

Introducción

Las enzimas catalizan reacciones químicas transformando sustratos en productos. Durante el siglo XX, las enzimas fueron percibidas como catalizadores altamente específicos, sin embargo, esta percepción cambió con el descubrimiento de que pueden. Esta capacidad de catalizar varias funciones químicas se conoce como la promiscuidad de la enzima. *Escherichia coli* contiene al menos 404 enzimas promiscuas. La relevancia de la promiscuidad radica tanto en su papel como mecanismo de evolución de la función enzimática, así como en la necesidad de su detección para la corrección de modelos de flujo metabólico y la determinación de efectos secundarios en drogas farmacológicas. A pesar de su frecuencia e importancia aun se esta en el proceso de entender las causas y las características observables de la promiscuidad enzimática.

Para estudiar la promiscuidad es necesario contar con una definición, algunos autores emplean el término promiscuidad para describir actividades enzimáticas distintas a la función principal [khersonsky_enzyme_2010] otros lo ven como una actividad secundaria fortuita [copley_enzymes_2003] que pudo aparecer de forma accidental o inducida artificialmente [hult_enzyme_2007]. Otros mas, cuando una enzima puede operar sobre un amplio rango de sustratos, prefieren llamarla multiespecífica [khersonsky_enzyme_2010]. A la acción de realizar distintas funciones catalíticas, ya sea al catalizar varias reacciones químicas o bien una misma reacción en sustratos diferentes se le conoce como promiscuidad enzimática [obrien_catalytic_1999]. Existen varios tipos de promiscuidad enzimática.

Por sustrato cuando la reacción es la misma pero se lleva a cabo en distintos sustratos ejemplo la familia PriA [baronagomez_occurrence_2003] y la familia de betalactamasas [risso_phenotypic_2014]

Catalítica cuando la enzima utiliza diferentes mecanismos de reacción y/o residuos catalíticos, e.g. la quimotripsina puede catalizar reacciones de amidasa y fosfotriesterasa en un mismo sitio activo. [obrien_catalytic_1999] Por condiciones del entorno, cuando la enzima cambia su conformación dependiendo de las condiciones químicas y físicas presentes como pH, temperatura, solventes orgánicos y salinidad e.g. algunas lipasas pueden actuar como sintetizadoras de ésteres en lugar de hidrolasas en presencia de solventes orgánicos [hult_enzyme_2007, kumari_preparation_2007].

Este trabajo se enfocara a la promiscuidad por sustrato, entendiendo así que la enzima es capaz de catalizar la misma reacción química en al menos dos sustratos. La promiscuidad por sustrato es importante en términos evolutivos, por ejemplo la enzyme commission number (EC) separa las enzimas en clases, a cada enzima se asignan 4 dígitos, los tres primeros corresponden a la reacción y el último al sustrato; el mayor número de sustratos (4306 clases) que de reacciones químicas (234 en el tercer nivel) sugiere que la mayor variación evolutiva se da a nivel de sustrato y no de reacción [li_computational_2004]. Otra evidencia de la importancia de la multiespecificidad por sustrato esta en el descubrimiento de las superfamilias, enzimas mecanística y estructuralmente relacionadas que divergen en su afinidad por sustrato [glasner_evolution_2006, baier_evolution_2016]

