

BIOINFORMÁTICA: HISTORIA Y PERSPECTIVAS FUTURAS

Por: **Emiliano Barreto Hernández** M.Sc.

Profesor Asociado

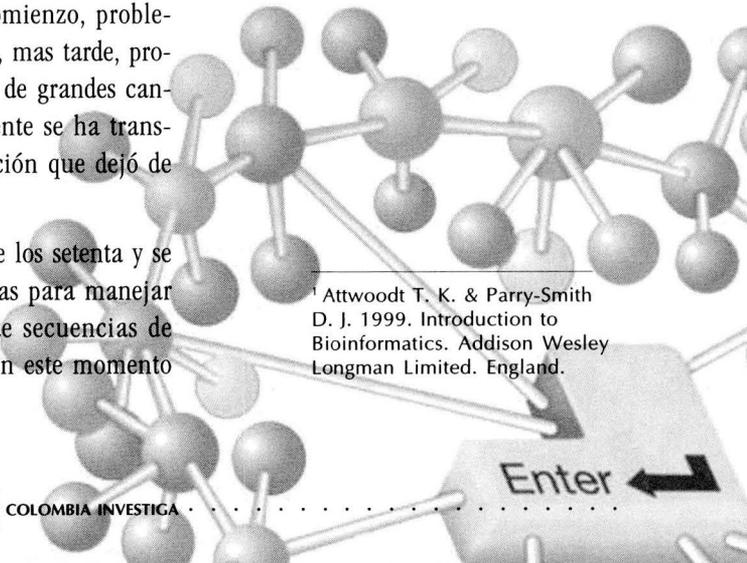
Director Grupo de Bioinformática. Instituto de Biotecnología
Universidad Nacional de Colombia - Sede de Bogotá

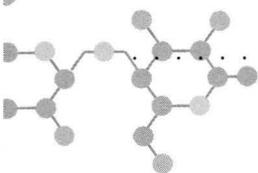
El término bioinformática tiene sus orígenes en la década de los setenta y se relacionaba con las aplicaciones computacionales diseñadas para manejar y analizar secuencias de origen biológico, principalmente secuencias de proteínas o del ADN proveniente de diversos organismos. En la actualidad, se extiende a áreas muy variadas que van desde tareas simples, como la adquisición directa de los datos a partir de los ensayos de secuenciación de DNA o de proteínas cuando se utilizan técnicas como espectrofotometría de masas, hasta el desarrollo de software para el almacenamiento y análisis de los datos, lo que implica en muchos casos el desarrollo de algoritmos que requieren de un conocimiento tanto matemáticos, como biológicos.

La bioinformática es una disciplina que surge de la interacción entre físicos, químicos, estadísticos, matemáticos y biólogos para solucionar, en un comienzo, problemas de cálculo con información biológica, mas tarde, problemas con el almacenamiento y análisis de grandes cantidades de datos extraídos de organismos y más recientemente se ha transformado en una disciplina con líneas propias de investigación que dejó de ser solo una herramienta para los biólogos.¹

El término bioinformática tiene sus orígenes en la década de los setenta y se relacionaba con las aplicaciones computacionales diseñadas para manejar y analizar secuencias de origen biológico, principalmente secuencias de proteínas o del ADN proveniente de diversos organismos. En este momento

¹ Attwood T. K. & Parry-Smith D. J. 1999. Introduction to Bioinformatics. Addison Wesley Longman Limited. England.





es difícil contar con una definición precisa de la bioinformática, quizás debido a que es una disciplina en crecimiento y plena diversificación. Algunas definiciones que se pueden encontrar en algunos libros de texto y en Internet son las siguientes:

“La bioinformática comprende los métodos matemáticos, estadísticos y computacionales que pretenden solucionar problemas biológicos usando secuencias de ADN y aminoácidos e información relacionada” (Tekai, citado por Counsell, 2002).

“La bioinformática se define como el uso de los computadores para almacenar, recuperar, analizar o predecir la composición o la estructura de las biomoléculas. Esas moléculas incluyen nuestro material genético (Ácidos nucleicos) y los productos de nuestros genes (proteínas)”.²

Sin embargo, la bioinformática o biología computacional en la actualidad no sólo se restringe al análisis molecular, sino que se puede extender a áreas como la enseñanza de las ciencias biológicas; al análisis estadístico aplicado a la experimentación en biología, a los sistemas para prueba y desarrollo de nuevos fármacos. Todas estas aplicaciones de la informática en biología han dado como resultado que el término bioinformática se asocie al conjunto de las aplicaciones para computador diseñadas para las ciencias biológicas.

