REAL ACADEMIA DE FARMACIA DE GALICIA

Discurso de Ingreso como
Académico Correspondiente

LA INTELIGENCIA ARTIFICIAL EN EL DESCUBRIMIENTO DE ANTIBIÓTICOS

Dr. D. César de la Fuente Núñez

Discurso de Presentación por
el Académico de Número

Ilmo. Sr. Dr. Germán Bou Arévalo

Santiago de Compostela,
22 de mayo de 2024

© César de la Fuente, Consellería de Sanidade y Real Academia de Farmacia de Galicia

Maquetación: Miguel A. Suárez.
ISBN: 978-84-128486-0-1.
Depósito legal: C 712-2024.

Índice

I DISCURSO DE PRESENTACIÓN 5

DISCURSO DE PRESENTACIÓNPOR EL ACADÉMICO DE NÚMERO Ilmo. Sr. Dr. Germán Bou Arévalo 7

II DISCURSO DE INGRESO 13

DISCURSO DE INGRESO EN LA REAL ACADEMIA DE FARMACIA DE GALICIA DEL ACADÉMICO CORRESPONDIENTE ELECTO Dr. D. César de la Fuente Núñez 15

PRÓLOGO 15

1. EL MUNDO EN 2050 21

   1.1. La resistencia a los antibióticos, un problema global 21
   1.2. Las posibilidades de la inteligencia artificial 24

2. ¿PUEDE CREARSE UN ANTIBIÓTICO CON ORDENADORES? 27

3. *MACHINE LEARNING* PARA EL DISEÑO DE FÁRMACOS 37

   3.1. Ventajas: reducción de tiempo y de costes 37
   3.2. Minería de proteínas para el descubrimiento de antibióticos 38
   3.3. El proteoma humano, una fuente de antibióticos 40
   3.4. Péptidos encriptados: un nuevo tipo de moléculas 43
   3.5. Desextinción molecular: las moléculas como fuentes de documentación histórica 47
   3.6. panCleave 49
   3.7. ¿Cómo funcionan los antiguos y los nuevos antibióticos? 51

3.8. *Deep learning* para el descubrimiento de antibióticos 54
3.9. ¿Por qué la desextinción molecular? 56

4. PRÓXIMOS PASOS 59

4.1. Minería del árbol de la vida 59
4.2. Aceleración del descubrimiento de antibióticos con inteligencia artificial 60

5. BIBLIOGRAFÍA 65

I

DISCURSO DE PRESENTACIÓN

DISCURSO DE PRESENTACIÓNPOR EL ACADÉMICO DE NÚMERO
Ilmo. Sr. Dr. Germán Bou Arévalo

PARA EL INGRESO DEL Dr. César de la Fuente Núñez

Ilmo. Sr. Presidente de la Academia de Farmacia de Galicia,
Autoridades
Ilmas. e Ilmos. Sras. y Sres. Académicas y Académicos,
Señoras y Señores,

Para mí, es un honor hacer la presentación del Dr. César de la Fuente Núñez como Académico Correspondiente. Agradezco a la Real Academia de Farmacia de Galicia el encargo de esta tarea, que me resulta particularmente grata.

El Dr. César de la Fuente Núñez, coruñés de nacimiento, es licenciado en Biotecnología por la Universidad de León y doctor en microbiología e inmunología por la Universidad de British Columbia (Canadá), gracias a una beca de la Fundación "La Caixa". Ha realizado estancias de investigación y formación en destacados centros como el Massachusetts Institute of Technology -MIT- (Estados Unidos), gracias a una beca de la Fundación Ramón Areces.

Después de su etapa en el MIT, fue contratado en la Universidad de Pennsylvania (UPenn), la institución en la que sigue trabajando desde entonces. En la actualidad, es profesor catedrático y lidera el Grupo Machine Biology, que tiene como objetivo principal desarrollar nuevos antibióticos mediante ordenadores.

César de la Fuente es autor o coautor de más de 150 publicaciones, incluidos artículos en Science, Cell Host Microbe, Nature Biomedical Engineering, Nature Communications, PNAS, ACS Nano, Cell, Nature Chemical Biology, y Advanced Materials, así como de numerosas patentes.

Las investigaciones del Dr. de la Fuente Núñez se centran en el uso de ordenadores (inteligencia artificial) para descubrir clases de antibióticos completamente nuevas que no compartan mecanismos de acción con los preexistentes. Hay que clarificar que dentro de la gran cantidad de familias distintas y clases moleculares de antibióticos que existen en la actualidad, su trabajo se centra en un particular tipo de moléculas basadas en péptidos o secuencias de aminoácidos.

Su objetivo de investigación es utilizar el poder de las máquinas o computadoras para acelerar los descubrimientos en biología y medicina. Específicamente, fue pionero en el desarrollo del primer antibiótico (basado en secuencias peptídicas) diseñado por ordenador con eficacia en modelos animales, demostrando la aplicabilidad de la inteligencia artificial (IA) en el descubrimiento de antibióticos y ayudando a lanzar este campo emergente. Su laboratorio también ha estado a la vanguardia en el desarrollo de métodos computacionales para la búsqueda de antimicrobianos a partir de proteomas, lo que ha llevado al descubrimiento revolucionario de un mundo completamente nuevo de antimicrobianos. Estos esfuerzos se han orientado a explorar el

proteoma humano como fuente de antibióticos por primera vez y han acelerado drásticamente el tiempo necesario para descubrir candidatos preclínicos, de años a horas, en esta particular clase de moléculas.

El grupo de de la Fuente también fue el primero en encontrar moléculas terapéuticas en organismos extintos, iniciando el campo de la desextinción molecular. La desextinción molecular ya ha dado lugar a candidatos a antibióticos preclínicos. Los avances adicionales de su laboratorio incluyen el diseño de algoritmos para el descubrimiento de antibióticos y la reprogramación de venenos para convertirlos en antimicrobianos, así como la creación de nuevos materiales antimicrobianos que evitan la resistencia a los mismos.

El Dr. de la Fuente es un investigador altamente citado y ha recibido más de 70 premios nacionales e internacionales. Es miembro electo del American Institute for Medical and Biological Engineering (AIMBE), convirtiéndose en uno de los más jóvenes en ser admitido. Fue reconocido por MIT Technology Review como uno de los principales innovadores del mundo por "digitalizar la evolución para fabricar mejores antibióticos". Fue seleccionado como primer ganador del Premio Langer y como Líder Emergente en Química de ACS Kavli, Profesor Distinguido de ASM, Profesor de la Fundación Waksman y recibió el Premio 35 Under 35 de AIChE, el Premio al Joven Investigador de la Sociedad de Ingenieros Profesionales Hispanos, y el Premio ACS Enfermedades Infecciosas Joven Investigador. También recibió el premio Thermo Fisher, así como el EMBS Academic Early Career Achievement Award "Por el desarrollo pionero de nuevos antibióticos diseñados utilizando principios de computación, ingeniería y biología".

Más recientemente, el Dr. de la Fuente recibió el prestigioso Premio Princesa de Girona de Investigación Científica, el Premio ASM de Investigación Biotecnológica y Aplicada en la Carrera Temprana, y el Premio de Cátedra Rao Makineni de la American Peptide Society.

El Dr. de la Fuente ha impartido más de 250 conferencias como ponente invitado, incluidas numerosas conferencias magistrales. Es coautor de un influyente libro sobre aprendizaje automático para el descubrimiento de fármacos y sus descubrimientos científicos han dado lugar a múltiples patentes y artículos científicos, como se ha mencionado.

A su capacidad investigadora, se une que, además, es un hombre muy familiar, como se desprende del prólogo de su discurso. Yo lo conocí personalmente hace años (allá por el 2011, perdona César si las fechas no son precisas, y permíteme que te llame César) en una pequeña estancia de pocos meses que realizó en nuestro laboratorio en La Coruña, siendo posteriormente testigo de la gran evolución científica de este investigador en un lapso de tiempo relativamente breve.

En su excelente discurso de ingreso César nos hablará de la problemática de la resistencia a los antibióticos y de algunas de las soluciones que pueden plantearse. A modo de resumen final, y como ejemplo desarrollado por su propio grupo en Estados Unidos, el uso de la IA para diseñar nuevos antimicrobianos basados en secuencias peptídicas, con sus ventajas e inconvenientes. Nos comentará también como ha realizado la búsqueda de estas nuevas moléculas con algoritmos informáticos novedosos a partir de genomas/proteomas de seres humanos y de animales extintos y los prometedores resultados preclínicos obtenidos a la fecha. No quiero ahondar más en su discurso de

ingreso que estoy seguro César presentará de manera magistral. El protagonismo hoy es suyo.

Por todo lo expuesto, es un placer introducir a César de la Fuente Núñez como académico correspondiente de la Real Academia de Farmacia de Galicia, a la que también felicito por tan excelente incorporación. Estoy seguro, César, de que ayudarás a mejorar la misión de nuestra Academia con tus conocimientos, y a su progreso científico para, de esta manera, seguir contribuyendo a su compromiso con la sociedad.

Muchas gracias y enhorabuena César.

II

DISCURSO DE INGRESO

DISCURSO DE INGRESO EN LA REAL ACADEMIA DE FARMACIA DE GALICIA DEL ACADÉMICO CORRESPONDIENTE ELECTO **Dr. D. César de la Fuente Núñez**

PRÓLOGO

Ilmo. Sr. Presidente de la Academia de Farmacia de Galicia,
Autoridades
Ilmas. e Ilmos. Sras. y Sres. Académicas y Académicos,
Señoras y Señores,

Quiero comenzar mi intervención agradeciendo a la Real Academia de Farmacia de Galicia, a su presidente, el Ilmo. Sr. Dr. Ángel Concheiro, a las académicas y académicos que apoyaron mi candidatura, la Dra. Carmen Álvarez Lorenzo, la Dra. María José Alonso y el Dr. Germán Bou Arévalo, y a todos sus miembros, la acogida en esta ilustre Institución como Académico correspondiente. Es un honor y una responsabilidad que asumo con emoción e ilusión y que, sin duda, será un impulso para seguir trabajando en el desarrollo de la ciencia en español, de la que soy un orgulloso abanderado cada vez que se presenta la ocasión.

Dedico este reconocimiento a mi padre, Raúl, que nos dejó demasiado pronto, y a mi madre, Chicha. Ellos me enseñaron a levantarme cada vez que caigo. En la vida hay pocas cosas más

importantes. También a mis hermanas, Milena y Vânia, a mi hermano, Manoel (Pipo), y a mi pareja, Lily. Sin su apoyo incondicional, amor y cariño, no estaría aquí ahora mismo junto a ustedes.

El recuerdo de mi padre, transmitiéndome la sensación de descubrir algo por primera vez, sigue aún vívido y me acompaña con nostalgia. Ese momento íntimo, precioso y emocionante en el que sabes algo que nadie más conoce. Quizás en ese instante decidí ser científico. Y el resto es historia.

Es difícil expresar en palabras el privilegio que supone este reconocimiento, que, desde luego, no podía ni soñar cuando era niño y sentía una profunda curiosidad por el mundo que me rodeaba. Siempre digo que todas las personas nacemos con un alma científica, pero tendemos a perder esa curiosidad a medida que crecemos. He de confesar que todos estos años me he esforzado para proteger esa semilla y no crecer demasiado, porque esa era la única manera de conservar el hambre por aprender, por saber, por comprender. Ese es el espíritu que ha guiado y sigue guiando mi formación y mi trabajo.

Es verdad que siempre he sentido un interés especial por comprender la biología en un nivel fundamental. Al principio, pensaba que eso podría ser una manera de mejorar el mundo. Sin embargo, creo que no les cuento nada que no sepan si les digo que esto no es fácil. La biología es compleja, desafiante y divertida, y también puede parecer caótica. No puede entenderse como una disciplina única, ya que requiere la conexión de patrones y conceptos de todos los dominios: química, física, ingeniería… A mí me atrajo ese caos y me llevó a tratar de comprender, diseñar y trabajar en esta área científica.

Durante mis estudios de Biotecnología en la Universidad de León, decidí intentar aprender lo máximo posible para llegar a comprender la biología. Para conseguirlo, me adentré en todas las disciplinas relacionadas: física, química, matemáticas, informática, estadística, biotecnología, microbiología e ingeniería. Fue en esa etapa cuando comencé a tomar contacto con la investigación y tuve un flechazo con las bacterias. Además, aprendí a utilizar la química para crear medicamentos con mis propias manos. Se puede decir que, en mi paso por la universidad leonesa, empezó a germinar la inquietud que se fue materializando más tarde.

Cuando terminé mi licenciatura, con poco más de veinte años, seguía decidido a comprender la biología a un nivel fundamental y tomé la decisión de realizar un doctorado. Para ello, me mudé a Canadá, donde había vivido de niño junto a mi familia, porque quería estudiar en la Universidad de British Columbia (UBC). He de decir que no tenía un plan B, por lo que me considero muy afortunado de haber conseguido el plan inicial. Tuve suerte, y me aceptaron en UBC, una de las mejores universidades del mundo. En aquel momento, tampoco hablaba mucho inglés, así que tuve que aprender rápido. Por entonces, en las universidades norteamericanas había muy pocos científicos y científicas que fueran hispanohablantes nativos (desafortunadamente, esto no ha cambiado mucho). Todo era un reto, pero soy muy testarudo. Ni el síndrome del impostor, ni las barreras idiomáticas iban a impedir que me convirtiera en científico.

Durante mi doctorado, me centré en la biología invisible, es decir, aquella que tiene lugar a microescala (la microbiología). Estudié las bacterias para tratar de comprender su funcionamiento y cómo podían volverse dañinas para los humanos. También

comencé a experimentar con péptidos, que son, junto con las proteínas, los caballos de batalla de la vida y las biomoléculas menos programables que existen.

