

MODELAJE DE PROTEÍNAS DE *Mycobacterium tuberculosis* RELACIONADAS A SU DEFENSA FRENTE AL ESTRÉS OXIDATIVO

CHRISTIAN SOLIS CALERO¹

RESUMEN

En el presente trabajo se describe el uso de procedimientos bioinformáticos e información genómica de acceso libre por Internet, para identificar y caracterizar proteínas que participan en la defensa de *Mycobacterium tuberculosis* (MT) frente al estrés oxidativo, con el propósito de obtener nuevos blancos farmacológicos en este patógeno. Las secuencias de aminoácidos de las proteínas correspondientes a los genes *katG*, *ahpC*, *ahpD*, *sodA*, *sodC* y *mgtC* fueron comparadas con el genoma de *Mycobacterium tuberculosis* H37Rv a través del servidor del National Center for Biotechnology Information, determinando su posible redundancia en el genoma. Paralelamente, se obtuvo información sobre sus residuos más conservados evolutivamente por alineamiento múltiple global usando el programa MAXHOM, así como modelos de sus estructuras tridimensionales usando los métodos de «modelaje molecular comparativo» a través del servidor SWISS-MODEL o modelaje por el método de «reconocimiento de plegamiento» (threading) a través del metaservidor GENESILICO. En base al análisis genómico se puede concluir que las proteínas codificadas por los genes *katG*, *ahpD* y *mgtC* son buenos candidatos como blancos farmacológicos debido a que no tienen similitud significativa con alguna proteína codificada por genes presentes en el genoma humano y no presentan redundancia en el genoma del *Mycobacterium tuberculosis*.

Palabras clave: *Mycobacterium tuberculosis*, estrés oxidativo, bioinformática, genoma, modelaje molecular

SUMMARY

This work uses bioinformatics tools and free online access to genomic information in order to identify and characterize proteins related to oxidative stress defense of *Mycobacterium tuberculosis*, in order to discover new appropriate pharmacological targets against this pathogen. Sequences of amino acids of proteins expressed with that function, products of the expression of the genes *katG*, *sodA*, *sodC*, *ahpC*, *ahpD* and *mgtC* were compared with the genome of *Mycobacterium tuberculosis* H37Rv through the NCBI server National Center for Biotechnology Information), determining their possible redundancy in the genome. Parallely, information was obtained above of their evolutive conserved sequences by global multiply alignment using MAXHOM program, as well models of their three-dimensional structures by comparative modeling method through the Swiss-model server, or using «recognition of folding» method (threading) through the GENESILICO metasever. By genomic analysis, proteins codified by *katG*, *ahpD* and *mgtC* genes would be very good candidates to be pharmacological targets due to absence of redundancy in the genome of *Mycobacterium tuberculosis* as well as similarity with any codified protein by human genes.

Key words: *Mycobacterium tuberculosis*, oxidative stress, bioinformatics, genome, molecular modeling

INTRODUCCIÓN

Para el diseño de nuevos fármacos, una de las nuevas estrategias es el llamado diseño racional, que comporta un conocimiento previo de la diana sobre la que se quiere actuar. La información sobre las dianas y blancos farmacológicos se encuentra de modo primario residente en el genoma, por ello una de las claves en esta estrategia es traducir esta información genómica usando varias herramientas, entre las que destaca la Bioinformática,

que permitan la identificación de dianas concretas y el posterior desarrollo de candidatos a fármacos. En el caso de los patógenos, las enzimas que catalizan los pasos limitantes de ciertas reacciones exclusivas de procariotas, son dianas adecuadas, fáciles de «extraer» una vez conocida la secuencia genómica. Seleccionar estas dianas resulta todo un reto para las empresas a fin de llevar a cabo esta investigación de una forma rápida y sin necesidad de mucha inversión.

