

# De la secuencia de un genoma bacteriano a la identificación de candidatos vacunales

Daniel Yero Corona.

Instituto Finlay. Centro de Investigación-Producción de Vacunas. Ave. 27 No. 19805. La Lisa. A.P. 16017, C.P. 11600. Ciudad de La Habana, Cuba. E-mail: dyero@finlay.edu.cu

Cada día hay más secuencias genómicas completas de un gran número de patógenos humanos. La disponibilidad de estas secuencias ha cambiado completamente el panorama para el desarrollo de vacunas, introduciendo una nueva línea de pensamiento en este proceso. Esta metodología comienza por la secuencia genómica y mediante un análisis computacional se predicen aquellos antígenos más probables a ser candidatos vacunales. Por ejemplo, con el uso de herramientas bioinformáticas se puede hacer un pesquijaje *in silico* de aquellas proteínas expuestas en la superficie bacteriana, con vistas a identificar antígenos candidatos vacunales. La confirmación *in vitro* de estos resultados de localización celular y el uso de modelos animales para evaluar la inmunogenicidad de los candidatos acota finalmente el número de candidatos definidos por la computadora. A este proceso, aplicado por primera vez a *Neisseria meningitidis* serogrupo B, se le denomina vacunología inversa. La genómica también brinda información sobre la biología y la virulencia de especies patógenas mediante la genómica comparativa. En este trabajo de revisión, describimos cómo la genómica puede ser usada en la identificación de nuevos candidatos vacunales.

**Palabras clave:** Genómica, vacunas, bioinformática, vacunología inversa, *Neisseria meningitidis*.

## Introducción

Recientemente ha habido dos innovaciones importantes en el campo de las vacunas. La primera ha sido el advenimiento de las técnicas modernas de biología molecular que han permitido la invención de importantes vacunas para la humanidad. Ejemplo de ellas son las vacunas recombinantes como la vacuna contra la hepatitis B, basada en la proteína de la envoltura viral expresada y purificada a partir de levaduras (1). La segunda revolución en el diseño de vacunas es el resultado del uso de la genómica. La genómica es en general el conjunto de herramientas para analizar o visualizar toda la información genética de un organismo o una célula en un momento dado. La genómica integra el análisis de las secuencias de ADN y la expresión de los genes, así como la anotación en bases de datos.

Desde la primera publicación de la secuencia completa del genoma de un organismo vivo (*Haemophilus influenzae*) (2), numerosos reportes han aparecido en la literatura donde se describe la secuencia genómica de muchas otras bacterias patógenas. Por ejemplo: *Helicobacter pylori*, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Chlamydia trachomatis*, *Treponema pallidum*, *Mycoplasma genitalum*, *Borrelia burgdorferi*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Escherichia coli*, etcétera (3,4). Hasta ahora se han publicado alrededor de 175 secuencias completas y más de 400 genomas de otros microorganismos están siendo secuenciados. Una lista completa y actualizada de todos los genomas secuenciados o en estudio la podemos encontrar disponible por Internet [<http://wit.integratedgenomics.com/GOLD/>]. Esta gran cantidad de información ha contribuido con la ayuda de la informática, a la búsqueda de genes de interés, un proceso comúnmente referido como minería de genomas. La aplicación de la informática a la biología se

considera una nueva rama de la investigación denominada Bioinformática. Esta disciplina se define en la actualidad como el estudio del contenido y flujo de la información en sistemas y procesos biológicos. Ahora es posible revisar la secuencia de un genoma, seleccionar genes específicos, determinar si un gen tiene una función determinada, diseñar oligonucleótidos para amplificar fragmentos codificantes por PCR y predecir localización subcelular, talla molecular, solubilidad, punto isoeléctrico, etcétera, de una proteína. Todo esto podemos hacerlo desde una computadora (análisis *in silico*) antes de llevar a cabo cualquier experimento en el laboratorio (5).

## Vacunología inversa

El desarrollo de una vacuna es en los inicios el proceso de identificar aquellas estructuras únicas capaces de generar respuesta inmune protectora. Usando técnicas tradicionales, este proceso lleva comúnmente a ensayos de prueba y error y a la eliminación de algunos candidatos en estudio. El análisis bioinformático de los genomas bacterianos permite la identificación racional de las proteínas que pudieran estar expuestas al sistema inmune y así minimiza grandemente los ensayos de prueba y error en el diseño de vacunas de subunidades. Por ejemplo, para las bacterias patógenas extracelulares los análisis *in silico* se centran en la identificación de proteínas localizadas en la superficie bacteriana expuestas al sistema inmune (6).

