustó de la siguiente manera: Base de datos = colección de nucleótidos no redundantes (nr/nt), Optimizar para = secuencias altamente similares (Megablast), Umbral esperado = 10, Tamaño semilla = 28, Puntuación igual/desigual = 1,-2, Costos de hueco = linear, la versión del programa BLASTN 2.8.0 (Morgulis *et al.*, 2008).

## Estimación de variabilidad génica

Para cada gen alineado se obtuvo el *cluster* de datos de cada especie de microorganismo evaluado. Con el fin de determinar la variabilidad génica y la tasa de sustitución sinónima y no sinónima de los genes de la ruta metabólica isomerasa a nivel de especie, se procedió a utilizar el programa DNAsp versión 6 (DNA Sequence Polymorphism) evaluando los sitios polimórficos y estimador de neutralidad mediante el algoritmo de Tajima's.

## Búsqueda de la enzima UxaA en las bases de datos

En la búsqueda de las bases de datos RCSB Protein Data Bank (Berman *et al.*, 2012) no se encontró estructura cristalizada ni resuelta para la enzima UxaA para las bacterias seleccionadas en el total de proteo bacterias estudiadas, las cuales fueron *Lactococcus lactis*, *Escherichia coli*, *Erwinia carotovora* y *Dickeya dadantii*. Las secuencias primarias de la enzima UxaA para cada una de las especies fue descargada de las bases de datos Uniprot y Embl, las cuales fueron almacenadas en formato FASTA (Stoesser *et al.*, 2002; UniProt Consortium, 2010).

## Alineamiento de secuencias primarias

Se tomaron las secuencias primarias almacenadas para la enzima UxaA y se procedió a realizar el alineamiento con el *software* Jalview, seleccionando el histograma normalizado sin consenso, asimismo se hizo el cálculo de índice de

similaridad ID a partir de las herramientas T-Coffee 2.0 y Clustal X para estudiar las regiones de conservación potencial para los fragmentos estructurales de las cuatro enzimas en comparación, las regiones de alta conservación se muestran en un diagrama consenso a partir del *Logomat* con coloración *clustal*, para un alineamiento por pares de secuencias de aminoácidos y obtener los respectivos porcentajes de similaridad de las secuencias primarias (Crooks *et al.*, 2004; Larkin *et al.*, 2007; Waterhouse *et al.*, 2009).

## Modelos de predicción estructural para la enzima UxaA

Se tomaron las secuencias primarias de las cuatro enzimas UxaA de las especies de bacterias *Lactococcus lactis*, *Escherichia coli*, *Erwinia carotovora* y *Dickeya dadantii* y se procedió a modelar su estructura tridimensional a partir de las secuencias primarias con los algoritmos de predicción estructural de I-Tasser que hacen un ensamblaje por homología proteica y cada uno de estos fragmentos estructurales se une por hilamiento (*Threading*), adopta la conformación estructural debida a la herramienta REMO de enlaces de hidrógeno, lo que favorece la conformación espacial del modelo y lo agrupa bajo los parámetros de C-score, número de iteraciones (*decoys*) y densidad de *cluster* (*cluster density*), generando un archivo de salida sin refinamiento para cada uno de los modelos, los cuales se seleccionaron para el modelo con mejor puntuación para cada tipo de bacteria (Li y Zhang, 2009; Roy *et al.*, 2010; Yang *et al.*, 2015; Zhang, 2008).

## Refinamiento estructural de los modelos estructurales

Los modelos obtenidos por el predictor de estructura con los resultados bajo la más alta calificación, fueron refinados a partir de los *softwares* Modrefiner y fragmentos guiados de dinámica molecular (FG-MD), los cuales generaron modelos de refinamiento estructural que se validaron posteriormente, teniendo en cuenta campos de fuerza a los que se sometió el modelo, buscando mejorar la orientación espacial del esqueleto carbonado, consideración de enlaces de hidrógeno y los rotámetros de los aminoácidos en las cadenas laterales de la enzima UxaA (Xu y Zhang, 2011; Zhang *et al.*, 2011).

## Validación estructural de los modelos de la enzima UxaA

Cada uno de los modelos obtenidos con el predictor de estructura sin refinar y refinados, fue validado de manera energética y estereoquímica a partir de algoritmos bioinformáticos que permiten generar comparaciones de medidores de parámetros globales y locales de los modelos, para los valores energéticos se hizo uso del suite de Swiss Assessment Structure con la función de puntaje QMEAN6 como un valor en rango normalizado que tiene en cuenta la combinación lineal de seis términos para los modelos, los cuales compilan el comportamiento del modelo en las torsiones angulares, afectación de campos de fuerza externos, valoración de las interacciones de las cadenas laterales como enlaces de hidrogeno, efectos en la solvatación y accesibilidad del solvente. Asimismo, se hizo

uso del cálculo del gráfico de Ramachandran con la herramienta Rampage para evaluar cada residuo de aminoácido según la orientación de los ángulos diedrales Phi ( $\phi$ ) y Psi ( $\psi$ ), seleccionando el modelo que considere la cantidad de residuos en la región favorable y permitida, los cuales acogen en gran proporción de residuos en el cuadrante, en donde se tiene estructura secundaria alfa-hélice (Artimo *et al.*, 2012; Benkert *et al.*, 2008; Guex and Peitsch, 1997; Studer *et al.*, 2019) becoming an extensible and integrative portal accessing many scientific resources, databases and software tools in different areas of life sciences. Scientists can henceforth access seamlessly a wide range of resources in many different domains, such as proteomics, genomics, phylogeny/evolution, systems biology, population genetics, transcriptomics, etc. The individual resources (databases, web-based and downloadable software tools.

## Visualizadores estructurales

Los visualizadores permiten realizar los alineamientos y dar una idea de los modelos estructurales ensamblados y creados, los alineamientos se hacen con el *software* Chimera U.C.S.F. versión 1.11 (Pettersen *et al.*, 2004), para ello se usa la comparación estructural con la herramienta Matchmaker, para los parámetros de matriz de sustitución de bloques BLOSUM62 y el algoritmo de programación dinámica que produce un alineamiento global para proteínas, ácidos nucleicos y canales de proteína transmembrana, para una identificación y localización topológica se hace uso del *software* Deepview/Swiss-PdbViewer versión



3.7 (Guex and Peitsch, 1997). El cálculo de la desviación estándar RMSD, el coeficiente de similaridad QH y el índice de similaridad se obtienen con un multialineamiento STAMP (Multi alignment of protein structure and sequences), visualizado y mediado por el *software* VMD que viene en el paquete NAMD/VMD de la universidad de Illinois (Humphrey *et al.*, 1996; Phillips *et al.*, 2005).

codificantes de proteínas es la relación entre el número de sustituciones no sinónimas por sitio no sinónimo y el número de sustituciones sinónimas por sitio sinónimo, también conocido como dN/dS o Ka/Ks. Las sustituciones se determinan comparando la secuencia ORF de genes con la secuencia de un gen de referencia, estimando procesos de selección purificadora en valores dN/dS más bajos, mientras que la selección positiva, que aumenta dN, dará como resultado valores de dN/dS incrementados (Ish-Am *et al.*, 2015). Estos datos fueron consignados en la **tabla 1**.