Durante el proceso de infección por *Mycobacterium tuberculosis* (MT) en el huésped, los macrófagos generan especies químicas reactivas de oxígeno y

¹ Facultad de Farmacia y Bioquímica - Universidad Nacional Mayor de San Marcos.

nitrógeno que restringen el desarrollo del patógeno (1), el cual, para contrarrestar este efecto, ha desarrollado varios mecanismos de defensa. Dentro de las especies químicas generadas por los macrófagos, destacan los intermedios reactivos de Nitrógeno (RNI) que incluyen NO, NO₂, y NO₃, porque se ha demostrado que juegan un rol importante en controlar el crecimiento del MT en el huésped (2, 3). La habilidad de MT para sobrevivir al ataque de los RNI involucra un grupo de proteínas, las que por su importancia para la supervivencia de la bacteria son blancos potenciales para nuevos fármacos a desarrollar para eliminar a este patógeno del organismo, de allí el interés en este trabajo de caracterizarlas y validarlas como blancos farmacológicos usando las herramientas bioinformáticas disponibles de acceso libre en Internet.

MATERIAL Y MÉTODOS

Búsquedas en bases de datos de secuencias. De la base de datos NR (no redundante) GenBank del NCBI (National Center for Biotechnology Information) localizado en la siguiente dirección (<http://www.ncbi.nlm.nih.gov/>) utilizando el programa ENTREZ, se obtuvo registros de secuencias de aminoácidos en formato FASTA, correspondientes a proteínas relacionadas a la defensa de MT frente al estrés oxidativo (Tabla 1).

Análisis genómico. Las secuencias de aminoácidos de las proteínas seleccionadas fueron comparadas con los genomas de MT y *Homo sapiens* usando los programas TBLASTN y contra la lista de proteínas referenciadas de ambos organismos usando el programa BLASTP a través de los servidores del NCBI, para determinar su presencia, localización, así como la posible redundancia de sus correspondientes genes en estos genomas. La información sobre el papel de las proteínas en el metabolismo de MT fue obtenida en el Tuberculist Web Server (Instituto Pasteur: <http://genolist.pasteur.fr/TubercuList/>).

Caracterización de las secuencias de aminoácidos. A través del servidor PREDICT PROTEIN <http://cubic.bioc.columbia.edu/predictprotein>, se calculó:

- Estructura secundaria de las proteínas usando el programa PROFSEC
- Búsqueda de motivos usando el programa PROSITE
- Alineamiento múltiple global usando el programa MAXHOM.
- Predicción del % de residuos de aminoácidos expuestos (en más del 16% de su superficie), midiendo la accesibilidad del solvente con el programa PROFSEC.

Predicción de las estructuras tridimensionales. Previamente se realizó una búsqueda por similitud en la base de datos de estructuras 3D del NCBI usando el programa BLASTP. De presentar la secuencia problema una similitud significativa con la secuencia de alguna proteína de estructura 3D conocida se realizó la aproximación a sus estructuras tridimensionales por modelaje molecular comparativo a través del servidor SWISS-MODEL (<http://www.expasy.ch>), en su defecto la aproximación se realizó por el método de «reconocimiento de plegamiento» («threading») a través del metaservidor para predicción de plegamiento GENESILICO (<http://genesilico.pl/meta>). La calidad de las estructuras tridimensionales fue evaluada usando el programa VADAR 1.4 (<http://redpoll.pharmacy.ualberta.ca/vadar>). El análisis de los modelos de las estructuras tridimensionales obtenidas en busca de los residuos de aminoácidos más conservados, se realizó usando el visor de estructuras 3D RASMOL y la información obtenida por el alineamiento múltiple global obtenido con el programa MAXHOM usando varias secuencias similares de otras especies.

RESULTADOS Y DISCUSIÓN

Al analizar el genoma de MT con el fin de desarrollar nuevos fármacos, el procedimiento a seguir pasa primero por el uso de herramientas bioinformáticas para identificar y caracterizar los genes de interés adicionalmente, usando dicha información se puede dar la utilización de técnicas de análisis de la expresión diferencial de genes (DNA microarrays), acopladas con la disciplina emergente de la química combinatoria. Un aspecto clave en la infección causada por MT es su supervivencia en el interior de los macrófagos, ya que estas células son capaces de destruir más del 90% de los bacilos (4). Para sobrevivir MT debe evitar los mecanismos bactericidas de estas células, una de las estrategias es a través de la detoxificación de los reactivos intermediarios del oxígeno o del nitrógeno, es decir defenderse del estrés oxidativo al que es sometido por el huésped.

