

# PREDICCIÓN DE PROTEÍNAS DE NOVO

USANDO INTELIGENCIA ARTIFICIAL PARA LA INHIBICIÓN DEL FACTOR NF-KB EN EL CÁNCER GÁSTRICO

## DE NOVO PROTEIN PREDICTION USING ARTIFICIAL INTELLIGENCE FOR THE INHIBITION OF NF-KB FACTOR IN GASTRIC CANCER

Diego Laverde-Lomas<sup>1</sup>

**E-mail:** [dlaverde3525@uta.edu.ec](mailto:dlaverde3525@uta.edu.ec)

**ORCID:** <https://orcid.org/0009-0001-8012-0701>

Adahir Sarabia-Galarza<sup>2</sup>

**E-mail:** [asarabia@alumni.usfq.edu.ec](mailto:asarabia@alumni.usfq.edu.ec)

**ORCID:** <https://orcid.org/0009-0008-7646-908X>

Cinthia Galarza-Galarza<sup>3</sup>

**E-mail:** [ckgalarza@pucesa.edu.ec](mailto:ckgalarza@pucesa.edu.ec)

**ORCID:** <https://orcid.org/0000-0001-6822-7875>

Cristian Galarza-Galarza<sup>1</sup>

**E-mail:** [cf.galarza@uta.edu.ec](mailto:cf.galarza@uta.edu.ec)

**ORCID:** <https://orcid.org/0000-0001-7890-5043>

<sup>1</sup> Universidad Técnica de Ambato. Ecuador.

<sup>2</sup> Universidad San Francisco de Quito. Ecuador.

<sup>3</sup> Pontificia Universidad Católica del Ecuador. Ecuador.

**Cita sugerida (APA, séptima edición)**

Laverde-Lomas, D., Sarabia-Galarza, A., Galarza-Galarza, C., & Galarza-Galarza, C. (2025). Predicción de proteínas de Novo usando inteligencia artificial para la inhibición del factor NF-KB en el cáncer gástrico. *Revista UGC*, 3(S2), 39-48.

**Fecha de presentación:** 20/04/2025

**Fecha de aceptación:** 18/05/2025

**Fecha de publicación:** 01/06/2025

### RESUMEN

Abordamos las limitaciones de los tratamientos actuales contra el cáncer gástrico y exploramos el potencial de la Inteligencia Artificial (IA) en el desarrollo de estrategias personalizadas mediante la predicción de proteínas de novo diseñadas para inhibir el Factor Nuclear de las cadenas ligeras kappa de las células B activas (NF-KB), asociado con este tipo de cáncer. Se predijeron 30 proteínas con características similares a las almacenadas en bases de datos estándar y experimentales. Estas proteínas fueron evaluadas en términos de estabilidad y capacidad de plegamiento mediante simulaciones de dinámica molecular, analizando energías de interacciones intra e intermoleculares. Se realizó un acoplamiento molecular entre los genes y factores de transcripción regulados por NF-KB y las proteínas predichas. Se calcularon variables termodinámicas como Energía Libre de Gibbs, Constante de disociación, Entalpía, Capacidad Calorífica y Entropía, comparándolas con los complejos del Inhibidor Alpha de kappa B y las subunidades p65/p50 de NF-KB, con el fin de verificar la afinidad de interacción proteína-proteína y su conformación estructural. Los resultados mostraron alta afinidad y selectividad en las interacciones evaluadas, concluyendo que las proteínas de novo predichas mediante IA representan una alternativa prometedora para el diseño de nuevos fármacos y tratamientos dirigidos contra el cáncer gástrico.

**Palabras clave:**

Proteínas de novo, inteligencia artificial, cáncer gástrico, factor NF-KB.

### ABSTRACT

We address the limitations of current gastric cancer treatments and explore the potential of Artificial Intelligence (AI) in developing personalized strategies by predicting de novo proteins designed to inhibit the Nuclear Factor kappa light chain of active B cells (NF-KB), associated with this type of cancer. Thirty proteins with similar characteristics to those stored in standard and experimental databases were predicted. These proteins were evaluated in terms of stability and folding capacity by molecular dynamics simulations, analyzing energies of intra- and intermolecular interactions. Molecular docking was performed between the genes and transcription factors regulated by NF-KB and the predicted proteins. Thermodynamic variables such as Gibbs Free Energy, Dissociation Constant, Enthalpy, Heat Capacity and Entropy were calculated, comparing them with the complexes of the Alpha inhibitor of kappa B and the p65/p50 subunits of NF-KB, in order to verify the protein-protein interaction affinity and its structural conformation. The results showed high affinity and selectivity in the evaluated interactions, concluding that the de novo proteins predicted by AI represent a promising alternative for the design of new drugs and treatments directed against gastric cancer.

**Keywords :**

De novo proteins, artificial intelligence, gastric cancer, NF-KB factor.

## INTRODUCCIÓN

A pesar de los esfuerzos de la comunidad médica, el cáncer se ha convertido en una de las principales causas de muerte a nivel global, provocando alrededor de 10 millones de defunciones en el año 2020 (World Health Organization [WHO], 2022). Entre los diferentes tipos de cáncer, el cáncer gástrico es considerado como el tercer tipo de cáncer con mayor mortalidad en el mundo debido al incremento de su incidencia en poblaciones menores de 40 años de países europeos, asiáticos y latinoamericanos, en donde, la tasa de supervivencia de los pacientes suele ser del 30 % durante los primeros 5 años tras su diagnóstico. Además, su desarrollo presenta una gran influencia de factores genéticos y ambientales que incluyen: el alto consumo de alcohol, el tabaquismo, el alto consumo de sales e infecciones por *Helicobacter pylori* y virus de Epstein-Barr, lo que produce en un crecimiento anormal de células tumorales en el revestimiento del estómago, lo que comúnmente no presenta síntomas hasta etapas muy avanzadas de la enfermedad, donde ya puede extenderse hasta órganos subyacentes como el hígado y el páncreas.

