

Estudio bioinformático del metabolismo de *Mycobacterium tuberculosis* bajo condiciones de hipoxia

Bioinformatic analysis of *Mycobacterium tuberculosis* metabolism under hypoxic conditions

Christian Solís¹, Luisa Negrón¹

¹ Instituto de Química Biológica, Microbiología y Biotecnología Marco Antonio Garrido Malo, Facultad de Farmacia y Bioquímica, Universidad Nacional Mayor de San Marcos. Lima, Perú.

Resumen

Objetivo: Predecir, usando métodos bioinformáticos, las vías metabólicas preferentemente activas en *Mycobacterium tuberculosis* (MT), bajo condiciones de hipoxia. **Diseño:** Análisis biológico. **Lugar:** Instituto de Química Biológica, Microbiología y Biotecnología Marco Antonio Garrido Malo, Facultad de Farmacia y Bioquímica, UNMSM. **Material biológico:** Genes de *Mycobacterium tuberculosis*. **Métodos:** Inicialmente se seleccionó 355 genes de MT H37Rv, cuya expresión ha sido demostrada que es inducida bajo condiciones de hipoxia, y 359 cuya expresión es reprimida. Usando la información de secuencia de los genes con expresión inducida y reprimida, se analizó de modo comparativo la posición en el genoma de cada grupo de genes, así como algunas propiedades fisicoquímicas (punto isoelectrico y momento hidrofóbico) de sus proteínas correspondientes. Posteriormente, a cada gen se le asignó una vía metabólica o regulatoria, usando la información sobre MT de la librería de genes y genomas de Kyoto (KEGG), y el procesamiento de secuencias, empleando el programa PATH-A. **Principales medidas de resultados:** Vías metabólicas en genes de *Mycobacterium tuberculosis*, bajo condiciones de hipoxia. **Resultados:** No se encontró diferencias entre los genes con expresión inducida y reprimida, en su distribución en el genoma de MT, así como en la distribución de los valores para las propiedades fisicoquímicas analizadas en sus productos. De los 355 genes con expresión inducida iniciales, solamente fue posible asignar al menos una vía metabólica a 95, usando KEGG, y 57, usando PATH-A. **Conclusiones:** El análisis comparativo de las vías metabólicas asignadas a los genes con expresión inducida y reprimida reveló que, bajo condiciones de hipoxia, se encuentran reprimidos muchos genes relacionados a vías metabólicas que implican gasto de ATP, encontrándose inducidos algunos genes cuyas proteínas participan en vías del metabolismo central, tales como el metabolismo del piruvato, glucólisis y ciclo del ácido cítrico.

Palabras clave: *Mycobacterium tuberculosis*; hipoxia; latencia; vías metabólicas.

Abstract

Objective: To predict by using bioinformatic tools *Mycobacterium tuberculosis* (MT) metabolic pathways under hypoxic conditions. **Design:** Biology analysis. **Setting:** Instituto de Química Biológica, Microbiología y Biotecnología Marco Antonio Garrido Malo Biological, Microbiologic and Biotechnologic Chemistry Institute, Faculty of Pharmacy and Biochemistry, UNMSM. **Biologic material:** *Mycobacterium tuberculosis* genes. **Methods:** The study began with the selection of 355 genes of MT H37Rv whose expression has been shown by other studies is induced by hypoxic conditions and 359 genes whose expression was repressed. Up and down expressed genes were comparatively analyzed, localizing genes of each group within the MT genome and predicting some physicochemical properties (isoelectric point and hydrophobic moment) for their protein products. In order to assign a metabolic or regulatory pathway to each gene, Kyoto Encyclopedia of Genes and Genomes (KEGG) and PATH-A sequence analysis tool were used. **Main outcome measures:** Metabolic pathways in *Mycobacterium tuberculosis* genes under hypoxia conditions. **Results:** From the initial 355 up expressed genes, it was possible to assign metabolic pathways to only 95 using KEGG and 57 using PATH-A. **Conclusions:** There were no differences between up and down expressed genes for their genome distribution and values for studied physicochemical properties of their protein products. The comparative analysis of the assigned metabolic pathways to down and up-expressed genes revealed that under hypoxic conditions several metabolic pathways related to ATP spent were down-expressed, being induced some genes whose proteins participate in central metabolism pathways such as the pyruvate metabolism, glycolysis and citric acid cycle.

