



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL  
INSTITUTO POLITÉCNICO NACIONAL**

Cinvestav Irapuato

**“Promiscuidad Bisustrato de la Isomerasa Fosforribosil Antranilato  
(PriA) en un Contexto de Modelos Metabólicos a Escala Genómica”**

Tesis que presenta

**Ing. Bioq. Ind. Naxhie López Reyes**

Para obtener el grado de

**Maestría en Ciencias**

con Especialidad en

**Biología de Plantas**

Co-dirección de Tesis

Dr. Martin Heil

Irapuato, GTO, México

Agosto 2023

El trabajo *in silico* desarrollado durante esta tesis de investigación, se realizó en el laboratorio de “Evolución de la Diversidad Metabólica” en la Unidad de Genómica Avanzada del Laboratorio Nacional de Genómica para la Biodiversidad (UGA-LANGEBIO) del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), y en la Universidad Autónoma Metropolitana, Unidad Cuajimalpa (UAM-C) en el “Laboratorio Virtual de Ingeniería de Procesos” parte del Dpto. de Procesos y Tecnología de la División de Ciencias Naturales e Ingeniería, en la CDMX; con importantes asesorías por parte del Dr. Roberto Olivares-Hernández y Dr. Francisco Barona-Gómez, además de contar con el apoyo del Dr. Martin Heil líder del laboratorio de “Ecología de Plantas” del Dpto. de Ingeniería Genética. Con financiamiento de la beca para estudios de posgrado por parte del Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) N° de CVU 1019506 y N° de Beca 772069.

"Engineers try to make things that do not exist in nature. Engineers stress invention."

*– Yuan-Cheng Fung*

"Science can amuse and fascinate us all, but it is engineering that changes the world."

*– Isaac Asimov*

|                         |     |
|-------------------------|-----|
| Agradecimientos.....    | i   |
| Índice de Figuras.....  | v   |
| Índice de Tablas.....   | v   |
| Abreviaturas.....       | vi  |
| Abreviaturas latín..... | vii |

# Índice

|  |    |
|--|----|
| Resumen.....   | 1  |
| Abstract.....  | 2  |
| 1. Introducción.....   | 3  |
| 1.1. Evolución de familias de enzimas.....   | 8  |
| 2. Antecedentes.....   | 13 |
| 2.1. Operón incompleto de Trp en Actinobacterias.....  | 13 |
| 2.2. Clases de genes <i>trpF</i> encontrados en Actinobacteria.....                            | 14 |
| 2.3. Regulación de las vías de síntesis de His y Trp en actinobacteria.....                    | 16 |
| 2.4. Modelo de promiscuidad enzimática bisustrato: enzima PriA.....                            | 17 |
| 2.5. Subfamilia enzimática de PriA.....  | 25 |
| 2.5.1. Transferencia horizontal de genes (SubHisA1).....                                       | 25 |
| 2.5.2. Co-existencia genómica de <i>trpF</i> y <i>priA</i> (PriB).....                         | 28 |
| 2.5.3. Pérdida de genes (SubHisA2 y SubTrpF).....  | 30 |
| 2.6. La promiscuidad enzimática para el diseño de rutas <i>de novo</i> .....                   | 31 |
| 2.7. El flujo metabólico relacionado con la evolución de genes.....                            | 33 |
| 3. Justificación.....  | 39 |
| 3.1. Formación de nuevas rutas metabólicas.....  | 42 |
| 4. Preguntas de Investigación.....   | 45 |
| 5. Objetivos.....  | 46 |
| 5.1. Objetivo General.....   | 46 |
| 5.2. Objetivos Específicos.....  | 46 |
| 6. Estrategia Experimental.....  | 47 |
| 6.1. Reconstrucción semiautomática de borradores de GSMM.....                                  | 47 |
| 6.1.1. Utilización de diferentes herramientas computacionales para la anotación funcional..... | 47 |
| 6.1.2. Fusión y combinación de los datos resultantes de las anotaciones funcionales.....       | 52 |

|        |  |     |
|--------|--|-----|
| 6.1.3. | Reconstrucción de borradores de GSMM con ModelSEED y “Gapfill” .....           | 53  |
| 6.1.4. | Validación temprana del borrador de GSMM .....                                 | 55  |
| 6.2.   | GSMM utilizados en este trabajo .....  | 57  |
| 6.3.   | Restricción de los GSMM con datos experimentales .....                         | 58  |
| 6.4.   | Análisis de sensibilidad de flujos relativos (Robustez y PhPP) .....           | 60  |
| 6.5.   | Fluxómica comparativa de L-aminoácidos .....                                   | 61  |
| 6.6.   | Mapas de flujos metabólicos.....   | 62  |
| 6.7.   | Contextos genómicos comparativos para los operones de His y Trp.....           | 62  |
| 7.     | Resultados.....  | 63  |
| 7.1    | Reconstrucción semiautomática de borradores de GSMM .....                      | 63  |
| 7.2.   | Comparación de análisis de sensibilidad de flujos relativos en tres GSMM ..... | 72  |
| 7.3.   | Fluxómica comparativa de L-aminoácidos .....                                   | 75  |
| 7.4.   | Mapas de flujos metabólicos.....   | 78  |
| 7.5.   | Contextos genómicos comparativos .....   | 83  |
| 8.     | Discusión de Resultados .....  | 85  |
| 9.     | Conclusiones .....   | 91  |
| 10.    | Perspectivas .....   | 93  |
| 11.    | Referencias Bibliográficas .....   | 95  |
| 12.    | Anexos.....  | 105 |
| 12.1.  | Anexo I.....   | 105 |
| 12.2.  | Anexo II.....  | 109 |
| 12.3.  | Anexo III.....   | 110 |
| 12.4.  | Anexo IV.....  | 112 |
| 12.5.  | Anexo V.....   | 115 |
| 12.6.  | Anexo VI.....  | 116 |

# Índice de Figuras

|   |    |
|---|----|
| <b>Figura. 2.2.1.</b> Filogenómica que muestra los tres subclados de clasificación de TrpF .....      | 16 |
| <b>Figura. 2.4.1.</b> Reordenamiento de Amadori .....   | 18 |
| <b>Figura. 2.4.2.</b> Estructura de enzima TIM o de barril de ocho subunidades $\beta/\alpha$ .....   | 21 |
| <b>Figura. 2.5.1.1.</b> Residuos catalíticos de PriA y SubHisA1 .....                                 | 27 |
| <b>Figura. 2.7.1.</b> <i>Flux balance analysis</i> (FBA).....   | 34 |
| <b>Figura. 2.7.2.</b> Principales pasos resumidos, involucrados en FBA.....                           | 36 |
| <b>Figura. 6.1.1.</b> Flujo de trabajo para la reconstrucción de 116 redes metabólicas de GSMM .....  | 57 |
| <b>Figura. 7.1.1.</b> Resumen de las características únicas que se encuentran en un objeto.....       | 64 |
| <b>Figura. 7.2.1.</b> Análisis de robustez para las reacciones ProFAR isomerasa y PRA isomerasa. .... | 74 |
| <b>Figura. 7.3.1.</b> Flujos calculados con FBA para la síntesis de dieciséis aminoácidos .....       | 77 |
| <b>Figura. 7.4.1.</b> Mapa de distribución de flujos metabólicos para <i>Sco</i> A (3)2.....          | 81 |
| <b>Figura. 7.4.2.</b> Mapa de distribución de flujos metabólicos para <i>Cgl</i> ATCC 13032.....      | 82 |
| <b>Figura. 7.5.1.</b> Contextos genómicos de <i>Sco</i> , <i>Mtu</i> y <i>Cgl</i> .....               | 84 |

# Índice de Tablas

|   |    |
|---|----|
| <b>Tabla 2.5.2.1.</b> Comparación de las eficiencias catalíticas.....           | 30 |
| <b>Tabla 7.1.1.</b> Borradores de GSMM obtenidos para los cuatro criterios..... | 70 |
| <b>Tabla. 7.3.1.</b> Reacciones y flujos de fluxómica comparativa.....          | 76 |

# Abreviaturas

**GSMM** · Modelos Metabólicos a Escala Genómica

***E. coli*** · *Escherichia coli*

**KEGG** · Kyoto Encyclopedia of Genes and Genomes

**GPR** · Gen-Proteína-Reacción

**PriA** · Isomerasa Fosforribosil Antranilato

**HisA** · Isomerasa 1-(5-fosforribosil)-5-[(5-fosforribosilamino)metilidenamino]imidazol-4-carboxamida

**TrpF** · Isomerasa N-(5'-fosforribosil)antranilato

**His** · L-Histidina

**Trp** · L-Triptófano

**FBA** · Análisis de Balance de Flujos

***Sco*** · *Streptomyces coelicolor*

***Mtu*** · *Mycobacterium tuberculosis*

***Cgl*** · *Corynebacterium glutamicum*

**PRA** · N'-(5'-fosforribosil)-antranilato

**THG** · Transferencia Horizontal de Genes

**LUCA** · Último Ancestro Común Universal

**BGC** · Agrupamiento de Genes Biosintético

**CDA** · Antibiótico Dependiente de Calcio

**ATP** · Adenosin Trifosfato

**PRPP** · Pirofosfato de Fosforribosilo

**ProFAR** · N'-[(5'-fosforribosil)-formimino]-5-aminoimidazol-4-carboxamida ribonucleótido

**PRFAR** · N'-[(5'-fosforribosil)-formimino]-5-aminoimidazol-4-carboxamida-ribonucleótido

**CdRP** · 1-(2-carboxi-fenilamino)-1'-desoxirribulosa-5'-fosfato

**TIM** · Triosa Fosfato Isomerasa

**PBS** · Sitio de Unión a Fosfato

**$k_{cat}$**  · Constante catalítica o número de rotación de una enzima

**$K_m$**  · Constante de Michaelis–Menten de una enzima

**EVB** · Simulaciones empíricas de enlaces de valencia

**ECA** · Escape del Conflicto Adaptativo

**SubHisA1** · Enzima subfuncionalizada mono funcional HisA1

**PriB** · Isomerasa Fosforribosil B

**Paur** · *Paenarthobacter aurences*

**SMg1** · *Streptomyces sp. Mg1*

**Ssvi** · *Streptomyces svisceus*

**Sipo** · *Streptomyces ipomoeae*

**DDC** · Duplicación-Degradación-Complementación

**SubHisA2** · Enzima subfuncionalizada monofuncional HisA2

**SubTrpF** · Enzima subfuncionalizada monofuncional TrpF

**PTS** · Sistema de Fosfotransferasa

**PEP** · Fosfoenolpiruvato

**DOE** · Departament of Energy

**CDS** · Secuencia de Codificación de Proteína

**CNN** · Redes Neuronales Convolucionales

**PAF** · Predicción Automática de la Función Enzimática

**SBO** · System Biology Ontology

**gDW** · Gramos de Peso Seco

**CGXII** · Medio de cultivo "*Corynebacterium glutamicum XII*"

**I + D** · Investigación y Desarrollo

**ARN** · Ácido Ribonucleico

## Abreviaturas latinas

**et al.** · *et alii* [y colaboradores]

**i.e.** · *id est* [esto es; es decir]

**e.g.** · *exempli gratia* [por ejemplo]

**sc.** · *scilicet* [respectivamente]

**ca.** · *circa* [alrededor de; cerca de]

**v. gr.** · *verbi gratia* [por ejemplo]

# Resumen

La isomerasa fosforribosil antranilato (PriA) es crucial en las rutas de biosíntesis de L-Histidina (His) y L-Triptófano (Trp), está ampliamente distribuida en el *phylum* Actinobacteria. PriA es una enzima promiscua, isomeriza dos sustratos diferentes en el mismo sitio catalítico, llevando a cabo dos funciones (otorgadas usualmente por dos genes diferentes: *hisA* y *trpF*). Los episodios evolutivos de subespecialización en una de estas funciones están documentados en la familia de enzimas PriA. Sin embargo, hasta antes del presente trabajo no se contaba con análisis acerca de si la promiscuidad enzimática podía ser modelada computacionalmente en los Modelos Metabólicos a Escala Genómica (GSMM) y cuál era la repercusión a nivel de biología de sistemas del “*metabolismo subterráneo*” (metabolismo que hace uso de enzimas promiscuas con reacciones secundarias) en los subsistemas de biosíntesis de L-aminoácidos. Los GSMM, herramientas de biología de sistemas utilizadas en biología sintética, son modelos matemáticos denominados estequiométricos que facilitan la investigación ya que predicen a nivel *in silico* una variedad de procesos metabólicos llevados a cabo por organismos de los tres dominios de la vida, otorgan una verdadera ventaja en cuanto a tiempo y experimentación, ya que intentan modelar computacionalmente metabolismos completa o parcialmente. En especial, los metabolismos de las bacterias son cambiantes ante la amplia diversidad del mundo físico y natural en los que se encuentran, la adaptación a un medio ambiente cambiante es una característica de los organismos donde la evolución enzimática toma un papel importante. PriA es considerada una enzima modelo en el estudio de la evolución enzimática. En este trabajo se realizó la reconstrucción semiautomática de borradores de GSMM dentro de la plataforma *Systems Biology Knowledgebase* “Kbase” a partir de veintinueve genomas actinobacterianos en los cuales se ha reportado con anterioridad la presencia de la enzima PriA y sus subespecializaciones con diferentes grados de promiscuidad. Lo anterior fue complementado, con análisis de contextos genómicos con el algoritmo CORASON; análisis de sensibilidad de flujos relativos como lo es “robustez” y fluxómica comparativa, con la herramienta COBRA Toolbox 3.0, utilizando GSMM previamente reportados de actinobacterias modelo que contienen a PriA y una subespecialización de esta. Los resultados mostraron que los flujos de biosíntesis de His son mayores en comparación con los flujos de biosíntesis de Trp, en las actinobacterias seleccionadas. En los análisis de robustez se observó un comportamiento inversamente proporcional cuando se variaron en magnitud los flujos de síntesis de las dos reacciones involucradas en la promiscuidad de PriA. Se llegó a la conclusión que los GSMM sí pueden modelar la promiscuidad enzimática a nivel de los subsistemas de biosíntesis de L-aminoácidos, lo cual depende de la anotación de la promiscuidad de la enzima en los genomas, lo cual debe ir acompañado de una adecuada y refinada reconstrucción del GSMM asegurando su calidad; si lo anterior no se cumple, los modelos difícilmente simularan *de novo* la promiscuidad enzimática.

# Abstract

The isomerase phosphoribosyl anthranilate (PriA) is crucial in the biosynthetic pathways of L-Histidine (His) and L-Tryptophan (Trp), it is widely distributed across the *phylum* Actinobacteria. PriA is a promiscuous enzyme, it isomerizes two different substrates at the same catalytic site, carrying out two functions (usually given by two different genes: *hisA* and *trpF*). Evolutionary episodes of subspecialization in one of these functions are documented within the PriA family enzymes. However, until before this work, analyzes about the incorporation of enzymatic promiscuity in the Genome Scale Metabolic Models (GSMM) was not done and how these models could reflect the enzyme promiscuity genomically annotated, nor which was the repercussion at a systems biology level of the “*underground metabolism*” (metabolism that uses promiscuous enzymes with secondary reactions) in L-aminoacid biosynthesis. GSMMs are systems biology tools used in synthetic biology, these stoichiometric mathematical models facilitate and predict *in silico* level a variety of metabolic processes carried out by organisms from the three domains of life. These models provide a real advantage in terms of time and experimentation, as they aim to modeling complete or partial pathways within metabolism, wich is specific to each organism and is constantly changing in the face of a wide diversity of the physical and natural world. Adaptation to a changing environment is a characteristic of organisms where enzymatic evolution plays an important role. PriA is considered a model enzyme in the study of enzymatic evolution. In this work, the semiautomatic reconstruction of drafts of GSMM was carried out from twenty-nine actinobacterial genomes in which the presence of the promiscuous enzyme PriA has been previously reported, within the Systems Biology Knowledgebase "Kbase" platform. These GSMM were further analyzed by looking into genomic contexts, carried out with the CORASON algorithm; sensitivity analysis of relative flux such as "robustness" and comparative fluxomics in three actinobacterial model organisms (containing PriA), with COBRA Toolbox 3.0, analyzing their metabolisms with previously reported GSMM. The results show that “His” biosynthetic fluxes were higher compared to Trp biosynthetic fluxes, in all selected actinobacteria. In robustness analysis, an inverse proportional behavior was observed when synthesis fluxes of two PriA reactions fluctuated in magnitude. It was concluded that the GSMM can model enzyme promiscuity at the subsystem level of L-aminoacid biosynthesis, which depends on the annotation of enzyme promiscuity in the genomes, which must be followed by an adequate and refined reconstruction of the GSMM ensuring its quality; if the above is not accomplished, the models hardly will simulate the enzymatic promiscuity *de novo*.

# 1. Introducción

La “materia oscura” metabólica abarca procesos bioquímicos de los cuales el conocimiento aún es escaso, lo que limita la comprensión absoluta del metabolismo, e.g. el descubrimiento de mecanismos de enfermedades como el cáncer, ante la secuenciación de células tumorales [Lewis & Abdel-Haleem 2013] y el desarrollo de medicamentos derivados de productos naturales de plantas como la vinblastina [Zhang et al. 2022], son ejemplos que han logrado dilucidar parte de este metabolismo oscuro.

Las brechas en el conocimiento metabólico obstaculizan el avance de aplicaciones de bioingeniería, como la creación de fábricas de células las cuales tienen un trasfondo sustentable para la producción ecológica de productos químicos y farmacéuticos. Los principales ejemplos de materia oscura metabólica incluyen *i)* el metabolismo subterráneo resultante de la actividad enzimática promiscua [Rosenberg & Commichau, 2019], *ii)* los productos naturales de plantas que aún permanecen desconocidos y sus vías de biosíntesis no caracterizadas, y *iii)* el daño celular de los metabolitos [Haas et al. 2022].

Por lo que resulta fundamental generar hipótesis sobre posibles funciones bioquímicas que guíen el descubrimiento experimental de funciones enzimáticas y productos naturales.

Los modelos metabólicos a escala genómica (GSMM, por sus siglas en inglés) están posicionados desde las últimas décadas, como herramientas matemáticas estequiométricas útiles y versátiles, que facilitan y predicen a nivel *in silico* una variedad de procesos metabólicos llevados a cabo por organismos vivos de los tres dominios de la vida, otorgando una ventaja importante en cuanto a tiempo y experimentación.

Estas herramientas de biología de sistemas utilizadas en biología sintética intentan modelar metabolismos completos o parciales, con extenuantes curaciones de los metabolitos celulares y reacciones enzimáticas; son específicos y diferentes para cada organismo, así como lo son las diferencias genómicas entre ellos; los metabolismos de cada individuo son cambiantes ante la amplia diversidad del mundo físico y natural en los que se encuentran en un determinado momento

(adaptabilidad), estas condiciones ambientales también pueden ser modelables, al restringir los modelos computacionalmente con datos obtenidos experimentalmente [Guzmán et al. 2015].

Los GSMM permiten llevar a cabo un análisis cuantitativo de la relación entre fenotipo y genotipo por medio de la integración de “*Big Data*” de origen ómico (e.g. datos genómicos, metabolómicos, transcriptómicos, epigenómicos, traductómicos, interactómicos y proteómicos), los cuales se han incorporado para obtener una comprensión sistémica que abarca bioprocesos completos [Chung et al. 2021; Passi et al. 2021].

Si bien los diferentes tipos de datos ómicos pueden formar una base para generar importantes hipótesis sobre la materia oscura metabólica, con frecuencia es difícil vincular estos datos con las funciones metabólicas [Galpering & Koonin, 2010], y persisten importantes lagunas en el conocimiento bioquímico y metabólico. e.g. el 25 % de las proteínas de *Escherichia coli* (*E. coli*), uno de los organismos modelo mejor estudiados, no tienen una función asignada [Keseler et al. 2021]; y casi 10,000 metabolitos en la *Kyoto Encyclopedia of Genes and Genomes* (KEGG) se consideran huérfanos, lo que significa que no están integrados en ninguna reacción bioquímica dentro de KEGG [Kanehisa et al. 2023; Mohammadi Peyhani et al. 2022].

Los GSMM básicos intentan coleccionar toda la información metabólica disponible del sistema biológico que se desea estudiar, recopilando el conocimiento de las vías catabólicas y anabólicas. Cada reacción catalizada por una enzima codificada en el genoma del organismo a manera de secuencias de proteínas crea en los GSMM las reglas de asociación *gen-proteína-reacción* (GPR). Por medio de los balances de materia de cada metabolito que compone el conjunto de reacciones enzimáticas, y la suposición de un estado pseudo-estacionario (i.e. el flujo de metabolitos “sustrato” que entran a la reacción enzimática es el mismo que el flujo de metabolitos “producto” que salen de la reacción), se puede obtener un modelo matemático que consiste en un conjunto de ecuaciones lineales que se representan en una matriz escasa cuyas entradas son los coeficientes estequiométricos de las reacciones que componen el metabolismo (*para mayor detalle ver la sección 6.1.4.*).

Estos modelos matemáticos de la red metabólica proporcionan predicciones cuantitativas de los flujos metabólicos los cuales son relacionados con el crecimiento o la aptitud celular en función de las asociaciones GPR. Los GSMM pueden integrar de manera efectiva los diferentes tipos de datos

ómicos, para validar redes metabólicas que se pueden usar en tres amplios aspectos: *i*) comprender el metabolismo de los sistemas biológicos, y su interacción con otros organismos (e.g. organismos patógenos, ecología microbiana e interacción entre especies); *ii*) identificar dianas terapéuticas potenciales en organismos patógenos; y *iii*) el diseño de sistemas biológicos con características manipulables preferenciales que de otro modo no existirían en la naturaleza (ingeniería metabólica) [Passi et al. 2021].

Los GSMM ayudan a comprender los mecanismos moleculares en un organismo e identifican nuevos procesos que podrían ser contrarios al conocimiento aparentemente obvio del fenómeno biológico en cuestión; también han facilitado la comprensión de la fisiología y estructura biológica a escala genómica de diversos microorganismos [Van't Hof et al. 2022].

En este trabajo de tesis se utilizaron los GSMM de un modo poco común, al evaluar el fenómeno biológico de promiscuidad enzimática ante diferentes sustratos en la biosíntesis de L-aminoácidos por miembros del *phylum* Actinobacteria, en específico el caso de la enzima promiscua PriA, denominada así por Barona-Gómez & Hodgson (2003), esta enzima isomeriza dos sustratos diferentes en pasos intermedios de las rutas de biosíntesis de His y Trp.

En primera instancia se reconstruyeron veintinueve borradores de GSMM básicos semiautomáticamente, haciendo uso de herramientas como: Kbase, anotación funcional automática (como RAST, KEGG, Prokka, y algunas otras anotaciones que hacen uso de “*Machine & Deep learning*” como DeepEC y NetGo 3.0), reconstrucción con ModelSeed y una validación de calidad con MEMOTE. Ahora se dispone de dichos borradores de veintinueve bacterias pertenecientes al *phylum* Actinobacteria para una posterior curación manual y análisis que superan a este trabajo de tesis; dicho ejercicio me permitió relacionarme con la reconstrucción de las redes metabólicas de los GSMM y apreciar la exigente tarea que es la obtención de un solo modelo matemático [Thiele & Palsson, 2010].

Por otra parte, se exploró la promiscuidad enzimática computacionalmente con *Análisis de Balance de Flujos* (FBA, por sus siglas en inglés) [Orth et al. 2010], empleando la herramienta COBRA Toolbox 3.0 en la plataforma de programación MATLAB® 2022 [Heirendt et al. 2019], utilizando tres GSMM que se encontraban reportados y curados manualmente al comienzo de esta tesis. Dichos GSMM se encuentran hasta la fecha, y dependiendo del modelo, entre su quinta o sexta generación

de actualización de curación, siendo afinados y mejorados progresivamente, publicándose en cada ocasión.

Los GSMM de las cepas actinobacterianas del organismo modelo *Streptomyces coelicolor* (*Sco*) A3(2) y el patógeno *Mycobacterium tuberculosis* (*Mtu*) H37Rv, además del microorganismo catalogado como el “caballo de batalla” de la sobreproducción de L-aminoácidos en biotecnología industrial *Corynebacterium glutamicum* (*Cgl*) ATCC 13032, fueron considerados para este trabajo de tesis.

Para dichos organismos se cuenta con GSMM, estos últimos representan un beneficio en el ámbito científico y tecnológico, debido a que se trata de cepas bacterianas ampliamente utilizadas en novedosos procesos biotecnológicos de biosíntesis de productos de alto valor agregado (al menos *Sco* y *Cgl*), muy útiles en la industria química, farmacéutica, y agroalimentaria, proporcionando la oportunidad de evaluar las condiciones ambientales y genéticas en un marco predictivo computacional; cabe recalcar que dichas cepas seleccionadas tienen parámetros cinéticos reportados de la enzima promiscua PriA y una subespecializada [Noda-García et al. 2010; Due et al. 2011; Noda-García et al. 2013].

Además del tipo de promiscuidad de sustrato; las enzimas también pueden exhibir promiscuidad de reacción y promiscuidad de cofactores; e.g. una aplicación biotecnológica del uso de sitios promiscuos existe en un mecanismo alternativo al usualmente conocido, de inhibición de la metanogénesis en arqueas metanogénicas con la administración del fármaco lovastatina (metabolito especializado principalmente del género *Aspergillus*) en rumiantes y seres humanos. La lovastatina es capaz de inhibir dos enzimas intermedias en la vía metanogénica, interfiriendo en la adecuada biosíntesis de metano; lo anterior debido a que la lovastatina se une promiscuamente al sitio de unión del cofactor F<sub>420</sub> de las dos enzimas con una afinidad mucho mayor a la observada con su cofactor natural F<sub>420</sub> [López-Reyes et al. 2022].

Se ha reportado que la enzima PriA existe como una enzima generalista intermedia bifuncional (*dicho concepto se describe con mayor detalle más adelante en la sección 1.1*) en ciertos organismos del *phylum* Actinobacteria, debido a que algunos géneros miembros de este *phylum* han perdido en su trayectoria evolutiva al gen *trpF* y por consecuencia a su enzima codificada N’-(5'-fosforribosil)-

antranilato (PRA) isomerasa, la cual es parte del operón completo de Trp encontrado en el *phylum* proteobacteria específicamente en la clase  $\gamma$ -proteobacteria.

Ante la carencia del gen *trpF*, los organismos adaptaron un homólogo evolucionado del gen *hisA* (~50 % identidad de secuencia) involucrado hasta entonces solamente en la ruta de síntesis de His, esto fue demostrado con análisis filogenéticos [Noda-García et al. 2015].

HisA evolucionó hacia la bifuncionalidad de catálisis e isomerización de dos sustratos diferentes (promiscuidad de sustrato), evolucionada hacia la enzima bifuncional PriA, la cual lleva a cabo la misma reacción enzimática de oxidorreducción intramolecular, tanto en la vía de síntesis de Trp, como en la vía de síntesis de His, ambas reacciones involucran la isomerización del fragmento fosforribosil idéntico en ambos sustratos [Fig. 2.4.2.b].

Las enzimas de los genes *trpF* e *hisA* exhiben estructuras de barril  $(\beta/\alpha)_8$  similares. Jürgens & cols. (2000) generaron mutantes en el gen *hisA*, las enzimas derivadas de la mutagénesis y la selección aleatoria podían catalizar la reacción TrpF tanto *in vitro* como *in vivo* al insertar el gen *hisA* mutado en *E. coli*, revelando que un solo intercambio de aminoácidos podía establecer la actividad de TrpF en el andamio HisA.

Por otra parte, se ha reportado que especies de género *Chlamydia* (*phylum* Chlamydiota) contienen un homólogo de TrpF con función GTP ciclohidrolasa I (GCH1), la cual isomeriza guanosina-5'-trifosfato (GTP) hacia 7,8-dihidroneopterina 3'-trifosfato en la ruta de biosíntesis de tetrahidrofolatos, ante la ausencia del gen *folE* que codifica una enzima de síntesis de folato. Estas especies han perdido otros genes de síntesis del operón de Trp (por lo que no sintetizan Trp), lo que indica que su actividad TrpF se conserva como una función promiscua, en lugar de biológica [Adams et al. 2014].

Existen diversos miembros de la familia enzimática PriA encontrados en diferentes organismos del *phylum* actinobacteria, que han evolucionado en diferente medida hacia una subespecialización de PriA, lo que conlleva a la denominación de subfamilias PriA, cada una con diferente rol de afinidad por los sustratos involucrados y origen evolutivo, estas subfamilias se han denominado enzimas: PriB, SubHisA1, SubHisA2 y SubTrpF.

En este trabajo de tesis se evaluó una de las subfamilias (SubHisA1) y la PriA bonafide con cada uno de los GSMM seleccionados. *Sco* y *Mtu* contienen en su genoma la enzima PriA bonafide con un comportamiento completamente bifuncional promiscuo; *Cgl* representa la subfamilia de subfuncionalización SubHisA1, esta última es una enzima con función HisA monofuncional con afinidad por un único sustrato, habiendo perdido la capacidad de isomerizar el sustrato PRA, ocasionado por la adquisición de un gen *trpF* en el genoma de *Cgl* por transferencia horizontal de genes (THG) proveniente de las  $\gamma$ -proteobacterias, completando el operón incompleto de Trp que predomina en las actinobacterias.

## 1.1. Evolución de familias de enzimas

Las proteínas han estado evolucionando durante ca. 3,700 millones de años, fragmentos de proteínas muy pequeños y funcionales pueden haber sido reclutados antes de la aparición del proteoma del último ancestro común universal (LUCA, por sus siglas en inglés) para dar lugar a sistemas metabólicos primitivos [Longo et al. 2020; Jayaraman et al. 2022].

La evolución de familias de enzimas son secuencias de proteínas que divergen con el tiempo, estos cambios pueden estar relacionados con la *deriva génica* o la *adaptación*; en la *deriva génica* los cambios en la secuencia se producen debido a una distinción de células bacterianas aleatoria en la misma población (i.e. algunas contendrán mutaciones otras no), en la enzima en evolución se preserva la estructura y la función de la proteína, llevando a cabo una selección de purificación o purga la cual puede impulsar una reducción en la frecuencia de ciertas mutaciones (alelos) que reducen la actividad con sustratos promiscuos dentro de una población determinada, esta última suele denominarse “selección negativa”; en cambio en la *adaptación* ocurren cambios en las propiedades de las proteínas, incluida la adquisición de nuevas actividades bioquímicas en una “selección positiva” donde existe el enriquecimiento de algunas mutaciones que aumentan la unión o la eficiencia catalítica con el ligando/sustrato nuevo o deseado (esta cualidad es ampliamente utilizada en ingeniería de proteínas). La *selección* da forma a los rasgos de las proteínas, incluida su actividad bioquímica (e.g. afinidad,

catálisis, alosterismo y selectividad) y propiedades biofísicas (e.g. plegamiento y termoestabilidad) [Tawfik, 2014].

Las transiciones en la evolución de las proteínas se pueden clasificar en *microtransiciones* donde existe divergencia de nuevas funciones mientras se mantiene la arquitectura original (plegamiento) y las características importantes del sitio activo (divergencia dentro de familias y superfamilias de proteínas) como ocurre en el caso de la familia de enzimas PriA con sus subespecializaciones (*ver sección 2.5*); y en *macrotransiciones* en las cuales las transiciones ocurren entre diferentes plegamientos, incluidas la aparición de los primeros plegamientos de proteínas [Longo et al. 2020; Jayaraman et al. 2022].

La duplicación de genes proporciona materia prima para nuevas proteínas y puede proporcionar una ventaja adaptativa *per se*, al aumentar la carga de proteína y, por lo tanto, aumentar una función promiscua débil preexistente; los genes duplicados pueden evolucionar hacia una nueva función enzimática (neofuncionalización) que no estaba presente en el gen ancestral pre-duplicado como ocurre en *Sco* ante el reclutamiento en un agrupamiento de genes biosintéticos (BGC, por sus siglas en inglés) de un homólogo de *priA* involucrado en la biosíntesis de un antibiótico dependiente de calcio (CDA, por sus siglas en inglés) (Hodgson, 2000). Por otra parte, a partir de un ancestro bifuncional (generalista o promiscuo) pueden surgir dos genes especialistas (subfuncionalización o divergencia antes de la duplicación) [Näsvall et al. 2012].

Es importante señalar que la evolución de una enzima divergente, no se trata solo de mejorar la función catalítica. A menudo también se requiere la evolución de una nueva regulación transcripcional, ya que la regulación transcripcional de la enzima ancestral frecuentemente será inapropiada para la nueva enzima divergente [Noda-García et al. 2017]. La transcripción de genes parálogos es general y sustancialmente diferente [Li et al. 2021].

La promiscuidad, tema central de esta tesis, se relaciona con la preexistencia coincidente de funciones catalíticas que pueden servir de punto de partida para nuevas funciones. Si tales funciones

latentes y promiscuas se someten a selección, dan lugar a intermediarios generalistas bifuncionales (e.g. enzima bisustrato PriA). Tras la duplicación de genes, los intermedios generalistas se dividen, dando lugar a dos especialistas, cada uno de los cuales realiza una función (subfuncionalización) [Khersonsky & Tawfik 2010]. Aunque la duplicación y el paso de *generalistas* a *especialistas* es una tendencia general, también ocurre el proceso inverso de pérdida de genes y el paso de *especialistas* a *generalistas* [Noda-García et al. 2018].

Los efectos de las mutaciones no solo en diferentes genes, sino también dentro del mismo gen/proteína pueden ser no aditivos, es decir, epistáticos. La epistasis tiene un profundo impacto en la evolución en general y en la evolución de las proteínas en particular. Un predominio de la epistasis significa que muchas (probablemente la mayoría) de las mutaciones que eventualmente se fijan en las proteínas en evolución son perjudiciales por sí mismas (durante la deriva génica y la adaptación) [Starr & Thornton, 2016]. Por lo tanto, su fijación puede ocurrir a través de dos mecanismos alternativos: una mutación deletérea se acumula transitoriamente y luego es seguida por una mutación compensadora; alternativamente, las mutaciones que se acumulan inicialmente como neutras permiten que las mutaciones deletéreas se reparen en una etapa posterior (mutaciones habilitadoras, permisivas) [Tóth-Petróczy & Tawfik, 2013].

La habilitación y la compensación (y, por lo tanto, la *epistasis*) pueden ser: *locales* o *específicas*, i.e. las mutaciones deletéreas compensatorias y habilitantes ocurren en un par específico de residuos (típicamente, en dos residuos en contacto, e.g. dentro de los sitios activos); también pueden ser *globales* o *inespecíficas* donde una mutación dada puede habilitar o compensar una variedad de mutaciones perjudiciales diferentes (e.g. mutaciones estabilizadoras que pueden compensar muchas mutaciones desestabilizadoras diferentes) [Dellus-Gur et al. 2015].

Las mutaciones casi siempre afectan a más de un rasgo de la proteína (pleiotropía) y, a menudo, de formas aberrantes. La *epistasis* y las *compensaciones* son elementos clave que dan forma a las trayectorias de la evolución de las proteínas [Starr & Thornton, 2016; Miton & Tokuriki, 2016].

Se conocen varios tipos de *compensaciones* evolutivas con respecto a las proteínas, v. gr. en la compensación entre una función original y una nueva función es probable que una mutación que mejora a una nueva función en evolución disminuya la función original. e.g. tres mutaciones que aumentan la actividad de TrpF en la enzima bifuncional ineficiente HisA/TrpF (*hisA* [dup13-15/D10G]) descrita por Newton & cols. (2017) disminuye 35 veces la actividad de HisA, a un nivel indetectable. La fuerte compensación entre las actividades HisA y TrpF minimiza la inhibición de la actividad TrpF en evolución por parte del sustrato nativo para HisA en células en las que están presentes ambos sustratos. Sin embargo, las actividades promiscuas a veces pueden mejorarse sustancialmente sin perjudicar significativamente la actividad nativa [Copley, 2020].

Una fuerte compensación impone una neofuncionalización, i.e. la duplicación debe ocurrir para completar la divergencia y la especialización (llamado; *escape del conflicto adaptativo*) [Sikosek et al. 2012]. En muchos casos, esta compensación es inicialmente débil, lo que permite la divergencia hacia un intermedio generalista bifuncional (podría explicar el origen de PriA). La magnitud de las compensaciones originales y nuevas tiende a variar a lo largo de las trayectorias adaptativas, comenzando con compensaciones débiles que dan lugar a cambios generalistas intermedios, cambiando eventualmente a fuertes compensaciones a medida que avanza la selección, produciendo así un nuevo especialista (típicamente después de la duplicación) [Mallik et al. 2022].

En el tipo de compensación entre estabilidad y actividad, la mayoría de las mutaciones disminuyen la estabilidad de la proteína y, por lo tanto, conducen a un plegamiento incorrecto, la agregación y/o la proteólisis. Las mutaciones de una nueva función/actividad son aún más desestabilizantes, lo que hace que su acumulación dependa de mutaciones habilitadoras/compensatorias [Tóth-Petróczy & Tawfik, 2013].

En cuanto al tipo de compensación de velocidad y afinidad, una mutación que mejora la eficiencia catalítica de una enzima puede reducir su especificidad; de manera similar, la mejora en la afinidad hacia un ligando también puede aumentar la reactividad cruzada con otros ligandos diferentes [Noda-García et al. 2018].

Las optimizaciones evolutivas, incluida la optimización de proteínas, están sujetas a fuertes rendimientos decrecientes, i.e. las mutaciones tempranas otorgan grandes ventajas por mutación, pero a medida que mejora el nuevo rasgo en la evolución, la mejora por mutación disminuye [Tokuriki et al. 2012].

Es posible que las mutaciones genómicas necesarias para conferir una nueva función no estén presentes en una población; sin embargo, pueden surgir por mecanismos no genéticos mediados por errores en la replicación, transcripción y traducción (mutaciones fenotípicas). Por lo tanto, la nueva próxima función ya está presente, como una variación fortuita y latente a nivel fenotípico dentro de genotipos idénticos (variabilidad fenotípica); la variación fenotípica existe en una población genéticamente idéntica debido al ruido asociado con los procesos biológicos no genéticos mencionados, los cuales se observan en todos los niveles biológicos de organización, desde proteínas individuales hasta organismos completos [Rotem et al. 2010].

La preexistencia de actividades proteicas como funciones promiscuas latentes es actualmente una hipótesis bien establecida y entendida a nivel atómico. Existen aspectos adicionales a la variabilidad fenotípica que subyacen a la aparición del más apto. Una hipótesis, es que la variabilidad fenotípica (y posiblemente también los cambios genéticos) es inducida directamente por desafíos ambientales [Levis & Pfennig, 2016].

## 2. Antecedentes

### 2.1. Operón incompleto de Trp en Actinobacterias

Los operones son grupos de genes bajo el control de las mismas secuencias promotoras, son un sello distintivo de la estructura del genoma procariota y la regulación génica. Una proporción significativa de los genes del genoma procariota están organizados en operones, es ampliamente aceptado que los operones representan una ventaja por múltiples razones, incluida una mejor co-regulación de los genes [Shieh et al. 2015].

Los operones se forman por reordenamientos, incluidos inserciones, deleciones y duplicaciones, las últimas pueden explicar la evolución de algunos operones como el de His, pero los genes en los operones no necesariamente tienen estructuras similares. Los pares de genes (duplicaciones) pueden terminar en el mismo operón mediante reordenamientos aleatorios, sin embargo, es más probable que terminen separados.

Ha sido ampliamente reportado que el gen *trpC* del operón de biosíntesis de Trp está fusionado con el gen *trpF* en las  $\gamma$ -proteobacterias. Muchos de los dominios catalíticos de las enzimas que forman parte del operón de Trp exhiben una clara homología en la identidad de aminoácidos de proteínas que tienen diferentes especificidades de sustrato y participan en diferentes vías. Desde una perspectiva evolutiva, esto es de interés con respecto a cómo se han ensamblado las enzimas de la ruta de Trp (e.g. a través de la duplicación de genes y la alteración del sustrato) mediante el reclutamiento de homólogos de otras rutas o que la ruta Trp ha sido fuente de genes reclutados para funcionar en otras vías, o inclusive que un gen homólogo con un historial reciente de función en otra vía se ha recombinado para reemplazar un gen de la vía Trp o viceversa.