### Resultados y discusión

La medida básica de la presión de selección que actúa sobre las secuencias

**Tabla 1.**

*Sitios polimórficos, prueba de neutralidad, diversidad nucleotídica y tasa de sustitución sinónima y no sinónima de los genes de la ruta metabólica isomerasa a nivel de especie: n: número de muestra; d.n.: diversidad nucleotídica; Tajima's D: prueba de neutralidad; Pi Ka: sustitución no-sinónima; Pi Ks: sustitución sinónima; Pi (Ka/Ks): relación sinónima y no-sinónima*

Microorganismo	KdgA					
	n	d.n.	Tajima's D	Pi Ka	Pi Ks	Pi (Ka/Ks)
<i>Dickeya dadantii</i>	2	0,003	ND	0,002	0,006	0,364
<i>Erwinia carotovora</i>	9	0,006	-0,507	0,006	0,005	1,128
<i>Escherichia coli</i>	851	0,001	-2,411	0,001	0,000	1,533
<i>Lactococcus lactis</i>	39	0,016	0,481	0,017	0,015	1,135
Microorganismo	KdgK					
	n	d.n.	Tajima's D	Pi Ka	Pi Ks	Pi (Ka/Ks)
<i>Dickeya dadantii</i>	2	0,005	ND	0,005	0,005	1,064
<i>Erwinia carotovora</i>	9	0,005	-1,740	0,005	0,005	0,980
<i>Escherichia coli</i>	847	0,001	-1,869	0,002	0,004	0,562
<i>Lactococcus lactis</i>	27	0,015	-1,760	0,014	0,020	0,706
Microorganismo	uxaA					
	n	d.n.	Tajima's D	Pi Ka	Pi Ks	Pi (Ka/Ks)
<i>Dickeya dadantii</i>	2	0,006	ND	0,006	0,007	0,895
<i>Erwinia carotovora</i>	9	0,004	0,438	0,003	0,007	0,368

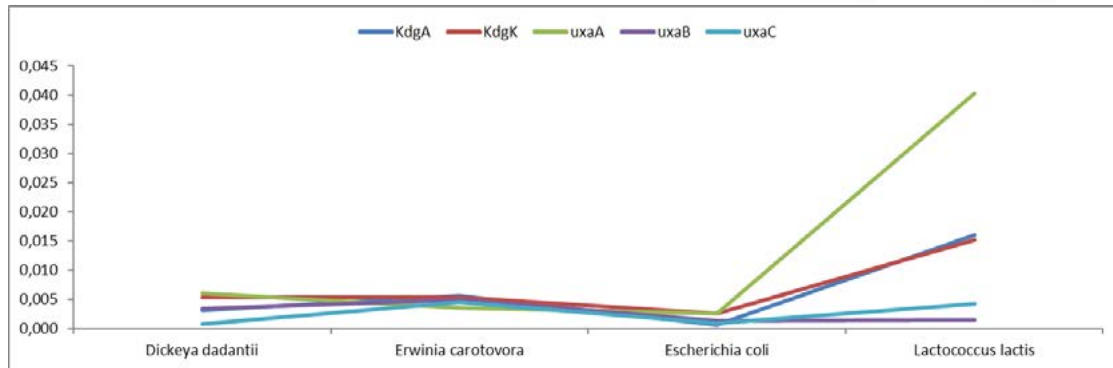
<i>Escherichia coli</i>	848	0,003	-1,892	0,002	0,004	0,475
<i>Lactococcus lactis</i>	3	0,040	ND	0,040	0,047	0,853
<b>Microorganismo</b>						
<i>uxaB</i>						
	<b>n</b>	<b>d.n.</b>	<b>Tajima's D</b>	<b>Pi Ka</b>	<b>Pi Ks</b>	<b>Pi (Ka/Ks)</b>
<i>Dickeya dadantii</i>	2	0,003	ND	0,004	0,002	1,716
<i>Erwinia carotovora</i>	9	0,005	-0,56675	0,004	0,007	0,614
<i>Escherichia coli</i>	798	0,001	-2,57737	0,001	0,001	1,055
<i>Lactococcus lactis</i>	2	0,001	ND	0,001	0,003	0,288
<b>Microorganismo</b>						
<i>uxaC</i>						
	<b>n</b>	<b>d.n.</b>	<b>Tajima's D</b>	<b>Pi Ka</b>	<b>Pi Ks</b>	<b>Pi (Ka/Ks)</b>
<i>Dickeya dadantii</i>	2	0,001	ND	0,001	0,000	0,000
<i>Erwinia carotovora</i>	9	0,005	-0,422	0,004	0,005	0,889
<i>Escherichia coli</i>	812	0,001	-2,164	0,001	0,002	0,353
<i>Lactococcus lactis</i>	26	0,004	-1,759	0,004	0,005	0,837

Como se observa en la **tabla 1**, se realizó una caracterización intraespecie de cada gen implicado en la ruta isomerasa en los microorganismos modelo, se determinaron los indicadores de diversidad nucleotídica, neutralidad con la prueba de Tajima y relación sinónima y no sinónima. Con los valores de diversidad nucleotídica para los genes *uxaC*, *uxaB*, *kdgA* y *kdgK* se evidencia que no se presenta alta variabilidad genética dentro de la especie, mientras que el gen *uxaA* resultó en un valor de diversidad nucleotídica mayor. De manera preliminar,

para este gen (*uxaA*) se observa un evento de especiación. Así mismo, la relación sinónima y no-sinónima muestra que para los genes *kdgK* y *uxaB* en *D. dadantii* y *kdgA* en *E. carotovora*, *E. coli* y *L. lactis* el valor evolutivo fue mayor a 1, indicando eventos de variantes en regiones funcionales del gen que pueden resultar en cambios estructurales a nivel proteico. La comparación del valor de diversidad nucleotídica de los cinco genes que codifican la ruta isomerasa se muestra en la **figura 1**.

**Figura 1.**

Comparación de la diversidad nucleotídica de los genes de la ruta metabólica isomerasa *uxaC*, *uxaB*, *uxaA*, *kdgK* y *kdgA* a nivel de especie



En esta figura se hace un análisis comparativo de la variabilidad genética a nivel interespecie de cada gen evaluado en los microorganismos modelo. Las especies *D. dadantii*, *E. carotovora*, y *E. coli* no mostraron una alta diversidad nucleotídica, caso contrario, *L. lactis* presentó variabilidad en los genes *uxaA*, *kdgK*, y *kdgA*, siendo el gen *UxaA* el que presentó mayor variabilidad nucleotídica, por lo que se planteó realizar un estudio teórico estructural del gen más variable a nivel proteico.