Key words: *Mycobacterium tuberculosis*; hypoxia; latency; metabolic pathways.

INTRODUCCIÓN

El *Mycobacterium tuberculosis* (MT) es el agente causal de la tuberculosis, una enfermedad infecciosa crónica que es la responsable aproximadamente de 2 a 3 millones de muertes anuales⁽¹⁾. Se considera que un tercio de la población mundial se encuentra infectado por MT y de ella solo 10% llega a presentar la enfermedad⁽²⁾. En el resto de la población infectada, el microorganismo se encuentra en un estado de latencia, el cual le permite al patógeno sobrevivir, esperando una oportunidad cuando las defensas inmunológicas de su hospedero disminuyan, para poder reactivarse y causar la enfermedad bajo su forma clínica⁽³⁾. Uno de los modelos experimentales para el estudio de la la-

tencia es el estudio de MT cultivado bajo condiciones de hipoxia, que simula parcialmente el ambiente anaerobio, hostil al patógeno, en el interior del hospedero, cuando la bacteria se encuentra latente. En el presente estudio se hace un análisis del contexto biológico de la actividad de las proteínas productos de la expresión génica bajo condiciones de hipoxia, mediante la asignación de vías metabólicas a cada uno de sus genes, usando la base de datos KEGG y programas como PATH-A. Teniendo en consideración este conocimiento, se plantea un modelo de las vías metabólicas preferentes bajo condiciones de hipoxia en el MT.

Este conocimiento es importante, porque puede ser la base para estudios

posteriores de búsqueda de nuevos antimicrobianos específicos para el patógeno bajo el estado de latencia.

MÉTODOS

La información de expresión génica usada provino de los trabajos de Sherman y col.⁽⁴⁾, en la Universidad de Washington (Seattle, WA-USA), y puede ser obtenida de la dirección <http://schoolniklab.stanford.edu/projects/tb.html>. La distribución en el genoma de MT H37Rv de los genes cuya expresión fue inducida y reprimida se representó gráficamente usando el programa *Genome Pixelizer* 2.2.7⁽⁵⁾. Para ello, primero se obtuvo el archivo GenBank para todos los genes del

MT H37Rv, del servidor FTP del NCBI. Posteriormente, usando esa información, así como programas escritos en perl, se extrajo la información de la posición genómica para cada gen cuya expresión fue reprimida o inducida. Se construyó con ello el archivo de entrada apropiado para el programa *Genome Pixelizer*, procediendo entonces a su ejecución, obteniendo como archivo de salida el gráfico correspondiente.

Para analizar de manera comparativa algunas de las características fisicoquímicas de los productos de los genes cuya expresión es inducida o reprimida por las condiciones de hipoxia, se usó los programas IEP y HMOMENT del paquete de programas EMBOSS (6), y las secuencias de aminoácidos de todas las proteínas productos de esos genes se obtuvo para cada una de las proteínas los valores predichos de punto isoeléctrico (pI) y momentos hidrofóbicos máximos (μH) (7). Las secuencias de aminoácidos fueron previamente obtenidas del servidor FTP del NCBI y la extracción de las secuencias correspondientes solamente a los genes con expresión reprimida o inducida se realizó con el programa *fastcmd* del paquete de programas *NCBI.toolkit* (8). Para la extracción de la información relevante de los archivos de salida de estos programas, se escribió varios programas *scripts* con el lenguaje perl (9). Se realizó todas estas operaciones bajo el sistema operativo Linux.

La información de las vías metabólicas fue obtenida de la base de datos KEGG (10). Los identificadores de vías metabólicas de KEGG fueron asignadas a cada gen de MTm cuya expresión fue inducida o reprimida bajo condiciones de hipoxia, para su análisis comparativo. Solo las secuencias de las proteínas de MT productos de genes con expresión inducida, a los que se le pudo asignar una vía metabólica, fueron seleccionadas, para su posterior procesamiento bioinformático. El planteamiento de un modelo de las relaciones entre las vías metabólicas que se presentan de modo preferente bajo con-