A medida que se acercaba el final de mi doctorado, crecía mi insatisfacción por nuestra incapacidad para programar la biología. Todo dependía de una experimentación minuciosa, de prueba y error, con poca sistematización o previsibilidad. Fue entonces, hace ya más de una década, cuando vislumbré que los ordenadores revolucionarían la biología.

En 2015, ya convertido en doctor, me reclutaron para el MIT, la meca de la investigación en Inteligencia Artificial (IA) en ese momento. Aunque se estaba aplicando a otras áreas, prácticamente nadie había pensado que la IA pudiera utilizarse en un campo tan complejo como la biología. De hecho, se pensaba que era imposible y que faltaban décadas para poder hacerlo. Casi parecía ciencia ficción. Pero yo decidí intentarlo y me planteé una pregunta muy sencilla y fundamental: ¿puede una máquina crear un antibiótico? Una pregunta sobre la que he seguido trabajando y a la que hemos ido añadiendo complejidad desde el laboratorio que he creado en la Universidad de Pennsylvania (UPenn), el De la Fuente Lab. A lo largo de este discurso, compartiré con ustedes algunas de las respuestas que hemos encontrado y de los retos que seguimos planteando en estos nuevos campos que hemos tenido la suerte de crear.

Mi sueño es trasladar nuestros descubrimientos científicos del laboratorio al mundo, de manera que lleguen a las personas y salven vidas, porque no olvido que estamos trabajando con problemas reales, con infecciones bacterianas que están matando a millones de personas en todo el mundo. Eso es lo que impulsa mi trabajo día a día.

A continuación, paso a la parte académica de mi discurso, en la que esbozaré las líneas más importantes del camino emprendido para poder descubrir y diseñar antibióticos con inteligencia artificial.

1. EL MUNDO EN 2050

1.1. La resistencia a los antibióticos, un problema global

En la última década, hemos sido capaces de lograr importantes avances que han contribuido al desarrollo del uso de la Inteligencia Artificial (IA) y de los ordenadores para el descubrimiento de antibióticos. Se trata de un área emergente en la que nuestra máxima motivación ha sido y sigue siendo el problema de la resistencia a los antibióticos. Este es un reto para la salud mundial, ya que afecta a todos los rincones del planeta. Desde hace décadas no se han descubierto nuevos antibióticos y las bacterias son cada vez más resistentes a los que ya tenemos. Actualmente, la resistencia a los antibióticos es la causa de más de un millón de muertes al año; una cifra que va en aumento, y se estima que, en 2050, puede matar anualmente a 10 millones de personas, superando cualquier otra causa de mortalidad en el mundo.

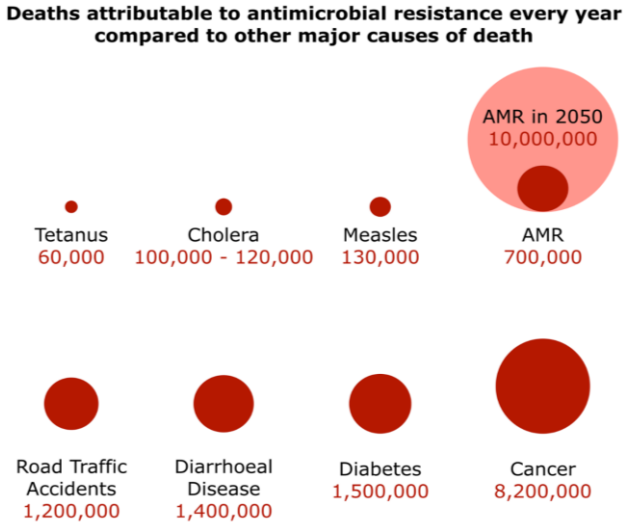


Figura 1. Muertes atribuidas a la resistencia a los antibióticos en comparación con las principales causas de mortalidad (Fuente: *Review on Antimicrobial Resistance* 2014; *Nature Reviews Drug Discovery* 6:28)

Por tanto, se trata de un problema muy generalizado que se está convirtiendo en una de las máximas preocupaciones por su constante agravamiento. Si tomamos esa estimación de 10 millones de muertos anuales para 2050 y realizamos un cálculo rápido, comprobamos que eso supondría aproximadamente una muerte cada tres segundos. Este es el futuro hacia el que nos dirigimos, a menos que encontremos soluciones para acelerar el descubrimiento de antibióticos o de nuevos tipos de antimicrobianos que podamos utilizar para contrarrestar las infecciones bacterianas.

Desde el año 2019, estoy al frente del De la Fuente Lab, en el que lidero el Grupo *Machine Biology* (de biología computacional) de la Universidad de Pennsylvania. Allí, tengo la suerte de poder trabajar con personas que proceden de muchos países distintos, también de muchos entornos y con formas de pensar diferentes. Para que se hagan una idea, en estos momentos, en mi laboratorio trabaja gente que se ha formado en informática, en química, en microbiología y en biología sintética, y todas esas personas colaboran para tratar de resolver algunos de estos grandes problemas causados por la resistencia a los antibióticos. Es un privilegio trabajar con este equipo cada día y muchas de las cosas que les voy a contar hoy son fruto de su trabajo.

El concepto de *Machine Biology* surge, básicamente, de nuestra creencia de que las máquinas y los ordenadores pueden ayudar a acelerar los descubrimientos en esa intersección con la biología y la medicina. La motivación de nuestro pensamiento, que puede resultar revolucionaria, se basa en que creemos que la IA es ideal para realizar descubrimientos moleculares debido a una serie de tendencias. La primera es la potencia de cálculo cada vez mayor. De acuerdo con la ley de Moore, el número de transistores que podemos introducir en un chip se duplica cada 18 meses

aproximadamente. En términos prácticos, esto se traduce en el hecho de que un superordenador que solía ocupar una habitación entera hace solo 30 ó 40 años, ahora se puede llevar en el bolsillo, en forma de teléfono móvil. Es decir, hay mucha potencia de cálculo y se puede procesar mucha información de una forma muy rápida y sin precedentes en la historia de la humanidad.

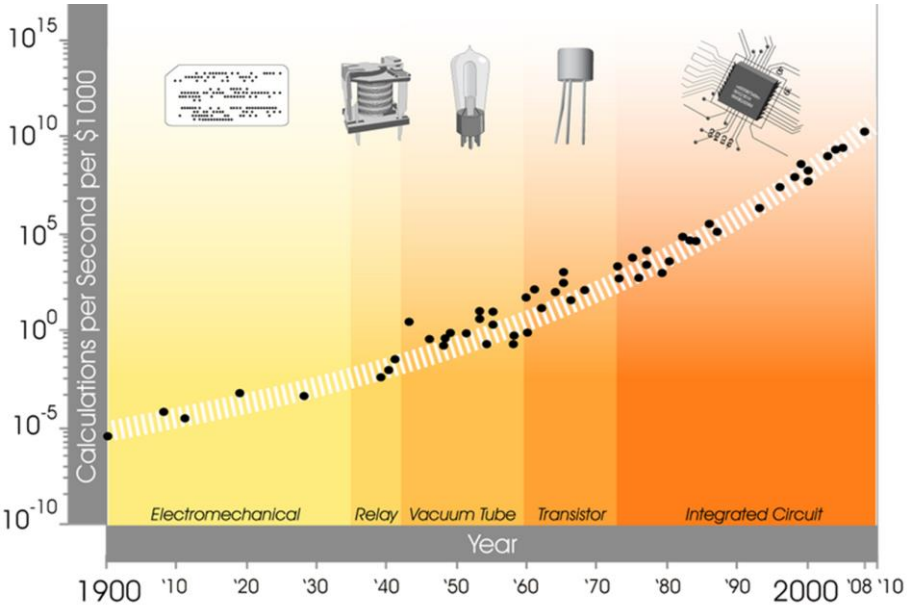


Figura 2. Evolución de la potencia de cálculo (Fuente: *Nat Rev Drug Discov.* 2012;11:191-200)

Además, ahora, en biología, también tenemos la capacidad de generar muchos datos gracias a los avances en la automatización y en los ensayos de experimentación de alto rendimiento. Esto hace posible generar toneladas de datos con las que luego alimentar a los ordenadores, de forma que el procesamiento de toda esa información permite construir modelos predictivos y generativos para crear y predecir nuevas moléculas.

1.2. Las posibilidades de la inteligencia artificial

Llevo mucho tiempo reflexionando sobre esto. Cuando todavía estaba haciendo mi doctorado en la Universidad de British Columbia (Canadá), antes de ir al MIT, anoté en un papel algunas ideas que tenía en ese momento sobre cómo podrían ayudarnos los ordenadores en este campo. Aún guardo ese trozo de papel, en el que escribí algunos de los planteamientos que hoy voy a compartir con ustedes. Verán que algunas de aquellas propuestas conceptuales tienen aún vigencia y comprobarán su relevancia a lo largo de este discurso. Pero me gustaría destacar algo que pensé en aquel momento, la idea que inspiró lo que esbocé en aquella hoja: que los ordenadores podrían ayudarnos a explorar el espacio de secuencias. Algo que era imposible cuando estaba haciendo mi doctorado en Canadá porque, básicamente, la biología era extremadamente meticulosa y compleja. Si tratabas de mutar un sistema biológico concreto como una bacteria, una molécula, una proteína o un péptido, siempre tenías que volver al inicio. No había una forma sistemática de hacerlo. Sin embargo, eso no fue impedimento para elucubrar las posibilidades que podría ofrecer la computación. De hecho, ya en ese instante, tuve una epifanía. Pensé que los ordenadores podrían ayudarnos a explorar el espacio de secuencias de una manera muy eficiente, aunque entonces era imposible hacerlo experimentalmente.

Lo segundo que apunté en aquel papel fue que las máquinas podrían contribuir a generar moléculas completamente nuevas y diferentes de las que existían en la naturaleza, lo que podría ser de gran ayuda para abordar la resistencia a los antimicrobianos y también otros problemas de nuestra sociedad. La otra cosa que anoté fue que los ordenadores podrían ser útiles para minar la biología con eficacia. Por biología me refiero a toda la información

biológica a la que tenemos acceso, incluidos genomas, proteomas, etc. Soñé que, quizás, algún día podríamos explorar toda la información biológica como fuente de nuevas moléculas. Más adelante, verán cómo se relacionan algunos de estos conceptos que registré cuando aún estaba escribiendo mi tesis doctoral y cómo se convirtieron en el germen de todo lo que vino a continuación.

Al terminar mi doctorado, ya les he contado que me contrataron en el MIT, que, por aquel entonces, era la meca de la IA. Los algoritmos se estaban utilizando, principalmente, en áreas ajenas a la biología. Por ejemplo, para los sistemas de reconocimiento de patrones faciales o el reconocimiento de voz que usamos todos los días en nuestros teléfonos. Pero el consenso generalizado era que la IA no sería útil en biología porque era demasiado compleja. Se pensaba que era demasiado caótica para que un ordenador fuera realmente útil. Sin embargo, hicimos caso omiso de ese escepticismo y decidimos poner en marcha un proyecto con algunos de nuestros grandes colaboradores. Así, iniciamos esta senda en la que continuamos avanzando, partiendo de una pregunta fundamental: ¿se puede utilizar un ordenador para crear un antibiótico?

2. ¿PUEDE CREARSE UN ANTIBIÓTICO CON ORDENADORES?

El inicio del proyecto supuso un camino lleno de desafíos, en el que tuvimos que convivir también con nuestras propias dudas. A cada paso, pensábamos que las cosas no iban a funcionar porque nadie las había ejecutado antes. De hecho, considerábamos que podríamos fracasar miserablemente, pero, incluso barajando esa posibilidad, decidimos seguir adelante. Decidimos lanzarnos al abismo, a los límites del conocimiento y de lo que era posible entonces.

Empezamos por preguntarnos cómo podríamos enseñar a un ordenador a crear algo nuevo a nivel molecular, a generar diversidad, a innovar. Después de pensarlo mucho y de reflexionar conjuntamente con nuestros y nuestras colegas, decidimos intentar imitar el mejor motor que tenemos en la Tierra para generar diversidad e innovar a cualquier nivel: la propia evolución. Lo que hicimos fue entrenar al ordenador para que aplicara el algoritmo evolutivo de Darwin. Pero, en nuestro caso, contábamos con una gran ventaja frente a lo observado por el naturalista británico: en lugar de esperar millones de años para que una molécula evolucionase y se optimizara, las máquinas nos permitirían comprimir esa escala de tiempo y lograrlo en cuestión de horas.

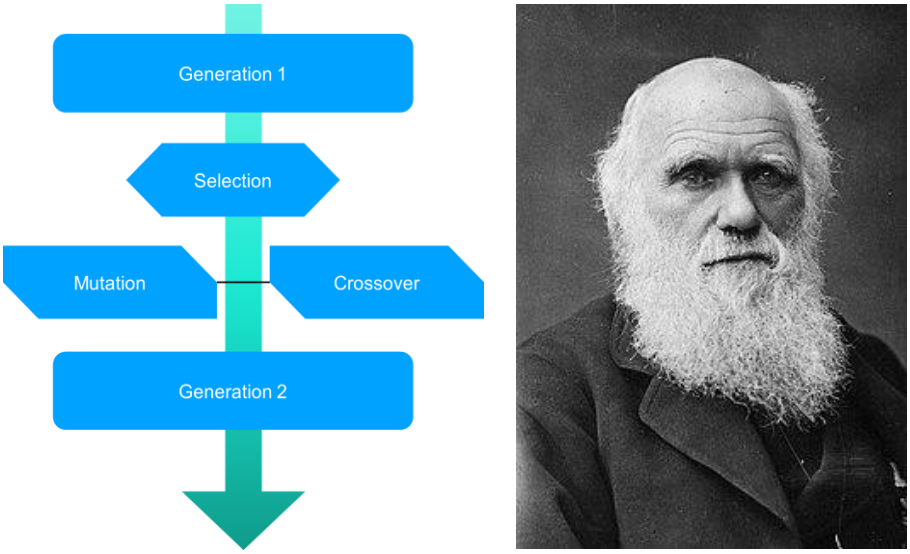


Figura 3. Algoritmo evolutivo de Darwin (Imagen: Wikimedia Commons)

Por tanto, partiendo de esa premisa evolutiva, empezamos a trabajar con una población inicial de pequeñas moléculas de la naturaleza (proteínas llamadas péptidos) y codificamos su complejidad química en el código binario de unos y ceros, para que el ordenador entendiera esas moléculas. A continuación, entrenamos al ordenador para que ejecutase la evolución. Como recordarán de las clases de biología, los principales pasos de la evolución son la mutación, la selección y la recombinación. Estos fueron los pasos sobre los que construimos este bucle de retroalimentación iterativa para obtener esencialmente la evolución en tiempo real. En la siguiente imagen (Figura 4), se puede ver una de las simulaciones, en la que, al aumentar el número de iteraciones, el algoritmo es capaz de evolucionar las moléculas a valores óptimos de aptitud (*fitness*) que se correlacionan con la actividad antibiótica prevista por la máquina. Todo esto es impulsado por la función de aptitud (*fitness function*) que establecimos para este estudio.