Un candidato vacunal que induzca protección conferida por la inmunidad humoral debe cumplir los siguientes criterios: Estar conservado antigénicamente entre la mayoría de los aislamientos clínicos; inducir anticuerpos funcionales; proteger en un modelo animal; y en última instancia ser seguro y eficaz para uso en humanos. Los métodos clásicos para encontrar este tipo de candidatos han sido, por ejemplo, la inmunización de animales con

células enteras del microorganismo patógeno, el uso del suero policlonal generado en la búsqueda de proteínas inmunorreactivas mediante Western blot, seguido de la purificación de estas proteínas, la secuenciación de algunos fragmentos y posteriormente la clonación, expresión recombinante y determinación de la secuencia nucleotídica del gen que codifica para el candidato (Figura 1). Este proceso consume gran cantidad de tiempo y recursos y rinde un número limitado de posibles candidatos dentro de un período definido de tiempo.

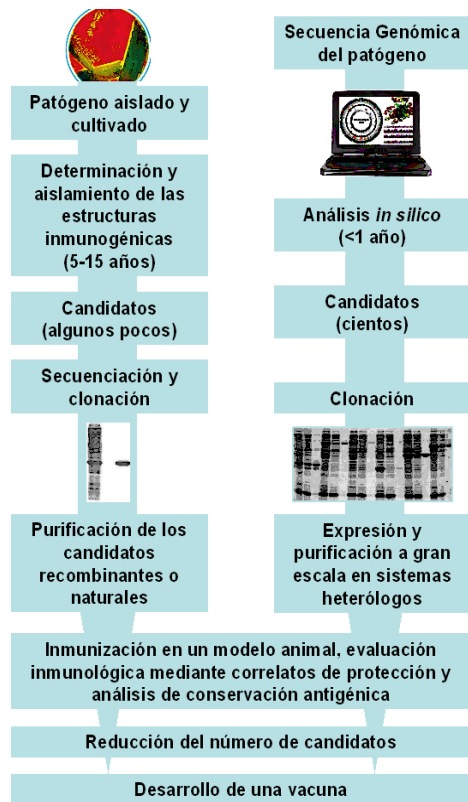


Figura 1. Diagrama de flujo que muestra el proceso de desarrollo de una vacuna mediante métodos tradicionales (izquierda) o mediante el uso de la genómica y la bioinformática (derecha). El proceso se acelera grandemente usando la genómica en la etapa de la determinación de los posibles candidatos.

La bioinformática tiene la capacidad de acelerar este proceso de descubrimiento permitiendo la rápida generación de una colección de candidatos vacunales potenciales para posteriores análisis en el laboratorio. Adicionalmente, este enfoque permite la evaluación de candidatos que normalmente son producidos por el patógeno en cantidades pequeñas y escapan al ser detectados por métodos tradicionales o clásicos. El enfoque *in silico* de un genoma en la búsqueda de candidatos vacunales parte del análisis de la secuencia nucleotídica, lo que permitirá detectar los candidatos, los cuales son posteriormente clonados, expresados y purificados como proteínas recombinantes, usados para inmunizar animales de experimentación y finalmente evaluados según su capacidad de inducir una respuesta inmune protectora. A este proceso, comparado con los enfoques

clásicos, se le denomina vacunología inversa (Figura 1), término introducido por Rino Rappuoli en el año 2001 (7). La vacunología es la ciencia que combina el estudio del sistema inmune con la diversidad antigénica de los patógenos y aquellas formulaciones que pueden modularlo para prevenir o curar enfermedades.

### El análisis en la computadora (*in silico*)

La primera etapa en el análisis de un genoma bacteriano involucra la identificación de todos los genes que codifican para proteínas y la asignación de la función probable a cada una de estas secuencias. Ese proceso se conoce como anotación del genoma. La anotación de los genes identificados se hace, por ejemplo, usando BLAST (herramienta de búsqueda de secuencias similares por alineamiento) (8). El programa BLAST busca secuencias homólogas en dos grandes bases de datos de secuencias: el GenBank y el SWISS-PROT. La búsqueda usando BLAST está disponible en Internet a través del Centro Nacional de Información Biotecnológica de los Estados Unidos (NCBI) [<http://www.ncbi.nlm.nih.gov/>]. Normalmente la anotación de un genoma la encontramos disponible por Internet cuando esta secuencia es publicada y guardada en las bases de datos. Pero debemos tener en cuenta que muchos genomas no se anotan correctamente y posteriores reanotaciones son publicadas sobre las ya existentes (9). Por tanto, muchos análisis *in silico* para identificar candidatos vacunales parten de una anotación llevada a cabo por los propios investigadores involucrados en el diseño de la vacuna en cuestión (10).