En el *phylum* Actinobacteria, con los primeros hallazgos en las secuencias de los genomas completos de las bacterias *Sco* y *Mtu*, se detectó que el fragmento del gen *trpF* no se encontraba presente entre sus secuencias [Hodgson, 2000; Barona-Gómez & Hodgson, 2003]. En la mayor parte de los genomas conocidos del *phylum* Actinobacteria, dentro del operón de Trp, *trpF* es un gen ausente

en ciertos clados de este linaje [Xie et al. 2003]. Se ha demostrado que un gen similar a *hisA*, denominado *priA*, codifica una enzima con especificidad dual en *Sco*, lo que genera una isomerasa de biosíntesis bifuncional en las rutas de síntesis tanto de His como de Trp [Barona-Gómez & Hodgson, 2003], como anteriormente he mencionado.

A partir de este hallazgo, se comenzó a investigar en más genomas del *phylum* Actinobacteria que carecían del gen *trpF* si también cumplían con esta bifuncionalidad de la enzima HisA en los pasos intermedios de la biosíntesis de los dos L-aminoácidos involucrados.

Se ha reportado que las estructuras de los operones, como la del operón Trp, son inestables, como se dedujo de observaciones de inserciones atípicas, dispersión de genes y alteración de la codificación del orden de los genes dentro de los operones. La caracterización de los operones como inestables connota susceptibilidad al deterioro, esto implicaría que las ventajas selectivas conferidas por tales operones deben ser débiles [Itoh et al. 1999].

## 2.2. Clases de genes *trpF* encontrados en Actinobacteria

Los homólogos del gen *trpF*, faltante en algunos clados de Actinobacteria que es complementado con el homólogo de *hisA* bisustrato (*priA*), con base en análisis genómicos comparativos y filogenéticos, se pueden categorizar de acuerdo con Verduzco-Castro & col. (2016) de la siguiente manera:

- I) *trpF-1*: subclado que representa al gen de biosíntesis de Trp xenólogo (*trpF*) corinebacteriano presente en las  $\gamma$ -proteobacterias adquirido por THG, por consecuencia en algunas especies del género *Corynebacterium* co-ocurre una PriA subfuncionalizada (SubHisA1), que ha perdido su función de PRA isomerasa [Noda-García et al. 2013; Noda-García et al. 2015].
- II) *trpF-2*: subclado que solo contiene genes parálogos de *trpF* de un limitado número de géneros actinobacterianos, entre ellos *Nocardia*, *Actinokineospora* y algunos *Streptomyces*. Los análisis de los contextos genómicos mostraron sintenia en el entorno de estos

homólogos de *trpF*, los loci incluyen homólogos biosintéticos de Trp (*trpD*, *trpA* y *trpC*), así como un gen homólogo de *ectC* anotado funcionalmente como L-ectoína sintasa, que juntos forman un operón. Este operón se encuentra dentro de un BGC de productos naturales, que en lugar de sintetizar Trp o L-ectoína, dirige la biosíntesis de un fragmento pirrolo-pirrol en la actinobacteria *Micromonospora sp.* TP-A0468 (M468) [Ma et al. 2013]. Este homólogo de la enzima similar a TrpF (gen *kstC3* o *trpF-2*) hasta donde se ha reportado, cataliza una isomerización abriendo el anillo fosforribosil de 4-*N'*-(5'-fosforribosil) 4-aminopirrol-2-carboxilato.

- III) *trpF-3*: subclado que incluye genes ortólogos variables no conservados de *trpF-1*, se encuentra en géneros lejanamente relacionados, como *Rhodococcus*, *Arthrobacter* o algunos *Streptomyces*. Al analizar la conservación de genes se reveló que casi todas las especies de *Arthrobacter* (100 %) y *Rhodococcus* (95 %) conservan un gen *trpF*, en *Streptomyces* el grado de conservación se reduce al 25 %. La presencia diferencial de este gen *trpF* sugiere un origen ancestral, es una enzima TrpF ortóloga. En el análisis del contexto genómico, los loci de *trpF* revelan falta de sintenia y ausencia adyacente de otros genes Trp. En la mayoría de las especies de *Streptomyces*, incluidas las que contienen un gen *trpF-3*, los genes *trp* tienden a agruparse con los de His. Existe la co-aparición de *trpF-3* y *priA* en estos géneros de actinobacterias. Con base en análisis cristalográficos estructurales y parametrización cinética se concluyó que la subfamilia de homólogos *priA* presentes en organismos cuyos genomas codifican un gen *trpF-3* se denominaría PriB (de fosforribosil isomerasa B) la cual tiene una disminución de la eficiencia catalítica de PRA isomerasa en varios ordenes de magnitud en comparación con la PriA bonafide [Verduzco-Castro et al. 2016].

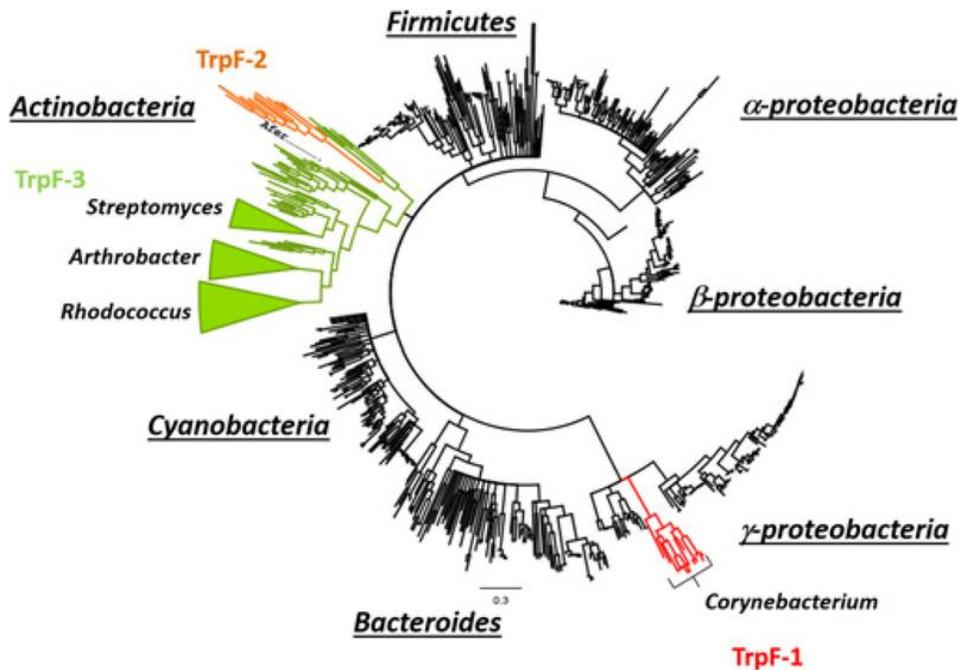


Figura. 2.2.1. Filogenómica que muestra los tres subclados de clasificación de TrpF con diferentes orígenes evolutivos, encontrados en diferentes grupos bacterianos. [Acorde con Verduzco-Castro et al. 2016].

### 2.3. Regulación de las vías de síntesis de His y Trp en actinobacteria

La biosíntesis de Trp es la ruta de L-aminoácidos más costosa desde el punto de vista bioquímico ya que requiere la entrada de eritrosa-4-fosfato, ATP, pirofosfato de fosforribosilo (PRPP), dos moléculas de fosfoenolpiruvato, L-glutamina y L-serina.

Por lo tanto, generalmente se esperaría una regulación eficiente (como en el caso de *E. coli*), pero esta eficiencia no está reflejada en el *phylum* Actinobacteria, donde la poca, o nula, regulación suele acontecer principalmente en las enzimas donde se divide el flujo de la vía metabólica, con una regulación alostérica mínima por parte del Trp [Palazzotto et al. 2015]; de la misma manera para el caso de la síntesis de His [Hodgson, 2000], sin embargo, este rubro aún permanece en gran parte inexplorado.

Se ha demostrado que cierta actinobacteria de suelo comun, como lo es *Sco* sintetiza Trp en todo su ciclo de vida ya que no solo lo utiliza como componente de su crecimiento (i.e. biomasa), sino que además lo utiliza en múltiples procesos biológicos como la síntesis de CDA [Kempter et al 1997].

La aparente deficiencia de regulación es ocasionada debido a que las bacterias de este *phylum* suelen desarrollarse en nichos ecológicos diversos: suelos, microbiomas, superficies intrahospitalarias, donde los L-aminoácidos no suelen encontrarse disponibles para su absorción, por lo cual no es necesaria una regulación estricta de síntesis de L-aminoácidos, ya que siempre deben sintetizarlos y han evolucionado economizando su genoma y metabolismo a una regulación poco estricta [Hodgson, 2000].

Algunas actinobacterias mantienen diferentes estados fisiológicos, o de desarrollo, en los que las demandas que afectan la vía Trp pueden ser más complicadas que simplemente detectar la disponibilidad de Trp para la síntesis de proteínas. Estos estados, a menudo involucran vías especializadas que coexisten con la biosíntesis primaria de Trp. Estas rutas especializadas están codificadas en parte o en su totalidad por duplicados divergentes (expansión de genomas) de genes Trp cuya expresión se desencadena por una variedad de señales temporales y ambientales (metabolismo especializado), e.g. para biosintetizar un pigmento o antibiótico determinado que se deriva en parte de la ruta Trp (e.g. en *Sco*).

## 2.4. Modelo de promiscuidad enzimática bisustrato: enzima PriA

La mayoría de las bacterias pueden sintetizar su propio Trp ya que este aminoácido solo lo sintetizan naturalmente bacterias, plantas y algunos hongos; el enantiómero *levo* es el biológicamente activo. La biosíntesis de L-aminoácidos no es parte de las rutas centrales catabólicas, se ha reportado metabolismo subterráneo (metabolismo que hace uso de enzimas promiscuas para llevar a cabo la síntesis de productos, usando metabolitos endógenos [D'Ari & Casadesus, 1998]) involucrado en la biosíntesis de este tipo de vías [Richs & Commichau, 2021].

En las rutas de biosíntesis más comúnmente encontradas para His y Trp en procariontes, con dos genes diferentes, *hisA* y *trpF* *sc.*, codifican dos enzimas intermedias en ambas rutas: la isomerasa N'-[(5'-fosforribosil)-formimino]-5-aminoimidazol-4-carboxamida ribonucleótido, (ProFAR) (gen *hisA*; EC 5.3.1.16) y la isomerasa PRA (gen *trpF*; EC 5.3.1.24), que isomerizan diferentes sustratos en las rutas de biosíntesis de los L-aminoácidos mencionados.

El sustrato ProFAR (biosíntesis de His) es isomerizado, produciendo N'-[(5'-fosforribosil)-formimino]-5-aminoimidazol-4-carboxamida-ribonucleótido (PRFAR) y, el sustrato PRA (biosíntesis de Trp) es isomerizado, produciendo 1-(2-carboxi-fenilamino)-1'-desoxirribulosa-5'-fosfato (CdRP), ambas reacciones catalizan bajo el mismo mecanismo: un reordenamiento de Amadori [Fig. 2.4.1.], este reordenamiento deriva a través de un mecanismo ácido-base de Schiff, utilizando residuos de aspartato, y en el caso de TrpF como pares ácido-base [Henn-Sax et al. 2002].

En su estructura, ambas enzimas de sustrato único y la bisustrato PriA se pliegan en barriles de ocho subunidades  $\beta/\alpha$  [Fig. 2.4.2.a.] [Due et al. 2011]. Las enzimas HisA, TrpF y PriA son sistemas modelo de estudio para la evolución de la especificidad y actividad enzimática.

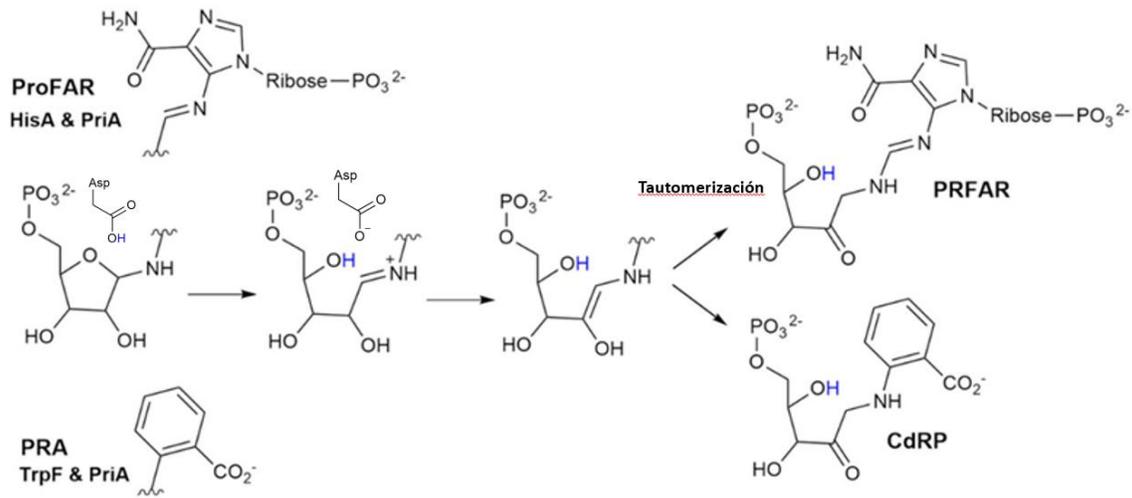


Figura. 2.4.1. Reordenamiento de Amadori que conduce a la isomerización de los sustratos ProFAR y PRA por HisA, TrpF y PriA, se observa el ácido aspártico donante de protón. [Modificado de Henn-Sax et al. 2002]

La estructura terciaria de barril de ocho subunidades  $\beta/\alpha$  (también llamada TIM, por la enzima triosa fosfato isomerasa) es un plegamiento común y versátil de proteínas de un solo dominio que es adoptado por una gran cantidad de enzimas en la naturaleza. Este tipo de plegamiento se ha utilizado como modelo para dilucidar la base estructural de la termoestabilidad de proteínas (i.e. en su entorno natural, las enzimas deben ser estables para mantener sus estructuras nativas, pero también deben ser flexibles para permitir cambios conformacionales durante la catálisis), también se han usado en

estudios para transformar actividades catalíticas o especificidades de sustrato mediante diseño racional o evolución dirigida [Richard, 2019].

El plegamiento canónico TIM consiste en dos estructuras secundarias, hojas  $\beta$  plegadas paralelas cerradas de ocho hebras, que forman un barril central, el cual está rodeado por ocho  $\alpha$ -hélices, las asas o “*loops*” unen las cadenas  $\beta$  con las hélices  $\alpha$  aledañas.

Los residuos del sitio activo están ubicados en la cara catalítica del barril, que comprende los extremos carboxi-terminal y amino-terminal de las cadenas  $\beta$ . Por el contrario, los *loops* que unen las  $\alpha$ -hélices con las hebras  $\beta$  posteriores, se encuentran en la cara opuesta del barril, y son importantes para estabilizar el plegamiento [Fig. 2.4.2.a]. Esta “división de actividades” entre las dos caras del barril  $\beta/\alpha$  hace posible cambiar la actividad catalítica con algunas mutaciones sin comprometer la estabilidad de la proteína, característica muy útil al ser utilizada como andamio en ingeniería de proteínas [Höcker et al. 2001; Richard, 2019].

Dada la especificidad de sustrato dual fisiológicamente relevante, PriA ha sido acogida como un sistema modelo para estudiar la evolución de las enzimas barril TIM. Sin embargo, debido a que los dos sustratos difieren en un factor de 2:1 en peso molecular, los sitios activos de HisA y TrpF sc., difieren sustancialmente en tamaño total. Mientras que el sustrato ProFAR de HisA es más grande con dos grupos fosfato terminales en posiciones simétricas en el sitio activo del plegamiento de barril TIM de PriA, se une a través de los extremos carboxi-terminal (Ser 81 en *Sco*) y amino-terminal (Arg 19 en *Sco*) [Fig. 2.4.2.a], el sustrato PRA de TrpF más pequeño se une a través de un solo “sitio de unión a fosfato” (PBS, por sus siglas en inglés) N-terminal en el sitio activo conservado de enzimas de barril TIM [Wright et al 2008; Due et al. 2011; Longo et al. 2020].

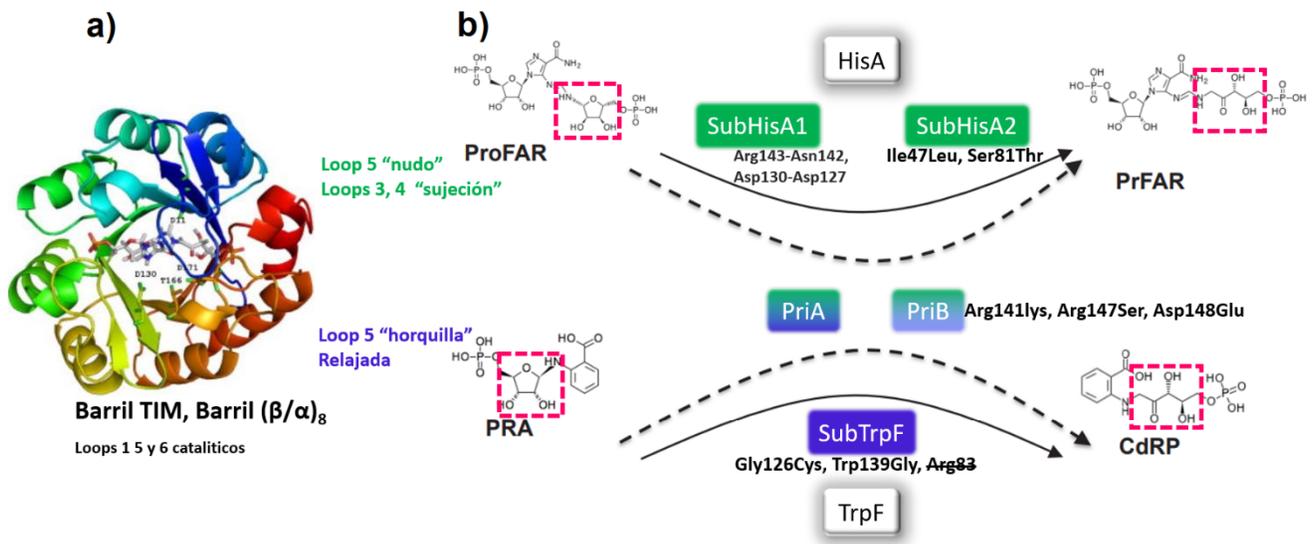
En una comparación por Romero-Rivera et al. (2022) con “*molecular docking*” de tres enzimas, una TrpF, una PriA y una HisA encontradas en diferentes organismos, observaron que la cavidad del sitio activo de la enzima TrpF es más compacta que las cavidades de PriA e HisA las cuales tienen sitios activos más grandes y flexibles. Comparando, los sustratos ProFAR y PRA [Fig. 2.4.2.b] se observó que tienen volúmenes de 829 Å<sup>3</sup> y 559 Å<sup>3</sup> sc.

Esto confirma datos estructurales que indican que la cavidad del sitio activo de una enzima TrpF es demasiado compacta para acomodar el sustrato más grande y voluminoso ProFAR. El sitio

activo de PriA es el más flexible de los tres, los datos estructurales indican que PriA es capaz de reorganizar significativamente su sitio activo (en particular, la conformación del *loop* 5) al acomodar los diferentes sustratos ProFAR y PRA [Romero-Rivera et al. 2022].

Todas las proteínas de barril TIM poseen *loops*, cuya diversidad conformacional juega un papel importante en la regulación de la especificidad y la función [Newton et al. 2017; Kadumuri & Vadrevu 2018]. El centro de la enzima PriA y de sus subespecializaciones (*ver sección 2.5.1 a la 2.5.3*) es rígido, mientras que los *loops* exhiben una flexibilidad importante [Liao et al. 2018]. Estos *loops* flexibles pueden variar en longitud, pero normalmente su función es sujetar, envuelven el sustrato, aislando y cerrando el sitio activo del solvente circundante [Wierenga, 2001]. Los *loops* catalíticos de la PriA de *Mtu* son: *loop* 1 (residuos 15–33), *loop* 5 (residuos 141–151) y *loop* 6 (residuos 172–181) [Romero-Rivera et al. 2022]. Estos tres *loops* experimentan conformaciones abierta (de no sujeción) y cerrada (de sujeción) dependiendo del sustrato que se une. Los *loops* catalíticos pueden variar dependiendo de la bacteria.

Las interacciones entre un *loop* de sujeción crucial [Fig. 2.4.2.a] que decora el sitio activo y los grupos fosfato no reactivos de los sustratos en estas enzimas desencadenan cambios conformacionales catalíticamente críticos y efectivos en el *loop* de sujeción, facilitando transiciones energéticamente desfavorables como son conformaciones abiertas inactivas a conformaciones cerradas activas que son energéticamente favorables para llevar a cabo la catálisis [Liao et al. 2018]. El grupo fosfato forma puentes de hidrógeno con las cadenas laterales Arg-85 del *loop* 3 y Thr-105 del *loop* 4 en PriA de *Mtu* [Due et al. 2011].



**Figura. 2.4.2.** a) Estructura de enzima TIM o de barril de ocho subunidades  $\beta/\alpha$ , con el sustrato ProFAR acoplado, muestra extremo C-terminal (rojo) y N-terminal (azul), así como las cadenas laterales de aminoácidos involucrados en la catálisis. Se muestran los *loops* claves. [Wright et al. 2008]. b) Reacción de isomerización de la enzima PriA, mostrando el reordenamiento de Amadori (apertura del anillo fosfo-ribosil, encuadros rosas punteados) [Modificado de Noda-García et al. 2015]. También se muestra las subfamilias de PriA con las mutaciones de intercambio de residuos en los sitios de unión a sustrato que confieren la subespecialización, para el caso de SubTrpF la Arg83 importante en PriA está ausente [Noda-García et al. 2013; Verdusco-Castro, 2016; Juárez-Vázquez et al. 2017].

Este tipo de interacciones activan la isomerización de PriA por cambios conformacionales controlados por ligando. Los *loops* 3 y 4 actúan de manera similar como "*loops de sujeción*", lo que permite que estas enzimas alcancen conformaciones catalíticamente activas, relevantes para la isomerización del sustrato más grande. Este efecto no está presente cuando el sustrato PRA más pequeño se une al sitio activo, ya que carece de un grupo fosfato no reactivo para interactuar con estos *loops* [Due et al. 2011].

El *loop* 5 funcionalmente importante en PriA de *Mtu* adopta una estructura similar a una horquilla  $\beta$  torcida (que consta de dos hebras), y asume, dependiendo de cuál de los dos sustratos esté presente, una conformación relajada sin movimiento (relacionado con PRA) o una conformación parecida a un nudo retorcido (relacionado con ProFAR) que abre y cierra el sitio activo [Due et al. 2011].

En el caso del arreglo conformacional del sustrato ProFAR, dos residuos del *loop* 5, Arg-143 y Trp-145, anclan el fragmento aromático de la molécula de sustrato (anillo de imidazol) fijándola y sujetándola a la cavidad del sitio activo (interacciones de apilamiento), estos residuos intercambian sus posiciones al momento de la sujeción, Trp apunta hacia el sitio activo y Arg hacia el solvente, o a la inversa si esta relajada [Due et al. 2011].

Es importante destacar que, además de la sujeción del sustrato ProFAR, se ha postulado que Arg-143 parece participar en el posicionamiento del Asp-175 catalítico (*loop* 6) [Fig. 2.4.1.] para isomerizar PRA, pero no desempeña ese papel para isomerizar ProFAR. El reordenamiento del *loop* 5 requerido para la unión de ProFAR evita que Arg-85 sujete el segundo grupo de fosfato, lo que conduce a una preferencia por la isomerización de PRA, en cuanto a movimientos de los *loops* (Due et al. 2011).

Datos de ángulos diédricos de cadena lateral en PriA de *Mtu*, los cuales evalúan la movilidad conformacional de la cadena lateral de los residuos de las enzimas, mostraron que las cadenas laterales de Trp y Arg son altamente flexibles conformacionalmente en la enzima sin ligando. La cadena lateral Arg-143 está a solo 4.8 Å del residuo de sujeción Arg-85 (distancia entre los dos átomos de carbono de la cadena lateral) [Romero-Rivera et al. 2022].

Lo anterior crea una repulsión electrostática entre las dos cadenas laterales de Arg, desestabilizando tanto el *loop* 5 como la interacción entre ProFAR y la cadena lateral Arg-85; es probable que la interacción de estos residuos con el sustrato más grande ProFAR desempeñe un papel importante en la capacidad de la enzima para unirse e isomerizarlo. Por el contrario, cuando PRA se une a PriA, las cadenas laterales Arg-83/Arg-85 aumentan su flexibilidad conformacional, mostrando aproximadamente el mismo espacio conformacional que en la enzima sin ligando [Romero-Rivera et al. 2022].

Los valores de  $k_{\text{cat}}$  observados para los sistemas de isomerización de ProFAR y PRA con PriA de *Mtu* y *Sco* son extremadamente bajos (relativamente ineficientes), del orden de  $\sim 10 \text{ s}^{-1}$  (o menos) a pesar de catalizar una reacción intrínsecamente fácil; en *Mtu* la reacción HisA con  $k_{\text{cat}}$ :  $0.23 \text{ s}^{-1}$  y la reacción TrpF con  $k_{\text{cat}}$ :  $3.6 \text{ s}^{-1}$ ; para el caso de la PriA de *Sco* la reacción HisA con  $k_{\text{cat}}$ :  $1.3 \pm 0.2 \text{ s}^{-1}$  y la

reacción TrpF con  $k_{\text{cat}}$ :  $3.4 \pm 0.09 \text{ s}^{-1}$  [Noda-García et al. 2010; Due et al. 2011]. Sin embargo, en el caso de *E. coli*, TrpF tiene una  $k_{\text{cat}}$  más alta que cualquiera de las enzimas mencionadas ( $k_{\text{cat}}$ : 30-40  $\text{s}^{-1}$ ).

En el reordenamiento de Amadori [Fig. 2.4.1.] que ocurre entre la reacción de apertura del anillo catalizada por la enzima y una tautomerización ceto-enol no enzimática (la cual es rápida y ocurre espontáneamente a 25 °C), la reacción de apertura del anillo es probablemente el paso químico más lento en este ciclo catalítico de la enzima y por consecuencia el paso limitante de la velocidad [Hommel et al. 1995; Dubey et al. 2021], sin embargo existen otros factores que intervienen en la velocidad de la reacción.

Romero-Rivera & col. (2022) reportaron una energía libre de activación de  $18.3 \pm 0.5 \text{ kcal} \cdot \text{mol}^{-1}$  para PriA de *Mtu* en su conformación pro-ProFAR, y para la conformación pro-PRA una energía libre de activación tentativa de  $23.3 \pm 1.2 \text{ kcal} \cdot \text{mol}^{-1}$  con base en simulaciones empíricas de enlaces de valencia (EVB, por sus siglas en inglés) de la apertura inicial del anillo de ribosa [Fig. 2.4.2.b]. El sustrato PRA tiene mucha más libertad conformacional en el sitio activo de la enzima PriA en comparación con el sustrato ProFAR, debido a lo anterior es capaz de exhibir una gran cantidad de conformaciones diferentes en el complejo de Michaelis y esto dificultó un cálculo preciso de la energía libre de activación en pro-PRA. Los autores resolvieron esta falta de precisión evaluando algunas mutaciones, dilucidando que la energía de activación es menor para el sustrato PRA que para ProFAR. También reportaron que el cierre completo del *loop* 1 es una conformación catalíticamente competente y es esencial para una isomerización eficiente de PRA [Romero-Rivera et al. 2022].

Cuando además se toma en cuenta la potencial participación de la dinámica de los *loops* (previamente mencionado) en la determinación de la  $k_{\text{cat}}$  experimental para la isomerización de ProFAR y PRA por las enzimas de interés, no solamente interviene el paso químico rápido que ocurre en el sitio activo de la enzima, sino que además los movimientos lentos conformacionales cerrados y abiertos de los *loops* limitan y controlan la velocidad en estas enzimas. La dinámica particularmente compleja de los *loops* dicta la selectividad de la enzima PriA. Y los cambios conformacionales activados por los ligandos estabilizan en parte el *loop* 5 en PriA consumando la isomerización de la reacción [Mhashal et al. 2020; Richard, 2022].

Los cambios conformacionales controlados por ligandos, y activados en diferentes *loops*, sugieren que estos *loops* evolucionan independientemente del barril estructural, proporcionando un punto de partida para la aparición y divergencia de nuevas actividades enzimáticas [Romero-Romero et al. 2021].

En experimentos de evolución en tiempo real que otorgaron actividad de PRA isomerasa a una HisA, se llevó a cabo efectivamente la bioingeniería de los *loops* "naturales" al modificar las conformaciones de los *loops* clave en los sitios activos [Näsval et al 2012; Ochoa-Leyva et al. 2013; Newton et al. 2017].

Acorde con Dubey & Singh (2021), el residuo Arg-15 (*loop* 1), el cual está cerca del sitio catalítico de una enzima HisA bifuncional, tiene aumentada la flexibilidad de dicho *loop* lo cual favorece la bifuncionalidad.

Se ha detectado actividad TrpF promiscua en enzimas HisA especializadas ancestrales y existentes, con valores de  $k_{cat}$  que van desde  $10^{-4}$  a  $10^{-2} s^{-1}$  [Plach et al. 2016]. HisA se ha convertido en TrpF por evolución dirigida, donde 3000 generaciones de evolución en laboratorio produjeron un amplio conjunto de mutaciones (e.g. la introducción de la sustitución Leu169Arg en HisA, la delección de Leu-169 o la duplicación de los residuos 13-15 Val-Val-Arg más la sustitución Asp10Gly, inducen la actividad de TrpF [Newton et al. 2017]) en HisA específica de ProFAR las cuales dieron como resultado enzimas HisA especializadas, entre ellas, enzimas TrpF especializadas y enzimas bifuncionales similares a PriA [Näsval et al. 2012].

Se ha argumentado que un gen similar a *priA* podría haber sido el ancestro evolutivo común tanto para HisA como para TrpF [Jürgens et al. 2000; Barona-Gómez & Hodgson 2003]. La reconstrucción de secuencias ancestrales también se ha utilizado para sugerir que los antiguos precursores de HisA probablemente eran bifuncionales y que esta bifuncionalidad persistió durante al menos un período de 2 mil millones de años [Planch et al. 2016]. Sin embargo, HisA y PriA utilizan los *loops* 1, 5 y 6 para facilitar la actividad, mientras que TrpF carece de un análogo del *loop* 1 e isomeriza PRA utilizando solo dos *loops* catalíticos 3 y 6 [Due et al. 2011].

Resultados de Romero-Rivera & cols. (2022) sugieren que una trayectoria evolutiva a partir de un ancestro existente similar a PriA hacia un TrpF sería sorprendentemente compleja. Para que se

llevara a cabo lo anterior, el *loop 1* se debería truncar (y no extenderse, como cuando HisA evolucionó artificialmente hacia TrpF [Newton et al 2017]), además sería necesario perder la interdependencia de este *loop*, lo que plantea la pregunta de qué camino evolutivo debería tomar un precursor similar a PriA para evolucionar a TrpF, mientras se elimina por completo el *loop 1*.

## 2.5. Subfamilia enzimática de PriA

El grupo del laboratorio de Diversidad de la Evolución Metabólica en UGA-LANGEBIO del CINVESTAV-IPN Irapuato, ha reportado hasta la fecha cuatro subfamilias de PriA con diferentes antecedentes y procesos evolutivos de subespecialización para cada uno de los cuatro casos.

Aunque la mayoría de los procariontes tienen enzimas HisA y TrpF separadas, la enzima PriA bifuncional de actinobacteria, se cree pudo evolucionar a partir de enzimas HisA promiscuas para catalizar pasos análogos en la biosíntesis de His y Trp [Jürgens et al. 2000; Barona & Hodgson, 2003], lo cual recientemente ha sido debatido por Romero-Rivera & cols. (2022), como se mencionó al final de la sección anterior. A su vez PriA se ha vuelto a especializar varias veces, con diferentes mecanismos evolutivos precedentes que se explican a continuación.

### 2.5.1. Transferencia horizontal de genes (SubHisA1)

En procariontes, la THG se ha propuesto como el mecanismo principal para la expansión de las familias de proteínas existentes [Arnold et al. 2022].

La ausencia de un gen *trpF* en Actinobacteria no se conserva universalmente en este phylum, ya que existe un operón de Trp completo que incluye un gen *trpF* (*trpF-1*, ver sección 2.2-I) en un subclado del género *Corynebacterium*. Acorde con Noda-García & cols. (2013), las actinobacterias, *Corynebacterium diphtheriae*, *Corynebacterium efficiens*, *Cgl*, *Corynebacterium matruchotii* y *Corynebacterium striatum*, adquirieron por THG el operón de Trp de vía completa, estos homólogos de PriA carecen completamente de la actividad PRA isomerasa.

Análisis de organización genética y filogenética de este operón completo demostraron que dicha vía metabólica se adquirió en bloque de las  $\gamma$ -proteobacterias. En este operón completo, el gen *trpF*, se fusiona con el gen *trpC* río abajo de las vías, una característica distintiva de las  $\gamma$ -proteobacterias [Xie et al. 2003]. Planteando la hipótesis de que la adquisición del operón completo provocó la pérdida de los genes *trp* corinebacterianos originales, luego de un desplazamiento de genes homólogos que provocó que la sintenia en este locus fuera casi imposible de reconocer [Xie et al. 2004].

Se ha propuesto que la adquisición de este operón Trp completo ocurrió después de la THG [Noda-García et al. 2013], y se ha demostrado que este gen *trpF* adquirido por THG se correlaciona con la falta de afinidad por PRA de los homólogos de PriA presentes en estos organismos. El producto resultante de este proceso es una PriA que se convirtió en una HisA monofuncional, o “SubHisA1”, nombre que refleja la subfuncionalización que sufrió la enzima.

Noda-García & cols. (2013) descubrieron que la especificidad alterada de SubHisA1 se debía tanto a mutaciones en el sitio activo como a una disminución de la flexibilidad conformacional, mostrando la importancia de la THG en la configuración de la evolución y el metabolismo.

La existencia del subclado del género *Corynebacterium* con un operón completo de Trp adquirido por THG se correlaciona con el deterioro del grupo de genes His-Trp, el deterioro incluye la pérdida de los genes *trpB* y *trpA*, y la mutación de *trpC*. Como consecuencia, los genes *his* y *trp* en estos organismos están separados por alrededor de 1000 kb al menos en *Cgl* [Fig. 7.4.1] [Noda-García et al. 2013].

Las mutaciones de los residuos que intervienen en la catálisis de SubHisA1 en comparación con PriA, detectadas estructuralmente fueron: Arg143 es remplazado por Asn142 en SubHisA1, y Asp127 en SubHisA1, se desplaza dos posiciones hacia el extremo amino-terminal quedando como Asp130, ambas ocurren a nivel del *loop* 5 [Noda-García et al. 2013] [Fig. 2.4.2.b]; las mutaciones de sustitución de residuos ocurren en Ser81Gly, His50Phe, Asp11Val, Asp130Val, Arg139Thr y Asp171Val en PriA bonafide (*Sco* & *Mtu*) y SubHisA1 (*Cgl*) sc., lo cual se observó en el transcurso de la presente tesis con la alineación de secuencias de aminoácidos de las proteínas en cuestión con MEGA 11 [Tamura et al. 2021], como se muestra a continuación en la Fig. 2.5.1.1.

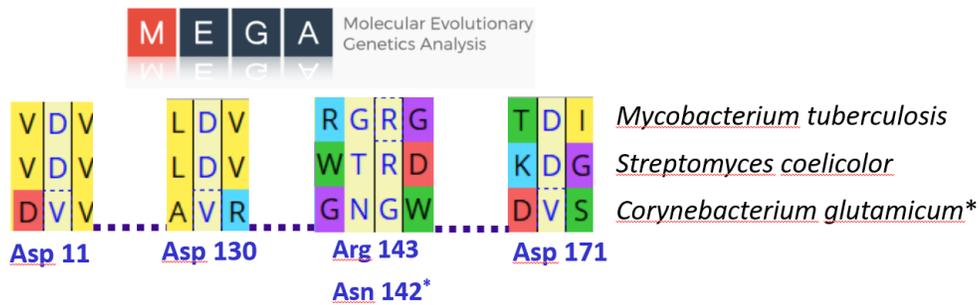


Figura. 2.5.1.1. Residuos catalíticos de PriA y SubHisA1. Análisis con MEGA 11 [Tamura et al. 2021].

Una explicación ante la subfuncionalización de SubHisA1, está dada por el modelo de “*Escape del Conflicto Adaptativo*” (ECA), dicho modelo propone que la selección positiva optimiza cada actividad después de la duplicación de una enzima multifuncional. ECA es un modelo estricto, que propone que las actividades en proteínas multifuncionales no pueden optimizarse mediante selección natural, porque las mutaciones que mejoran una función pueden ser perjudiciales para otras funciones [Des Marais & Rausher, 2008].

ECA plantea que después de la duplicación o amplificación de genes (expansión del genoma), estas mutaciones ofrecen una ventaja selectiva, y cada duplicado está sujeto a una selección positiva que los adapta a diferentes actividades [Des Marais & Rausher, 2008]. De manera similar, los residuos que mejoran las actividades de una enzima transferida horizontalmente estarán sujetos a selección positiva si mejoran la aptitud del huésped, mientras que los residuos que solo afectan las actividades no utilizadas estarán sujetos a selección neutral [Glesner et al. 2020].

ECA puede explicar la evolución de una enzima HisA (SubHisA1) especializada a partir de una PriA bifuncional [Glesner et al. 2020]. La especificidad dual de PriA está mediada por tres *loops* del sitio activo que adoptan diferentes conformaciones para cada reacción [Noda-García et al 2010, Due et al. 2011]. La subfuncionalización de SubHisA1 se produjo porque dichos *loops* adoptaron conformaciones diferentes y son menos dinámicos (selección neutral) que en las enzimas bifuncionales PriA [Noda-García et al 2013].

Para los fines de esta tesis la subvariante SubHisA1 es importante, ya que se evaluó con el GSMM de *Cgl* ATCC 13032.

### 2.5.2. Co-existencia genómica de *trpF* y *priA* (PriB)

En el género *Streptomyces* y *Paenarthrobacter*, existe la subfamilia enzimática denominada fosforribosil isomerasa B (gen *priB*), con bajas eficiencias catalíticas para la función TrpF de PriA (en este caso PriB) debido a la co-existencia de una copia extra, ortóloga de *trpF* (*trpF-3*, ver sección 2.2-III) localizada fuera del contexto genómico inmediato de los operones de His y Trp, y es posible que apoye la biosíntesis de Trp [Verduzco-Castro, 2016].

Las enzimas PriB de cinco organismos, *Paenarthrobacter aurences* (*Paur*), *Streptomyces* sp. *Mg1* (*SMg1*), *Streptomyces svisceus* (*Ssvi*), *Streptomyces ipomoeae* (*Sipo*) y *Streptomyces* sp. *C*, se caracterizaron, demostrando la conservación de la especificidad de sustrato dual (ProFAR isomerasa y PRA isomerasa). Los parámetros cinéticos de las enzimas PriB se ven afectados por la presencia de TrpF-3. La actividad de PRA isomerasa de las enzimas PriB varían de  $0.0022 \mu\text{M}^{-1} \cdot \text{s}^{-1}$  a  $0.14 \mu\text{M}^{-1} \cdot \text{s}^{-1}$  [Verduzco-Castro, 2016].

El análisis de los parámetros cinéticos de PriB revela que la variabilidad observada parece estar relacionada con  $k_{\text{cat}}$  más lentas, más que con valores de  $K_m$  más grandes [Tabla. 2.5.2.1.]. Esto sugiere que la compensación de las actividades de la PRA isomerasa observada no está relacionada con la especificidad de sustrato.

La co-evolución de las actividades PRA isomerasa codificadas por los homólogos *trpF-3* y *priB* mostró una compensación funcional. La baja actividad de la PRA isomerasa de PriB parece estar compensada por enzimas TrpF más activas, ya que sus actividades combinadas alcanzan niveles similares [Tabla. 2.5.2.1.] a los encontrados en las PriA bonafide (de *Sco* y *Mtu*).

Los residuos no conservados presentes en el *loop* 5 Arg141Lys, Arg147Ser, Asp148Glu con sustituciones de residuos en PriA y PriB sc. son responsables de los cambios funcionales de hasta dos órdenes de magnitud, en los organismos caracterizados [Verduzco-Castro, 2016].