Si bien existen familias de enzimas con alta especificidad por sustrato, otras familias como el citocromo P450 [bloom_neutral_2007, nath_quantitative_2008] y las beta lactamasas [zou_evolution_2015] son promiscuas. Es posible que la visión previa de alta especificidad se deba a que las primeras rutas metabólicas estudiadas pertenecen al metabolismo central, donde la especificidad puede haber sido favorecida por presiones de selección [firn_darwinian_2009]. Esta visión ha cambiado debido al conocimiento de mas enzimas con multifuncionalidad [jia_multifunctional_2013], sin afectar la eficiencia catalítica por la función primaria [aharoni_evolvability_2005]. En 1976 el interés por la promiscuidad comenzó por su influencia en la evolución de la función enzimática [jensen_enzyme_1976, pandya_enzyme_2014], las aproximaciones variaron desde la aparición de la síntesis funcional [dean_mechanistic_2007], cuando la disponibilidad de genomas permitió la combinación de análisis filogenéticos con técnicas de biología molecular, bioquímica y biofísica (Fig 1). En 2003 la biofísica de las proteínas entra en escena al postularse que la diversidad conformacional durante la dinámica molecular debe incidir en la aceptación de distintos sustratos. Recientemente se ha investigado su papel en efectos secundarios en drogas farmacológicas [nobeli_protein_2009, hopkins_drug_2009, nath_quantifying_2010, von_eichborn_promiscuous_2011, zhang_multidimensional_2012, zou_evolution_2015]. Entre 2005 y 2010 se avanza del estudio de una sola familia enzimática hacia el interés por propiedades globales, por ejemplo dado un genoma se investiga

la distribucion de familias promiscuas en subsistemas metabolicos. En estos años, surge el desarrollo de indices que reflejen las características bioquímicas de enzimas promiscuas. En 2010, comienzan los intentos por desarrollar un metodo computacional de prediccion de promiscuidad. Desde 2012 a la fecha, a la par que las aproximaciones bioinformaticas se multiplican, se desarrollan investigaciones de aspectos biofisicos, bioquimicos y evolutivos de enzimas promiscuas reafirmando que todos estos aspectos estan relacionados al fenomeno. En las siguientes secciones se describiran trabajos importantes sobre la relacion que guarda la promiscuidad con expansiones genomicas y flexibilidad molecular. Ademas se hablara sobre analisis bioquimicos y metabolicos para la descripcion del fenomeno.

Funcion biologica de la promiscuidad enzimatica

¿Por que existe la promiscuidad enzimatica? Se tiene evidencia de dos papeles biologicos: el primero proporcionar robustez a la red metabolica de un organismo mediante redundancia de reacciones de otras enzimas; el segundo permitir plasticidad evolutiva, es decir materia prima para la adaptacion a variaciones ambientales [aharoni_evolvability_2005, @sanchez-ruiz_promiscuity_2012, @martinez-nunez_lifestyle_2015] mediante la adquisicion de nuevas funciones quimicas. Respecto a la robustez, se probó que sobreexpresar enzimas promiscuas puede rescatar perdidas genicas [patrick_multicopy_2007]. De 104 knockout sencillos de genes esenciales para E. coli K-12, 20% de las auxotrofias pudieron ser suprimidas por la sobreexpresion de plasmidos que contenian enzimas promiscuas. Otro ejemplo que aporta a la robustez es PriA, enzima de la ruta de histidina que realiza en la ruta del triptofano la reaccion E.C. 5.3.1.24 [baronagomez_occurrence_2003]. En cuanto a la plasticidad se propone que para que la promiscuidad pueda dar origen a la aparicion de nuevas funciones la actividad promiscua debe proveer una ventaja fisiologica inmediata para poder ser seleccionada positivamente, ademas una vez que una funcion promiscua se vuelva relevante se debe poder mejorar mediante pocas mutaciones derivando en el intercambio entre la actividad promiscua y la principal [khersonsky_enzyme_2010].

Aun cuando el producto de la promiscuidad genera metabolitos que no se integran al metabolismo central de la celula, su efecto es positivo ya que estos metabolitos podrian colaborar a la adaptacion al entorno participando por ejemplo en una relacion de simbiosis o de competencia con otros organismos. Este tipo de metabolitos, por lo general, no son dañinos [notebaart_network-level_2014, @linster_metabolite_2013, @khanal_differential_2015] y pueden servir como bloques de construccion para vias metabolicas nuevas [ma_unconventional_2013, @adams_promiscuous_2014, @soskine_mutational_2010]. La respuesta inmediata de adaptacion de un organismo podria ser una consecuencia de su grado de promiscuidad.