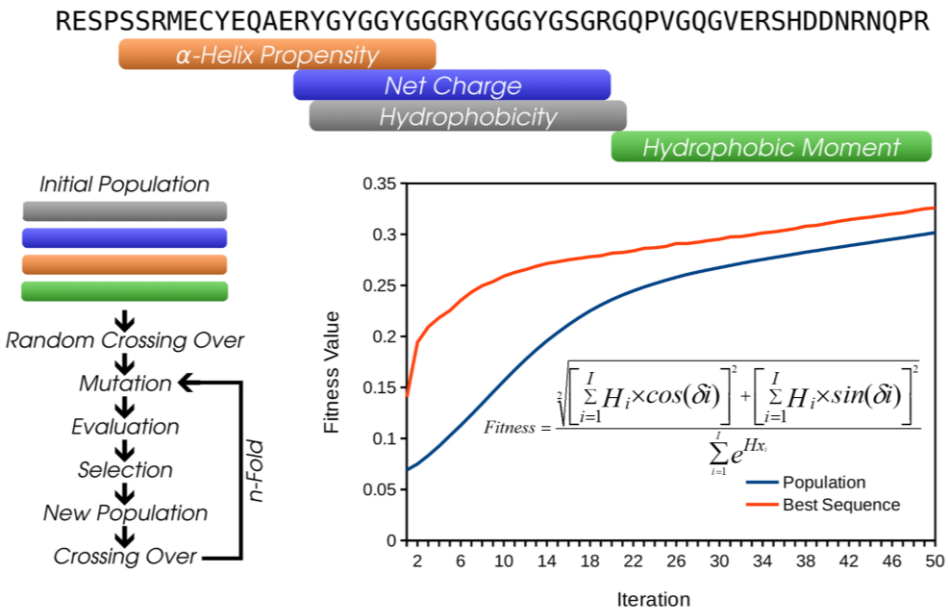


Figura 4. Ejemplo de una de las simulaciones
(Fuente: Porto et al. *Nature Communications,* 2018, 9:1490)

Este proyecto, que comenzamos alrededor del año 2015, tardó en materializarse en publicaciones. Fue en 2018 cuando salió el primer artículo porque necesitamos muchas iteraciones antes de empezar a publicar los resultados, ya que era algo que no se había hecho con anterioridad y todo era completamente nuevo. Este algoritmo de optimización también nos permitió explorar nuevas secuencias que antes no se encontraban en la naturaleza, dando lugar a moléculas con proporciones de aminoácidos diferentes a las que solemos encontrar en la biología de forma natural. Resultó muy emocionante ver esta generación de una nueva diversidad de secuencias, hechas por ordenador.

Figura 5. Ejemplo de las nuevas moléculas generadas (Colaboración con Franco Lab, Hancock Lab y otros)

El ordenador dio lugar a estas nuevas moléculas, generando péptidos con estructuras alfa-helicoidales, pero solo era una predicción que la máquina proponía porque pensaba que cada una de estas moléculas, cada uno de estos compuestos propuestos, podría ser un gran antibiótico. Sin embargo, no podíamos confiar en las suposiciones del ordenador, sobre todo teniendo en cuenta que nos encontrábamos en los pasos iniciales del proyecto y de este campo entero.

Por eso, tuvimos que idear métodos que nos permitieran validar si las propuestas realizadas por el ordenador eran correctas o no. Para ello, decidimos sintetizar químicamente cada una de esas moléculas generadas por el ordenador, utilizando una síntesis química de fase sólida, y probarlas con bacterias a las que teníamos acceso en el laboratorio. Estas bacterias eran, además, clínicamente relevantes para poder comprobar el efecto de esas moléculas sobre ellas y ver si eran capaces de matarlas. Llevamos a cabo una serie de pruebas de cribado que dieron como resultado una molécula a la

que llamamos guavanina, que fue la más exitosa, ya que conseguía matar bacterias de forma muy potente en dosis muy bajas, logrando esterilizar completamente los cultivos bacterianos en el laboratorio. Además, a través de la resonancia magnética nuclear (RMN), logramos dilucidar la estructura 3D de la molécula y hacernos una idea de su conformación tridimensional. Como se imaginarán, esto todo fue un avance realmente importante porque pudimos diseñar algo en el ordenador que era capaz de matar bacterias en el laboratorio, para optimizarlo en un tiempo récord.

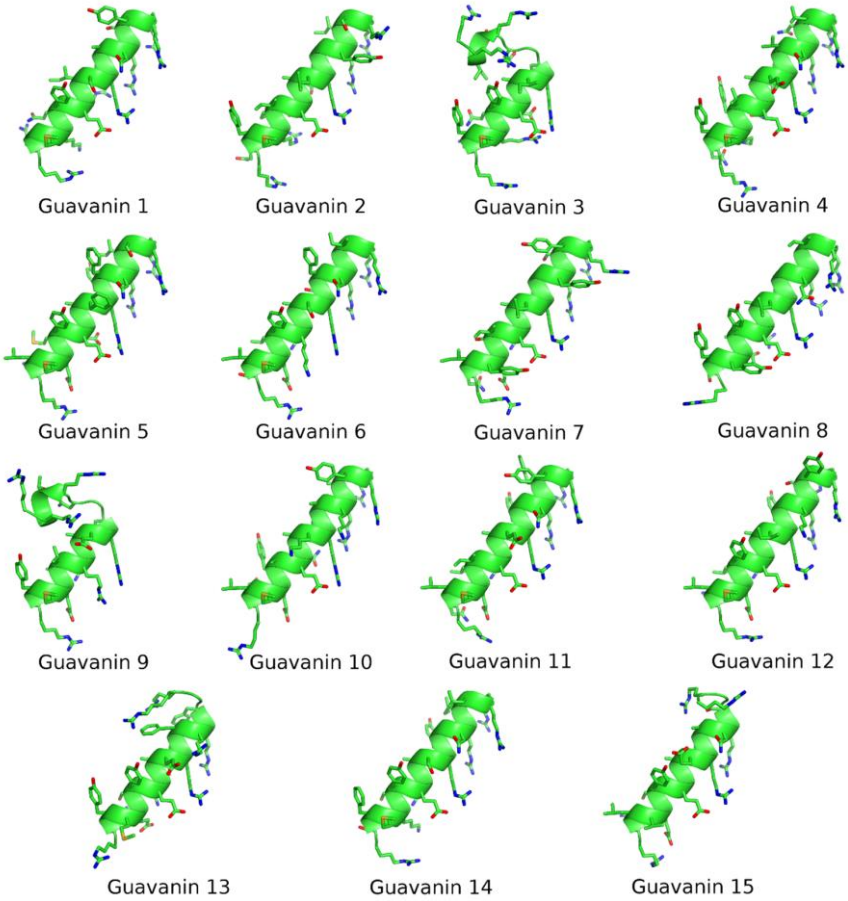


Figura 6. Moléculas generadas por el ordenador que podían funcionar como posibles antibióticos (Fuente: Porto et al. *Nature Communications,* 2018, 9:1490)

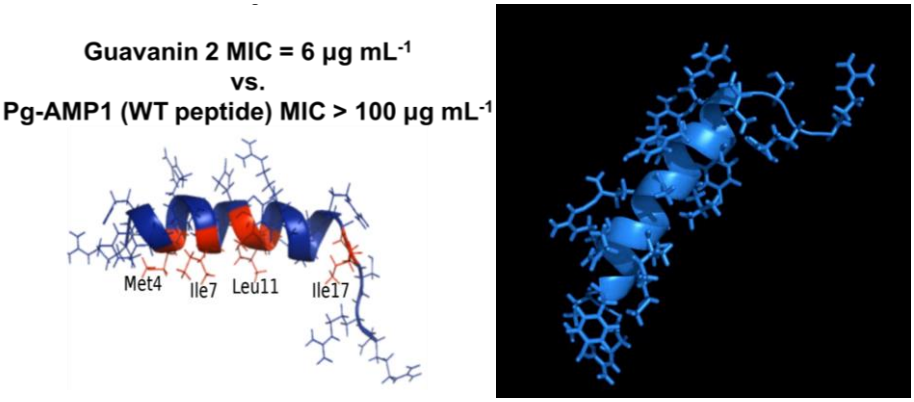


Figura 7. Datos experimentales de las pruebas realizadas y estructura 3D de guavanina (Fuente: Porto et al. *Nature Communications,* 2018, 9:1490)

Como comenté, una vez que habíamos sintetizado la molécula guavanina, pudimos validar en el laboratorio su actividad antimicrobiana, que resultó muy prometedora contra patógenos bacterianos clínicamente relevantes. Eso supuso un gran paso para todas las personas que trabajábamos en el proyecto, porque seguíamos pensando que todo iba a fracasar en cualquier momento. A continuación, decidimos averiguar cómo esa molécula conseguía matar a las bacterias, es decir, cuál era su mecanismo de acción. Para tratar de dilucidarlo, llevamos a cabo una serie de experimentos muy interesantes con los que obtuvimos los siguientes resultados.

Uno de los aspectos más reseñables que descubrimos es que, mientras que los antibióticos de este tipo descritos anteriormente mataban a las bacterias polarizando la membrana, la guavanina hacía lo contrario: mataba a las bacterias hiperpolarizando la membrana en lugar de despolarizándola. Esto fue realmente sorprendente, ya que, ninguno de los humanos (científicos y científicas) que estábamos trabajando en el proyecto había descrito estos conocimientos mecanicistas en la función de aptitud (*fitness function*) utilizada. Por eso, no pudimos explicarlo en el primer artículo que salió en 2018.

Pero, al pensar retrospectivamente en ese experimento, creo que ese pudo ser uno de los primeros ejemplos de las propiedades emergentes de la IA en la biología. Con esto quiero decir que, a veces, los algoritmos conducen a propiedades emergentes que los humanos que diseñan y escriben los programas informáticos no habíamos previsto. Creo que este fue uno de esos ejemplos iniciales en el complejo mundo de la biología.

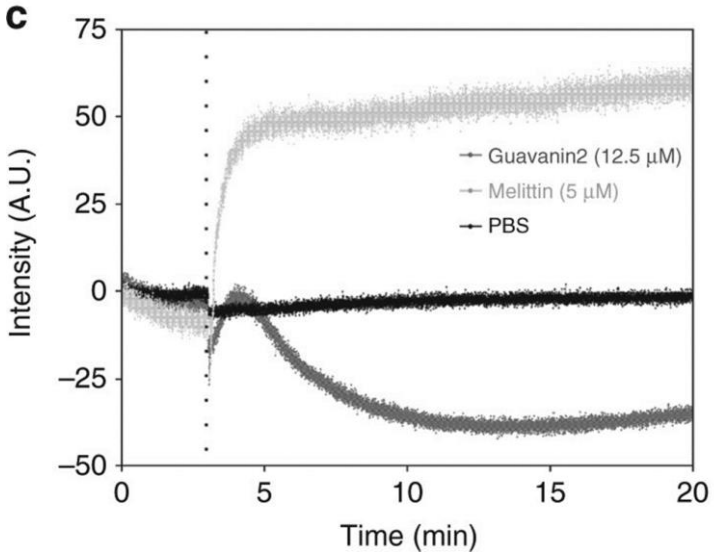


Figura 8. Resultados de las pruebas sobre el mecanismo de la molécula guavanina (Fuente: Porto et al. *Nature Communications,* 2018, 9:1490)

A continuación, quisimos comprobar si la molécula guavanina era capaz de reducir las infecciones en un modelo de ratón, porque queríamos ver la aplicabilidad de todo lo que estábamos creando en el ordenador. De esta forma, desarrollamos un modelo de infección de ratón, el de la infección de absceso cutáneo, e infectamos a los ratones con *Pseudomonas aeruginosa*, que es un patógeno Gram negativo.

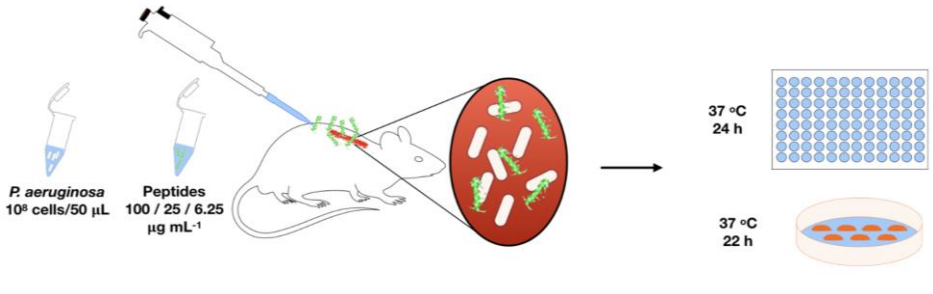


Figura 9. Resultados de las pruebas sobre el mecanismo de la molécula guavanina (Fuente: Porto et al. *Nature Communications,* 2018, 9:1490)

Lo que ocurrió fue que cuando se trataba a los animales con la molécula guanavina, la infección se reducía significativamente en comparación con el grupo de ratones de control no tratados. También tratamos a los ratones con dos de los compuestos con los que el programa informático inició el proceso de optimización, pero no estaban optimizados y no funcionaron bien. Sirvieron como controles negativos.

