

Algunas consideraciones acerca de la depuración de modelos metabólicos a escala genómica
basado en algoritmos computacionales

Some considerations about debugging a genome-scale metabolic models based on computational
algorithms

Autores: Ing. Hendy Maier Pérez Barrera; Lic. Yaniel Jesús Barceló Fernández; Ing. Yarlenis
Pacheco Suarez

Centros de procedencia: Universidad de Ciencias Pedagógicas "Rafael María de Mendive";
Universidad "Hermanos Saíz Montes de Oca", de Pinar del Río

Email: hendy@ucp.pr.rimed.cu; yaniel@ucp.pr.rimed.cu; yarle@upr.edu.cu

Resumen:

El desarrollo de la biología molecular ha permitido rápidos avances en la comprensión de las funciones y principios de funcionamiento de las células y organismos unicelulares a nivel molecular. Uno de los nuevos campos de investigación que surge de este panorama es la Biología de Sistemas, teniendo como una de sus piedras angulares la reconstrucción de los modelos metabólicos a escala genómica. En la actualidad, esta reconstrucción es un tema complejo, debido, fundamentalmente, a que no está automatizado. Consecuentemente, el presente trabajo pretende realizar algunas consideraciones acerca de la depuración de modelos metabólicos a escala genómica basado en algoritmos computacionales, intención que posee un alto valor metodológico, al aportar un posible cómo a una problemática social de gran envergadura.

Palabras Claves: Biología de Sistemas, reconstrucción de modelos metabólicos

Abstract:

The development of molecular biology has allowed rapid progress in understanding the functions and principles of operation of cells and unicellular organisms at the molecular level. One of the new areas of research that emerges from this picture is Systems Biology, having as one of its reconstruction of genome-scale metabolic models cornerstones. Currently, this reconstruction is a complex issue, because basically it is not automated. Consequently, this paper attempts to make some considerations about debugging a genomic scale metabolic models based on computational algorithms, an intention that has a high methodological value, how to make possible a major social problem.

Keywords: Systems Biology, reconstruction of metabolic models

Acerca de la reconstrucción de modelos metabólicos a escala genómica.

Los avances en la biología molecular y en la infraestructura disponible para el desarrollo de investigaciones en este campo, han dado lugar a una creciente acumulación de datos biológicos. La información resultante, expuesta en grandes repositorios, incluyen ramas que van desde la genómica, la transcriptómica, la proteómica, hasta la interactómica, la metabolómica y la fluxómica (Benson et al., 1993). El creciente flujo de la información requiere el uso de técnicas innovadoras para su visualización, modelación, interpretación y análisis. Las ciencias de la computación, las ciencias de la información, la ingeniería, las matemáticas, la estadística, entre otras han sido aplicadas a la biología, emergiendo así la disciplina hoy conocida como bioinformática. En este escenario, la Biología de Sistema ha surgido como una prometedora ciencia que pretende entender cuantitativamente cómo las propiedades de los sistemas biológicos se pueden enfocar en función de las características y de las interacciones entre sus componentes macromoleculares (Snoep et al., 2006). En esta área, las vías metabólicas y sus capacidades constituyen los objetos de estudios prioritarios (Papin et al., 2003).

La reconstrucción de los modelos metabólicos a escala genómica es un proceso que tiene como objetivo permitir la simulación del metabolismo celular del organismo en estudio. Este procedimiento permite la integración de información genómica con actividades metabólicas observadas a través de experimentos fenotípicos y otros datos "ómicos" para obtener conocimiento biológico oculto y que pudiera ser de otro modo difícil de obtener (Fang et al., 2011). La proximidad de las simulaciones al comportamiento in vivo, dependerá de la calidad de los datos biológicos así como de la calidad de estos modelos a gran escala. Muchos trabajos referenciales nos pueden instruir y dirigir en esta labor (Förster et al., 2003; Notebaart et al., 2006; Feist et al., 2009; Thiele and Palsson, 2010; Triana et al., 2012).