Existen muchas proteínas que tienen estructuras faltantes y esto se debe a las

limitantes a nivel experimental, por lo que se utilizan predictores de estructura para generar modelos artificiales en el estudio de diferencia genética o análisis funcionales a nivel metabólico. Para este estudio se utiliza el predictor de estructura I-Tasser, el cual cuenta con un algoritmo de predicción de estructura con base en plantillas por homología y reconocimiento de plegamiento tridimensional, considerando la secuencia primaria de la proteína *UxaA* para las bacterias seleccionadas según los resultados de diversidad genética, los resultados para los modelos de I-Tasser se presentan en la **tabla 2**.

**Tabla 2.**

Parámetros de modelamiento molecular de *UxaA* para los microorganismos seleccionados

Microorganismo	Modelo	C-score	Número de decoys	Densidad del cluster
<i>Lactococcus lactis</i>	Modelo 1	-2,01	600	0,0590
	Modelo 2	-3,30	218	0,0163
	Modelo 3	-3,77	148	0,0102
	Modelo 4	-3,82	118	0,0097
	Modelo 5	-2,63	86	0,0319
<i>Escherichia coli</i>	Modelo 1	-1,87	N/A	N/A
	Modelo 2	-3,63	N/A	N/A
	Modelo 3	-3,50	N/A	N/A
	Modelo 4	-2,67	N/A	N/A
	Modelo 5	-2,44	N/A	N/A
<i>Erwinia carotovora</i>	Modelo 1	-1,03	600	0,1562
	Modelo 2	-2,27	519	0,0452
	Modelo 3	-2,57	96	0,0337
	Modelo 4	-2,60	82	0,0325
	Modelo 5	-3,76	104	0,0102
<i>Dickeya dadantii</i>	Modelo 1	-1,06	600	0,1515
	Modelo 2	-2,18	598	0,0490
	Modelo 3	-2,69	78	0,0295
	Modelo 4	-4,43	66	0,0052
	Modelo 5	-2,80	66	0,0265

Según la tabla anterior, los parámetros más importantes están relacionados con la capacidad de generar modelos con alta confianza, esto se debe a funciones de puntaje que conciben los modelos cuyas predicciones convergen a las disposiciones espaciales con valor más positivo para C-score en un rango de [-5,2], que es uno de los criterios de selección para clasificar los modelos predichos, otro indicativo es el mayor número de *decoys* (iteraciones de

computación) y una alta densidad de *cluster*, lo que indica para ambos parámetros que hay una mayor cantidad de conformeros que adoptaron esa disposición espacial. Según los resultados obtenidos para las bacterias *Lactococcus lactis*, *Escherichia Coli* y *Erwinia carotovora* el mejor modelo fue el 1, en cambio *Dickeya dadantii* presentó el 2 como el modelo para realizar la caracterización bioinformática. En la **tabla 2** se presenta la caracterización de los modelos sin refinar y

con refinamiento FG-MD y Modrefiner para validarlos energética y estereoquímicamente, con el objetivo de evitar solapamientos

atómicos, malos enlazamientos y ángulos de torsión inadecuados que puedan fracturar la estructura proteica.

**Tabla 3.**

Validación energética y estereoquímica de los modelos de la enzima UxaA para las bacterias seleccionadas

Microorganismo	Modelo	Software	QMEAN6	Ramachandran plot		
				Región favorable	Región permitida	Región prohibida
<i>Lactococcus lactis</i>	Sin refinar	I-TASSER	0,294	343 (69,4 %)	95 (19,2 %)	56 (11,3 %)
	Refinado	FG-MD	0,301	345 (19,8 %)	98 (19,8 %)	51 (10,3 %)
		ModRefiner	0,453	434 (87,9 %)	50 (10,1 %)	10 (2,0 %)
<i>Escherichia coli</i>	Sin refinar	I-TASSER	0,222	339 (68,8 %)	92 (18,7 %)	62 (12,6 %)
	Refinado	FG-MD	0,230	547 (94,5 %)	29 (5,0 %)	3 (0,5 %)
		ModRefiner	0,380	424 (86 %)	50 (10,1 %)	19 (3,9 %)
<i>Erwinia carotovora</i>	Sin refinar	I-TASSER	0,247	321 (65,0 %)	125 (25,3 %)	48 (9,7 %)
	Refinado	FG-MD	0,249	321 (65,0 %)	123 (24,9 %)	50 (10,1 %)
		ModRefiner	0,374	413 (83,6 %)	67 (13,6 %)	14 (2,8 %)
<i>Dickeya dadantii</i>	Sin refinar	I-TASSER	0,233	342 (69,2 %)	107 (21,7 %)	45 (9,1 %)
	Refinado	FG-MD	0,228	353 (71,5 %)	101 (20,4 %)	40 (8,1 %)
		ModRefiner	0,353	437 (88,5 %)	42 (8,5 %)	15 (3,0 %)

Los modelos seleccionados con las mejores funciones de puntajes fueron refinados con las herramientas de fragmentos guiados de dinámica molecular (FG-MD) y Modrefiner, que someten los modelos a campos de fuerza buscando mejorar el enlazamiento y las torsiones diedrales, de la disposición espacial de los aminoácidos en los enlaces peptídicos. El mejor modelo calificado se usará como representante de la enzima UxaA como modelo hipotético para las bacterias seleccionadas. Según los resultados, el mejor diseño del modelo se logró con el

algoritmo de I-Tasser y el modelo creado por hilamiento fue refinado por el software Modrefiner para cada una de las bacterias. Para *Lactococcus lactis* los resultados mostraron que para un valor normalizado de Qmean6 en límite inferior de 0,0 a superior 1,0, su valor fue de 0,453 y a nivel estereoquímico con los gráficos de Ramachandran muestran un porcentaje muy alto de los residuos en la región favorable de 87,9 % con 434, en la región permitida fue de 10,1 % con 50 aminoácidos y 2,0 % para 10 aminoácidos. Para *Escherichia coli* el valor de Qmean6 fue de 0,380 y estereoquímico