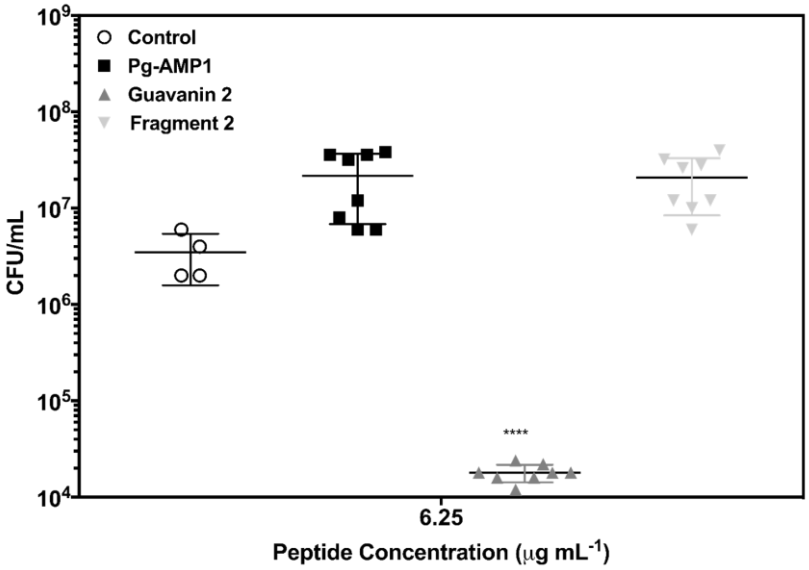


Figura 10. Resultados de las pruebas de la molécula guavanina en el modelo de ratón (Fuente: Porto et al. *Nature Communications,* 2018, 9:1490)

Sin duda, este experimento fue el que nos convenció, el que nos abrió los ojos a este nuevo campo emergente y corroboró nuestra idea de que los ordenadores y la IA podían utilizarse para crear nuevos tipos de antibióticos, de fármacos y terapias potencialmente nuevos, que no solo eran prometedores *in vitro,* sino también en modelos preclínicos de infección en ratones.

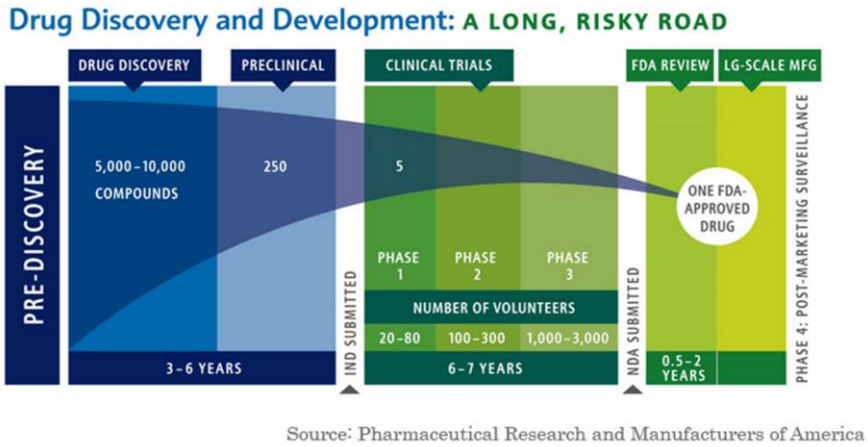
Desde que publicamos el artículo en 2018, muchos otros grupos de investigación se han unido a este campo emergente y han publicado trabajos sobre el desarrollo de métodos de IA para el descubrimiento y el desarrollo de antibióticos.

3. *MACHINE LEARNING* PARA EL DISEÑO DE FÁRMACOS

3.1. Ventajas: reducción de tiempo y de costes

Después de mi etapa en el MIT, me contrataron en la Universidad de Pennsylvania (UPenn), la institución en la que sigo trabajando desde entonces. Cuando empecé en UPenn, decidí abordar otra cuestión fundamental: ¿pueden utilizarse los ordenadores para acelerar el descubrimiento de antibióticos extrayendo toda la información de la biología?

Con información biológica, me refiero a genomas, proteomas, metagenomas y ese tipo de material orgánico. La idea estaba motivada por el hecho de que el descubrimiento de nuevos antibióticos es un proceso muy lento. Se puede tardar hasta seis años en llegar a nuevos candidatos antibióticos preclínicos y es, además, muy costoso. En EE.UU. el coste de desarrollo de un nuevo antibiótico supera los dos mil millones de dólares, que equivale al presupuesto con el que cuentan la NASA o SpaceX para llevar un cohete o un transbordador espacial al espacio exterior o a la luna. Por eso, pensé en la IA para ayudar a acelerar los procedimientos y ser capaces de reducir no solo el tiempo, sino también el coste asociado al descubrimiento de antibióticos.



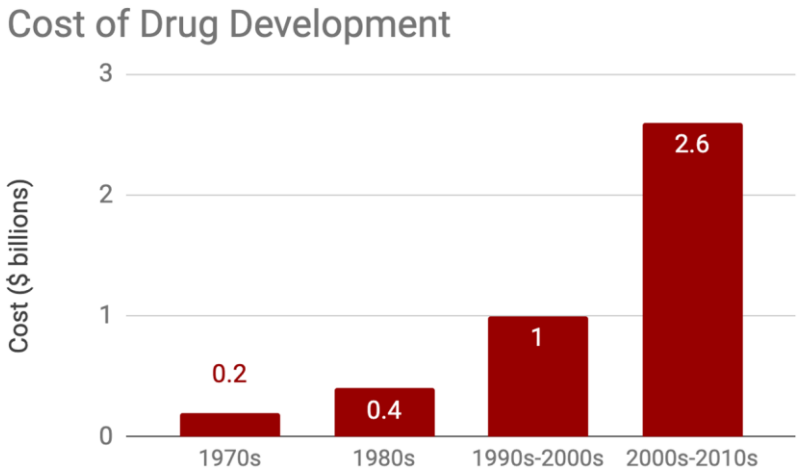


Figura 11. Tiempo y coste del descubrimiento y desarrollo de nuevos antibióticos (Fuente: *Scientific American*)

3.2. Minería de proteínas para el descubrimiento de antibióticos

Como he mencionado antes, cuando estaba en el MIT, aprendí los fundamentos de los algoritmos de IA para el reconocimiento de voz e imágenes que usaban allí. Al iniciar el

proyecto en UPenn, en el laboratorio pensamos que esos algoritmos serían ideales para abordar, desde la perspectiva de minar la biología, el descubrimiento de nuevos patrones. Por supuesto, en nuestro caso, en lugar de intentar reconocer expresiones faciales o patrones sonoros, lo que buscábamos era reconocer patrones moleculares codificados en genomas y proteomas que pudieran corresponderse con potenciales moléculas antibióticas. Esto volvía a ser un mundo completamente nuevo.

Empezamos con proyectos muy sencillos, en los que desarrollamos algoritmos de reconocimiento de patrones que podían encontrar fragmentos de péptidos escondidos como mensajes ocultos en proteínas individuales. En la imagen (Figura 12), se puede ver una proteína en tres dimensiones, que se muestra en dos dimensiones con el código de aminoácidos que aparece en el eje Y del gráfico. Lo que hacen los algoritmos es ir a través de este código de aminoácidos e identificar las regiones que, en sus predicciones, figuran como posibles secuencias antimicrobianas. En el gráfico, estas secciones están coloreadas en función de la puntuación antimicrobiana calculada (los colores más cálidos se corresponden con la función antimicrobiana más alta). Una vez más, se trata solo de una predicción que se ha impulsado algorítmicamente, pero es una predicción muy poderosa porque nos permite minar proteínas de manera individual e identificar, dentro de la proteína, fragmentos o regiones que el algoritmo ha reconocido como buenos antimicrobianos. Por ejemplo, la estructura amarilla de la imagen de la izquierda sería una de esas secuencias.

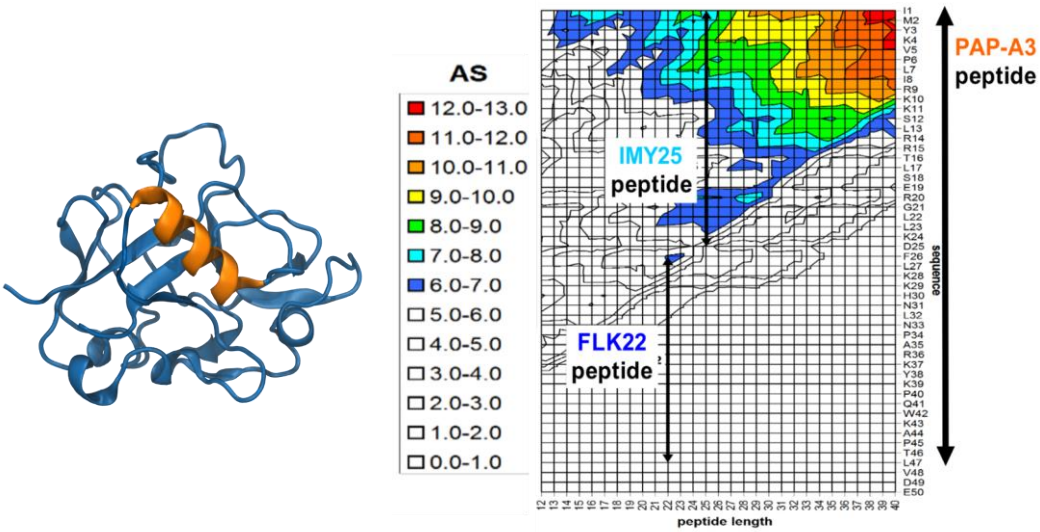


Figura 12. Minería de proteínas para el descubrimiento de antibióticos (Fuente: Pane et al. *ACS Synthetic Biology,* 2018, 7:2105-2115)

Una vez identificadas esas proteínas y secuencias, utilizando métodos químicos y de biología sintética, podemos extraer esa región en el laboratorio y tratar de desarrollarla para que se convierta en algo potencialmente útil. Así lo hicimos con algunos colaboradores y colaboradoras en varios proyectos en los que utilizamos algoritmos muy sencillos para extraer proteínas individuales como fuente de moléculas antimicrobianas ocultas. Entonces, no sabíamos a dónde nos llevaría todo esto.

3.3. El proteoma humano, una fuente de antibióticos

Con los avances en la potencia de cálculo y de los algoritmos computacionales, nos dimos cuenta de que podíamos pasar de la extracción de proteínas individuales a la de proteomas enteros. Un proteoma está compuesto por todas las proteínas

producidas por un organismo, que a la vez están codificadas en el genoma. Esto fue lo que nos animó a embarcarnos, por primera vez, en un proyecto muy ambicioso que jamás se había intentado antes. Decidimos tratar de minar el proteoma humano como fuente de antibióticos. Este trabajo, que alcanzó notoriedad y se convirtió en la portada de la revista *Nature Biomedical Engineering* en 2022, abrió un nuevo horizonte intelectual. En su desarrollo, no solo exploramos las 20.000 proteínas prototípicas, sino que también tuvimos en cuenta las isoformas. Como resultado, el algoritmo minó 42.000 proteínas, lo que equivale a un total de unos 100 millones de péptidos dentro de todas esas secuencias de nuestro propio cuerpo.

Si les soy sincero, basándonos en nuestros cálculos iniciales, habíamos previsto que este proceso nos llevaría más tiempo; muchas horas, tal vez varios días. Recuerdo perfectamente el momento en el que pusimos en marcha el algoritmo. Cuando, al cabo de una hora, los miembros del equipo me dijeron que la exploración ya había concluido, me quedé estupefacto. Les dije: "Esto no es posible. Volved al inicio para asegurarnos de que no hay errores en el código. Aseguraos de que todo esté correcto."

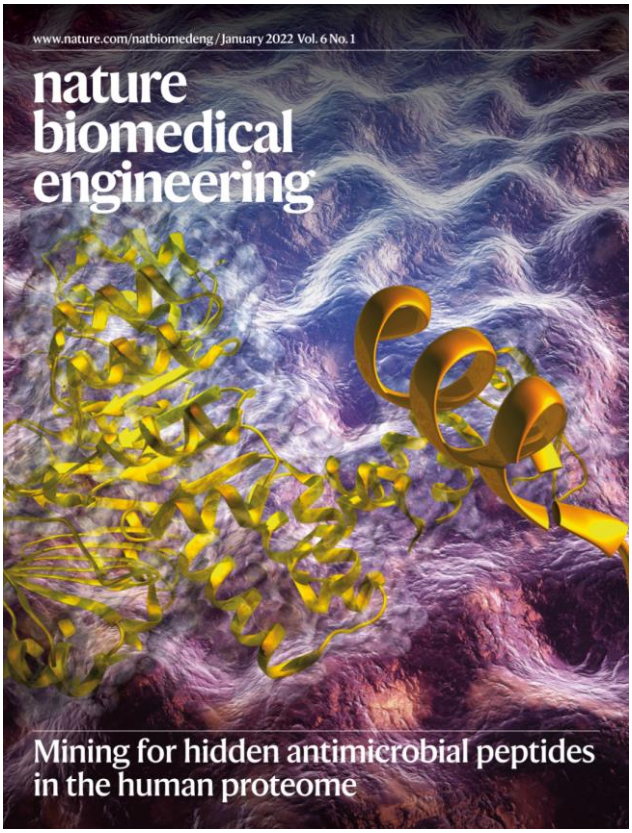


Figura 13. Portada de *Nature Biomedical Engineering* con el proyecto de minería a gran escala (*Nature Biomedical Engineering*, 2022)

El equipo entonces realizó una comprobación muy minuciosa y no detectó ningún problema. Todo estaba en orden. Entonces volvimos a realizar los cálculos. Básicamente, habíamos sobreestimado la cantidad de tiempo que nos llevaría el procedimiento porque nuestro algoritmo era realmente muy simple. Era computacionalmente muy barato. En aquel momento, solo contábamos con 48 CPUs en el laboratorio, e, incluso así, el proceso fue increíblemente rápido, pero con resultados notables.