Este proceso, no automatizado e iterativo, presupone el trabajo a largo plazo de un especialista utilizando la información contenida en diversas bases de datos biológicas, con el objetivo de organizar la lista de reacciones metabólicas específicas para un organismo (Förster et al., 2003). De acuerdo con varios protocolos descritos en la literatura se requieren de tres a cuatro pasos fundamentales para la reconstrucción a escala genómica de modelos metabólicos (Thiele and Palsson, 2010; Triana et al., 2013). Este proceso pasa por generar un borrador del modelo metabólico del organismo en cuestión, seguido de un refinamiento exhaustivo (desde el principio o partir de ese paso se debe tener en cuenta el formato computacional para la escritura del modelo), así como de una validación del mismo a través de las capacidades de la red metabólica. La curación iterativa del modelo que incluye el análisis de huecos en la red así como el ajuste metabólico para su llenado, la eliminación de ciclos fútiles o ciclos de sustratos (Voet and Voet, 2004) y las vías extremas tipo III (Schilling et al., 2000), entre otras, son parte crucial en el proceso de depurado. Se han reportado varios métodos que inciden en el refinamiento de la red (Osterman and Overbeek, 2003; Kharchenko et al., 2004; Kharchenko et al., 2006; Chen and Vitkup, 2006; Green and Karp, 2004; Kumar et al., 2007). El primer método publicado fue SMILEY, que basado en la programación lineal, identifica el mínimo de reacciones que necesitan ser agregadas al modelo metabólico desde una base de datos universal de reacciones, para permitir un acercamiento en el crecimiento que debe ser logrado (Reed et al., 2003). Kumar y colaboradores, proponen dos métodos, GapFind, para identificar los metabolitos no producidos y GapFill, para resolver los huecos con el mínimo de modificaciones en el modelo metabólico (Kumar et al., 2007). Karp y colaboradores, desarrollaron un método, llamado MetaFlux, que basándose en la programación MILP (del inglés "Mixed Integer Linear Programming") corrige el conjunto de reacciones, metabolitos relacionados con la ecuación de biomasa, nutrientes y procesos de secreción celular. El método genera los modelos metabólicos directamente de la base de datos de vías metabólicas y genomas gestionada por el Pathways Tools, paquete informático desarrollado por este equipo (Latendresse et al., 2012).

Sin embargo, no existen algoritmos soportados sobre criterios estadísticos y ortológicos (estudio de homología de secuencias genéticas) que permitan lograr una completitud para dichos modelos. Esta razón hace aún más engorroso el proceso de depuración si se tiene en cuenta que conlleva realizar un estudio de la reversibilidad de las reacciones, metabolitos desconectados y reacciones bloqueadas asociadas a los mismos, así como la inclusión de reacciones que fueron estudiadas para otros organismos y que se puede inferir, mediante enfoques ortológicos, su presencia también para los organismos de estudio. Este desafío incluye, por tanto, la posibilidad de identificar, rellenar o eliminar aquellos metabolitos que no pueden ser producidos por ninguna reacción o importados a través de las vías metabólicas en el modelo y los metabolitos que no son consumidos por ninguna reacción en la red o que no son exportados hacia ninguna ruta existente, así como la eliminación de ciclos internos que no son termodinámicamente factibles, a saber, los ciclos fútiles.

Particularidades de la Biología de sistemas.

La Biología Molecular de las últimas décadas se ha basado en la teoría que asume el camino directo existente entre genes, proteínas y función biológica, así como la presencia de respuestas predeterminadas del sistema a perturbaciones externas. Aunque este tipo de investigación ha dado lugar a gran cantidad de conocimiento, no proporciona información acerca de cómo integran las células estos datos de forma que se genere un tipo de respuesta u otro. A pesar de que la Biología de Sistemas se considera una nueva disciplina, el abordaje del estudio de los procesos biológicos como sistemas se trató por Wiener en 1948 en lo que se llamó en aquel momento la cibernética. La Biología de Sistemas ha sido descrita por investigadores como Leroy Hood con detalle, aunque el término ya se empleó por primera vez en 1968 por teóricos como Mesarovic (Mesarovic, 1968)