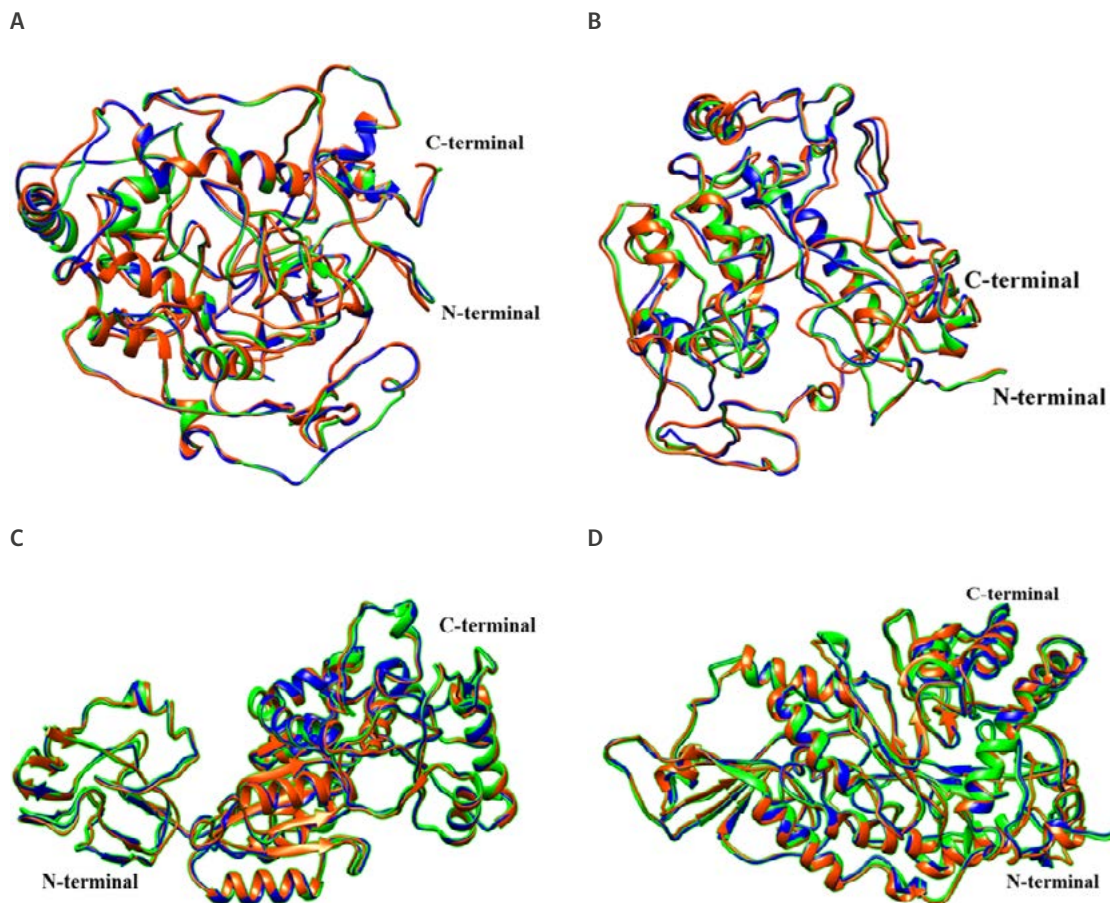


con residuos en región favorable de 86,0 % para 424, región permitida de 10,1 % para 54 residuos y 3,9 %, 19 residuos. La bacteria *Erwinia carotovora* tuvo puntaje energético de Qmean6 de 0,374 y estereoquímico en la región permitida de 83,6 % para 413 aminoácidos, 13,6 % para 67 aminoácidos en la región favorable y 2,8 % para 14 aminoácidos en la región prohibida.

Finalmente, *Dickeya Dadantii* presenta energéticamente un valor de Qmean6 de 0,353 y estereoquímico de 88,5 % para 437 aminoácidos en la región favorable, 8,5 % para 42 aminoácidos en la región permitida y 3,0 % en la región prohibida. En la **figura 2**. Se presenta el alineamiento global con el algoritmo de Needleman Wunsch para la proteína UxaA.

**Figura 2.**

Alineamiento tridimensional global de los modelos de UxaA para las bacterias con el modelo sin refinar (verde), modelo refinado FG-MD (azul) y modelo refinado Modrefiner (naranja): A) UxaA *Lactococcus lactis*; B) UxaA *Escherichia coli*; C) UxaA *Erwinia carotovora*; D) UxaA *Dickeya dadantii*