Una explicación al fenómeno evolutivo de PriB, recurriendo nuevamente a un modelo de evolución de proteínas, está dado por el modelo de “*Duplicación-Degradación-Complementación*” (DDC), el cual propone que cada copia de un gen multifuncional se especialice al perder todas sus funciones menos una [Innan & Kondrashov 2010].

Conceptualmente, la pérdida de función (*degradación*) es neutral, pero la medición de la presión selectiva a través de la relación de tasas no sinónimas a sinónimas (dn/ds) probablemente indique una selección purificadora, relajada o débil, porque solo el subconjunto de sitios que contribuyen únicamente a la actividad perdida podría evolucionar neutralmente, mientras que los sitios que contribuyen a la actividad retenida estarían bajo selección purificadora [Innan & Kondrashov 2010].

Este modelo fue desarrollado con la regulación de genes en mente, proponiendo que las mutaciones en las regiones reguladoras alteran los patrones de expresión de los genes duplicados [Force, 1999]. Sin embargo, en las enzimas que catalizan múltiples reacciones en el mismo sitio activo, es probable que las mutaciones afecten el plegamiento y la estabilidad de las proteínas más que otras propiedades, y estas enzimas multifuncionales comparten un subconjunto de residuos catalíticos que se requieren para todas las actividades [De Pristo et al 2005]. Estas restricciones limitan el número de sitios donde podrían ocurrir mutaciones neutrales que afectan solo una actividad [Glasner et al. 2020].

DDC explica el caso de complementación ocurrido en PriB y TrpF-3. La actividad TrpF de PriB se volvió menos eficiente en especies de microorganismos que tenían una enzima TrpF eficiente, pero mantuvieron una mayor eficiencia manteniendo su bifuncionalización en especies con enzimas TrpF ineficientes (*Sipo*) [Tabla 2.5.2.1.]. Esta compensación funcional es consistente con la compensación por la pérdida de la función por neutralidad (*degradación*) del modelo DDC en una copia del gen [Glasner et al. 2020].

| M.O.        | TrpF-3 ( $\mu\text{M}^{-1}\cdot\text{s}^{-1}$ ) | PriB [TrpF] ( $\mu\text{M}^{-1}\cdot\text{s}^{-1}$ ) | PriB [HisA] ( $\mu\text{M}^{-1}\cdot\text{s}^{-1}$ ) |
|-------------|---|--|--|
| <i>Paur</i> | 3.3   | 0.014  | 0.9  |
| <i>SMg1</i> | 1.25  | 0.0022   | 0.069  |
| <i>Ssvi</i> | 0.088   | 0.067  | 0.18   |
| <i>Sipo</i> | <b>0.19</b>                                     | 0.14   | 0.21   |

Tabla 2.5.2.1. Comparación de las eficiencias catalíticas (mediciones *in vitro*) de la enzima TrpF-3, y catálisis de HisA y TrpF de PriB, en los microorganismos *Paur*, *SMg1*, *Ssvi* y *Sipo*. [Tomado de Verduzco-Castro et al. 2016].

### 2.5.3. Pérdida de genes (SubHisA2 y SubTrpF)

La pérdida de genes puede promover la evolución de la especificidad de sustrato hacia la subfuncionalización. Experimentos de cinéticas *in vitro* revelaron que en algunas especies del género *Actinomyces*, los homólogos de *priA* codifican enzimas monofuncionales (i.e. han perdido su promiscuidad característica en alguno de los dos sustratos) [Juárez-Vázquez, 2017]. El establecimiento de filogenias y la observación de los contextos genómicos de *priA* de miembros del género *Actinomyces* (especies que pertenecen principalmente al microbioma de mucosas en seres humanos), permitió entender la modificación de la promiscuidad.

Se observó la pérdida y ganancia de genes en la proximidad genómica del gen *priA*. Existen clados de *Actinomyces* donde genes pertenecientes al operón de His no están en la proximidad genómica de PriA, comprobando con cinéticas *in vivo* que la actividad catalítica HisA está ausente en el clado, se denominó a este tipo de enzimas SubTrpF (con mutaciones diferenciales con PriA sc., Gly126Cys, Trp139Gly y Arg83 ausente) debido a que únicamente se encuentran dotadas con la función catalítica TrpF. Similarmente, existen otros clados que perdieron genes del operón de Trp en la proximidad de PriA, estas enzimas se subfuncionalizaron a la función HisA, denominándose SubHisA2 (con mutaciones diferenciales con PriA sc., Ile47Leu, Ser81Thr); SubHisA2 muestra cierta actividad de isomerasa PRA residual, debido a la arquitectura de su sitio activo casi completamente conservada [Juárez-Vázquez et al. 2017].

El análisis de los genomas de *Actinomyces* evidenció la conservación de algunas rutas biosintéticas completas, y otras no conservadas e incompletas; lo anterior confiere a estos organismos la capacidad, o limitante, de proliferar en diferentes conjuntos de nutrientes ambientales e implica que estos microorganismos han estado expuestos a una amplia gama de condiciones ambientales y presiones de selección [Juárez-Vázquez et al. 2017].

La cercanía filogenética de las especies de *Actinomyces* funcionalmente diversas sugiere una rápida y fuerte capacidad de adaptación a las diversas condiciones presentes en el cuerpo humano. El proceso de pérdida de genes, asociado con la relajación de la selección purificadora, es el impulsor clave de esta estrategia de adaptación. De esta serie de datos se aprecia que en *Actinomyces* la especiación corresponde con el cambio de la promiscuidad en PriA, conforme con la pérdida y ganancia de genes adyacentes.

Los casos de subfuncionalización de SubHisA2 y SubTrpF donde ambas actividades de PriA parecen estar disminuyendo por la deriva genética. Desde la perspectiva del modelo DDC (descrito en la sección 2.5.2.), el género *Actinomyces* que ha experimentado la pérdida de genes, evolucionando con la relajación de la selección purificadora, en clados de especies con pérdida de una de las actividades de PriA, se correlaciona con la pérdida de la vía de biosíntesis del L-aminoácido correspondiente; sin embargo, la eficiencia de la actividad conservada también disminuyó, lo que sugiere que ambas actividades podrían perderse eventualmente [Juárez-Vázquez et al. 2017; Glasner et al. 2020].

## **2.6. La promiscuidad enzimática para el diseño de rutas *de novo***

El uso de enzimas promiscuas ha mostrado un gran potencial para la construcción de vías *de novo*, en las cuales las enzimas faltantes de la vía de biosíntesis pueden compensarse mediante la introducción de enzimas promiscuas reutilizadas [Notebaart et al. 2018].

Se ha estimado que el 37 % de todas las enzimas en *E. coli* pueden actuar sobre múltiples sustratos y catalizan el 65 % de las reacciones metabólicas conocidas [Nam et al. 2012].

Aprovechando el gran potencial de las enzimas promiscuas derivadas de *E. coli*, ahora es posible producir productos químicos de alto valor agregado utilizando vías biosintéticas *de novo*, antes estos productos solo se podían obtener por procesos petroquímicos.

A continuación, presento algunos ejemplos.

Cen & cols. (2021) reportaron la producción de 1,5-pentanodiol (un elemento básico en la síntesis de poliésteres y poliuretanos) por primera vez en una *E. coli* modificada, utilizaron la estrategia de introducir vías biosintéticas *de novo*. Seleccionaron enzimas promiscuas de una amplia gama de sustratos, en especial, enzimas aldehído deshidrogenasas heterólogas de diversos organismos para la completa producción *de novo* de 1,5-pentanodiol a partir de glucosa.

De manera análoga, el 4-amino-1-butanol (un precursor importante de polímeros biodegradables) se produjo por primera vez en *Cgl* mediante la introducción de vías de biosíntesis *de novo* que comprendieron las enzimas aldehído deshidrogenasa y putrescina aminotransferasa de *E. coli*, las cuales tienen una amplia afinidad por diferentes sustratos [Prabowo et al. 2020].

Enzimas promiscuas de algunas otras especies, además de las mencionadas, también han sido ampliamente utilizadas para construir vías *de novo*. Por ejemplo, la enzima mevalonato difosfato descarboxilasa de *Saccharomyces cerevisiae*, de la cual se tenía conocimiento previo que exhibía una actividad promiscua, se modificó para crear una vía de mevalonato que evita el paso de formación de un metabolito tóxico que consume ATP isopentenil difosfato para la producción de isopentanol (biocombustible y precursor de productos químicos industriales como el isopreno) en *E.coli* [Kang et al. 2017].

De manera similar, se han logrado sintetizar algunos compuestos aromáticos prenilados en *E.coli*, entre ellos los ácidos orselínico, divarinólico y olivetólico, aprovechando la promiscuidad de la enzima preniltransferasa aromática NphB diseñada a partir de una cepa de *Streptomyces sp. CL190*, con modelos estructurales de proteínas y diseño racional, para obtener mutantes variantes de NphB con mayor actividad y selectividad de sustrato [Qian et al. 2019].

El 1,6-hexametilendiamina (monómero precursor en la producción del nailon-6,6, el cual es uno de los dos nailons más comunes en la industria textil y plástica) se produjo a partir de la conversión

*in vitro* de ácido adípico mediante catálisis con enzimas transaminasas y reductasas de ácidos carboxílicos, las cuales aceptan sustratos con grupos carboxilo, carbonilo y aminas terminales. Estas enzimas se modificaron para crear variantes mutantes con actividades mejoradas, lo que resultó en una transformación completa dada en sucesivas reacciones químicas (cascada) del mismo ciclo de ácido adípico a 1,6-hexametilendiamina [Fedorchuk et al. 2020].

De los ejemplos que he citado en los párrafos anteriores referentes a la explotación del potencial de enzimas promiscuas; la gran mayoría utilizo a *E.coli.* como chasis; en solo uno de los ejemplos utilizaron a *Cgl* como chasis, y en otro ejemplo utilizaron una enzima promiscua proveniente de un *Streptomyces*, estos dos últimos ejemplos involucraron al *phylum* actinobacteria.

## 2.7. El flujo metabólico relacionado con la evolución de genes

El flujo metabólico se correlaciona con la proporción de tasas de sustitución no sinónimas a sinónimas de los nucleótidos en la secuencia de un gen. Los genes que codifican enzimas, las cuales transportan flujos metabólicos elevados han sufrido mayores restricciones selectivas durante su evolución, mientras que la selección purificadora es más relajada en genes que codifican enzimas que transportan flujos metabólicos bajos [Nam et al. 2012; Juárez-Vázquez et al. 2017].

En ausencia de un conocimiento detallado de los parámetros cinéticos de las reacciones individuales de las vías de los organismos que permitiría reconstruir un modelo cinético computacional, el flujo puede predecirse computacionalmente a través de FBA [Orth et al. 2010].

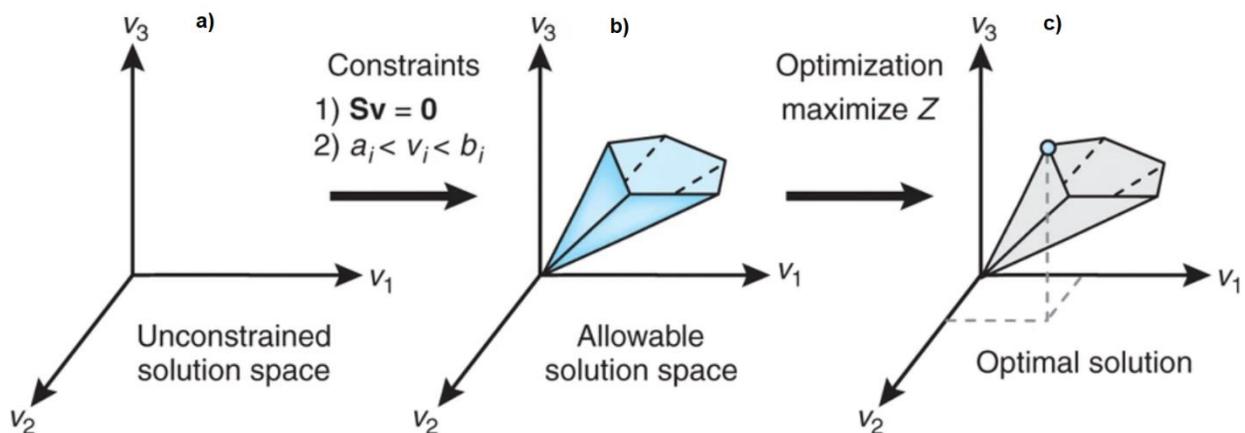
FBA es un enfoque matemático que se emplea extensamente, el cual permite analizar la distribución de flujos metabólicos en una red metabólica. El flujo metabólico se fundamenta en la distribución de carbono dentro de una red metabólica, sus unidades son  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$  (i.e. la cantidad de materia en milimoles de carbono por gramos de biomasa seca en el tiempo) [Orth et al. 2010]. Esta información parte de los balances de materia de los metabolitos internos que al aplicarse el estado pseudo-estacionario se genera un sistema de ecuaciones lineales. Para lograr una solución del sistema, la aproximación de FBA transforma el problema a un problema de optimización.

Utilizando el algoritmo de programación lineal para calcular la distribución de flujo que maximiza o minimiza la función objetivo la cual es la tasa de crecimiento, esta solución se logra dentro del espacio de flujos permitidos (espacio solución) definido por las restricciones impuestas de las ecuaciones de balance de masa y los límites de reacción [Loghmani et al. 2022].

Con este recurso matemático es posible analizar con los GSMM abordando la distribución de carbono en las redes metabólicas.

FBA se utiliza para determinar la distribución de flujo aplicando la hipótesis de un estado pseudo-estacionario (i.e. la concentración de los metabolitos intracelulares se mantiene constante en el periodo de tiempo en el que se realizan los cálculos de la distribución de flujos), y el principio de balance de masa planteando un sistema de ecuaciones lineales, dada una o varias rutas metabólicas en un sistema completo del metabolismo celular (la red metabólica) [Nam et al. 2012].

Para estimar la distribución de flujo fisiológico en estado pseudo-estacionario, se deben aplicar procedimientos de optimización. La distribución de flujo óptima dentro de una red metabólica se define como aquella que maximiza o minimiza una función objetivo particular (e.g. la tasa de crecimiento celular) [Kauffman et al. 2003].



**Figura. 2.7.1. Flux balance analysis (FBA).** a) Espacio de solución sin restricciones, la distribución de flujo de una red biológica puede estar en cualquier punto de un espacio de solución. → Cuando se aplican a una red las restricciones de balance de masa impuestas por la matriz estequiométrica  $S$  (etiqueta 1) y las restricciones de capacidad impuestas por los límites inferior y superior ( $a_i$  y  $b_i$  sc.) (etiqueta 2). b) Se define un espacio de solución permisible. La red puede adquirir cualquier distribución de flujo dentro de este espacio, pero las restricciones niegan los puntos fuera de este espacio (espacio nulo). → A través de la optimización de una función objetivo. c) FBA puede identificar una única distribución de flujo óptima que se encuentra en un borde del espacio de solución permitido. [Modificado de Orth et al. 2010].

En el estado pseudo-estacionario, el flujo a través de cada reacción esta dado por:

$$S \cdot \vec{v} = 0 \quad (2.1)$$

Donde  $\vec{v}$  es el vector de tasas de reacción (valores del flux) y  $S$  es la matriz estequiométrica de dimensión  $m \times n$ ; con  $m$  metabolitos y  $n$  reacciones. El sistema de ecuaciones lineales 2.1 se resuelve con algoritmos de programación lineal para encontrar la distribución de flujos  $\vec{v}$ . Los GSMM suelen ser grandes por lo que contienen más reacciones que metabolitos, por lo que es posible que exista más de una solución posible para estas ecuaciones, las cuales son difíciles de abarcar experimentalmente. Debido a esto, se planteó el problema de optimización lineal denominado FBA.

Resolver las ecuaciones para predecir la tasa máxima de crecimiento requiere definir una función objetivo:

$$Z = c^T \cdot \vec{v} \quad (2.2)$$

Donde  $c^T$  es un vector de peso que indica cuánto contribuye cada reacción ( $\vec{v}$ ) al objetivo. En la práctica, cuando solo se desea maximizar o minimizar una reacción, como la producción de biomasa,  $c$  es un vector de ceros con un valor de 1 en la posición de la reacción de interés.

Por lo tanto, en el ejemplo de crecimiento, la función objetivo sería:

$$Z = \vec{v}_{biomasa} \quad (2.3)$$

Donde  $c$  tiene un valor de 1 en la reacción de biomasa [Feist & Palsson 2010].

Las restricciones proporcionan dos caracteres, primero son ecuaciones que equilibran las entradas y salidas de las reacciones y segundo son desigualdades que imponen límites al sistema.

La matriz estequiométrica impone restricciones de equilibrio de flujo (i.e. masa) en el sistema, asegurando que la cantidad total de cualquier compuesto que se produzca debe ser igual a la cantidad total que se consume en el estado pseudo-estacionario [Orth et al. 2010].

A cada reacción se puede asignar un límite superior ( $u$ ) e inferior ( $l$ ), que definen los flujos máximos y mínimos permitidos para la reacción [Raman & Chandra, 2009].

$$lb_j \leq \vec{v}_j \leq ub_j \quad (2.4)$$

Estos equilibrios y límites definen el espacio de distribuciones de flujo permitidas de un sistema (i.e. las tasas a las que cada reacción consume o produce cada metabolito).

Las restricciones incluso se pueden utilizar para simular la desactivación de genes al limitar las reacciones a flujo cero. Se pueden agregar más restricciones de diferente índole, como las proporcionadas por los datos ómicos [Raman & Chandra, 2009; Loghmani et al. 2022].

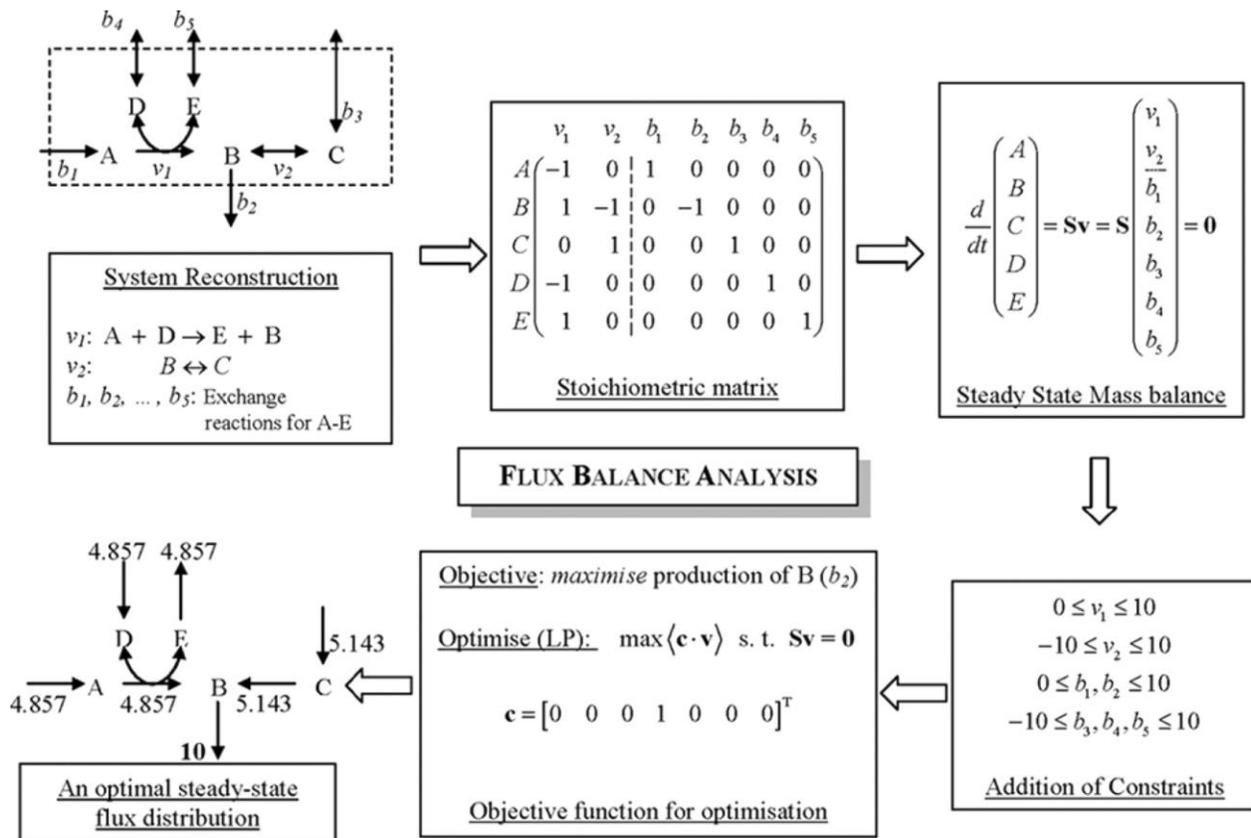


Figura. 2.7.2. Principales pasos resumidos, involucrados en FBA. [Tomado de Raman & Chandra, 2009]

FBA produce predicciones que deben verificarse. Los estudios experimentales se utilizan como parte del proceso de reconstrucción del modelo y para validar las predicciones del modelo. Los estudios han demostrado que las tasas de crecimiento de *E. coli* en varios sustratos diferentes

predichas por FBA concuerdan bien con las obtenidas por mediciones experimentales [Edwards et al. 2001].

Acorde con Nam & cols. (2012), en la  $\gamma$ -proteobacteria *E. coli*, el arquea *Methanosarcina barkeri*, la levadura *Saccharomyces cerevisiae* y la microalga *Chlamydomonas reinhardtii*, hallaron con ayuda de los GSMM y un método de muestreo de Monte Carlo basado en cadenas de Markov, que la especificidad de las enzimas se correlaciona con las funciones globales de la red metabólica. Las enzimas especializadas (como las del metabolismo central) mantuvieron un flujo más alto en promedio que las enzimas generalistas (que manejan eficiencias pequeñas y no son parte del metabolismo central), cuando se simularon cambios ambientales para cada organismo, era menos probable que las enzimas generalistas cambiaran el flujo entre las condiciones de crecimiento, la magnitud del flujo puede contribuir a la especialización enzimática, así como también puede contribuir la esencialidad del gen. La alta afinidad de sustrato por las enzimas esenciales podría mitigar la competencia entre sustratos en la síntesis de los componentes necesarios de la biomasa, independientemente del nivel de flujo. Nam & cols. (2012) detectaron que las enzimas esenciales tienen valores de  $K_m$  más bajos y, por lo tanto, mayor afinidad por el sustrato.

El crecimiento celular rara vez depende directamente del flujo a través de enzimas generalistas, mientras que muchas enzimas especialistas fueron esenciales para el crecimiento en las 174 condiciones de medios diferentes probados [Nam et al. 2012]. El flujo a través de enzimas especialistas es considerablemente más sensible al cambio ambiental, mientras que los flujos de enzimas generalistas varían menos. La regulación metabólica fue menos frecuente entre los generalistas, en consonancia con la menor necesidad de cambiar el flujo a través de sus reacciones en entornos dinámicos. Las enzimas que exhiben una regulación metabólica más extensa tienden a haber desarrollado una mayor especificidad enzimática [Nam et al. 2012].

La evolución no ha convergido hasta un punto en el que las enzimas metabólicas sean todas especialistas, esta convergencia se ha visto obstaculizada en parte por la menor esencialidad, el flujo más pequeño y los requisitos regulatorios reducidos de las enzimas generalistas, incluidas las que son multifuncionales y las que exhiben la promiscuidad de sustrato. Nam & cols. (2012), demostraron en

cuatro organismos que la biología de sistemas permite dilucidar las presiones de selección sistémicas del metabolismo que no son evidentes a nivel de una sola enzima.

Los cambios en la regulación de la expresión de proteínas también pueden afectar la capacidad de evolución de las proteínas. Los niveles de expresión y la concentración de proteína se correlacionan con las tasas evolutivas: cuanto mayor es la cantidad de proteína en la célula, más lenta es la tasa de evolución [Echave et al. 2016].

### 3. Justificación

A causa de la actual dependencia de materiales obtenidos a partir de combustibles fósiles, existe un interés creciente por el desarrollo de métodos sustentables para reemplazarlos y proporcionar enfoques alternativos para producir compuestos químicos con aplicación en campos como medicina (farmacéutica), productos naturales, cosmética, polímeros y materiales (e.g. bioplásticos), combustibles, productos químicos finos y a granel, aminoácidos, y agro-alimentos [Cho et al 2022].

Un enfoque prometedor es el desarrollo y diseño de biofábricas de células microbianas para producir tales compuestos de valor agregado a partir de materias primas renovables (biomasa) como medios de cultivo. Específicamente, la red metabólica de estas especies de microorganismos se puede rediseñar y modificar para optimizar la producción de los compuestos deseados [Yim et al. 2011; Hadadi & Hatzimanikatis, 2015; Nielsen & Keasling, 2016; Wang et al. 2017; Ko et al. 2020].

Un informe de *Market Research Future* en 2019, estimó que el alcance del mercado mundial de productos químicos de base biológica creció a una tasa promedio anual de 10.47 % para alcanzar 97,200 millones de USD en 2023 [Cho et al 2022].

China en particular ha implementado múltiples regulaciones para crear una industria química más sustentable y aumentar gradualmente los salarios [acorde con "*The 13th Five-Year Plan for Economic and Social Development of the People's Republic of China, 2016-2020*"]. Lo anterior ha dado incentivos para el desarrollo de procesos de fermentación económicamente competitivos y ecológicamente sostenibles [Revuelta et al. 2016].

Existen varias razones por las que la fermentación es más sostenible desde el punto de vista medioambiental: *i)* los organismos pueden utilizar recursos renovables (biomasa) como materiales de partida para la síntesis de los compuestos de interés; *ii)* la síntesis enzimática es más eficiente y

selectiva que la síntesis química; y *iii*) en contraste con la síntesis química, se forman subproductos mucho menos tóxicos y peligrosos para el medio ambiente [Ahmann & Dorgan, 2007]. Por lo tanto, el desarrollo de procesos de producción biotecnológicos novedosos y más eficientes es atractivo y no hay indicios de que esto vaya a cambiar en un futuro cercano. Por lo que, se necesitan con urgencia métodos para desentrañar y aprovechar completamente el potencial de los microorganismos para producir todo tipo de compuestos de valor agregado [Rosenberg & Commichau, 2019].

Se han desarrollado múltiples enfoques de ingeniería metabólica para mejorar la producción de compuestos de valor agregado (Okano et al. 2018; Pontrelli et al. 2018). Uno de los enfoques más ampliamente aplicado es la introducción y expresión de reacciones metabólicas heterólogas (i.e. reacciones y genes externos de un organismo diferente), que generan rutas biosintéticas heterólogas con el uso de proteínas heterólogas (o proteínas recombinantes) que permiten la producción de compuestos no nativos de valor agregado por parte del huésped [Lechner et al. 2016; Pickens et al. 2011]; algunas otras estrategias utilizadas en ingeniería metabólica son: *i*) delección de genes que desvían el flujo hacia la biosíntesis de interés, eliminando reacciones competidoras de flujo; *ii*) modificaciones que alteran el equilibrio redox; *iii*) sobreexpresión de genes (a través de inducción de promotores fuertes e inducibles); *iv*) modificaciones relacionadas con el consumo de fuentes de carbono (e.g. en bacterias el sistema de fosfotransferasa (PTS) dependiente del fosfoenolpiruvato (PEP) [Roseman, 1969], implica un proceso de acoplamiento energético que fosforila glucosa durante su transporte. Este sistema de importación de la fuente de carbono consume hasta el 50 % de todo el PEP producido por la célula, el PEP es un metabolito del metabolismo central involucrado en importantes vías metabólicas, es un precursor importante de Trp y los demás aminoácidos aromáticos; las mutaciones PtsG y PtsH son las más recurrentes para evitar el consumo de PEP durante la importación de glucosa, aumentando la disponibilidad de PEP para la sobreproducción industrial de Trp [Gu et al. 2012; Liu et al. 2017]); *v*) eliminación de rutas de degradación del compuesto de interés; *vi*) modificación de los transportadores transmembranales del compuesto de valor agregado; *vii*) control de la tasa de crecimiento; *viii*) reducción de la formación de subproductos de fermentación (e.g. ácido acético y CO<sub>2</sub>) y; *ix*) aumento de la biodisponibilidad de precursores limitantes del compuesto de interés.

Se han propuesto diversos enfoques computacionales para facilitar la ingeniería de rutas metabólicas a través de reacciones heterólogas utilizando GSMM (Ko et al. 2020; Wang et al. 2017). Los GSMM que están disponibles para los microorganismos más comúnmente utilizados en ingeniería metabólica pueden predecir con precisión cómo la adición de enzimas heterólogas y otras de las estrategias mencionadas afectan la producción de metabolitos [Pharkya et al. 2003; Pharkya et al. 2004].

Aunque el vasto repertorio de reacciones heterólogas disponibles proporciona un gran potencial para la producción de compuestos de valor agregado, existen varias limitaciones. En primer lugar, las enzimas heterólogas a menudo requieren cofactores específicos que el huésped no puede proporcionar, lo que dificulta el adecuado funcionamiento de la ruta biosintética (Boynton et al. 1996). En segundo lugar, la expresión de genes heterólogos puede conducir a un estrés celular debido a la sobreproducción de proteínas o la acumulación de intermediarios tóxicos [Gill et al. 2000; Martin et al. 2003]. En tercer lugar, los microorganismos que contienen genes heterólogos se consideran “*organismos modificados genéticamente*” (GMO, por sus siglas en inglés), cuya aplicación comercial, especialmente en la industria alimentaria, depende de las diferentes normas jurídicas regentes de los países [Kovács et al. 2022].

La gran robustez del metabolismo celular se debe a la redundancia, la regulación y la estrecha interacción del metabolismo y todos los demás procesos celulares. El metabolismo ha evolucionado para salvaguardar el crecimiento y el mantenimiento de las células, cuando se intenta diseñar el metabolismo para redirigir los flujos metabólicos hacia un metabolito específico, la regulación dentro de la célula luchará por mantener la homeostasis y, por lo tanto, contrarrestar los esfuerzos de ingeniería. El metabolismo subterráneo al ser el metabolismo nativo, que se puede modificar levemente y ser una herramienta viable y llamativa que coadyuvaría en la ingeniería metabólica de microorganismos.

Enfoques de ingeniería de proteínas para biocatálisis, evolución dirigida, quimioinformática y metagenómica, hacen uso del conocimiento de la promiscuidad enzimática en procariontes para optimizar el diseño de cepas sobreproductoras de compuestos de interés industrial.

### 3.1. Formación de nuevas rutas metabólicas

Un posible enfoque para superar las limitaciones que impone la expresión de genes heterólogos en los genomas procariontes para las cepas biofábricas, es explotar el metabolismo subterráneo (i.e. el repertorio de actividades enzimáticas secundarias en una célula) [D'Ari & Casadesús 1998].

Además de las actividades nativas predominantes, la mayoría de las enzimas muestran actividades secundarias débiles mediante las cuales, ciertos sustratos afines se convierten en productos, no obstante, a tasas bajas, debido a la limitada especificidad de sustrato de la enzima [Rosenberg & Commichau, 2019].

El efecto fisiológico de tales reacciones subterráneas es ignorado la mayor parte de las veces en el diseño de cepas, debido a su cinética ineficiente, a pesar de que usualmente se pueden mejorar con unas pocas mutaciones (Khersonsky & Tawfik 2010; Notebaart et al. 2014; Romero-Rivera et al. 2022). Dichas modificaciones genéticas permiten que el metabolismo subterráneo contribuya a la optimización del funcionamiento celular (e.g. crecimiento microbiano o adaptación a nuevas condiciones nutricionales) [Notebaart et al. 2014; Cam et al. 2016; King et al. 2017; Notebaart et al. 2018].

Las actividades subterráneas se han utilizado para el desarrollo de rutas metabólicas *de novo* en algunos casos de estudio como se mencionó en la *sección 2.6*.

Recientemente, dicha estrategia ha surgido como una alternativa a las enzimas heterólogas, osea el reclutamiento de actividades secundarias (actividades promiscuas) de las enzimas del huésped para establecer nuevas vías biosintéticas. Aunque estas actividades enzimáticas "subterráneas" ineficientes son frecuentes, son pocos los estudios que exploraran si pueden servir como un reservorio

importante para la ingeniería de rutas, se ha explorado en mayor parte en el organismo modelo *E. coli*, pero no en el *phylum* actinobacteria.

Kovács & cols. (2022) estimaron con GSMM los potenciales teóricos de las reacciones subterráneas para diseñar nuevas vías biosintéticas en *E. coli*, encontraron con los análisis *in silico*, que la contribución de reacciones bioquímicas catalizadas por actividades enzimáticas subterráneas, mejoran frecuentemente la producción de compuestos con importancia industrial, incluyendo algunos casos donde las actividades subterráneas son indispensables para la biosíntesis. Las reacciones subterráneas y heterólogas tienen potenciales teóricos similares cuando contribuyen a la producción de compuestos de valor agregado.

La mayoría de estas nuevas capacidades se pueden lograr mediante la adición de una o dos reacciones subterráneas a la red nativa, lo que sugiere que solo se deben mejorar algunas actividades secundarias durante la implementación.

Se ha reportado que la contribución de las reacciones subterráneas a la producción de compuestos de valor agregado es comparable a la de las reacciones de enzimas heterólogas del mismo compuesto de interés, lo que subraya su potencial biotecnológico; este estudio demostró, que la explotación de las actividades enzimáticas subterráneas puede ser una herramienta prometedora de ingeniería metabólica en el desarrollo de cepas industriales [Pontrelli et al. 2018], al incrementar el rendimiento de producción de numerosos compuestos industrialmente relevantes e incluso permitir la producción de nuevos compuestos que no pueden ser producidos por la red metabólica nativa [Kovács et al. 2022].

La producción de un determinado compuesto de valor agregado a través del metabolismo subterráneo a menudo depende de una o unas pocas reacciones subterráneas únicamente. Esto implica que sería suficiente diseñar solo unas pocas actividades secundarias enzimáticas para cualquier compuesto objetivo dado [Kovács et al. 2022].

La ingeniería de un pequeño número de pasos enzimáticos probablemente beneficiaría una estrategia de ingeniería metabólica exitosa, especialmente porque se trata de actividades del lado débil de la enzima.

Las mutaciones que aumentan la actividad subterránea podrían diseñarse mediante diversas técnicas de edición del genoma, como los métodos basados en CRISPR-Cas y MAGE [Wang et al. 2009; Jakočiūnas et al. 2016; Csörgő et al. 2020]. La evolución adaptativa de laboratorio ha demostrado tener éxito en el aumento de las actividades subterráneas [Pontrelli et al. 2018; Guzmán et al. 2019]; a su vez se puede aplicar conjuntamente edición y evolución [Pontrelli et al. 2018; Wannier et al. 2020].

El uso de reacciones subterráneas junto con evolución adaptativa no conlleva la introducción de ADN específico de otros organismos, por lo que no son GMO.

Las actividades enzimáticas subterráneas pueden tener un potencial limitado en comparación con algunas actividades enzimáticas heterólogas en un área particular de aplicación (e.g. muchos metabolitos industrialmente relevantes se producen a través del metabolismo especializado de plantas, y es poco probable que se lleguen a producir mediante actividades subterráneas de enzimas microbianas del huésped) [Nielsen & Keasling 2016; Zhang et al. 2022]. Lo anterior sugiere que combinar los dos repertorios de reacción puede ser ventajoso para aplicaciones industriales [Ko et al. 2020].

Las actividades enzimáticas subterráneas no deberían interferir con la estructura de la red metabólica nativa para volverse bioquímica y fisiológicamente funcionales. Por esta razón, podría ser un desafío afinar las actividades subterráneas *in vivo*, especialmente si se necesita una ingeniería de proteínas extensa [Porokhin et al. 2021]. Normalmente se utilizan en la actualidad tres o más estrategias para mejorar los rendimientos de las cepas biofábricas en ingeniería metabólica.

El conocimiento de las actividades subterráneas aún es rudimentario y existe el potencial de explotación de muchas más actividades secundarias de las que se conocen actualmente que podrían reclutarse para nuevas vías.

Existen referencias recientes y prometedoras de quimiinformática, estructura enzimática y “*deep learning*” que predicen reacciones metabólicas secundarias [Carbonell et al. 2014; Koch et al. 2017; Amín et al. 2019; Mou et al. 2021; Patra et al. 2023]. El desarrollo de este conocimiento ampliara aún más el conocido “*espacio de reacción metabólica*” de las especies (i.e. el número total de reacciones metabólicas que podrían estar potencialmente activas en una especie) [Tyzack et al. 2019; Hafner & Payne et al. 2021].

Las reacciones subterráneas pueden ser utilizadas como materia prima para construir nuevas vías, sin embargo, las actividades promiscuas de las enzimas en la red metabólica existente también pueden afectar negativamente la producción de compuestos objetivo [Kim & Copley 2012]. e.g. se ha demostrado que las actividades promiscuas de enzimas fosfatasas de *E. coli* desvían el flujo de una ruta biosintética heteróloga de terpenoides, reduciendo su eficiencia [Wang et al. 2018].

El enfoque de explotación del metabolismo subterráneo (promiscuo) puede proporcionar nuevos entornos de reducción de costos industriales, creando rendimientos similares o incluso mayores a los ya establecidos [Kovács et al. 2022]. Una manera de abordar el problema es mediante el uso de herramientas de modelado del metabolismo celular, como los GSMM.

## 4. Preguntas de Investigación

- I. ¿Un GSMM logra modelar la promiscuidad enzimática?
- II. Y en todo caso que así sea... ¿Cuál es su alcance?
- III. ¿Por qué algunas enzimas evolucionan para convertirse en especialistas, mientras que otras conservan características generalistas? desde un enfoque de biología de sistemas, al comparar los flujos de los subsistemas de biosíntesis de L-aminoácidos.

# 5. Objetivos

## 5.1. Objetivo General

Reconstruir redes metabólicas a escala genómica de diferentes organismos del *phylum* Actinobacteria, y analizar la promiscuidad enzimática de sustrato de PriA bifuncional y una variante, en tres GSMM publicados, de cepas en las cuales se han reportado con anterioridad, parámetros cinéticos experimentales de PriA y variantes.

## 5.2. Objetivos Específicos

- I. Reconstruir borradores de GSMM de manera semiautomática.
- II. Analizar la sensibilidad de flujos relativos en tres GSMM comparándolos y evaluando la repercusión de la promiscuidad enzimática de PriA y su variante SubHisA1.
- III. Analizar los GSMM a nivel sistema, vía, sustrato, precursores e intermediarios.
- IV. Contextos genómicos comparativos del gen *priA* y variante SubHisA1 con el algoritmo CORASON en tres genomas actinobacterianos.

# 6. Estrategia Experimental

## 6.1. Reconstrucción semiautomática de borradores de GSMM

Se utilizó la plataforma interactiva en línea de la base de conocimientos de biología de sistemas (KBase) del *department of energy* (DOE) [Arkin et al. 2018], para reconstruir los borradores preliminares de GSMM de veintinueve microorganismos (*para mayor descripción de dichos organismos, ver Anexo I, Tabla I.c.*).

El proceso de reconstrucción semiautomático de los borradores de GSMM se realizó siguiendo parcialmente los pasos recomendados por Thiele & Palsson (2010). Con los siguientes pasos generales:

- i) Utilización de diferentes herramientas computacionales para la anotación funcional de cada uno de los genomas.
- ii) Fusión y combinación de los datos resultantes de la anotación funcional.
- iii) Reconstrucción del borrador de modelo usando el planteamiento ModelSEED y “Gapfill” para permitir el crecimiento celular y otorgar continuidad (llenado de brechas) a las vías incompletas.
- iv) Validación temprana del borrador con la herramienta MEMOTE. La **Fig. 6.1.1** es un esquema del flujo de trabajo que se realizó para este apartado. A continuación, se describe cada uno de los pasos.

### 6.1.1. Utilización de diferentes herramientas computacionales para la anotación funcional

Se recopilaron los genomas completos de referencia de veintinueve organismos actinobacterianos del [Centro Nacional de Información Biotecnológica](#) (NCBI), (*ver Anexo I, Tabla I.c.*). Posteriormente en Kbase (marco computacional en línea) [Arkin et al. 2018], se creó la Narrativa

titulada "[Promiscuos PriA](#)", disponible en la liga, a la cual se puede acceder si se es usuario de Kbase y con permiso previamente solicitado.