Un segundo paso hacia la identificación de los candidatos *in silico* es el relacionado con la localización celular del producto de los genes anotados. Las proteínas una vez que son sintetizadas en la célula se localizan en diferentes compartimentos celulares donde ellas realizan sus funciones; para el caso de la célula procariota estos compartimentos pueden ser: el citoplasma, la membrana citoplasmática o el medio extracelular. En el caso particular de las bacterias gramnegativas podemos adicionar el medio periplasmático y la membrana externa. La localización celular de una proteína está predeterminada por su secuencia primaria, por eso se han desarrollado numerosos algoritmos computacionales que interpretan el posible destino final de una proteína a partir de su secuencia (Tabla 1). La gran mayoría de los programas disponibles se enfocan en la predicción de proteínas de membrana ya que son las más interesantes en el diseño de una droga o una vacuna.

El programa SignalP 3.0 [<http://www.cbs.dtu.dk/services/SignalP/>] predice la presencia y localización del sitio de corte del péptido señal en secuencias aminoacídicas de diferentes organismos incluyendo eucariontes (11, 12). El programa LipoP 1.0 [<http://www.cbs.dtu.dk/services/LipoP/>] discrimina entre diferentes péptidos señales aquellos que predicen la formación de una lipoproteína (13), lo que es un indicio de localización en membrana. Otros programas como PSORT [<http://psort.nibb.ac.jp/>] integran diferentes algoritmos que predicen entre todas las variantes de localización celular para organismos de diferentes reinos (14).

**Tabla 1. Ejemplo de algunos programas o bases de datos disponibles en Internet para la manipulación de secuencias de ADN y proteínas con el objetivo de buscar posibles candidatos vacunales.**

Objetivo	Nombre	Dirección (URL)
<b>Búsqueda de secuencias</b>		
Genomas completos	TIGR	<a href="http://www.tigr.org/">http://www.tigr.org/</a>
	Sanger Institute	<a href="http://www.sanger.ac.uk/">http://www.sanger.ac.uk/</a>
	GOLD database	<a href="http://wit.integratedgenomics.com/GOLD/">http://wit.integratedgenomics.com/GOLD/</a>
	GenBank	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
Por homología	BLAST	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>
<b>Búsqueda de genes (ORFs)*</b>		
ORFs en bacteria	SMS	<a href="http://bioinformatics.org/sms2/orf_find.html">http://bioinformatics.org/sms2/orf_find.html</a>
ORFs en eucariontes	GENSCAN	<a href="http://genes.mit.edu/GENSCAN.html">http://genes.mit.edu/GENSCAN.html</a>
Por homología	blastx	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>
<b>Estructura secundaria</b>		
Hélices alfa	TMHMM 2.0	<a href="http://www.cbs.dtu.dk/services/TMHMM/">http://www.cbs.dtu.dk/services/TMHMM/</a>
Barriles-beta	TBBpred	<a href="http://www.imtech.res.in/raghava/tbbpred/">http://www.imtech.res.in/raghava/tbbpred/</a>
	PredictProtein**	<a href="http://www.embl-heidelberg.de/predictprotein/">http://www.embl-heidelberg.de/predictprotein/</a>
<b>Localización celular</b>		
Péptido señal	SignalP 3.0	<a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>
Lipoproteínas	LipoP 1.0	<a href="http://www.cbs.dtu.dk/services/LipoP/">http://www.cbs.dtu.dk/services/LipoP/</a>
	PSORT**	<a href="http://psort.nibb.ac.jp/">http://psort.nibb.ac.jp/</a>
<b>Identificación de epitopos</b>		
Epitopos B y T	IMTECH***	<a href="http://www.imtech.res.in/bic/">http://www.imtech.res.in/bic/</a>
Unión a MHC I y II	RANKPEP	<a href="http://www.mifoundation.org/Tools/rankpep.html">http://www.mifoundation.org/Tools/rankpep.html</a>

\* Con estos programas se pueden buscar ORFs (del inglés: *Open Reading Frames*) en secuencias cortas de ADN y no en genomas completos. Para esto existen programas más complejos como por ejemplo ARTEMIS. [<http://www.sanger.ac.uk/Software/Artemis/>] que puede ser descargado libremente de Internet.

\*\* Este tipo de servidor manipula secuencias de proteínas utilizando un gran número de programas basados en algoritmos diferentes, dando un resultado más integral y confiable.

\*\*\* Servicio bioinformático del Instituto de Tecnología Microbiana de la India, con numerosos programas para el diseño de vacunas de subunidad basado en la identificación de epitopos B y T.