En términos más prácticos, lo que este enfoque computacional nos permitió fue descubrir nuevas moléculas,

muchos miles de lo que llamamos antibióticos encriptados, es decir, que están ocultos o estaban previamente ocultos dentro de nuestro propio cuerpo. Estos mensajes moleculares ocultos en nuestras propias proteínas quizás cumplan un papel en la inmunidad o como antimicrobianos de cualquier tipo que previamente no se conocía. Constituyen un nuevo tesoro, descubierto por ordenadores, y tenemos muchas ganas de seguir descifrando sus secretos.

3.4. Péptidos encriptados: un nuevo tipo de moléculas

Al descubrir esos miles de antibióticos encriptados en nuestras proteínas, lo que hicimos fue caracterizarlos y sintetizarlos para tratar de aprender de ellos. En esos nuevos experimentos, observamos que mostraban un amplio espectro de actividad antimicrobiana, ya que eran capaces de matar bacterias Gram-negativas y Gram-positivas. Esto nos llevó a pensar que podrían desempeñar un papel en la inmunidad del huésped. Además, comprobamos que muchas de esas secuencias podrían ser cortadas por proteasas de la proteína madre, que suele tener una función no relacionada.

La proteína madre puede desempeñar un papel en el sistema nervioso, digestivo, cardiovascular, etc. Pero, en nuestros experimentos, observamos que estos productos escindidos de esa proteína tenían un papel antimicrobiano que indica que, quizás, cumplan un rol en el sistema inmunológico. Barajamos varias hipótesis al respecto y hay una en la que ya estamos pensando más seriamente. Creemos que las proteínas no inmunes están desempeñando un papel en el sistema inmune, algo que no se había reconocido hasta el momento. Esto establece un vínculo entre el sistema inmunológico y muchos otros sistemas del cuerpo, como el sistema nervioso, por ejemplo.

A partir de ahí, llevamos a cabo más experimentos y los resultados mostraron que si tomas algunas de estas moléculas ocultas o codificadas de la misma zona del cuerpo, por ejemplo, del torrente sanguíneo o del cerebro, y las combinas en cócteles, establecen sinergias entre sí para atacar a los patógenos. Es decir, potencian la actividad de las otras moléculas para atacar a los patógenos relevantes. Esto concuerda con la idea que teníamos de que, en caso de infección, muchas de estas moléculas codificadas se liberan en nuestro organismo para proporcionar una primera línea de defensa contra los patógenos invasores.

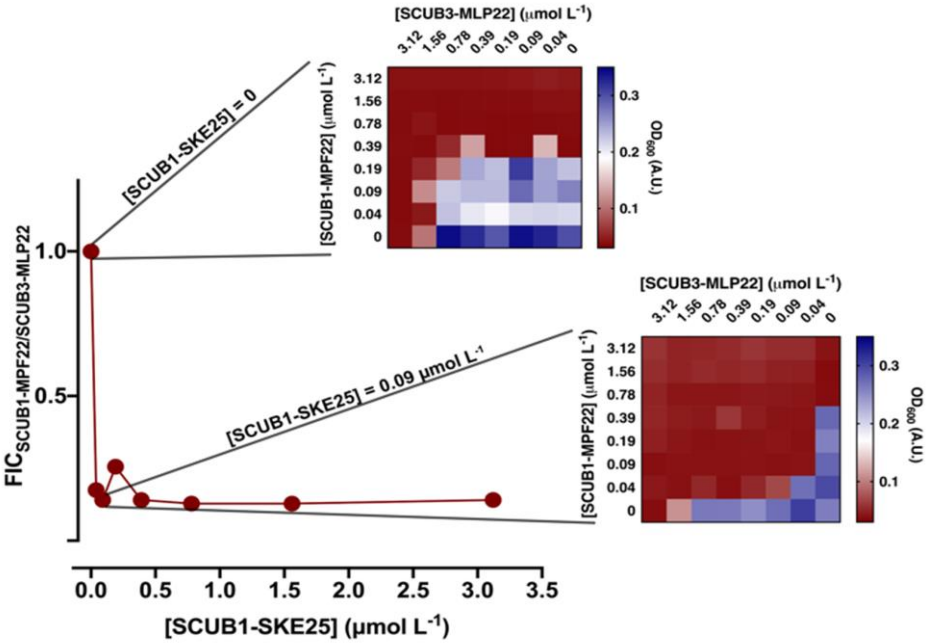


Figura 14. Resultados de los experimentos realizados con cócteles de las moléculas encriptadas para matar a bacterias patógenas (Fuente: Torres et al. *Nature Biomedical Engineering,* 2022)

También observamos, a través de un experimento de desarrollo de resistencia, que a las bacterias les cuesta mucho desarrollar resistencia a estas secuencias codificadas. Expusimos cada

día un patógeno bacteriano llamado *Acinetobacter baumannii* a la polimixina B, que es un antibiótico convencional, y después a tres de las secuencias codificadas del proteoma humano (SCUB1-SKE25, SCUB3-MLP22 y SCUB1-MPF22). Los resultados mostraron que, con el tiempo, este patógeno desarrollaba resistencia a la polimixina B (en color gris en el gráfico de la Figura 15), pero no a los tres compuestos descubiertos en el proteoma humano. Por tanto, resulta alentador que las bacterias no desarrollen fácilmente resistencia a estas moléculas de nuestro propio cuerpo.

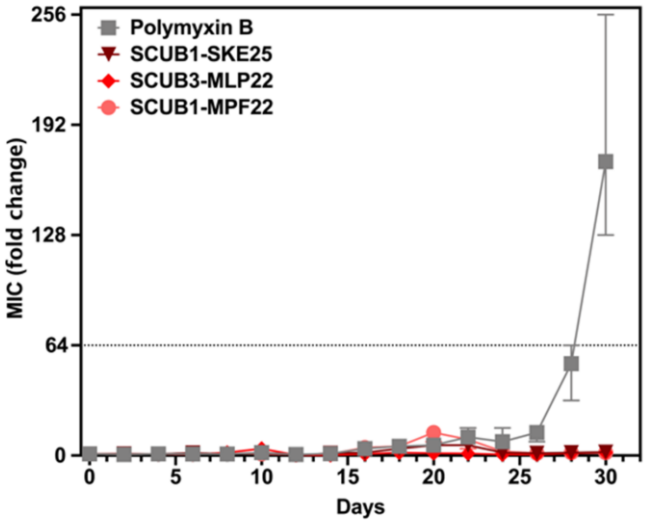


Figura 15. Resultados del experimento de desarrollo de resistencia (Fuente: Torres et al. *Nature Biomedical Engineering,* 2022)

A la vista de estos resultados, decidimos comprobar la eficacia antiinfecciosa de estas moléculas y su seguridad en dos modelos preclínicos diferentes de infección en ratones. El primero era un modelo de infección cutánea por absceso, similar al que he presentado antes. Realizamos varios experimentos diferentes, incluyendo el uso de distintos patógenos en los ratones (*P. aeruginosa* PA01 y *A. baumannii* ATCC19606). En general, se

observó una disminución de la infección en monoterapia con un péptido o el otro, y cuando se combinaron en un cóctel (en color púrpura en el gráfico de la Figura 16), se produjo una reducción más significativa de la infección. Por tanto, se evidenciaban las interacciones sinérgicas que también habíamos observado *in vitro*.

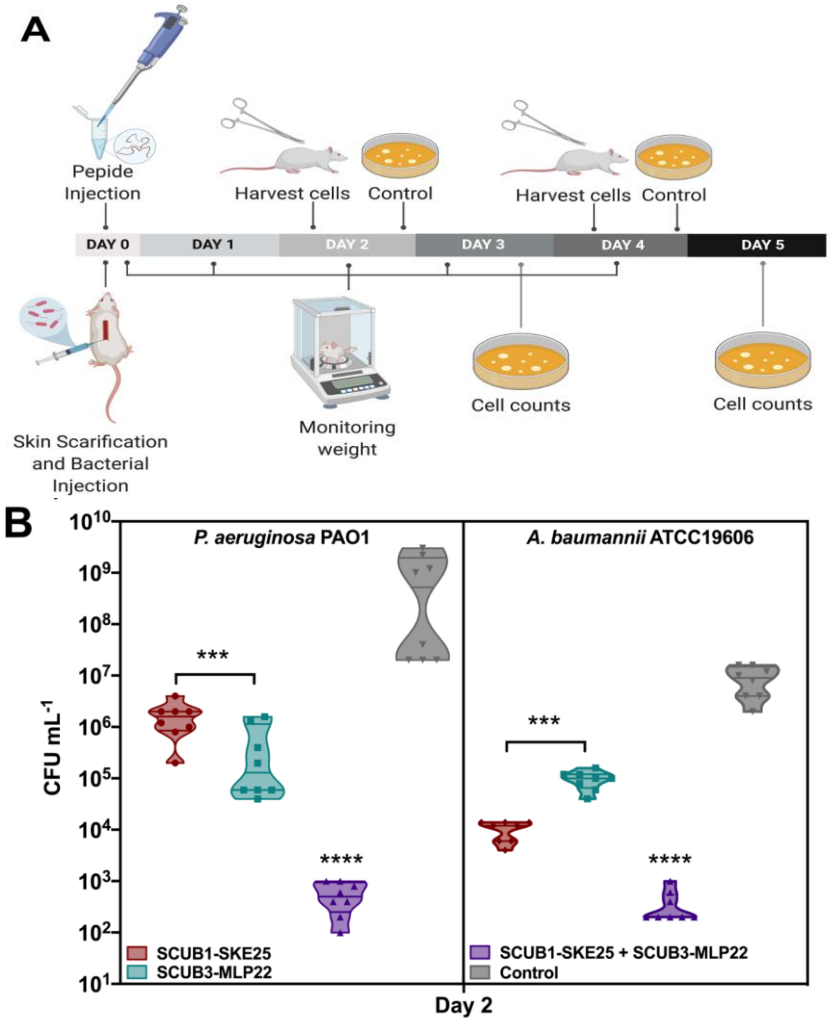


Figura 16. Resultados del experimento de la infección cutánea por absceso en ratones (Fuente: Torres et al. *Nature Biomedical Engineering,* 2022)

El otro modelo preclínico de ratón que usamos fue el de una infección profunda del muslo, que es el candidato habitual que se suele utilizar para las propuestas que pretenden superar la evaluación de la FDA en Estados Unidos. Al igual que en el modelo anterior, se observó un efecto bactericida bastante pronunciado y muy buenos perfiles de seguridad de estos compuestos en el contexto de este modelo de ratón.

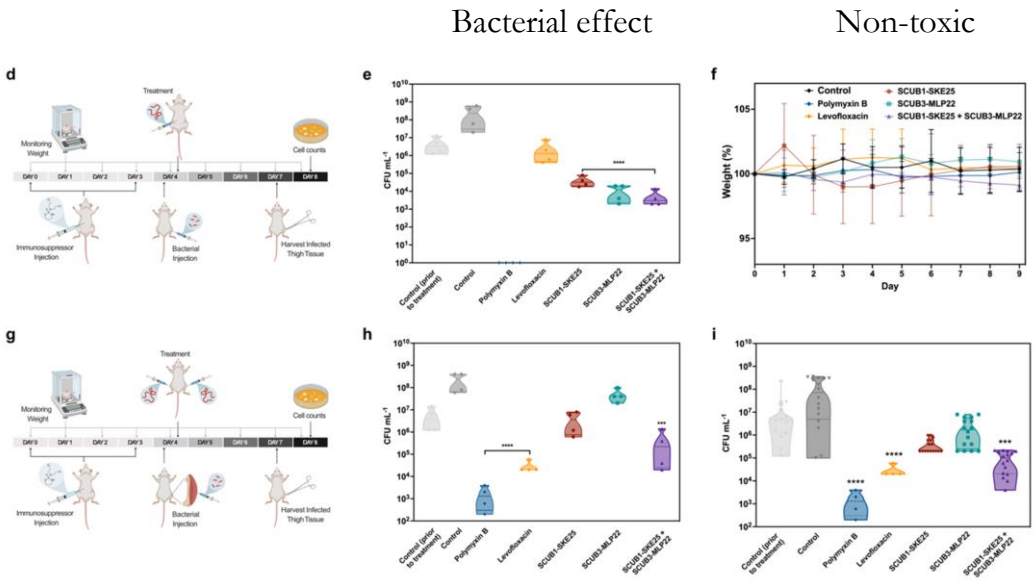


Figura 17. Resultados del experimento de la infección profunda del muslo en ratones (Fuente: Torres et al. *Nature Biomedical Engineering,* 2022)

3.5. Desextinción molecular: las moléculas como fuentes de documentación histórica

Durante el desarrollo de todos los experimentos anteriores, encontramos miles de moléculas en el genoma y proteoma humanos, y hemos continuado caracterizando muchas de ellas, incluyendo, por ejemplo, muchas de las secuencias peptídicas del plasma humano.