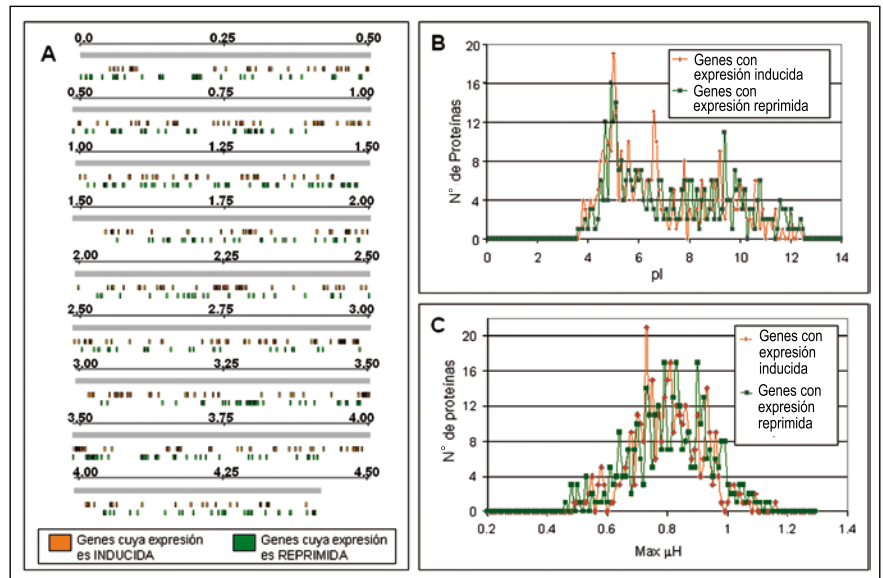


Figura 1. Análisis comparativo de genes con expresión inducida y reprimida bajo condiciones de hipoxia. A. Localización en el genoma de MT H37Rv (distancia física dentro del genoma en megabases). B. Distribución de valores de punto isoeléctrico en las proteínas productos de la expresión génica. C. Distribución de valores del momento hidrofóbico máximo (Max μH), en las proteínas productos de la expresión génica.

diciones de hipoxia se realizó usando el programa *PATH-A* (11).

RESULTADOS

La localización en el genoma de MT H37Rv de los genes cuya expresión fue inducida o reprimida se muestra en la Figura 1A. La distribución de los genes expresados e inducidos se da alrededor de todo el genoma y no se aprecia visualmente diferencias entre ambos gru-

pos de genes. La distribución del punto isoeléctrico y momento hidrofóbico máximo para ambos grupos de proteínas fue semejante (Figuras 1B y 1C).

De los 355 genes con expresión inducida bajo condiciones de hipoxia, solamente a 95 se les pudo asignar una vía metabólica de acuerdo a KEGG y 57 usando el programa *PATH-A*. En la Figura 2 se presenta un gráfico comparativo de las vías metabólicas asignadas a los genes de MT, cuya expresión se indujo