La Narrativa descrita por Arkin y cols. (2018), "es un documento interactivo, dinámico y que permanece en el transcurso del tiempo, creado por usuarios de Kbase, que promueve la ciencia abierta, reproducible y colaborativa. Se trata de la interfaz de usuario principal de KBase, la *interfaz Narrativa*. Desde dicha interfaz, la cual se basa en la plataforma Jupyter, los usuarios pueden cargar datos privados, buscar y recuperar datos de referencia públicos extensos, acceder a datos compartidos por otros, compartir sus datos con otros, seleccionar y ejecutar aplicaciones en sus datos, ver y analizar los resultados de esas aplicaciones y registrar bitácoras con interpretaciones junto con los pasos del análisis." Estas actividades se llevan a cabo dentro de un entorno tipo "*cuaderno*" de apuntar y hacer clic. Cuando un usuario comienza un nuevo experimento computacional en KBase, crea un nuevo "*cuaderno*" (denominado **Narrativa** en KBase) para guardar este experimento. Cada acción realizada por un usuario aparece como una "*celda*" en la Narrativa. Las celdas de la aplicación muestran los parámetros de entrada elegidos para la aplicación y los resultados del análisis. Las celdas desplegadas permiten a los usuarios agregar texto con formato y figuras a una Narrativa para describir el proceso de desarrollo (tipo bitácora) detrás del flujo de trabajo científico que se está elaborando [Arkin et al. 2018] (*se muestra una captura de pantalla en el Anexo II del entorno Kbase Narrativa*).

Dentro de la Narrativa interactiva que se creó, titulada "*Promiscuos PriA*" se ubicaron los mismos genomas de NCBI, los cuales se encontraban disponibles como objetos dentro de la plataforma Kbase. Una vez ubicados los genomas y llamados dentro de la Narrativa, se descargó de Kbase un archivo fasta con las secuencias de proteínas (o CDS) utilizando la aplicación "Text Reports - Genome" para asegurar que los ID's de los genes importados (con anotaciones funcionales externas) coincidan con los ID's de genes en el objeto del genoma de la Narrativa de Kbase.

Griesemer & cols. (2018) promueven y explican que la combinación de múltiples herramientas de anotación funcional puede mejorar en gran medida la cobertura del genoma y el tamaño de la red metabólica, especialmente para organismos no modelo y vías metabólicas no centrales. Las herramientas de anotación tienden a ser incompletas e inconsistentes; incorporar anotaciones de varias herramientas puede lograr una mejor cobertura del genoma y/o obtener una anotación

consenso de mejor calidad [Griesemer et al. 2018]. Esto último es especialmente importante para genomas que son evolutivamente distantes (como es el caso de actinobacterias) de organismos modelo bien estudiados como *E. coli*.

Se utilizaron cinco herramientas diferentes para la anotación funcional. Con los archivos fasta que contienen las regiones codificantes (coding sequence, CDS) se procedió a efectuar tres anotaciones automáticas externas a Kbase (Blast Koala, DeepEC y NetGO 3.0). Con los objetos de los genomas y aplicaciones dentro de la Narrativa de Kbase se realizaron dos anotaciones RAST y Prokka para los veintinueve genomas procariotas. A continuación, se describen estas cinco herramientas.

**RAST** es un servidor *web* (disponible en [RAST](#)) para anotar funcionalmente genomas de bacterias y arqueas, basado en la homología de secuencias con funciones descritas. En la versión clásica de RAST cuando se sube un nuevo genoma a RAST, se llama a los genes y se realizan las anotaciones funcionales en comparación con la colección FIGfam (i.e. subsistemas de proteínas derivadas de colecciones de familias de proteínas funcionalmente relacionadas). Si el genoma se hace público, se aloja dentro de SEED y sus proteínas pueblan la colección FIGfam. Este ciclo de anotación ha demostrado ser una solución robusta y escalable al problema de anotar el número exponencialmente creciente de genomas. Hasta el año 2014, más de 12,000 usuarios en todo el mundo habían anotado más de 60,000 genomas distintos utilizando RAST [Overbeek et al. 2014]. Sin embargo, para este trabajo de tesis; la Narrativa de Kbase dispone de la aplicación “Annotate Microbial Genome with RASTtk - v1.073”, para anotar con RAST cualquier genoma, en este caso los veintinueve genomas, quedando el objeto genómico con su correspondiente anotación funcional; se trata del mismo método sin necesidad de utilizar el servidor RAST en línea [Arkin et al. 2018]. RAST en Kbase genera ID’s de funciones que son coherentes con la nomenclatura del conjunto de modelos metabólicos del subsistema ModelSEED (el cual utiliza este tipo de datos de anotación para asociar genes con reacciones bioquímicas y luego crear un modelo de FBA) que se construirán más adelante en este flujo de trabajo.

**Prokka** es otra herramienta de anotación funcional integrada en KBase, se basa en herramientas externas de predicción de características funcionales para identificar las coordenadas de características genómicas dentro de los *contigs*. Hace uso de herramientas como Prodigal [Hyatt,

2010], RNAmmer [Lagesen et al. 2007], Aragorn [Laslett & Canback, 2004], SignalP [Petersen et al. 2011] e Infernal [Kolbe & Eddy, 2011], todas ellas, excepto Prodigal, proporcionan coordenadas y etiquetas apropiadas para describir la característica funcional de los genes [Seemann, 2014]. Dentro de la Narrativa de Kbase está disponible la aplicación “Annotate Assembly and Re-annotate Genomes with Prokka - v1.14.5”, para anotar con Prokka cualquier genoma, en este caso los veintinueve genomas, quedando el objeto con su anotación funcional correspondiente, sin necesidad de utilizar la tubería computacional de Prokka en Linux [Arkin et al. 2018].

KEGG (**Blast KOALA**) es una herramienta de anotación funcional automática interna de KEGG [Kanehisa et al. 2023], (disponible como un servidor *web* en [Blast KOALA](#)). Las funciones moleculares se almacenan en la base de datos de KO (Ortología KEGG) que contiene ortólogos de genes/proteínas que han sido caracterizados experimentalmente. Blast KOALA asigna números K (identificadores KO) a los datos de secuencia del genoma mediante búsquedas de similitud en BLAST, contra un conjunto no redundante de genes KEGG [Kanehisa et al. 2016]. Una vez obtenidos los resultados de cada genoma subido al servidor de Blast KOALA, y con dichos datos de anotación funcional convertidos en archivos de formato TSV, se incorporaron a la Narrativa “*Promiscuos PriA*” de Kbase, en los correspondientes objetos genómicos existentes utilizando la aplicación “Import Annotations from Staging” [Arkin et al. 2018].

**DeepEC** es un marco computacional ejecutable en Python basado en aprendizaje profundo (*deep learning*) que predice con alta calidad y rendimiento, números de la comisión de enzimas (EC) previstos para 33,942,253 secuencias de proteínas, los EC son esenciales para comprender con precisión las funciones catalíticas de las enzimas en la construcción de redes metabólicas a escala del genoma de alto rendimiento. En el procedimiento de anotación funcional de un genoma, a la secuencia de proteína para un gen metabólico se le asignan números EC que corresponden a un esquema de clasificación numérica de enzimas basado en patrones de reacciones químicas relevantes. El número EC sirve para asociar una secuencia de proteínas con reacciones químicas relevantes. El número EC consta de cuatro números de nivel, con cada número separado por un punto (i.e. A. B. C. D), los números del primero al cuarto corresponden a: clase, subclase, sub-subclase y número de serie, sc. [Ryu et al. 2019].

DeepEC consta de tres redes neuronales convolucionales (CNN, por sus siglas en inglés) independientes que realizan 3 tareas de clasificación diferentes para una sola secuencia de enzima proporcionada como entrada. La primera CNN, denominada CNN-1, clasifica si una secuencia de proteína dada es una proteína enzimática o no enzimática. La segunda (CNN-2) y tercera (CNN-3), predicen los números EC de tercer y cuarto nivel, sc. [Ryu et al. 2019]. Una enzima puede tener múltiples números de EC si es una enzima promiscua (*ver sección 2.4, para los dos EC de PriA*). Los resultados de la anotación funcional para cada genoma se obtuvieron en la consola Linux UBUNTU 7.10 con Python 3.6. (el código fuente abierto se encuentra en [DeepEC](#)). Posterior, con los datos de la anotación funcional, contenidos en un archivo formato TSV se incorporaron a la Narrativa “*Promiscuos PriA*” de Kbase en los correspondientes objetos de los genomas existentes, utilizando la aplicación “Import Annotations from Staging” [Arkin et al. 2018].

**NetGO 3.0** es un método que predice anotaciones GO (“*Gen Ontology*”), (disponible como un servidor *web* en [NetGO 3.0](#)); lleva a cabo la *predicción automática de la función* (PAF) de las proteínas, se considera de última generación, ya que integra información de múltiples fuentes para mejorar el rendimiento. Sin embargo, utiliza principalmente proteínas con anotaciones funcionales respaldadas experimentalmente y, por lo tanto, ignora información valiosa de una gran cantidad de proteínas. Incorpora un *modelo de lenguaje de proteínas* para aprender (basado en *machine learning*) representaciones informativas de secuencias de proteínas basadas únicamente en la “auto-supervisión” de 250 millones de secuencias de proteínas que abarcan la diversidad evolutiva [Rives et al. 2021]. NetGO 3.0 es la versión mejorada de NetGO 2.0 mejorando ampliamente el rendimiento de PAF, con el reemplazo de *secuencias de redes neuronales recurrentes* (Seq-RNN) por el *modelado a escala evolutiva de regresión logística* (LR-ESM) basado en el *modelo de lenguaje de proteínas* para entrenarlo [Yao et al. 2021]. Envié las secuencias de los veintinueve genomas por e-correo (debido a que el servidor *web* solo acepta secuencias de un solo gen, y no de genomas completos) al servidor del Centro de Biología de Sistemas Computacionales, ISTBI, de la Universidad de Fudan, en Shanghái, China; recibiendo la respuesta con la anotación funcional NetGO 3.0 de los genomas por parte del servidor de la universidad al cabo de 1-3 días por el mismo medio. Una vez obtenidos los resultados de NetGO 3.0 de cada genoma, los datos de la anotación funcional contenidos en un archivo formato

TSV, se incorporaron a la Narrativa “*Promiscuos PriA*” de Kbase en los correspondientes objetos de los genomas existentes utilizando la aplicación “Import Annotations from Staging” [Arkin et al. 2018].

### 6.1.2. Fusión y combinación de los datos resultantes de las anotaciones funcionales

Una vez incorporados los datos resultantes de las cinco herramientas de anotación funcional en la Narrativa de Kbase en sus correspondientes objetos genómicos; se utilizó la aplicación “Merge Metabolic Annotations” para efectuar una fusión de las múltiples anotaciones en una sola anotación consenso con cuatro diferentes criterios.

Los cuatro criterios son:

- i) **Unión**; con los parámetros  $weight=1$ ,  $threshold=0.5$ , manteniendo todas las anotaciones por encima del “*threshold*”. Se **unen** las cinco anotaciones funcionales. Esto genera la mayor cantidad de reacciones y genes anotados, pero también incluye la mayor cantidad de falsos positivos y, a menudo, asigna múltiples funciones por gen.
- ii) **Regla del mayor**; con los parámetros  $weight=1$ ,  $threshold=0.5$ , manteniendo solo la anotación de puntuación más alta. Esto produce la misma cantidad de genes anotados que el criterio Unión, e incluye aún una considerable cantidad de falsos positivos para aquellos genes anotados por una sola fuente. Pero las anotaciones tienden a ser más confiables con la cantidad de fuentes que cubren un solo gen. Este criterio puede perder funciones secundarias para genes que son verdaderamente multifuncionales.
- iii) **Dos o más**; con los parámetros  $weight=1$ ,  $threshold=1.5$ , manteniendo todas las anotaciones por encima del “*threshold*”. Este criterio solo incluye las anotaciones funcionales que concuerdan en al menos dos fuentes diferentes. Esto puede ser demasiado restrictivo, pero genera un núcleo de reacciones anotadas de manera más confiable.
- iv) **Prioridad**; con los parámetros  $threshold=0.5$ , el  $weight$  corresponde a que cada fuente de anotación obtiene un peso igual a dos veces el de la siguiente fuente más confiable ( $16 > 8 > 4 > 2 > 1$ , siendo 16 el de mayor confiabilidad) manteniendo solo la anotación de puntuación más alta. Este criterio se limita con la confiabilidad en algunas fuentes de anotaciones sobre otras, el esquema

de puntuación asigna a cada gen la anotación de la fuente más confiable entre las que anotan dicho gen. Este criterio puede perder funciones secundarias para genes que son verdaderamente multifuncionales. Se eligió el siguiente orden de confiabilidad de mayor a menor RAST > Koala KEGG > NetGO 3.0 > DeepEC > Prokka.

Para elegir entre cuales fuentes de anotación eran más confiables sobre otras, se consideró a RAST como la más confiable debido a que a pesar de que DeepEC demostró ser mejor que RAST acorde con Ryu & cols. (2019), ModelSEED basa la reconstrucción de los borradores de GSMM en los términos de RAST; Koala KEGG se considera una de las bases de datos más completas y actualizadas hasta el día de hoy [Kanehisa et al. 2023]; NetGO 3.0 se colocó una posición por arriba de DeepEC debido a que la actualización más reciente del servidor de NetGO fue en 2021 [Yao et al. 2021] a diferencia de la base de datos de DeepEC del 2019 [Ryu et al. 2019]; la base de datos Prokka se colocó en último lugar debido a que los datos y herramientas que utiliza datan de antes del año 2014 [Seemann, 2014].

La actualización es importante en la anotación funcional, debido a que constantemente existen nuevos tipos de secuenciación y datos que revelan el descubrimiento de funciones de los genes/enzimas, que eran desconocidas.

Los cuatro criterios de fusión de anotaciones que se han descrito se basaron en el tutorial de Kbase titulado "[Importing annotation sources for metabolic modeling](#)" (2021) disponible en esta liga.

Después de la anotación funcional con las cinco diferentes herramientas dentro de Kbase, se utilizó la aplicación "Compare Metabolic Annotations" la cual realiza una comparación en paralelo de las múltiples anotaciones funcionales (funciones metabólicas) mapeadas en los genomas, observando así qué anotación funcional y fusión consenso con cada uno de los cuatro criterios, proporciona más información a diferentes niveles (*ver 7. Resultados sección 7.1*).

### **6.1.3. Reconstrucción de borradores de GSMM con ModelSEED y "Gapfill"**

Una vez con los genomas anotados, con disponibilidad de cuatro genomas (debido a los cuatro criterios) con diferente fusión de anotación para cada uno de los veintinueve genomas iniciales (un total 116 genomas). Se construyeron 116 borradores de GSMM basados en cada uno de los 116

genomas anotados, con la aplicación de Kbase “Build Metabolic Model” que engloba el enfoque de ModelSEED [Henry et al. 2010; Seaver et al. 2021].

Los borradores de modelos metabólicos preliminares se componen de *i)* una red de reacciones completa con asociaciones GPR, *ii)* energía libre de Gibbs predicha de los valores de reacción y, *iii)* la reacción de biomasa. La reacción de biomasa incluye cofactores, lípidos y componentes de la pared celular; y es específica del organismo, se basa en una plantilla de reacción de la biomasa, que utilizan los subsistemas SEED y las anotaciones funcionales RAST para asignar componentes de biomasa no universales (e.g. cofactores, componentes de la pared celular, entre otros), que representan funciones biológicas únicas, exhibidas por un conjunto grande de organismos o específico de un pequeño conjunto de organismos [Seaver et al. 2021].

Para que un componente de biomasa específico de un organismo se agregue a la reacción de biomasa, en su genoma debe contener los subsistemas y anotaciones adecuados especificados en la plantilla. Las asociaciones GPR representan el mapeo entre las reacciones bioquímicas y los roles funcionales estandarizados asignados a los genes durante la anotación RAST [Overbeek et al. 2014]. Este mapeo permite que se diferencie entre casos en los que los productos proteicos de varios genes forman un complejo para catalizar una reacción y los casos en los que los productos proteicos de varios genes pueden catalizar la misma reacción de forma independiente.

El borrador del modelo incluye todas las reacciones asociadas con una o más enzimas codificadas en el genoma que se identifican en las anotaciones. Además, se agregan reacciones espontáneas (i.e. catálisis no enzimáticas) durante este paso. Todas las plantillas utilizadas para la reconstrucción de los modelos se pueden encontrar en Github.

En las continuas actualizaciones del enfoque ModelSEED automatizado, se ha mejorado la automatización del cálculo de la *producción de ATP del organismo* en el procedimiento de reconstrucción del modelo al construir modelos de metabolismos centrales, probar la producción adecuada de ATP a partir de este núcleo y luego garantizar que la producción de ATP no estalle incorrectamente al expandir el modelo central a un GSMM que abarca además metabolismo de subsistemas no centrales, como serían los subsistemas de biosíntesis de L-aminoácidos [Seaver et al. 2021].

El enfoque de llenado de brechas (“*Gapfill*”), garantiza que un modelo no produzca ATP en exceso. “*Gapfill*” es el proceso mediante el cual la aplicación de Kbase “Build Metabolic Model” identifica el conjunto mínimo de reacciones bioquímicas a agregar a un borrador de GSMM para permitirle producir biomasa en un medio específico [Latendresse, 2014]. Para la construcción en específico de los 116 borradores, se eligió el medio mínimo mineral suplementado con glucosa como fuente de carbono.

La calidad de un proyecto de reconstrucción de un modelo metabólico depende de la integridad del genoma anotado utilizado para la reconstrucción preliminar. Debido al hecho de que la mayoría de los genomas no están completamente anotados, los borradores de los modelos metabólicos suelen contener lagunas que impiden la producción de algunos componentes de la biomasa. En este paso, el algoritmo de optimización “*Gapfill*” identifica el conjunto mínimo de reacciones que deben agregarse a cada modelo para llenar estos vacíos [Dreyfuss et al. 2013; Latendresse, 2014].

Las reacciones que se utilizan para el llenado de lagunas se seleccionan de la base de datos bioquímicos de ModelSEED [Seaver et al. 2021]. Esta base de datos contiene reacciones estandarizadas, en equilibrio en cuanto a masa y carga, para condiciones acuosas a pH neutro. La base de datos de reacciones de ModelSEED integra el conocimiento bioquímico contenido en KEGG, MetaCyc, EcoCyc, Plant BioCyc, Plant Metabolic Networks y Gramene [Seaver et al. 2021]. “*Gapfill*” se lleva a cabo para garantizar que cada borrador preliminar de un GSMM sea capaz de simular el crecimiento celular.

#### **6.1.4. Validación temprana del borrador de GSMM**

Una vez construidos los 116 borradores, se utilizó la herramienta MEMOTE versión 0.11.1. [Lieven & Moritz et al. 2020], para validar los borradores reconstruidos. MEMOTE es una plataforma en línea que pone a prueba medidas estandarizadas de modelos metabólicos o sus borradores y genera puntuaciones de control de calidad que van desde el 0 % para una calidad del modelo deficiente hasta el 100 % para una calidad del modelo excelente [Lieven & Moritz et al. 2020].

Las medidas que generan las puntuaciones evalúan la consistencia y las anotaciones del modelo dentro de diferentes categorías. Estas categorías incluyen información básica sobre el modelo, los metabolitos y las reacciones, el grado de anotaciones para metabolitos, reacciones, genes y términos SBO (“*system biology ontology*”, por sus siglas en inglés,). MEMOTE también comprueba la presencia de GPR, una función de biomasa realista, el metabolismo energético y la topología adecuada de la red. Además de estos puntajes individuales de las diferentes subcategorías, MEMOTE también informa un puntaje general. Este puntaje general representa una medida general de qué tan bueno es el modelo dentro de todas las categorías individuales.

Para valorar la consistencia estequiométrica del modelo o borrador, MEMOTE evalúa la estequiometría, balances de masa y carga, la conectividad de los metabolitos y los flujos ilimitados en el medio predeterminado. Dentro de la evaluación de las anotaciones, MEMOTE verifica la presencia y conformidad de varias bases de datos y la presencia de términos SBO específicos. Todas las categorías se puntúan individualmente, y la puntuación de la calidad general otorgada por MEMOTE se calcula en función de la suma y peso de las diferentes puntuaciones de las categorías individuales [Lieven & Moritz et al. 2020; Feierabend & Renz et al. 2021].

Finalmente se seleccionaron los veintinueve mejores borradores de GSMM, con base en la puntuación general arrojada por MEMOTE para 116 borradores (4 borradores por genoma), se redujo el número de borradores, considerando solo aquellos más confiables y con mayor puntaje (i.e. porcentaje), para cada organismo, no considerando a los de menor puntaje.

Los veintinueve borradores de GSMM con mayor puntaje MEMOTE, de los organismos del *phylum* actinobacteria en los cuales se ha detectado y reportado la enzima promiscua PriA y subespecializaciones, están disponibles en la Narrativa de Kbase “*Promiscuos PriA*”.

Los borradores pueden ser utilizados para el siguiente paso en la reconstrucción de un GSMM, que es una curación manual extensa y a detalle, de revisión bibliográfica de cada uno de los metabolismos de los organismos y la introducción de dichos datos en el borrador [Thiele & Palsson, 2010]. Dicha curación manual no se efectuó para los borradores aquí fabricados, debido a que la curación de un solo modelo requiere un mínimo de 6 meses o más tiempo, en conjunto con la colaboración de personal académico, debido a lo anterior, esta curación manual de los veintinueve

genomas escapa del tiempo considerado para este trabajo de tesis (Al final de la sección 7.1 del capítulo 7. Resultados, explico con mayor detalle en que consiste dicha curación manual). En la Fig. 6.1.1 se esquematiza un resumen del flujo de trabajo que se siguió para la reconstrucción de los 116 borradores de GSMM, descrita a lo largo de la sección 6.1.



Figura. 6.1.1. Flujo de trabajo para la reconstrucción de 116 redes metabólicas de GSMM a partir de genomas actinobacterianos. Se incluye el orden cronológico de reconstrucción y la referencia de cada herramienta utilizada.

## 6.2. GSMM utilizados en este trabajo

Bibliográficamente se localizaron tres GSMM reportados, previamente reconstruidos y curados manualmente, de organismos reportados dentro de los veintinueve organismos que contienen a la enzima PriA o alguna de sus subespecializadas. Dos GSMM de estos organismos contienen en sus genomas la enzima PriA bonafide, el modelo iAA1259 del organismo *Sco* A3(2) contiene 1908 reacciones, 1472 metabolitos y 1259 genes [Amara et al. 2018] y el modelo iEK1011 del organismo *Mtu* H37Rv que contiene 1228 reacciones, 998 metabolitos y 1011 genes [Kavvas et al. 2018]; además

del modelo iCGB21FR del organismo *Cgl* ATCC 13032 conteniendo 1539 reacciones, 1044 metabolitos y 805 genes [Feierabend & Renz et al. 2021] que contiene la enzima subespecializada SubHisA1.

Los GSMM que se tomaron en cuenta, son de las últimas versiones, ya que se debe considerar que dichas versiones continuamente se actualizan y mejoran con nuevas herramientas *in silico* que no se encontraban disponibles al momento de reconstruir las versiones anteriores, también fueron actualizados con nuevos datos metabólicos experimentales reportados para ciertas enzimas, genes, metabolitos y reacciones con los que no se contaba en las versiones anteriores, aumentando el conocimiento de dichas cepas, ocasionado por la secuenciación masiva del genoma de última generación y nuevas técnicas bioinformáticas (e.g. herramientas de integración de datos, bases de datos y diferentes algoritmos de optimización). Lo anterior, permite observar que los tres modelos seleccionados para esta tesis son GSMM de última generación, muy bien curados, con resultados y condiciones computacionales validadas experimentalmente en laboratorio húmedo [Kavvas et al. 2018; Amara et al. 2018; Feierabend & Renz et al. 2021]. Los tres GSMM fueron sometidos al control de calidad MEMOTE (*para los resultados obtenidos véase Tabla 7.1.1. de 7. Resultados*).

### 6.3. Restricción de los GSMM con datos experimentales

Con la herramienta COBRA Toolbox 3.0 [Heirendt et al. 2019] dentro de la interfaz del software MATLAB® 2022 versión 9.13 (R2022b), se llamó a cada uno de los tres GSMM, se procedió a desarrollar un código de comandos (o *script*) para cada uno, conteniendo las restricciones derivadas de datos con mediciones experimentales (*Ver Anexo III*).

Como se describió en las **Fig. 2.7.1.** y **Fig. 2.7.2.**, los GSMM se restringen para obtener resultados confiables con base a datos experimentales reales, usualmente provenientes de cinéticas de crecimiento, consumo de oxígeno, producción de dióxido de carbono o producción de algún metabolito excretado de interés.

El modelo iAA1259 del organismo *Sco* A3(2), se restringió con datos experimentales reportados por Melzoch & cols. (1997), donde determinan la producción del antibiótico actinorodina por *Sco*

A3(2), en un cultivo de quimiostato (ver **Anexo IV**, tabla IV.a y IV.b, para los detalles de las restricciones de este GSMM).

El modelo iEK1011 del organismo *Mtu* H37Rv, se restringió con datos experimentales reportados por los mismos autores que reconstruyeron el GSMM; Kavvas & cols. (2018), reportan los requerimientos de consumo de sustratos para el adecuado crecimiento *in vivo* de *Mtu* en los macrófagos de los alveolos pulmonares, ya que se trata de un organismo patógeno intracelular obligado, aerobio con preferencia por tejidos bien oxigenados. Sus fuentes de nitrógeno están compuestas por nitrato, aspartato, asparagina, glutamato, urea y glutamina. Los ácidos grasos se eligieron como la principal fuente de carbono *in vivo* debido a la gran cantidad de evidencia bibliográfica que respalda la afirmación de que *Mtu* utiliza ácidos grasos derivados del huésped. La alanina se incluye como nutriente debido a evidencia que describe que se encuentra en abundantes cantidades dentro de los macrófagos alveolares y se importa a la bacteria [Beste et al. 2013].

En el caso del GSMM de *Mtu* H37Rv, las restricciones incluyen consumo de los aminoácidos L-glutamina, L-glutamato, L-alanina, L-aspartato y L-asparagina, sin embargo y debido a que en esta tesis se evalúa la biosíntesis de dieciséis L-aminoácidos comparables entre los tres GSMM seleccionados, solo se aplicó la restricción de L-alanina con base en los reportes experimentales de Shi (2021), donde cultiva a *Mtu* en medio mínimo con el aporte de un solo aminoácido, evaluando veinte aminoácidos por separado, para el cultivo con L-alanina como única fuente nitrógeno en concentraciones de 0.5, 1, 2, 4 g/L el organismo crece bastante bien (+++), con el aporte único de dicho L-aminoácido; por lo que se decidió restringir el modelo solo con este aminoácido proporcionando el límite inferior de consumo en  $1 \text{ mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ , observando si el modelo del organismo lograba la biosíntesis de alguno de los otros aminoácidos que se consumían inicialmente en las restricciones proporcionadas por Kavvas & cols. (2018). El flujo para la reacción VPAMTr correspondiente a la enzima valina-piruvato aminotransferasa se cerró (con restricción de cero) para lograr un flujo único en una sola dirección, ya que dicha reacción corresponde a una reacción reversible en la ruta síntesis de L-valina, esta estrategia se utiliza para una mejor simulación de la distribución de flujos. (ver **Anexo IV**, tablas IV.c y IV.d, para mayores detalles de las restricciones de este GSMM).

El modelo iCGB21FR del organismo *Cgl* ATCC 13032, se restringió con datos experimentales reportados por Graf & cols. (2020), donde cultivaron el organismo en medio CGXII (D-glucosa es la fuente de carbono predominante en este medio mínimo. Sin embargo, *Cgl* puede utilizar glucosa, fructosa, sacarosa, ribosa, gluconato, piruvato, acetato, lactato y propionato como fuente de carbono, en condiciones aeróbicas [Unthan & Grünberger et al. 2014]) creciendo en condiciones de quimiostato limitadas en glucosa con  $\mu = 0.2 \text{ h}^{-1}$ ,  $0.3 \text{ h}^{-1}$  y  $0.4 \text{ h}^{-1}$  [Graf et al. 2020]. Para los flujos de las reacciones HSTPTr y PGM que corresponden a las enzimas Histidinol fosfato transaminasa y Fosfoglicerato mutasa sc. se cerraron (con restricción de cero) para lograr un flujo único en una sola dirección, ya que dichas reacciones corresponden a reacciones reversibles, esta estrategia se utiliza para una mejor simulación de la distribución de flujos, como ya había mencionado con anterioridad (en el GSMM de *Mtu*). (ver **Anexo IV**, tablas IV.e y IV.f, para mayores detalles de las restricciones de este GSMM).

#### **6.4. Análisis de sensibilidad de flujos relativos (Robustez y PhPP)**

Una vez con los GSMM restringidos en la herramienta COBRA Toolbox 3.0 [Heirendt et al. 2019] dentro de la interfaz del software MATLAB® 2022 versión 9.13 (R2022b) 64-bit (Win-64)., se procedió a efectuar simulaciones y diferentes análisis.

El FBA codificado en COBRA Toolbox 3.0 se utilizó como algoritmo de optimización. Además de realizar el cálculo de la distribución de flujos, para analizar el metabolismo, se llevaron a cabo diferentes análisis basados en restricciones, tales como: delección de genes [Edwards & Palsson, 2000a], robustez [Edwards & Palsson, 2000b] y fenotipos de fase plana (PhPP, por sus siglas en inglés) [Edwards et al. 2002], para explorar el metabolismo con diferentes análisis.

El análisis de robustez permite el cálculo de cómo cambia un objetivo de interés (e.g. tasa de crecimiento) a medida que el flujo a través de una reacción específica de interés varía en magnitud. El rango de valores de reacción varía dentro de los límites mínimo y máximo para la reacción de interés y la función objetivo se maximiza usando FBA de acuerdo con cada valor de flujo establecido de la reacción que se quiere evaluar [Becker et al. 2007].

Para este trabajo de tesis las dos reacciones que se evaluaron fueron las de las enzimas ProFAR isomerasa y PRA isomerasa *versus* la reacción objetivo de tasa de crecimiento (o biomasa), además de las dos enzimas contrapuestas, aunque ProFAR isomerasa y PRA isomerasa pueden estar anotadas funcionalmente bajo el mismo gen en los genomas se trata de dos reacciones diferentes en las matrices estequiométricas de los GSMM.

Este análisis proporciona un gráfico que permite apreciar la variación en magnitud de los flujos de la reacción de interés y su repercusión en el flujo de la reacción objetivo contrapuesta, cuando aumentan o disminuyen los flujos (*ver sección 7.2*).

El análisis de PhPP es un análisis de robustez doble, con el cual se pueden evaluar tres reacciones en cuestión al mismo tiempo. Este análisis proporciona un gráfico 3D con escala degradada de colores (los colores representan a la biomasa), el cual permite apreciar la variación de los flujos de las tres reacciones comprometidas, los resultados obtenidos con esta herramienta no se muestran en esta tesis, debido a su falta de relevancia.

Para el análisis de robustez y PhPP ejecute el código fuente publicado en [GitHub](#) [Vlasov, 2018] en COBRA Toolbox 3.0 en la interfaz de MATLAB® 2022.

## **6.5. Fluxómica comparativa de L-aminoácidos**

Se calculó la distribución de los flujos metabólicos con FBA para los tres GSMM seleccionados y restringidos. Se compararon los flujos intracelulares para dieciséis aminoácidos en los tres GSMM, los tres metabolismos son fisiológicamente diferentes, crecen en condiciones diferentes, y el tamaño de sus genomas difiere, por lo que las reacciones involucradas en las rutas de biosíntesis de L-aminoácidos no suelen ser las mismas; lo anterior imposibilita la homogenización comparativa de los flujos de síntesis de los veinte aminoácidos, por lo que solo se consideraron los L-aminoácidos que mantuvieran las mismas reacciones de síntesis en los tres GSMM, lo anterior aseguro su comparación; el flujo de dichas reacciones corresponde al de la enzima/reacción (última reacción de la ruta) de

síntesis del L-aminoácido en cuestión (*Para mayor detalle de las dieciséis reacciones seleccionadas y comparadas entre los tres GSMM, véase sección 7.3. de 7. Resultados*).

## 6.6. Mapas de flujos metabólicos

Con los modelos restringidos, y la simulación de FBA se calculó la distribución de flujos metabólicos; se esquematizaron las rutas de biosíntesis en cuestión, representando los flujos de las reacciones del metabolismo central parcialmente y de las vías de biosíntesis de His y Trp en mapas de flujos metabólicos para los tres GSMM seleccionados.

Los mapas de flujos se elaboraron utilizando la herramienta *Escher* [King et al. 2015]. [Escher](#) es una aplicación *web* es de uso gratuito, disponible en la liga, que permite ilustrar vías metabólicas de forma semiautomática, y cargar datos de diferente índole, en este caso se trató de datos de fluxómica. *Escher* permite mostrar los valores de los flujos de cada reacción en cada mapa, muestra los nombres abreviados (de acuerdo con identificadores de la plataforma “[BiGG Models](#)”), de metabolitos (involucrados en cada reacción) y reacciones. *BiGG Models* es una base de conocimiento bioquímico, genético y genómico, la cual integra, estandariza y comparte la información contenida de 75 GSMM seleccionados manualmente y de alta calidad [King et al. 2016].

## 6.7. Contextos genómicos comparativos para los operones de His y Trp

A manera de complementar y visualizar los operones de His y Trp en los genomas de los tres organismos modelo utilizados en los GSMM seleccionados. Se utilizó de manera poco usual, el algoritmo CORASON [Navarro-Muñoz et al. 2020] al comparar los tres genomas de *Sco*, *Mtu* y *Cgl* (*ver Anexo I, Tabla I.c.*) anotándolos funcionalmente en el servidor *web* RAST (fuera de Kbase), utilizando la secuencia de la enzima PriA como semilla; dentro de la interfaz Docker en la consola de UBUNTU 7.10 Linux, de acorde con el protocolo y código fuente publicados en [GitHub](#).

# 7. Resultados

## 7.1 Reconstrucción semiautomática de borradores de GSMM

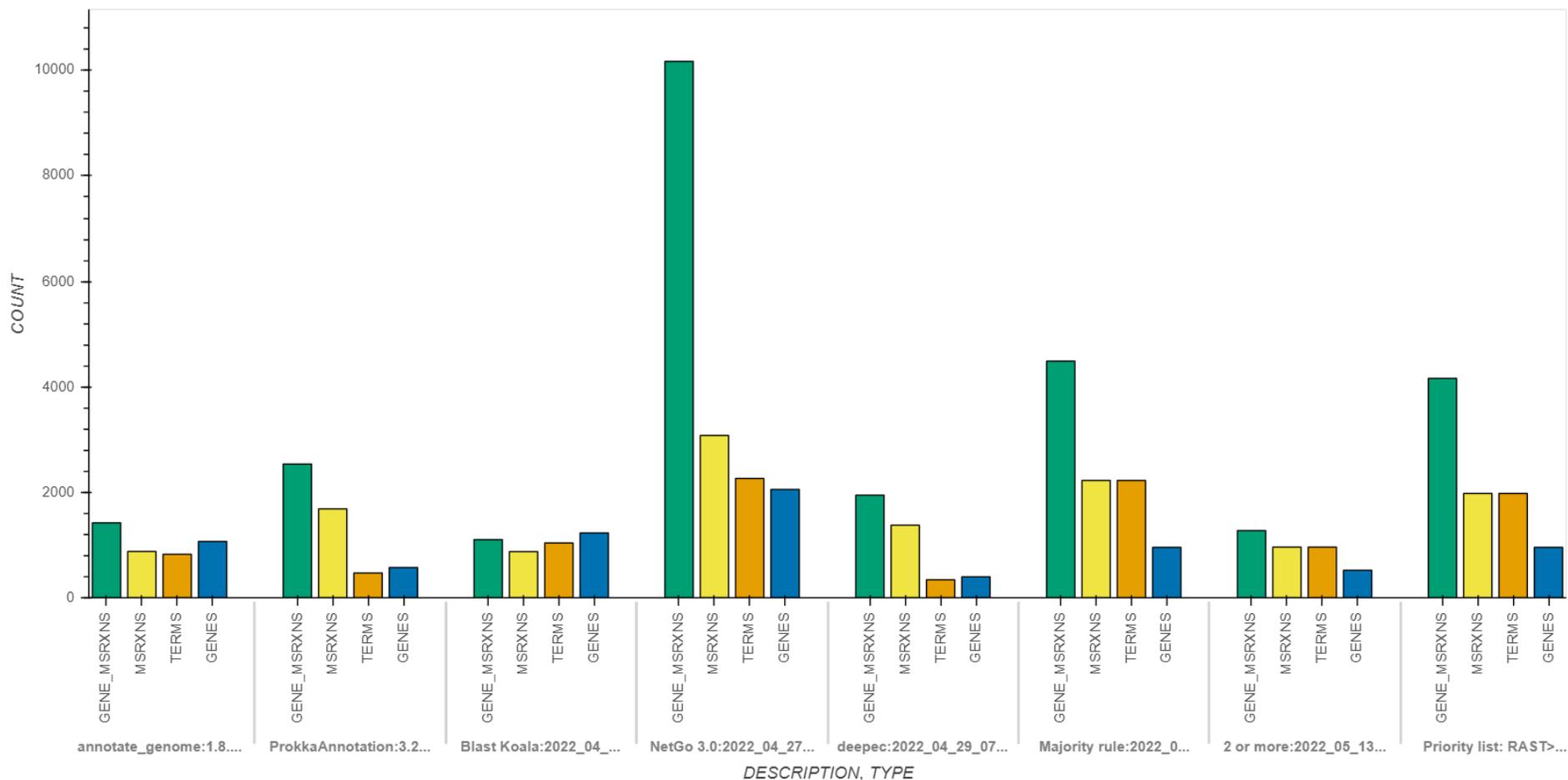
Después de la anotación funcional con cinco diferentes herramientas cargadas en Kbase, se obtuvieron informes comparativos con la aplicación “Compare Metabolic Annotations”, tanto de las cinco herramientas de anotación funcional (*sección 6.1.1.*) como de la fusión con cada uno de los cuatro criterios (*sección 6.1.2.*).

Uno de los informes gráficos [**Fig. 7.1.1.**] arrojados por dicha aplicación, el cual solo se muestra a modo de ejemplo, para el genoma de *Cgl* ATCC 13032, con el número de cada una de las características que detecta la aplicación, para cada herramienta de anotación funcional o fusión con criterios; estas características fueron:

- i)* El número de genes anotados con términos de la ontología (i.e. dependiendo si se trata de GO, KO, EC, etc.).
- ii)* El número total de términos de anotación únicos.
- iii)* A cuántas reacciones ModelSEED únicas se asignan estos términos.
- iv)* El número de pares únicos de genes/reacciones ModelSEED (importante para la reconstrucción de los borradores de GSMM con el enfoque ModelSEED) [Seaver et al. 2021].

Un informe con los mismos datos, pero diferente aspecto, se encuentra en el **Anexo I Grafico I.a.**

Informes gráficos como los que se muestran a continuación están disponibles en la Narrativa “*Promiscuos PriA*” en Kbase, para cada uno de los veintinueve objetos/genomas con los que se reconstruyeron los borradores de GSMM de este trabajo de tesis.



**Figura. 7.1.1. Resumen de las características únicas que se encuentran en un objeto del genoma de *Cgl* ATCC 13032, agrupadas por fuente de anotación y tipo de fusión con los cuatro criterios (las descripciones se truncan a los primeros 20 caracteres). RAST (annotate\_genome), Prokka (ProkkaAnnotation), Blast Koala, NetGo 3.0 y DeepEC, además de las fusiones Unión, Regla de Mayor (Majority rule), Dos o más (2 or more) y Prioridad (Priority). Anidada en cada fuente de anotación y fusión se muestra a modo comparativo las características: número de genes, de términos únicos (TERMS), reacciones ModelSEED únicas (MSRXNS), y pares únicos de gen/reacción ModelSEED (GENE\_MSrxNS). Visualización obtenida con Python Bokeh.**

Algunos tipos de anotaciones funcionales, como los ortólogos KEGG (KO) y la ontología génica (GO), incluyen muchos términos que no se asignan directamente a las reacciones metabólicas, ya que no los contiene ModelSEED o son huérfanos [Seaver et al. 2021; MohammadiPeyhani et al. 2022; Kanehisa et al. 2023].

La cantidad de reacciones ModelSEED únicas puede ser mayor que la cantidad de términos únicos (como ocurre en el caso DeepEC, Prokka y NetGO), esto debido a que se está asignando a más de una reacción los términos únicos de la anotación.

