

LA BIOINFORMÁTICA COMO UNA FILOSOFÍA DE LA CIENCIA BIOLÓGICA MODERNA:

EL CASO DE LA BIOQUÍMICA ESTRUCTURAL

Por:  Paulina Cortes Hernández ·  Benjamín Padilla Morales · Lenin Domínguez Ramírez

RESUMEN

La bioinformática es comúnmente entendida como el estudio de secuencias genómicas, datos masivos derivados de sistemas biológicos y, en algunos casos, como el modelaje y predicción del comportamiento de estos sistemas. Sin embargo, la bioinformática moderna es más que una disciplina, es una interdisciplina de aplicación biológica. El presente trabajo no es una revisión exhaustiva de esta interdisciplina, más bien pretende enfocarla como una herramienta adicional en el arsenal del pensamiento científico de las ciencias biológicas, como una filosofía práctica. Una manifestación de esta aseveración es la dependencia inextricable de la bioinformática en la computación y la continua demanda por procesadores más poderosos, así como mayores capacidades de almacenamiento de datos. En el presente trabajo, tomaremos a la bioquímica estructural –una rama paralela de la bioinformática– como un ejemplo del impacto de esta interdisciplina en un área específica y de tanto impacto como la genómica.

1970

EL TÉRMINO BIOINFORMÁTICA SE CREÓ PARA DESCRIBIR UN CAMPO INTERDISCIPLINARIO QUE DESARROLLA Y EMPLEA MÉTODOS COMPUTACIONALES PARA ENTENDER DATOS BIOLÓGICOS.

AÑO
1934

La BE nació cuando John Desmond Bernal y Dorothy Crowfoot colocaron un cristal de la proteína pepsina en un haz de rayos X.

AÑO
1936

Max Perutz se embarcó a determinar la estructura de la hemoglobina utilizando el patrón de difracción de rayos X.



Figura 1. John Kendrew y Max Perutz con el modelo de la mioglobina (MRC Laboratory of Molecular Biology).

ESTE
PROCESO
LE LLEVÓ

30

AÑOS,
POR DOS
RAZONES

Por un problema intrínseco a la difracción de rayos X, la incapacidad de medir simultáneamente amplitudes y fases.



La ausencia de computadoras potentes en esa época.

AÑO
1960

Se construyó el modelo de la hemoglobina.



Las densidades electrónicas fueron calculadas a mano y graficadas sobre «acetatos» (material de dibujo translúcido). Cada acetato contenía las coordenadas X y Y.

BIOLOGÍA ESTRUCTURAL (BE),
NACIDA DE LA DIFRACCIÓN
DE RAYOS X

◆ PALABRAS CLAVE

Bioinformática • Biología estructural
• Computación • Big data • Bioquímica
estructural

◆ ABSTRACT

Currently, bioinformatics is conceived as the study of genomic sequences, massive data obtained from biological systems and, in some cases, as modeling and prediction of said biological systems. Nevertheless, modern bioinformatics is much broader than that. It should be viewed as an interdisciplinary field with applications to Biology. The current work is not an exhaustive bioinformatics review but a reminder that this tool is part of the scientist's toolkit to understand biological sciences but it is also a practical philosophy. The fact that bioinformatics is interdependent and tied to the continuously growing computing power and storage capacity is a clear manifestation of this notion. Thus, to focus on the subject, we will discuss a branch of bioinformatics: structural biochemistry. Structural biology is an example of the profound impact Bioinformatics has had on Biochemistry in a fashion similar to the emergence of Genomics.

◆ KEYWORDS

Bioinformatics • Structural biology • Computers • Big data • Structural biochemistry

◆ INTRODUCCIÓN

En sus orígenes (c. 1970), el término bioinformática se creó para describir un campo interdisciplinario que desarrolla y emplea métodos computacionales para entender datos biológicos. Desde sus orígenes ha sido –en efecto– interdisciplinaria pues involucra la combinación de disciplinas como: ciencias de la computación, estadística, matemáticas, ingeniería y, desde luego, biología (Hogeweg, 2011). Es en tiempos más recientes (c. 1980) que se volvió un término asociado directamente al análisis de secuencias de genomas, de expresión de genes, y más recientemente, asociado con la biología sintética (c. 2000). Los esfuerzos por secuenciar genomas completos, anotarlos y extrapolar su contenido a los fenotipos del organismo completo, han hecho de la bioinformática moderna una piedra filosofal, una herramienta con capacidades únicas de procesar información masiva (Big Data). Tomemos, por ejemplo, los 3.2 billones de nucleótidos que forman el genoma del *Homo sapiens*. ¿Cómo correlacionar las secuencias de los 30,000 genes que contiene con sus respectivas funciones, predecir el estado de un organismo en un punto en el tiempo (cuándo desarrollará un padecimiento específico) e, idealmente, ser capaces de modificar ese estado predicho? Una vez logrado esto para un solo genoma, repitamos el proceso para los más de seis billones de seres humanos sobre la faz del