En la tabla 1 se presenta una lista de las proteínas de MT seleccionadas para este estudio que participan en la defensa frente al estrés oxidativo, entre ellas destaca el sistema Catalasa-peroxidasa codificada por el gen *KatG* que defiende al patógeno de las especies reactivas de oxígeno (1, 5, 6). Esta catalasa-peroxidasa, es responsable de la conversión de la Isoniazida a su forma biológicamente activa contra el patógeno (7). Las mutaciones en *katG* que codifican para una catalasa/ peroxidasa incapaz de convertir la isoniácida a su forma acti-

va, confieren resistencia a este fármaco, entonces queda como protector antioxidante el sistema alkyl hidropéroxido reductasa (*AhpC/AhpD*) que es capaz de detoxificar peróxidos orgánicos y, probablemente, también peróxido de hidrógeno, su actividad se atribuye a la presencia en su estructura de un par de residuos de cisteína (Cys) (8). Dos genes en MT codifican las enzimas superóxido dismutasas, *SodA* [Fe] y *SodC* [Cu-Zn]. Estas proteínas parecen jugar un papel importante protegiendo a la bacteria de la acción de los intermediarios reactivos del oxígeno. *SodA* a diferencia de *SodC*, es una proteína dependiente de la proteína SecA2 para su secreción (9). Bacterias mutantes de los genes que codifican estas proteínas han mostrado ser susceptibles a los mecanismos

bactericidas de macrófagos peritoneales activados con interferón gamma (INF- γ) usando ratón como modelo experimental (10). Otra proteína incluida en este estudio es una proteína C transportadora de magnesio codificada por el gen *mgtC*, la disrupción de este gen provoca un crecimiento alterado del microorganismo en macrófagos derivados de monocitos humanos y menor supervivencia en ratones BALB/c infectados por vía venosa, lo cual indica que el gen *mgtC* juega un papel importante en el crecimiento de MT en macrófagos humanos y en su virulencia en ratones (11). El transporte de cationes como el Mg^{2+} parece ser uno de los mecanismos que le permite a MT sobrevivir dentro de un fagosoma ligeramente ácido y limitado en Mg^{2+} (12).

TABLA 1. CARACTERÍSTICAS BIOQUÍMICAS DE LAS PROTEÍNAS ESTUDIADAS

GEN	Proteína	Número de aminoácidos	Peso molecular (Daltons)	Punto Isoeléctrico	Acceso a la secuencia (ENTREZ)
katG	Sistema Catalasa-peroxidasa	740	80600.37	4.8700	NP_216424.1
sodA	Superóxido dismutasa A	207	23032.65	6.4323	NP_218363.1
sodC	Superóxido dismutasa C	240	23843.13	6.4255	NP_214946.1
ahpC	Sistema alkyl hidropéroxido	195	21565.05	4.2810	NP_216944.1
ahpD	reductasa	177	18779.40	7.0572	NP_216945.1
mgtC	Proteína C transportadora de Mg	234	24805.22	8.4887	NP_216327.1

En la tabla 2 se presenta el análisis genómico, que permite validar las proteínas en estudio como posibles blancos farmacológicos. En este análisis se evalúa la redundancia, que puede presentarse cuando para una proteína hay en el genoma más de un segmento con similitud, lo que puede ser una señal que para la actividad de esa proteína existen isoformas codificadas por distintos genes bastante similares pero no iguales, lo que dificultaría el hallar un inhibidor para su actividad, ya que los sitios de unión al ligando no serían exactamente los mismos presentándose diferentes valores de afinidad y de actividad intrínseca. De acuerdo a los resultados obtenidos de las 6 proteínas analizadas, todas, excepto la correspondiente al gen *ahpC*, no presentan