Para el caso de la anotación NetGO 3.0, y las fusiones con criterios Unión, Prioridad y Regla del Mayor, destaca el alto número de pares únicos de genes con reacciones ModelSEED (color verde), en comparación con las demás características, es muy probable que existan muchos falsos positivos para esta característica. Al parecer NetGO 3.0 esta asignando muchas reacciones en algunos genes, esto puede estar relacionado con la posible promiscuidad enzimática de dichos genes.

En los mapas de calor [Fig. 7.1.2.], otro informe gráfico concerniente a las diferentes anotaciones funcionales y fusiones con los cuatro criterios, hacen notar las concordancias (azul-verde) tanto en columnas como en filas, y discordancias (canario-rojo) solamente columnas, entre las diferentes anotaciones funcionales y criterios de fusión de anotaciones de los genomas, de nuevo solo se muestra los resultados para el organismo *Cgl* ATCC 13032 a modo de ejemplo en este apartado de resultados.

Los mapas de calor “a” y “b” reflejan las concordancias y discordancias entre el número (barra de colores) de los genes anotados con términos de la anotación funcional o tipo de fusión. En la Fig. 7.1.2.a. se observa que las anotaciones Blast Koala vs NetGO 3.0 tuvieron el mayor número de genes concordantes (>1,400) entre ambos tipos de anotación; DeepEC tiene la menor cantidad (<600) de concordancia de genes contra todos los demás tipos de anotaciones y fusiones con criterios. Con referencia a los criterios de fusión, el criterio “dos o más” tuvo el menor número de concordancias de genes (<800) contra todos los demás tipos de anotaciones y fusiones con criterios; el criterio “regla

del mayor” tuvo el mayor número de concordancias de genes (>1,200) en comparación con los demás tipos de fusiones con criterios.

En la **Fig. 7.1.2.b.** se observa que el tipo de anotación NetGO 3.0 tiene el mayor número de genes discordantes (>1,500); y DeepEC tuvo el menor número (0) de discordancias, estos dos últimos resultados son congruentes con los de la **Fig. 7.1.1.**, donde se observa que NetGO 3.0 tiene el mayor número (2,962) de genes anotados con términos de la ontología y DeepEC tiene el menor número de genes (472) anotados con términos de la ontología. Aunque NetGO 3.0 tuvo el mayor número de genes concordantes y DeepEC el menor, las discordancias parecerían contradictorias en una primera mirada, sin embargo esto se explica debido a que el número de genes anotados con NetGO 3.0 es mucho mayor que los genes anotados con DeepEC (una relación aproximada de 1 : 6.3 sc.), por lo que los pocos genes de DeepEC son concordantes con todas las demás anotaciones, al parecer los genes encontrados en las demás anotaciones no se encuentran en DeepEC y por esta razón aparece como el de menor concordancia, esto es evidente en el mapa de calor de discordancia.

Los mapas de calor “c” y “d” reflejan las concordancias y discordancias entre el número (barra de colores) de a cuántas reacciones ModelSEED únicas se han asignado los términos de la anotación funcional o tipo de fusión. En la **Fig. 7.1.2.c.** se observa que la anotación NetGO 3.0 vs el criterio de fusión Unión tuvieron el mayor número (>3000) de concordancia de reacciones ModelSEED con asignación de términos de la anotación funcional o tipo de fusión, esto es debido a NetGO 3.0 (3,413) y Unión (5,061) tienen el mayor número de reacciones ModelSEED únicas, en comparación con las demás anotaciones funcionales y criterios de fusión, por consecuencia existen más coincidencias/concordancias; en contraparte la anotación RAST tiene la menor cantidad (<1000) de concordancia de reacciones ModelSEED únicas en comparación con las demás anotaciones funcionales o tipos de fusión, esto es debido a que RAST (902) tiene el menor número de reacciones ModelSEED únicas, en comparación con las demás anotaciones funcionales y criterios de fusión, es por esta razón que existen menos coincidencias/concordancias para la anotación RAST.

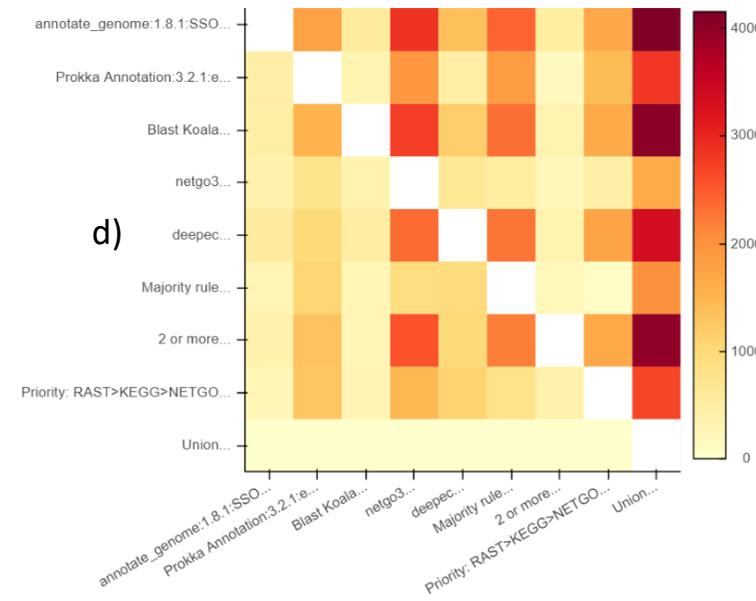
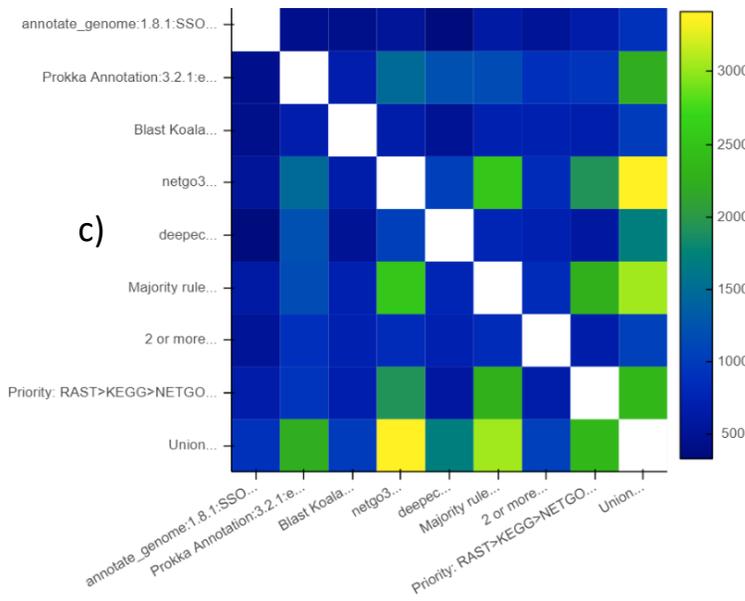
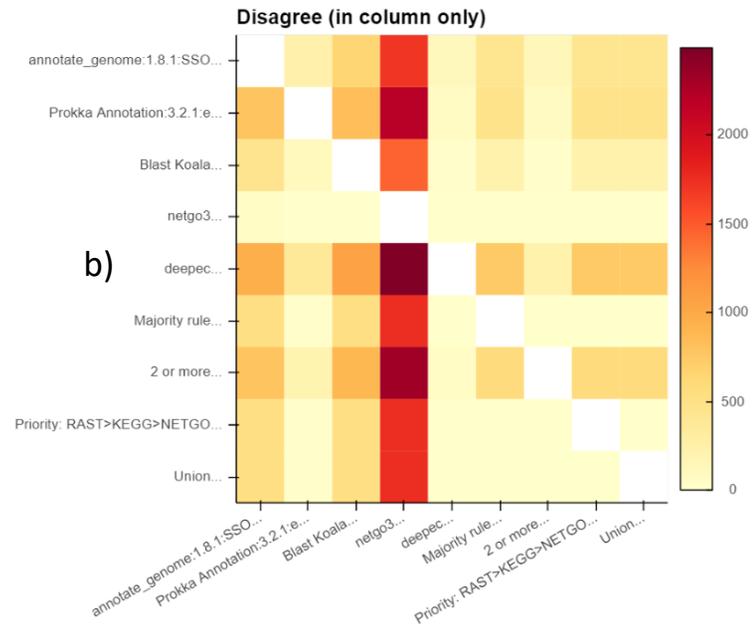
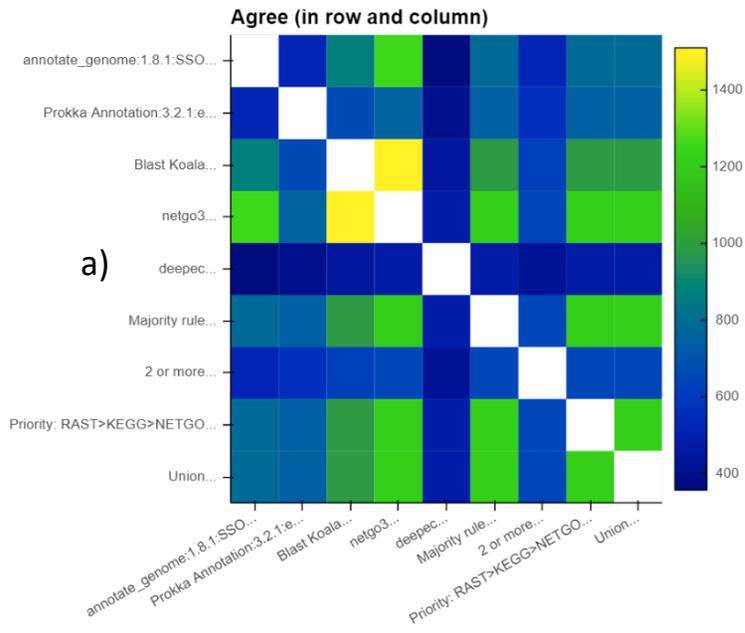
En la **Fig. 7.1.2.d.** se observa que el tipo de criterio de fusión Unión tiene el mayor número de reacciones ModelSEED únicas discordantes, principalmente con RAST (4,159), Blast Koala (4,055) y la fusión “dos o más” (4,012), que son los que tienen el menor número de reacciones ModelSEED únicas

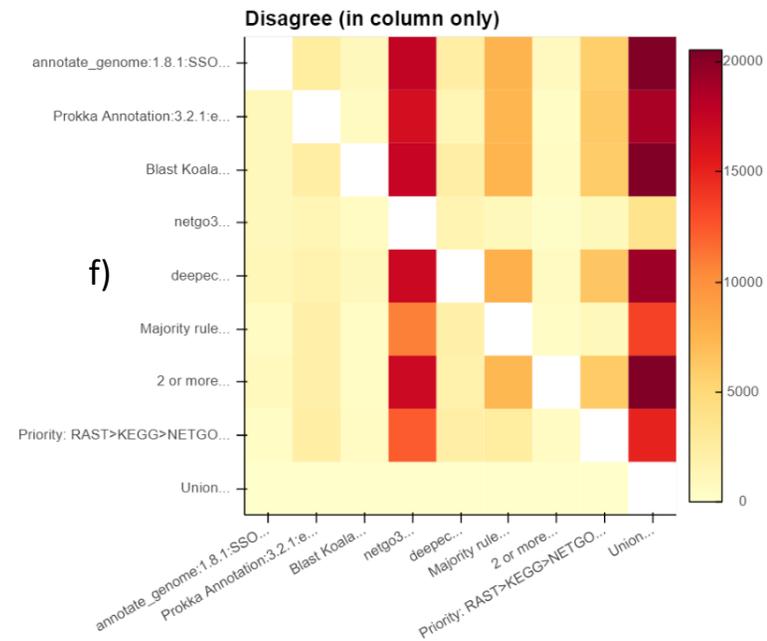
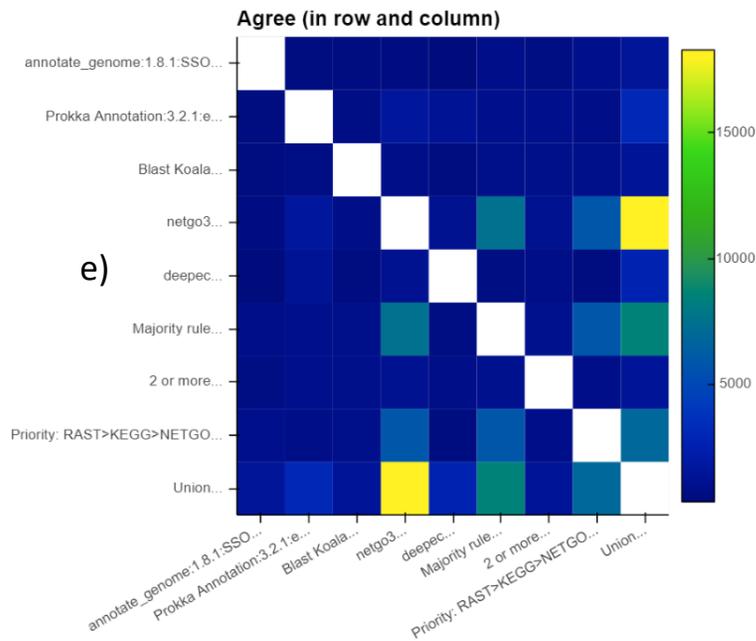
[Fig. 7.1.1.]. Los menores discordantes son RAST, Blast Koala y la fusión “dos o más”, que ocurre similar al caso de DeepEC en el mapa de calor de discordancia de la Fig. 7.1.2.b.

Los mapas de calor “e” y “f” reflejan las concordancias y discordancias entre el número (barra de colores) de pares únicos de genes/reacciones ModelSEED dependiendo de la anotación funcional o tipo de fusión.

En la Fig. 7.1.2.e. se observa que la anotación NetGO 3.0 vs el criterio de fusión Unión tuvieron el mayor número (>15,000) de concordancia de pares únicos de genes/reacciones ModelSEED, esto es debido a que NetGO 3.0 (18,264) y el criterio Unión (21,969) tienen el mayor número de pares únicos de genes/reacciones ModelSEED, en comparación con las demás anotaciones funcionales y criterios de fusión, por consecuencia existen más coincidencias/concordancias, sin embargo el criterio Unión tiende a asignar muchas reacciones a un gen y de ahí su alto índice de falsos positivos, por lo que este tipo de criterio no es el mejor; en contraparte la anotación RAST, Blast Koala y el criterio “dos o más” tienen la menor cantidad (<5000) de concordancia de pares únicos de genes/reacciones ModelSEED en comparación con las demás anotaciones funcionales o tipos de fusión, esto es debido a que RAST (1,489), Blast Koala (1,422) y “dos o más” (1,443) tienen los menores índices de pares únicos de genes/reacciones ModelSEED, en comparación con las demás anotaciones funcionales y criterios de fusión, es por esta razón que existen menos coincidencias/concordancias para estos.

En la Fig. 7.1.2.f. se observa que el tipo de criterio de fusión Unión y la anotación NetGO 3.0 tienen el mayor número de pares únicos de genes/reacciones ModelSEED discordantes, entre un rango de 10,000 a 20,000 pares principalmente con RAST, Blast Koala, DeepEC y la fusión “dos o más”, que son los que tienen el menor número de pares únicos de genes/reacciones ModelSEED [Fig. 7.1.1.]. Los menores discordantes son RAST, Blast Koala y el criterio “dos o más”, que ocurre similar a la Fig. 7.1.2.b. y Fig. 7.1.2.d.





**Figura. 7.1.2. Gráficos de mapas de calor de:** GENES (a,b), reacciones ModelSEED únicas (c,d), y pares únicos de gen/ reacción ModelSEED (e,f), que muestran las concordancias por pares (izquierda-azul) y las discordancias (derecha-rojo) entre los eventos de anotación en el objeto genómico de *Cgl* ATCC 13032. En el lado izquierdo los mapas son simétricos y muestran los números de anotaciones totales que se encuentran en los eventos de fila y columna (barra de color). Del lado derecho el mapa no es simétrico mostrando el recuento de tipos de anotaciones únicos, tomando en cuenta solo el evento de columna, pero que faltan en el evento de fila. Las descripciones se truncan a los primeros 20 caracteres. **Nomenclatura:** RAST (`annotate_genome`), Regla de Mayor (`Majority rule`), Dos o más (`2 or more`) y Prioridad (`Priority list`). Visualización obtenida con Python Bokeh.

En la **Tabla 7.1.1.** se encuentran cuatro borradores uno por cada criterio de fusión aplicado, solo para tres organismos modelo con los cuales se cuenta con un GSMM previamente publicado, estos presentados a manera comparativa, para ejemplificar los resultados obtenidos al construir las 116 redes metabólicas.

| Cepa /Ref.                               | Criterio | RXS  | MET  | GENES | RXN GF | MEMOTE |
|--|----------|------|------|-------|--------|--------|
| <b><i>Mtu</i></b>                        | Majority | 1489 | 1613 | 1205  | 64     | 75%    |
|  | 2OrMore  | 1063 | 1161 | 637   | 67     | 93%    |
|  | Priority | 1487 | 1583 | 1241  | 60     | 92%    |
|  | Union    | 2349 | 2301 | 1351  | 43     | 74%    |
| <b>Kavvas et al. 2018</b>                | Manual   | 1228 | 998  | 1011  |        | 73%    |
| <b><i>Sco</i></b>                        | Majority | 1573 | 1834 | 1667  | 76     | 75%    |
|  | 2OrMore  | 1633 | 1841 | 1546  | 75     | 75%    |
|  | Priority | 1868 | 1990 | 1860  | 59     | 92%    |
|  | Union    | 2525 | 2542 | 1928  | 47     | 73%    |
| <b>Amara et al. 2018</b>                 | Manual   | 1908 | 1472 | 1259  |        | 26%    |
| <b><i>Cgl</i></b>                        | Majority | 1335 | 1495 | 924   | 61     | 75%    |
|  | 2OrMore  | 945  | 989  | 523   | 72     | 92%    |
|  | Priority | 1248 | 1418 | 964   | 67     | 75%    |
|  | Union    | 1543 | 1675 | 997   | 76     | 73%    |
| <b>Feierabend &amp; Renz et al. 2021</b> | Manual   | 1539 | 1044 | 805   |        | 87%    |

**Tabla 7.1.1. Borradores de GSMM obtenidos para los cuatro criterios.** Regla del mayor (Majority), dos o más (2OrMore), prioridad (Priority) y Union, con su respectivo número de reacciones (RXN), metabolitos (MET), GENES, reacciones de llenado de brechas “Gapfill” (RXN GF) y la puntuación en porcentaje de la validación del control de calidad MEMOTE, solamente para los organismos modelo *Mtu*, *Sco* y *Cgl*, para tener el comparativo de los reportados previamente. En verde los GSMM reportados y utilizados en este trabajo de tesis, con sus respectivos datos. En azul el borrador de GSMM del criterio con mayor puntaje general obtenido en MEMOTE.

La tabla muestra que en el organismo ***Mtu*** se obtuvo el mayor puntaje/porcentaje general en el control de calidad MEMOTE para el criterio “dos o más” con 93 %, mostrando la menor cantidad de metabolitos, reacciones y genes, pero el mayor número de reacciones “gapfill” (i.e. se necesitó mayor cantidad de reacciones introducidas para vincular las reacciones no conectadas en las rutas metabólicas) en comparación con los demás criterios de fusión (borradores), la menor cantidad de los componentes proporcionó al borrador del modelo de mayor consistencia (i.e. qué tan bien está conectada la red), mientras más pequeña pueda ser la red, mayor sencillez y conectividad; sin embargo al momento de observar el GSMM manual de Kavvas et al. 2018 se observa que supera en cuanto a

reacciones y genes al del “*mejor*” borrador, además que el puntaje al someter el GSMM en MEMOTE es baja (73 %) en comparación con el borrador.

Para el caso de *Sco* la mejor puntuación general del control de calidad MEMOTE lo obtuvo el borrador del criterio de fusión “prioridad” (92 %), aquí la cantidad de los componentes no fue la menor, pero tampoco la mayor, al compararlo con el GSMM manual de Amara et al. 2018 (MEMOTE 26 %) el número de reacciones es hasta cierto punto similar pero los metabolitos y genes son diferentes.

En el caso de *Cgl* la mejor puntuación general del control de calidad MEMOTE lo obtuvo el borrador del criterio de fusión “dos o más” (92 %), mostrando la menor cantidad de metabolitos, reacciones y genes en comparación con los demás borradores, la menor cantidad de los componentes proporciono al borrador del modelo mayor consistencia; al compararlo con el GSMM con curación manual de Feierabend & Renz et al. 2021, el número de reacciones y genes es muy disímil, sin embargo, la puntuación MEMOTE de este GSMM de 87 % es buena.

Una cuestión importante que mencionar acerca de los datos en la **Tabla 7.1.1.** es que, los borradores de los diferentes criterios de fusión son solo eso “borradores”, son redes metabólicas las cuales falta curar manualmente, por lo que una comparación con los GSMM curados manualmente y reportados previamente, no tiene gran relevancia. Sin embargo, MEMOTE puede otorgar una diferenciación entre los cuatro borradores de cada organismo, que puede ser útil para dilucidar a partir de cuál de ellos (mayor porcentaje) se puede partir para llevar a cabo la extenuante y detallista curación manual. (*para los resultados de los borradores de GSMM con los mejores puntajes de MEMOTE para las 29 actinobacterias, ver Tabla I.a. Anexo I*).

Los borradores de GSMM requieren un refinamiento manual, el cual es necesario dado que suele haber enzimas que están asociadas con múltiples reacciones, con diferentes sustratos y/o cofactores. La información sobre la utilización de sustratos y cofactores se puede obtener de estudios bioquímicos específicos de organismos y también se pueden incluir en bases de datos específicas de organismos (e.g. Ecocyc o Actinobase [Keseler et al. 2021; Feeney et al. 2022]), se puede utilizar la

literatura disponible, experimentos de crecimiento, fenotipado, estructuras de proteínas, genómica comparativa, datos ómicos, datos exometabólicos, *Knock out* de genes, entre otros para hallar dicha información. Esta parte del proceso de curación puede llevar mucho tiempo y ser laboriosa, ya que puede ser difícil encontrar la información necesaria. A menudo, esto requiere una búsqueda bibliográfica intensiva. Es importante prestar mucha atención ya que la inclusión falsa de sustratos o cofactores puede cambiar mucho el comportamiento *in silico* (i.e. el potencial predictivo) de la reconstrucción. Los datos concernientes a las reacciones enzimáticas, los cuales se deben revisar con detenimiento y detalle son: *i)* Función del gen, *ii)* Función de la proteína, *iii)* Mecanismo de reacción, *iv)* Especificidad de cofactor/sustrato, y *v)* Direccionalidad de la reacción [Thiele & Palsson, 2010].

## 7.2. Comparación de análisis de sensibilidad de flujos relativos en tres GSMM

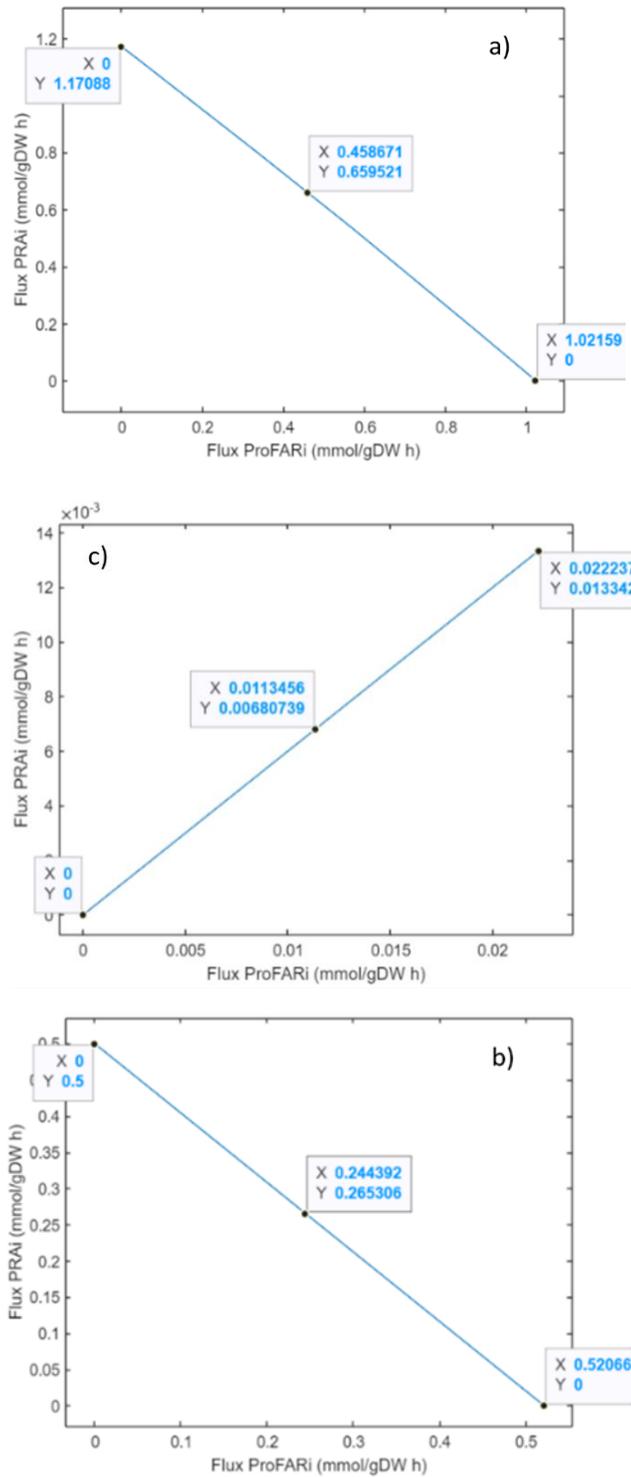
Después de aplicar las restricciones con datos experimentales (ver **Anexo IV**) al GSMM iEK1011 del organismo *Mtu* H37Rv, iCGB21FR del organismo *Cgl* ATCC 13032, y iAA1259 del organismo *Sco* A3(2). Se llevaron a cabo simulaciones en COBRA Toolbox 3.0 para evaluar la repercusión de la promiscuidad enzimática de PriA bonafide presente en *Mtu* y *Sco*; y su variante SubHisA1 presente en *Cgl*.

En el gráfico [Fig. 7.2.1.] concerniente al análisis de sensibilidad de flujos relativos, robustez, se observa cómo cambia el flujo de una reacción objetivo de interés (en este caso PRA isomerasa) a medida que el flujo a través de otra reacción específica de interés varía en magnitud (para tal caso ProFAR isomerasa). Se muestran los gráficos para cada uno de los tres GSMM con los que se trabajaron. El rango de valores del flujo de las reacciones varía dentro de los límites mínimo y máximo (como se aprecia en la Fig. 7.2.1) para la reacción de interés y, la función objetivo se maximiza usando FBA de acuerdo con cada valor de flujo establecido de la reacción de interés, aunque ProFAR isomerasa y PRA isomerasa pueden estar anotadas funcionalmente bajo el mismo gen en los genomas, se trata de dos reacciones diferentes en las matrices estequiométricas de los GSMM. Lo anterior explica las observaciones realizadas al evaluar con este tipo de análisis de sensibilidad de flujos relativos.

Puede existir el caso en contra parte, que una reacción pertenezca a dos enzimas diferentes con dos genes diferentes que las codifican como es el caso de PykF y PykA (piruvato quinasa II y I sc.) en el organismo *E.coli*; similarmente la duplicación de genes en el genoma de *Sco*, de dos enzimas piruvato quinasa que surgieron por un antiguo evento de duplicación, donde cada copia desarrollo diferentes cinéticas enzimáticas, Pyk1 tiene una  $k_{cat}$  20 veces mayor que Pyk2 (4,703  $s^{-1}$  en comparación con 215  $s^{-1}$  sc.) y, sin embargo, ambos genes (*pyk1* y *pyk2*) están expresados constitutivamente [Schniete et al. 2018]. Lo anterior sugiere que las familias de genes en expansión pueden ayudar a mantener la funcionalidad celular durante la perturbación del metabolismo, como la limitación de nutrientes y/o la producción de metabolitos especializados, proporcionando robustez a las vías en la que están involucradas dichas enzimas.

En la **Fig. 7.2.1.a.** se observa que al variar los flujos de carbono en magnitud de las dos reacciones donde está involucrada la enzima PriA, los flujos reflejan un gráfico de comportamiento inversamente proporcional (i.e. cuando asciende uno descende el otro y viceversa, por consecuencia las reacciones son dependientes una de la otra); ocurre de manera similar en el grafico **Fig. 7.2.1.b.** Estos dos gráficos corresponden a los organismos *Sco* y *Mtu* sc. donde se ha reportado con anterioridad que contienen en sus genomas al gen de la enzima PriA promiscua bisustrato, el comportamiento de este par de gráficos refleja la promiscuidad, ya que es la misma cantidad de enzima que está siendo utilizada por dos sustratos diferentes y no puede estar ocupada por ambos sustratos al mismo tiempo por lo que si asciende la ocupación de la enzima (flujo en la reacción) por alguno de los dos sustratos necesariamente descenderá el flujo de la otra reacción.

Para el caso del grafico **7.2.1.c.** del organismo *Cgl* que contiene en su genoma a la subespecializada de PriA denominada SubHisA1, donde la promiscuidad dual esta subfuncionalizada solo para la reacción ProFAR isomerasa, ya que contiene un gen *trpF* y por consiguiente una enzima PRO isomerasa fuera de la posible promiscua PriA; el comportamiento de los flujos es creciente lineal, esto se traduce a, que se trata de dos enzimas diferentes para cada reacción y la saturación de la enzima no depende una de la otra, este comportamiento tiene un máximo de flujo debido a que la



**Figura 7.2.1. Análisis de robustez para las reacciones ProFAR isomerasa y PRA isomerasa.** Se muestra para los tres GSMM: **a.** *Sco A3(2)*; **b.** *Mtu H37Rv*; **c.** *Cgl ATCC 13032*, exponiendo puntos máximos y mínimos de los posibles flujos y un punto aleatorio intermedio. Los flujos se expresan en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . Gráficos obtenidos en MATLAB® R2022b COBRA Toolbox [Heirendt et al. 2019].

biosíntesis de His y Trp no es infinita; la biosíntesis de los dos L-aminoácidos forman parte de la biomasa y su síntesis con dicho fin ocurre durante la fase exponencial primordialmente.

### 7.3. Fluxómica comparativa de L-aminoácidos

Se calculó la distribución de los flujos metabólicos con FBA para los tres GSMM seleccionados y restringidos. Se compararon los flujos intracelulares para dieciséis aminoácidos en los tres GSMM, los tres metabolismos son inherentemente diferentes, por lo que las reacciones involucradas en la síntesis de L-aminoácidos no suelen ser las mismas, lo anterior imposibilita la homogenización comparativa de los flujos de reacciones diferentes entre los tres metabolismos, por lo que solo se consideraron flujos de las mismas reacciones encontradas en los tres GSMM involucrados en la biosíntesis de **dieciséis** L-aminoácidos comparables.

En la **Tabla. 7.3.1.** se observan los datos recopilados para la comparación de las reacciones correspondientes a dieciséis L-aminoácidos. Las reacciones están homogenizadas i.e. se trata de la misma enzima en los tres GSMM seleccionados y restringidos, con la finalidad de comparar los flujos de las enzimas que biosintetizan dieciséis L-aminoácidos. Los ID's de reacción, nombres de enzimas y ecuaciones estequiométricas están anotados acorde con la nomenclatura de BIGG Models [King et al. 2016]; se muestra el nombre del aminoácido que sintetiza; así como los valores numéricos de los flujos para cada reacción en los tres modelos (*Sco*, *Mtu* y *Cgl*).

Para el caso de síntesis de L-Prolina la única reacción encontrada para su síntesis en *Cgl* fue P5CRx que utiliza la coenzima NADPH reducida, a diferencia de P5CR de *Sco* y *Mtu* que usa la forma no fosfatada de la coenzima NADH reducida, sin embargo, se trata de la misma enzima codificada por el mismo gen *proC*. Además, se ha considerado el aminoácido no esencial, proteinogénico, no quiral (i.e. sin enantiómeros) glicina.

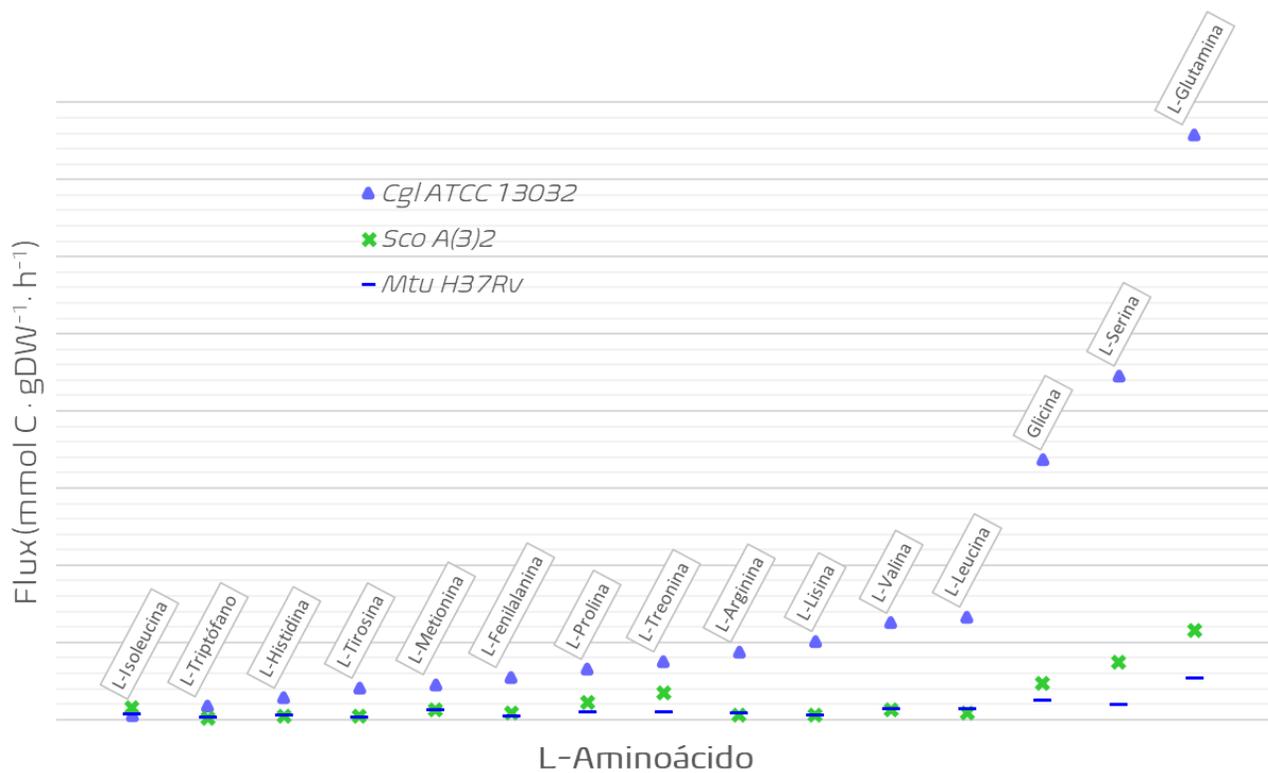
| ID's reacción   | Nombre de enzima                         | L-Aminoácido sintetizado | Estequiometría de la reacción  | Sco    | Cgl    | Mtu    |
|-----------------|--|--------------------------|--|--------|--------|--------|
| TRPS1           | Triptófano sintasa                       | L-Triptófano             | $3ig3p\_c + ser\_L\_c \rightleftharpoons g3p\_c + h2o\_c + trp\_L\_c$                  | 0.0020 | 0.0168 | 0.0005 |
| HISTD           | Histidinol deshidrogenasa                | L-Histidina              | $h2o\_c + histd\_c + 2.0 nad\_c \rightleftharpoons 3.0 h\_c + his\_L\_c + 2.0 nadh\_c$ | 0.0048 | 0.0280 | 0.0019 |
| TYRTA           | Tirosina transaminasa                    | L-Tirosina               | $34hpp\_c + glu\_L\_c \rightleftharpoons akc\_c + tyr\_L\_c$                           | 0.0042 | 0.0408 | 0.0008 |
| DAPDC           | Diaminopimelato descarboxilasa           | L-Lisina                 | $26dap\_M\_c + h\_c \rightleftharpoons co2\_c + lys\_L\_c$                             | 0.0054 | 0.1015 | 0.0020 |
| ARGSL           | Argininosuccinato liasa                  | L-Arginina               | $argsuc\_c \rightleftharpoons arg\_L\_c + fum\_c$                                      | 0.0061 | 0.0875 | 0.0031 |
| PHETA1          | Fenilalanina transaminasa                | L-Fenilalanina           | $glu\_L\_c + phpyr\_c \rightleftharpoons akc\_c + phe\_L\_c$                           | 0.0081 | 0.0548 | 0.0012 |
| LEUTA           | Leucina transaminasa                     | L-Leucina                | $4mop\_c + glu\_L\_c \rightleftharpoons akc\_c + leu\_L\_c$                            | 0.0084 | 0.1332 | 0.0052 |
| METS            | Metionina sintasa                        | L-Metionina              | $5mthf\_c + hcys\_L\_c \rightleftharpoons h\_c + met\_L\_c + thf\_c$                   | 0.0122 | 0.0455 | 0.0045 |
| VALTA           | Valina transaminasa                      | L-Valina                 | $3mob\_c + glu\_L\_c \rightleftharpoons akc\_c + val\_L\_c$                            | 0.0133 | 0.1251 | 1.0806 |
| ILETA           | Isoleucina transaminasa                  | L-Isoleucina             | $3mop\_c + glu\_L\_c \rightleftharpoons akc\_c + ile\_L\_c$                            | 0.0149 | 0.0053 | 0.0859 |
| P5CR (Sco, Mtu) | Pirrolina-5-carboxilato reductasa        | L-Prolina                | $1pyr5c\_c + 2.0 h\_c + nadph\_c \rightleftharpoons nadp\_c + pro\_L\_c$               | 0.0225 | 0.0654 | 0.0035 |
| P5CRx (Cgl)     | Pirrolina-5-carboxilato reductasa (nadh) | L-Prolina                | $1pyr5c\_c + 2.0 h\_c + nadh\_c \rightleftharpoons nad\_c + pro\_L\_c$                 |        |        |        |
| THRS            | Treonina sintasa                         | L-Treonina               | $h2o\_c + phom\_c \rightleftharpoons pi\_c + thr\_L\_c$                                | 0.0341 | 0.0750 | 0.0035 |
| GHMT2r          | Glicina hidroximetiltransferasa          | Glicina                  | $ser\_L\_c + thf\_c \rightleftharpoons gly\_c + h2o\_c + mlthf\_c$                     | 0.0472 | 0.3370 | 0.0232 |
| PSP_L           | Fosfoserina fosfatasa                    | L-Serina                 | $h2o\_c + pser\_L\_c \rightleftharpoons pi\_c + ser\_L\_c$                             | 0.0740 | 0.4456 | 0.0307 |
| GLNS            | Glutamina sintetasa                      | L-Glutamina              | $atp\_c + glu\_L\_c + nh4\_c \rightleftharpoons adp\_c + gln\_L\_c + h\_c + pi\_c$     | 0.1156 | 0.7583 | 0.0528 |
| ASPTA           | Aspartato transaminasa                   | L-Aspartato              | $glu\_L\_c + oaa\_c \rightleftharpoons akc\_c + asp\_L\_c$                             | 0.1319 | 0.8820 | 0.9599 |

**Tabla. 7.3.1. Reacciones y flujos de fluxómica comparativa.** Señalados en azul y verde, las enzimas que sintetizan Trp e His sc. Las enzimas están ordenadas con referencia a la columna de flujos de Sco de menor a mayor. La nomenclatura de la columna de estequiometría de la reacción acorde con BIGG Models es la siguiente: **3ig3p\_c**: C<sup>1</sup>-(3-indolil)-glicerol 3-fosfato; **ser\_L\_c**: L-Serina; **g3p\_c**: Gliceraldehído 3-fosfato; **h2o\_c**: Agua; **trp\_L\_c**: L-Triptófano; **histd\_c**: L-Histidinol; **nad\_c**: Nicotin adenin dinucleótido; **h\_c**: Protón; **his\_L\_c**; **nadh\_c**: Nicotin adenin dinucleótido reducido; **34hpp\_c**: 3-(4-hidroxifenil)piruvato; **glu\_L\_c**: L-Glutamato; **akc\_c**: 2-oxoglutarato; **tyr\_L\_c**: L-tirosina; **26dap\_M\_c**: Meso-2,6-diaminoheptanodioato; **co2\_c**: Dióxido de carbono; **lys\_L\_c**: L-lisina; **argsuc\_c**: N(omega)-(L-Arginino)succinato; **arg\_L\_c**: L-Arginina; **fum\_c**: Fumarato; **phpyr\_c**: Fenilpiruvato; **phe\_L\_c**: L-Fenilalanina; **4mop\_c**: 4-metil-2-oxopentanoato; **leu\_L\_c**: L-Leucina; **5mthf\_c**: 5-metil-tetrahydrofolato; **hcys\_L\_c**: L-homocisteína; **met\_L\_c**: L-Metionina; **thf\_c**: 5,6,7,8-tetrahydrofolato; **3mob\_c**: 3-Metil-2-oxobutanoato; **val\_L\_c**: L-Valina; **3mop\_c**: (S)-3-metil-2-oxopentanoato; **ile\_L\_c**: L-Isoleucina; **1pyr5c\_c**: 1-pirrolina-5-carboxilato; **nadph\_c**: Nicotin adenin dinucleótido fosfato reducido; **nadp\_c**: nicotin adenin dinucleótido fosfato; **pro\_L\_c**: L-Prolina; **phom\_c**: O-fosfo-L-homoserina; **pi\_c**: Fosfato inorgánico; **thr\_L\_c**: L-Treonina; **gly\_c**: Glicina; **mlthf\_c**: 5,10-metil-tetrahydrofolato; **pser\_L\_c**: O-fosfo-L-serina; **atp\_c**: Adenosín trifosfato; **nh4\_c**: Amonio; **adp\_c**: Adenosín difosfato; **gln\_L\_c**: L-Glutamina; **oaa\_c**: Oxaloacetato; **asp\_L\_c**: L-Aspartato; la "c" terminal en la abreviatura de cada metabolito indica que se trata de un metabolito celular (i.e. intracelular).