A pesar de que no se cuenta con una única definición de la disciplina, esta si cuenta con una historia mas o menos bien entendida, siendo notable que buena parte de los

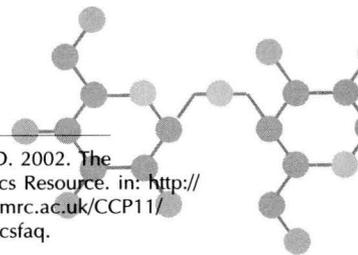
desarrollos están asociados con los avances de la biología molecular. Eventos históricos como el descubrimiento de la doble hélice de la estructura química del ADN³, la secuenciación de la primera proteína, la insulina bovina⁴, la secuenciación del primer ácido nucleico, el tRNA alanina de levadura,⁵ dieron paso a la aparición de esquemas de sistematización como el “Atlas de secuencia y estructura de proteínas”⁶ que recabó la información disponible hasta ese momento; el Protein Data Bank (PDB) en 1972 uno de los primeros sistemas de almacenamiento electrónico⁷ y el desarrollo del SWISSPROT que es una de las bases de datos con información estructural sobre proteínas mas completa y una de las más consultadas a través de Internet.⁸

En sus inicios y hasta aproximadamente mediados de los ochentas, la bioinformática tuvo una fuerte orientación a las aplicaciones, a la publicación de nuevos programas para enfrentar viejos y nuevos problemas y al diseño de bases de datos y sistemas de búsqueda en estas, enfoque que favoreció su desarrollo y de la misma forma el enorme crecimiento de los datos almacenados obtenidos de experimentos de biología molecular.

La necesidad de depósitos centrales de datos donde los investigadores registraran, compararan y obtuvieran las secuencias relacionadas con sus proyectos, de forma que se evitara la duplicación de esfuerzos y se contribuyera a avanzar en el conocimiento, llevo en la década de los 80s a la creación del GenBank, el La-

boratorio Europeo de Biología Molecular EMBL y el banco de datos de ADN del Japón DDBJ, los cuales por medio de la *International Nucleotide Sequence Database Collaboration*, mantienen unificada la base de datos de secuencias reportadas de ADN a través del intercambio y actualización diaria la de información.

El crecimiento exponencial de los datos, que han pasado de 606 secuencias de ADN almacenadas en 1982 a unos 17.5 millones hoy en día (Fig.1), fue impulsado por el desarrollo de la técnica “Reacción en Cadena de la Polimerasa” (PCR) en el



² Counsell, D. 2002. The Bioinformatics Resource. in: <http://www.hgmp.mrc.ac.uk/CCP11/bioinformaticsfaq>.

³ Watson, J. & Crick, F. 1953. Molecular Structure of Nucleic Acids, A Structure for Deoxyribose Nucleic Acid. *Nature* 171:737-738.

⁴ Sanger, F. 1956. The structure of insulin. *Currents in Biochemical Research* (ed. D.E. Green), Wiley Interscience, New York.

⁵ Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T. & Marguisee, M. 1965. The base sequence of yeast alanine transfer RNA. *Science*, 147: 1462-1465.

⁶ Dayhoff, M.O. 1966. Atlas of protein sequences and structure. National Biomedical Research Foundation, Silver Springs.

⁷ Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, D. F. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. 1977. Protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112, 535-542.

⁸ Bairoch, A. & Boeckmann, B. 1991. The SWISS-PROT protein sequence data bank. *Nucleic Acids Res.*, 19, 2247-2249 (Supplement).

año de 1986⁹ y por el mejoramiento instrumental de equipos como los secuenciadores automáticos de alto rendimiento, los cuales a su vez contribuyeron al desarrollo de proyectos de secuenciación como el del genoma humano¹⁰ y de otros organismos importantes en las áreas de la salud y la industria haciendo que no solo se requirieran sistemas eficientes de almacenamiento, sino nuevos modelos para la comparación de secuencias, más rápidos y estadísticamente confiables como los algoritmos BLAST¹¹ y FAST.¹²

A mediados de la década de los noventa, la bioinformática dio un giro en su enfoque, cuando deja de estar orientada a las aplicaciones y perfilarse más como una disciplina investigativa, en la que se utilizan las he-