una señal de redundancia (Figura 1). Los criterios para identificar los segmentos son los valores de E (probabilidad de que el alineamiento obtenido sea producto del azar) que deben ser superiores a 10^{-6} , los valores de Score (una medida de la significancia estadística) que deben ser superiores a 50 bits, y la longitud del alineamiento (se muestra en la columna posición del alineamiento en la proteína de la tabla 2), el análisis de este último parámetro es importante porque frecuentemente alineamientos muy cortos presentan valores E altos, lo que no es confiable. Se comparó también las secuencias de las proteínas frente a la información de secuencia del genoma humano. Es un criterio de validación que las proteínas del MT, que presenten una similitud

significativa con alguna proteína humana, deben ser descartadas como posibles blancos farmacológicos ya que es probable que los antagonistas de es-

tas proteínas presenten también algún grado de unión específica a proteínas codificadas por genes del genoma humano, lo que, a la larga, se traduci-

TABLA 2. EVALUACIÓN DE LAS PROTEÍNAS SELECCIONADAS EN BASE A DATOS GENÓMICOS

PROTEÍNA	Segmentos similares			Parámetros (BLASTP) Genoma MT H37TRV		Parámetros (BLASTP) Genoma Humano	
	Gen similar	Posición del alineamiento en la proteína (aminoácidos)	Posición del gen el genoma de MT (nucleótidos)	Score	E value	Score	E value
Sistema Catalasa-peroxidasa KatG	<i>katG</i>	1-740	2153889 - 2156111	1494	0.0	No se encontró similitud	
Superóxido dismutasa A sodA	<i>sodA</i>	1-207	4320704 - 4321327	434	e-123	207	3e-54
Superóxido dismutasa C sodC	<i>sodC</i>	1-240	519600 - 520322	451	e-128	50	2e-06
ahpC: alkyl hidroperóxido reductasa	<i>ahpC</i>	1-195	2726193 - 2726780	399	e-113	122	1e-28
	<i>ahpE</i>	49-146	2510715 - 2511176	52	2e-08		
	<i>bcpB</i>	49-170	1807298 - 1807762	48	4e-07		
	<i>bcp</i>	4-144	2837684 - 2838157	44	7e-06		
ahpD: alkyl hidroperóxido reductasa	<i>ahpD</i>	1-177	2726806 - 2727339	322	9e-90	No se encontró similitud	
Proteína C transportadora de Mg (<i>mgtC</i>)	<i>mgtC</i>	1-234	2053443 - 2054147	462	e-131	No se encontró similitud	

ría en efectos adversos para el paciente de usar esos antagonistas como fármacos contra el MT. Los resultados de este estudio (tabla 2) muestran que

las proteínas codificadas por los genes *katG*, *ahpD*, *mgtC* no presentan ninguna similitud significativa con proteína alguna de origen humano.

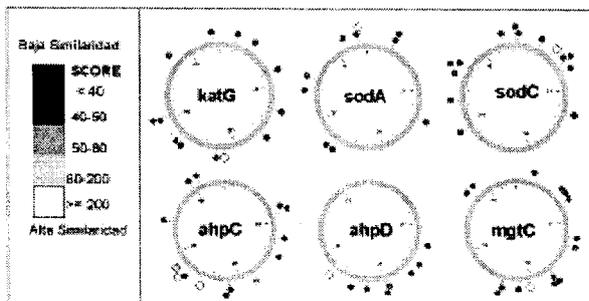


Figura 1. Análisis genómico de la redundancia en el genoma del MT para las proteínas estudiadas. Cada punto alrededor del genoma representado como un círculo señala la similitud encontrada y la escala de grises la significancia de ésta.