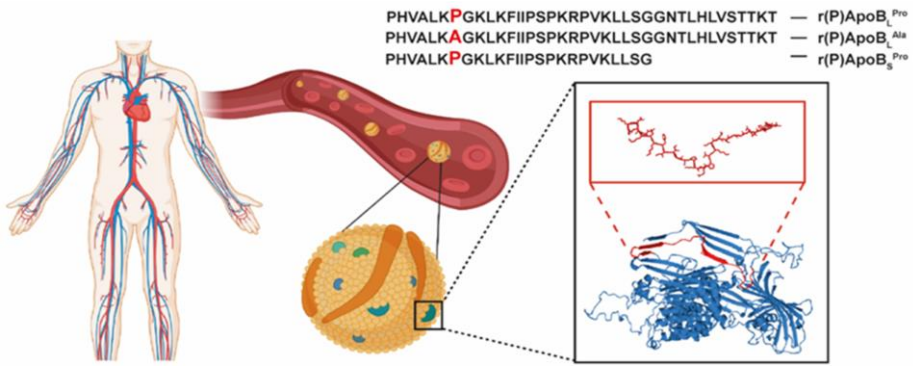


Figura 18. Secuencias antimicrobianas peptídicas descubiertas en el plasma humano (Fuente: Cesaro et al. *ACS Nano,* 2022)

Estamos trabajando también con otras secuencias del sistema nervioso y de otros sistemas del cuerpo humano. Sin embargo, creo que, conceptualmente, lo más importante de este trabajo es que dio lugar a muchas ideas dentro del laboratorio. Así surgió la hipótesis de que, si habíamos encontrado estas moléculas en el proteoma humano, lo más probable era que no solo pudiéramos encontrarlas allí, sino que tal vez estuvieran extendidas por todo el árbol de la vida, y decidimos investigarlo. Como el punto de partida de nuestro trabajo había sido el proteoma humano, nos pareció lógico intentar buscar secuencias similares en nuestros parientes más cercanos, los Neandertales y los Denisovanos.

Figura 19. Ejemplos de Neandertal y Denisovano (Fuentes: Shutterstock.com; Artículos relevantes: Maasch et al. *Cell Host & Microbe,* 2023; Maasch et al. *bioRxiv,* 2022)

Para abordar todo esto, se nos ocurrió el concepto de "desextinción molecular". Con él, queríamos hacer referencia a la noción de devolver a la vida moléculas del pasado o a lo largo de la historia evolutiva para abordar problemas actuales, incluido el problema de la resistencia a los antibióticos y quizá también muchos otros problemas biomédicos a los que nos enfrentamos hoy en día.

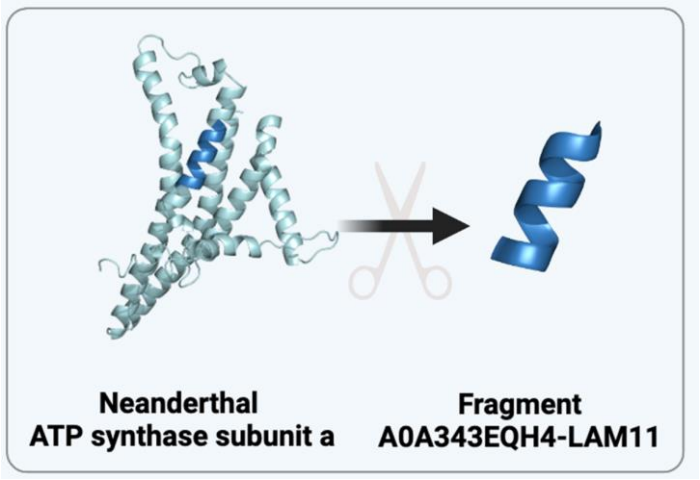


Figura 20. Desextinción molecular: un nuevo marco para el descubrimiento de moléculas (Fuentes: Maasch et al. *Cell Host & Microbe,* 2023; Maasch et al. *bioRxiv,* 2022)

3.6. panCleave

Para poder realizar la desextinción molecular de manera exitosa, tuvimos que desarrollar un algoritmo nuevo, que llamamos panCleave. Se trata de una herramienta informática que utiliza el aprendizaje automático (*machine learning* en inglés) para realizar proteólisis computacional, es decir, fragmenta las secuencias de proteínas humanas en péptidos. Les voy a explicar de manera sucinta su funcionamiento. En primer lugar, tomamos las proteínas o el proteoma de los seres humanos, en este caso, de los extintos y

de los modernos, estos últimos actúan como grupo control. Entonces, el algoritmo selecciona pequeños trozos de proteínas y luego se aplican los filtros de aprendizaje automático y de expertas y expertos humanos de nuestro laboratorio para predecir si esos fragmentos podrían actuar como antimicrobianos o no.

Digo que se aplica el filtro humano porque algo que hacemos con mucha seriedad en nuestro laboratorio es que siempre, para informar lo que vamos a sintetizar y lo que vamos a caracterizar a nivel experimental, se realiza un trabajo colaborativo entre la máquina y los investigadores e investigadoras del laboratorio. Es decir, la toma de decisiones de la máquina siempre está supervisada por un ser humano. Pensamos que esta colaboración estrecha entre la inteligencia de las máquinas y el ingenio humano es el presente y el futuro de la humanidad.

En la imagen (Figura 21), se pueden ver tres ejemplos de proteomas arcaicos y de proteomas modernos de algunas de las moléculas antimicrobianas que descubrió el algoritmo. Por ejemplo, la proteína que se ve más a la derecha en la fila de los homínidos arcaicos proviene de los Neandertales. Es una proteína neandertal. Es una ATP sintasa subunidad A, que es la responsable de producir ATP, la molécula de energía. El algoritmo fue capaz de encontrar el antibiótico (coloreado en azul en la imagen) en esa proteína. Se trata de una estructura alfa-helicoidal, a la que rebautizamos, tras su descubrimiento, como Neandertalina. Es la primera molécula terapéutica descrita en un organismo extinto, ya que después comprobamos que era eficaz en modelos preclínicos de ratón. Como se imaginarán, este también fue un descubrimiento muy importante para nuestro equipo. Un hito dentro del nuevo campo de la desextinción molecular.

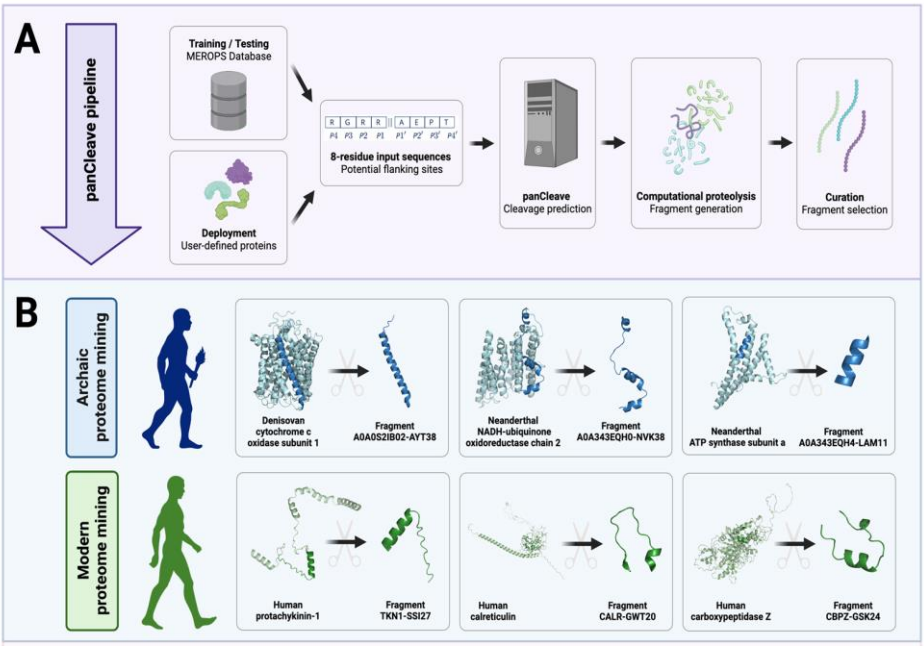


Figura 21. Desextinción molecular realizada por el algoritmo panCleave ((Fuentes: Maasch et al. *Cell Host & Microbe,* 2023; Maasch et al. *bioRxiv,* 2022)

No obstante, en esta etapa del proyecto, esto era solo una predicción computacional, así que lo que tuvimos que hacer fue resucitar muchas de estas moléculas recurriendo a la química. Hablo de resurrección porque, por lo que sabemos, muchas de ellas no se expresan en los organismos vivos de hoy en día. Son restos de antiguos organismos que ya no se expresan en la biología actual.

3.7. ¿Cómo funcionan los antiguos y los nuevos antibióticos?

Ahora voy a contarles un par de cosas que hemos aprendido de nuestro trabajo sintetizando muchas de estas moléculas a lo largo de la evolución y comparando las modernas con las antiguas. Algo que hemos observado es que las secuencias modernas tienden a

dirigirse a las bacterias a través de la membrana externa, mientras que las antiguas lo hacen a través de la membrana citoplasmática. Tienen mecanismos de acción diferentes. Esto es solo un ejemplo. Creemos que la comparación de moléculas a lo largo de la evolución puede ayudarnos a descubrir nueva biología que nos ayudará a entender la biología del pasado, pero también la biología del presente, y tal vez a predecir lo que va a suceder en la biología en el futuro.

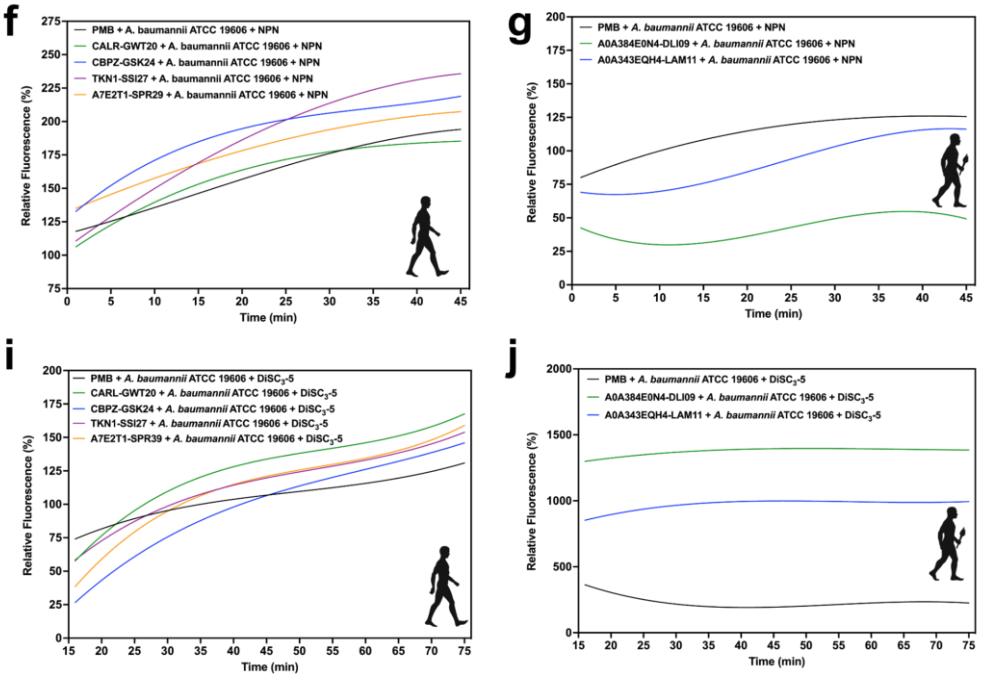


Figura 22. Diferentes mecanismos de acción de las moléculas antiguas y modernas (Fuentes: Maasch et al. *Cell Host & Microbe,* 2023; Maasch et al. *bioRxiv,* 2022)

Además, la molécula Neandertalina fue la que mostró la mayor eficacia antiinfecciosa en los dos modelos preclínicos diferentes de ratón: el modelo de absceso cutáneo y el modelo de infección de muslo. En definitiva, fue un gran éxito, ya que su eficacia fue similar a la del antibiótico polimixina B de tratamiento estándar, al menos en el modelo de absceso cutáneo.

Desde un punto de vista más conceptual, este trabajo fue la clave para convencernos de que era posible desarrollar métodos de IA para descubrir nuevas terapias en organismos extintos. Así, decidimos emprender un nuevo estudio y, como habíamos trabajado con esos tres proteomas (el del humano moderno, el de los Neandertales y el de los Denisovanos), era lógico continuar esta línea de investigación abordando todos los organismos extintos conocidos por la ciencia para comprobar si podíamos encontrar secuencias similares codificadas en todos ellos.

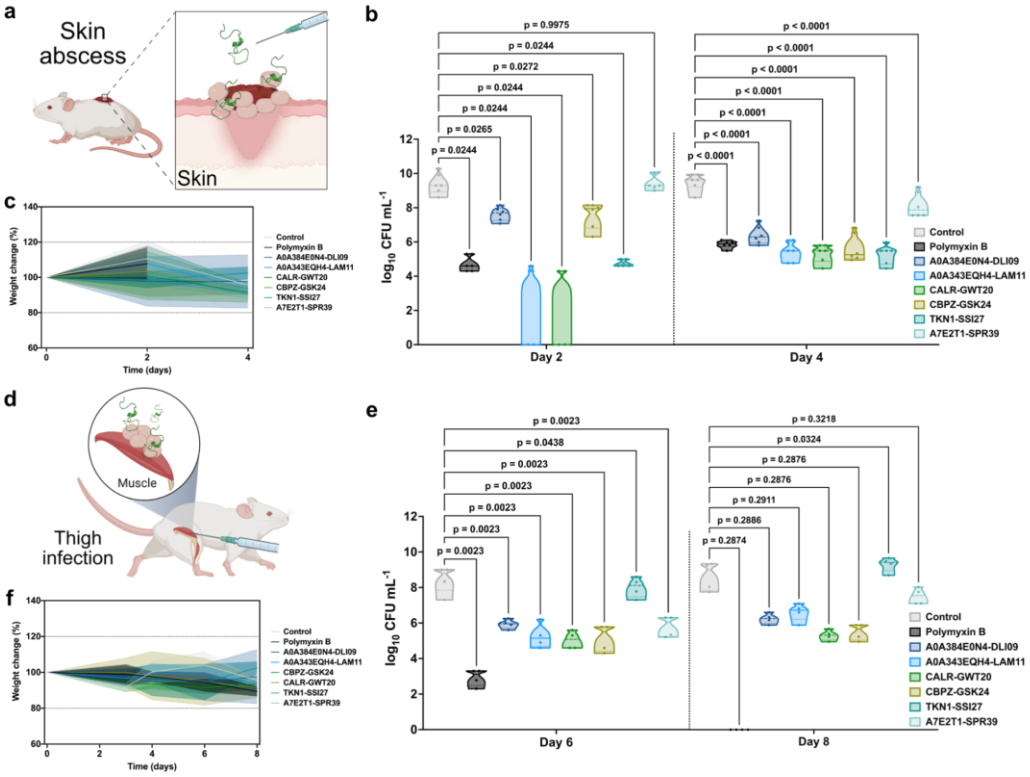


Figura 23. Eficacia de las distintas moléculas antimicrobianas en los dos modelos de ratón (Fuentes: Maasch et al. *Cell Host & Microbe,* 2023; Maasch et al. *bioRxiv,* 2022)

3.8. *Deep learning* para el descubrimiento de antibióticos

Este nuevo paso implicaba cientos de proteomas diferentes, por lo que tuvimos que desarrollar un modelo de aprendizaje automático más potente. Así nació APEX, un modelo de *deep learning* que diseñamos para tratar de abordar este nuevo reto. Lo utilizamos para muestrear lo que llamamos el extintoma, que es esencialmente el proteoma de todos los organismos extintos conocidos por la ciencia y que describe la literatura. No pudimos resistirnos a crear un nuevo término finalizado en -oma.

