



Facultad de ingeniería
Escuela de Ingeniería Civil en Bioinformática

Implementación de una estrategia basada en métodos teórico-computacionales para la búsqueda y diseño de nuevos inhibidores en la enzima ureasa de *Helicobacter pylori*.

Memoria para optar al grado de Ingeniero Civil en Bioinformática y Licenciatura en Ciencias de la Ingeniería.

Alumna: Paulina V. Valenzuela Hormazábal.
Tutor: Dr. Daniel M. Bustos Guajardo.
Co-tutor: Dr. Erix W. Hernández Rodríguez.
Profesor informante: Dr. Jans Alzate-Morales.

Talca – Chile.

2021

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Talca, 2022

AGRADECIMIENTOS

Está culminando esta larga pero bella etapa de mi vida que veía muy lejana, ha sido difícil pero con mucho esfuerzo, estudio y sobretodo el apoyo de muchas personas lo he logrado. Quiero comenzar agradeciendo a mi profesor tutor Dr. Daniel Bustos ya que gracias a su ayuda y paciencia mi trabajo de tesis se logró con mayor facilidad. Sus consejos siempre útiles cuando las ideas no querían fluir en mi mente, además creyó en mí y formó parte importante durante el último año en mi vida profesional con sus múltiples consejos, enseñanzas y ayuda sin importar la hora.

También quiero agradecer a mi profesor co-tutor, el Dr. Erix Hernández, ya que al igual que el profesor Daniel siempre estuvo ahí presente y pendiente en que me podía ayudar y orientar en mi proceso de tesis. Agradecer también a los profesores en particular a los que forman parte del minor de Bioinformática Estructural que gracias a ellos, hoy estoy aquí agradezco inmensamente los conocimientos rigurosos y precisos, mis conocimientos son gracias a ustedes. Profesor Jans Alzate-Morales estoy muy agradecida por darme la posibilidad de ser su ayudante durante varios semestres y poder aprender cada vez más del bonito mundo de la bioinformática estructural y querer buscar e investigar más allá de los conocimientos entregados.

A mi familia (padres y hermana) ustedes son mi motor que me impulsaron a no bajar los brazos y no rendirme nunca, que luchara por mis sueños, son quienes han estado siempre a mi lado en los días y noches más difíciles siempre apoyándome en todo momento, soportando mi mal humor a final de cada semestre y mis llantos cuando creía que ya no podía más. Padres ahora que está terminando este proceso les quiero dedicar este logro a ustedes como una meta más lograda. Estoy orgullosa de haberlos elegido como mis padres y que estén a mi lado en este momento tan importante. Gracias por creer siempre en mí.

Además quiero dedicar este logro a mi abuela Hilda que estuvo durante todos estos años alentándome a no bajar los brazos y siempre empoderándome con sus consejos. Hoy no se encuentra con nosotros, pero sé que desde el cielo me está acompañando y espero que se encuentre orgullosa que su nieta al fin terminó su carrera y se dedicará a lo que tanto le gusta y pero ella nunca entendió ni una palabra de lo que yo decía pero aun así me escuchaba y tomaba atención como si entendiera todo.

A mi pareja, Claudio me ha acompañado prácticamente desde comienzo en este camino universitario, siendo mi compañero de carrera y pololo siempre ha estado presente apoyándome, ayudando a no bajar los brazos y que si me propongo una meta debo cumplirla y ser la mejor. Días y noches estudiando juntos y enseñándome todos los *tips* de programación para que me fuera bien y aprobara esos cursos tanto odiaba pero que tanto me han servido. Y no puedo dejar de mencionar que siempre estaba listo para un abrazo después de una prueba independiente de cómo me hubiera ido. Gracias por el apoyo y amor incondicional.

Y por último, y no menos importante, mis amigos y amigas que son poquitos pero son increíbles, los mejores compañeros de viaje, que si bien está terminando esta aventura y soy la última en terminar este proceso, ha sido maravilloso junto a ustedes, y no puedo dejar de recordar cuantas mañanas, tardes, noches, almuerzos y horas de trabajo, nos juntamos para estudiar, hablar de la vida de nuestras proyecciones y del futuro y siempre apoyándonos en todo. Hoy toca cerrar un capítulo hermoso en nuestras vidas y no puedo dejar de agradecer por todo el apoyo, abrazos, cariños y horas de estudio. Gracias por estar siempre ahí, los quiero.