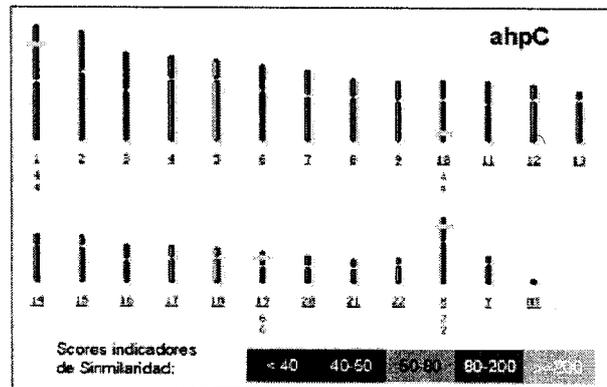


Figura 2. Similitud de la proteína codificada por el gen *ahpC* de *Mycobacterium tuberculosis* a proteínas codificadas por genes del genoma humano

La tabla 3 muestra los datos obtenidos en los cálculos de predicción de la estructura secundaria de las proteínas en estudio. Se señala además los métodos que se han usado para obtener información de sus estructuras tridimensionales, así como la evaluación de la calidad de estos modelos a través del programa VADAR 1.4 (13). La información sobre la estructura tridimensional obtenida de modo experimental para las proteínas codificadas por los genes *katG*, *ahpD*, *SodA* y *SodC*, la ob-

tuvimos de la base de datos PDB («Protein Data Bank»).

Las otras proteínas no presentan estructuras determinadas experimentalmente por lo que se obtuvo modelos de sus estructuras tridimensionales por los métodos de predicción de «modelado por homología» en el caso de los productos del gen *ahpC*, y el método de «reconocimiento de plegamiento» (threading) que se aplicó en el caso del

TABLA 3. CARACTERÍSTICAS DE LAS ESTRUCTURAS 3D DE LAS PROTEÍNAS SELECCIONADAS

Proteína	% de Estructuras secundarias (PROFsec)			Método de obtención de la Estructura tridimensional	Calidad de la estructura (Vadar 1.4)
	α hélices	Hojas- β	lazos		
Sistema Catalasa-peroxidasa (KatG)	32.43	8.51	59.05	Difracción por rayos X (Código PDB: 1SJ2)	5/10
Superóxido dismutasa A <i>sodA</i>	53.62	9.18	37.20	Difracción por rayos X (Código PDB: 1IDS)	7/10
Superóxido dismutasa C <i>sodC</i>	12.50	25.42	62.08	Difracción por rayos X (Código PDB: 1PZS)	7/10
Sistema alkyl hidroperóxido reductasa (<i>ahpC</i>)	19.49	18.97	61.54	Modelaje comparativo (SWISS MODEL)	6/10
Sistema alkyl hidroperóxido reductasa (<i>ahpD</i>)	80.23	0	19.77	Difracción por rayos X (Código PDB: 1GU9)	7/10
Proteína C transportadora de magnesio (<i>mgtC</i>)	48.72	19.23	32.05	Threading (MGENTHREADER)	6/10

gen *mgtC*, usando el Metaservidor GeneSilico. En este último caso, se seleccionó el modelo generado por el programa MGenTHREADER, por presentar los valores más confiables respecto a otros métodos de threading como 3dpssm, Bioinbgu, Ffasm03, Ffasm2.0, Pcons, Sam-T02 y sparks.

Estos dos tipos de métodos usan información que proviene, básicamente, de otras proteínas cuya estructura se ha determinado experimentalmente. Los métodos de «modelado por homología» se basan en la similitud que presenta la secuencia de aminoácidos de la proteína de interés con las secuencias de proteínas cuya estructura tridimensional ya se ha determinado experimentalmente (14). Cuando la secuencia de aminoácidos de la proteína analizada presenta baja similitud con proteínas cuya estructura 3D se conoce, no se puede aplicar el modelado por homología, por lo que se uso el método «threading». En este método se

busca que estructuras de conocida estructura 3D, presenten el mejor modelo de plegamiento al que corresponda la secuencia problema (15). La estrategia general de estos programas, como el MGenTHREADER, consiste en evaluar la secuencia problema en una serie de plegamientos y encontrar en cual de ellos esta secuencia se «encaja» o «se encuentra bien». MGenTHREADER combina varios métodos incluyendo alineamiento, potenciales de solvatación y potenciales entre pares, finalmente evalúa el alineamiento usando una red neuronal (16). Con la Proteína C transportadora de Mg, (*mgtC*) sólo se obtuvo un modelo con valores confiables usando MGenTHREADER, pero que no pudo validarse con otros métodos ni de modelaje por homología y threading por lo que se constituye en un objeto de estudio para el afinamiento del modelo por métodos propios de la Química Computacional usando la mecánica molecular o cuántica, métodos que involucran un