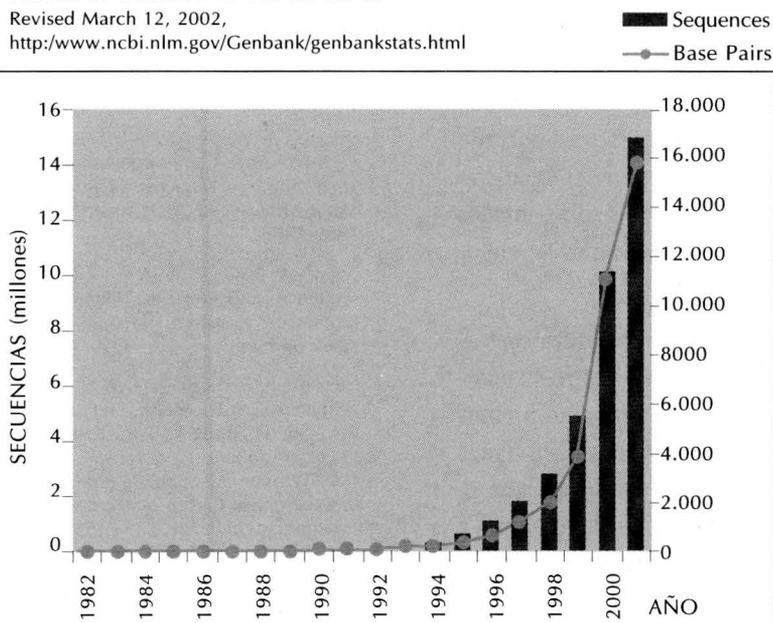
rramientas computacionales para realizar estudios en áreas como la organización de los genomas, el “descubrimiento” de genes, la relación entre las mutaciones y la alteración de la función biológica ó la evolución de la diversidad de los diferentes organismos, entre muchas otras. Adquirió un enfoque de mayor pensamiento biológico dirigido a responder preguntas de esta índole no sólo en cuanto a la manera en que se manejan, guardan y analizan grandes cantidades de datos¹³ sino impulsado el desarrollo de disciplinas como la genómica, la proteómica y la filogenia.

Con el cambio de enfoque de la bioinformática, se estableció el perfil del profesional dedicado a esta disciplina quien debe tener destrezas en

el manejo de diversos sistemas operativos, además de conocimientos profundos sobre bioquímica y biología molecular, con capacidad de interactuar en grupos interdisciplinarios dada la complejidad de los temas que se tratan y que requieren de la intervención de varias disciplinas.¹⁴

No obstante, existe una vieja discusión sobre si los investigadores dedicados a la bioinformática deben ser individuos altamente entrenados en el desarrollo y manejo de herramientas computacionales en búsqueda de problemas biológicos por resolver ó son investigadores con preguntas biológicas que, a través de un entrenamiento en el manejo de herramientas computacionales puede encontrar las respuestas.¹⁵

FIGURA 1. CRECIMIENTO DEL GENBANK.
Revised March 12, 2002,
<http://www.ncbi.nlm.gov/Genbank/genbankstats.html>



⁹ Mullis, K. B. 1990. Target Amplification for DNA analysis by the polymerase chain reaction. *Ann. Biol. Clin.* 48(8): 579-82.

¹⁰ The International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

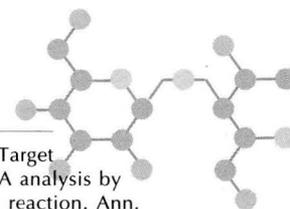
¹¹ Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. "Basic local alignment search tool." *J. Mol. Biol.* 215 403-410.

¹² Pearson, W. R. & Lipman, D. J. 1988. Improved Tools for Biological Sequence Analysis. *Proc. Nat. Acad. Sci.* 85:2444 -2448.

¹³ Attwood T. K. & Parry-Smith D. J. 1999. *Op cit.*

¹⁴ Tisdall, J. D. 2001. *Beginning Perl for Bioinformatics.* O'Reilly & Associates Inc. USA.

¹⁵ Pearson, W. 2000. Training for bioinformatics and computational biology. *Bioinformatics*, 17(9):761-762.



Al dar una mirada a los investigadores que trabajan en bioinformática estos, en su gran mayoría, son biólogos, químicos, bioquímicos y médicos que han tenido que aprender por sí mismos elementos de informática y computación para poder satisfacer las necesidades de desarrollo o usar herramientas de software o bases de datos que se presentan en las investigaciones que se realizan en el área biológica. Sin embargo, como consecuencia de los avances y la importancia del área, en los últimos años se han vinculado cada vez más profesionales en informática, computación, matemáticas y estadística que aportan su experiencia a los equipos multidisciplinarios que conforman los grandes centros de bioinformática.