Relacion del pangenoma con la promiscuidad enzimatica

El pangenoma es el contenido génico total de un linaje taxonómico. Las familias génicas de un pangenoma pueden clasificarse según su frecuencia de presencia/ausencia en cada genoma del linaje. De acuerdo a esta clasificación los principales grupos de familias génicas en un pangenoma son el *core*, el *shell* y el *cloud* también conocido como *dispensable* o *accessory* genome. El *core genome* es el conjunto de familias con presencia en todos los genomas del linaje. Por ejemplo, tanto la secuencia de la subunidad 16s de rRNA así como los diversos genes ribosomales suelen estar en el *core* de la gran mayoría de linajes bacterianos. El *shell genome* es el grupo de familias presentes en la mayoría de los genomas pero no en todos. En el *shell* se ubican por ejemplo familias que estaban en el *core genome* pero que algunas bacterias del linaje sufrieron una dinámica de pérdida génica. Mientras que el *cloud genome* o *dispensable genome* es aquel grupo de familias que sólo ocurre en unos cuantos genomas del linaje.

En el Dominio Bacteria se estima que entre al rededor de 200 secuencias están altamente conservadas. Si bien no están en el *core*, estas secuencias están compartidas por 90% de los genomas [halachev_calculating_2011]. El *core* depende de los genomas seleccionados, entre menos amplio sea el rango filogenético electo mayor será el tamaño del *core*. Dada su conservacion el *core genome* puede utilizarse para trazar mejores relaciones filogeneticas que las obtenidas con el uso exclusivo de marcadores como la subunidad 16s del RNA ribosomal o el gen rpoB. El pangenoma es el conjunto complemento del core genome, es decir todas aquellas secuencias que estan ausentes de uno o mas organismos del grupo y por lo tanto no son necesarias para todos, sino

solo posiblemente para el organismo que las posee. Como en el pangenoma la presion de seleccion esta relajada respecto al core-genome [firn_darwinian_2009] es el conjunto donde la plasticidad genomica tiene facilidades para desarrollarse.

Esta idea puede restringirse a subsistemas metabolicos para identificar genes cuyas enzimas estan en proceso de cambio de funcion quimica, por ejemplo, en este trabajo se encontro que el gen *trpF* esta presente en solo 49 de 290 genomas analizados del genero *Streptomyces* por lo que se encuentra en el pangenoma de triptofano de este genero taxonomico, posiblemente adquiriendo una nueva funcion [ma_unconventional_2013]. Para evitar problemas tecnicos del calculo del pangenoma existen otros modelos de medicion de variabilidad del genomica entre especies bacterianas [kislyuk_genomic_2011].

Modelos bioinformaticos de promiscuidad

Con el fin de reducir la inversion en el proceso de experimentacion, se han implementado en los ultimos años algoritmos computacionales para predecir promiscuidad enzimatica [carbonell_molecular_2010, cheng_global_2012, nagao_prediction_2014, noda-garcia_insights_2015, garcia-seisdedos_probing_2012]. Estos procedimientos cuentan con un conjunto de aprendizaje, unos descriptores del conjunto, una fase de ajuste de parametros y finalmente una prediccion. En 2010, Carbonell propone un algoritmo de soporte vectorial basado en subsecuencias de distinto tamaño que llama huellas moleculares. En este trabajo aplicado sobre 500,000 proteinas reportadas en la enciclopedia de Kyoto de genes y genomas (KEGG) se reporta 85% de exito en deteccion de enzimas promiscuas anotadas en KEGG. En 2012, Cheng compara los metodos de random forest y soporte vectorial en 6799 proteinas provenientes de la base de datos Universal Protein Resource (UniProt). Las enzimas son descritas con subsecuencias de aminoacidos incorporando ademas características biofisicas como polaridad. Se utiliza como grupo de control a familias de enzimas donde nunca se ha reportado una enzima promiscua.