planeta. Sin los paradigmas fundacionales de la bioinformática, *id est*, la homología de las secuencias de genes y su relación con la evolución, esta tarea sería inimaginable. Si tomamos como guía (como hipótesis nula) la primera parte del título de este texto, podremos empezar a apreciar el impacto de la bioinformática en otros aspectos de la biología. Como un ejemplo específico retomaremos la historia de una disciplina madura y que cobra mayor importancia cada día: la biología estructural (BE). En las siguientes secciones describiremos la BE, y las técnicas que proveen las estructuras que son su materia prima: difracción de rayos X, resonancia magnética nuclear y microscopía electrónica. Todas ellas son técnicas bioinformáticas de acuerdo a la definición de Hogeweg (2011) que en conjunto evidencian el valor de la bioinformática como una filosofía y no como una herramienta inmutable.

◆ BIOLOGÍA ESTRUCTURAL, NACIDA DE LA DIFRACCIÓN DE RAYOS X

Si se juzgara el impacto de la BE en la bioquímica moderna, tomando como única referencia un texto moderno de bioquímica (por ejemplo, la *Bioquímica* de Lehninger), descubriríamos que sólo hay tres páginas donde se describen los métodos básicos de la BE, aunque el resto del libro está repleto de ilustraciones de estructuras y modelos de proteínas, ácidos nucleicos,

lípidos y sacáridos. En las páginas de ese mismo libro no se encontrará una definición del término «BE». Una definición moderna es, necesariamente, amplia: es el marco en el que se interpretan las medidas y las ideas de la biología molecular (Campbell, 2002). Pero, ¿qué es concretamente la BE y cuál es su relación con la bioinformática? La BE nació cuando John Desmond Bernal y Dorothy Crowfoot colocaron un cristal de la proteína pepsina en un haz de rayos X (c. 1934). En ese entonces se pensaba que las proteínas eran coloides sin estructura. Para el año 1936, Max Perutz se embarcó a determinar la estructura de la hemoglobina utilizando el patrón de difracción de rayos X (figura 2); este proceso le llevó 30 años (1962) por dos razones: 1) por un problema intrínseco a la difracción de rayos X, la incapacidad de medir simultáneamente amplitudes y fases (esto es una limitante física, los rayos X no pueden ser enfocados como la luz visible); 2) la ausencia de computadoras potentes en esa época, ya que los cálculos para convertir la información del patrón de difracción en un modelo tridimensional requieren de la transformada de Fourier. Mientras que es posible resolverla «a mano», el proceso se vuelve más intenso a mayor número de átomos en el modelo. Max Perutz y John Kendrew resolvieron las dos primeras estructuras mediante difracción de rayos X: la hemoglobina y la globina. Esta tarea les llevó tanto tiempo que

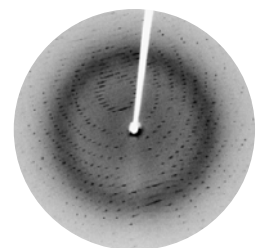


Figura 2. Patrón de difracción de rayos X de la proteasa 3Clpro del virus del SARS a 2.1 Angstroms de resolución. Esta fue capturada por la exposición de una placa fotográfica colocada tras el cristal de la proteína. En la parte superior y cruzando hasta el centro se observa el stopper que previene que la placa se sobreexponga. Debido a que los rayos X no se pueden enfocar, los puntos representan el espacio recíproco del cristal, cada punto contiene información de intensidad (qué tan opacos son) y de frecuencia (la posición relativa al centro indica la resolución, a mayor distancia mayor resolución). Imagen original por Jeff Dahl, reproducida bajo la licencia de Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic y 1.0 Generic.