mayor costo computacional en tiempo, pero debido al bajo peso molecular de esta proteína (234 aminoácidos) respecto a las otras proteínas en estudio, ello se puede hacer factible sobre todo para zonas de la proteína en los que no hay ninguna semejanza a los plegamientos conocidos.

Los sitios de unión a ligandos, así como los sitios activos, presentan generalmente residuos que han sido conservados en la evolución, debido a su importancia para la función de la proteína, por ello es posible detectarlos por alineamiento múltiple. Con ese propósito se usó el programa MAXHOM, usando los alineamientos múltiples se pudo identificar estos residuos para luego buscar su posición en la estructura 3D de la proteína usando un programa de visualización de estructuras 3D como RASMOL. En la Figura 4 se presentan los resultados correspondientes a la alquil hidroperóxido reductasa codificada por el gen *ahpC*; destaca el hecho de que varios residuos altamente conservados en el alineamiento múltiple como la Asp 88, Ser 86, Ser 89 y la His 93 se presentan muy cerca una de la otra y las tres primeras formando parte de un «loop» o lazo proteico, estructura secundaria que muy frecuentemente aparte de conectar otros tipos de estructuras secundarias, está relacionada a los sitios activos de las enzimas (17). Se hace énfasis en estos residuos porque son caracte-

ción en la estructura 3D de la proteína usando un programa de visualización de estructuras 3D como RASMOL. En la Figura 4 se presentan los resultados correspondientes a la alquil hidroperóxido reductasa codificada por el gen *ahpC*; destaca el hecho de que varios residuos altamente conservados en el alineamiento múltiple como la Asp 88, Ser 86, Ser 89 y la His 93 se presentan muy cerca una de la otra y las tres primeras formando parte de un «loop» o lazo proteico, estructura secundaria que muy frecuentemente aparte de conectar otros tipos de estructuras secundarias, está relacionada a los sitios activos de las enzimas (17). Se hace énfasis en estos residuos porque son caracte-