En la **Tabla. 7.3.1.** se observan las diferencias y similitudes entre los flujos de carbono en la biosíntesis de L-aminoácidos, en las tres actinobacterias en cuestión.

Las simulaciones con el modelo de **Cgl** mostró valores de flujos altos para los dieciséis L-aminoácidos en comparación con **Sco** y **Mtu**, lo anterior es coherente dado que **Cgl** es el caballo de batalla en la biosíntesis de L-aminoácidos a nivel industrial y es de esperar el aumento en dichos flujos. Las simulaciones con el GSMM de **Cgl** reportó los mayores flujos de carbono, en orden de mayor a menor, para L-glutamina, L-serina, y glicina; los menores flujos de menor a mayor para L-isoleucina, Trp e His; **Cgl** es un sobreproductor industrial especialmente de L-glutamato (de ahí el nombre de la bacteria), el cual es encargado de proporcionar el sabor "umami" muy popular en los aditivos

alimentarios, el metabolismo de *Cgl* biosintetiza este L-aminoácido naturalmente en múltiples vías (en reacciones no comparables con los otros dos GSMM), en algunas de ellas no es el metabolito principal de síntesis, por lo que el interés en las modificaciones del microorganismo ya sea a nivel de diseño o proceso son numerosas, los descubrimientos acerca de su fisiología sobreproductora de L-aminoácidos continua develándose, a pesar de la serie de estudios con los que ya se cuenta [Ganguly, 2023] (Para mayor desglose de los flujos de este GSMM ver **Anexo V inciso b**).



**Figura. 7.3.1. Flujos calculados con FBA para la síntesis de dieciséis aminoácidos intracelulares en tres GSMM.** Los flujos se expresan en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . El orden de los aminoácidos está colocado de izquierda a derecha de menor a mayor sc. conforme a los flujos de *Cgl* ATCC 13032; para los flujos de *Sco* A (3)2 y *Mtu* H37Rv se encuentran en la misma columna de la etiqueta marcada, sin estar en el mismo orden. Las diferentes figuras de los marcadores permiten diferenciar entre los flujos en la misma columna.

Al modelo de *Mtu* se asignó la restricción de consumo de L-alanina como única fuente de nitrógeno (ver sección 6.3 de la estrategia experimental) para observar la biosíntesis de los demás aminoácidos, durante la simulación de FBA, el modelo reportó los tres mayores flujos de mayor a menor para L-glutamina, glicina y L-serina; los menores flujos de menor a mayor que reportó la

simulación fueron para los aminoácidos Trp, L-tirosina, L-fenilalanina e His, se observa similarmente como en *Cgl* que Trp e His se encuentran entre los flujos más pequeños (*Para mayor desglose de los flujos de este GSMM ver Anexo V inciso c*).

Para el caso de las simulaciones con el GSMM de *Sco*, reportó los menores flujos de menor a mayor para los L-aminoácidos Trp, L-Tirosina e His; y los mayores flujos de mayor a menor para L-glutamina, L-serina y glicina (*Para mayor desglose de los flujos de este GSMM ver Anexo V inciso a*).

## 7.4. Mapas de flujos metabólicos

El mapa de flujos metabólicos en el cual se esquematizan vías metabólicas, donde el valor del flujo de las reacciones involucradas, así como los metabolitos involucrados de cada reacción, se muestra en la **Fig. 7.4.1** para el modelo iAA1259 de *Sco* A(3)2; se observa como el flujo de la vía de biosíntesis de Trp que se muestra en color azul-morado tiene una relación de **1 : 2.44** con respecto al flujo de la vía de biosíntesis de His que se muestra en color verde; se observan las diferencias en peso de las líneas que representan las reacciones, mostrando mayor flujo (peso) en la vía de biosíntesis de His de  $0.00481 \text{ mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$  con respecto al de Trp con un valor de  $0.00197 \text{ mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . Se aprecia la convergencia de ambas rutas metabólicas en las reacciones PRM1CI (His) y PRA1I (Trp) representando la promiscuidad enzimática. Se distingue cada uno de los metabolitos consumidos y producidos en cada reacción.

**Nota:** para explicar mejor los mapas metabólicos de ahora en adelante, en cada enzima o metabolito comentado en esta sección, apunto la abreviatura de los mapas metabólicos entre paréntesis inmediatamente después.

Se visibiliza con nombres en negritas las vías principales del metabolismo central como glucolisis, pentosas fosfato, el desvío del flujo hacia el ciclo de los ácidos tricarboxílicos con la enzima piruvato quinasa (PYK) (sin mostrar el ciclo completo, solo se esquematiza); la ruta pre-corismática

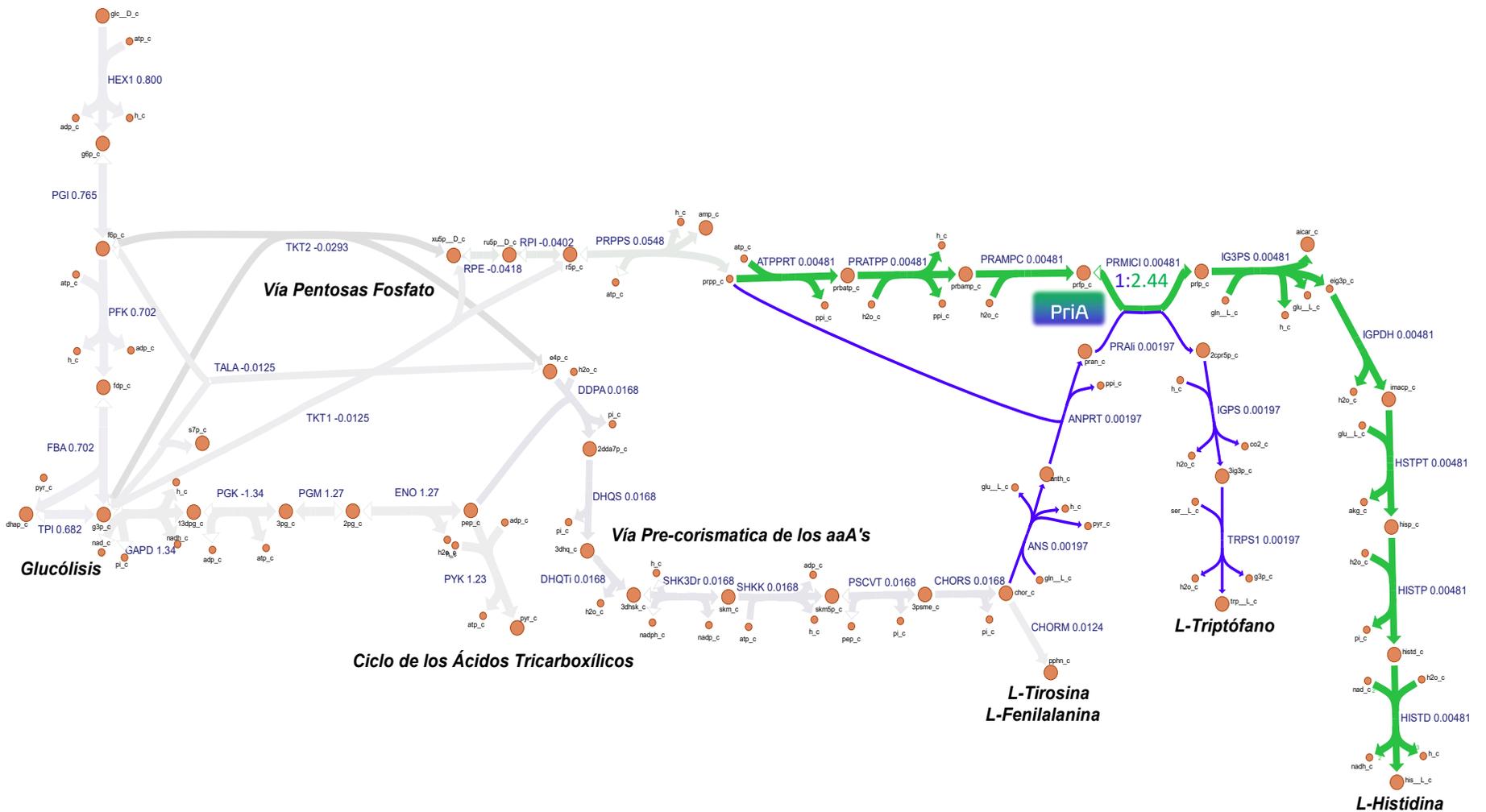
común para los tres aminoácidos aromáticos la cual comienza con la enzima 3-desoxi-D-arabinoheptulosonato 7-fosfato sintetasa (DDPA) que conjunta la eritrosa-4-fosfato (e4p\_c) metabolito proveniente de la vía de las pentosas fosfato y el fosfoenolpiruvato (pep\_c) metabolito proveniente de la glucólisis, esta ruta común de los aminoácidos aromáticos concluye con la formación de corismato (chor\_c) por la enzima corismato sintasa (CHORS), a partir del chor\_c se bifurca el flujo hacia la biosíntesis de L-tirosina y L-fenilalanina por la enzima corismato mutasa (CHORM) que a partir de dos pasos enzimáticos separados se sintetizan los dos L-aminoácidos (vía no mostrada en el mapa de flujo, solo esquematizada), el desvío del flujo hacia la biosíntesis de Trp ocurre en cinco pasos enzimáticos comenzando con la enzima antranilato sintasa (ANS) a partir de chor\_c y culminando con la enzima triptófano sintasa (TRPS1), en un paso intermedio de esta vía se muestra la convergencia de la promiscuidad enzimática de la enzima PriA con las reacciones fosforribosilantranilato isomerasa (PRAli) que se une al sustrato N-(5-fosfo-D-ribosil)antranilato (pran\_c) y produce 1-(2-carboxi-fenil-amino)-1-desoxi-D-ribulosa 5-fosfato (2cpr5p\_c). PriA una misma enzima codificada por un gen utilizada en dos reacciones enzimáticas diferentes de dos vías de biosíntesis de dos L-aminoácidos diferentes con flujos pequeños.

La ruta de biosíntesis de His (verde) comienza con el metabolito 5-fosfo-alfa-D-ribosa 1-difosfato o pirofosfato de fosforribosil (prpp\_c) proveniente de la vía de pentosas fosfato y ATP (atp\_c) con la enzima ATP fosforribosiltransferasa (ATPPRT) y a partir de ocho pasos enzimáticos más culmina con la biosíntesis de His con la enzima histidinol deshidrogenasa (HISTD) con una molécula de agua (h2o\_c) reduciendo una molécula de NAD (nad\_c) a NADH (nadh\_c); en un paso intermedio de esta vía se encuentra la promiscuidad enzimática de la enzima PriA con la reacción 1-(5-fosforribosil)-5-[(5-fosforribosil-amino)metilidenamino]imidazol-4-carboxamida isomerasa (PRMICI) que se une al sustrato 1-(5-fosforribosil)-5-[(5-fosforribosilamino) metilidenamino] imidazol-4-carboxamida (prfp\_c) y produce 5-[(5-fosfo-1-desoxiribulosa-1-ilamino)metilidenamino]-1-(5-fosforribosil)imidazol-4-carboxamida (prlp\_c).

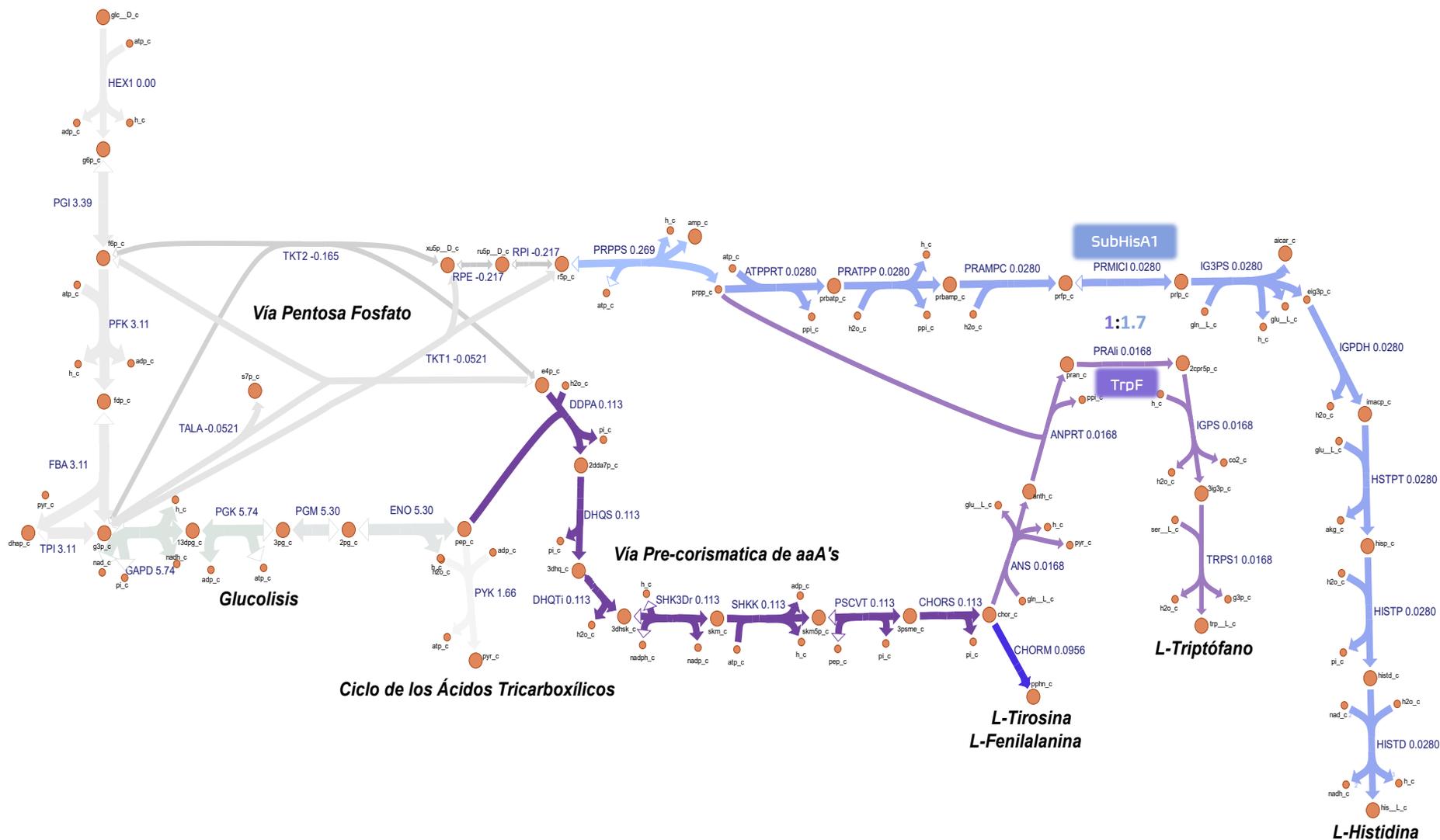
El mapa de distribución de flujos metabólicos del organismo *Mtu* H37Rv se encuentra en el **Anexo VI**, no se creyó conveniente añadirlo a esta sección ya que su comportamiento es muy similar

a *Sco* y sigue tendencias muy parecidas a pesar de sus metabolismos disimiles, a continuación, se muestra a modo comparativo el mapa de flujo de *Cgl*.

En la **Fig. 7.4.2.** se aprecia la distribución de flujos para biosíntesis de His y Trp en las simulaciones con el modelo iCGB21FR del organismo *Cgl* ATCC 13032, los flujos y el mapa son diferentes al de *Sco* con la diferencia mostrada con respecto a la vía pre-corismatica común para aminoácidos aromáticos en color morado, donde se muestra que a partir de chor\_c se bifurca la vía hacia la síntesis conjunta de L-tirosina y L-fenilalanina (color azul-morado) y hacia la biosíntesis de Trp (color lila); en cuanto a la ruta de síntesis de His y Trp en cuestión de numero de pasos enzimáticos son los mismos que para el mapa de flujo de la **Fig 7.4.1.** con los mismos metabolitos; en las reacciones PRMICI y PRAli se observa la separación de las vías, donde no comparten una enzima promiscua como es el caso para los organismos *Sco* y *Mtu*, lo anterior ocasionado por que *Cgl* presenta la THG descrita con anterioridad y por consecuencia contiene en su genoma una enzima derivada del gen *trpF* (etiqueta lila) reacción PRMICI, además de la enzima PriA subfuncionalizada a HisA (etiqueta azul claro) en la ruta de síntesis de His reacción PRAli. Se observan los flujos de la vía de biosíntesis de Trp de  $0.0168 \text{ mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$  e His de  $0.0280 \text{ mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$  con una relación de flujos de **1 Trp : 1.7 His**, ambas vías diferenciadas por color y peso, con mayor flujo en la vía de His en comparación con Trp, observado también en las simulaciones de los otros dos GSMM de actinobacteria.



**Figura. 7.4.1. Mapa de distribución de flujos metabólicos para *Sco A (3)2*.** Los nombres de cada reacción y metabolito están abreviados de acorde con la plataforma BIGG Models [King et al. 2016]. Al lado de cada reacción se encuentra el valor del flujo en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . Se anota la zona correspondiente a cada vía metabólica central. Las vías de síntesis de His (verde) y Trp (azul-morado) se diferencian por el peso de la línea y el color. Se observan las reacciones de promiscuidad resaltadas con la etiqueta bicolor “PriA”, mostrando la relación entre el valor de los flujos de ambas vías identificados por el color. El mapa se elaboró con la herramienta Escher [King et al. 2015].



**Figura. 7.4.2. Mapa de distribución de flujos metabólicos para *Cgl* ATCC 13032.** Los nombres de cada reacción y metabolito están abreviados acorde con la plataforma BIGG Models [King et al. 2016]. Al lado de cada reacción se encuentra el valor del flujo en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . Se anota la zona correspondiente a cada vía metabólica central. Las vías de síntesis de His (azul claro) y Trp (lila) se diferencian por el peso de la línea y el color. Se observan las reacciones resaltadas con la etiqueta de color azul claro “SubHisA1” y color lila “TrpF”, mostrando la relación numérica de los flujos de ambas vías identificados por el color. El mapa se elaboró con la herramienta Escher [King et al. 2015].

## 7.5. Contextos genómicos comparativos

Se compararon los contextos genómicos adyacentes a la secuencia del gen *priA* (marcada en rojo en la **Fig. 7.5.1.**) con el algoritmo CORASON [Navarro-Muñoz et al. 2020], en los genomas de los tres GSMM (*Sco*, *Mtu* y *Cgl*) evaluados en las secciones anteriores; dicho algoritmo fue creado originalmente para la detección de BGC's compartidos filogenéticamente entre diferentes genomas comparando la adyacencia genómica a partir de una secuencia semilla.

En el caso particular de la presente tesis ha sido común utilizar herramientas computacionales en la evaluación de características genómicas y metabólicas para las cuales no se ha generalizado su uso, en esta tesis se utilizaron exploratoriamente en el fenómeno de promiscuidad enzimática. Siguiendo esta vertiente se utilizó el algoritmo CORASON para exponer la adyacencia genómica de los genes contiguos al gen *priA* (semilla), los cuales forman parte de los operones para biosíntesis de Trp (morados-azules) e His (verdes).

En la **Fig. 7.5.1.** se observan los operones de biosíntesis de His y Trp, así como la direccionalidad de los genes. La secuencia del gen *priA* se encuentra entre los genes pertenecientes al operón de His, dado que *priA* es un homólogo del gen *hisA* perteneciente a este operón, el cual evolucionó a una bifuncionalidad, además la proximidad de *priA* al menos en *Mtu* y *Sco* al operón de Trp, puede explicar el pluriempleo metabólico bifuncional [Noda-García et al. 2013; Copley, 2020]. Ante la ausencia del gen *trpF* en el operón de Trp, la secuencia y enzima PriA es secuestrada por su función catalítica de isomerización hacia ambas reacciones y por la semejanza encontrada en ambos sustratos de la ribosafosfato.

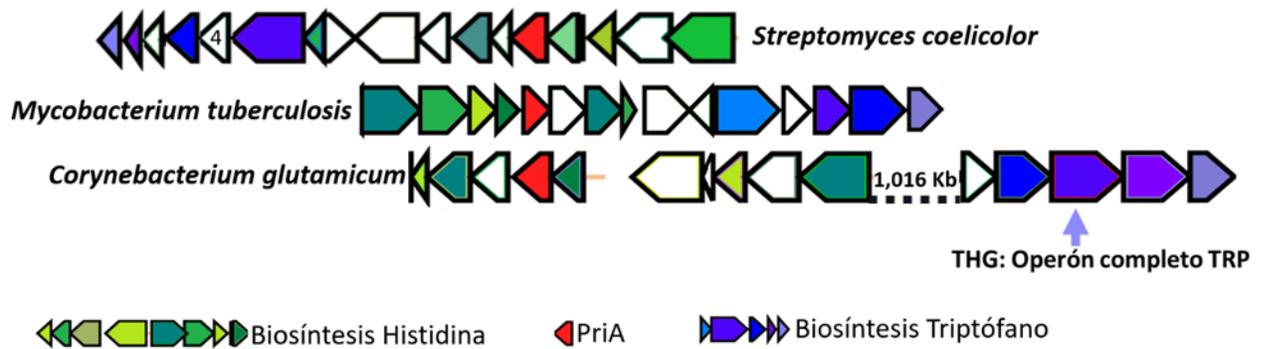


Figura. 7.5.1. Contextos genómicos de *Sco*, *Mtu* y *Cgl*. Los operones de Trp e His en *Sco* y *Mtu* se encuentran en la adyacencia genómica de *priA*. Para el caso de la adyacencia en *Cgl* se muestra la THG representada por la distancia entre los operones de His y Trp en Kb. Elaborado con algoritmo CORASON [Navarro-Muñoz et al. 2020].

Para el caso del organismo *Cgl* se observa el operón de His con la secuencia de *priA* que evolucionó a una subfuncionalización ocasionada porque el genoma de este organismo contiene el evento de THG a partir de las  $\gamma$ -proteobacterias y por consecuencia contiene una copia del gen *trpF* al adquirir el operón completo [Noda-García et al. 2013]; se ha observado que cuando ocurren dicho tipo de eventos, los genes se localizan distantes [Glesner et al. 2020], en este caso los genes del operón de Trp se localizan a 1.016 Mpb de los genes del operón de His.

## 8. Discusión de Resultados

El resultado de la anotación funcional de “*machine learning*” NetGO 3.0 [Fig. 7.1.1.], permitió visualizar el alto número de pares únicos de gen/reacción ModelSEED (GENE\_MSXNS), en comparación con el resto de las anotaciones funcionales, este resultado sugiere que es probable que más de una reacción este siendo atribuida a un solo gen, si esto es cierto implicaría reacciones de enzimas promiscuas, hay que tomar en cuenta que NetGO 3.0 utiliza principalmente proteínas con anotaciones funcionales respaldadas experimentalmente [Rives et al. 2021] y, por lo tanto, ignora información valiosa *in silico* de una gran cantidad de proteínas que no cuentan con experimentación; por lo que dicho resultado atípico de GENE\_MSXNS se puede deber a falsos positivos primordialmente.

Numerosas bases de datos contienen una gran cantidad de reacciones duplicadas. Acorde con Mohammadi Peyhani & cols. (2022) observaron que la mayor proporción de reacciones únicas corresponde para KEGG (97 %), seguido de BKMS (95 %). Reactome y BiGG tuvieron el porcentaje más bajo de reacciones únicas (49 % y 50 %, sc.), lo que indica que muchas de las reacciones son duplicados estructurales. Dicha evaluación cuantitativa de la singularidad de la reacción expone la naturaleza heterogénea de las bases de datos bioquímicas, lo que sugiere que la cantidad de entradas proporcionadas por los anfitriones de la base de datos debe manejarse con cuidado al comparar bases de datos y destaca la importancia de la evaluación de la calidad [Mohammadi Peyhani et al. 2022]. Por lo tanto, es importante verificar si las reacciones recopiladas están o no balanceadas por elementos y asociadas con un número EC, para esta última característica se utilizó la herramienta DeepEC en la anotación funcional realizada.

Las redes metabólicas reconstruidas en este trabajo de tesis, obtuvieron puntuaciones altas durante la validación con MEMOTE [Lieven & Moritz et al. 2020; Sulheim et al. 2020], e.g. el GSMM reportado por Feierabend & Renz et al. (2021) de *Cgl* ATCC 13032 después de la curación manual obtuvo una puntuación general en MEMOTE de 87 %; el GSMM consenso de Sco (ya que utilizaron la información de los dos últimos GSMM publicados anteriormente) restringido con datos de proteómica

reportado por Sulheim & cols. (2020) alcanzo una puntuación general en MEMOTE de 77 %, que es superior a la lograda por cualquier modelo anterior de *Sco*.

A pesar de que las reconstrucciones elaboradas en este trabajo de tesis, solo es el primer paso en la reconstrucción de un GSMM, las redes metabólicas reconstruidas mostraron porcentajes generales entre 90.79 % - 93 % (donde una puntuación cercana a 100 % representa una excelente red metabólica y una puntuación cercana a 0 % representa una muy mala red metabólica), esto permite observar que dichas redes construidas con ModelSEED [Henry et al. 2010; Seaver et al. 2021] están conectadas i.e. son consistentes y cuentan con la información básica para metabolitos, reacciones, genes, términos SBO, presencia de GPR, una función de biomasa realista, energética del metabolismo y topología adecuada de la red. Por lo anterior se consideran buenos borradores de GSMM que ahora se encuentran disponibles para una curación manual posterior, es muy probable que los puntajes generales en MEMOTE disminuyan al someter a esta validación los GSMM una vez curados manualmente, debido a que los GSMM no incluyen en su totalidad las reacciones de todas las enzimas codificadas en los genomas de los organismos que se intentan modelar, es por esta razón que se hace uso de la herramienta “Gapfill” para llenar los espacios vacíos en las vías metabólicas que se crean. Cabe señalar que la calidad de “Gapfill” depende en gran medida de la calidad de la secuencia genómica y su anotación, y otro aspecto que suele ocurrir son los cortes del alineamiento de las secuencias para asignar función.

El análisis de sensibilidad de flujos relativos, robustez, permitió observar la ocupación de la misma cantidad de enzima PriA por dos sustratos diferentes al mismo tiempo, en los organismos *Sco* y *Mtu*, que contienen en sus genomas a dicha enzima.

En la reconstrucción de los GSMM con los que se trabajó en esta tesis (al menos *Sco* y *Mtu*) [Amara et al 2018; Kavvas et al. 2018] se encuentra registrado en las GPR, la promiscuidad de PriA, donde especifican la existencia de dos reacciones atribuidas a un solo gen, para estos casos el gen no está nombrado como *priA* sino como *hisA*, pero contiene dicha doble función.

Acorde con los análisis de sensibilidad realizados, si existiera una mayor formación de Trp disminuiría la formación de His y viceversa; sin embargo cuando se observaron los flujos de FBA tanto

en *Sco* como en *Mtu* y *Cgl* la relación de biosíntesis de Trp : His, el flujo siempre fue mayor para His que para Trp, a pesar de que la estructura de la enzima PriA (descrita en la sección 2.4.) tenga preferencia por el sustrato PRA (Trp), ya que le es más sencillo a la enzima unirse a PRA, pero debido a que maneja velocidades de formación de producto más rápidas en comparación a cuando se une a ProFAR (His), esto ocasionado por el movimiento lento de apertura y cierre de los *loops* catalíticos en presencia de ProFAR en comparación con los movimientos dinámicos y más rápidos de los *loops* catalíticos en presencia de PRA, el tiempo de residencia del sustrato PRA en la enzima PriA es menor al observado con el sustrato ProFAR [Romero-Rivera et al. 2022; Due et al. 2011].

La diferencia de flujos entre ambas rutas puede estar relacionada a su vez con la relajación en la regulación de His en comparación con Trp (que es uno de los aminoácidos más caros celularmente hablando, por la cantidad de metabolitos y monedas energéticas que requiere para su síntesis); si es cierto que la regulación de L-aminoácidos en actinobacteria es poca, o casi nula [Hodgson, 2000], en comparación por ejemplo con el organismo bacteriano modelo *E. coli* donde su regulación ocurre a múltiples niveles (transcripcional, traduccional, enzimático y en un contexto metabólico local y global) [Santos-Zavaleta et al. 2019; Keseler et al. 2021].

En bacterias Gram positivas y Gram negativas, el nivel intracelular de Trp libre y la disponibilidad de tRNATrp cargado son señales reguladoras que controlan la expresión de los genes biosintéticos de Trp [Gutiérrez-Preciado et al. 2007]. En particular, el análisis de los mecanismos utilizados para regular la expresión de los genes *trp* en bacterias Gram positivas revelaron la existencia de dos estrategias reguladoras principales; la primera, descrita en *Bacillus subtilis* y sus parientes más cercanos, implica un mecanismo de atenuación transcripcional mediado por el producto del gen *mtrB*, i.e. la proteína TRAP de unión a ARN activada por Trp [Yanofsky, 2007]; la segunda estrategia, restringida a los Firmicutes, se basa en el mecanismo T-box, esto permite que una secuencia líder de ARN reconozca y responda a tRNATrp sin carga, regulando así la expresión de los genes *trp* [Merino et al. 2008].

En el caso de la regulación de Trp en actinobacterias, e.g. en el actinomiceto filamentoso *Sco* A3(2) Gram positivo, los genes biosintéticos de Trp están organizados en operones policistrónicos (*trpCXBA* y *trpC2D2GE2*) y/o existen como genes individuales (*trpE3*, *trpE* y *priA*) [Palazzotto et al.

2015]; los genes *trp* tienen parálogos que se mapean en operones separados, los genes *trpC2D2GE2* se encuentran además dentro del grupo de genes responsables de la producción de CDA, un antibiótico lipopeptídico cíclico que contiene residuos L-Trp y D-Trp [Kempter et al 1997]. Cabe señalar que la expresión de los genes biosintéticos *trp* no está controlada negativamente por Trp en *Sco* y promueve la diferenciación morfológica y fisiológica de la acinobacteria [Palazzotto et al. 2015], que es diferente a lo que se observa en otras bacterias [Merino et al. 2008]. De hecho, la expresión de los genes *trp* está regulada en respuesta a los requerimientos generales de crecimiento (biomasa) de la célula [Hu et al. 1999], y su suplementación al medio mínimo puede estimular la transcripción de los genes *trp*, incluidos los presentes en el grupo de genes responsables de la biosíntesis de CDA. *trpM* (antes llamado *trpX*) del operon *trpCXBA* es un componente clave en la modulación de la ruta anabólica de Trp, es un péptido multiempleo (o “*moon-light*”) involucrado en la modulación de la disponibilidad de precursores de Trp, principalmente serina e indol [Palazzotto et al. 2016], además estimula el crecimiento y la producción del antibiótico actinorrodina, y disminuye la producción de CDA, que se obtienen a través de la modulación de varias vías metabólicas [Vassallo et al 2020].

En *Cgl*, el operón His está regulado a través del mecanismo de atenuación transcripcional “*T-box*” [Jung et al. 2010]. Los riboswitches “*T-box*” están ampliamente distribuidos en actinobacteria y firmicutes, donde controlan la expresión de numerosos genes que codifican varias aminoacil-tRNA sintetasas, transportadores y enzimas de biosíntesis de aminoácidos [Vitreschak et al. 2008]. Los elementos reguladores de la caja T se caracterizan por un conjunto de estructuras de ARN conservadas, una de las cuales contiene una secuencia específica, que es complementaria a la secuencia anticodón del ARNt [Kreuzer & Henkin, 2018], la unión de este par de elementos, dependiendo de su estado de carga, promueve la formación del terminador transcripcional o estabiliza una estructura de ARN alternativa (anti-terminador).

HisR un factor de transcripción recientemente descrito por Ashniev & cols. (2022) encontrado en 20 genomas de actinobacteria; controla tres loci de genes independientes: el operón His, el transportador HisXYZ\* e *hisR*. En contraste, el regulón HisR en dos especies actinobacterianas auxotróficas de His, *Collinsella intestinalis* y *Cryptobacterium curtum*, contiene solo los genes *hisXYZ\** e *hisR* [Ashniev et al. 2022].

Además, encontraron ortólogos de *hisR* en 22 genomas de actinobacteria de los órdenes de Coriobacteriales y Eggertellales, dichos ortólogos nunca se encontraron en grupos cromosómicos de genes de biosíntesis y transporte de His. Una comparación de la estructura terciaria del represor TrpR (i.e. el atenuador de Trp en *E.coli*), un mecanismo bien caracterizado de unión a ADN dependiente de Trp, vs con una estructura de HisR recientemente disponible de *Eubacterium eligens* (un firmicutes) reveló una similitud estructural y de secuencia entre sus sitios de unión al ADN carboxi-terminal y con diferencias en los sitios de unión amino-terminal [Ashniev et al. 2022].

La descripción dada en los últimos párrafos, intenta mostrar la complejidad generalizada en bacterias en la regulación biosintética de Trp, inclusive en actinobacteria; y por qué en las tres actinobacterias tocadas en esta tesis el flujo para biosíntesis de Trp es dos o tres veces menor en comparación con el flujo para biosíntesis de His, este último L-aminoácido no debe vencer la compleja regulación para su síntesis como en el caso de Trp, a pesar de que para la síntesis de His se ha reportado casos aislados de regulación por atenuación en actinobacterias como describieron Ashniev y cols. (2022).

La diferencia de flujos se aprecia perfectamente en los mapas de distribución de flujos metabólicos de la *sección 7.4.* y en la fluxómica comparativa de la *sección 7.3.*; ambos metabolismos de síntesis (His y Trp) están relacionados en la mayor parte de las actinobacterias; la retención del gen *trpM* en el genoma de *Sco* sugiere que este péptido podría estar evitando conflictos metabólicos relacionados con la ambigüedad del sustrato de la promiscuidad de PriA [Palazzotto et al. 2015] y; la similitud de estructura y secuencia de HisR y TrpR como mecanismos de atenuación en ambos operones [Ashniev et al. 2022], puede estar sugiriendo un posible origen evolutivo común, al menos del sistema regulatorio que parece estar en constante evolución.

Requisitos regulatorios reducidos y cambiantes dependientes del entorno celular (*Sco*) [Jin et al. 2023], además de flujos pequeños, parecen caracterizar a las enzimas generalistas [Nam et al. 2012], tal es el caso de la promiscuidad bisustrato de PriA. El análisis de fluxómica comparativa llevado a cabo en este trabajo, permitió explorar en este sentido, al observar que los flujos de las rutas de biosíntesis de Trp e His en tres actinobacterias (en especial Trp), fueron los más pequeños de los dieciséis aminoácidos evaluados.

Las actividades enzimáticas con función biológica bien definida están bajo selección positiva o purificadora porque contribuyen a la aptitud del organismo. Las actividades enzimáticas secundarias (metabolismo subterráneo) no contribuyen a la aptitud a menos que se produzcan cambios ambientales o genómicos, como mutaciones o amplificación de genes. Así también, los residuos que mejoran las actividades de una enzima transferida horizontalmente estarán sujetos a selección positiva si mejoran la aptitud del huésped, mientras que los residuos que solo afectan las actividades no utilizadas (secundarias) estarán sujetos a selección neutral [Glesner et al. 2020].

Se ha demostrado que el metabolismo subterráneo salvaguarda la biosíntesis de metabolitos celulares esenciales; e.g. propiedades promiscuas de enzimas de las rutas de biosíntesis de L-aminoácidos como L-Serina y L-Treonina (genes *serA*, *serC* y *thrB*) en *E. coli* y *Bacillus subtilis* son útiles en rutas subterráneas (i.e. reacciones secundarias de las enzimas que coadyuvan) para biosíntesis de metabolitos muy importantes e imprescindibles para las células, como es el caso de la vitamina B6 [Richs & Commichau, 2021]. Por lo que, naturalmente el metabolismo procariota hace uso del metabolismo subterráneo ante cambios ambientales y necesidad de síntesis de metabolitos importantes para su supervivencia [Guzmán et al. 2015]; la enzima PriA está involucrada en la biosíntesis de dos L-aminoácidos completamente necesarios para el crecimiento y desarrollo de los microorganismos aquí evaluados, los cuales han evolucionado en diferentes circunstancias, subespecializándose dependiendo de los nichos ecológicos y ambientes en los que crecen.

## 9. Conclusiones

Los MMEG pueden simular la promiscuidad de un enzima en un análisis de robustez, el gráfico proporcionado por este análisis refleja un comportamiento inversamente proporcional vs un comportamiento lineal cuando se trata de una subespecialización de la misma enzima. Sin embargo, esto solo puede ocurrir si en la anotación funcional está definida tal promiscuidad en las GPR, de lo contrario se debe modificar el modelo de tal manera que se incluya el uso de una misma enzima en dos o más reacciones diferentes presentes en el GSMM.

La distribución de flujos de carbono para los organismos que muestran un comportamiento PriA bonafide (*Sco* & *Mtu*), suelen ser mayores en la vía de biosíntesis de His que en la vía de Trp, sin embargo, ambas vías reflejan flujos mucho más pequeños en comparación con los flujos de los demás L-aminoácidos sintetizados por las actinobacterias aquí tocadas.

El comportamiento subespecializado de PriA hacia SubHisA1 en *Cgl* muestra un comportamiento de la distribución de flujos de la síntesis de los dos L-aminoácidos casi a la par, pero sin ser iguales, esto ocasionado por la THG a partir de las  $\gamma$ -proteobacterias completando el operón de Trp, modificando el sistema de regulación de *Cgl*.

A partir de los análisis que efectúe puedo clasificar las enzimas del metabolismo de actinobacterias en dos principales grupos desde la perspectiva de biología de sistemas:

**Enzimas Especialistas**, las cuales cumplen con ciertas cualidades, suelen manejar flujos de carbono del metabolismo central, así como manejar flujos altos (esto representa una presión selectiva, al resultar ser enzimas importantes y necesarias para el organismo, dichas enzimas están evolucionando muy lentamente), normalmente tienen fuerte regulación a nivel genético y alostérico, sus flujos cambian dependiendo del medio de cultivo en que estén creciendo, derivan frecuentemente de genes esenciales y manejan parámetros cinéticos muy eficientes.

**Enzimas generalistas o promiscuas**, suelen cumplir las cualidades de estar presentes en vías metabólicas accesorias no centrales (e.g. en biosíntesis de aminoácidos y cofactores), suelen manejar

distribuciones de flujo bajos, las enzimas evolucionan más rápidamente en comparación con las del metabolismo central, suelen tener poca o nula regulación alostérica, suelen manejar flujos constantes independientemente del medio de cultivo en que estén creciendo, suelen tener parámetros cinéticos muy poco eficientes. PriA y variantes entran en esta categoría de enzimas.