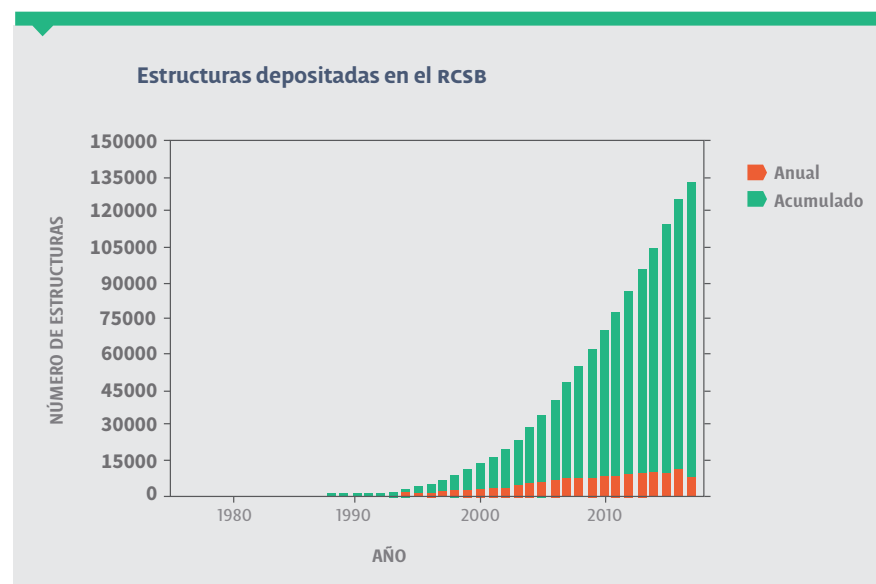


Figura 3. Número de estructuras resueltas y depositadas en la base de datos del RCSB-PDB de manera acumulada y por año. En esta gráfica se incluyen estructuras resueltas por difracción de rayos X, resonancia magnética nuclear, microscopía electrónica y otras. No incluye modelos que no estén basados en datos experimentales. La tendencia de la gráfica muestra que el número de estructuras depositadas ha crecido exponencialmente.

no se esperaban grandes avances en el área. Sin embargo, para 1973, se habían resuelto diez estructuras, para 1983, 27 y en 1993, un total de 922. Actualmente, 55 años después, hay un total de 118,744 (figura 3). Estas estructuras son depositadas en una base de datos de proteínas (RCSB-PDB) y están disponibles de manera gratuita para todos los usuarios con acceso a internet. Desde el 2000 se resuelven y depositan alrededor de cien mil estructuras por año.

Al igual que la bioinformática, la BE cobró auge por dos razones: la disponibilidad de fuentes de rayos X y la capacidad de computación moderna. Mientras que la primera ha dependido de aceleradores de partículas (sincrotrones) asociados a grandes universidades, la computación ha sido más accesible a cualquier público. Usemos de contraste la manera en que Max Perutz construyó el modelo de la hemoglobina en 1960: las densidades electrónicas fueron calculadas a mano y graficadas sobre «acetatos» (material de dibujo translucido). Cada acetato contenía las coordenadas X y Y. Con ayuda de soportes metálicos e hilos, se creó una pila tridimensional que representaba la coordenada Z. Con esta pila como guía, se cortaron rebanadas de corcho para tener un modelo de tres dimensiones. El modelo de la mioglobina de John Kendrew, por su resolución atómica, requirió de un andamiaje más elaborado (figura 1). Hace quince años, todo este proceso se realizaba de manera completamente computacional

en ordenadores especializados, los ordenadores Silicon Graphics y un sistema operativo basado en Unix. Hace diez años, las CPU de Intel o AMD se volvieron lo suficientemente baratas y poderosas para realizar el mismo trabajo. Además, el éxito de los sistemas operativos Linux y sus compiladores gratuitos permitieron la migración de las aplicaciones necesarias para la generación de modelos de manera más rápida y de mejor calidad. Desde hace más de cinco años, las computadoras portátiles o laptops son capaces de procesar, en horas, estructuras de la complejidad de la hemoglobina e incluso se han migrado las aplicaciones a sistemas operativos de Microsoft.

El impacto de la computación y las telecomunicaciones en la determinación de estructuras de proteínas por rayos X (XRD) es tal, que hoy un cristal puede mandarse por correo a Berkeley, difractarse remotamente (mientras el usuario observa en su computadora o teléfono móvil) e incluso preprocesar los datos para determinar su calidad. Y, sin embargo, la metodología de XRD no es la única disponible.