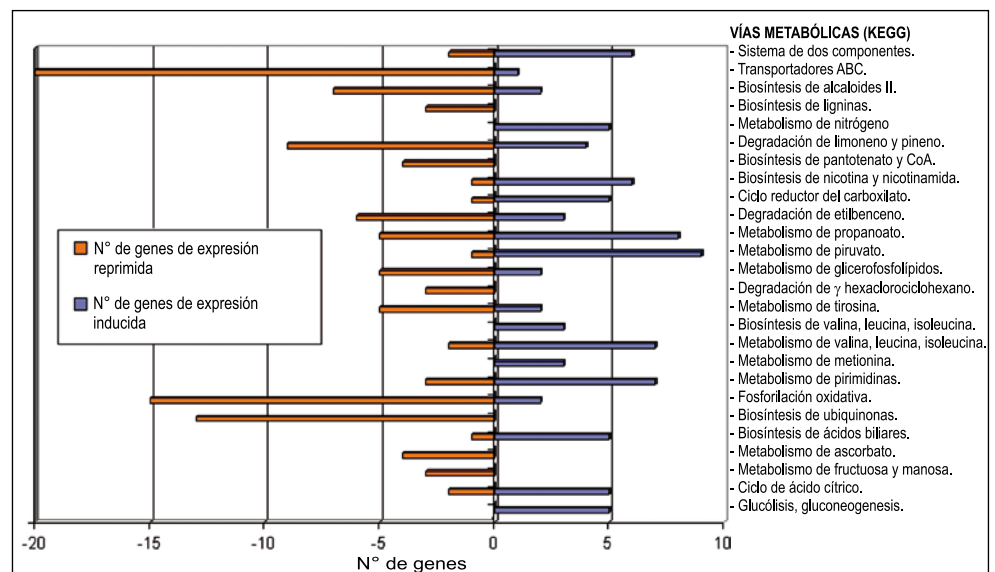


Figura 2. Asignación diferencial de vías metabólicas de acuerdo a KEGG, a genes cuya expresión se encuentra inducida y reprimida bajo condiciones de hipoxia.

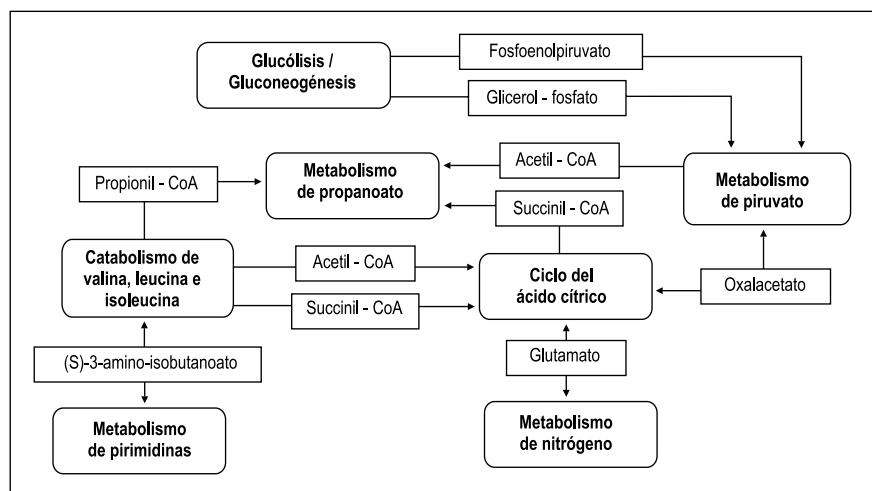


Figura 3. Relaciones entre las vías metabólicas predichas para proteínas cuyos genes se expresan bajo condiciones de hipoxia.

o reprimió bajo condiciones de hipoxia. Basado en los resultados obtenidos con el programa PATH-A, se desarrolló un modelo de las relaciones entre las vías metabólicas predichas para proteínas cuyos genes se expresan bajo condiciones de hipoxia (Figura 3).

DISCUSIÓN

Los datos derivados de los experimentos de *microarrays* son una fuente de información que puede ser procesada de diversos modos. La primera hipótesis que manejamos fue la de una distribución física diferente entre los genes cuya expresión es inducida y reprimida. Como se observa en la Figura 1A, la distribución de los genes expresados e inducidos se da alrededor de todo el genoma de MT H37Rv y no se aprecian visualmente diferencias entre ambos grupos de genes.

Otra hipótesis que se planteó es que posiblemente por ser genes cuyas proteínas productos de su expresión presentan entornos fisicoquímicos diferentes, como el pH y el potencial de oxidación; las propiedades fisicoquímicas de ambos grupos de proteínas también serían diferentes. Como ejemplo de cambios en el entorno fisicoquímico del patógeno en el transcurso de la infección, podemos mencionar el hecho de que en el ambiente del fagosoma, una vez interiorizado el bacilo en el macrófago, el MT se ve sometido gradualmente a una acidificación de su pH, que va de 6,5 hasta valores menores

a 5,5^(12,13). Como se representa en las Figuras 1B y 1C, la distribución del punto isoeléctrico (pI) y momento hidrofóbico máximo ($\mu\text{H Max}$) para ambos grupos de proteínas fue bastante semejante; eso podría hacernos inferir que otros parámetros diferentes al pI y al $\mu\text{H Max}$ u otras estrategias, como la presencia de chaperoninas⁽¹⁴⁾, probablemente podrían explicar la adaptación funcional de ambos grupos de proteínas a diferentes entornos fisicoquímicos.