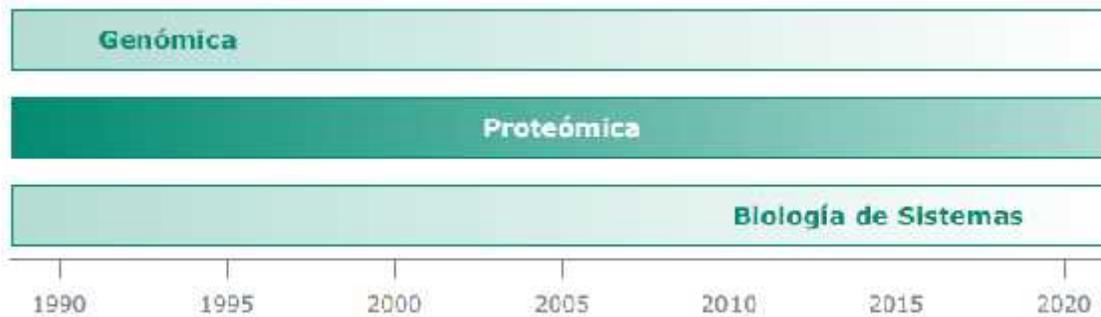


Figura 1 Evolución de la Biología de Sistemas. (Url2)

La Biología de sistemas es el campo de investigación interdisciplinario de los procesos biológicos en el que las interacciones de los elementos, internos y externos, que influyen en el desarrollo del proceso se representan mediante un sistema matemático. Este enfoque global permite comprender integrativamente el funcionamiento de los sistemas biológicos y profundizar en el entendimiento de cómo sus interacciones internas y externas (con otros sistemas) conlleva a la aparición de nuevas propiedades y/o procesos.

Convencionalmente, en el estudio de los procesos biológicos se utiliza el método científico clásico, que se basa en la confirmación o refutación de una hipótesis al confrontarle con los resultados experimentales. La Biología sistémica utiliza, sin embargo, un enfoque distinto, basado en la modelización matemática de los procesos en estudio. Como resultado de la simulación, al poner a funcionar los modelos matemáticos con los que se representa el proceso, se obtiene una serie de predicciones del estado de dicho proceso biológico que corresponderían a los resultados experimentales esperados. Durante las simulaciones, la red de interacciones entre los elementos que componen el proceso biológico se representa mediante un sistema de ecuaciones diferenciales. Los valores de las características de dichos elementos a distintos tiempos y bajo diversas condiciones experimentales (simuladas) son predecibles debido a que la dinámica (es decir, los cambios) del estado de ese sistema modelado es calculable matemáticamente. La Biología de sistemas es un área interdisciplinaria en la que participan biólogos, químicos, matemáticos y físicos, entre otros.

Esta disciplina emplea una estrategia diferente a las aproximaciones empíricas tradicionales, por medio del estudio de sistemas biológicos en sus diferentes niveles, desde células y redes celulares a organismos completos. La Biología de Sistemas implica el mapeo de rutas, interacción de proteínas y genes, así como el de los circuitos de organismos a nivel celular, y de organismo completo, todo ello integrado en un modelo informático.

Una de las piedras angulares de la biología de sistemas es la reconstrucción de los modelos metabólicos a escala genómica. Actualmente, esta reconstrucción es un tema complejo, debido fundamentalmente a que no está automatizado, por lo que el trabajo se desarrolla de forma manual por los Biólogos.

Reconstrucción de la red metabólica a escala genómica.

La reconstrucción de la red metabólica y la simulación permite una visión en profundidad de los mecanismos moleculares de un organismo en particular. En particular, estos modelos se correlacionan con el genoma fisiología molecular. Una reconstrucción rompe vías metabólicas en sus respectivas reacciones y enzimas, y los analiza en la perspectiva de toda la red. En términos simplificados, una reconstrucción recoge toda la información relevante metabólica de un organismo y lo compila en un modelo matemático. Validación y análisis de las reconstrucciones pueden permitir la identificación de las características clave del metabolismo tales como el rendimiento de crecimiento, distribución de recursos, la robustez de la red, y la esencialidad del gen. Este conocimiento se puede aplicar para crear novela la biotecnología.