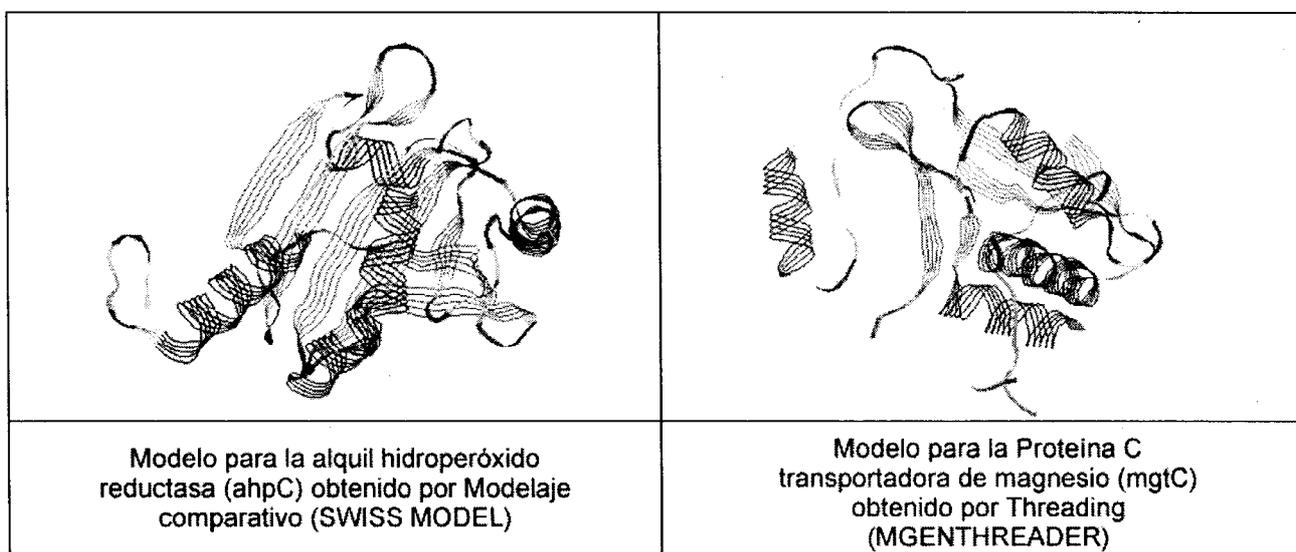


Figura 3. Modelos de la estructura tridimensional obtenidos por modelaje molecular usando secuencias de aminoácidos

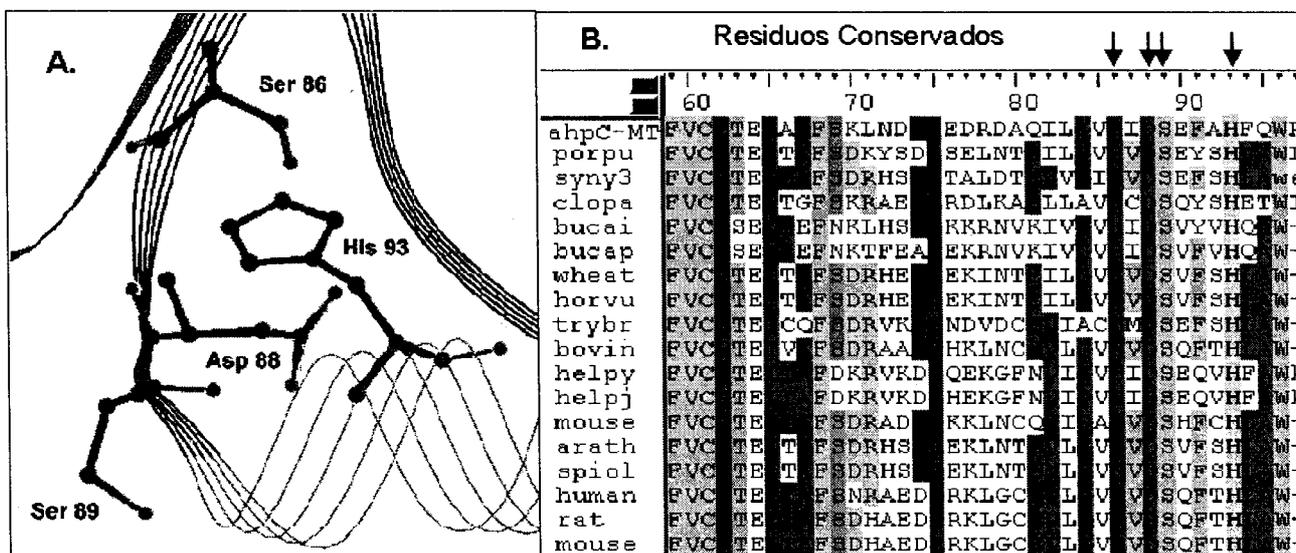


Figura 4. Análisis del Modelo de la alquil hidroperóxido reductasa codificada por el gen *ahpC*. A: Residuos conservados característicos de centros activos enzimáticos. B: Alineamiento múltiple obtenido con el programa MAXHOM mostrando los residuos conservados.

rísticos de los centros activos de las enzimas y porque los otros residuos conservados correspondientes a aminoácidos aromáticos y alifáticos que forman bloques en el alineamiento, son más bien característicos de los núcleos o «core» de las proteínas, regiones de la proteína muy importantes para la estabilidad de la estructura de la proteína aunque no están relacionados de manera directa a su rol catalítico (18).