● RESONANCIA MAGNÉTICA NUCLEAR

La resonancia magnética nuclear (RMN) tiene una historia muy rica que inició con el estudio de haces de átomos en el vacío y su respuesta a campos magnéticos y exposición a radiofrecuencias (c. 1926-45) (Becker, 1993). De manera breve, la RMN es capaz de distinguir átomos in-



La resonancia magnética nuclear (RMN) tiene una historia muy rica que inició con el estudio de haces de átomos en el vacío y su respuesta a campos magnéticos y exposición a radiofrecuencias.



Las aplicaciones biológicas de dos métodos principales se hicieron manifestar: Correlated Spectroscopy (COSY) y Nuclear Overhauser Enhancement Spectroscopy (NOESY).



La estructura tridimensional de la primera proteína, determinada por RMN, fue publicada.

dividuales (comúnmente hidrógenos, pero para proteínas nitrógeno y carbono) por su respuesta a un estímulo de radiofrecuencia en un campo magnético controlado.

Fue hasta la década comprendida entre 1976 y 1985 que las aplicaciones biológicas de dos métodos principales se hicieron manifestar: Correlated Spectroscopy (COSY) y Nuclear Overhauser Enhancement Spectroscopy (NOESY). Como se describió para XRD, la computación jugó un papel fundamental pues, para realizar la asignación del espectro de RMN requirió del desarrollo de métodos sofisticados de análisis de datos, así como el uso coordinado de dinámica molecular para evaluar la energética de los modelos. Aun así, fue hasta 1985 que la estructura tridimensional de la primera proteína, determinada por RMN, fue publicada. Dada la complejidad de las biomoléculas, al día de hoy, la RMN está limitada a masas moleculares de 50 kDa (para proteínas significa 500 residuos), pero tiene una gran ventaja sobre la cristalografía: no se requieren cristales y las biomoléculas se analizan en solución.

● MICROSCOPIA ELECTRÓNICA

Aunque el primer microscopio electrónico (ME) fue desarrollado en 1930, fue hasta 1960 que se empezó a aplicar en la biología. El ME funciona de manera muy similar a los microscopios ópticos, excepto que reemplaza la luz visible por

electrones acelerados que vibran a altas frecuencias en una columna de vacío; la muestra que se observe tendrá que introducirse en dicha columna. Las lentes del microscopio óptico son lentes electromagnéticas que son capaces de enfocar estos electrones. Por ello, la resolución teórica del ME alcanza el nivel atómico. Con materiales inertes (*id est.*, aleaciones de metales) se pueden observar incluso átomos individuales. Con macromoléculas hay que considerar su fragilidad, un haz de electrones es capaz de evaporar una muestra de proteína y eso es sin considerar el vacío. El desarrollo de la tinción negativa, la cual rodea a las proteínas con una sal pesada que simultáneamente las protege del vacío y les da un gran contraste frente al haz de electrones, fue el primer paso para las aplicaciones en macromoléculas. Posteriormente, fue con la introducción de la criomicroscopía que esta técnica logró establecerse como una herramienta más de la BE (Gebel Berg, 2015). En la criomicroscopía se emplea una capa de hielo amorfo, a temperatura de nitrógeno líquido (77 K o -196 °C), que cumple las funciones de la tinción negativa y preserva mejor la estructura nativa de las macromoléculas. Sus primeros hitos tomaron lugar en 1975, cuando Richard Henderson y Nigel Unwin resolvieron la estructura de la bacteriorodopsina, una proteína de membrana, a 7 Angstroms de resolución, usando difracción de electrones¹. Para 1990, con ayuda

¹ Un ME es capaz de enfocar los electrones tomando, literalmente, fotos. Pero también es capaz de no enfocarlos y difractarlos sobre la muestra. De esta manera, no sufre la limitación de los rayos X pues se pueden obtener las amplitudes y fases para determinar una estructura.

AL IGUAL QUE LA BIOINFORMÁTICA, LA BE COBRÓ AUQUE POR DOS RAZONES: LA DISPONIBILIDAD DE FUENTES DE RAYOS X Y LA CAPACIDAD DE COMPUTACIÓN MODERNA.

de mejores microscopios, mejores algoritmos y computadoras más rápidas, se alcanzaron resoluciones de 3.5 Angstroms utilizando análisis de imagen de partícula única². Cabe resaltar que esta resolución es baja para los estándares de la XRD. Sin embargo, la XRD de proteínas de membrana es casi imposible en ambientes naturales (*id est.*, lípidos) y éste no es el caso para la ME. Al igual que la RMN, la ME no requiere de cristales. Al día de hoy resoluciones de 2 Angstroms, casi tan altas como las de XRD, son posibles mediante ME³.