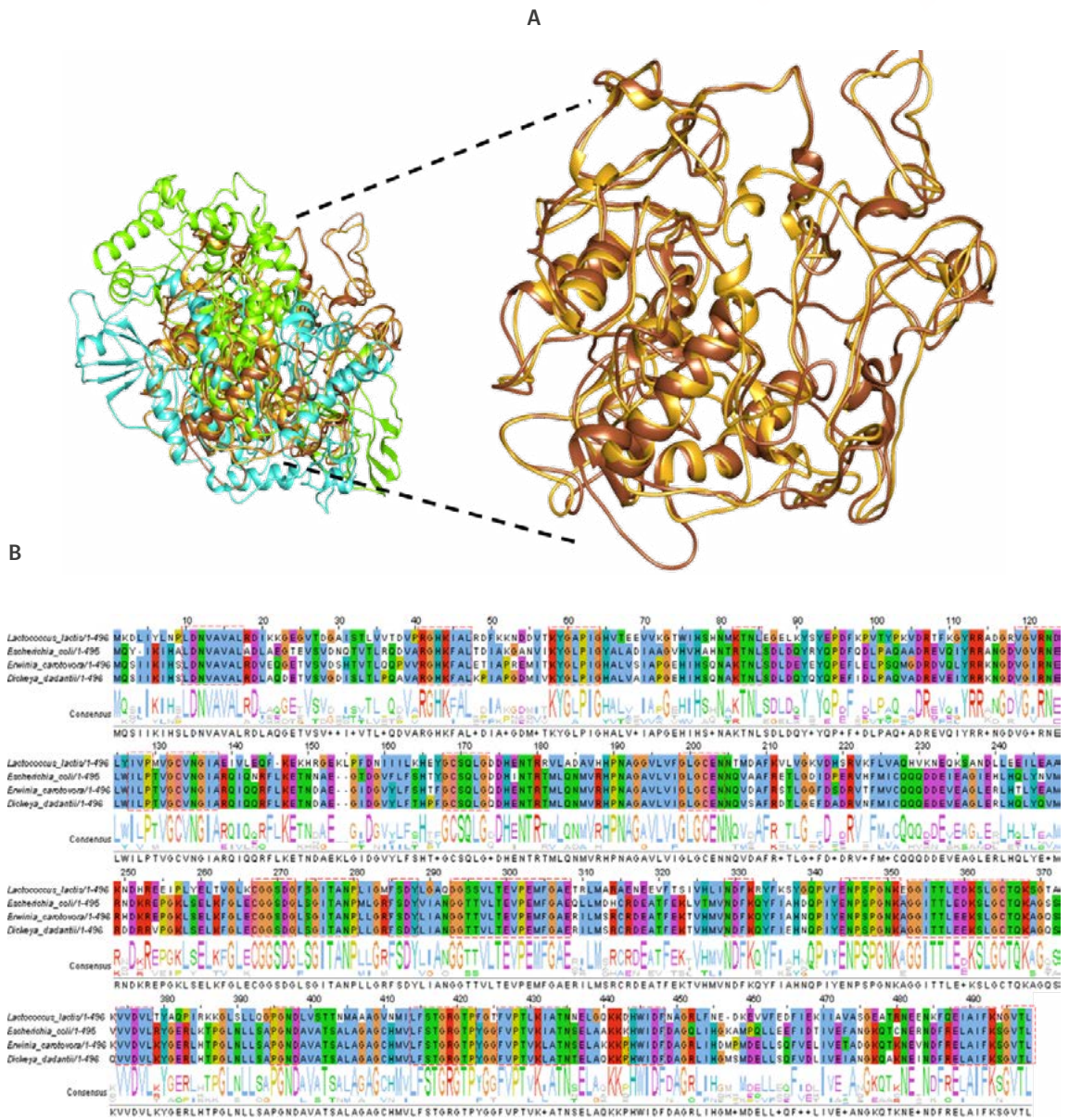


En la **figura 2** se puede observar a nivel tridimensional la comparación entre los modelos alineados a nivel global y se presentan pequeñas variaciones entre los modelos, esto se da debido al efecto del refinamiento considerando mejoras en los enlaces de hidrógeno, torsiones de ángulos y acomodamiento de los ángulos diedrales, de igual forma los patrones de interacción por enlaces de hidrógeno fueron restringidos para ángulos de  $45^\circ$  y una distancia de enlace de  $2,7\text{\AA}$ . Para los modelos con puntajes más altos que se presentaron con refinamiento Modrefiner, se seleccionaron para realizar un alineamiento global considerando cada uno de sus aminoácidos en la secuencia primaria y como bloques de ensamble se utilizó un multialineamiento con el algoritmo Needleman-Wunsch con una matriz Blosum 62 generando la comparación espacial que fue depurada para las proteínas que sí presentaron alineamiento y en *zoom* se muestran en la región derecha de la **figura 3**.

Como se puede observar, para el alineamiento a nivel tridimensional de la **figura 3A**, las pequeñas variaciones en los genes en regiones específicas generan cambios drásticos en la tridimensionalidad de la enzima, que no permite un buen alineamiento tridimensional y por eso las estructuras de los modelos predichos a nivel *in silico* no coinciden completamente en el multialineamiento para las especies bacterianas evaluadas. Al depurar las estructuras y considerar las que presentan un buen alineamiento con RMSD menor a  $2,00\text{\AA}$ , en el caso del alineamiento por pares se obtuvo un RMSD de  $1,81\text{\AA}$ , que es un valor muy pequeño de desviación cuantitativa y que se correlaciona con la forma estructural de UxaA *Erwinia carotovora* (dorado) y UxaA *Dickeya dadantii* (marrón); para este alineamiento también se midieron los valores de índice de similaridad  $Q_{ii}$  cuyo valor normalizado fue de  $0,68$  que muestra de manera cuantitativa un buen alineamiento entre ambas enzimas UxaA.

Figura 3.

Alineamiento global para las bacterias seleccionadas de la ruta isomerasa: A) multilineamiento global de las proteínas de UxaA refinadas, UxaA *Lactococcus lactis* (verde), UxaA *Escherichia coli* (azul ciano), UxaA *Erwinia carotovora* (dorado), UxaA *Dickeya dadantii* (marrón); B) alineamiento con Jalview de las secuencias primarias del gen UxaA



Las secuencias primarias de las enzimas UxaA fueron alineadas por pares a partir del *software* Jalview como se presenta en la **figura 3B**, este alineamiento se hizo teniendo en cuenta toda la secuencia primaria

en formato Fasta y se utilizó una matriz Blosum62 de archivo plano, calculando el logomat para las regiones más conservadas, encontrando un histograma con un eje que depende del tamaño relativo de las letras



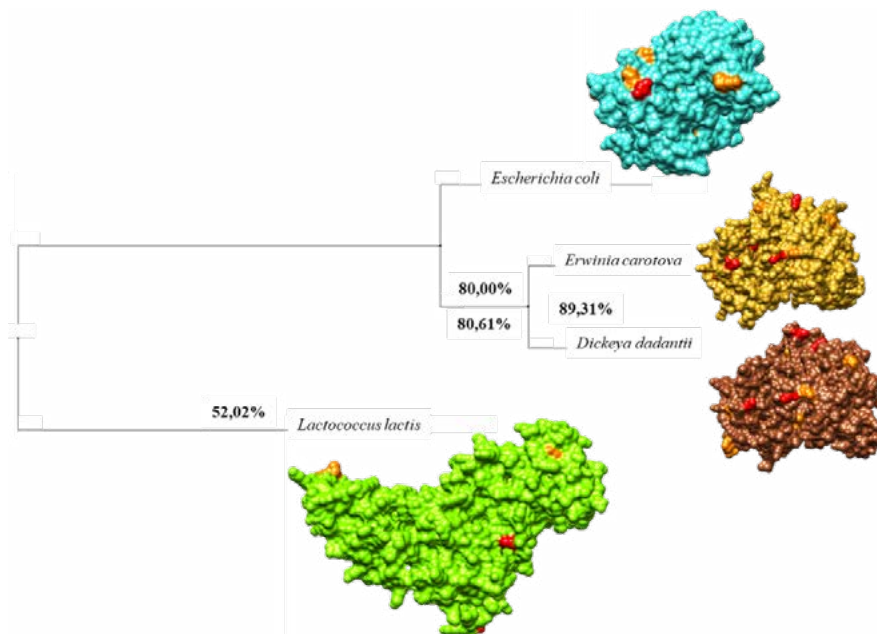
que componen las secuencias primarias. Asimismo, se puede observar que algunas regiones en las vecindades del N-terminal para todas las estructuras tienen un impacto importante en la distribución espacial del modelo, ya que esas variaciones son causantes de los cambios que no permiten los alineamientos tridimensionales para cada enzima específica de cada bacteria, de igual forma el dominio funcional se mantiene conservado, dado que la enzima UxaA procesa el sustrato D-ácido altrónico a 2-dehidro-3-deoxi-D-gluconato y todas las enzimas para

las bacterias consideradas tienen esa ruta bioquímica de transformación, pero con diversos niveles de transformación en su cinética enzimática.

Respecto a los resultados obtenidos con anterioridad, al intentar entender por qué ocurren esos alineamientos entre las enzimas UxaA de cada una de las especies de bacterias, se procede a hacer un análisis de diferencia genética, con el fin de establecer las correlaciones entre las enzimas y los resultados se presentan en la **figura 4**.

**Figura 4.**

Relación filogenética a nivel estructural y porcentajes de identidad para la enzima UxaA de las bacterias evaluadas



Como se muestra en las **figuras 3 y 4** vemos que la UxaA de *Lactococcus lactis* (verde) es la que peor alineamiento tridimensional presenta y esto se da porque en términos filogenéticos para la enzima,

es la que se encuentra más alejada, según los resultados de árbol filogenético, la enzima UxaA de *Escherichia coli* (azul ciano) presenta fragmentos que consideran un buen multialineamiento pero tiene muchas

regiones que no coinciden con las demás y eso se puede observar en lo alejado que está de las enzimas UxaA de *Erwinia carotovora* (dorado) y de la UxaA *Dickeya dadantii* (marrón), las cuales son muy cercanas entre sí, además, ambas bacterias se consideran saprofitas y al serlo son patógenas de plantas que presentan pectinas en su pared vegetal y esto obliga a tener la ruta isomerasa muy encendida a diferencia de los hongos que tiene la ruta reductiva. Esto conlleva a estudiar y comparar los niveles de expresión de cada uno de los genes, ya que allí puede estar destacada la importancia enzimática y entender si la cinética enzimática se ve favorecida por esa diversidad o no.