Un aspecto no considerado en estos metodos es que hay familias de enzimas con alta identidad de secuencia entre sus miembros, con cambios bruscos en promiscuidad, debidos por ejemplo a la dinamica genomica [noda-garcia_evolution_2013, verdel-aranda_molecular_2015], lo que dificulta que considerar solo la secuencia conlleve a buenos predictores de promiscuidad. Cuando se obtiene una prediccion positiva utilizando los modelos existentes, lo que significa es que dada esa secuencia, en su familia se conoce previamente un elemento promiscuo y que ademas sus subsecuencias de cierto tamaño son suficientemente similares. Estos enfoques no pueden predecir de novo, en familias donde la promiscuidad no ha sido previamente detectada experimentalmente, pues no consideran aspectos evolutivos ni mecanisticos de las enzimas.

Otra limitante a los enfoques descritos es que mezclan en su conjunto de entrenamiento fenomenos distintos de promiscuidad. Cheng p. g. incluye enzimas moonlight que si bien poseen funciones adicionales a la catalizacion, son distantes a las enzimas promiscuas [copley_enzymes_2003]. Ademas en ambos casos mezclan en el mismo conjunto enzimas bacterianas y eucariotas, con lo que si existia una huella basada en secuencia entonces esta puede diluirse por la gran distancia taxonomica entre estos grupos (Tabla 1).

Promiscuidad in vitro y promiscuidad in vivo

La ganancia de promiscuidad no solo puede entenderse como la capacidad de convertir mas sustratos [carbonell_molecular_2010], sino tambien como la mejora de la capacidad catalitica respecto a ellos. El I-index [nath_quantitative_2008], esta definido como un rango de valores entre 0 y 1 que tiende a 1 entre mas parecida sea la actividad de la enzima sobre distintos sustratos, la capacidad catalitica es medida en terminos del cociente de Michaelis - Menten $\frac{K_{cat}}{K_m}$. El indice ha sido utilizado para predecir la afinidad por sustrato del citocromo P450 [nath_quantifying_2010]. Una limitante del indice I es que se deben conocer los sustratos a los que la enzima es afin; sin embargo se puede sospechar que una enzima ha ganado promiscuidad aun sin conocer sus potenciales sustratos. Otro punto a señalar es que las variables K_{cat} , K_m son mediciones realizadas in vitro y no se consideran todos los sustratos presentes in vivo. Para solventar esta dificultad e investigar variaciones de sustratos nativos se pueden buscar productos similares a los ya conocidos por medio de analisis metabolicos [nesvizhskii_analysis_2007] como los empleados en la deteccion de rutas no conservadas en la

biosíntesis de productos naturales [medema_computational_2015]. En particular para este fin se ha utilizado espectrometría de masas MS/MS, [nesvizhskii_analysis_2007,campbell_biophysical_2012] combinada con molecular networking para identificar productos similares [yang_molecular_2013,kocher_mass_2007]

El papel de la dinámica molecular en la promiscuidad

La estructura tridimensional de una proteína es obtenida mediante previa purificación y cristalización. Aunque mucho se ha hablado de la relación estructura función, al cristalizar se obtienen estados conformacionales homogéneos, que bien pueden no ser la única conformación que adopta la proteína en solución. [james_conformational_2003]. En particular en el problema de promiscuidad, se ha observado que la variación funcional no queda obviamente reflejada en la variación estructural, lo que sugiere un rol significativo para la dinámica molecular [parisi_conformational_2015,noda-garcia_evolution_2013,zou_evolution_2015]. Se postula que un aspecto de la dinámica molecular relevante para la diversificación de especificidad por sustrato es el número de conformeros [javier_zea_protein_2013]. Por ejemplo, en la actinobacteria *Corynebacterium diphtheriae* parece que el contexto genómico correlaciona con pérdida de promiscuidad de PriA ya que al poseer el genoma una copia de *trpF*, la enzima perdió esta función química conservando solo la función EC 5.3.1.16 correspondiente a la ruta de histidina. Esta sub-funcionalización se refleja en la pérdida de estados conformacionales cambiando desde 1 estado en *C. diphtheriae* hasta 4 presentes en la dinámica de PriA de *M. tuberculosis* [noda-garcia_evolution_2013].