En general, el proceso de construcción de una reconstrucción es la siguiente:

- ✓ Elaborar una reconstrucción
- ✓ Acotar el modelo
- ✓ Convertir modelo en una representación matemática/computación
- ✓ Evaluar y depurar modelo a través de la experimentación

Una reconstrucción metabólica proporciona una plataforma altamente matemática, estructurada para entender la biología de sistemas de vías metabólicas en el organismo. La integración de las vías metabólicas, bioquímicas con secuencias del genoma unannotated rápidamente disponibles ha desarrollado lo que se llama modelos genoma escala metabólica. En pocas palabras, estos modelos se corresponden con genes metabólicos vías metabólicas. En general, cuanto más información acerca de la fisiología, la bioquímica y la genética está disponible para el organismo diana, mejor será la capacidad predictiva de los modelos

reconstruidos. Mecánicamente hablando, el proceso de reconstrucción de redes metabólicas procariotas y eucariotas es esencialmente la misma. Dicho esto, reconstrucciones eucariotas son generalmente más difícil debido al tamaño de los genomas, la cobertura de los conocimientos, y la multitud de compartimentos celulares. El primer modelo de genoma escala metabólica fue generada en 1995 por *Haemophilus influenzae*. El primer organismo multicelular, *C. elegans*, fue reconstruido en 1998. Desde entonces, se han formado muchas reconstrucciones. Para obtener una lista de las reconstrucciones que se han convertido en un modelo y validada experimentalmente.

La elaboración de una reconstrucción

➤ Recursos

Debido a que la escala de tiempo para el desarrollo de reconstrucciones es tan reciente, la mayoría de las reconstrucciones se han construido de forma manual. Sin embargo, ahora, hay un buen número de recursos que permiten el montaje semiautomático de estas reconstrucciones que son utilizados por el tiempo y el esfuerzo necesarios para una reconstrucción. Una reconstrucción rápida inicial puede ser desarrollada utilizando los recursos de forma automática, en combinación con las enciclopedias como MetaCyc, y luego actualizar manualmente mediante el uso de recursos como PathwayTools. Estos métodos semiautomáticos permiten un borrador rápido que se creará al mismo tiempo a los ajustes de sintonía fina requeridos una vez que se encontraron nuevos datos experimentales. Es sólo de esta manera que el campo de las reconstrucciones metabólicas mantendrá al día con el siempre creciente número de genomas anotada.

➤ Bases de datos

- Enciclopedia de Kyoto de genes y genomas: una base de datos de bioinformática que contiene información sobre los genes, proteínas, reacciones, y las vías. La sección de Organismos KEGG, que se divide en eucariotas y procariotas, abarca muchos organismos para los cuales la información genética y el ADN se pueden buscar introduciendo el enzima de su elección.
- BioCyc, EcoCyc y MetaCyc: BioCyc Es una colección de 1.000 bases de datos vía/genoma, con cada base de datos dedicada a un solo organismo. Por ejemplo, EcoCyc es una base de datos de bioinformática muy detallada sobre el genoma y la reconstrucción metabólica de *Escherichia coli*, incluyendo descripciones completas de *E. coli* y las vías de señalización de la red reguladora. La base de datos EcoCyc puede servir como un paradigma y modelo para cualquier reconstrucción. Además, MetaCyc, una enciclopedia de las vías y enzimas metabólicas definidas experimentalmente, contiene 1.500 rutas metabólicas y 8.700 reacciones metabólicas.
- ENZIMA: Una base de datos de nomenclatura de las enzimas. Después de buscar una enzima en particular sobre la base de datos, este recurso le da la reacción que es catalizada. Enzima tiene enlaces directos a otras bases de datos de genes/enzima/literatura como KEGG, BRENDA y PUBMED.
- BRENDA: base de datos de enzimas integral que permite una enzima que se debe buscar por nombre, número CE u organismo.
- BiGG: Una base de conocimientos del genoma escala metabólica reconstrucciones red bioquímica, genética y genómicamente estructurados.
- metaTIGER: Es una colección de perfiles metabólicos e información phylogenomic en un taxonómicamente diversa gama de eucariotas, que proporciona nuevos medios para ver y comparar los perfiles metabólicos entre organismos.