Estas tres metodologías son el pilar moderno de la BE, por ejemplo, se puede determinar la estructura de una proteína junto con un fármaco, o cuando se identifica una mutación causante de una enfermedad, se puede sugerir cómo la mutación afecta la función de dicha proteína. Aunque esta información es en extremo valiosa, sin embargo, es estática; es decir, semejante a una sola fotografía del galope de un caballo. Una sola fotografía nos puede dar mucha información general del caballo (sus dimensiones, el largo de su zancada), pero carece de información del proceso del galope en sí (... y no nos dice ¡qué tan rápido corre!). Mejores y más rápidas técnicas de difracción se han desarrollado, sin embargo, hay otros métodos que usan los modelos generados experimentalmente para extrapolar su comportamiento dinámico, como la simulación por mecánica molecular. La mecánica molecular es tan antigua como la BE (c. 1938) (Wilson y Crawford, 1938), y también floreció con la disponibilidad de computación de alto poder alrededor de los ochenta. En pocas palabras, la mecánica molecular es una aplicación de la mecánica clásica (también llamada física newtoniana) para modelar y simular el comportamiento atómico de diversos sistemas. Inicialmente, cuando los cálculos se realizaban «a mano», se aplicó a sistemas químicos simples (uno clásico es el ión bifluoruro, FHF-) donde se podían realizar cálculos más rigurosos (mecánica cuántica o mecánica semiempírica). Posteriormente, en la descripción de la estructura de la proteína BPTI (Bovine Pancreatic Trypsin Inhibitor), en 1977, se usaron por primera vez los algoritmos del premio Nobel Martin Karplus⁴ en las supercomputadoras disponibles en las universidades y se aplicó la mecánica cuántica a problemas biológicos. En

El término Bioquímica Computacional fue plasmado en

2001

en el libro *Computational Biochemistry and Biophysics* (Becker, MacKerell, Roux, & Watanabe, 2001). Se refiere específicamente al uso de la mecánica molecular, semiempírica o cuántica, para entender la naturaleza de los fenómenos biológicos a nivel atómico.

mi opinión, es en ese artículo donde nació de la Bioquímica Computacional (BC).

• BIOQUÍMICA COMPUTACIONAL, HEREDERA DE LA BE

La bioquímica computacional (BC) ha recorrido un camino semejante a la bioinformática moderna, aunque ha ido de la mano de muchos experimentos que confirman o refutan sus resultados. El término BC fue plasmado en 2001 en el libro *Computational Biochemistry and Biophysics* (Becker, MacKerell, Roux, y Watanabe, 2001). Se refiere específicamente al uso de la mecánica molecular, semiempírica o cuántica, para entender la naturaleza de los fenómenos biológicos a nivel atómico.

Tal como se describió para la bioinformática, la BC surgió de la confluencia de varias disciplinas,azonadas con la programación y la computación moderna. Primero, la bioquímica *per se*, con la descripción de la estructura del material genético, ADN, y los modelos de la catálisis enzimática, había alcanzado una cima en importancia y desarrollo. Segundo, la descripción de la estructura tridimensional de las proteínas, el desarrollo de la mecánica molecular y su aplicación a las proteínas abrieron un panorama inexplorado que llevaría a novedosas hipótesis para entender la vida misma. Y, finalmente, la accesibilidad al poder de cómputo, necesario

para simular estas proteínas (y otras macromoléculas), permitieron la fundación de esta área.

La observación inmediata, *id est.*, las vibraciones de la BPTI, hicieron palpable la naturaleza dinámica de la bioquímica a nivel molecular. Sin el conocimiento bioquímico para interpretar los resultados de la mecánica molecular, la aplicación de la BC sería limitada. Es decir, parafraseando las palabras de Martin Karplus⁴: las simulaciones no están reemplazando a los experimentos, sino que son también herramientas para entender dichos experimentos y extenderlos. Mientras que la difracción de rayos X, RMN y ME han recorrido un largo camino y madurado a disciplinas sólidas, la bioquímica estructural como parte de la bioinformática sigue evolucionando. Un ejemplo contemporáneo de las aplicaciones de la bioquímica computacional está en el diseño de fármacos. Para el diseño de fármacos se requieren estructuras de XRD o NMR de las macromoléculas blanco, una librería (computacional) de moléculas orgánicas que se usarán de señuelo, un set de programas de cómputo que realicen *docking*, simulación y cálculo de la energética de las interacciones y, finalmente, gran poder de cómputo⁵ y tiempo. Desde luego que un diseño de fármacos exitoso tendrá que ser acompañado de pruebas experimentales (por ejemplo, farmacología) u otras pruebas estructurales (XRD o espectroscopia).