### Conclusiones

La biología estructural es una herramienta muy útil que nos permite explicar comportamientos y entender sistemas vivos a nivel funcional, para correlacionar aspectos importantes como la incapacidad de una levadura como *Saccharomyces cerevisiae*, la cual no puede catabolizar el ácido D-galacturónico o el monómero el D-galacturonato para obtener energía. De igual forma, a partir de la búsqueda en la literatura se encuentran reportes que la ruta isomerasa asociada a las proteobacterias, consideran a la *Escherichia coli* como el modelo por excelencia de ingeniería genética; *Lactococcus lactis* tiene publicaciones de cinéticas enzimáticas considerando la ruta isomerasa y es un microorganismo GRAS (Generally recognized as safe), lo que se define como un microorganismo que puede ser usado en alimentos, dada su inocuidad. De igual forma la *Dickeya dadantii* y la

*Erwinia carotovora*, fueron seleccionadas como buenas candidatas dada su cercanía a nivel filogenético, ya que por su naturaleza saprofita tienen la ruta metabólica isomerasa muy activada y permite catabolizar este tipo de sustratos presentes en los residuos agroindustriales.

Se hace el estudio de diversidad genética para los cinco genes implicados en la ruta isomerasa y que hacen parte del metabolismo de estas bacterias seleccionadas, obteniendo que el gen que presentaba mayor diversidad es UxaA. Respecto a los modelos obtenidos por predicción de estructura, se hizo el modelamiento molecular, en el entendimiento del gen y su funcionalidad, se hizo un estudio de diferencia genética, se encontraron alineamientos tridimensionales parecidos a la diferencia genética encontrada para la enzima UxaA, siendo bastante representativa la enzima con una cinética más eficiente que por evolución natural, optimizan el procesamiento de las pectinas como residuos agroindustriales y que nos permite ver cómo dos microorganismos saprofitos pueden tener rutas metabólicas, que podrían proponerse para transfectar una levadura y favorecer conversiones a rutas fermentativas.

### Agradecimientos

SM-B agradecimiento a la Escuela de Microbiología y al proyecto ganador de Capital Semilla Pepticol. Homenaje y reconocimiento al profesor Gabriel de Jesús Bedoya (Q.E.P.D) por aceptar el aporte investigativo en el grupo de investigación de Genética Molecular (GenMol). AS-O fue financiado por la beca de doctorados



nacionales para Colombia, Francisco José de Caldas con el proyecto FP44842-124-2017 y la Corporación Universitaria Lasallista Grupo de Investigación en Alimentos GRIAL.

## Referencias

- Al-Karadaghi, S. (2017). *Experimental methods in structural biology: Protein crystallization and X-ray crystallography*. Protein Structure. <http://www.proteinstructures.com/Experimental/experimental-methods.html>.
- Alimardani-Theuil, P.; Gainvors-Claisse, A. and Duchiron, F. (2011). Yeasts: An attractive source of pectinases– From gene expression to potential applications: A review. *Process Biochemistry*, 46(8), 1525–1537. <https://doi.org/10.1016/j.procbio.2011.05.010>.
- Álvarez, Y.; Soto-Ospina, A.; Villegas-Ianau, A.; Araque, P.; Bedoya, G. D. J. y Velásquez, M. (2023). Modelamiento molecular y análisis estructural de la variante I684S en la enzima TYK2 y su relación con la propensión a la psoriasis. *Ciencias Biomédicas*, 47(182), 8–21.
- Apweiler, R.; Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Alam-Faruque, Y.; Alpi, E.; Antunes, R.; Arganiska, J.; Casanova, E. B.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Chan, W. M.; Chavali, G.; Cibrian-Uhalte, E.; Da Silva, A.; Zhang, J. (2014). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 42(D1), 191–198. <https://doi.org/10.1093/nar/gkt1140>.
- Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; De Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernández, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40(W1), 597–603. <https://doi.org/10.1093/nar/gks400>.
- Benkert, P.; Tosatto, S. C. E. and Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, 71(1), 261–277. <https://doi.org/10.1002/prot.21715>.
- Benz, J. P.; Protzko, R. J.; Andrich, J. M. S.; Bauer, S.; Dueber, J. E. and Somerville, C. R. (2014). Identification and characterization of a galacturonic acid transporter from *Neurospora crassa* and its application for *Saccharomyces cerevisiae* fermentation processes. *Biotechnology for Biofuels*, 7(1). <https://doi.org/10.1186/1754-6834-7-20>.
- Berman, H. M.; Kleywegt, G. J.; Nakamura, H. and Markley, J. L. (2012). The protein data bank at 40: Reflecting on the past to prepare for the future. *Structure*, 20(3), 391–396. <https://doi.org/10.1016/j.str.2012.01.010>.

- Biz, A.; Sugai-Guérios, M. H.; Kuivanen, J.; Maaheimo, H.; Krieger, N.; Mitchell, D. A. and Richard, P. (2016). The introduction of the fungal d-galacturonate pathway enables the consumption of d-galacturonic acid by *Saccharomyces cerevisiae*. *Microbial Cell Factories*, 15(1), 1-11. <https://doi.org/10.1186/s12934-016-0544-1>.
- Cavasotto, C. N. and Phatak, S. S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug Discovery Today*, 14(13-14), 676-683. <https://doi.org/10.1016/j.drudis.2009.04.006>.
- Choi, I. S.; Kim, J. H.; Wi, S. G.; Kim, K. H. and Bae, H. J. (2013). Bioethanol production from mandarin (Citrus unshiu) peel waste using popping pretreatment. *Applied Energy*, 102, 204-210. <https://doi.org/10.1016/j.apenergy.2012.03.066>.
- Crooks, G.; Hon, G.; Chandonia, J. and Brenner, S. (2004). NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res*, 14, 1188-1190. <https://doi.org/10.1101/gr.849004.1>.
- Dakal, T. C.; Kumar, R. and Ramotar, D. (2017). Structural modeling of human organic cation transporters. *Computational Biology and Chemistry*, 68, 153-163. <https://doi.org/10.1016/j.compbiolchem.2017.03.007>.
- Edwards, M. C. and Doran-Peterson, J. (2012). Pectin-rich biomass as feedstock for fuel ethanol production. *Applied Microbiology and Biotechnology*, 95(3), 565-575. <https://doi.org/10.1007/s00253-012-4173-2>.
- Guex, N. and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, 18(15), 2714-2723. <https://doi.org/10.1002/elps.1150181505>.
- Gullón, B.; Gómez, B.; Martínez-Sabajanes, M.; Yáñez, R.; Parajó, J. C. and Alonso, J. L. (2013). Pectic oligosaccharides: Manufacture and functional properties. *Trends in Food Science and Technology*, 30(2), 153-161. <https://doi.org/10.1016/j.tifs.2013.01.006>.
- Guex, N. and Peitsch, M. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. <https://doi.org/10.1002/elps.1150181505>.
- Güven, G.; Erginel-Unaltuna, N.; Samancı, B.; Gulec, C.; Hanagasi, H. and Bilgic, B. (2019). A patient with early-onset Alzheimer's disease with a novel PSEN1 p.Leu424Pro mutation. *Neurobiology of Aging*, 1, 2-5. <https://doi.org/10.1016/j.neurobiolaging.2019.05.014>.
- Gwanpua, S. G.; Van Buggenhout, S.; Verlinden, B. E.; Christiaens, S.; Shpigelman, A.; Vicent, V.; Kermani, Z. J.; Nicolai, B. M.; Hendrickx, M. and Geeraerd, A. (2014). Pectin modifications and the role of pectin-degrading enzymes during postharvest softening of Jonagold apples. *Food Chemistry*, 158,