Entre las técnicas para su tratamiento se encuentran: la radioterapia y la quimioterapia que son las más comunes durante el desarrollo de ensayos clínicos, sin embargo, este tipo de técnicas se encuentran limitadas por su incapacidad para actuar en secciones específicas (Joshi & Badgwell, 2021), en donde, a pesar de que es habitual contrarrestar la acción tumoral de dicha enfermedad aún no se demuestran avances significativos en su eliminación definitiva. En este sentido, el uso de procedimientos altamente invasivos para su diagnóstico y tratamiento se presenta como una de las pocas alternativas ante este tipo de pronósticos, es por ello por lo que la mayoría de las estrategias empleadas se basan en una combinación simultánea de varias terapias como: la cirugía tradicional y laparoscópica, quimioterapia, radioterapia, inmunoterapia e incluso el uso de fármacos. Sin embargo, dichos tratamientos a pesar de que llegan a contrarrestar la acción tumoral del organismo presentan una gran limitación al momento de actuar en zonas o secciones específicas (Joshi & Badgwell, 2021), por lo que resulta imperativa la implementación de nuevas estrategias diagnósticas que faciliten la interacción de los fármacos con determinados receptores de la célula.

Por tal motivo, la implementación de alternativas como la terapia proteica enfocada al diseño de proteínas *de novo* se presenta como una posible estrategia para el tratamiento del cáncer debido a la facilidad para generar secuencias proteicas altamente estables y destinarlas a la interacción o unión con receptores de células cancerígenas específicas. Ante este hecho, se plantea el desarrollo de proteínas *de novo* a través de un enfoque computacional con la ayuda de herramientas de inteligencia artificial para la inhibición del factor NF- $\kappa$ B como medida

introdutoria para la generación de terapias alternativas contra el cáncer gástrico. Se establece un punto de partida en investigaciones relacionadas con la comprensión de estructuras proteicas, inteligencia artificial en la biomedicina y desarrollo de nuevos fármacos para tratar diferentes enfermedades.

El cáncer gástrico probablemente sea una de las enfermedades más antiguas reconocidas por la sociedad humana incluso llegando a presentar varios registros dentro de distintas civilizaciones, sin embargo, no fue considerada como tal hasta el descubrimiento de su fisiopatología (alteraciones en la mucosa gástrica y desarrollo de úlceras duodenales) durante la época del renacimiento, en donde, gran parte de los conocimientos actuales fueron desarrollados.

Actualmente se la considera la neoplasia más recurrente en el estómago y el tercer tipo de cáncer con mayor mortalidad (1 de cada 12 muertes) por debajo del cáncer de pulmón y colorrectal, además, suele ser más frecuente en hombres (1 de cada 54) que en mujeres (1 de cada 126) de regiones rurales de Asia Oriental, Europa Oriental y América Latina, siendo esta última región, donde se presenta la mayor parte de los pacientes jóvenes (López M. J. et al., 2023).

Recientemente, se han desarrollado diversos procedimientos (biomarcadores circulantes, microARN, entre otros) para el diagnóstico temprano del cáncer gástrico, sin embargo, aun presentan una baja sensibilidad y especificidad durante la presentación de sus resultados.

En el presente artículo se estudia la relación que existe entre el factor NF- $\kappa$ B y el desarrollo de cáncer gástrico, además de una posible estrategia para la inhibición de dicho factor través de la inclusión de conceptos clave acerca del desarrollo de proteínas *de novo* por medio métodos computacionales y el enorme potencial de la inteligencia artificial dentro de la industria biomédica.

El Factor Nuclear de las cadenas ligeras kappa de las células B activas (NF- $\kappa$ B), se refiere al factor de transcripción de proteínas que actúa como regulador del sistema inmune innato y que presenta una enorme influencia durante la proinflamación y neuroprotección del organismo. Es responsable de la señalización biológica compleja del cuerpo, resultante de la activación de su vía de señalización (I $\kappa$ B quinasa o IKK) por diversos estímulos: señales de estrés, antígenos inmunes, citoquinas inflamatorias o productos microbianos que permiten al complejo IKK trasladarse hacia el núcleo, en donde, la poli ubiquitinación y fosforilación de NF- $\kappa$ B permite la degradación proteasomal de dicha proteína y la liberación de sus dímeros para su traslocación (Mulero et al., 2019).

NF- $\kappa$ B actúa como regulador central de la respuesta inmune e inflamatoria del organismo, por lo tanto, si existe una desregulación durante la activación de dicho factor se pueden presentar varias alteraciones (mutaciones

## INTRODUCCIÓN

A pesar de los esfuerzos de la comunidad médica, el cáncer se ha convertido en una de las principales causas de muerte a nivel global, provocando alrededor de 10 millones de defunciones en el año 2020 (World Health Organization [WHO], 2022). Entre los diferentes tipos de cáncer, el cáncer gástrico es considerado como el tercer tipo de cáncer con mayor mortalidad en el mundo debido al incremento de su incidencia en poblaciones menores de 40 años de países europeos, asiáticos y latinoamericanos, en donde, la tasa de supervivencia de los pacientes suele ser del 30 % durante los primeros 5 años tras su diagnóstico. Además, su desarrollo presenta una gran influencia de factores genéticos y ambientales que incluyen: el alto consumo de alcohol, el tabaquismo, el alto consumo de sales e infecciones por *Helicobacter pylori* y virus de Epstein-Barr, lo que produce un crecimiento anormal de células tumorales en el revestimiento del estómago, lo que comúnmente no presenta síntomas hasta etapas muy avanzadas de la enfermedad, donde ya puede extenderse hasta órganos subyacentes como el hígado y el páncreas.

Entre las técnicas para su tratamiento se encuentran: la radioterapia y la quimioterapia que son las más comunes durante el desarrollo de ensayos clínicos, sin embargo, este tipo de técnicas se encuentran limitadas por su incapacidad para actuar en secciones específicas (Joshi & Badgwell, 2021), en donde, a pesar de que es habitual contrarrestar la acción tumoral de dicha enfermedad aún no se demuestran avances significativos en su eliminación definitiva. En este sentido, el uso de procedimientos altamente invasivos para su diagnóstico y tratamiento se presenta como una de las pocas alternativas ante este tipo de pronósticos, es por ello por lo que la mayoría de las estrategias empleadas se basan en una combinación simultánea de varias terapias como: la cirugía tradicional y laparoscópica, quimioterapia, radioterapia, inmunoterapia e incluso el uso de fármacos. Sin embargo, dichos tratamientos a pesar de que llegan a contrarrestar la acción tumoral del organismo presentan una gran limitación al momento de actuar en zonas o secciones específicas (Joshi & Badgwell, 2021), por lo que resulta imperativa la implementación de nuevas estrategias diagnósticas que faciliten la interacción de los fármacos con determinados receptores de la célula.