En la Figura 24, se puede ver el esquema en el que se recogen todos los representantes de la historia evolutiva que muestreamos en esta fase, incluyendo el Holoceno y el Pleistoceno. Se han destacado aquellos organismos en los que pudimos encontrar algunos de los principales candidatos a antibióticos que el algoritmo fue capaz de descubrir. Como se puede observar, hay pingüinos antiguos, elefantes antiguos, el mamut lanudo. También un perezoso gigante que tiene una historia muy interesante. De hecho, Charles Darwin descubrió restos de este perezoso gigante en una de sus expediciones a la Patagonia. Además, muestreamos muchos otros organismos, criaturas del pasado, algunas de ellas se habían extinguido hace cientos de miles de años. Este trabajo fue un *preprint* del que se hicieron eco muchos medios de comunicación y acaba de ser publicado en la revista *Nature Biomedical Engineering*.

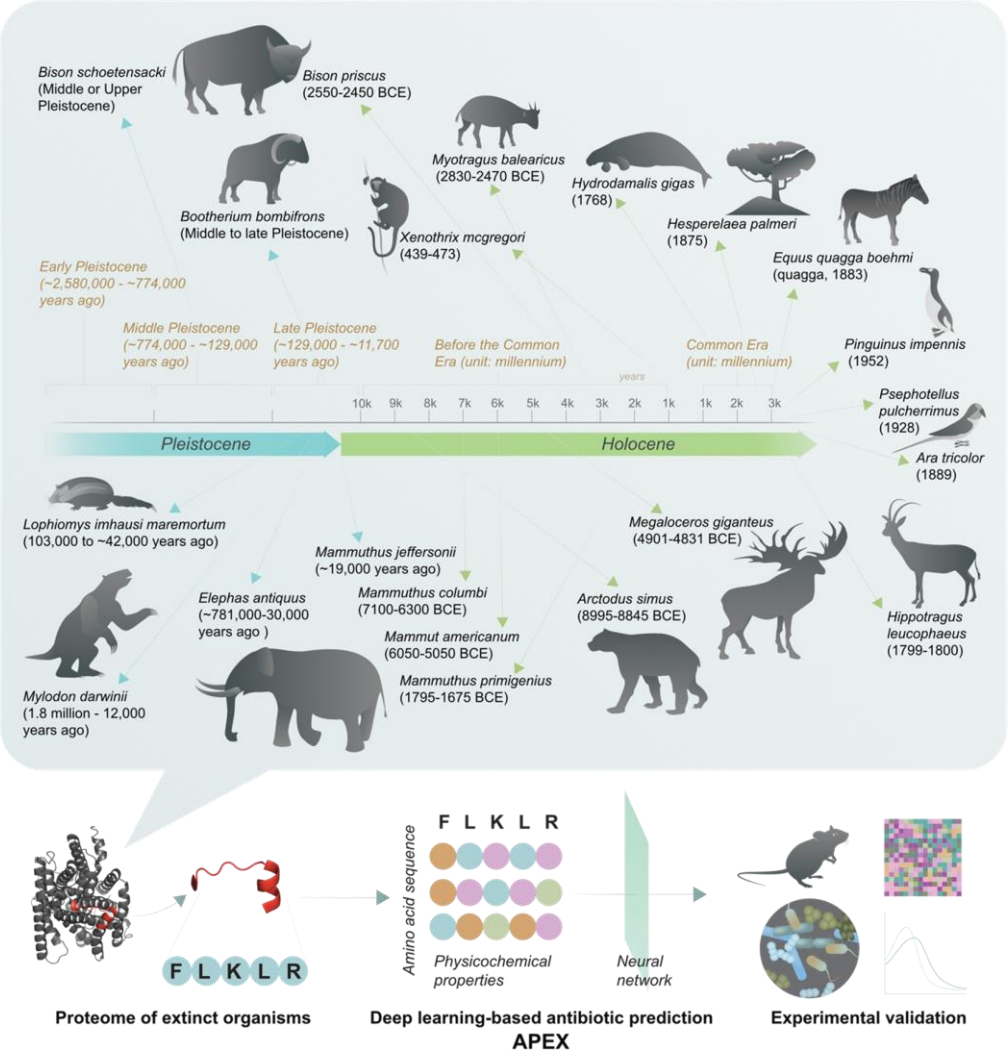


Figura 24. Esquema del extintoma muestreado con inteligencia artificial (Fuentes: Wan et al. *Nature Biomedical Engineering,* 2024; Wan et al. *bioRxiv,* 2023)

Lo que hicimos después fue sintetizar muchas de estas sustancias con métodos químicos y probarlas *in vitro* y también en modelos preclínicos de ratón. De nuevo, en dos modelos diferentes de ratones preclínicos. Los animales extintos que proporcionaron los antimicrobianos más prometedores fueron el mamut lanudo, la antigua vaca marina, el perezoso gigante y el alce gigante.

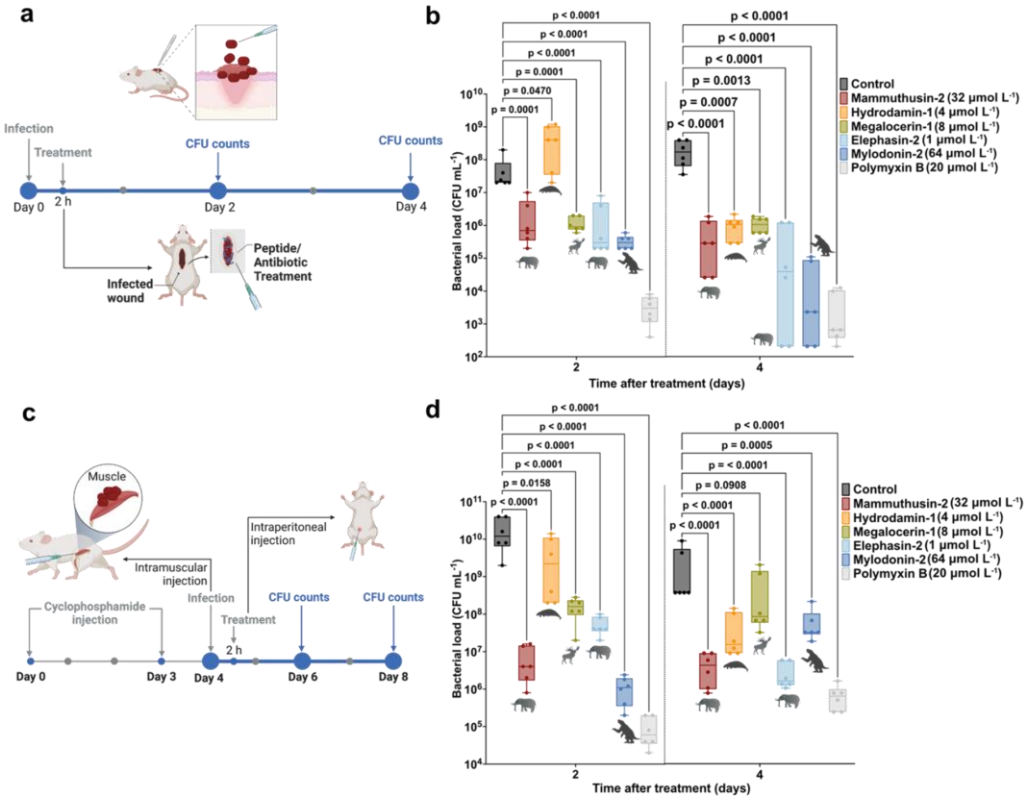


Figura 25. Resultados de los experimentos con el extintoma en los modelos de ratón de relevancia preclínica (Fuentes: Wan et al. *Nature Biomedical Engineering,* 2024; Wan et al. *bioRxiv,* 2023)

El descubrimiento de todas estas moléculas en organismos antiguos nos ha obligado a idear una nueva nomenclatura para nombrarlas, por lo que hemos establecido la convención de denominarlas en función del organismo que las produce. Así, por ejemplo, la Mammutusina proviene del mamut lanudo, y así con el resto (véase Figura 25).

3.9. ¿Por qué la desextinción molecular?

Puede que se estén preguntando ¿por qué la desextinción molecular? ¿Por qué molestarse en ir tan hacia atrás? La razón más

importante radica en que ya nos está permitiendo explorar un nuevo espacio de secuencias que antes no se rastreaba con ningún método. Además, creemos que, en el futuro, podrá contribuir a desentrañar una nueva biología, como ya he comentado. En el caso de nuestro propio trabajo, nos ha ayudado a ampliar nuestra visión de la vida, su diversidad molecular y espacio de secuencias. Hemos podido encontrar nuevas regiones del espacio de secuencias que son diferentes de lo que vemos en la biología viva. Intuimos también que muchas de estas moléculas que estamos descubriendo podrían desempeñar un papel en la inmunidad a lo largo de la evolución. En este sentido, esto puede ayudarnos a aprender más sobre el sistema inmunitario en el tiempo, nuestro propio sistema inmunitario, y cómo podría evolucionar en el futuro en respuesta a patógenos, brotes y pandemias. Por último, simplemente creemos que traer de vuelta moléculas del pasado puede ayudarnos a abordar muchos de los problemas a los que nos enfrentamos hoy en día, incluida la resistencia a los antibióticos, pero quizá también a otros retos.

Me gustaría también comentar un par de aspectos interesantes, que son ramificaciones de nuestro estudio. El primero de ellos es un planteamiento desde una perspectiva bioética. Esto se me ocurrió cuando empezamos a descubrir algunas de estas moléculas de organismos antiguos. Entonces me planteé qué significaba resucitar estas moléculas y si estaba bien hacerlo. Porque, por lo que sabemos, muchas de ellas ya no se producen, ya no se expresan en la biología viva actual, en los organismos vivos, en la naturaleza que nos rodea. Por eso, decidimos asesorarnos con personas expertas en bioética, que saben mucho más que nuestro equipo sobre este tema, para asegurarnos de que cualquier innovación que propongamos se haga con responsabilidad.

El segundo aspecto tiene que ver con las patentes y es bastante curioso. Cuando descubrimos estas moléculas, fui a la oficina de patentes de la Universidad de Pennsylvania y les dije: "Sé que las secuencias naturales no son patentables, ¿verdad? Pero ¿qué ocurre con las secuencias o moléculas que solían existir hace cientos de miles de años y ya no están presentes en la biología y la naturaleza actual? ¿Son patentables?" Creo que no hace falta decir que me miraron como si hubiera perdido la cabeza y tuve que explicarles el proyecto con más detalle para que lo comprendieran. Traigo esta reflexión aquí porque mi pregunta ha abierto una nueva área del derecho de patentes, ya que los expertos y expertas legales no están seguros de si las moléculas y secuencias extinguidas deben ser patentables. Además, he escrito un artículo de opinión con un colega que es abogado de patentes para examinar este problema de forma equilibrada, analizando los pros y los contras de patentar algunas de estas moléculas frente a no hacerlo. Es una cuestión sobre la que se seguirá trabajando en un futuro inmediato.

4. PRÓXIMOS PASOS

4.1. Minería del árbol de la vida

En nuestros próximos pasos, tenemos previsto explorar toda la información biológica disponible en el mundo. Esto incluye todos los genomas y proteomas. Queremos analizarlos para tratar de encontrar nuevas moléculas antibióticas u otras moléculas que pudieran ser de utilidad. Estamos iniciando ya colaboraciones en el ámbito del cáncer y también en otras áreas. Pero, tan solo en el ámbito de los proteomas, son muchísimas; estamos hablando de millones de proteínas. Por eso, tenemos que seleccionar los tipos de proyectos que vamos a abordar y establecer prioridades. Como primero iniciamos este campo explorando las eucariotas, observando a los humanos modernos y a los antiguos, decidimos seguir explorando el árbol de la vida y ahora estamos analizando la materia oscura bacteriana y la materia oscura de las arqueas (células procariotas) para ver si podemos encontrar nuevas moléculas de utilidad codificadas en toda esa información biológica. Para ello, contamos con varios colaboradores y colaboradoras de todo el mundo.