➤ Herramientas para la integración de datos

- Herramientas Pathway: Un paquete de software de bioinformática que ayuda en la construcción de bases de datos vía/genoma, tales como EcoCyc. Desarrollado por Peter Karp y asociados en el Grupo Internacional de Bioinformática SRI, Pathway Herramientas comprende varias unidades separadas. En primer lugar, patológica toma un genoma anotado de un organismo e infiere vías metabólicas probables para producir una nueva base de datos vía/genoma.
- ERGO: Un servicio basado en suscripción desarrollado por Genómica Integrada. Se integra datos de todos los niveles, incluyendo los datos genómicos bioquímicos, la literatura y análisis de alto rendimiento en una red de fácil uso integral de las rutas metabólicas y nonmetabolic.
- Modelo SEED: Un recurso en línea para el análisis, la comparación, la reconstrucción, y la preservación de los modelos de genoma escala metabólica. Los usuarios pueden enviar secuencias

del genoma del sistema de anotación RAST, y la anotación resultante se puede canalizar de forma automática en la SEED modelo para producir un proyecto de modelo metabólico. La semilla Modelo construye automáticamente una red de reacciones metabólicas, asociaciones gen-proteína-reacción para cada reacción, y una reacción composición de biomasa para cada genoma para producir un modelo de metabolismo microbiano que puede ser simulado utilizando análisis de flujo de equilibrio.

- MetaMerge: algoritmo para la conciliación de semi-automáticamente un par de reconstrucciones de redes metabólicas existentes en un único modelo de red metabólica.

➤ Herramientas para la Literatura

- PUBMED: Se trata de una biblioteca en línea desarrollado por el Centro Nacional de Información sobre Biotecnología, que contiene una gran colección de revistas médicas. Usando el enlace proporcionado por la enzima, la búsqueda puede ser dirigida hacia el organismo de interés, recuperando así la literatura sobre la enzima y su uso en el interior del organismo.

Metodología para elaborar una reconstrucción

Una reconstrucción se edifica mediante la compilación de datos de los recursos anteriores. Herramientas de bases de datos tales como KEGG y BioCyc se pueden utilizar en combinación unos con otros para encontrar todos los genes metabólicos en el organismo de interés. Estos genes se compararán con los organismos estrechamente relacionados que ya han desarrollado reconstrucciones para encontrar los genes y las reacciones homólogas. Estos genes homólogos y reacciones se realizan a través de las reconstrucciones conocidos para formar el proyecto de reconstrucción del organismo de interés. Herramientas tales como ERGO pueden recopilar datos en vías que pueden ser utilizados entonces por el Modelo de SEED para formar una red de vías metabólicas y no metabólicas. Estas redes son luego verificados y refinados antes de convertirse en una simulación matemática.

El aspecto predictivo de metabólicos bisagras de reconstrucción en la capacidad de predecir la reacción bioquímica catalizada por una proteína con la secuencia de aminoácidos de esa proteína como una entrada, y para deducir la estructura de una red metabólica basa en el predicho conjunto de reacciones. Una red de enzimas y metabolitos se redactó relacionar secuencias y funciones. Cuando una proteína no caracterizada se encuentra en el genoma, su secuencia de aminoácidos se compara primero a las de las proteínas previamente caracterizadas para la búsqueda de homología. Cuando se encuentra una proteína homóloga, se considera que las proteínas que tienen un ancestro común y sus funciones se infiere como algo similar. Sin embargo, la calidad de un modelo de reconstrucción depende de su capacidad para inferir con precisión fenotipo directamente de la secuencia, por lo que esta estimación aproximada de la función de la proteína no será suficiente. Un número de algoritmos y recursos bioinformática se han desarrollado para el refinamiento de las asignaciones basadas en homología de secuencia de funciones de la proteína:

- InParanoid: Identifica eucariotas orthologs por mirar sólo en-paralogs.
- CDD: Recursos para la anotación de las unidades funcionales en las proteínas. Su colección de modelos de dominio utiliza la estructura 3D para proporcionar información sobre las relaciones de secuencia/estructura/función.
- InterPro: Proporciona un análisis funcional de las proteínas mediante la clasificación en familias y la predicción de dominios y sitios importantes.
- STRING: Base de datos de interacciones de proteínas conocidas y previstas.

Una vez que las proteínas se han establecido, más información acerca de la estructura de la enzima, las reacciones catalizadas, sustratos y productos, mecanismos, y mucho más se puede adquirir a partir de bases de datos como KEGG y NC-IUBMB. Metabólica reconstrucciones precisas requieren información adicional acerca de la reversibilidad y la dirección fisiológica preferida de una reacción catalizada por la enzima que puede provenir de la base de datos BRENDA.