Por tal motivo, la implementación de alternativas como la terapia proteica enfocada al diseño de proteínas *de novo* se presenta como una posible estrategia para el tratamiento del cáncer debido a la facilidad para generar secuencias proteicas altamente estables y destinarlas a la interacción o unión con receptores de células cancerígenas específicas. Ante este hecho, se plantea el desarrollo de proteínas *de novo* a través de un enfoque computacional con la ayuda de herramientas de inteligencia artificial para la inhibición del factor NF- $\kappa$ B como medida

introdutoria para la generación de terapias alternativas contra el cáncer gástrico. Se establece un punto de partida en investigaciones relacionadas con la comprensión de estructuras proteicas, inteligencia artificial en la biomedicina y desarrollo de nuevos fármacos para tratar diferentes enfermedades.

El cáncer gástrico probablemente sea una de las enfermedades más antiguas reconocidas por la sociedad humana incluso llegando a presentar varios registros dentro de distintas civilizaciones, sin embargo, no fue considerada como tal hasta el descubrimiento de su fisiopatología (alteraciones en la mucosa gástrica y desarrollo de úlceras duodenales) durante la época del renacimiento, en donde, gran parte de los conocimientos actuales fueron desarrollados.

Actualmente se la considera la neoplasia más recurrente en el estómago y el tercer tipo de cáncer con mayor mortalidad (1 de cada 12 muertes) por debajo del cáncer de pulmón y colorrectal, además, suele ser más frecuente en hombres (1 de cada 54) que en mujeres (1 de cada 126) de regiones rurales de Asia Oriental, Europa Oriental y América Latina, siendo esta última región, donde se presenta la mayor parte de los pacientes jóvenes (López M. J. et al., 2023).

Recientemente, se han desarrollado diversos procedimientos (biomarcadores circulantes, microARN, entre otros) para el diagnóstico temprano del cáncer gástrico, sin embargo, aun presentan una baja sensibilidad y especificidad durante la presentación de sus resultados.

En el presente artículo se estudia la relación que existe entre el factor NF- $\kappa$ B y el desarrollo de cáncer gástrico, además de una posible estrategia para la inhibición de dicho factor través de la inclusión de conceptos clave acerca del desarrollo de proteínas *de novo* por medio métodos computacionales y el enorme potencial de la inteligencia artificial dentro de la industria biomédica.

El Factor Nuclear de las cadenas ligeras kappa de las células B activas (NF- $\kappa$ B), se refiere al factor de transcripción de proteínas que actúa como regulador del sistema inmune innato y que presenta una enorme influencia durante la proinflamación y neuroprotección del organismo. Es responsable de la señalización biológica compleja del cuerpo, resultante de la activación de su vía de señalización (I $\kappa$ B quinasa o IKK) por diversos estímulos: señales de estrés, antígenos inmunes, citoquinas inflamatorias o productos microbianos que permiten al complejo IKK trasladarse hacia el núcleo, en donde, la poli ubiquitinación y fosforilación de NF- $\kappa$ B permite la degradación proteasomal de dicha proteína y la liberación de sus dímeros para su traslocación (Mulero et al., 2019).

NF- $\kappa$ B actúa como regulador central de la respuesta inmune e inflamatoria del organismo, por lo tanto, si existe una desregulación durante la activación de dicho factor se pueden presentar varias alteraciones (mutaciones

oncogénicas, daños en los tejidos y ADN) que permiten la carcinogénesis gástrica, supresión de la apoptosis, metástasis e inestabilidad genómica que puede incluir cierto tipo de resistencia a medicamentos destinados al tratamiento del cáncer (Chaithongyot et al., 2021).

Cada uno de los miembros de la familia NF- $\kappa$ B (RelA, RelB, NF $\kappa$ B1 y NF $\kappa$ B2) tienen un papel clave durante la supervivencia, proliferación y expresión de distintos tipos de células del sistema nervioso e inmunológico (Medeiros et al., 2021). Por ende, cualquier alteración o desregulación en su activación puede ocasionar daños en los tejidos y ADN de la zona en que se localizan, permitiendo la carcinogénesis, supresión de la apoptosis, metástasis, inestabilidad genómica y algún tipo de resistencia a medicamentos contra este tipo de cáncer (Chaithongyot et al., 2021). Por esta razón, varios estudios establecen nuevas estrategias basadas en la inhibición selectiva de dicho factor por medio de la generación de proteínas inhibitorias que comúnmente suelen estar acompañadas de diversas terapias convencionales para un tratamiento más efectivo (Mirzaei et al., 2021).

Actualmente, la evolución continua de herramientas y métodos computacionales ha permitido diseñar diferentes tipos de proteínas por medio de redes neuronales y modelos de aprendizaje profundo facilitando la construcción de secuencias proteicas y la estabilización de sus conformaciones tridimensionales. Estos nuevos tipos de proteínas son desarrollados a partir de un enfoque de ensamblaje de fragmentos con o sin la ayuda de plantillas, de este modo, cada una de las nuevas estructuras proteicas contarán con características similares a una proteína natural.

Sin embargo, la mayoría de estas proteínas *de novo* no presentan una función biológica, por lo que, suele ser un gran desafío destinar una función específica a sus diversas formas conformacionales e interacciones con los monómeros disponibles (Baker, 2019). Por esta razón, su principal aplicación se enfoca a la comprensión del aspecto y capacidad de plegamiento de las proteínas, logrando predecir estructuras más estables frente a factores de desnaturalización que cuentan con un enorme potencial terapéutico, incluso podríamos hablar de nuevas aplicaciones en un futuro cercano como es: la generación de nanomateriales e interfaces biocompatibles.