Las variaciones en la expresión génica están relacionadas de un modo directo con cambios en el metabolismo. Cuando los cultivos de MT son sometidos a condiciones de hipoxia, se ha observado cambios en la expresión génica. Por ello, el interés en el presente estudio de predecir a partir del análisis de esos cambios los posibles cambios a nivel del metabolismo del patógeno, que expliquen la fisiología de este. Estudios previos han demostrado que, cuando los cultivos de MT son sometidos a la reducción de la concentración de oxígeno, pueden pasar a un estado de no replicación y ser resistentes a los agentes antituberculosos⁽¹⁵⁾; de ello, la importancia de estos análisis.

En muchos trabajos para el análisis de los resultados de los experimentos de expresión génica, se considera a cada gen expresado de manera individual o como parte de un grupo de genes con similares modos de expresión en varios experimentos. En el presente trabajo se considera

otra forma de análisis de los datos de expresión génica; el análisis no considera un gen individual o en un solo grupo, sino cada gen puede formar parte de diferentes grupos, que además son vías metabólicas y regulatorias, establecidas previamente en MT, a partir de numerosos trabajos previos. Esta forma de agrupar a los genes por vías metabólicas tiene un mayor sentido biológico, ya que en las células cada gen puede codificar información para una proteína que puede participar en varias vías metabólicas a la vez. Para este propósito, nos fue de valiosa ayuda la biblioteca de genes y genomas de Kyoto (KEGG)⁽¹⁰⁾ y el programa PATH-A⁽¹¹⁾. En KEGG se pone a disposición de los usuarios la información sobre las vías metabólicas de manera interactiva o a través de un servidor FTP, donde la información se obtiene como archivos de texto, de los que se puede extraer la información de interés del investigador, luego de un procesamiento mediante programas *scripts* escritos con ese propósito, siendo PERL el lenguaje de programación usado por su sintaxis apropiada para la manipulación de información en forma de textos. En cambio, PATH-A usa de forma directa información de secuencia para predecir vías metabólicas en base a técnicas de aprendizaje automático *machine learning* y análisis de secuencias (SVM, BLAST y HMM), que permiten primero la anotación funcional de las secuencias y posteriormente su agrupamiento en una vía metabólica determinada. El mayor número de pasos que toma el procesamiento de la información con PATH-A respecto a KEGG explicaría porqué con esta técnica se realizó asignaciones de vías metabólicas a un menor número de proteínas.

En la Figura 2 se observa que el metabolismo de piruvato es una de las vías en las que hay una mayor diferencia entre en número de genes con expresión inducida y reprimida; ello puede tener relación con el hecho de que esta vía es un punto central entre el metabolismo de carbohidratos y lípidos. Se observa también que, en algunos casos, para una misma vía metabólica hay genes cuya expresión es inducida y reprimida; esto se explica porque incluso para una misma función enzimática se pueden presentar en MT

dos o más proteínas (isoformas) cuyos genes se encuentren diferencialmente expresados. Respecto a los genes cuya expresión se reprime bajo condiciones de hipoxia, se observa que para los transportadores ABC se asigna un alto número de estos genes; ello se puede explicar porque forman parte de un sistema de transporte de sustratos del exterior hacia el citoplasma, que requiere la hidrólisis de ATP⁽¹⁶⁾. Bajo condiciones de hipoxia, en las que la disponibilidad de energía bajo la forma de ATP se encuentra limitada, la expresión de esos genes se encuentra reprimida, a fin de evitarle a MT mayores gastos de ATP, que los que le son indispensables.

Es necesario agregar que, para el análisis comparativo de las vías metabólicas bajo condiciones de hipoxia, solo se ha considerado a las proteínas cuyos genes han tenido su expresión inducida o reprimida; por lo que, algunas vías metabólicas con un aparente bajo número de proteínas asignadas, pueden también incluir proteínas cuya expresión es constitutiva y no varían ante las condiciones de hipoxia; esto se aplica sobre todo para las vías metabólicas troncales, como las del ciclo del ácido cítrico y la glucólisis/gluconeogénesis.