La explotación del metabolismo subterráneo en actinobacterias puede coadyuvar a obtener altos rendimientos en la biosíntesis de productos de alto valor agregado, sin necesidad de la introducción de genes heterólogos externos, disminuyendo de esta manera la utilización de GMO's en la industria biotecnológica, teniendo en cuenta la normatividad jurídica con respecto a este tipo de organismos alrededor del planeta, el conocimiento generado en este trabajo de tesis puede ser útil en apoyar esta idea.

# 10. Perspectivas

La promiscuidad enzimática y su repercusión a nivel sistémico en los metabolismos (metabolismo subterráneo), como he descrito a lo largo de este trabajo de tesis, tiene un alto potencial de explotación para amplias aplicaciones en el área de la biotecnología industrial, en las cuales vale la pena la I + D.

Los análisis y revisión que he realizado en este trabajo de tesis se pueden ampliar con uso de algoritmos de inteligencia artificial, biología cuántica, análisis de estructura enzimática, molecular docking, ingeniería de proteínas, evolución dirigida y adaptativa, quimioinformática y metagenómica, herramientas que existen hasta la fecha y continúan desarrollándose, para la detección del metabolismo subterráneo.

En particular la enzima promiscua PriA, tiene expansiones genómicas (i.e duplicaciones, triplicaciones, etc.) en el metabolismo especializado dentro del *phylum* actinobacteria, las cuales se pueden seguir detectando con análisis filogenómicos (tipo EvoMining) y GSMM con experimentación con los procedimientos descritos por Guzmán & cols. (2015); de esta manera es posible ampliar el conocimiento de la evolución enzimática que experimenta la familia de enzimas PriA. Dado que es un modelo de estudio, comprender su constante evolución a diferentes niveles del sistema metabólico resulta interesante.

La regulación transcripcional en actinobacterias especialmente en la biosíntesis de L-aminoácidos, también es un campo fértil dadas las novedosas estrategias de ingeniería metabólica para mejorar los rendimientos de las fábricas de células y actualmente permanece en gran parte inexplorado.

El dar continuidad a la curación de los borradores de los GSMM elaborados en esta tesis, y poder restringirlos con datos experimentales es una posibilidad que ampliaría el conocimiento de dichos metabolismos.

En lo personal, indagar acerca del metabolismo subterráneo en microorganismos, con sus posibles aplicaciones en biotecnología industrial, resulta un camino emocionante, al ampliar las estrategias disponibles de ingeniería metabólica, con mínimas modificaciones del metabolismo nativo (metabolismo subterráneo) de los organismos, dichas estrategias están comenzando a tomar un rumbo muy prometedor.

Además, pensar en la posible inclusión de las reacciones secundarias de enzimas que se conocen, en el modelado de los metabolismos (GSM), resulta todo un reto, pero a su vez una infinita posibilidad de avance en el modelado de procesos metabólicos biosintéticos cada vez más completos.

# 11. Referencias Bibliográficas

- Adams N.E., Thiaville J.J., Proestos J., Juárez-Vázquez A.L., McCoy A.J., Barona-Gómez F., Iwata-Reuyl D., de Crecy-Lagard V. & Maurelli A.T. (2014). Promiscuous and adaptable enzymes fill "holes" in the tetrahydrofolate pathway in *Chlamydia* species. *MBio*. 5. e01378–14.
- Ahmann D., Dorgan J.R. (2007). Bioengineering for pollution prevention through development of biobased materials and energy. State of the art science report. Office of Research and Development National Center for Environmental Research.
- Amara A., Takano E., & Breitling R. (2018). Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism. *BMC genomics*, 19(1): 519.
- Amin S.A., Chavez E., Porokhin V., Nair N.U., Hassoun S. (2019). Towards creating an extended metabolic model (EMM) for *E. coli* using enzyme promiscuity prediction and metabolomics data. *Microb Cell Fact*. 18(1): 109.
- Arkin A.P., Cottingham R.W., Henry C.S., Harris N.L., Stevens R.L., Maslov S., Dehal P., Ware D., Perez F., Canon S., Sneddon M.W., Henderson M.L., Riehl W.J., Murphy-Olson D., Chan S.Y., Kamimura R.T., Kumari S., Drake M.M., Brettin T.S., Glass E.M., Chivian D., Gunter D., Weston D.J., Allen B.H., Baumohl J., Best A.A., Bowen B., Brenner S.E., Bun C.C., Chandonia J.M., Chia J.M., Colasanti R., Conrad N., Davis J.J., Davison B.H., DeJongh M., Devoid S., Dietrich E., Dubchak I., Edirisinghe J.N., Fang G., Faria J.P., Frybarger P.M., Gerlach W., Gerstein M., Greiner A., Gurtowski J., Haun H.L., He F., Jain R., Joachimiak M.P., Keegan K.P., Kondo S., Kumar V., Land M.L., Meyer F., Mills M., Novichkov P.S., Oh T., Olsen G.J., Olson R., Parrello B., Pasternak S., Pearson E., Poon S.S., Price G.A., Ramakrishnan S., Ranjan P., Ronald P.C., Schatz M.C., Seaver S.M.D., Shukla M., Sutormin R.A., Syed M.H., Thomason J., Tintle N.L., Wang D., Xia F., Yoo H., Yoo S., Yu D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol*. 36(7): 566-569.
- Arnold B.J., Huang IT. & Hanage W.P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 20: 206-218.
- Ashniev G.A., Sernova N.V., Shevkoplias A.E., Rodionov I.D., Rodionova I.A., Vitreschak A.G., Gelfand M.S., Rodionov D.A. (2022). Evolution of transcriptional regulation of histidine metabolism in Gram-positive bacteria. *BMC Genomics*. 23 (Suppl 6): 558.
- Barona-Gómez F. & Hodgson D.A. (2003). Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep* 4, 296–300.
- Becker S.A., Feist A.M., Mo M.L., Hannum G., Palsson B.Ø., Herrgard M.J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc*. 2(3): 727-38.
- Beste D.J., Nöh K., Niedenfür S., Mendum T.A., Hawkins N.D., Ward J.L., Beale M.H., Wiechert W., McFadden J. (2013). <sup>13</sup>C-flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chem Biol*. 20(8): 1012-21.
- Bokeh Development Team (2018). Bokeh: Python library for interactive visualization URL [https://docs.bokeh.org/en/latest/docs/user\\_guide/server.html](https://docs.bokeh.org/en/latest/docs/user_guide/server.html).
- Boynton Z.L., Bennet G.N., Rudolph F.B. Cloning, sequencing, and expression of clustered genes encoding beta-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. *J Bacteriol*. 178(11): 3015-24.
- Burgard A. P., Pharkya P. & Maranas C. D. (2003). OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnology and Bioengineering*. 84(6), 647–657.

- Cam Y., Alkim C., Trichez D., Trebosc V., Vax A., Bartolo F., Besse P., François J.M., Walther T. (2016). Engineering of a Synthetic Metabolic Pathway for the Assimilation of (d)-Xylose into Value-Added Chemicals. *ACS Synth Biol.* 5(7): 607-18.
- Carbonell P., Parutto P., Herisson J., Pandit S.B., Faulon J.L. (2014). XTMS: pathway design in an eXTended metabolic space. *Nucleic Acids Res.* 42 (Web Server issue): W389-94.
- Cen X., Liu Y., Chen B., Liu D., Chen Z. (2021). Metabolic engineering of *Escherichia coli* for de novo production of 1,5-pentanediol from glucose. *ACS Synth. Biol.* 10 (1), 192–203
- Central Committee of the Party of China. (2016). The 13th Five-Year Plan for Economic and Social Development of the People's Republic of China, 2016-2020. Central Compilation and Translational Press.
- Cho J.S., Kim G.B., Eun H., Moon C.W., Lee S.Y. (2022). Designing Microbial Cell Factories for the Production of Chemicals. *JACS.* 2(8):1781-1799.
- Chung C.H., Lin D.W., Eames A., Chandrasekaran S. (2021). Next-Generation Genome-Scale Metabolic Modeling through Integration of Regulatory Mechanisms. *Metabolites.* 11(9): 606.
- Copley S. (2020). Evolution of new enzymes by gene duplication and divergence. *FEBS J* 287, 1262– 1283.
- Csőrgő B., Nyerges A., Pál C. (2020). Targeted mutagenesis of multiple chromosomal regions in microbes. *Curr Opin Microbiol.* 57:22-30.
- D'Ari R., Casadesús J. (1998). Underground metabolism. *Bioessays.* 20(2): 181-186.
- De Pristo M.A., Weinreich D.M. & Hartl D.L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet.* 6: 678–687.
- Dellus-Gur E., Elias M., Caselli E., Prati F., Salverda M.L., de Visser J.A., Fraser J.S., Tawfik D.S. (2015). Negative Epistasis and Evolvability in TEM-1  $\beta$ -Lactamase--The Thin Line between an Enzyme's Conformational Freedom and Disorder. *J Mol Biol.* 427(14): 2396-409.
- Des Marais D.L. & Rausher M.D. (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454, 762–765.
- Dreyfuss J.M., Zucker J.D., Hood H.M., Ocasio L.R., Sachs M.S., Galagan J.E. (2013). Reconstruction and Validation of a Genome-Scale Metabolic Model for the Filamentous Fungus *Neurospora crassa* Using FARM. *PLOS Computational Biology.* 9: e1003126.
- Dubey, K. D.; Singh, W. Simulations Reveal the Key Role of Arg15 in the Promiscuous Activity in the HisA Enzyme. *Org. Biomol. Chem.* 2021, 19, 10652– 10661,
- Due A. V., Kuper, J., Geerlof A., von Kries J. P., Wilmanns M. (2011). Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3554–3559. doi: 10.1073/pnas.1015996108
- Echave J., Spielman S.J., Wilke C.O. (2016). Causes of evolutionary rate variation among protein sites. *Nat Rev Genet.* 17(2): 109-121.
- Edwards J. S. & Palsson B. Ø. (2000a). Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC bioinformatics.* 1(1):1.
- Edwards J. S., Ibarra R., Palsson B. Ø. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol.* 19: 125-130.
- Edwards J. S., Ramakrishna R., Palsson B. Ø. (2002). Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnology and bioengineering.* 77(1): 27-36.
- Edwards J.S. & Palsson, B. Ø. (2000b). Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnology Progress.* 16(6): 927-39.

- Fedorchuk T. P., Khusnutdinova A. N., Evdokimova E., Flick R., Di Leo R., Stogios P., Savchenko A., Yakunin A. F. (2020). One-pot biocatalytic transformation of adipic acid to 6-aminocaproic acid and 1,6-hexamethylenediamine using carboxylic acid reductases and transaminases. *J. Am. Chem. Soc.* 142(2): 1038–1048.
- Feeney M.A., Newitt J.T., Addington E., Algora-Gallardo L., Allan C., Balis L., Birke A.S., Castaño-Espriu L., Charkoudian L.K., Devine R., Gayraud D., Hamilton J., Hennrich O., Hoskisson P.A., Keith-Baker M., Klein J.G., Kruasuwan W., Mark D.R., Mast Y., McHugh R.E., McLean T.C., Mohit E., Munnoch J.T., Murray J., Noble K., Otani H., Parra J., Pereira C.F., Perry L., Pintor-Escobar L., Pritchard L., Prudence S.M.M., Russell A.H., Schniete J.K., Seipke R.F., Sélem-Mojica N., Undabarrena A., Vind K., van Wezel G.P., Wilkinson B., Worsley S.F., Duncan K.R., Fernández-Martínez L.T., Hutchings M.I. (2022) ActinoBase: tools and protocols for researchers working on *Streptomyces* and other filamentous actinobacteria. *Microb Genom.* 8: mgen000824.
- Feierabend M., Renz A., Zelle E., Nöh K., Wiechert W., Dräger A. (2021). High-Quality Genome-Scale Reconstruction of *Corynebacterium glutamicum* ATCC 13032. *Front Microbiol.* 12:750206.
- Feist A.M. & Palsson B.Ø. (2010). The biomass objective function. *Curr Opin Microbiol.* 13: 344–349.
- Force A., Lynch M., Pickett F.B., Amores A., Yan Y.L. & Postlethwait J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 151: 1531–1545.
- Galperin M. Y. & Koonin E. V. (2010). From complete genome sequence to ‘complete’ understanding? *Trends Biotechnol.* 28: 398–406.
- Ganguly S. (2023). The pivotal role of *Corynebacterium glutamicum* in l-Glutamic acid fermentation: A concise review. *Biocatal. Agric. Biotechno.* 47: 102578.
- Gill R.T., Valdes J.J., Bentley W.E. (2000). A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metab Eng.* 2(3): 178-89.
- Glasner M.E., Truong D.P., Morse B.C. (2020). How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. *FEBS J.* 287:1323–1342.
- Graf M., Haas T., Teleki A., Feith A., Cerff M., Wiechert W., Nöh K., Busche T., Kalinowski J., Takors R. (2020). Revisiting the Growth Modulon of *Corynebacterium glutamicum* Under Glucose Limited Chemostat Conditions. *Front Bioeng Biotechnol.* 8: 2296-4185.
- Griesemer M., Kimbrel J.A., Zhou C.E. Navid A., D’haeseleer P. (2018). Combining multiple functional annotation tools increases coverage of metabolic annotation. *BMC Genomics* 19, 948.
- Gu P., Yang, F., Kang, J., Wang, Q., Qi, Q., 2012. One-step of tryptophan attenuator inactivation and promoter swapping to improve the production of L-tryptophan in *Escherichia coli*. *Microbial Cell Factories* 11, 30.
- Gutiérrez-Preciado A., Yanofsky C., Merino E. (2007). Comparison of tryptophan biosynthetic operon regulation in different Gram-positive bacterial species. *Trends Genet.* 23(9): 422–426.
- Guzmán G.I., Sandberg T.E., LaCroix R.A., Nyerges Á., Papp H., de Raad M., King Z.A., Hefner Y., Northen T.R., Notebaart R.A., Pál C., Palsson B. Ø., Papp B., Feist A.M. (2019). Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol Syst Biol.* 15(4): e8462.
- Guzmán G.I., Utrilla J., Nurk S., Brunk E., Monk J.M., Ebrahim A., Palsson B. Ø., Feist A.M. (2015). Model-driven discovery of underground metabolic functions in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 112(3): 929-34.
- Haas D., Thamm A.M., Sun J., Huang L., Sun L., Beaudoin G.A.W., Wise K.S., Lerma-Ortiz C., Bruner S.D., Breuer M., Luthey-Schulten Z., Lin J., Wilson M.A., Brown G., Yakunin A.F., Kurilyak I., Folz J., Fiehn O., Glass J.I., Hanson A.D., Henry C.S., de Crécy-Lagard V. (2022). Metabolite Damage and Damage Control in a Minimal Genome. *mBio.* 13(4): e0163022.
- Hadadi N. & Hatzimanikatis V. (2015). Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. *Curr Opin Chem Biol.* 28: 99-104.

- Hafner J., Payne J., MohammadiPeyhani H., Hatzimanikatis V., Smolke C. (2021). A computational workflow for the expansion of heterologous biosynthetic pathways to natural product derivatives. *Nat Commun.* 12(1): 1760.
- Heirendt L., Arreckx S., Pfau T., Mendoza S.N., Richelle A., Heinken A., Haraldsdottir H.S., Wachowiak J., Keating S.M., Vlasov V. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc.* 14(3): 639-702.
- Henn-Sax M., Thoma R., Schmidt S., Hennig M., Kirschner K., Sterner R. (2002). Two ( $\beta\alpha$ )8-Barrel Enzymes of Histidine and Tryptophan Biosynthesis Have Similar Reaction Mechanisms and Common Strategies for Protecting Their Labile Substrates. *Biochemistry.* 41: 12032-12042.
- Henry C.S., DeJongh M., Best A.A., Frybarger P.M., Lindsay B., Stevens R.L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol.* 28: 977-982.
- Höcker B., Jürgens C., Wilmanns M., Sterner R. (2001). Stability, catalytic versatility and evolution of the ( $\beta\alpha$ )8-barrel fold. *Current Opinion in Biotechnology.* Volume 12, Issue 4, 376-381.
- Hodgson D.A. (2000). Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv. Microb. Physiol.*, 42: 47-238.
- Hommel U., Eberhard, M., Kirschner K. (1995). Phosphoribosyl Anthranilate Isomerase Catalyzes a Reversible Amadori Reaction. *Biochemistry.* 34: 5249– 5439.
- Hu D.S., Hood D.W., Heidstra R., Hodgson D.A. (1999). The expression of the *trpD*, *trpC* and *trpBA* genes of *Streptomyces coelicolor* A3(2) is regulated by growth rate and growth phase but not by feedback repression. *Mol Microbiol.* 32: 869–80.
- Hyatt D., Chen G.L., LoCascio P.F, Land M.L., Larimer F.W., Hauser L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 11:119.
- Innan H. & Kondrashov F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 11, 97–108.
- Itoh T., Takemoto K., Mori H., Gojobori T. (1999). Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. Evol.* 16:332-346.
- Jakočiūnas T., Jensen M.K., Keasling J.D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Metab Eng.* 34: 44-59.
- Jayaraman V., Toledo-Patiño S., Noda-García L., Laurino P. (2022). Mechanisms of protein evolution. *Protein Sci.* 31(7): e4362.
- Jin S., Hui M., Lu Y., Zhao Y. (2023). An overview on the two-component systems of *Streptomyces coelicolor*. *World J Microbiol Biotechnol.* 39(3): 78.
- Juárez-Vázquez A.L., Edirisinghe J.N., Verduzco-Castro E.A., Michalska K., Wu C., Noda-García L., Babnigg G., Endres M., Medina-Ruiz S., Santoyo-Flores J. et al. (2017). Evolution of substrate specificity in a retained enzyme driven by gene loss. *ELife.* 6: e22679.
- Jung S., Chun J.Y., Yim S.H., Lee S.S., Cheon C.I., Song E., Lee M.S. (2010). Transcriptional regulation of histidine biosynthesis genes in *Corynebacterium glutamicum*. *Can J Microbiol.* 56(2): 178–87.
- Jürgens C., Strom A., Wegener D. Hettwer S., Wilmanns M., Sterner R. (2000). Directed evolution of a ( $\beta\alpha$ )8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. USA*97: 9925-9930.
- Kadumuri R. V., Vadrevu R. (2018). Diversity in  $\alpha\beta$  and  $\beta\alpha$  Loop Connections in TIM Barrel Proteins: Implications for Stability and Design of the Fold. *Interdiscip. Sci. Comput. Life Sci.* 10: 805-812.
- Kanehisa M., Furumichi M., Sato Y., Kawashima M., Ishiguro-Watanabe M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research.* 51(D1): D587–D592.

- Kanehisa M., Sato Y., Morishima K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol.* 428(4):726-731.
- Kang A., Meadows C. W., Canu N., Keasling J. D., Lee T. S. (2017). High-throughput enzyme screening platform for the IPP-bypass mevalonate pathway for isopentenol production. *Metab. Eng.* 41: 125-134.
- Kauffman K.J., Prakash P., Edwards J.S. (2003). Advances in flux balance analysis. *Curr Opin Biotechnol.* 14:491-496.
- Kavvas E.S., Seif Y., Yurkovich J.T., Norsigian C., Poudel S., Greenwald W.W., Ghatak S., Palsson B. Ø., Monk J.M. (2018). Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Syst Biol.* 12(1): 25.
- Kempton C., Kaiser D., Haag S., Nicholson G., Gnau V., Walk T., Gierling K. H., Decker H., Zahner H., Jung G., Metzger J. W. (1997). CDA: calcium-dependent antibiotic from *Streptomyces coelicolor* A3(2) containing unusual residues. *Angew. Chem. Int. Ed. Engl.* 36:498-501.
- Keseler I.M., Gama-Castro S., Mackie A., Billington R., Bonavides-Martínez C., Caspi R., Kothari A., Krummenacker M., Midford P.E., Muñoz-Rascado L., Ong W.K., Paley S., Santos-Zavaleta A., Subhraveti P., Tierrafria V.H., Wolfe A.J., Collado-Vides J., Paulsen I.T., Karp P.D. (2021). The EcoCyc Database in 2021. *Front Microbiol.* 12: 711077.
- Khersonsky O., Tawfik D.S. (2010). Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu Rev Biochem.* 79: 471-505.
- Kim J. & Copley S.D. (2012). Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc Natl Acad Sci USA.* 109(42): E2856-64.
- King J.R., Woolston B.M., Stephanopoulos G. (2017). Designing a New Entry Point into Isoprenoid Metabolism by Exploiting Fructose-6-Phosphate Aldolase Side Reactivity of *Escherichia coli*. *ACS Synth Biol.* 6(7): 1416-1426.
- King Z.A., Dräger A., Ebrahim A., Sonnenschein N., Lewis N.E., Palsson B. Ø. (2015). Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput Biol.* 11(8): e1004321.
- King Z.A., Lu J., Dräger A., Miller P., Federowicz S., Lerman J.A., Ebrahim A., Palsson B. Ø., Lewis N.E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44(D1): D515- D522.
- Ko Y.S., Kim J.W., Lee J.A., Han T., Kim G.B., Park J.E., Lee S.Y. (2020). Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem Soc Rev.* 49(14): 4615-4636.
- Koch M., Duigou T., Carbonell P., Faulon J.L. (2017). Molecular structures enumeration and virtual screening in the chemical space with RetroPath 2.0. *J Cheminform.* 9(1): 64.
- Kolbe D.L. & Eddy S.R. (2011). Fast filtering for RNA homology search. *Bioinformatics.* 27: 3102-3109.
- Kovács S.C., Szappanos B., Tengölics R., Notebaart R.A., Papp B. (2022). Underground metabolism as a rich reservoir for pathway engineering. *Bioinformatics.* 38(11): 3070-3077.
- Kreuzer K.D. & Henkin T.M. (2018). The T-Box Riboswitch: tRNA as an Effector to Modulate Gene Regulation. *Microbiol Spectr.* 6(4): 6-4.
- Lagesen K., Hallin P., Rødland E.A., Staerfeldt H.H., Rognes T., Ussery D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35(9): 3100-3108.
- Laslett D. & Canback B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11-16.
- Latendresse M. (2014). Efficiently Gap-filling reaction networks. *BMC Bioinformatics.* 15: 225.
- Lechner A., Brunk E., Keasling J.D. (2016). The Need for Integrated Approaches in Metabolic Engineering. *Cold Spring Harb Perspect Biol.* 8(11): a023903.
- Levis N.A., Pfennig D.W. (2016). Evaluating “plasticity-first” evolution in nature: Key criteria and empirical approaches. *Trends Ecol Evol.* 31(7): 563-574.

- Lewis N. E. & Abdel-Haleem A. M. (2013). The evolution of genome-scale models of cancer metabolism. *Front. Physiol.* 4: 237.
- Li A., Li L., Zhang Z., Li S., Wang W., Guo X. (2021). Noncoding variation and transcriptional plasticity promote thermal adaptation in oysters by altering energy metabolism. *Mol Biol Evol.* 38(11): 5144-5155.
- Liao Q., Kulkarni Y., Sengupta U., Petrović D., Mulholland A. J., van der Kamp M. W., Strodel B., Kamerlin S. C. L. (2018). Loop Motion in Triosephosphate Isomerase is not a Simple Open and Shut Case. *J. Am. Chem. Soc.* 140: 15889-15903.
- Lieven C., Beber M.E., Olivier B.G., Bergmann F.T., Ataman M., Babaei P., Bartell J.A., Blank L.M., Chauhan S., Correia K., Diener C., Dräger A., Ebert B.E., Edirisinghe J.N., Faria J.P., Feist A.M., Fengos G., Fleming R.M.T., García-Jiménez B., Hatzimanikatis V., van Helvoirt W., Henry C.S., Hermjakob H., Herrgård M.J., Kaafarani A., Kim H.U., King Z., Klamt S., Klipp E., Koehorst J.J., König M., Lakshmanan M., Lee D.Y., Lee S.Y., Lee S., Lewis N.E., Liu F., Ma H., Machado D., Mahadevan R., Maia P., Mardinoglu A., Medlock G.L., Monk J.M., Nielsen J., Nielsen L.K., Nogales J., Nookaew I., Palsson B. Ø., Papin J.A., Patil K.R., Poolman M., Price N.D., Resendis-Antonio O., Richelle A., Rocha I., Sánchez B.J., Schaap P.J., Malik Sheriff R.S., Shoaie S., Sonnenschein N., Teusink B., Vilaça P., Vik J.O., Wodke J.A.H., Xavier J.C., Yuan Q., Zakhartsev M., Zhang C. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol.* 38(3): 272-276.
- Liu L., Chen S., Wu J., (2017). Phosphoenolpyruvate: glucose phosphotransferase system modification increases the conversion rate during L-tryptophan production in *Escherichia coli*. *Journal of Industrial Microbiology and Biotechnology* 44, 1385–1395.
- Loghmani S.B., Veith N., Sahle S., Bergmann F.T., Olivier B.G., Kummer U. (2022). Inspecting the Solution Space of Genome-Scale Metabolic Models. *Metabolites.* 5.12(1): 43.
- Longo L.M., Petrović D., Kamerlin S.C.L., Tawfik D.S. (2020). Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes. *Proc Natl Acad Sci U S A.* 117(10): 5310-5318.
- López-Reyes N., Ábrego-García A., Poggi-Varaldo, H. M. (2022). Mecanismos de inhibición de la metanogénesis con lovastatina y análisis de la expresión génica. *Revista Internacional de Contaminación Ambiental.* 38: 58-67.
- Ma H.M., Zhou Q., Tang Y.M., Zhang Z., Chen Y.S., He H.Y., Pan H.X., Tang M.C., Gao J.F., Zhao S.Y., Igarashi Y., Tang G.L. (2013). Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in kosinostatin biosynthesis. *Chem. Biol.* 20. 796-805
- Mallik S., Tawfik D.S., Levy E.D. (2022). How gene duplication diversifies the landscape of protein oligomeric state and function. *Curr Opin Genet Dev.* 76: 101966.
- Martin V.J., Pitera D.J., Withers S.T., Newman J.D., Keasling J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol.* 21(7): 796-802.
- MATLAB. (2022). versión 9.13 (R2022b). Natick, Massachusetts: MathWorks Inc.
- Melzoch K., de Mattos M.J., Neijssel O.M. (1997). Production of actinorhodin by *Streptomyces coelicolor* A3(2) grown in chemostat culture. *Biotechnol Bioeng.* 54(6): 577-82.
- Merino E., Jensen R.A., Yanofsky C. (2008). Evolution of bacterial Trp operons and their regulation. *Curr Opin Microbiol.* 11: 78-86.
- Mhashal A. R., Romero-Rivera A., Mydy L. S., Cristobal J. R., Gulick A. M., Richard J. P., Kamerlin S. C. L. (2020). Modeling the Role of a Flexible Loop and Active Site Side Chains in Hydride Transfer Catalyzed by Glycerol-3-phosphate Dehydrogenase. *ACS Catal.* 10: 11253-11267.
- Miton C.M. & Tokuriki N. (2016). How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.* 25: 1260-1272.
- MohammadiPeyhani H., Hafner J., Sveshnikova A., Viterbo V., Hatzimanikatis V. (2022). Expanding biochemical knowledge and illuminating metabolic dark matter with ATLASx. *Nat Commun.* 13(1): 1560

- Mou Z., Eakes J., Cooper C.J., Foster C.M., Standaert R.F., Podar M., Doktycz M.J., Parks J.M. (2021). Machine learning-based prediction of enzyme substrate scope: Application to bacterial nitrilases. *Proteins*. 89(3): 336-347.
- Nam H., Lewis N.E., Lerman J.A., Lee D.H., Chang R.L., Kim D., Palsson B. Ø. (2012). Network context and selection in the evolution to enzyme specificity. *Science*. Aug 31; 337 (6098): 1101-4.
- Näsvall J.; Sun L.; Roth J. R.; Andersson D. I. (2012). Real-Time Evolución of New Genes by Innovation, Amplification and Divergence. *Science*. 338: 384-387.
- Navarro-Muñoz J.C., Selem-Mojica N., Mullowney M.W., Kautsar S.A., Tryon J.H., Parkinson E.I., De Los Santos E.L.C., Yeong M., Cruz-Morales P., Abubucker S., Roeters A., Lokhorst W., Fernandez-Guerra A., Cappelini L.T.D., Goering A.W., Thomson R.J., Metcalf W.W., Kelleher N.L., Barona-Gómez F., Medema M.H. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol*. 16(1): 60-68.
- Newton M.S., Guo X., Soderholm A., Nasvall J., Lundstrom P., Andersson D.I., Selmer M. & Patrick W.M. (2017). Structural and functional innovations in the real-time evolution of new (betaalpha)<sub>8</sub> barrel enzymes. *Proc Natl Acad Sci USA*. 114: 4727-4732.
- Nielsen J. & Keasling J.D. (2016). Engineering Cellular Metabolism. *Cell*. 164(6): 1185-1197.
- Noda-García L., Camacho-Zarco A.R., Medina-Ruiz S., Gaytan P., Carrillo-Tripp M., Fulop V., Barona-Gómez F. (2013). Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Mol. Biol. Evol* 30: 2024–2034
- Noda-García L., Camacho-Zarco A.R., Verdel-Aranda K., Wright H., Soberon X., Fulop V. & Barona-Gómez F. (2010). Identification and analysis of residues contained on beta → alpha loops of the dual-substrate (beta alpha)<sub>8</sub> phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci* 19, 535-543.
- Noda-García L., Juárez-Vázquez A.L., Ávila-Arcos M.C., Verduzco-Castro E.A., Montero-Morán G., Gaytán P., Carrillo-Tripp M., Barona-Gómez F. (2015). Insights into the evolution of enzyme substrate promiscuity after the discovery of (β $\alpha$ )<sub>8</sub> isomerase evolutionary intermediates from a diverse metagenome. *BMC Evol. Biol*. 15: 107.
- Noda-García L., Liebermeister W., Tawfik D.S. (2018). Metabolite-Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks. *Annu Rev Biochem*. 87:187-216.
- Noda-García L., Romero-Romero M.L., Longo L.M., Kolodkin-Gal I., Tawfik D.S. (2017). Bacilli glutamate dehydrogenases diverged via coevolution of transcription and enzyme regulation. *EMBO Rep*. 18(7):1139-1149.
- Notebaart R. A., Kintsjes B., Feist A. M., Papp B. (2018). Underground metabolism: network-level perspective and biotechnological potential. *Curr. Opin. Biotechnol*. 49: 108-114.
- Notebaart R.A., Szappanos B., Kintsjes B., Pál F., Györkei Á., Bogos B., Lázár V., Spohn R., Csörgő B., Wagner A., Ruppín E., Pál C., Papp B. (2014). Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci. USA*. 111(32): 11762-7.
- Ochoa-Leyva A., Montero-Morán G., Saab-Rincón G., Brieba L. G., Soberón X. (2013). Alternative Splice Variants in TIM Barrel Proteins from Human Genome Correlate with the Structural and Evolutionary Modularity of this Versatile Protein Fold. *PLoS One*. 8: e70582.
- Ofaim S., Zarecki R., Porob S., Gat D., Lahav T., Kashi Y., Aly R., Eizenberg H., Ronen Z., Freilich S. (2020). Genome-scale reconstruction of *Paenarthrobacter aurescens* TC1 metabolic model towards the study of atrazine bioremediation. *Sci Rep*. 10(1): 13019.
- Okano K., Uematsu G., Hama S., Tanaka T., Noda H., Kondo A., Honda K. (2018). Metabolic Engineering of *Lactobacillus plantarum* for direct L-Lactic Acid Production From Raw Corn Starch. *Biotechnol J*. 13(5): e1700517.
- Orth J. D., Thiele I., Palsson B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, 28(3): 245-248.
- Overbeek R., Olson R., Pusch G.D., Olsen G.J., Davis J.J., Disz T., Edwards R.A., Gerdes S., Parrello B., Shukla M., Vonstein V., Wattam A.R., Xia F., Stevens R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 42(Database): D206-14.

- Palazzotto E., Gallo G., Renzone G., Giardina A., Suter A., Silva J., Vocat C., Botta L., Scaloni A., Puglia A.M. (2016). TrpM, a Small Protein Modulating Tryptophan Biosynthesis and Morpho-Physiological Differentiation in *Streptomyces coelicolor* A3(2). PLoS One. 11(9): e0163422.
- Palazzotto E., Renzone G., Fontana P., Botta L., Scaloni A., Puglia A.M., Gallo G. (2015). Tryptophan promotes morphological and physiological differentiation in *Streptomyces coelicolor*. Appl Microbiol Biotechnol. 99(23): 10177-89.
- Passi A., Tibocha-Bonilla J.D., Kumar M., Tec-Campos D., Zengler K., Zuniga C. (2021). Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data. Metabolites. 12(1): 14.
- Patra P., Disha B.R., Kundu P., Das M., Ghosh A. (2023) Recent advances in machine learning applications in metabolic engineering. Biotechnology Advances. 62, 108069.
- Petersen T.N., Brunak S., von Heijne G., Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8(10): 785-786.
- Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 2004 Nov;14(11):2367-76.
- Pharkya P., Burgard A.P., Maranas C.D. (2003). Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol Bioeng. 84(7): 887-99. doi: 10.1002/bit.10857.
- Pickens L.B., Tang Y., Chooi Y.H. (2011). Metabolic engineering for the production of natural products. Annu Rev Chem Biomol Eng. 2: 211-36.
- Plach M.G., Reisinger B., Sterner R. & Merkl R. (2016). Long-term persistence of Bi-functionality contributes to the robustness of microbial life through exaptation. PLoS Genet. 12: e1005836.
- Pontrelli S., Chiu T.Y., Lan E.I., Chen F.Y., Chang P., Liao J.C. (2018). *Escherichia coli* as a host for metabolic engineering. Metab Eng. 50: 16-46.
- Porokhin V., Amin S.A., Nicks T.B., Gopinarayanan V.E., Nair N.U., Hassoun S. (2021). Analysis of metabolic network disruption in engineered microbial hosts due to enzyme promiscuity. Metab Eng Commun. 12: e00170.
- Prabowo C. P. S.; Shin J. H.; Cho J. S.; Chae T. U.; Lee S. Y. (2020). Microbial production of 4-amino-1-butanol, a four-carbon amino alcohol. Biotechnol. Bioeng. 117(9): 2771-2780.
- Raman K., Chandra N. (2009). Flux balance analysis of biological systems: applications and challenges, Briefings in Bioinformatics. 10(4): 435-449.
- Revuelta J.L., Buey R. M., Ledesma-Amaro R., Vandamme E.J. (2016). Microbial biotechnology for the synthesis of (pro)vitamins, biopigments and antioxidants: challenges and opportunities. Microb. Biotechnol. 9: 564-567.
- Richard J. P. (2019). Protein Flexibility and Stiffness Enable Efficient Enzymatic Catalysis. J. Am. Chem. Soc. 141: 3320-3331.
- Richard J. P. (2022). Enabling Role of Ligand-Driven Conformational Changes in Enzyme Evolution. Biochemistry. 61(15): 1533-1542.
- Richts B., Commichau F.M. (2021). Underground metabolism facilitates the evolution of novel pathways for vitamin B6 biosynthesis. Appl Microbiol Biotechnol. 105(6):2297-2305.
- Rives A., Meier J., Sercu T., Goyal S., Lin Z., Liu J., Guo D., Ott M., Zitnick L., Ma J., Fergus R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences. 118(15): e2016239118.
- Romero-Rivera A., Corbella M., Parracino A., Patrick W.M. & Lynn Kamerlin S. C. (2022). Complex Loop Dynamics Underpin Activity, Specificity, and Evolvability in the ( $\beta\alpha$ )8 Barrel Enzymes of Histidine and Tryptophan Biosynthesis. JACS. 2(4): 943-960.
- Romero-Romero S., Kordes S., Michel F., Höcker B. (2021). Evolution, Folding, and Design of TIM Barrels and Related Proteins. Curr. Opin. Struct. Biol. 68: 94-104.
- Roseman S. (1969). The Transport of Carbohydrates by a Bacterial Phosphotransferase System. J Gen Physiol. 54: 138-184.

- Rosenberg J., Commichau F.M. (2019) Harnessing underground metabolism for pathway development. *Trends Biotechnol.* 37: 29-37.
- Rotem E., Loinger A., Ronin I., Levin-Reisman I., Gabay C., Shoresh N., Biham O., Balaban N.Q. (2010). Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proc Natl Acad Sci U S A.* 107(28): 12541-12546.
- Ryu J.Y., Kim H.U., Lee S.Y. (2019). Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. *Proc Natl Acad Sci U S A.* 116(28): 13996-14001.
- Santos-Zavaleta A., Salgado H., Gama-Castro S., Sánchez-Pérez M., Gómez-Romero L., Ledezma Tejeida D., García-Sotelo J.S., Alquicira-Hernández K., Muñiz-Rascado L.J., Peña-Loredo P., Ishida-Gutiérrez C., Velázquez-Ramírez D.A., Del Moral-Chávez V., Bonavides-Martínez C., Méndez-Cruz C.-F., Galagan J., Collado-Vides J. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Research.* 47: D212–D220.
- Schniete J.K., Cruz-Morales P., Selem-Mojica N., Fernández-Martínez L.T., Hunter I.S., Barona-Gómez F., Hoskisson P.A. (2018). Expanding Primary Metabolism Helps Generate the Metabolic Robustness To Facilitate Antibiotic Biosynthesis in *Streptomyces*. *mBio.* 9(1): e02283-17.
- Seaver S.M.D., Liu F., Zhang Q., Jeffryes J., Faria J.P., Edirisinghe J.N., Mundy M., Chia N., Noor E., Beber M.E., Best A.A., DeJongh M., Kimbrel J.A., D'haeseleer P., McCorkle S.R, Bolton J.R., Pearson E., Canon S., Wood-Charlson E.M., Cottingham R.W., Arkin A.P., Henry C.S. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research.* 49(D1): D575-D588.
- Shi W. (2021). Activity of Pyrazinamide against *Mycobacterium tuberculosis* at Neutral pH in PZA-S1 Minimal Medium. *Antibiotics.* 10: 909.
- Shieh Y.W., Minguez P., Bork P., Auburger J.J., Guilbride D.L., Kramer G., Bukau B. (2015). Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science.* 350: 678-680.
- Sikosek T., Chan H.S., Bornberg-Bauer E. (2012). Escape from adaptive conflict follows from weak functional trade-offs and mutational robustness. *Proc Natl Acad Sci U S A.* 109(37): 14888-14893.
- Starr T.N. & Thornton J.W. (2016). Epistasis in protein evolution. *Protein Sci.* 25: 1204-1218.
- Sulheim S., Kumelj T., Van Dissel D., Salehzadeh-Yazdi A., Du C., Van Wezel G.P., Nieselt K., Almaas E., Wentzel A., Kerkhoven E.J. (2020). Enzyme-Constrained Models and Omics Analysis of *Streptomyces coelicolor* Reveal Metabolic Changes that Enhance Heterologous Production. *iScience.* 23(9): 101525.
- Tamura K., Stecher G., Kumar S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38: 3022-3027.
- Tawfik D.S. (2014). Accuracy-rate tradeoffs: How do enzymes meet demands of selectivity and catalytic efficiency? *Curr Opin Chem Biol.* 21:73-80.
- Thiele I., Palsson B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.* 5: 93-121.
- Tokuriki N., Jackson C..J, Afriat-Jurnou L., Wyganowski K.T., Tang R., Tawfik D.S. (2012). Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun.* 3:1257.
- Tóth-Petróczy Á., Tawfik D.S. (2013). Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol Biol Evol.* 30(4): 761-771.
- Tyzack J.D., Ribeiro A.J.M., Borkakoti N., Thornton J.M. (2019). Exploring Chemical Biosynthetic Design Space with Transform-MinER. *ACS Synth Biol.* 8(11): 2494-2506.