- 283-291. <https://doi.org/10.1016/j.foodchem.2014.02.138>.
- Higgins, M. K. and Lea, S. M. (2017). On the state of crystallography at the dawn of the electron microscopy revolution. *Current Opinion in Structural Biology*, 46, 95-101. <https://doi.org/10.1016/j.sbi.2017.06.005>.
- Huisjes, E. (2013). *Towards fermentation of galacturonic acid-containing feedstocks with Saccharomyces cerevisiae*. <http://repository.tudelft.nl/view/ir/uuid:c3b977ac-dbf1-4c62-8ceb-23212ff0dcc0/>.
- Huisjes, E. H.; De Hulster, E.; Van Dam, J. C.; Pronk, J. T. and Van Maris, A. J. A. (2012). Galacturonic acid inhibits the growth of *Saccharomyces cerevisiae* on galactose, xylose, and arabinose. *Applied and Environmental Microbiology*, 78(15), 5052-5059. <https://doi.org/10.1128/AEM.07617-11>.
- Huisjes, E. H.; Luttkik, M. A. H.; Almering, M. J. H.; Bisschops, M. M. M.; Dang, D. H. N.; Kleerebezem, M.; Siezen, R.; Van Maris, A. J. A. and Pronk, J. T. (2012). Toward pectin fermentation by *Saccharomyces cerevisiae*: Expression of the first two steps of a bacterial pathway for D-galacturonate metabolism. *Journal of Biotechnology*, 162(2-3), 303-310. <https://doi.org/10.1016/j.jbiotec.2012.10.003>.
- Humphrey, W.; Dalke, A. and Schulten, K. (1996). VMD- Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- Ish-Am, O.; Kristensen, D. M. and Ruppin, E. (2015). Evolutionary conservation of bacterial essential metabolic genes across all bacterial culture media. *PLoS ONE*, 10(4), 1-15. <https://doi.org/10.1371/journal.pone.0123785>.
- Iyer, S.; Acharya, K. R. and Subramanian, V. (2019). Prediction of structural consequences for disease causing variants in C21orf2 protein using computational approaches. *Journal of Biomolecular Structure and Dynamics*, 37(2), 465-480. <https://doi.org/10.1080/07391102.2018.1429313>.
- Jacobs, C.; Lambourne, L.; Xia, Y. and Segrè, D. (2017). Upon accounting for the impact of isoenzyme loss, gene deletion costs anticorrelate with their evolutionary rates. *PLoS ONE*, 12(1), 1-15. <https://doi.org/10.1371/journal.pone.0170164>.
- Johansson-Åkhe, I.; Mirabello, C. and Wallner, B. (2019). Predicting protein-peptide interaction sites using distant protein complexes as structural templates. *Scientific Reports*, 9(1), 1-13. <https://doi.org/10.1038/s41598-019-38498-7>.
- Kuivanen, J.; Biz, A. and Richard, P. (2019). Microbial hexuronate catabolism in biotechnology. *AMB Express*, 9(1), 1-11. <https://doi.org/10.1186/s13568-019-0737-1>.
- Kuivanen, J.; Mojzita, D.; Wang, Y.; Hilditch, S.; Penttilä, M.; Richard, P. and Wiebe, M. G. (2012). Engineering filamentous

- fungi for conversion of d-galacturonic acid to L-galactonic acid. *Applied and Environmental Microbiology*, 78(24), 8676-8683. <https://doi.org/10.1128/AEM.02171-12>.
- Larkin, M.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; López, R.; Thompson, J. D.; Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), 2947-2948. <https://doi.org/10.1093/bioinformatics/btm404>.
- Li, Y. and Zhang, Y. (2009). REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins: Structure, Function and Bioinformatics*, 76(3), 665-676. <https://doi.org/10.1002/prot.22380>.
- Marsiglia, D. E.; Ojeda, K. A.; Ramírez, M. C. and Sánchez, E. (2016). Pectin extraction from cocoa pod husk (*Theobroma cacao* L.) by hydrolysis with citric and acetic acid. *International Journal of ChemTech Research*, 9(7), 497-507.
- Maury, J.; Asadollahi, M. A.; Møller, K.; Schalk, M.; Clark, A.; Formenti, L. R. and Nielsen, J. (2008). Reconstruction of a bacterial isoprenoid biosynthetic pathway in *Saccharomyces cerevisiae*. *FEBS Letters*, 582(29), 4032-4038. <https://doi.org/10.1016/j.febslet.2008.10.045>.
- Mojzita, D.; Wiebe, M.; Hilditch, S.; Boer, H.; Penttilä, M. and Richard, P. (2010). Metabolic engineering of fungal strains for conversion of D-galacturonate to meso-Galactarate. *Applied and Environmental Microbiology*, 76(1), 169-175. <https://doi.org/10.1128/AEM.02273-09>.
- Morgulis, A.; Coulouris, G.; Raytselis, Y.; Madden, T. L.; Agarwala, R. and Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics*, 24(16), 1757-1764. <https://doi.org/10.1093/bioinformatics/btn322>.
- Nielsen, J.; Larsson, C.; Van Maris, A. and Pronk, J. (2013). Metabolic engineering of yeast for production of fuels and chemicals. *Current Opinion in Biotechnology*, 24(3), 398-404. <https://doi.org/10.1016/j.copbio.2013.03.023>.
- Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C. and Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612. <https://doi.org/10.1002/jcc.20084>.
- Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L. and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781-1802. <https://doi.org/10.1002/jcc.20289>.