La Inteligencia Artificial en Biomedicina describe al conjunto de algoritmos complejos que imita la cognición humana junto a su capacidad de adaptabilidad y aprendizaje para la búsqueda, recopilación e interpretación de información dentro de un entorno computacional. En cambio, dentro del campo médico se refiere al sistema de asistencia complementaria destinado a brindar ayuda en la resolución de problemas en donde, su potencial aplicación radica en la atención médica multidisciplinaria:

- Aprendizaje automático para el diagnóstico de enfermedades.
- Procesamiento de datos para la evaluación de riesgos médicos.
- Procesamiento natural del lenguaje para la predicción y vigilancia de enfermedades

Las técnicas de diseño de fármacos: cribado virtual de estructuras y ligandos, evaluación de parámetros de toxicidad, comprensión de mecanismos moleculares, actividad fisicoquímica y liberación de compuestos farmacológicos han logrado solventar los altos costos que presenta el estudio y predicción de estructuras proteicas *in vitro* e *in vivo* destinadas a mejorar el desarrollo de nuevos medicamentos mediante el refinamiento o modelado computacional de sus estructuras tridimensionales y la optimización e identificación de nuevas dianas terapéuticas para el tratamiento del cáncer. Por lo tanto, este tipo de herramientas representan tanto el presente como el futuro en el diseño de fármacos y predicción de proteínas, por lo que es necesario conocer algunas herramientas de inteligencia artificial tales como: AlphaFold desarrollada por Google's Deep Mind, Deepchem presentada por Google's TensorFlow framework o Rfdifusión elaborada por Rossetta.

## MATERIALES Y MÉTODOS

Para la predicción precisa de proteínas, se buscó la secuencia y estructura tridimensional del factor NF- $\kappa$ B, como base para la construcción de inhibidores que presenten la capacidad de interactuar con este, también se tomaron las diferentes mutaciones con las que se relaciona dentro del cáncer gástrico. Se utilizó diversos repositorios de datos, tales como: PDB (Protein Data Bank), Uniprot y DogSiteScorer para la visualización y descarga de sus secuencias y estructuras proteicas cristalizadas, seleccionando como organismo modelo "*Homo sapiens*".

Para la construcción proteica de doble hélice  $\beta$  se empleó la metodología dispuesta por Marcos et al., (2018), se especificó que la longitud de las hebras mantuviera un rango de 5 a 7 residuos y de 3 a 5 residuos para los bucles de arco  $\beta$ , también se permitió un rango de desplazamiento de 0 a 2 para hebras  $\beta$  de distinto arco junto a un desplazamiento para los arcos  $\beta$  de 0 a 4, mientras que para las hebras pertenecientes a un mismo arco se mantuvieron sin ningún cambio, resultando en la formación de estructuras estrechas (tipo sándwich) para aquellas proteínas con un pequeño desplazamiento y para grandes desplazamientos estructuras anchas (tipo barril); se incluyó enlaces disulfuro entre la primera y última cadena- $\beta$  para disminuir la entropía y aumentar la estabilidad termodinámica del diseño.

Para construcciones que contengan los elementos bucle-hélice-bucle (LHL) su elaboración se llevó a cabo mediante el método descrito por (Pan et al., 2020), que excluye todas las estructuras redundantes y comienza con

un pliegue de proteína que permitió insertar bucles de 5 residuos o menos en los extremos de cada espacio y en las hélices de 2 a 5 residuos de longitud eliminando toda secuencia del bucle que entre en conflicto con el pliegue de la proteína inicial junto al exceso de residuos presente.

Se utilizó “templates” con el fin de obtener una adecuada conformación proteica similar a las estructuras naturales, se utilizó el programa BLAST en conjunto con el repositorio de datos NCBI, estos permiten identificar diferentes “clusters” pertenecientes a bases de datos experimentales que presenten un alineamiento homólogo casi idéntico (mayor al 70%) con las secuencias anteriormente investigadas. Se generó un alineamiento “múltiple” y “pareado” entre los “clusters” y la “query sequence”, utilizando los ficheros que contienen dichas secuencias en el software Jalview, esto permitió seleccionar los fragmentos con un alto índice de conservación junto a los segmentos disponibles para un mejor consenso en la predicción de las proteínas de novo.

Se utilizó herramientas de IA tales como: AlphaFold, Deepchem y RFDiffusion que generaron modelos 3D con una marcada similitud con las proteínas naturales. Para “AlphaFold” se estableció como entrada el nombre del factor propuesto y se seleccionó solo aquellos que muestran un acceso directo a su estructura cristalizada dentro de la base de datos “Uniprot” pertenecientes al organismo *H. sapiens*. Para “Deepchem”, se estableció una sintaxis que permitió establecer la longitud de las secuencias proteicas con la finalidad de modelarla en colaboración de los servidores de Google Colab de “AlphaFold” (Bellamkonda, 2023). Finalmente, “RFDiffusion” ejecuta sus propias líneas de comando disponibles en su servidor de Google Colab, se establece los parámetros de diseño: longitud de la proteína, generación de dímeros, oligómeros o monómeros junto al número de variantes que se quieren diseñar, estableciendo que por cada herramienta se genere 5 a 10 estructuras (Watson & Juergens, 2023).

Se utilizó “PyMOL”, para explorar la composición estructural de una proteína, para lo cual utilizamos los comandos: “File; Run script” con el propósito de visualizar de forma completa la estructura, composición, motivos y dominios estructurales, familias de proteínas, entre otras características que componen dichas moléculas (Mooers & Brown, 2021).

Se analizó la utilidad de estas proteínas en el proceso de interacción proteína-proteína, examinando su conformación y plegamiento a través de técnicas computacionales. Utilizamos herramientas tales como, Chemistry at HARvard Macromolecular Mechanics conocida como CHARMM que permite: leer distintas estructuras proteicas, analizar propiedades de equilibrio dinámicas y estructurales presentes en las proteínas predichas (Brooks et al., 1983). Se calculó energías de referencia, capacidad de plegamiento e interacciones residuo-residuo (Jo et al., 2008).

Para efectuar esta tarea se debe eliminar sitios irrelevantes, para esto se utilizó el programa ChimeraX v1.6.1 (Pettersen et al., 2021), luego se utilizó el servicio Protein Plus para la predicción de sitios de unión apropiados dentro de estas estructuras previamente procesadas, se establecieron parámetros “DrugScore” entre 0.5 a 1 de significancia (Michel et al., 2019). Para el acoplamiento molecular se empleó AutoDock Vina, se estableció las coordenadas de los sitios de unión por medio de mapeos estructurales y se estableció como parámetro 10 predicciones de sitios de unión que fueron identificados mediante la opción “DogSiteScorer”.