Las condiciones de hipoxia en las que se obtuvo experimentalmente los datos de expresión génica⁽⁴⁾ simulan en parte el ambiente anaerobio hostil del patógeno en el interior del granuloma, ambiente en el que el patógeno entra en un estado de latencia. El desarrollo de agentes antituberculosos en contra del patógeno bajo este estado fisiológico sería valioso, porque evitaría que el microorganismo retorne a su fase activa de crecimiento, donde se manifiesta la enfermedad y produce sus más graves efectos. Por ello, la información obtenida en el presente trabajo sobre las vías metabólicas preferentemente activas bajo estas condiciones, es importante para un posterior estudio de selección de blancos farmacológicos específicos para el estado

de latencia. Un antimicobacteriano diseñado para bloquear la actividad de una proteína, que participa en alguna de estas vías metabólicas de un modo indirecto, no solo podría alterar el funcionamiento de la vía metabólica, sino también la supervivencia del patógeno en estado de latencia.

AGRADECIMIENTOS

A la Universidad Internacional de Andalucía (España) y a la Red Iberoamericana de Bioinformática por las oportunidades de entrenamiento y formación en bioinformática y química computacional. Un agradecimiento especial al Consejo Superior de Investigaciones de la Universidad Nacional Mayor de San Marcos, por el financiamiento parcial de la presente investigación.

REFERENCIAS BIBLIOGRÁFICAS

1. Dye C, Watt CJ, Bleed DM, Hosseini SM, Raviglione MC. Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. *JAMA*. 2005;293(22):2767-75.
2. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med*. 2003;163(9):1009-21.
3. Aaron L, Saadoun D, Calatroni I, Launoy O, Memain N, Vincent V, et al. Tuberculosis in HIV-infected patients: a comprehensive review. *Clin Microbiol Infect*. 2004;10(5):388-98.
4. Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc Natl Acad Sci USA*. 2001;98(13):7534-9.
5. Kozik A, Michelmore R. Computational approach to select the set of ESTs with a single BLAST hit to Arabidopsis genome [monografía en Internet]. University of California at Davis, 2002 [acceso 24 Jul 2006]. Disponible en: http://cgpdb.ucdavis.edu/COS_Arabidopsis/
6. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16(6):276-7.
7. Eisenberg D, Wilcox W, McLachlan AD. Hydrophobicity and amphiphilicity in protein structure. *J Cell Biochem*. 1986;31(1):11-7.

8. Shiryev SA, Papadopoulos JS, Schäffer AA, Agarwala R. Improved BLAST searches using longer words for protein seeding. *Bioinformatics*. 2007;23(21):2949-51.
9. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res*. 2002;12(10):1611-8.
10. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006;34:D354-7.
11. Pireddu L, Szafron D, Lu P, Greiner R. The path-A metabolic pathway prediction web server. *Nucleic Acids Res*. 2006;34(Web Server issue):W714-9.
12. Oh YK, Straubinger RM. Intracellular fate of *Mycobacterium avium*: use of dual-label spectrofluorometry to investigate the influence of bacterial viability and opsonization on phagosomal pH and phagosome lysosome interaction. *Infect Immun*. 1996;64:319-25.
13. Sturgill-Koszycki S, Schlesinger PH, Chakraborty P, Haddix PL, Collins HL, Fok AK, et al. Lack of acidification in *Mycobacterium phagosomes* produced by exclusion of the vesicular proton-ATPase. *Science*. 1994;263(5147):678-81.
14. Qamra R, Mande SC, Coates AR, Henderson B. The unusual chaperonins of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*. 2005;85(5-6):385-94.
15. Hu Y M, Butcher PD, Sole K, Mitchison DA, Coates ARM. Protein synthesis is shut down in dormant *Mycobacterium tuberculosis*. *FEMS Microbiol Lett*. 1998;158:139-45.
16. Murugasu-Oei B, Tay A, Dick T. Upregulation of stress response genes and ABC transporters in anaerobic stationary-phase *Mycobacterium smegmatis*. *Mol Gen Genet*. 1999;262(4-5):677-82.

Manuscrito recibido el 27 de agosto de 2008 y aceptado para publicación el 15 de agosto de 2008.

Correspondencia:
Dra. Luisa Negrón Ballarte
Facultad de Farmacia y Bioquímica,
Universidad Nacional Mayor de San Marcos.
Jr. Puno 1002. Lima 1, Perú
Correo-e: luisane@terra.com.pe