Durante los últimos 15 años las necesidades y el crecimiento tanto en número como en conocimiento en el área de la bioinformática, llevaron a la realización de congresos (Conferencia de la Sociedad Internacional de Biología Computacional ISCB, Conferencia Internacional de Investigación sobre Biología Molecular Computacional RECOMB, etc.), a la edición de revistas seriadas (Journal of Computational Biology, Cambios, Bioinformatics, etc.), y, con el desarrollo del Internet, la puesta en funcionamiento de servidores www, como ExPASy (Expert Protein Analysis System) (www.expasy.org)¹⁶, uno de los primeros dedicado a las ciencias biológicas.

La Bioinformática ha evolucionado hacia la organización de los investigadores y grupos en grandes cen-

tros dedicados al desarrollo del área, como: el Centro Nacional para la Información en Biotecnología (NCBI), el Instituto Europeo de Bioinformática (EBI), el Instituto Suizo de Bioinformática (SIB), el Centro de Bioinformática de Pekín (CBI), etc., que se conformaron para reunir las experiencias y conocimientos de los especialistas en bioinformática, de tal forma que esta capacidad acumulada, sirviera no solo para hacer mejores desarrollos, sino que se pudiera dar asesoría y capacitación a los investigadores que requieran de la utilización de herramientas de la bioinformática.

La experiencia de la creación de los centros, como la utilización de las redes de telecomunicación y las necesidades de servicios especializados y, de capacitación y asesoría de los investigadores del área de la biología molecular, llevaron a la conformación de grandes redes de carácter internacional como la APBioNet (red de bioinformática Asia – Pacífico), la EMBnet (red europea de biología molecular) o ICCBnet (red centro internacional para la cooperación en bioinformática auspiciada por la UNESCO), que mantienen interconectados a los expertos de los diferentes centros, y que garantizan la disponibilidad a través de la red y al dominio público de todos los datos, experiencias y recursos informáticos que se generan día a día logrando un verdadero aporte a la socialización del conocimiento.

ÁREAS DE APLICACIÓN

Las áreas en las que se desenvuelve actualmente la bioinformática son muchas y muy variadas, van desde tareas simples como la adquisición directa de los datos a partir de los ensayos de secuenciación de DNA o de proteínas, (cuando se utilizan técnicas como espectrofotometría de masas), hasta el desarrollo de software para el almacenamiento y análisis de los datos, lo que implica en muchos casos el desarrollo de algoritmos que requieren tanto de un conocimiento matemáticos y biológico.

Dentro de las áreas en las que se desenvuelve la bioinformática están, la genómica, la proteómica, la farmacogenética y la filogenia.

La genómica se ocupa de la caracterización molecular de genomas completos. En esta área la bioinformática se ha enfocado en la caracterización y localización de las secuencias que conforman el ADN (genómica estructural) y, en la recolección sistemática de información sobre la función de los genes (genómica funcional).¹⁷ El objeto último es tratar de llenar los vacíos de conocimiento existentes entre la secuencia de un gen y su función, pasando de la

¹⁶ Appel, R. D., Bairoch A. & Hochstrasser, D. F. 1994. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server, Trends Biochem. Sci. 19:258-260.

¹⁷ Cowley, A., W. 1998. The Emergence of Physiological Genomics. INABIS 98 - 5th Internet World Congress on Biomedical Sciences., Canadá.

evaluación individual de los genes, y coadyuda al estudio de todos los genes expresados por una célula en un momento determinado (transcriptoma).

Dentro de este campo algunas de las tareas más importantes que se desarrollan son las siguientes:

Captura y análisis preliminar de datos de secuencia y patrones de bandas, a través de la interdependencia entre instrumentos como secuenciadores de ADN, capaces de capturar imágenes, con software especializado que permite la depuración (eliminando señales de ruido y contaminación), análisis y almacenamiento electrónico de las secuencias o de los patrones de bandas obtenidos.

Comparación y búsqueda rápida de secuencias, fundamental para encontrar, en las bases de datos posibles, secuencias relacionadas con una secuencia problema, usando algoritmos de comparación como BLAST¹⁸ o FASTA¹⁹ y hardware especializado especialmente diseñado para este propósito.

Ensamblaje de secuencias de ADN, a partir de fragmentos de 100 a 300 pares de bases que se deben solaparse y en los que se deben detectar los errores en el proceso de secuenciación y las secuencias repetidas, entre otras. Este proceso es especialmente útil en la secuenciación de cADN o como paso obligado en los diferentes métodos de secuenciación parcial o total de genomas.