Hoy por hoy, hay aún un largo camino por recorrer para hacer de la BC una digna heredera de la BE y también para cimentar sus vínculos con la bioinformática y la genética moderna. Sólo así nos moveremos algo más cerca al futuro utópico donde a partir de la secuencia de un gen podremos predecir una función, predecir las alteraciones causadas por mutaciones específicas (diagnóstico) y diseñar un tratamiento a nivel molecular.

CONCLUSIONES

En este trabajo presentamos a la bioinformática moderna no como una disciplina, sino como una filosofía de la ciencia, una plantilla del paradigma de la ciencia y como una interdisciplina. Es nuestra opinión, evidenciada por la evolución de la biología estructural, que es una filosofía pues involucra la resolución de preguntas biológicas tomando herramientas de muchas otras disciplinas. Además, enfatizamos que no es un producto acabado sino que sigue

cambiando, mejorando. Presentamos además el caso de la bioquímica computacional que nació de la interdisciplina (computo, química, BE) y tiene la capacidad en potencia de incluir a la biología celular y a las ciencias de la salud. Es crucial en la educación profesional incluir los elementos necesarios para alcanzar ese objetivo, mediante la bioinformática moderna.

REFERENCIAS

- Becker, E. D. (1993). A brief history of nuclear magnetic resonance. *Analytical Chemistry*, 65(6), 295A-302A.
- Becker, O. M., MacKerell, A. D., Jr, Roux, B., & Watanabe, M. (2001). *Computational Biochemistry and Biophysics*. (O. M. Becker, A. D. MacKerell Jr, B. Roux, & M. Watanabe, Eds.) (1st ed.). CRC Press.
- Campbell, I. D. (2002). Timeline: the march of structural biology. *Nature Reviews Molecular Cell Biology*, 3(5), 377-381. <http://doi.org/10.1038/nrm800>
- Gebel Berg, E. (2015, September 23). Breaking the Crystal Ceiling. *ACS Central Science*, 283-285. <http://doi.org/10.1021/acscentsci.5b00295>
- Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology*, 7(3), e1002021. <http://doi.org/10.1371/journal.pcbi.1002021>
- Wilson, E. B., Jr., & Crawford, B. L., Jr. (1938). The Secular Equation for Molecular Vibrations. *The Journal of Chemical Physics*, 6(4), 223-223. <http://doi.org/10.1063/1.1750232>



Paulina Cortes Hernández

Médico cirujano y doctora en Ciencias Bioquímicas por la UNAM. Actualmente es investigadora asociada en el Centro de Investigación Biomédica de Oriente (CIBIOR), del IMSS en Metepec, Puebla. Conduce estudios que permiten entender las bases moleculares de las enfermedades humanas y el papel de la estructura de las proteínas en la enfermedad. Ha trabajado en enfermedades mitocondriales, enfermedades metabólicas y en cáncer. paulina.cortes.hernandez@gmail.com



Benjamín Padilla Morales

Biólogo, egresado de la UDLAP y actualmente estudiante de MBA en la UDLAP. Colaborador del laboratorio de Modelaje Químico-Biológico y Visualización de la UDLAP en proyectos enfocados en determinar el comportamiento estructural de proteínas *in-silico* y detección computacional de patrones de especiación-extinción en artrópodos. Colaboró con el Departamento de Matemáticas en el diseño de la propuesta del Centro en Aprendizaje de Matemáticas. benjamin.padillams@udlap.mx



Lenin Domínguez Ramírez

Doctor en Ciencias, Licenciado en Investigación Biomédica Básica. Profesor de la UDLAP. Conduce investigación orientada a la aplicación de herramientas computacionales para el rediseño de proteínas y la búsqueda de nuevos fármacos, por ejemplo contra la influenza. julio.dominguez@udlap.mx

2. Este análisis usa sólo información de imágenes enfocadas, no requiere difracción.

3. Premio Nobel de Química en 2017.

4. Premio Nobel de Química 2013.

5. Un cluster de supercómputo, como el Laboratorio Nacional de Supercómputo, LNS.