Por otro lado, existe otro gran número de programas que predicen la localización basados en la estructura tridimensional que adoptarán las proteínas y en predicciones de estructura secundaria (Tabla 1). Es muy típico que en bacterias gramnegativas las proteínas de membrana interna estén mayormente formadas por dominios con estructura de hélice alfa, mientras que las de membrana externa por motivos de láminas beta que en muchos casos forman los denominados barriles beta (15,16). Muchos de estos programas analizan los perfiles de hidrofobicidad de cada segmento de una proteína para así predecir si estos segmentos atraviesan la membrana (17).

Para el caso de patógenos intracelulares donde la respuesta celular es la que determina la protección, la identificación de proteínas de membrana no es un requisito para reducir el número de candidatos. En cambio la identificación *in silico* de epitopos T pudiera indicarnos cuáles son los genes con mayor probabilidad de codificar para proteínas inmunoprotectoras.

El servidor RANKPEP disponible por Internet [<http://www.mifoundation.org/Tools/rankpep.html>] permite la identificación de

péptidos que se unen a las moléculas del complejo mayor de histocompatibilidad de clase I (MHC I) o de clase II (MHC II). Adicionalmente este programa hace un análisis de posibles sitios de corte por el proteosoma y predice cuáles ligandos del MHC I provienen de una digestión proteolítica por el proteosoma, lo que aumenta el rigor de la predicción (18). Estas predicciones permiten diseñar vacunas basadas en genes o péptidos quiméricos.

Finalmente, gracias a que existe un gran número de microorganismos con sus genomas secuenciados, incluso dentro de una misma especie, se puede verificar si los genes candidatos detectados están conservados. Para ello localizamos las secuencias ortólogas en los genomas relacionados y usando herramientas de alineamiento como ClustalW [<http://www.ebi.ac.uk/clustalw/>] determinamos el grado de conservación de cada candidato.

#### Genómica Comparada

El análisis *in silico* de genomas completos tiene la potencialidad de proveernos con las bases para un entendimiento global de la genética,

la bioquímica, la fisiología y la patogénesis de un microorganismo dado. Como el número de genomas secuenciados se incrementa anualmente, es posible ahora comparar un grupo significativo de secuencias genómicas entre bacterias relacionadas evolutivamente. En particular, el análisis de variabilidad genética entre un patógeno y especies no patogénicas relacionadas. Este análisis genera rápidamente una colección de genes potencialmente responsables de la adquisición de virulencia y por tanto con implicaciones prácticas en el diseño de vacunas. Primeramente la genómica comparada nos da una guía en la búsqueda de proteínas para uso en vacunas de subunidad. Por otro lado, nos da una base racional para la atenuación estable y segura de vacunas vivas o vectores vivos para ser usados como portador de un candidato vacunal.

La comparación de los genomas de cepas patogénicas de *E. coli* con la cepa de laboratorio K-12 mostró evidencias de una estructura en mosaico de estos genomas. Es decir, el cromosoma tiene una estructura básica común interrumpida por múltiples segmentos de ADN portando diferentes genes codificando para factores de virulencia. Diferentes combinaciones de estas “islas” de genes confiere a cada cepa de *E. coli* sus características fenotípicas y sus potencialidades patogénicas (19,20). La virulencia de muchos patógenos está frecuentemente correlacionada con la presencia de estos segmentos de ADN comúnmente llamados “islas de patogenicidad”. Usualmente estos segmentos son adquiridos por transferencia horizontal de genes y están ausentes en las especies no patogénicas (21).

La genómica comparada también brinda importante información sobre la evolución de la patogénesis bacteriana. Nos descifra una serie de mecanismos diferentes que el microorganismo emplea para cambiar su apariencia antigénica. Como ejemplo tenemos, la variación de fase (22), la duplicación génica (23), y la tendencia a perder funciones rudimentales (24, 25).

Recientemente se ha enfatizado en la secuenciación genómica de organismos estrechamente relacionados (26), lo que puede ser usado en principio para monitorear la variabilidad génica dentro de una especie. En particular, algunos ejemplos de comparación genómica entre aislamientos clínicos de un mismo patógeno han sido reportados (27, 28). El análisis comparativo de los genomas de *H. pylori* cepa J99, aislada en Estados Unidos en 1994 de un paciente con úlcera duodenal, y de la cepa 26695, aislada en Inglaterra antes de 1997 de un paciente con gastritis, reveló un importante número de genes cepa específicos (27). Estos genes pudieran estar involucrados en la severidad de la infección por *H. pylori*.