- Unthan S., Grünberger A., van Ooyen J., Gätgens J., Heinrich J., Paczia N., Wiechert W., Kohlheyer D., Noack S. (2014). Beyond growth rate 0.6: What drives *Corynebacterium glutamicum* to higher growth rates in defined medium. *Biotechnol Bioeng.* 111(2): 359-71.
- Van't Hof M., Mohite O.S., Monk J.M., Weber T., Palsson B. Ø., Sommer M.O.A. (2022). High-quality genome-scale metabolic network reconstruction of probiotic bacterium *Escherichia coli* Nissle 1917. *BMC Bioinformatics.* 23(1): 566.
- Vassallo A., Palazzotto E., Renzone G., Botta L., Faddetta T., Scaloni A., Puglia A.M., Gallo G. (2020). The *Streptomyces coelicolor* Small ORF *trpM* Stimulates Growth and Morphological Development and Exerts Opposite Effects on Actinorhodin and Calcium-Dependent Antibiotic Production. *Front Microbiol.* 11: 224.
- Verduzco-Castro, E., Michalska, K., Endres, M., Juárez-Vázquez, A., L., Noda-García, L., Chang, C., Barona-Gómez, F. (2016). Co-occurrence of analogous enzymes determines evolution of a novel ( $\beta\alpha$ )8-isomerase sub-family after non-conserved mutations in flexible loop. *The Biochemical Journal*, 473(9): 1141-1152.
- Vitreschak A.G., Mironov A.A., Lyubetsky V.A., Gelfand M.S. (2008). Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA.* 14(4): 717-35.
- Vlasov V. (2018). COBRA.tutorials/analysis/robustnessPhPP/tutorial\_robustnessPhPP.m. Systems Biochemistry Group, LCSB, University of Luxembourg. Disponible en: [https://github.com/opencobra/COBRA.tutorials/blob/master/analysis/robustnessPhPP/tutorial\\_robustnessPhPP.m](https://github.com/opencobra/COBRA.tutorials/blob/master/analysis/robustnessPhPP/tutorial_robustnessPhPP.m)
- Wang H.H., Isaacs F.J., Carr P.A., Sun Z.Z., Xu G., Forest C.R., Church G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature.* 460(7257): 894-898.
- Wannier T.M., Nyerges A., Kuchwara H.M., Czikkely M., Balogh D., Filsinger G.T., Borders N.C., Gregg C.J., Lajoie M.J., Rios X., Pál C., Church G.M. (2020). Improved bacterial recombineering by parallelized protein discovery. *Proc Natl Acad Sci USA.* 117(24): 13689-13698.
- Wierenga R. K. (2001). The TIM-Barrel Fold: A Versatile Framework for Efficient Enzymes. *FEBS Lett.* 492: 193-198.
- Wright H., Noda-García L., Ochoa-Leyva A., Hodgson D.A., Fülöp V., Barona-Gómez F. (2008). The structure/function relationship of a dual-substrate ( $\beta\alpha$ )8-isomerase. *Biochem Biophys Res Commun.* 365(1): 16-21.
- Xie G., Bonner C.A., Song J., Keyhani N.O., Jensen R.A. (2004). Inter-genomic displacement via lateral gene transfer of bacterial *trp* operons in an overall context of vertical genealogy. *BMC Biol.* 2: 15
- Xie G., Keyhani N.O., Bonner C.A., Jensen R.A. (2003). Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev.* 67: 303-342.
- Yanofsky C. (2007). RNA-based regulation of genes of tryptophan synthesis and degradation in bacteria. *RNA.* 13: 1141-1154.
- Yao S., You R., Wang S., Xiong Y., Huang X., Zhu S. (2021) NetGO 2.0: improving large-scale protein function prediction with massive sequence, text, domain, family and network information. *Nucleic Acids Research.* 49(W1): W469-W475.
- Yim H., Haselbeck R., Niu W., Pujol-Baxley C., Burgard A., Boldt J., Khandurina J., Trawick J.D., Osterhout R.E., Stephen R., Estadilla J., Teisan S., Schreyer H.B., Andrae S., Yang T.H., Lee S.Y., Burk M.J., Van Dien S. (2011). Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol.* 7(7): 445-52.
- Zhang J., Hansen L.G., Gudich O., Viehrig K., Lassen L.M.M., Schrübbers L., Adhikari K.B., Rubaszka P., Carrasquer-Alvarez E., Chen L., D'Ambrosio V., Lehka B., Haidar A.K., Nallapareddy S., Giannakou K., Laloux M., Arsovska D., Jørgensen M.A.K., Chan L.J.G., Kristensen M., Christensen H.B., Sudarsan S., Stander E.A., Baidoo E., Petzold C.J., Wulff T., O'Connor S.E., Courdavault V., Jensen M.K., Keasling J.D. (2022). A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature.* 609(7926): 341-347.

# 12. Anexos

## 12.1. Anexo I

**Tabla I.a. Resultados de los borradores de GSMM con los mejores puntajes de MEMOTE para cada cepa y criterio.** El identificador en la columna “Borrador de GSMM” corresponde a la letra inicial del género de la cepa seguido de las primeras letras de la especie, seguido después del guion bajo por el criterio (en este caso solo se anotó el borrador que obtuvo la mejor puntuación general en MEMOTE). La notación de criterio 2o+ corresponde a “dos o más”. Se muestra el número de reacciones que fueron añadidas por “Gapfill” (llenado de brechas), así como el número de reacciones (Rxn), metabolitos (Met) y Genes de cada borrador de GSMM. Los borradores de GSMM de color verde ya contaban con un GSMM al inicio de esta tesis, pero aun así decidí crear un borrador de *novo* con fines comparativos.

| Borrador de GSMM | Rxn  | Met  | Genes | Gapfill | MEMOTE |
|------------------|------|------|-------|---------|--------|
| Sco_Prioridad    | 1868 | 1990 | 1860  | 59      | 92%    |
| Mtuber_2o+       | 1063 | 1161 | 637   | 67      | 93%    |
| Cjeike_2o+       | 892  | 939  | 408   | 90      | 92.22% |
| Cdiph_2o+        | 890  | 918  | 441   | 78      | 92.09% |
| Ceffic_2o+       | 979  | 1027 | 524   | 74      | 92.10% |
| Cstria_2o+       | 951  | 976  | 480   | 83      | 92.07% |
| Cglut_2o+        | 945  | 989  | 523   | 72      | 92%    |
| Cmatru_2o+       | 683  | 761  | 380   | 83      | 92.53% |
| Camycola_2o+     | 851  | 941  | 444   | 73      | 92.78% |
| Aferro_2o+       | 765  | 895  | 420   | 68      | 92.30% |
| Blongum_2o+      | 768  | 901  | 343   | 59      | 92.51% |
| Msmeg_Prioridad  | 1827 | 1946 | 2072  | 75      | 90.79% |
| Paure_Prioridad  | 1536 | 1684 | 1544  | 63      | 91.11% |
| Sglobis_2o+      | 1079 | 1197 | 803   | 72      | 92.01% |
| SMg1_2o+         | 1050 | 1138 | 828   | 59      | 92.02% |
| Ssvic_2o+        | 1119 | 1219 | 840   | 73      | 91.98% |
| Sipomo_2o+       | 1110 | 1206 | 873   | 62      | 92.04% |
| SspC_2o+         | 1061 | 1148 | 761   | 82      | 92.10% |
| Bgallcum_2o+     | 785  | 906  | 334   | 87      | 92.31% |
| Badolesc_2o+     | 726  | 838  | 351   | 68      | 92.35% |
| Aurogeni_2o+     | 792  | 889  | 415   | 90      | 92.35% |
| A171_2o+         | 770  | 846  | 401   | 97      | 92.46% |
| Aoris_2o+        | 780  | 852  | 407   | 98      | 92.38% |
| Acardi_Prioridad | 929  | 1048 | 680   | 63      | 92.17% |
| Aodonto_2o+      | 730  | 827  | 382   | 87      | 92.37% |
| Avacc_2o+        | 763  | 837  | 397   | 65      | 92.19% |
| A848_2o+         | 758  | 826  | 353   | 87      | 92.27% |
| A180_2o+         | 720  | 815  | 356   | 67      | 92.44% |
| Agraev_2o+       | 659  | 776  | 321   | 97      | 92.33% |

**Tabla. I.b.** Comparación de las diferentes anotaciones funcionales y cuatro criterios de fusión para el genoma de *Cgl* ATCC 13032 con los datos para genes, términos únicos, reacciones ModelSEED únicas, pares únicos de gen/reacción ModelSEED. Informe obtenido de Kbase.

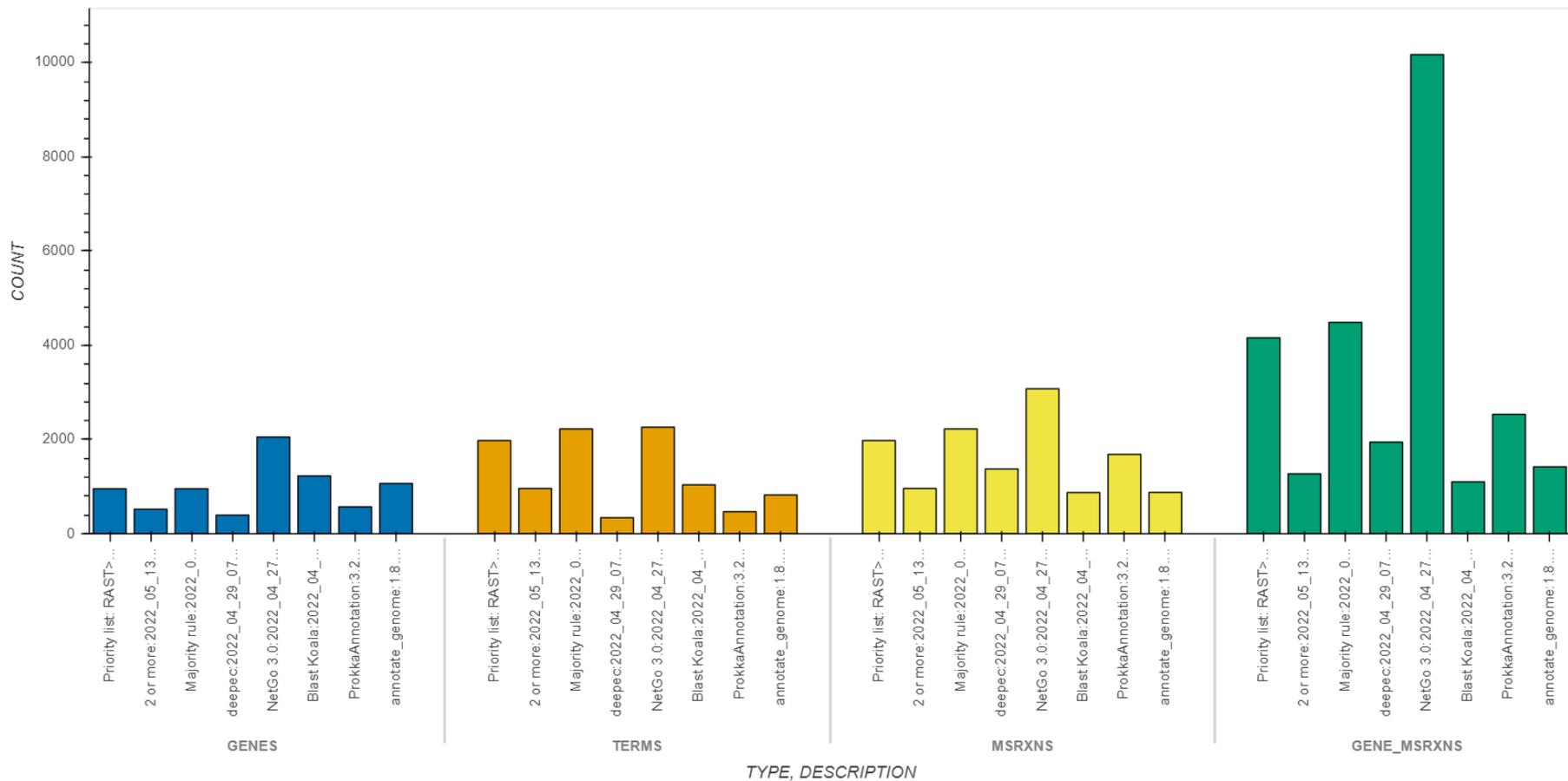
Compare Annotations Summary

| Description      | Timestamp           | KBase App                  | Ontology | Genes | Unique Terms | Unique ModelSEED rxns | Unique Gene/ModelSEED rxn pairs |
|------------------|---------------------|----------------------------|----------|-------|--------------|-----------------------|---------------------------------|
| annotate_genome  | 2022-04-20T09:37:52 | annotate_genome v1.8.1     | SSO      | 1313  | 946          | 902                   | 1489                            |
| ProkkaAnnotation | 2022_04_20_12_10_34 | Prokka Annotation v3.2.1   | EC       | 763   | 590          | 2233                  | 3122                            |
| Blast Koala      | 2022_04_21_13_56_53 | Import Annotations v1.2.11 | KO       | 1509  | 1266         | 1006                  | 1422                            |
| netgo3           | 2022_04_28_07_37_02 | Import Annotations v1.2.11 | GO       | 2962  | 2736         | 3413                  | 18264                           |
| deepec           | 2022_04_29_07_29_08 | Import Annotations v1.2.11 | EC       | 472   | 400          | 1703                  | 2699                            |
| Majority rule    | 2022_05_09_18_37_38 | Merge Annotations v1.2.11  | MSRXN    | 1215  | 3048         | 3048                  | 8476                            |
| 2 or more        | 2022_05_09_22_30_23 | Merge Annotations v1.2.11  | MSRXN    | 647   | 1049         | 1049                  | 1443                            |
| Priority         | 2022_05_09_22_35_03 | Merge Annotations v1.2.11  | MSRXN    | 1215  | 2352         | 2352                  | 6910                            |
| Union            | 2023_01_31_10_50_57 | Merge Annotations v1.2.11  | MSRXN    | 1215  | 5061         | 5061                  | 21969                           |

**Tabla I.c.** Veintinueve cepas reportadas conteniendo en sus genomas a la enzima PriA o alguna de sus subespecializaciones. Es visible la referencia para cada cepa, el tamaño de genoma en Mbp, CDS, identificador taxonómico de la cepa, RefSeq, el lugar del aislamiento original de la cepa, y en algunos casos si lo tiene su uso biotecnológico. Las cepas de color **verde** ya contaban con un GSMM al inicio de esta tesis; para el caso de la cepa *Paenarthrobacter aurescens* TC1, el GSMM reportado por Ofaim & Zarecki et al. 2020, no se consideró para este trabajo de tesis debido a que es un modelo que se encuentra en su primera generación, por lo que su curación manual y procesos de reconstrucción resultan insuficientes para los fines de este trabajo.

| Referencia/Cepa                                  | Enzima        | Tamaño<br>genoma<br>(Mbp) | Genes | Protein-<br>coding<br>(CDS) | GC % | Taxonomy<br>ID | RefSeq          | Accesión          | Lugar de aislamiento/Usos biotecnológico   |
|--|---------------|---------------------------|-------|-----------------------------|------|----------------|-----------------|-------------------|--|
| <b>Noda-García et al. 2010</b>                   |               |                           |       |                             |      |                |                 |                   |  |
| <i>Streptomyces coelicolor A3(2)</i>             | PriA          | 8.668                     | 7765  | 7491                        | 72   | 100226         | GCF_000203835.1 | NC_003888.3       | Vida libre, suelo, agua. 2/3 producción de antibióticos.                                     |
| <b>Due et al. 2011</b>                           |               |                           |       |                             |      |                |                 |                   |  |
| <i>Mycobacterium tuberculosis H37Rv</i>          | PriA          | 4.412                     | 4008  | 3906                        | 65.5 | 83332          | GCF_000195955.2 | NC_000962         | Cepa virulenta de H37 (paciente de hospital)   |
| <b>Noda-García et al. 2013</b>                   |               |                           |       |                             |      |                |                 |                   |  |
| <i>Corynebacterium jeikeium K411</i>             | PriA          | 2.477                     | 2179  | 2055                        | 61   | 306537         | GCF_000006605.1 | NC_007164         | Suelo y agua. Microbioma de piel humana. Aumenta en ambiente nosocomial                      |
| <i>Corynebacterium diphtheriae NCTC 13129</i>    | SubHisA1      | 2.489                     | 2334  | 2272                        | 53.5 | 257309         | GCF_000195815.1 | NC_002935         | Hemocultivo  |
| <i>Corynebacterium efficiens YS-314</i>          | SubHisA1      | 3.22                      | 2943  | 2816                        | 62.5 | 196164         | GCF_000011305.1 | NC_004369         | Suelo y vegetales. Productora de ácido glutámico   |
| <i>Corynebacterium striatum ATCC 6940</i>        | SubHisA1      | 2.829                     | 2634  | 2472                        | 57.2 | 525268         | GCF_000159135.1 | NZ_GG667603       | Diversas superficies y equipos médicos en entornos hospitalarios. (Infecciones nosocomiales) |
| <i>Corynebacterium glutamicum ATCC 13032</i>     | SubHisA1      | 3.309                     | 3080  | 2959                        | 53.8 | 196627         | GCF_000011325.1 | NC_003450.3       | Aguas residuales. Productora de ácido glutámico, aminoácidos y vitaminas                     |
| <i>Corynebacterium matruchotii ATCC 14266</i>    | SubHisA1      | 2.856                     | 2566  | 2484                        | 57   | 553207         | GCF_000175375.1 | NZ_ACSH02000008.1 | Microbioma oral humano   |
| <i>Corynebacterium amycolatum SK46</i>           | PriA          | 2.514                     | 2194  | 2112                        | 58.6 | 553204         | GCF_000173655.1 | NZ_ABZU01000048   | Microbioma vaginal humano (potencial probiótico)   |
| <b>Noda-García et al. 2015</b>                   |               |                           |       |                             |      |                |                 |                   |  |
| <i>Acidimicrobium ferrooxidans DSM 10331</i>     | HisA          | 2.158                     | 2097  | 2020                        | 68.3 | 525909         | GCF_000023265.1 | NC_013124         | Escorrentía de aguas termales en Islandia.   |
| <i>Bifidobacterium longum NCC2705</i>            | PriA          | 2.26                      | 1897  | 1781                        | 60   | 206672         | GCF_000007525.1 | NC_004307.2       | Heces humanas adulto ( Regula el microbioma intestinal)                                      |
| <b>Verduzco-Castro et al. 2016</b>               |               |                           |       |                             |      |                |                 |                   |  |
| <i>Mycolicibacterium smegmatis MC2 155</i>       | PriA          | 6.988                     | 6708  | 6528                        | 67.4 | 246196         | GCF_000015005.1 | NC_008596         | Smegma humano (genital)  |
| <i>Paenarthrobacter aurescens TC1</i>            | PriB          | 5.227                     | 4,828 | 4678                        | 62.4 | 290340         | GCF_000014925.1 | NC_008711         | Suelo en sitio de derrame de Dakota del Sur que contiene el herbicida atrazina               |
| <i>Streptomyces globisporus C-1027</i>           | PriA (TrpF-2) | 7.784                     | 7004  | 6740                        | 71.6 | 1172567        | GCF_000261345.2 | NZ_CP013738.1     | Produce C-1027, un potente antibiótico antitumoral   |
| <i>Streptomyces sp. Mg1</i>                      | PriB          | 8.716                     | 7822  | 7408                        | 72   | 465541         | GCF_000412265.2 | NZ_CP011664.1     | Suelo  |
| <i>Streptomyces sviveus ATCC 29083</i>           | PriB          | 9.313                     | 8459  | 7616                        | 70.5 | 463191         | GCF_000154965.1 | NZ_CM000951.1     | Antifúngico, antitumoral   |
| <i>Streptomyces ipomoeae 91-03</i>               | PriB          | 10.4                      | 9243  | 8556                        | 70.2 | 698759         | GCF_000317595.1 | NZ_AEJC01000001.1 | Agente del suelo causante de la pudrición de la batata (camote)                              |
| <i>Streptomyces sp. C</i>                        | PriB          | 8.46                      |       | 7672                        | 72.6 | 253839         | GCF_000158895.1 | NZ_GG657753       | Suelo. Producción de antibióticos  |
| <b>Juarez-Vázquez et al. 2017</b>                |               |                           |       |                             |      |                |                 |                   |  |
| <i>Bifidobacterium gallicum DSM 20093</i>        | PriA          | 2                         | 1566  | 1465                        | 57.5 | 561180         | GCF_000741205.1 | NZ_JGYW01000001.1 | Heces humanas (Regula el microbioma intestinal)  |
| <i>Bifidobacterium adolescentis L2 32</i>        | PriA          | 2.4                       | 2075  | 1963                        | 59   | 411481         | GCF_000154085.1 | NZ_DS264459       | Heces humanas (Regula el microbioma intestinal)  |
| <i>Actinomyces urogenitalis DSM 15434</i>        | PriA          | 2.7                       | 2344  | 2161                        | 68.5 | 525246         | GCF_000159035.1 | NZ_DS999640       | Tracto genitourinario humano   |
| <i>Actinomyces sp oral taxon 171 str F0337</i>   | PriA          | 3.097                     | 2628  | 2521                        | 68   | 706439         | GCF_000186965.1 | NZ_GL637675       | Cavidad oral humana  |
| <i>Actinomyces oris MG-1</i>                     | PriA          | 2.9                       | 2744  | 2389                        | 68   | 544580         | GCF_001682715.1 | NZ_Maub01000107   | Cavidad oral humana  |
| <i>Schaalia cardiffensis F0333</i>               | SubHisA2      | 2.231                     | 1831  | 1735                        | 61   | 888050         | GCF_005886435.1 | NZ_CP040505.1     | Cavidad oral humana  |
| <i>Schaalia odontolytica ATCC 17982</i>          | SubTrpF       | 2.394                     | 2071  | 1996                        | 65   | 411466         | GCF_000154225.1 | NZ_DS264586       | Intestino humano   |
| <i>Schaalia vaccimaxillae</i>                    | SubHisA2      | 2.338                     | 2073  | 1995                        | 57.5 | 183916         | GCF_000420425.1 | NZ_ATUX01000001   | Lesión en mandíbula de vacuno adulto   |
| <i>Actinomyces sp. oral taxon 848 str. F0332</i> | SubTrpF       | 2.52                      |       | 2063                        | 64.4 |                | GCF_000162895.1 | NZ_GG703882       | Cavidad oral humana  |
| <i>Actinomyces sp. oral taxon 180 str. F0310</i> | SubTrpF       | 2.373                     | 2063  | 1975                        | 66.7 | 888052         | GCF_000185285.1 | NZ_GL622194       | Cavidad oral humana  |
| <i>Actinomyces graevenitzi F0530</i>             | SubTrpF       | 2.108                     | 1797  | 1695                        | 57.5 | 1321817        | GCF_000466185.1 | NZ_KE951478       | Espudo expectorado de humano, cavidad oral humana  |

**Gráfico I.a.** Resumen de todas las características únicas que se encuentran en un objeto del genoma de *Cgl* ATCC 13032, agrupadas a modo comparativo con los mismos datos de la **Fig. 7.1.1**. Genes, Términos únicos (TERMS), Reacciones ModelSEED únicas (MSRXNS) y Pares únicos de gen/ Reacción ModelSEED (GENE\_MSXXNS). Anidada en cada característica se encuentran las nueve comparaciones para cada fuente de anotación funcional y criterio de fusión: RAST (annotate\_genome), Prokka (ProkkaAnnotation), Blast Koala, NetGo 3.0 y DeepEC, además de los criterios de fusión Unión, Regla de Mayor (Majority rule), Dos o más (2 or more) y Prioridad (Priority), las descripciones se truncan a los primeros 20 caracteres. Visualización de alta calidad obtenida con Python Bokeh.



## 12.2. Anexo II

Captura de pantalla mostrando cómo se visualiza la Narrativa del entorno Kbase.

The screenshot displays the Kbase web interface. At the top, the KBase logo is on the left, and the title 'Promiscuos PriA' is centered, with 'Created by: Naxhie López (naxhie)' below it. On the right, there are icons for help, kernel, share, and save, along with a user profile icon.

The main content area is divided into two sections. The top section, titled 'Narratives', shows a narrative card with a 'T' icon. The text of the narrative reads: 'This narrative was created for the creation of the metabolic models for the homologous and analogs of the bifunctional enzyme PriA involve in the biosynthesis of tryptophan and histidine of microbes of the genus Actinobacteria.'

The bottom section, titled 'Compare Metabolic Annotations', shows a job status of 'Finished with success at Jan 25, 2023 at 8:34pm'. It includes buttons for 'Reset', 'View Configure', 'Info', 'Job Status', and 'Result'. Below this, there are sections for 'Objects' (No objects created) and 'Report' (View report in separate window).

On the left side, there is a sidebar with 'DATA' and 'APPS' sections. The 'DATA' section lists three items: 'Cglut\_Majority\_2OrMore\_Priorit... v1' (Genome: Corynebacterium glutamicum ATCC 13032, Jan 31, 2023), 'Cjeike\_Majority\_2OrMore\_Priorit... v1' (Genome: Corynebacterium jeikeium K411, Jan 25, 2023), and 'Auroaeni Prioritv v1'. The 'APPS' section lists various categories with counts: Comparative Genomics (35), Expression (33), Genome Annotation (25), Genome Assembly (27), Host (1), and Metabolic Modeling (25).

## 12.3. Anexo III

### *Scripts* para COBRA Toolbox 3.0 en MATLAB® 2022

**Script I.** Restricciones, y análisis de robustez para el Modelo iAA1259 del organismo *Sco* A3(2).

```
initCobraToolbox
changeCobraSolver('gurobi','all', 1)
%
clc
clear
model= readCbModel('Sco_2018.xml')
solution_mat=optimizeCbModel(model)
printFluxVector (model,[solution_mat.v],1, 1);
%Restricción del modelo
%Tasa de consumo experimental de glucosa
model_1=changeRxnBounds(model,'EX_glc(e)', -0.8,'b');
solution_1=optimizeCbModel(model_1)
printFluxVector (model_1, [solution_1.v],1, 1);
% Tasa de consumo experimental de O2
model_2=changeRxnBounds(model_1,'EX_o2(e)', -2.5,'b');
solution_2=optimizeCbModel(model_2)
printFluxVector (model_2, [solution_2.v],1, 1);
% Tasa de producción experimental CO2
model_3=changeRxnBounds(model_2,'EX_co2(e)', 2.4,'b');
solution_3=optimizeCbModel(model_3)
printFluxVector (model_3, [solution_3.v],1, 1);
% Cerrar doble flujo de L-isoleucina
model_4=changeRxnBounds(model_3,'ILEDxi', 0,'b');
solution_4=optimizeCbModel(model_4)
printFluxVector (model_4, [solution_4.v],1, 1);
%Análisis simple de robustez PRMICii vs PRAIi (max)
[controlFlux, objFlux] = robustnessAnalysis(model_3, 'PRMICii', 50, 1, 'PRAIi', 'max')
```

**Script II.** Restricciones, y análisis de robustez para el Modelo iEK1011 del organismo *Mtu* H37Rv.

```
initCobraToolbox
changeCobraSolver('gurobi','all', 1)
%%
clc
clear
model= readCbModel('Mtuber_2018.xml')
solution_mat=optimizeCbModel(model)
printFluxVector (model,[solution_mat.v],1, 1);
%Restrictions for Physiological Media (in vivo modeling)
```

```

model_1=changeRxnBounds(model,'EX_ala_l_e', -1,'l');
model_1=changeRxnBounds(model_1,'EX_chsterol_e', -10,'l');
model_1=changeRxnBounds(model_1,'EX_cobalt2_e', -1000,'l');
model_1=changeRxnBounds(model_1,'EX_fe3_e', -5,'l');
model_1=changeRxnBounds(model_1,'EX_h_e', -1000,'l');
model_1=changeRxnBounds(model_1,'EX_hdca_e', -10,'l');
model_1=changeRxnBounds(model_1,'EX_no3_e', -1.5,'l');
model_1=changeRxnBounds(model_1,'EX_nodcoa_e', -10,'l');
model_1=changeRxnBounds(model_1,'EX_o2_e', -0.5,'l');
model_1=changeRxnBounds(model_1,'EX_ocdca_e', -10,'l');
model_1=changeRxnBounds(model_1,'EX_octscoa_e', -10,'l');
model_1=changeRxnBounds(model_1,'EX_ppa_e', -20,'l');
model_1=changeRxnBounds(model_1,'EX_ttdca_e', -10,'l');
model_1=changeRxnBounds(model_1,'EX_urea_e', -0.1,'l');
model_1=changeRxnBounds(model_1,'VPAMTr', 0,'b');
solution_1=optimizeCbModel(model_1)
printFluxVector(model_1, [solution_1.v],1, 1);
%Análisis simple de robustez PRAIi
[controlFlux, objFlux] = robustnessAnalysis(model_1,'PRMICI' , 50, 1, 'PRAIi', 'max')

```

**Script III.** Restricciones, y análisis de robustez para el Modelo iCGB21FR del organismo *Cgl* ATCC 13032.

```

initCobraToolbox
changeCobraSolver('gurobi','all', 1)
%
clc
clear
model= readCbModel('Cglut_2021.xml')
solution_mat=optimizeCbModel(model)
printFluxVector (model,[solution_mat.v],1, 1);
%Restriccion del Modelo D=0.3
model_1=changeRxnBounds(model,'EX_glc_D_e', -3.38595439507982,'b');
solution_1=optimizeCbModel(model_1)
printFluxVector (model_1, [solution_1.v],1, 1);
model_2=changeRxnBounds(model_1,'EX_o2_e', -7.02,'b');
solution_2=optimizeCbModel(model_2)
printFluxVector (model_2, [solution_2.v],1, 1);
model_3=changeRxnBounds(model_2,'EX_co2_e', 7.48,'b');
solution_3=optimizeCbModel(model_3)
printFluxVector (model_3, [solution_3.v],1, 1);
model_4=changeRxnBounds(model_3,'PGM', 0,'b');
solution_4=optimizeCbModel(model_4)
printFluxVector (model_4, [solution_4.v],1, 1);
model_5=changeRxnBounds(model_4,'HSTPTr', 0,'b');
solution_5=optimizeCbModel(model_5)
printFluxVector (model_5, [solution_5.v],1, 1);
%Análisis simple de robustez PRAIi(max) vs PRMICI
[controlFlux, objFlux] = robustnessAnalysis(model_5, 'PRMICI', 50, 1, 'PRAIi', 'max')

```

## 12.4. Anexo IV

### Restricción de modelos con flujos experimentales

**Tabla IV.a.** Valores de flujo propuestos utilizados para restringir el GSMM iAA1259 del organismo *Sco* A3(2), basado en estudios experimentales reportados por Melzoch & cols. (1997) para  $\mu = 0.06 \text{ h}^{-1}$ . Los flujos se expresan en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ .

| Reacción<br>(Restricción del modelo) | Flujo experimental<br>( $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ ) |
|--------------------------------------|---|
| Glucosa <i>consumo</i>               | 0.8   |
| O <sub>2</sub> <i>consumo</i>        | 2.5   |
| CO <sub>2</sub> <i>producción</i>    | 2.4   |

**Tabla IV.b.** Print de COBRA Toolbox 3.0 en MATLAB® 2022, una vez restringido el modelo, se aprecian los flujos experimentales y valor de tasa de crecimiento.

```

Biomass                0.06014
DM_4CRSOL              0.0002893
DM_5DRIB               0.004485
DM_AMOB               0.0005021
DM_DALUA              7.734e-05
EX_5mtr(e)            0.002008
EX_ca2(e)             -0.0002978
EX_cl(e)              -0.0002978
EX_co2(e)             2.4
EX_cobalt2(e)         -7.878e-05
EX_cu2(e)             -4.053e-05
EX_fe2(e)             -0.004268
EX_glc(e)             -0.8
EX_h(e)               7.21
EX_h2o(e)             3.08
EX_h2o2(e)            0.6924
EX_h2s(e)             0.07241
EX_k(e)               -0.01117
EX_mg2(e)             -0.0004963
EX_mn2(e)             -3.957e-05
EX_mobd(e)            -0.0002188
EX_nh4(e)             -0.5075
EX_ni2(e)             -1.846e-05
EX_o2(e)              -2.5
EX_pi(e)              -0.04491
EX_so4(e)             -0.09348
EX_zn2(e)             -1.949e-05

```

**Tabla IV.c.** Valores de flujo propuestos utilizados para restringir el GSMM iEK1011 del organismo *Mtu* H37Rv basado en el reporte de Kavvas & cols. (2018), del crecimiento de la actinobacteria en medio fisiológico (i.e. modelado *in vivo*). Los flujos son los límites inferiores y se expresan en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . Con restricción de L-alanina como única fuente de nitrógeno con base en reportes de Shi (2021) a pesar de que la restricción original de Kavvas & cols. (2018) contenía además consumo de L-glutamina, L-glutamato, L-aspartato y L-asparagina.

| Reacción (Restricción del modelo)                  | Límite de flujo inferior<br>( $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ ) |
|--|---|
| L-Alanina <i>consumo</i>                           | 1   |
| O <sub>2</sub> <i>consumo</i>                      | 0.5   |
| Colesterol <i>consumo</i>                          | 10  |
| Cobalto II <i>consumo</i>                          | 1000  |
| Hierro III <i>consumo</i>                          | 5   |
| Hexadecanoato <i>consumo</i>                       | 10  |
| H <sup>+</sup> <i>consumo</i>                      | 1000  |
| Nitrato <i>consumo</i>                             | 1.5   |
| Nonadecano CoA <i>consumo</i>                      | 10  |
| Octadecanoato <i>consumo</i>                       | 10  |
| Octacosanoil CoA <i>consumo</i>                    | 10  |
| Propionato <i>consumo</i>                          | 20  |
| Tetradecanoate <i>consumo</i>                      | 10  |
| Urea <i>consumo</i>                                | 0.1   |
| VPAMTr ( <i>Valina-piruvato aminotransferasa</i> ) | 0 ( <i>flujo cerrado</i> )  |

**Tabla IV.d.** Print en COBRA Toolbox 3.0 en MATLAB® 2022, una vez restringido el modelo, se aprecian los flujos dentro del límite inferior fijado en las restricciones.

|              |            |
|--------------|------------|
| EX_nh4_e     | 3.453      |
| EX_no3_e     | -1.5       |
| EX_o2_e      | -0.5       |
| EX_pi_e      | -0.03244   |
| EX_ala_L_e   | -1         |
| EX_asn_L_e   | -1         |
| EX_so4_e     | -0.002896  |
| EX_succ_e    | 1.602      |
| EX_cit_e     | -1         |
| EX_co2_e     | 4.396      |
| EX_val_L_e   | 1.075      |
| EX_coa_e     | 0.02466    |
| EX_ac_e      | 0.003065   |
| EX_fe3_e     | -0.0001428 |
| EX_urea_e    | -0.1       |
| EX_glyc_e    | -1         |
| EX_h2_e      | 0.002705   |
| EX_h2o_e     | 3.346      |
| EX_h_e       | -7.76      |
| BIOMASS_2    | 0.0255     |
| EX_nodcoa_e  | -0.006425  |
| EX_4hba_e    | 2.04e-05   |
| EX_octscoa_e | -0.01873   |
| EX_acald_e   | 0.6876     |
| EX_etoh_e    | -1         |

**Tabla IV.e.** Valores de flujo propuestos utilizados para restringir el GSMM iCGB21FR del organismo *Cgl* ATCC 13032, basados en estudios experimentales reportados por Graf & cols. (2020) para  $\mu = 0.3 \text{ h}^{-1}$ . Los flujos se expresan en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ .

| Reacción (Restricción del modelo)        | Flujo experimental<br>( $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ ) |
|--|---|
| Glucosa <i>consumo</i>                   | 3.3859  |
| O <sub>2</sub> <i>consumo</i>            | 7.02  |
| CO <sub>2</sub> <i>consumo</i>           | 7.48  |
| HSTPTr (Histidinol fosfato transaminasa) | 0 ( <i>flujo cerrado</i> )  |
| PGM (Fosfoglicerato mutasa)              | 0 ( <i>flujo cerrado</i> )  |

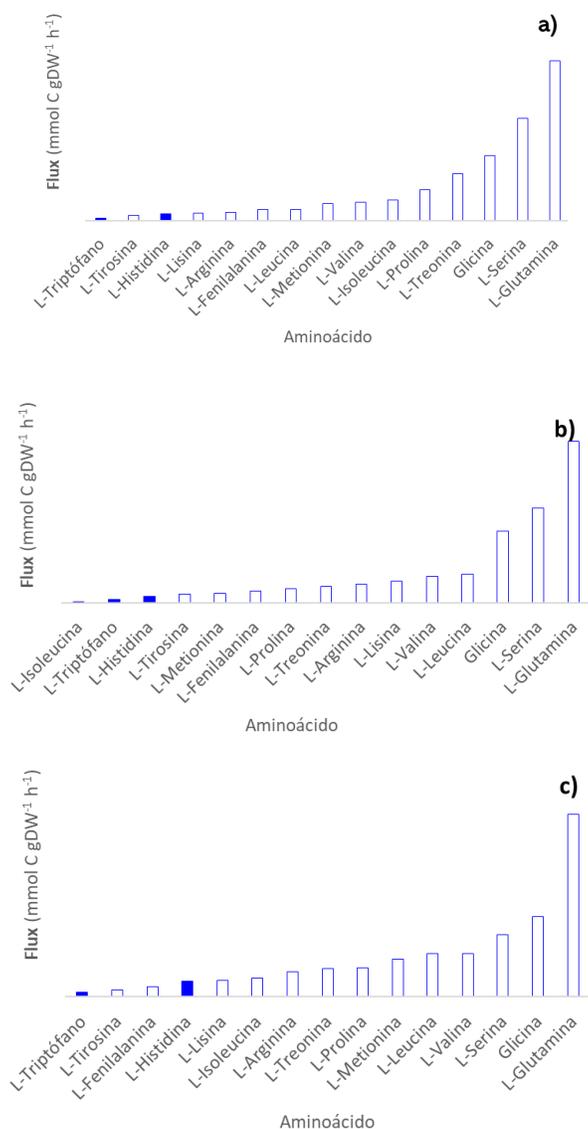
**Tabla IV.f.** Print en COBRA Toolbox 3.0 en MATLAB® 2022, una vez restringido el modelo iCGB21FR del organismo *Cgl* ATCC 13032, se aprecian los flujos medidos experimentalmente y valor de tasa de crecimiento.

|             |           |
|-------------|-----------|
| EX_glc__D_e | -3.386    |
| EX_h2o_e    | 9.685     |
| EX_h_e      | -10       |
| EX_cl_e     | -0.01872  |
| EX_pi_e     | -0.2554   |
| EX_nh4_e    | -3.276    |
| EX_gly_e    | 0.02601   |
| EX_fe3_e    | -0.01189  |
| EX_k_e      | -0.01698  |
| EX_ca2_e    | -0.01656  |
| EX_mg2_e    | -0.02731  |
| EX_mn2_e    | -0.01208  |
| EX_zn2_e    | -0.01015  |
| EX_succ_e   | 1.272     |
| EX_co2_e    | 7.48      |
| EX_cu2_e    | -0.01045  |
| EX_o2_e     | -7.02     |
| EX_fe2_e    | -0.01189  |
| EX_glyc1t_e | 0.7831    |
| EX_lac__L_e | 2.178     |
| EX_so4_e    | -0.08044  |
| SK_4hba_c   | 6.594e-05 |
| Growth      | 0.3053    |
| EX_na1_e    | -0.02887  |
| EX_ni2_e    | -0.01131  |
| EX_34hzbz_e | -1.803    |

## 12.5. Anexo V

### Distribución de flujos de biosíntesis de L-aminoácidos

**Gráfico V.a.** Desglose de flujos para cada uno de los dieciséis L-aminoácidos sintetizados para los organismos: **a) *Sco A (3)2***, **b) *Cgl ATCC 13032*** y **c) *Mtu H37Rv***; ordenados de menor a mayor; las barras rellenas corresponden a los L-aminoácidos **Trp** e **His**, las barras vacías a los demás 14 L-aminoácidos.



## 12.6. Anexo VI

**Figura. V.a. Mapa de distribución de flujos metabólicos para *Mtu* H37Rv.** Los nombres de cada reacción y metabolito están abreviados de acorde con la plataforma BIGG Models [King et al. 2016]. Al lado de cada reacción se encuentra el valor del flujo en  $\text{mmol C} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ . Se anota la zona correspondiente a cada vía metabólica central. El desvío de flujo hacia la síntesis de los aminoácidos aromáticos L-fenilalanina y L-tirosina, así como la vía de síntesis de His (en verde) y vía de síntesis de Trp (en azul-morado) se diferencian por el peso de la línea y el color. Se observan las reacciones involucradas en la enzima promiscua resaltadas con la etiqueta bicolor "PriA", mostrando la relación numérica de los flujos de ambas vías identificadas por color **1 Trp : 3.7 His**. El mapa se elaboró con la herramienta Escher [King et al. 2015].