Para las variables termodinámicas: Entalpía y Capacidad calorífica se asumió un sistema a presión y temperatura constantes (25 °C a 1 atm) igual para cada una de las interacciones obtenidas, mientras que las variables restantes se obtuvieron a través del proceso de acoplamiento molecular de las proteínas predichas.

Para la evaluación de la selectividad de las proteínas predichas durante la inhibición del factor propuesto, se categorizó a cada una de las interacciones proteína-proteína conforme la cantidad de genes o factores NF-κB con los que han logrado interactuar, aquellas estructuras con la menor cantidad de interacciones serán consideradas como “selectivas”. Para considerar la afinidad de una proteína se tomó en cuenta la “energía libre de Gibbs”, donde una  $\Delta G$  más negativa sugiere una interacción más fuerte y, por lo tanto, una mayor afinidad.

## RESULTADOS Y DISCUSIÓN

Es una prioridad el desarrollo de tratamientos con la capacidad de suprimir la expresión del factor NF-κB. Cada uno de los modelos contemplados en la Tabla 1 son el resultado de redes de aprendizaje altamente entrenadas para la generación de nuevas estructuras proteicas sin algún tipo de restricción o limitación en su plegamiento, por tal motivo, la verificación de su confiabilidad y efectividad se presentan como el punto de partida para la investigación de nuevas estrategias especializadas para este contexto.

Tabla 1. Proteínas generadas por las herramientas de Inteligencia Artificial en base a inhibidores del factor NF- κB.

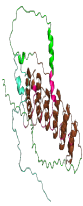
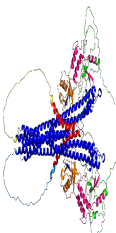
N.º	Códigos para la identificación de las proteínas predichas		
	AlphaFold	Deepchem	RFdiffusion
1	O15226	A7886	RF21A
2	P25963	EC19F	RF415AD
3	Q15653	963D2	RF715A1
4	O15111	92B1D	RF315AD
5	O14920	A26C8	RF215A2
6	Q8NI38	208E3	RF315BD
7	O00221	10FAD	RF51B
8	Q9BYH8	9061D	RF2M31
9	Q9Y6K9	533E3	RF3M31
10	Q9UBC1	07795	RF2M32


En total, se predijeron 30 proteínas de novo a partir de datos disponibles en repositorios (AlphaFold), conjuntos de datos experimentales (RFdiffusion) y estándar (Deepchem).

El progreso y evolución constante de las herramientas de IA han permitido determinar las mejores métricas, composiciones estructurales y funciones biológicas con la única finalidad de que respondan adecuadamente a las pruebas experimentales a las que serán sometidas. La predicción exitosa de la estructura T1008 ha demostrado que es posible generar modelos homólogos a estructuras PDB con un alto grado de precisión, pero con una clara diferencia entre los rendimientos obtenidos por una proteína de novo y una de origen natural.

A continuación, se puede visualizar un modelo basado en una amplia variedad de inhibidores y reguladores naturales de factor NF- $\kappa$ B, siendo el inhibidor NFKBIA uno de los “templates” más empleados en la construcción y predicción de estas estructuras, debido a que puede llegar a disminuir la inflamación causada por los complejos NF- $\kappa$ B/REL (Wang et al., 2024).

Tabla 2. Composición y Estructura de las proteínas predichas en herramientas de IA.

IA	Código	Composición de la proteína	Figuras
AlphaFold	Q15653	<p><b>Regiones</b></p> <ul style="list-style-type: none"> <li>Regiones desordenadas: Verde (149–193, 298–356)</li> </ul> <p><b>Dominios</b></p> <ul style="list-style-type: none"> <li>Dominio que contiene repeticiones Ankyrin: Rosado (56-302)</li> </ul> <p><b>Repeticiones</b></p> <ul style="list-style-type: none"> <li>ANK 1: Café (57-86)</li> <li>ANK 2: Café (93-122)</li> <li>ANK 3: Café (126-155)</li> <li>ANK 4: Café (206-235)</li> <li>ANK 5: Café (240-269)</li> <li>ANK 6: Café (273-302)</li> </ul>	
	A26C8	<p><b>Regiones</b></p> <ul style="list-style-type: none"> <li>Regiones artificiales: Verde (1-60, 565-625)</li> </ul> <p><b>Dominios</b></p> <ul style="list-style-type: none"> <li>Dominio catalítico de la serina/treonina quinasa: Rosado (59-139)</li> <li>Dominio similar a la ubiquitina (Ubl): Naranja (140-220)</li> <li>Dominio de dimerización del andamio IQBAL: Azul (219- 493)</li> <li>Dominio de unión a I-kappa-quinasa-beta NEMO: Rojo (536-565)</li> </ul>	

<b>RFdiffusion</b>	RF715A1	<p><b>Regiones</b></p> <ul style="list-style-type: none"> <li>Regiones artificiales: Verde (1-5, 305-359)</li> </ul> <p><b>Dominios</b></p> <ul style="list-style-type: none"> <li>Dominio catalítico de la serina/treonina quinasa: Rosado (13-242)</li> <li>Dominio similar a la ubiquitina (Ubl): Naranja (244-302)</li> </ul> <p><b>Familias</b></p> <ul style="list-style-type: none"> <li>Familia de función desconocida (DUF6088): Morado (255-313)</li> </ul>	
--------------------	---------	---	---

En la tabla 2, se observan las estructuras de las proteínas generadas por tres herramientas de IA, Q15653, A26C8 y RF715A1 para cada uno de los modelos respectivamente. Estas presentan una conformación Bucle-Hélice-Bucle (LHL), donde, la variedad en sus estructuras tridimensionales se ve determinada por la herramienta de IA utilizada, mientras que su composición está definida por el conjunto de datos destinado a dicho modelos. La mayoría de las proteínas predichas presentan composiciones similares entre sí, las que coinciden con datos cristalográficos de tipo estándar y experimental (E. H. Li et al., 2023). Los dominios y regiones presentados en las tres gráficas son los que mejor representan la composición general de las estructuras, ejemplo el dominio que contiene repeticiones de ankyrin, que describe a una de las familias de proteínas más comunes que puede encontrarse en todas las formas de vida, y tienen un papel clave en las interacciones proteína-proteína de iniciadores transcripcionales, transductores de señales, reguladores de desarrollo y toxinas.