La transferencia horizontal de genes es generalmente un fenómeno raro en el caso de las bacterias intracelulares como *Mycobacterium* spp. y Chlamydiaceae. En principio, la gran estabilidad de los genomas de las bacterias intracelulares aumenta las posibilidades de identificar proteínas blancas para ser usadas en una droga o una vacuna de amplio espectro. Sin embargo, la comparación de los genomas de dos cepas de *M. tuberculosis* (cepa 1551 virulenta aislada de paciente y cepa H37Rv adaptada al laboratorio), mostró más variabilidad que la que se esperaba inicialmente (28, 29). Tales divergencias pudieran tener importantes implicaciones para la patogénesis, la inmunidad y el diseño de vacunas, ya que estas

proteínas parecen estar implicadas en la interacción con el hospedero. El Instituto Sanger de genómica se encuentra preparando la secuencia completa del genoma de *Mycobacterium bovis* BCG (98% anotado). La cepa atenuada de *M. bovis* BCG es el componente activo de la única vacuna existente contra la tuberculosis. Este resultado junto con el ya publicado de genómica funcional comparada en BCG (30), nos revelarán todos los mecanismos involucrados en la atenuación de esta cepa vacunal y adicionará pistas al diseño de una vacuna más racional y efectiva contra la tuberculosis. En el Instituto Finlay se trabaja en la comparación genómica de especies virulentas y no virulentas dentro del género *Mycobacterium*.

### El ejemplo del meningococo

El primer organismo vivo en donde la genómica fue usada para la identificación de candidatos vacunales potenciales fue *N. meningitidis* serogrupo B (31). La compañía de vacunas Chiron en colaboración con el Instituto de Investigaciones Genómicas (TIGR) de los Estados Unidos, secuenció el genoma completo del meningococo de la cepa virulenta MC58 del serogrupo B (32). Durante los meses que duró la secuenciación de este genoma, y mientras se iban descifrando las secuencias nucleotídicas de todos los genes, un grupo de bioinformáticos detectó más de 600 candidatos vacunales mediante análisis *in silico* (Figura 2). De estos candidatos 350 fueron clonados y expresados en *E. coli* y posteriormente usados para inmunizar ratones (31). El antisuero obtenido fue evaluado mediante ELISA de células enteras y FACS para verificar que estas proteínas eran antígenos expuestos en la superficie del meningococo. Adicionalmente, a estos sueros se les determinó la actividad bactericida, parámetro que correlaciona con la protección en humanos (33). Finalmente 91 nuevos antígenos de superficie fueron identificados, de los cuales 29 indujeron anticuerpos bactericidas. Después de evaluar la variabilidad de las secuencias de estos candidatos entre numerosos aislamientos clínicos el número se redujo a siete (Figura 2).

En la Figura 2 se puede apreciar la reducción del número total de genes de un genoma a pocos candidatos vacunales usando metodologías *in silico*. También se aprecia la importancia de la confirmación *in vitro* de estos resultados de predicción y el uso de modelos animales para evaluar la inmunogenicidad de los candidatos. Datos tomados del trabajo publicado por Pizza y colaboradores (31).

A estos antígenos detectados a partir de la secuencia genómica del meningococo se les ha denominado con las siglas GNA (del inglés *Genome Derived Antigens*). Hasta la fecha, los GNA candidatos vacunales más promisorios han sido: GNA33, una proteína relacionada con el metabolismo del peptidoglicano y la virulencia (34); NadA, una proteína de superficie involucrada en la adherencia (35); y dos lipoproteínas GNA2132 (36) y GNA1870.

GNA1870 es una proteína de superficie y es considerada el candidato más atractivo para el desarrollo de una nueva vacuna antimeningocócica, ya que es capaz de inducir inmunidad protectora en animales de experimentación (37, 38). Esta proteína ha sido expresada a altos niveles en el meningococo para producir una

vacuna basada en vesículas de membrana externa (VME) donde este antígeno esté más representado (39). Este resultado demuestra las potencialidades de la genómica en el diseño y mejoramiento de las vacunas. La enfermedad meningocócica por el serogrupo B es un problema de salud en muchos países y la única vacuna existente basada en VME (VA-MENGOC-BC®) ha sido desarrollada por el Instituto Finlay (40). Esta vacuna ha mostrado una eficacia satisfactoria en niños reduciendo la tasa de mortalidad y morbilidad por *N. meningitidis* B en Cuba desde su aplicación (41). En nuestro país, en el Centro de Ingeniería Genética y Biotecnología (CIGB) en colaboración con el Instituto Finlay, se trabaja en la identificación de nuevos candidatos a partir de estudios de Genómica y Proteómica de la cepa vacunal cubana del meningococo. En el departamento que estudia la enfermedad meningocócica en el CIGB se evalúan diferentes composiciones basadas en nuevos antígenos identificados a partir de la secuencia genómica.