Modelo Refinamiento

Una reconstrucción metabólica inicial de un genoma es típicamente lejos de ser perfecto debido a la alta variabilidad y la diversidad de microorganismos. A menudo, bases de datos vía metabólicas tales como KEGG y MetaCyc tendrán "agujeros" dentro de sus reacciones metabólicas, lo que significa que hay una conversión

de un sustrato en un producto que no tiene una enzima para catalizar identificado que la conversión. Lo que también puede suceder en las reconstrucciones elaboradas semi-automática es que algunos caminos son falsamente predecirse y no en realidad no ocurren de la manera prevista. Debido a esto, una verificación sistemática se hizo con el fin de asegurarse de que no hay contradicciones están presentes y que todas las indicaciones que figuran son correctos y exactos. Por otra parte, la literatura previa se puede investigar con el fin de apoyar a cualquier información obtenida de una de las muchas reacción metabólica y bases de datos del genoma. Esto proporciona un nivel adicional de garantía para la reconstrucción que la enzima y la reacción que cataliza ocurren en realidad en el organismo.

Las nuevas reacciones que no están presentes en las bases de datos necesitan ser añadido a la reconstrucción. Este es un proceso iterativo que los ciclos entre la fase experimental y la fase de codificación. Como nueva información se encuentra sobre el organismo diana, el modelo se puede ajustar para predecir la salida metabólica y fenotípica de la célula. La presencia o ausencia de ciertas reacciones del metabolismo afectarán a la cantidad de sustancias reaccionantes/productos que están presentes para otras reacciones dentro de la vía en particular. Esto es porque los productos en una reacción pasan a convertirse en los reactivos para la otra reacción, es decir, productos de una reacción pueden combinar con otras proteínas o compuestos para formar nuevas proteínas/compuestos en presencia de diferentes enzimas o catalizadores. Por lo tanto, la verificación sistemática de la reconstrucción inicial sacará a la luz varias incoherencias que pueden afectar negativamente a la interpretación final de la reconstrucción, que es comprender con precisión los mecanismos moleculares del organismo. Por otra parte, el paso de simulación también se asegura de que todas las reacciones presentes en la reconstrucción están adecuadamente equilibrados. En resumen, una reconstrucción que es totalmente exacta puede conducir a un mayor conocimiento acerca de la comprensión del funcionamiento del organismo de interés.

Recientes avances en la reconstrucción de los modelos metabólicos.

Una red metabólica captura la inter-conversión de metabolitos a través de transformaciones químicas catalizadas por enzimas. Para este fin, un modelo metabólico describe la estequiometría de la reacción y su direccionalidad, el gen para la proteína asociadas a las reacciones (GPRS), reacción-orgánulo especificando su localización, información transportador / reacción de intercambio, transcripción / traducción reglamento y composición de la biomasa [Kumar A, Suthers PF, Maranas CD: MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinform 2012, 13:6.]. Al definir el espacio metabólica, un modelo metabólico puede evaluar fenotipo celular permisible bajo condiciones ambientales y / o genéticos específicos, según Saha R (2011).

El número de modelos metabólicos desarrollado en los últimos años es un testimonio de su creciente utilidad y la penetración en muchas áreas de la biotecnología y la biomedicina. Inicialmente, los modelos metabólicos se han utilizado para caracterizar los sistemas biológicos y desarrollar estrategias no intuitivas para una reingeniería de ellos para aumento de la producción de bioproductos valiosos, según Ranganathan S (2010). Más recientemente, los modelos se han desarrollado y aplicado por una variedad de objetivos, por ejemplo: drogas, enfermedades metabólicas, la identificación de objetivos, el estudio de la patogenicidad microbiana y el parasitismo.