Los datos relacionados con el estado de protonación y los residuos terminales se presentan en la Tabla 3, estos son relevantes en el análisis de la estabilidad de las proteínas de novo debido a que su presencia o ausencia durante el ensamblaje de la estructura tridimensional está ligado a su capacidad de plegamiento y retención. Se describe

los residuos N-terminal y C-terminal como regiones comunes de varias proteínas de unión al ADN con funciones reguladoras, donde, N-terminal presenta cierta flexibilidad conformacional y capacidad de unión con moléculas de señalización, mientras C-terminal al estar conformada por residuos hidrófobos, estabiliza la estructura y aumenta su afinidad con otras proteínas.

El estado de protonación de los residuos presentes en las proteínas predichas permite experimentar cierto grado de inestabilidad durante su plegamiento, incluso, los residuos que se encuentran asociados con dicho efecto tienden a separarse de su estructura principal con mayor frecuencia cuando se encuentran con dos o más de estos estados, sin embargo, este estado no se puede tomar como prueba irrefutable o concluyente de la estabilidad de dichas estructuras.

Las energías computadas se presentan como una medida de mayor confiabilidad para el análisis de estabilidad, debido a que representan aquellas interacciones intra e intermoleculares que mantienen unidas a este tipo de estructuras complejas, en estas conformaciones, las interacciones de enlace de hidrogeno asociadas con la energía de unión entre sus enlaces covalentes y de Van der Waals relacionadas con energías más débiles para mantener la estabilidad, tienen un papel clave.

Tabla 3. Datos y propiedades energéticas de las proteínas predichas por las herramientas de IA.

IA	Código	Datos básicos de la proteína						Energía Computada				
		Chain	Long	pH	Residuos	Protonación	Ion	Enlace	CMAP	VDWaals	Angular	
AF	Q15653	A	356	7,0	NTER CTER	-	NaCl	544,93	-13,17	-727,46	1147,65	
DC	A26C8	A B	625	7,0	NTER CTER	A (384ARG- RN1; 561ASP- ASPP; 5ASP- ASPP)	NaCl	4574,76	229,09	0,18x1014	7053,08	
RF	RF715A1	A	433	7,0	GLYP CTER	A (229LYS- LSN; 30ASP- ASPP)	NaCl	3,10x105	-33,67	7,24x106	1452,76	

En la tabla 3, se puede observar el estado de protonación expresado en la mayoría de los modelos diméricos predichos, lo que representa la presencia o ausencia de protones (H+) en cada uno de los residuos, además, se incluye la forma que adopta tras dicho proceso. Por lo tanto, se puede concluir, que estas herramientas de IA aún no se encuentran familiarizadas con el desarrollo de conformaciones más complejas en las proteínas.

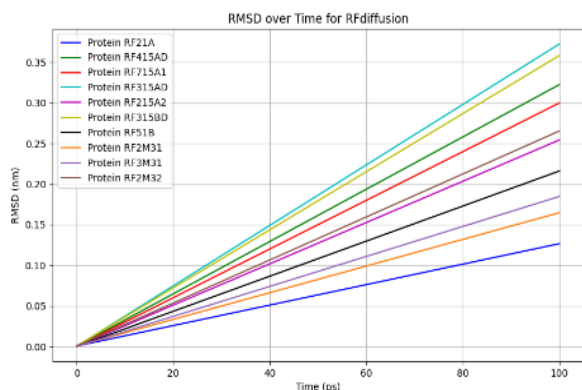
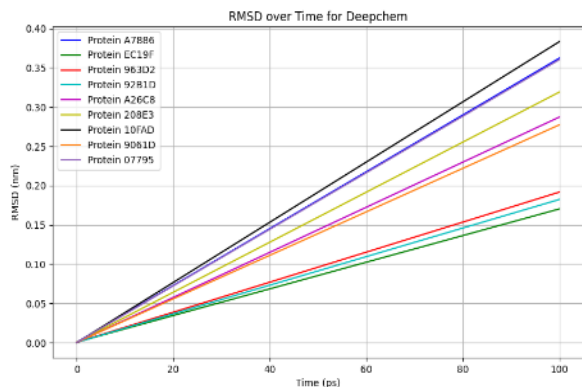
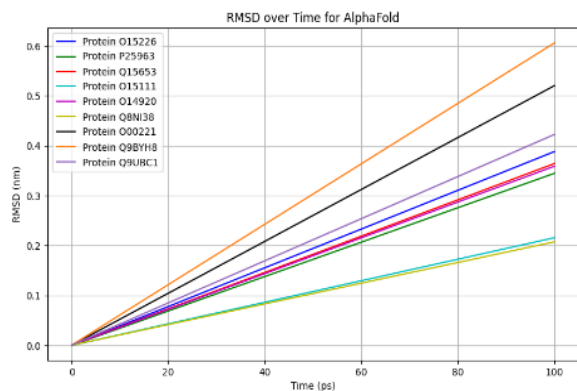


Figura 1. Gráficas RMSD de estabilidad de las proteínas predichas.

RMSD (Raíz de la desviación cuadrática media) muestra el cambio producido dentro de la conformación estructural de las proteínas predichas en un tiempo (100 picosegundos debido a la capacidad computacional usada en el

proceso). Los valores de RMSD visualizados en la Figura 1 muestran la estabilidad de las proteínas predichas, este es considerado como un indicador útil del estado de la estructura durante su plegamiento en las simulaciones realizadas. Por esta razón, varios estudios consideran al estado estable de una proteína dentro de un rango de fluctuación de entre 0,2 a 1,0 nm tras llegar a su equilibrio a los 100 o 200 picosegundos hasta finalizar entre 150 a 200 nanosegundos de dinámica molecular (Arshia et al., 2021; Maruyama et al., 2023), por tanto, se puede concluir que todas las proteínas predichas con fluctuación no mayor a 0,6 nm son estables durante los primeros 100 picosegundos.