Las regiones rígidas de una enzima proporcionan orientación adecuada con respecto a los grupos catalíticos, mientras que las regiones flexibles permiten al sitio activo adaptarse a los sustratos con diferentes formas y tamaños [copley_enzymes_2003]. Esta consideración sugiere que la flexibilidad del sitio activo es otra característica de la dinámica molecular a considerar para obtener información de la capacidad de ligación de una enzima a distintos sustratos [gatti-lafranchi_flexibility_2013]. Recientemente el índice de flexibilidad dinámica (dfi) se utilizó como una medida cuantitativa basado en la respuesta a perturbaciones de aminoácidos (PRS). Este índice se incrementa en regiones cercanas al sitio activo de beta lactamasas promiscuas respecto al correspondiente dfi de β -lactamasas especialistas existentes [zou_evolution_2015].

Modelo biológico diversidad de Actinobacteria

Al escoger un conjunto acotado para investigar familias de enzimas promiscuas se debe recordar que la funcionalidad es jerárquica por lo que para mejorar la anotación, es deseable reflejar el proceso evolutivo y restringirse a un grupo de organismos taxonómicamente relacionados [cruz-morales_phylogenomic_2016]. Actinobacteria es un phylum que posee promiscuidad tanto en el metabolismo periférico como en el core metabólico. Entre datos públicos (NCBI) y privados están disponibles alrededor de 1200 genomas no redundantes de especies de Actinobacteria. Como punto de partida, se han estudiado las relaciones filogenéticas y grupos de ortología [li_orthomcl_2003,waterhouse_orthodb_2013], en particular en Actinobacteria para identificar relaciones entre las familias del phylum, se obtuvieron árboles multilocus de entre 100 y 157 genomas [gao_phylogenetic_2012,sen_phylogeny_2014]. Estos estudios sugieren como separar los genomas disponibles para hacer el cálculo de grupos de ortología. Finalmente, se han realizado estudios de plasticidad genómica en *Streptomyces* considerando 5 y 17 organismos de los 300 genomas disponibles en la actualidad [zhou_genome_2012,kim_comparative_2015] donde reportan 2,018 familias en el core genome y 32,574 en el pangénoma.

Modelo metabólico biosíntesis de aminoácidos.

Al hacer el cálculo vemos que *Streptomyces*, un género del phylum Actinobacteria cuenta en su genoma con un promedio de 8316 secuencias codificantes según la especie. Gran parte de estas secuencias pueden ser agrupadas en subsistemas metabólicos como metabolismo de carbohidratos o de lípidos; de estos subsistemas uno de los más amplios es el metabolismo de aminoácidos con entre 429 y 910 secuencias según el organismo. La síntesis de aminoácidos es un subsistema presente en todas las especies pero con suficientes variaciones

que permiten hacer observaciones evolutivas. En un gran numero de Actinobacterias las rutas de histidina y triptofano de 7 y 11 pasos respectivamente convergen en una enzima bifuncional llamada PriA, que realiza tanto la funcion de HisA como la de TrpF [baronagomez_occurrence_2003]. La cantidad de familias en el subsistema de metabolismo de aminoacidos, su variabilidad, su conservacion entre distintos grupos taxonomicos y la existencia de estos ejemplos en Actinobacteria lo posicionan como un buen punto de partida para la busqueda de promiscuidad tanto de familias promiscuas como de miembros promiscuos de las mismas.

En las cuatro decadas de estudio de la promiscuidad enzimatica, hemos aprendido que es un fenomeno distribuido en distintos subsistemas metabolicos [nam_network_2012] y que su existencia puede deberse tanto al desarrollo de nuevas funciones para fines adaptativos [jensen_enzyme_1976,obrien_catalytic_1999,firn_darwinian_2009, como al rescate de una funcion perdida [patrick_multicopy_2007,baronagomez_occurrence_2003]. Por ello la dinamica de perdida y ganancia de genes asociada al contexto genómico en bacterias se relaciona con cambio en la funcion enzimatica [zhao_discovery_2013,zhao_prediction_2014,copley_evolutionary_2015]. Precisando, respecto a la ganancia de genes, se postula que la bifuncionalidad precede la duplicacion [hughes_evolution_1994,obrien_catalytic_1999]. Lo que implica que dada una duplicacion muy posiblemente previamente la promiscuidad estuvo presente [gerlt_divergent_2001,huang_enzyme_2012,noda-garcia_evolution_2013,risso_phenotypic_2014].