Detección de secuencias codificantes, como genes, intrones, exo-

nes, motivos, patrones, en general, obtenidas a través de herramientas de cómputo basadas en acumulación, sistematización y comparación de secuencias con funciones conocidas, cuyas características estructurales se han podido determinar usando alineamiento múltiple de secuencias y métodos estadísticos como los modelos ocultos de Markov. Lo anterior es importante porque ayuda a la implementación de sistemas predictivos de alta confiabilidad, cuyos resultados son utilizados por programas especializados en esta materia como: VEIL (Localizador de Intrones-Exones) usando un modelo de Markov (HMM) para encontrar genes en ADN eucariótico (<http://www.cs.jhu.edu/labs/compbio/veil.html>); MORGAN sistema integrado para localización de genes en secuencias de ADN de vertebrados, basándose en árboles de decisión, cadenas de Markov para reconocer sitios SPLICE y programación dinámica (<http://www.cs.jhu.edu/labs/compbio/morgan.html>);

GENSCAN programa diseñado para predecir estructuras genéticas completas, incluyendo exones, intrones, promotores y señales de poliadenilación en secuencias geonómicas (<http://genes.mit.edu/GENSCAN.html>).

Análisis de secuencias de ADN o ARN, permite la detección de regiones de baja complejidad; secuencias repetidas como microsatélites, minisatélites y, de secuencias no codificantes como señales de transcripción (elementos promotores), orígenes de replicación, etc. La predicción

y comparación de la estructura secundaria de tARN, rARN y partes de mARN (Fig. 2), hacen uso de métodos como la maximización en el apareamiento de bases, minimización de energía local o un análisis comparativo de las secuencias entre otros, es muy útil en la inferencia de relaciones filogenéticas, dado que contribuyen al conocimiento de los patrones de evolución molecular.

La comparación entre genomas, es recientemente una de las aplicaciones que requiere nuevas herramientas. Permite el análisis de las diferencias y similitudes reales entre los diversos organismos a los que se les ha secuenciado su genoma o aun mejor entre genomas de una misma especie.

Es así como la comparación entre el genoma de una cepa no patógena y otra patógena de *Escherichia Coli*, permitió identificar secuencias y plásmidos con capacidad infecciosa que no estaban presentes en la cepa no patógena, e identifico la transferencia horizontal de genes, que se relacionan con la patogenicidad.²⁰

La secuenciación de genomas microbianos, ha permitido la identi-

¹⁸ Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990.

¹⁹ Pearson, W. R., & Lipman, D. J. 1988. Op cit.

²⁰ Saier, M. H. Jr. 1998. Genome sequencing and informatics: New tools for biochemical discoveries. *Plant Physiol.* 117: 1129-1133.

ficación de nuevas proteínas, como la identificada en *Bacillus subtilis*, que es un péptido que controla la iniciación de la esporulación de tal forma que en varios microorganismos se han identificado familias de proteínas relacionadas con los procesos patogénicos o de resistencia a condiciones ambientales adversas. De otro lado, la bioinformática permite identificar nuevas vías metabólicas o vías no conocidas que los microorganismos tenían la capacidad de utilizar.

El conocimiento del genoma refleja el estilo de vida de los organismos y desde esta óptica se pueden realizar deducciones ecológicas e inferencias y utilizar estos microorganismos para bioremediación.²¹

Con la publicación de la secuencia completa de algunos genomas, la comprensión de cómo el medio millón de proteína humanas están codificadas en por lo menos 30.000 genes, constituye un largo y duro camino por recorrer. Un nuevo concepto fundamental llamado proteóma (conjunto de proteínas que se producen en la célula en un momento y condición dados) ha emergido y con él la proteómica dedicada a la comparación cualitativa y cuantitativa de proteómas en los diferentes estadios y condiciones de los procesos biológicos.

La bioinformática no solo ha contribuido al análisis y agrupamiento de la proteínas expresadas en un momento dado, usando los datos obtenidos por nuevas técnicas instrumentales como los microarreglos,

sino que, comparten con los ácidos nucleicos un sin número de desarrollos y herramientas en lo relacionado con la búsqueda y comparación rápida de secuencias y la predicción de regiones codificantes.

El estudio de la funcionalidad de las proteínas ha impulsado el desarrollo de herramientas que a través del análisis de las secuencias y los correspondientes datos de actividad biológica (obtenidos experimentalmente), ha permitido agruparlas por funciones y familias de proteínas (bases de datos secundarias como PROSITE, BLOCKS, PRINTS, etc.), generando patrones, usando alineamiento múltiple de secuencias y estrategias

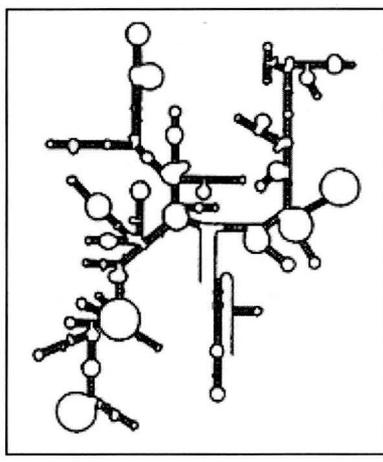


FIGURA 2. rARN DE 16S.

como los Modelos Ocultos de Markov utilizados, entre otros, para predecir la presencia de regiones topogénicas, transmembranales, antigénicas, repetidas o con posibilidad de tener sitios que sufran modificaciones post-traduccionales, en una secuencia problema.