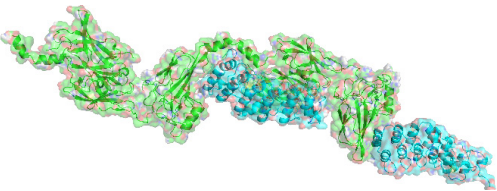
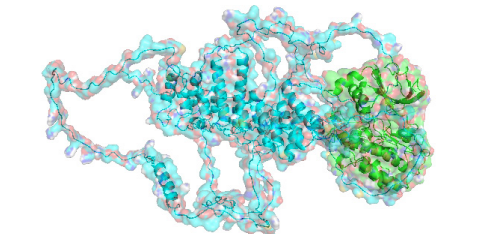
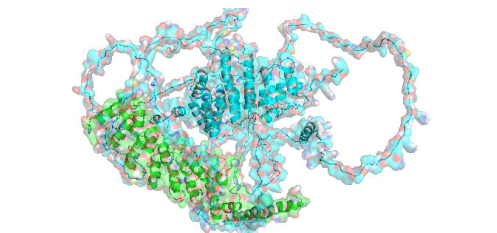
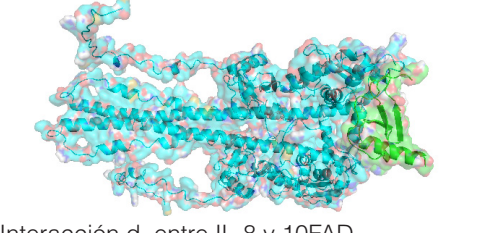
Tabla 4. Variables termodinámicas obtenidas en el Acoplamiento Molecular.

Docking y Acoplamiento Molecular		
Receptor	Mejor ligando	$\Delta G$ (kcal/mol)
Heterodímero NF- $\kappa$ B p50/p65 (Control)	Inhibidor I $\kappa$ B $\alpha$ (Control)	-15,11
Quinasa inductora de NF- $\kappa$ B (NIK)	Q9BYH8	-15,59
Subunidad p100 (NF $\kappa$ B2)	Q9BYH8	-18,07
Interleucina-8	10FAD	-11,93

El control pertenece al complejo I $\kappa$ B $\alpha$ /NF- $\kappa$ B, que a través de su acoplamiento e interacción específica produce una inhibición efectiva de la actividad del factor NF- $\kappa$ B, este es un punto de referencia para nuevos prospectos a inhibidores. Por tanto, todo acoplamiento con valores mayores a -15,11 kcal/mol de Energía libre de Gibbs ( $\Delta G$ ) y 8,4 pM de Constante de disociación (Kd) representan interacciones proteína-proteínas menos favorables o viceversa.

Estas variables y sus valores se encuentran asociados con la estabilidad y plegado de los acoplamientos, el valor de -30 kcal/mol obtenido en la Entalpía ( $\Delta H$ ) junto al valor de -119,5 cal/mol\*K de Capacidad calorífica (Cp) reflejan en cada uno de complejos proteicos un cierto grado de termo-estabilidad (25 °C) durante plegamiento de su estructura. La Entropía ( $\Delta S$ ) representa el desorden y la espontaneidad, valores mayores a 50 cal/mol\*K representarán un mayor desorden en su estructura y viceversa, en consecuencia, la mayoría de los acoplamientos evaluados tienen una buena predisposición para el acoplamiento en función del control evaluado.

Tabla 5. Información de afinidad de las proteínas predichas con sus ligandos.

Receptor	Proteínas predichas (ligandos)	Afinidad (kcal/mol)	Acoplamiento
Heterodímero NF- $\kappa$ B p50/p65 (Control)	Inhibidor I $\kappa$ B $\alpha$ (Control)	-15,11	 Interacción a, entre las proteínas control.
Quinasa inductora de NF- $\kappa$ B (NIK)	Q9BYH8	-15,59	 Interacción b, entre NIK y Q9BYH8.
Subunidad p100 (NF $\kappa$ B2)	Q9BYH8	-18,07	 Interacción c, entre la subunidad p100 y Q9BYH8.
Interleucina-8	10FAD	-11,93	 Interacción d, entre IL-8 y 10FAD.

De acuerdo con los mencionado por Siebenmorgen & Zacharias (2020), se sabe que tanto la afinidad de unión como la energía libre de unión ( $\Delta G$ ) se encuentran asociadas a las interacciones proteína-proteína y por ende sus valores llegan a ser iguales durante una evaluación de afinidades. En consecuencia, la estructura del complejo I $\kappa$ B $\alpha$ /NF- $\kappa$ B interacción a, junto a su energía libre de unión ( $\Delta G$ ) serán nuevamente el punto de referencia para los demás complejos obtenidos, dando como resultado que la proteína Q9BYH8 sea altamente afín en complejos como NIK y NF $\kappa$ B2(p100) observados tanto en la interacción b y c, mientras que en complejos como el de la interacción d, se puede observar que existe menor afinidad.

## CONCLUSIONES

A partir de las herramientas de IA AlphaFold, Deepchem y RFDiffusion se predijeron 30 proteínas de novo con conformación Bucle-Hélice-Bucle (LHL) y una composición similar a inhibidores y reguladores del factor NF- $\kappa$ B

presente en varios conjuntos de bases de datos estándar y experimental. Se encontraron principalmente: dominios de repeticiones de Ankyrin, dominios de dimerización, dominios similares a ubiquitina (Ubl) y dominios catalíticos serina/treonina quinasa junto a regiones artificiales o desordenadas que permiten mejorar sus interacciones proteína-proteína.

Utilizando técnicas de acoplamiento y docking molecular con las proteínas predichas y los genes regulados por el factor NF- $\kappa$ B presentes en el cáncer gástrico, se obtuvieron valores correspondientes a las variables termodinámicas: Energía libre de Gibbs ( $\Delta G$ ), Constante de disociación (Kd), Entalpía ( $\Delta H$ ), Capacidad calorífica (Cp) y Entropía ( $\Delta S$ ), las cuales comparadas con los valores impuestos por el complejo I $\kappa$ B $\alpha$ /NF- $\kappa$ B (control), permitieron concluir que la mayoría de complejos predichos tienen una mejor interacción proteína-proteína y plegamiento en comparación con el control sugerido.