Se han desarrollado tecnicas bioquimicas y metabolicas de medicion [nath_quantitative_2008], asi como algoritmos computacionales de prediccion de promiscuidad [carbonell_molecular_2010,cheng_global_2012,noda-garcia_insights_2015]. Un aspecto a mejorar dentro del modelado es la restriccion del conjunto de estudio a un grupo taxonomico tan reducido que exista congruencia en las familias de ortologia y a la vez tan amplio que permita observar efectos evolutivos; el phylum Actinobacteria ha probado tener ejemplos de promiscuidad. Si bien la secuencia no ha sido suficiente para la correcta prediccion de promiscuidad [verdel-aranda_molecular_2015,copley_evolutionary_2015], es posible que dentro de las tecnicas computacionales la flexibilidad durante la dinamica molecular este correlacionada con la promiscuidad de los miembros de una familia [james_conformational_2003,javier_zea_protein_2013,noda-garcia_evolution_2013,zou_evolution_2015].

Modelo biologico

De los mas de mil genomas actualmente disponibles de Actinobacterias, se seleccionaron 888 (correspondientes a 49 familias), que no estan excesivamente fragmentados; es decir con un estimado de al menos 5 genes por contig (Tabla 2). Estos genomas fueron divididos en tres grupos (http://pubseed.theseed.org/wc.cgi?request=show_otus&base=/homes/nselem/Data/CS), uno de ellos correspondiente a Streptomycetaceae, la familia con la mayor cantidad de genomas disponibles; los otros dos grupos siguieron la taxonomia propuesta por Gao & Gupta en 2012. En el grupo de 290 genomas de Streptomycetaceae 2,126,832 ORFS fueron clasificados en 288,390 familias; de las 919,292 ORF del grupo I de Actinobacteria resultaron 269,406 familias. Las relaciones taxonomicas fueron corroboradas con algoritmos propios basados en best bidirectional hits (BBH).

Subsistemas metabolicos

Los operones his y trp de histidina [fondi_evolution_2009] y triptofano [merino_evolution_2008] respectivamente, participantes del metabolismo de aminoacidos estan ampliamente distribuidos en los organismos bacterianos. En Actinobacteria la familia promiscua PriA participa en ambas rutas biosinteticas, para su estudio se han generado datos bioquimicos, genomicos y estructurales (Tabla 3). En bacterias gram negativas estan presentes los operones his y trp y en lugar de PriA su familia homologa HisA. PriA comprende un conjunto de subfamilias en Actinobacteria. En Streptomyces, el gen trpF se desplaza de la vecindad genomica de trp, con lo que el homologo de hisA gana promiscuidad aunque con baja actividad de TrpF, a esta subfamilia se le llama PriB [verduzco-castro_co-occurrence_2016]. En otras Actinobacterias trpF se pierde totalmente y la familia homologa de HisA, se vuelve promiscua [baronagomez_occurrence_2003] realizando tanto la funcion quimica correspondiente a HisA como la de TrpF. Finalmente en la familia subHisA se pierde la funcion TrpF debido posiblemente a la ganancia del operon trp completo [noda-garcia_evolution_2013] y en la familia subtrpF se conserva solo a la funcion TrpF debido a la perdida del operon his [Juarez vazquez

et al 2015 in prep]. Existen al menos 43 familias de Actinobacteria sin explorar respecto a la funcionalidad de PriA.

Modelos computacionales

Para desarrollar herramientas se adoptó el enfoque de los contenedores bioinformáticos, todo el código fue depositado y documentado en GitHub y distribuido a través de un contenedor Docker