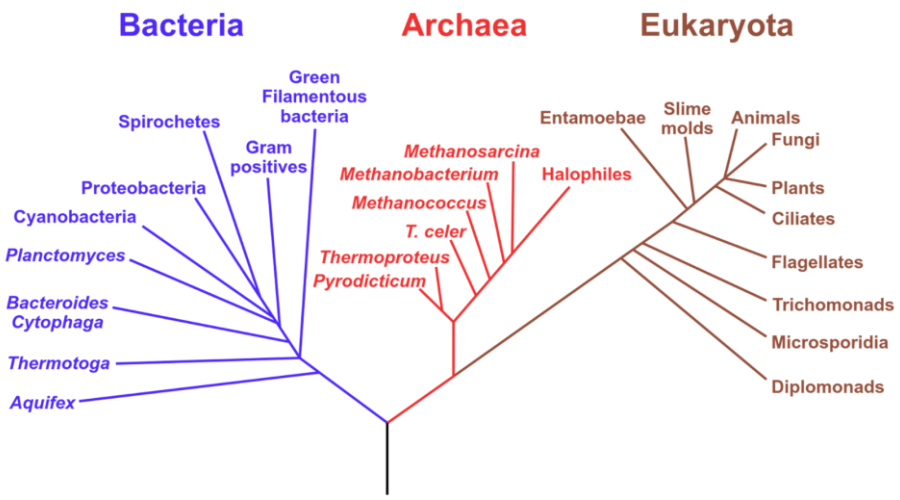


Figura 26. Minería del árbol de la vida con inteligencia artificial (Fuente: Wikimedia Commons)

4.2. Aceleración del descubrimiento de antibióticos con inteligencia artificial

Para terminar, me gustaría recordar lo que mencioné al principio sobre el tiempo de desarrollo de los antibióticos. Se puede tardar hasta seis años en encontrar candidatos clínicos con métodos tradicionales. Si nos fijamos en los avances que hemos logrado en el último lustro, hemos conseguido acelerar de manera sustancial ese proceso con la IA, y hemos demostrado que se pueden descubrir miles o cientos de miles de candidatos en cuestión de horas. Sin duda, esto es algo que no podría haber previsto hace cinco o seis años. Si me hubieran preguntado entonces si iba a ser posible, probablemente habría dicho que no. Realmente, ha sido un progreso fascinante. Por eso, he vivido con mucha emoción todo este proceso facilitado por los ordenadores, por los avances en los métodos de IA, que han logrado que las intuiciones que había anotado en aquel trozo de papel antes de finalizar mi doctorado se hayan hecho realidad.

Si hay una imagen que puede resumir el resultado de este trabajo del que estoy tan orgulloso es esta en la que aparecen algunos de los candidatos preclínicos que hemos sido capaces de descubrir con la IA (Figura 27). Todos ellos han sido validados en modelos preclínicos de ratón, así que estamos deseando ver qué depara el futuro para cada uno de ellos. Por supuesto, nuestro sueño es ver que alguno ayude a la gente en el futuro.

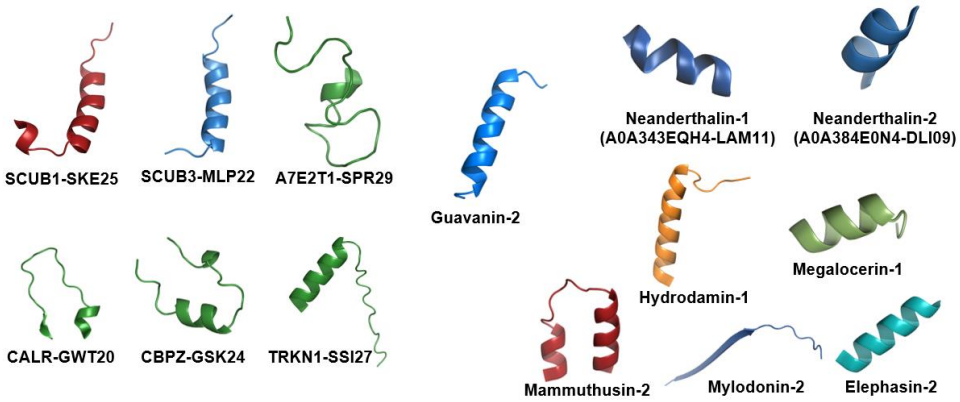


Figura 27. Candidatos preclínicos descubiertos con ordenadores en de la Fuente Lab (Fuentes: Torres et al. *Nature Biomedical Engineering,* 2022; Maasch et al. *Cell Host & Microbe,* 2023; Maasch et al. *bioRxiv,* 2022; Wan et al. *Nature Biomedical Engineering,* 2024; Wan et al. *bioRxiv,* 2023)

Por otra parte, debo decir que, desde que empecé con el laboratorio en la Universidad de Pennsylvania, hemos estado construyendo nuestro propio conjunto de datos internos de péptidos validados experimentalmente con sus respectivas puntuaciones antimicrobianas. Esta base de datos se ha construido utilizando condiciones estandarizadas (mismo pH, mismos medios), prestando mucha atención a la normalización de todas las condiciones previas para la generación de ese conjunto de datos internos. Muchos de los últimos métodos empleados en el laboratorio, incluyendo APEX, han sido entrenados en este

conjunto de datos internos. Con esto, quiero subrayar la importancia de disponer de buenos datos de entrenamiento en biología y química, porque son esenciales para que la IA tenga éxito en estos campos.

Así mismo, me parece importante recordar la juventud de esta línea de investigación porque los avances conseguidos con el uso de la IA para el descubrimiento de antibióticos en apenas un lustro han tenido gran repercusión. Permítanme que comparta el análisis retrospectivo que hemos realizado de una serie de publicaciones de este campo para presentar su evolución. En realidad, hasta 2018 cuando publicamos nuestro primer artículo, apenas había artículos en esta área, pero, desde entonces, como pueden ver en la imagen (Figura 28), las publicaciones han ido creciendo de forma casi exponencial.

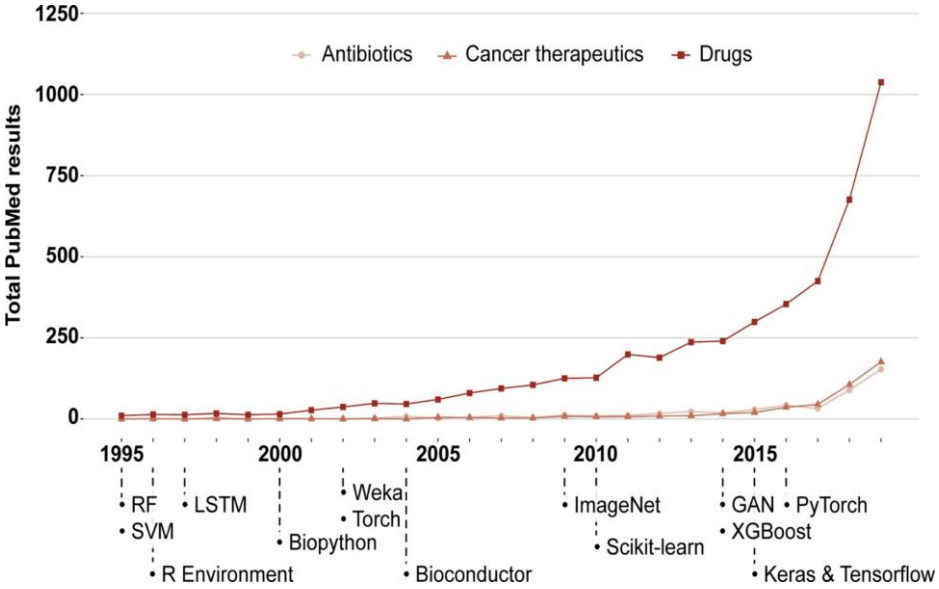


Figura 28. Análisis retrospectivo de las publicaciones sobre el uso de IA en el descubrimiento de antibióticos (Fuente: *Nature Communications Biology,* 2021; Otro artículo relevante: *Science,* 2023)

Se trata de un avance muy prometedor y estoy muy entusiasmado con el futuro de este campo. De hecho, siempre me gusta animar a las personas que están iniciando sus pasos en el mundo de la investigación, cuando muestran interés en los problemas de salud global y en su abordaje, a que se adentren en esta línea de investigación. Es un reto fascinante sobre el que trabajar y con un gran potencial de desarrollo.

Por último, me gustaría dar las gracias a todas las personas de nuestro laboratorio, las que trabajan actualmente en él, pero también a las que han pasado por el equipo, así como a las fantásticas personas con las que estamos colaborando a diario en todo el mundo y, por supuesto, a aquellas que contribuyen a la financiación de nuestro trabajo porque, sin ellas, no podríamos hacerlo. Me gustaría finalizar con una muestra de los hitos más importantes del De La Fuente Lab, en el que todas esas personas han ido sumando esfuerzos para alcanzarlos.

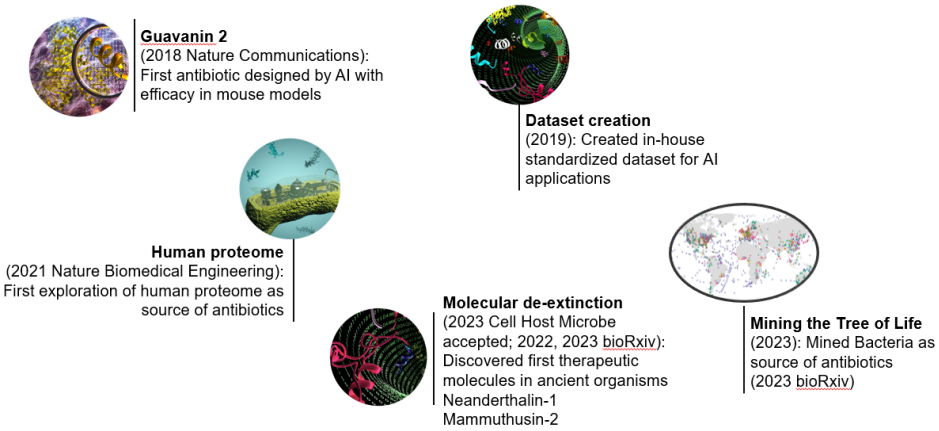


Figura 29. Hitos más importantes de de la Fuente Lab

5. BIBLIOGRAFÍA

Breidenstein, E.B.; de la Fuente-Núñez, C.; Hancock, R.E. Pseudomonas aeruginosa: all roads lead to resistance. *Trends Microbiol.* 2011 Aug;19(8):419-26. doi: 10.1016/j.tim.2011.04.005. Epub 2011 Jun 12. PMID: 21664819.

de la Fuente-Nuñez, C.; Cesaro, A.; Hancock, R.E.W. Antibiotic failure: Beyond antimicrobial resistance. Drug Resist Updat. 2023 Nov;71:101012. doi: 10.1016/j.drup.2023.101012. Epub 2023 Oct 18. PMID: 37924726.

Maasch, J.R.M.A.; Torres, M.D.T.; Melo, M.C.R.; de la Fuente-Núñez, C. Molecular de-extinction of ancient antimicrobial peptides enabled by machine learning. *Cell Host Microbe.* 2023 Aug 9;31(8):1260-1274.e6. doi: 10.1016/j.chom.2023.07.001. Epub 2023 Jul 28. PMID: 37516110.

Porto, W.F.; Irazazaba, l.L.; Alves, E.S.F.; Ribeiro, S.M.; Matos, C.O.; Pires, Á.S.; Fensterseifer, I.C.M.; Miranda, V.J.; Haney, E.F.; Humblot, V.; Torres, M.D.T.; Hancock, R.E.W.; Liao, L.M.; Ladram, A.; Lu, T.K.; de la Fuente-Núñez, C.; Franco, O.L. In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. *Nat Commun.* 2018 Apr 16;9(1):1490. doi: 10.1038/s41467-018-03746-3. PMID: 29662055; PMCID: PMC5902452.

Santos-Júnior, C.D.; Torres, M.D.T.; Yiqian, D.; Álvaro, R. del R.; Schmidt, T.S.B.; Hui, C.; Anthony, F.; Michael, K.; Chengkai, Z.; Amy, H., et al. 2024. Computational exploration of the global microbiome for antibiotic discovery. *Cell.* En prensa.

Torres, M.D.T.; de la Fuente-Núñez, C. Toward computer-made artificial antibiotics. *Curr Opin Microbiol.* 2019 Oct;51:30-38. doi: 10.1016/j.mib.2019.03.004. Epub 2019 May 11. PMID: 31082661.

Torres, M.D.T.; Melo, M.C.R.; Flowers, L.; Crescenzi, O.; Notomista, E.; de la Fuente-Núñez, C. Mining for encrypted peptide antibiotics in the human proteome. *Nat Biomed Eng.* 2022 Jan;6(1):67-75. doi: 10.1038/s41551-021-00801-1. Epub 2021 Nov 4.

Torres, M.D.T.; Brooks, E.; Cesaro, A.; Sberro, H.; Nicolaou, C.; Bhatt, A.S.; de la Fuente-Núñez, C. 2023. Human gut metagenomic mining reveals an untapped source of peptide antibiotics. bioRxiv. 10.1101/2023.08.31.555711.

Wan, F.; Kontogiorgos-Heintz, D.; de la Fuente-Núñez, C. Deep generative models for peptide design. *Digit Discov*. 2022 Mar 31;1(3):195-208. doi: 10.1039/d1dd00024a. PMID: 35769205; PMCID: PMC9189861.

Wan, F.; Wong, F.; Collins, J.J.; de la Fuente-Núñez C. Machine learning for antimicrobial peptide identification and design. *Nat Rev Bioeng*. 2024. https://doi.org/10.1038/s44222-024-00152-x

Wan, F.; Torres, M.D.T.; Peng, J.; de la Fuente-Núñez, C. 2024. Molecular de-extinction of antibiotics enabled by deep learning. *Nat Biomed Eng*. En prensa.

Wong, F.; de la Fuente-Núñez, C.; Collins, J.J. Leveraging artificial intelligence in the fight against infectious diseases. *Science*. 2023 Jul 14;381(6654):164-170. doi: 10.1126/science.adh1114. Epub 2023 Jul 13. PMID: 37440620; PMCID: PMC10663167.