Un aspecto muy importante en los estudios de funcionalidad de proteínas es el análisis y predicción de sus estructuras secundaria y terciaria, implementado métodos que utilizan en algunas aplicaciones técnicas estadística y en otras, inteligencia artificial a través del diseño e implementación las redes neuronales artificiales. Algunos de estos métodos están basados en predecir la estructura tridimensional directamente desde la secuencia problema, como es el caso del método Ab-initio²² que se apoya en el cálculo reiterado de parámetros asociados a la termodinámica de plegamiento de proteínas, buscando como resultado la secuencia que corresponde a su estado mínimo de energía libre (Fig. 3). También son muy utilizados los métodos de predicción que se basan en la búsqueda de similitudes de las secuencias problema con patrones bien estudiados presentes en las proteínas a las que se les ha determinado su estructura tridimensional.

El alineamiento múltiple de secuencia de proteína, ARN o ADN, los modelos de comparación de estructura tridimensionales o el análisis de las bases de datos de rutas metabólicas, han permitido avanzar en la reconstrucción de la evolución de las secuencias de ADN en sitios particulares del genoma, de familias de proteínas o de rutas metabólicas, con-

²¹ Saier, M. H. Jr. 1998. Op cit.

²² Sternberg M. 1997. Protein Structure Prediction A Practical Approach. Edited by Oxford University Press, London, Cap. 8.

tribuyendo a la creación de árboles taxonómicos, cada vez más confiables para la reconstrucción de la evolución de las especies vivientes, favoreciendo el desarrollo de las herramientas de bioinformática para el análisis filogenético.

Desde el punto de farmacológico, la bioinformática se enfoca en el descubrimiento y caracterización de los receptores de los fármacos, generalmente localizados en estructuras como ácidos nucleicos y proteínas. En este contexto se ha logrado mejorar su clasificación asociándolos con familias y subfamilias de secuencias bien conocidas, como las que se encuentran en bases de datos secundarias de proteínas.

La genómica y la proteómica junto con el conocimiento bioquímico que se ha alcanzado hasta ahora, están contribuyendo al entendimiento del funcionamiento y comportamiento bioquímico de los receptores, permitiendo el modelamiento predictivo de su funcionamiento en aspectos como los niveles de captación de fármaco y otros compuestos, intoxicación,

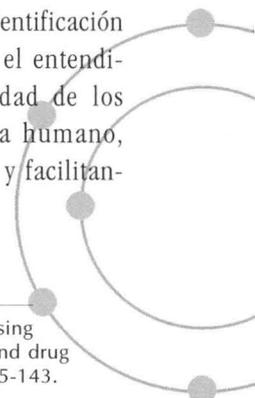
vida media, estratificación de las poblaciones de pacientes y otras variaciones asociadas a los genes.²³

El conocimiento generado sobre las estructuras primarias y tridimensionales de los receptores, ha permitido diseñar fármacos que pueden tener estructuras similares a las de las secuencias blanco (receptores) muchas veces más potentes y seguros predecir que variaciones del genoma influyen en la respuesta interindividual de la actividad de los fármacos (farmacogenética).

PERSPECTIVAS DE LA BIOINFORMÁTICA

La culminación del proyecto genoma humano, ha constituido el comienzo de una ardua labor de integración y análisis de los datos biológicos de todo tipo, que constantemente son almacenados en la multitud de bases de datos disponibles. La bioinformática enfrenta el reto de lograr una mayor armonización buscando la implementación de un lenguaje común para el manejo de los datos.²⁴

Una estrategia a utilizar podría ser la creación de un índice de genes en el cual cada proteína este asociada a la secuencia de nucleótidos de los que se deriva como lo plantea Jongeneel,²⁵ de tal forma que, en un futuro, los diseñadores de microarreglos podrían identificar las características de las secuencias de forma única y universal, usando un lenguaje común que ayudaría a establecer si las secuencias de genes están completas o incompletas. Este índice proporcionaría una base de datos de referencia única para la identificación de los genes, facilitaría el entendimiento de la complejidad de los transcritos del genoma humano, evitando la redundancia y facilitan-



²³ Searls, D. B. 2000. Using bioinformatics in gene and drug discovery. DDT 5(4): 135-143.