- Protzko, R. J.; Latimer, L. N.; Martinho, Z.; De Reus, E.; Seibert, T.; Benz, J. P. and Dueber, J. E. (2018). Engineering *Saccharomyces cerevisiae* for co-utilization of d-galacturonic acid and d-glucose from citrus peel waste. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-07589-w>.
- Putnik, P.; Bursać Kovačević, D.; Režek Jambrak, A.; Barba, F. J.; Cravotto, G.; Binello, A.; Lorenzo, J. M. and Shpigelman, A. (2017). Innovative “green” and novel strategies for the extraction of bioactive added value compounds from citruswastes: A review. *Molecules*, 22(5). <https://doi.org/10.3390/molecules22050680>.
- Richard, P. and Hilditch, S. (2009). D-Galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Applied Microbiology and Biotechnology*, 82(4), 597-604. <https://doi.org/10.1007/s00253-009-1870-6>.
- Rimola, A.; Alí-torres, J.; Rodríguez-Rodríguez, C.; Poater, J.; Matito, E. and Sodupe, M. (2011). Ab Initio Design of Chelating Ligands Relevant to Alzheimer’s Disease: Influence of Metalloaromaticity. *Journal of Physical Chemistry A*, 12659-12666.
- Rodríguez, R.; China, G.; López, N.; Pons, T. and Vriend, G. (1998). Homology modeling, model and software evaluation: three related resources. *Bioinformatics (Oxford, England)*, 14(6), 523-528. <https://doi.org/10.1093/bioinformatics/14.6.523>.
- Rorick, R.; Nahar, N. and Pryor, S. W. (2011). Ethanol production from sugar beet pulp using *Escherichia coli* KO11 and *Saccharomyces cerevisiae*. *Biological Engineering*, 3(4), 199-209. <https://doi.org/10.13031/2013.37124>.
- Roy, A.; Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725-738. <https://doi.org/10.1038/nprot.2010.5>.
- Schäfer, D.; Schmitz, K.; Weuster-Botz, D. and Benz, J. P. (2020). Comparative evaluation of *Aspergillus niger* strains for endogenous pectin-depolymerization capacity and suitability for d-galacturonic acid production. *Bioprocess and Biosystems Engineering*, 43(9), 1549-1560. <https://doi.org/10.1007/s00449-020-02347-z>.
- Shpigelman, A.; Kyomugasho, C.; Christiaens, S.; Van Loey, A. M. and Hendrickx, M. E. (2015). The effect of high pressure homogenization on pectin: Importance of pectin source and pH. *Food Hydrocolloids*, 43, 189-198. <https://doi.org/10.1016/j.foodhyd.2014.05.019>.
- Soni, N. and Madhusudhan, M. S. (2017). Computational modeling of protein assemblies. *Current Opinion in Structural Biology*, 44, 179-189. <https://doi.org/10.1016/j.sbi.2017.04.006>.



- Soto-Ospina, A.; Araque Marín, P.; Bedoya, G.; Sepúlveda-Falla, D. and Villegas Lanau, A. (2021). Protein Predictive Modeling and Simulation of Mutations of Presenilin-1 Familial Alzheimer's Disease on the Orthosteric Site. *Frontiers in Molecular Biosciences*, 8(June), 1-15. <https://doi.org/10.3389/fmolb.2021.649990>.
- Souffriau, B.; Den Abt, T. and Thevelein, J. M. (2012). Evidence for rapid uptake of d-galacturonic acid in the yeast *Saccharomyces cerevisiae* by a channel-type transport system. *FEBS Letters*, 586(16), 2494-2499. <https://doi.org/10.1016/j.febslet.2012.06.012>.
- Stoesser, G.; Baker, W.; Van den Broek, A.; Camon, E.; García-Pastor, M.; Kanz, C.; Kulikova, T.; Leinonen, R.; Lin, Q.; Lombard, V.; López, R.; Redaschi, N.; Stoehr, P.; Tuli, M. A.; Tzouvara, K. and Vaughan, R. (2002). The EMBL Nucleotide Sequence Database. *Nucleic Acids Research*, 30(1), 21-26. <https://doi.org/10.1093/nar/gki098>.
- Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J. and Schwede, T. (2019). QMEANDisCo – Distance Constraints Applied on Model Quality Estimation. *Bioinformatics*, (November), 1-7. <https://doi.org/10.1093/bioinformatics/btz828>.
- Tejada Moreno, J. A.; Villegas Lanau, A.; Madrigal Zapata, L.; Baena Pineda, A. Y.; Vélez Hernández, J.; Campo Nieto, O.; Soto Ospina, A.; Araque Marín, P.; Rishishwar, L.; Norris, E. T.; Chande, A. T.; Jordan, I. K. and Berrío, G. B. (2022). Mutations in SORL1 and MTHFDL1 possibly contribute to the development of Alzheimer's disease in a multigenerational Colombian Family. *PLoS ONE*, 17(7 July), 1-28. <https://doi.org/10.1371/journal.pone.0269955>.
- Tsukamoto, J.; Durán, N. and Tasic, L. (2013). Nanocellulose and bioethanol production from orange waste using isolated microorganisms. *Journal of the Brazilian Chemical Society*, 24(9), 1537-1543. <https://doi.org/10.5935/0103-5053.20130195>.
- UniProt Consortium. (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 40(2), D71-5. <https://doi.org/10.1093/nar/gks1068>.
- United Nations. (2022). *Food and Agriculture Organization of the United Nations. Statistical of Waste Disposal*. <https://www.fao.org/home/en/>.
- Van Maris, A. J. A.; Abbott, D. A.; Bellissimi, E.; Van den Brink, J.; Kuyper, M.; Luttik, M. A. H.; Wisselink, H. W.; Scheffers, W. A.; Van Dijken, J. P. and Pronk, J. T. (2006). Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: Current status. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 90(4), 391-418. <https://doi.org/10.1007/s10482-006-9085-7>.

- Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M. and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191. <https://doi.org/10.1093/bioinformatics/btp033>.
- Wilkins, M. R. (2009). Effect of orange peel oil on ethanol production by *Zymomonas mobilis*. *Biomass and Bioenergy*, 33(3), 538-541. <https://doi.org/10.1016/j.biombioe.2008.08.010>.
- Xu, D. and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, 101(10), 2525-2534. <https://doi.org/10.1016/j.bpj.2011.10.024>.
- Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J. and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12(1), 7-8. <https://doi.org/10.1038/nmeth.3213>.
- Zhang, J.; Liang, Y. and Zhang, Y. (2011). Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure*, 19(12), 1784-1795. <https://doi.org/10.1016/j.str.2011.09.022>.
- Zhang, L.; Thiewes, H. and Van Kan, J. A. L. (2011). The d-galacturonic acid catabolic pathway in *botrytis cinerea*. *Fungal Genetics and Biology*, 48(10), 990-997. <https://doi.org/10.1016/j.fgb.2011.06.002>.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40. <https://doi.org/10.1186/1471-2105-9-40>.