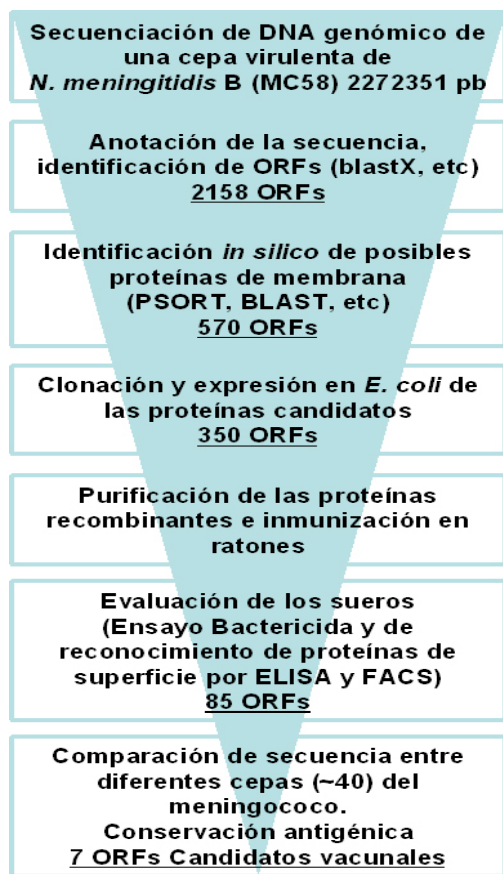


Figura 2. Diagrama que muestra el principio de la vacunología inversa usando el ejemplo del meningococo serogrupo B.

Después del éxito logrado con el meningococo, otros grupos han usado esta metodología para identificar candidatos vacunales contra algunos de los principales patógenos humanos. Ejemplos de bacterias que han sido estudiadas usando la vacunología inversa son: *Bacillus anthracis* (42), *S. pneumoniae* (43), *S. aureus* (44), *Chlamydia*

*pneumoniae* (45;46), *Porphyromonas gingivalis* (47), *Edwardsiella tarda* (48) y *M. tuberculosis* (49).

## Conclusiones

La investigación en vacunas entró en una nueva era cuando la primera secuencia de un genoma fue publicada en 1995. Actualmente el proceso de la secuenciación de un genoma se ha acelerado grandemente y alrededor de 4 millones de pares de bases pueden ser procesadas en sólo 2 días. Contamos ahora con una valiosa información, muchas veces disponible para todos, que debemos saber utilizar en el diseño racional de nuevos fármacos para prevenir o curar enfermedades. Existen ya vacunas en ensayos clínicos basadas en nuevos antígenos descubiertos mediante minería de genomas. Dentro de pocos años se espera una nueva generación de vacunas compuestas por proteínas derivadas de genomas o por epítomos T y B predichos *in silico*. Con este trabajo intentamos hacer converger dos novedosas disciplinas en el campo de la vacunología: la genómica microbiana y la biología computacional, con el objetivo de dar a conocer el estado del arte en el diseño de vacunas en la era posgenómica.

## Agradecimientos

Quisiera agradecer a Rolando Pajón, jefe del Departamento de Enfermedad Meningocócica del CIGB, por la lectura crítica de este documento y sus aportes valiosos al mismo.

## Referencias

- Hardy E, Martínez E, Diago D, Díaz R, González D, Herrera L. Large-scale production of recombinant hepatitis B surface antigen from *Pichia pastoris*. *J Biotechnol* 2000; 77(2-3):157-167.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269(5223):496-512.
- Cordwell SJ. Exploring and exploiting bacterial proteomes. *Methods Mol Biol* 2004; 266:115-135.
- Clayton RA, White O, Fraser CM. Findings emerging from complete microbial genome sequences. *Curr Opin Microbiol* 1998; 1(5):562-566.
- De Groot AS, Bosma A, Chinai N, Frost J, Jesdale BM, Gonzalez MA et al. From genome to vaccine: *in silico* predictions, *ex vivo* verification. *Vaccine* 2001; 19(31):4385-4395.
- Zagursky RJ, Olmsted SB, Russell DP, Wooters JL. Bioinformatics: how it is being used to identify bacterial vaccine candidates. *Expert Rev Vaccines* 2003; 2(3):417-436.
- Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 2001; 19(17-19):2688-2691.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25(17):3389-3402.
- Camus JC, Pryor MJ, Medigue C, Cole ST. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 2002; 148(Pt 10):2967-2973.
- Chakravarti DN, Fiske MJ, Fletcher LD, Zagursky RJ. Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* 2000; 19(6):601-612.