²⁴ Brazma, A. 2001. On the importance of standardization in life sciences. Bioinformatics, 17(2):113-114.

²⁵ Jongeneel, C. V. 2000. The Need for a human gene index Bioinformatics, 16(12):1059-1061.

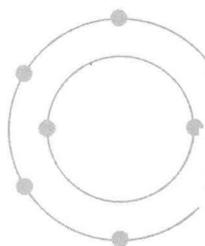
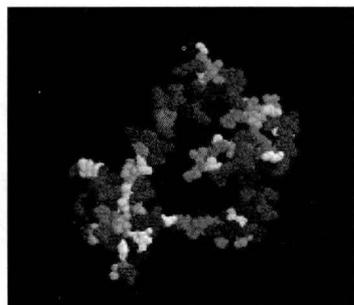
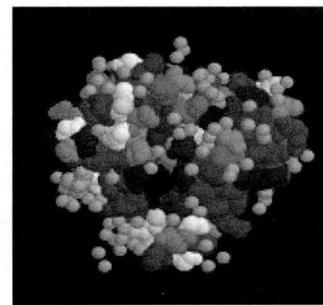


FIGURA 3. PREDICCIÓN DE ESTRUCTURA DE LA MIOGLOBINA DE ESPERMA DE BALLENA (153 RESIDUOS) USANDO EL MÉTODO AB-INITIO (SIMULACIÓN DE MONTE CARLO, ANILLAMIENTO Y ESTIMACIÓN DE LA ENERGÍA HIDROFÓNICA.



500 pasos de búsqueda



10.000 pasos de búsqueda

do la corrección de los datos disponibles.

Además de la indexación de los datos biológicos de toda índole, el mayor esfuerzo de los bioinformáticos estará dedicado al desarrollo y mejoramiento de los algoritmos que se han plasmado en las herramientas de cómputo que permitan la búsqueda, comparación y análisis de datos de diferentes tipos, además de estructurar modelos dinámicos que representen, no solo características puntuales del genoma o el proteoma, sino que permitan su integración con información bioquímica como rutas metabólicas, perfiles lipídicos, electrolitos, etc.

Otro campo en el que se asume un rápido desarrollo es el de comparación de genomas entre diferentes organismos, analizando integralmente sus similitudes y diferencias, avanzando así en el conocimiento de sus relaciones evolutivas.

Tomando como base la influencia de la teoría de probabilidades y la estadística en el desarrollo de las herramientas y en los criterios de decisión con los que cuenta la bioinformática hasta ahora, los nuevos desarrollos deberán incorporar aun más el pensamiento estadístico²⁶, dado que no sólo se tendrá una mayor cantidad de datos sino que estos serán de diferentes clases, representando diferentes momentos y condiciones en el ciclo celular e incluso diferentes organismos, como ocurre en los análisis de expresión génica donde la variación biológica hace difícil la identificación de las señales de información verdaderas.

En áreas específicas como la salud, por ejemplo, la información generada por la geonómica y la farmacogenética, será valiosa para la prevención, diagnóstico y el manejo de numerosas enfermedades cuya cura hasta la fecha ha sido difícil o imposible; enfermedades hereditarias, causantes de problemas de salud como las alteraciones cardiovasculares o el cáncer, entre otras, podrán ser previstas anticipadamente y tratadas, contrastando la información genética del paciente con los datos de condiciones de tratamiento específicas para sus características genotípicas.

La investigación del genoma de agentes patógenos, ayudará a comprender mejor la transmisión de las enfermedades, los mecanismos de su virulencia y la forma de evadir el sistema inmunológico del portador, contribuyendo al desarrollo de nuevas técnicas de diagnóstico, vacunas y agentes terapéuticos.

Un aspecto importante en el desarrollo futuro de la bioinformática es el comercial, dado que no solamente se considera como una área de desarrollo estratégico por los gobiernos de Estados Unidos y la Unión Europea, sino que la aparición de la industria genómica y el gran volumen de aplicación de las técnicas genómicas en la biología y la medicina, han creado un mercado de software, hardware y servicios bioinformáticos, estimado en \$2.0 billones de dólares para los próximos 5 años. Actualmente hay más de 50 compañías bioinformáticas (en su mayoría privadas) ofreciendo servicios y se espera, que

el número se incremente significativamente en los siguientes años, con una mayor participación del sector público, debido al fuerte apoyo estatal que se espera recibirá esta área.