La validación de alta calidad de los modelos, es fundamental no sólo para una recapitulación de las propiedades fisiológicas conocidas sino también mejorar su precisión de la predicción. Hacia este fin, se han desarrollado estrategias para incorporar otros procesos celulares, tales como genes, expresión de proteínas para comprender mejor la aparición de complejo celular fenotipos. Por ejemplo, los modelos a escala genómica de patógenos se han reconstruido para desarrollar nuevos fármacos en favor de combatir infecciones y también minimizar los efectos secundarios en la acogida. Un enfoque integrado del modelo de E. coli ha sido desarrollado mediante la combinación del Metabolismo con la expresión de genes (es decir, ME modelo) aumentar el alcance y la precisión de la modelo-computable fenotipos correspondientes a la acción crecimiento óptimo. Además, mediante la combinación de todos los componentes moleculares, así como sus interacciones, un conjunto de células modelo ha sido desarrollado para Mycoplasma genitalium, un patógeno humano, para estudiar previamente inexplorada comportamientos celulares incluyendo la asociación de proteínas de ADN y la correlación entre la iniciación de la replicación del ADN y la replicación en sí.

Los Modelos metabólicos pueden facilitar enormemente la evaluación de los posibles fenotipos metabólicos alcanzables por estos organismos. Por lo tanto, el rápido desarrollo de alta calidad metabólica, modelos y algoritmos para el análisis de su contenido son de importancia crítica. La generación automatizada de los modelos metabólicos a escala genómica, mejoras y aplicaciones aun es un proceso no automatizado que requiere el uso de técnicas innovadoras para su desarrollo.

Enfoques de la reconstrucción de modelo metabólicos

El proceso de reconstrucción de la red o del modelo metabólico sigue tres pasos principales (como se destaca en la figura 1).

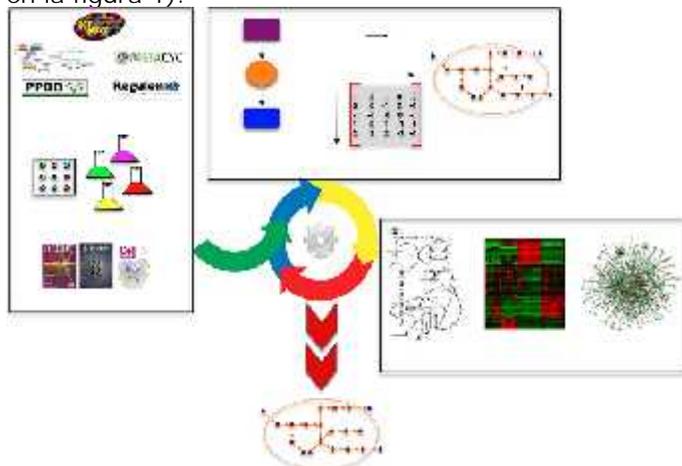


Figura 1. Pasos de la reconstrucción de la red o del modelo metabólico

Inicialmente, sobre la secuenciación y anotación de un genoma de interés, fuentes de la literatura y / o búsquedas de homología se utilizan para asignar la función a todos los marcos de lectura abierta (es decir, ORFs). Para cada función con una huella dactilar metabólica se asigna una transformación química específica. Por lo tanto, iterativamente marchando a lo largo de todo el genoma, una compilación de reacciones que abarcan toda la química del repertorio del organismo se puede lograr. Debe observarse que estos modelos no son necesariamente predictivo pero en cambio tienen una naturaleza de alcance porque nos permite evaluar lo que es metabólicamente factible. Restricciones reglamentarias sobre flujos de reacción se incorporan basados en la termodinámica (es decir, la reversibilidad de la reacción) y ómicas (es decir, transcriptómica / datos proteómicos) que pueden agudizar aún más las predicciones.

Uno de los pasos más críticos de la construcción de modelos metabólicos es establecer información GPR de un organismo específico a partir de bases de datos biológicos y / o fuentes de literatura. Para este fin, las bases de datos biológicos, tales como KEGG, SEED, MetaCyc, BKM-reaccionar, Brenda, Uniprot Expasy, PubChem, ChEBI y Propiedades físicas proporcionan información sobre reacciones / metabolitos asociados, enzimas y genes. Sin embargo, como se ilustra por Kumar (2012), las incompatibilidades en representación de los datos como metabolitos con nombres múltiples fórmulas químicas / bases de datos a través e, errores estequiometrias (es decir, elemental o cargar desequilibrios) y detalles incompletos atomistas (por ejemplo, ausencia de estereo-especificidad, y la presencia de R-grupo (s)) son los principales obstáculos para la rápida reconstrucción de nuevo modelos metabólicos de alta calidad mediante la combinación de información a partir de estas bases de datos. Recientemente, las bases de datos tales como MetRxn se han desarrollado para abordar estas cuestiones mediante la integración de la información (de los metabolitos y reacciones) de ocho de esas bases de datos.