11. Antelmann H, Tjalsma H, Voigt B, Ohlmeier S, Bron S, van Dijl JM et al. A proteomic view on genome-based signal peptide predictions. *Genome Res* 2001; 11(9):1484-1502.
12. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; 340(4):783-795.
13. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gramnegative bacteria. *Protein Sci* 2003; 12(8):1652-1662.
14. Rey S, Acab M, Gardy JL, Laird MR, deFays K, Lambert C et al. PSORTdb: a protein subcellular localization database for bacteria. *Nucleic Acids Res* 2005; 33 (Database issue):D164-D168.
15. Chen CP, Rost B. State-of-the-art in membrane protein prediction. *Appl Bioinformatics* 2002; 1(1):21-35.
16. Wimley WC. Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures. *Protein Sci* 2002; 11(2):301-312.
17. Chen Y, Yu P, Luo J, Jiang Y. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome* 2003; 14(12):859-865.
18. Reche PA, Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 2004; 56(6):405-419.
19. Welch RA, Burland V, Plunkett G, III, Redford P, Roesch P, Rasko D et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 2002; 99(26):17020-17024.
20. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 2001; 8(1):11-22.
21. Hacker J, Blum-Oehler G, Hochhut B, Dobrindt U. The molecular basis of infectious diseases: pathogenicity islands and other mobile genetic elements. A review. *Acta Microbiol Immunol Hung* 2003; 50(4):321-330.
22. Snyder LA, Butcher SA, Saunders NJ. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria spp.* *Microbiology* 2001; 147(Pt 8):2321-2332.
23. Alm RA, Bina J, Andrews BM, Doig P, Hancock RE, Trust TJ. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect Immun* 2000; 68(7):4155-4168.
24. Preston A, Parkhill J, Maskell DJ. The bordetellae: lessons from genomics. *Nat Rev Microbiol* 2004; 2(5):379-390.
25. Sakharkar KR, Dhar PK, Chow VT. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *Int J Syst Evol Microbiol* 2004; 54(Pt 6):1937-1941.
26. Whittam TS, Bumbaugh AC. Inferences from whole-genome sequences of bacterial pathogens. *Curr Opin Genet Dev* 2002; 12(6):719-725.
27. Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 1999; 397(6715):176-180.
28. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* 2002; 184(19):5479-5490.
29. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; 393(6685):537-544.
30. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999; 284(5419):1520-1523.
31. Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 2000; 287(5459):1816-1820.
32. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA et al. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 2000; 287(5459):1809-1815.
33. Holst J, Feiring B, Fuglesang JE, Hoiby EA, Nokleby H, Aaberge IS et al. Serum bactericidal activity correlates with the vaccine efficacy of outer membrane vesicle vaccines against *Neisseria meningitidis* serogroup B disease. *Vaccine* 2003; 21(7-8):734-737.
34. Adu-Bobie J, Lupetti P, Brunelli B, Granoff D, Norais N, Ferrari G et al. GNA33 of *Neisseria meningitidis* is a lipoprotein required for cell separation, membrane architecture, and virulence. *Infect Immun* 2004; 72(4):1914-1919.
35. Comanducci M, Bambini S, Brunelli B, Adu-Bobie J, Arico B, Capecchi B et al. NadA, a novel vaccine candidate of *Neisseria meningitidis*. *J Exp Med* 2002; 195(11):1445-1454.
36. Welsch JA, Moe GR, Rossi R, Adu-Bobie J, Rappuoli R, Granoff DM. Antibody to genome-derived neisserial antigen 2132, a *Neisseria meningitidis* candidate vaccine, confers protection against bacteremia in the absence of complement-mediated bactericidal activity. *J Infect Dis* 2003; 188(11):1730-1740.
37. Masignani V, Comanducci M, Giuliani MM, Bambini S, Adu-Bobie J, Arico B et al. Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870. *J Exp Med* 2003; 197(6):789-799.
38. Fletcher LD, Bernfield L, Barniak V, Farley JE, Howell A, Knauf M et al. Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein. *Infect Immun* 2004; 72(4):2088-2100.
39. Hou VC, Koeberling O, Welsch JA, Granoff DM. Protective antibody responses elicited by a meningococcal outer membrane vesicle vaccine with overexpressed genome-derived neisserial antigen 1870. *J Infect Dis* 2005; 192(4):580-590.
40. Sierra GV, Campa HC, Varcácel NM, García IL, Izquierdo PL, Sotolongo PF et al. Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba. *NIPH Ann* 1991; 14(2):195-207.
41. Rico CO, Pereira CC, Fernández AA. [The post-licensing efficacy of VA-MENGOC-BC in children under 6 in Holguin, Cuba. The first year of observation]. *Rev Cubana Med Trop* 1995; 47(1):59-64.
42. Ariel N, Zvi A, Grosfeld H, Gat O, Inbar Y, Velan B et al. Search for potential vaccine candidate open reading frames in the *Bacillus anthracis* virulence plasmid pXO1: in silico and in vitro screening. *Infect Immun* 2002; 70(12):6817-6827.
43. Wizemann TM, Heinrichs JH, Adamou JE, Erwin AL, Kunsch C, Choi GH et al. Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect Immun* 2001; 69(3):1593-1598.
44. Etz H, Minh DB, Henics T, Dryla A, Winkler B, Triska C et al. Identification of in vivo expressed vaccine candidate antigens from *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 2002; 99(10):6573-6578.
45. Capo S, Nuti S, Scarselli M, Tavarini S, Montigiani S, Mori E et al. *Chlamydia pneumoniae* genome sequence analysis and identification of