Debido al desarrollo comercial de la bioinformática, el futuro enfrentará grandes retos éticos, como el de acceso y la disponibilidad de la información, dado que a raíz de la publicación del genoma humano se ha centrado en la discusión de si los datos deben estar abiertos a toda la comunidad académica o si aquellos que los necesitan deben pagar para obtenerlos. Es necesario aclarar que el trabajo en bioinformática depende del acceso eficiente a los datos y que la publicación de los mismos sea parte del intercambio científico.²⁷

LA BIOINFORMÁTICA Y LAS PERSPECTIVAS PARA COLOMBIA

En general los grupos de investigación en Colombia cuentan con personal que ha recibido alguna formación circunstancial en el manejo de herramientas de la bioinformática a través de cursos (cortos y de posgrado) o de la asesoría parcial de los diversos sitios disponibles en la red que brindan acceso a herramientas y documentación. Algunas universidades colombianas con la colabora-

²⁶ Vingron, M. 2001. Bioinformatics needs to adopt statistical thinking. *Bioinformatics*, 17(5):389-390.

²⁷ States, D. 2001. Time to defend what we have won. *Bioinformatics*, 17(4):299.

ción de grupos de bioinformática del exterior han organizado cursos cortos y disponen de herramientas como bases de datos y software para sus trabajos particulares. Sin embargo, se puede vislumbrar que con la creciente utilización de técnicas de secuenciación en Colombia las necesidades cada vez van a ser mayores, requiriéndose investigadores con formación y dedicación específica en esta área, al igual que la formación de grupos y centros cuyo objeto de trabajo sea la bioinformática.

Algunos grupos y centros en la Universidad del Valle, la Universidad Nacional de Colombia Sede Bogotá, Genicafé, Universidad del Quindío, ICA, Corpogen, Instituto Nacional de Salud, Cevipapa, entre otros, han comenzado a interactuar convocados por el programa nacional de biotecnología de Colciencias, con el objeto de bosquejar una Red Colombiana de bioinformática, que se espera se desarrolle en un futuro no muy lejano, con el propósito de utilizar las capacidades individuales para el beneficio de la comunidad científica que requiere de servicios, asesoría y capacitación en bioinformática.

Es importante mencionar que con la creación de un centro de bioinformática la Universidad Nacional de Colombia Sede Bogotá (<http://>

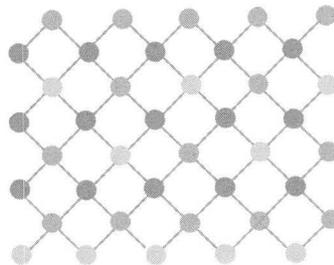
bioinf.ibun.unal.edu.co) dedicado a prestar servicios de asesoría y capacitación en el área, postulado como nodo Colombiano de la Red Europea de Biología Molecular (EMBnet), ha comenzado el proceso de compartir recursos que pueden ser utilizados por otros investigadores a través de Internet, recursos a los que seguramente se podrán unir los de los otros grupo y centros de investigación, integrados a través de la Red Colombiana de Bioinformática, lo que traerá como consecuencia una maximización de los escasos recursos con los que se cuenta y permiten hacer un planeamiento estratégico de los proyectos por realizar.

Considerando que en la época actual la bioinformática no sólo, se restringe al análisis de datos moleculares, sino que su ámbito incluye datos de interés biológico, para Colombia uno de los aspectos de interés es la aplicación de la bioinformática en los estudios de biodiversidad, por tratarse de un país megabiodiverso (el 0.7% de la superficie continental mundial posee cerca del 10% de la diversidad biológica) que cuenta con investigadores, a nivel nacional e internacional, que realizan aportes constantes a la caracterización de nuestras riquezas naturales, que se traducen en la genera-

ción de un gran volumen de datos que deben ser procesados y analizados.

Para enfrentar el reto del conocimiento completo de la biodiversidad en nuestro país es necesario iniciar varios frentes de acción de forma sistemática. La bioinformática ofrece las herramientas y los conceptos para apropiarse de este conocimiento mediante el diseño de bases de datos y herramientas para el análisis de la información acopiada.

La información sobre secuencias de ácidos nucleicos y proteínas de organismos distribuidos en Colombia está dispersa, cada grupo tiene sus propios datos y solo parte de ellos se encuentran en los bancos de datos mundiales. Es necesaria la creación y mantenimiento de un repositorio de toda esta información, que ayudaría a enfocar esfuerzos en los grupos biológicos menos estudiados o en aquellos con mayor potencial de uso, en áreas como la farmacéutica o la de alimentos. ■



BIBLIOGRAFÍA

- STANFORD CENTER FOR PROFESIONAL DEVELOPMENT, 2002. Certificate Program in Bioinformatics. University of Stanford.