En este sentido, resulta necesario elaborar un algoritmo soportado sobre criterios estadísticos y ortológicos (estudio de homología de secuencias genéticas) que permita lograr una completitud para dichos modelos, por lo que el autor del presente artículo se encuentra enfrascado en la elaboración de dicho algoritmo, el cual, una vez validado será presentado en próximos trabajos.

Referencias Bibliográficas:

- Benson, D. Lipman, D & Ostell, J. 1993. GenBank. Nucleic Acid Research. 21. 2963-2965.
- Chen, L. and Vitkup, D. (2006). Predicting genes for orphan metabolic activities using phylogenetic profiles. Genome Biol 2006, 7:R17.
- Edwards J.S., Ibarra R.U., Palsson B.Ø. (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19, 125–130.
- Fang, K., Zhao, H., Sun, Ch., Lam, C. M., Chang, S., Zhang, K., Panda, G., Godinho, M., Martins dos Santos, V. and Wang, J. (2011). Exploring the metabolic network of the epidemic pathogen Burkholderia cenocepacia J2315 via genome-scale reconstruction. Systems biology 1752-0509/5/83.
- Feist A.M., Herrgard M.J., Thiele I. et al. (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7, 129–143.

- Förster J., Famili I., Fu P. et al. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*13, 244–253.
- Green, M.L. and Karp, P.D. (2004). A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics*2004, 5:76.
- Kumar, V.S, Dasika, M.S and Maranas, C.D. (2007).Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 2007, 8:212 doi:10.1186/1471-2105-8-212.
- Kharchenko, P., Vitkup, D. and Church, G.M. (2004). Filling gaps in a metabolic network using expression information. *Bioinformatics* 2004, 20Suppl 1:1178-1185.
- Kharchenko, P., Chen, L., Freund, Y., Vitkup, D. and Church, G.M. 2006. Identifying metabolic enzymes with multiple types of association evidence. *BMC Bioinformatics* 2006, 7: 177.
- Latendresse, M., Krummenacker, M., Trupp, M., and Karp, P.D. (2012). Construction and completion of flux balance models from pathway databases. *Bioinformatics* Vol. 28 no. 3 2012, pages 388– 396. doi: 10.1093/bioinformatics/btr681).
- Notebaart R.A., van Enkevort F.H., Francke C. et al. (2006) Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics*. doi:10.1186/1471-2105-7-296.
- Osterman, A. and Overbeek, R. (2003). Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* 2003,7:238-251.
- Papin J.A., Price N.D., Wiback S.J. et al. (2003) Metabolic pathways in the post-genome era. *Trends BiochemSci* 28, 250–258.
- Reed JL, Vo TD, Schilling CH, Palsson B. (2003): An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome biology* 2003, 4(9):R54.51–R54.12.
- Snoep J.L., Bruggeman F., Olivier B.G. et al. (2006) Towards building the silicon cell: A modular approach. *BioSystems* 83, 207–216.
- Schilling C.H., Letscher D., Palsson B.Ø. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*203, 229–248.
- Thiele I., Palsson B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*5, 93–121.
- Triana J., Montagud A., Siurana M., Gamermann D., Torres J., Tena J., Fernández de Córdoba P., Urchueguía J. F. Reconstruction procedure to generate and evaluate genome-scale metabolic network models, use of *Synechococcus elongatus* PCC7942 as a case study. Summited at *Metabolic Network Reconstruction for the series Methods in Molecular Biology*, Springer.
- Voet D. and Voet J.G. (2004) *Biochemistry*. John Wiley & Sons. Hoboken.
- Saha R, Suthers PF, Maranas CD: *Zea mays* iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS ONE* 2011, 6:e21784.
- Ranganathan S, Suthers PF, Maranas CD: OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 2010, 6:e1000744
- Kumar A, Suthers PF, Maranas CD: MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinform* 2012, 13:6.