- HLA-A2-restricted CD8+ T cell epitopes recognized by infection-primed T cells. *Vaccine* 2005; 23(42):5028-5037.
46. Stephens RS, Lammel CJ. Chlamydia outer membrane protein discovery using genomics. *Curr Opin Microbiol* 2001; 4(1):16-20.
47. Ross BC, Czajkowski L, Hocking D, Margetts M, Webb E, Rothel L et al. Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. *Vaccine* 2001; 19(30):4135-4142.
48. Srinivasa Rao PS, Lim TM, Leung KY. Functional genomics approach to the identification of virulence genes involved in *Edwardsiella tarda* pathogenesis. *Infect Immun* 2003; 71(3):1343-1351.
49. Betts JC. Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. *IUBMB Life* 2002; 53(4-5):239-242.

## Identifying Vaccine Candidates from the Sequence of a Bacterial Genome

### Abstract

Whole genome sequence data are increasingly available for a wide range of human pathogens. The availability of these sequences has changed our approach to vaccine development entirely and introduced a new way of thinking in this process. This approach starts from the genomic sequence and, by computer analysis, predicts those antigens that are most likely to be vaccine candidates. The use of bioinformatic tools allows the comprehensive *in silico* screening of genome data for surface-expressed proteins, in order to identify candidate vaccine antigens. *In vitro* confirmation of surface location and the use of animal models to test immunogenicity further refine the list of proteins likely to be of use as vaccine antigens. This process, first applied to serogroup B *Neisseria meningitidis*, has been termed as reverse vaccinology. Genomics also enables valuable information concerning the biology and virulence of the pathogenic species to be extracted by comparative genomics. In this review, we describe how genomic approaches can be used to identify novel vaccine candidates.

**Keywords:** Genomics, Vaccines, Bioinformatics, Reverse Vaccinology, *Neisseria meningitidis*.



# SLIPE

SOCIEDAD LATINOAMERICANA DE INFECTOLOGÍA PEDIÁTRICA.  
FILIAL CUBA

### Sexta Jornada Internacional de Infectología Pediátrica

20 al 23 de Septiembre del 2006  
Hotel Meliá Habana  
Ciudad de La Habana, Cuba

#### Auspician:

Ministerio de Salud Pública, Consejo Nacional de Sociedades Científicas de la Salud, Instituto Finlay, Sociedad Latinoamericana de Infectología Pediátrica.

#### Temas principales:

► Actualidad en Vacunología; ► Dengue en las Américas; ► Resistencia Bacteriana. ¿Estamos a tiempo de controlarla?; ► SIDA Pediátrico. La experiencia cubana; ► Actualidad en las infecciones del recién nacido; ► Otitis Media. Debate sobre diagnóstico y manejo terapéutico; ► Micosis Profundas; ► Vigilancia y Control de la Infecciones Nosocomiales; ► Virus respiratorios emergentes; ► Sepsis. Bases inmunológicas y mejor manejo terapéutico; ► Infecciones oportunistas e inmunosupresión.

#### Modalidades de Presentación:

Simposios, mesas redondas, conferencias, videos y carteles: (90x 1,20 cm, orientación Vertical). Los mismos serán designados por el Comité Científico. Los resúmenes para carteles se aceptarán siempre y cuando cumplan con la temática del evento y previa aprobación por el Comité Científico. La fecha de entrega de resúmenes se extenderá desde el 1ro de febrero hasta el 30 de junio de 2006.

Los interesados en presentar sus trabajos enviarán: nombre y apellidos de los autores, institución y título del trabajo a la Presidenta del Comité Científico:

Dra. Ileana Álvarez Lam. **E-mail:** [arlert@infomed.sld.cu](mailto:arlert@infomed.sld.cu)

#### Contacto Comité Organizador:

Dra. Mabel González Alemán (Presidenta)

**E-mail:** [mabel.aleman@infomed.sld.cu](mailto:mabel.aleman@infomed.sld.cu)