Se concluye que las proteínas producto de los genes *katG*, *ahpD* y *mgtC* en base a los análisis bioinformáticos realizados, presentan las mejores características desde el punto de vista genómico para ser consideradas como blancos farmacológicos. Como se dispone en algunos casos de modelos de estructuras 3D obtenidos experimentalmente y en otros casos por Modelaje molecular, es factible usar esa información para posteriores ensayos computacionales y experimentales que permitan el diseño de nuevos ligandos antagonistas de estas proteínas, que luego de un proceso de optimización se puedan constituir en nuevos fármacos. La información obtenida por el trabajo «in silico» proporciona una aproximación complementaria que puede ser la base para el trabajo experimental «in vitro».

REFERENCIAS BIBLIOGRÁFICAS

- Manca C, Paul S., Barry CE., Freedman V.H, Kaplan G. Mycobacterium tuberculosis catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro. *Infect. Immun* 1999; 67:74-79
- Chan, J., Xing Y., Magliozzo RS. Bloom B.R. Killing of virulent Mycobacterium tuberculosis by reactive nitrogen intermediates produced by activated murine macrophages. *J. Exp. Med.* 1992; 175:1111-1122.
- MacMicking JD., North RJ., Lacourse R., Mudgett JS., Shah SK. Nathan C.F. Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc. Natl Acad. Sci. USA* 1997; 94:5243-5248
- Fenton MJ, Vermeulen MW. Immunopathology of tuberculosis: roles of macrophages and monocytes. *Infect Immun.* 1996; 64(3):683-90.
- Heym B., Zhang Y., Poulet S., Young D. Cole S.T. Characterization of the *katG* gene encoding a catalase-peroxidase required for the isoniazid susceptibility of Mycobacterium tuberculosis. *J. bacteriol* 1993; 175:4255-4259.
- Flynn JL. Chan J. Immunology of tuberculosis. *Annu. Rev. Immunol* 2001; 19:93-129.
- Koshkin A, Nunn C, Djordjevic S, Ortiz de Montellano P. The Mechanism of Mycobacterium tuberculosis Alkylhydroperoxidase AhpD as Defined by Mutagenesis, Crystallography, and Kinetics. *Journal Biol Chem* 2003; 278(32):29502-08
- Bryk R, Griffin P, Nathan C. Peroxynitrite reductase activity of bacterial peroxiredoxins. *Nature* 2000; 407: 211-15
- Braunstein M, Espinosa BJ, Chan J, Belisle JT, Jacobs WR. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis. *Mol Microbiol* 2003; 48(2):453-64.
- Piddington DL, Fang FC, Laessig T, Cooper AM, Orme IM, Buchmeier NA. Cu,Zn superoxide dismutase of Mycobacterium tuberculosis contributes to survival in activated macrophages that are generating an oxidative burst. *Infect Immun* 2001; 69(8):4980-7.
- Chacón O, Realpe T, Barletta R, Robledo J. Inactivación de genes de Mycobacterium tuberculosis y su potencial utilidad en la prevención y el control de la tuberculosis. *Biomédica* 2004; 24:165-87
- Buchmeier N, Blanc-Potard A, Ehrt S, Piddington D, Riley L, Groisman EA. A parallel intraphagosomal survival strategy shared by Mycobacterium tuberculosis and Salmonella enterica. *Mol Microbiol* 2000; 35(6):1375-82.
- Morris A., MacArthur M., Hutchinson, E. Stereochemical quality of protein structure coordinates. *Proteins* 1992; 12(4):345-364
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003; 31:3381-3385.
- Fischer D, Barret C., Bryson K. CAFASP-1: Critical Assessment of Fully Automated Structure Prediction Methods. *Proteins: Structure, Function and Genetics* 1999; 3:209-217
- Jones D. GenTHREADER: An Efficient and Reliable Protein Fold Recognition Method for Genomic Sequences. *J Mol Biol* 1999; 287(4):797-815.
- Branden C, Tooze J. Introduction to protein structure. Garland Publishing New York 1991.
- Claverie J-M. Notredame C. Building a Multiple Sequence Alignment. *Bioinformatics for Dummies*, New York, Wiley Publishing Inc., 2003; pp 279-314