En las interacciones proteínas-proteína y el complejo I $\kappa$ B $\alpha$ /NF- $\kappa$ B (control), se evaluaron los valores de afinidad mediante el análisis de la Energía libre de Gibbs ( $\Delta G$ ) y la selectividad mostrada por las proteínas de novo durante el acoplamiento molecular. Se encontró que los modelos predichos por AlphaFold son más afines y que los modelos predichos por Deepchem son más selectivos en sus interacciones con lgenes y factores de transcripción regulados por NF- $\kappa$ B.

## REFERENCIAS BIBLIOGRÁFICAS

- Arshia, A. H., Shadravan, S., Solhjoo, A., Sakhteman, A., & Sami, A. (2021). De novo design of novel protease inhibitor candidates in the treatment of SARS-CoV-2 using deep learning, docking, and molecular dynamic simulations. *Computers in Biology and Medicine*, 139(104967). <https://doi.org/10.1016/j.compbiomed.2021.104967>
- Baker, D. (2019). What has de novo protein design taught us about protein folding and biophysics? *Protein Science*, 28(4). <https://doi.org/10.1002/pro.3588>
- Bellamkonda, S. V. (2023). Applications of DeepChem with AlphaFold: Docking and protein-ligand interaction from protein sequence. DeepChemXAlphaFold. <https://deepchem.io/tutorials/deepchemxalphafold/>
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2), 187–217. <https://doi.org/10.1002/jcc.540040211>
- Chaithongyot, S., Jantaree, P., Sokolova, O., & Naumann, M. (2021). NF- $\kappa$ B in gastric cancer development and therapy. *Biomedicines*, 9(8). <https://doi.org/10.3390/biomedicines9080870>
- Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11), 1859–1865. <https://doi.org/10.1002/jcc.20945>
- Joshi, S. S., & Badgwell, B. D. (2021). Current treatment and recent progress in gastric cancer. *CA: A Cancer Journal for Clinicians*, 71(3). <https://doi.org/10.3322/caac.21657>
- López, M. J., Carbajal, J., Alfaro, A. L., Saravia, L. G., Zanabria, D., Araujo, J. M., Quispe, L., Zevallos, A., Buleje, J. L., Cho, C. E., Sarmiento, M., Pinto, J. A., & Fajardo, W. (2023). Characteristics of gastric cancer around the world. In *Critical Reviews in Oncology/Hematology*, 181. <https://doi.org/10.1016/j.critrevonc.2022.103841>
- Marcos, E., Chidyausiku, T. M., McShan, A. C., Evangelidis, T., Nerli, S., Carter, L., Nivón, L. G., Davis, A., Oberdorfer, G., Tripsianes, K., Sgourakis, N. G., & Baker, D. (2018). De novo design of a non-local  $\beta$ -sheet protein with high stability and accuracy. *Nature Structural and Molecular Biology*, 25(11), 1028–1034. <https://doi.org/10.1038/s41594-018-0141-6>
- Maruyama, Y., Igarashi, R., Ushiku, Y., & Mitsutake, A. (2023). Analysis of Protein Folding Simulation with Moving Root Mean Square Deviation. *Journal of Chemical Information and Modeling*, 63(5), 1529–1541. <https://doi.org/10.1021/acs.jcim.2c01444>
- Medeiros, M., Candido, M. F., Valera, E. T., & Brassesco, M. S. (2021). The multifaceted NF- $\kappa$ B: are there still prospects of its inhibition for clinical intervention in pediatric central nervous system tumors? In *Cellular and Molecular Life Sciences*, 78(17–18). <https://doi.org/10.1007/s00018-021-03906-7>
- Michel, M., Visnes, T., Homan, E. J., Seashore-Ludlow, B., Hedenström, M., Wiita, E., Vallin, K., Paulin, C. B. J., Zhang, J., Wallner, O., Scobie, M., Schmidt, A., Jenmalm-Jensen, A., Warpman Berglund, U., & Helleday, T. (2019). Computational and Experimental Druggability Assessment of Human DNA Glycosylases. *ACS Omega*, 4(7), 11642–11656. <https://doi.org/10.1021/acsomega.9b00162>
- Mirzaei, S., Zarrabi, A., Hashemi, F., Zabolian, A., Saleki, H., Ranjbar, A., Seyed Saleh, S. H., Bagherian, M., Sharifzadeh, S.omid, Hushmandi, K., Liskova, A., Kubatka, P., Makvandi, P., Tergaonkar, V., Kumar, A. P., Ashrafzadeh, M., & Sethi, G. (2021). Regulation of Nuclear Factor-KappaB (NF- $\kappa$ B) signaling pathway by non-coding RNAs in cancer: Inhibiting or promoting carcinogenesis? *Cancer Letters*, 509. <https://doi.org/10.1016/j.canlet.2021.03.025>
- Mooers, B. H., & Brown, M. E. (2021). Templates for writing PyMOL scripts. *Protein Science*, 30(1), 262–269. <https://doi.org/10.1002/pro.3997>
- Mulero, M. C., Wang, V. Y. F., Huxford, T., & Ghosh, G. (2019). Genome reading by the NF- $\kappa$ B transcription factors. *Nucleic Acids Research*, 47(19), 9967–9989. <https://doi.org/10.1093/NAR/GKZ739>
- Pan, X., Thompson, M. C., Zhang, Y., Liu, L., Fraser, J. S., Kelly, M. J. S., & Kortemme, T. (2020). Expanding the space of protein geometries by computational design of de novo fold families. *Science*, 369(6507), 1132–1136. <https://doi.org/10.1126/science.abc0881>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70–82. <https://doi.org/10.1002/pro.3943>
- Siebenmorgen, T., & Zacharias, M. (2020). Computational prediction of protein–protein binding affinities. In *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(3), e1448. <https://doi.org/10.1002/wcms.1448>
- Watson, J. L., & Juergens, D. (2023, October 3). RosettaCommons/RFdiffusion: Code for running RFdiffusion. GitHub. <https://github.com/RosettaCommons/RFdiffusion>
- World Health Organization. (2022). Cancer. World Health Organization. <https://www.who.int/es/news-room/factsheets/detail/cancer>