RESUMEN

El metabolismo del nitrógeno es fundamental para un adecuado estado de salud en los seres vivos. Además, el nitrógeno de desecho residual producido por el catabolismo de los aminoácidos es excretado como urea por el riñón. Esta molécula de desecho en los mamíferos es clave para el metabolismo en plantas, hongos y bacterias, ya que estos hidrolizan la urea en amonio para conseguir trazas de nitrógeno esenciales para su desarrollo. Para estas formas de vida, la enzima ureasa dependiente de níquel permite hidrolizar la urea y formar amonio.

Por otro lado, la urea es utilizada como fertilizante en la agricultura, debido a su rico contenido de nitrógeno. Las enzimas ureasa de diversos microorganismos distribuidos en el suelo hidrolizan los fertilizantes a base de urea, generando amoníaco que es volatilizado al medio ambiente y finalmente perdido. Esta pérdida de nitrógeno genera una reducción en los rendimientos de cultivo afectando a la economía en la agricultura. Además, la liberación de amoníaco provoca graves consecuencias medioambientales como lluvia ácida y gases de efecto invernadero. La combinación de inhibidores de ureasa (IU) con fertilizantes a base de urea ha surgido como una opción viable para disminuir la pérdida de nitrógeno, al inhibir la actividad catalítica de las enzimas ureolíticas distribuidas en el suelo.

Desde otro punto de vista, actualmente se estima que el 50% de la población mundial está infectada por *Helicobacter pylori* (*H. pylori*). En el caso de Chile, la tasa de infección es aún mayor, llegando alrededor de 70%. *H. pylori* es la bacteria ureolítica más patogénica para el ser humano, ya que tiene una rápida adaptación para sobrevivir en diferentes ambientes, incluso pudiendo subsistir en ambientes ácidos como el estómago. Una vez que esta bacteria ha colonizado el cuerpo, puede provocar numerosas patologías, entre ellas cáncer. El actual tratamiento consiste en antibióticos de amplio espectro. Sin embargo, *H. pylori* ha generado resistencia a estos fármacos de forma progresiva, por lo que la eficacia de la terapia está disminuyendo drásticamente. Además, este tipo de tratamiento genera efectos secundarios. De este modo, han surgido nuevas estrategias terapéuticas que consisten en el uso de IU. Numerosos esfuerzos están centrados en la búsqueda de nuevos inhibidores de ureasa de *H. pylori*, aquí denominada HpU, que sean potentes y selectivos a la vez, para que se pueda establecer una terapia eficaz para erradicar las infecciones por esta bacteria.

Esta tesis se centró en implementar un protocolo basado en el método de Cribado Virtual, específicamente en dos sus variantes: Cribado Virtual Basado en Estructura y Cribado Virtual Basado en Farmacóforo. Estas técnicas permitieron realizar una búsqueda en la base de datos de compuestos ZINC15 que posee una alta dimensionalidad. La metodología consistió en aplicar filtros sucesivos, entre los que se encuentran: generar un filtrado a la base de datos mediante reglas farmacocinéticas y fisicoquímicas, crear hipótesis de farmacóforos basados en núcleos de cumarinas, realizar estudios de acoplamiento molecular mediante diferentes algoritmos y por último, realizar cálculos de energía libre de unión a los complejos proteína-ligando, con el objetivo de seleccionar un número limitado de compuestos candidatos que tengan una mejor afinidad de unión a la enzima HpU.

ABSTRACT

The nitrogen metabolism is fundamental for an adequate state of health in living beings. Also, the nitrogen residual waste produced by the catabolism of amino acids is excreted as urea by the kidney. This waste molecule in mammals is key to metabolism in plants, fungi and bacteria, as they hydrolyze urea into ammonium to obtain trace amounts of nitrogen that is essential for their development. For these life forms, the nickel-dependent enzyme urease allows urea to be hydrolyzed to form ammonium.

On the other hand, urea is used as a fertilizer in agriculture, because of its rich nitrogen content. Urease enzymes from various microorganisms distributed in the soil hydrolyze urea-based fertilizers, generating ammonia that is volatilized into the environment and finally lost. This loss of nitrogen generates a reduction in crop yields, affecting the economics of agriculture. In addition, the release of ammonia causes serious environmental consequences such as acid rain and greenhouse gases. The combination of urease inhibitors (UI) with urea-based fertilizers has emerged as a viable option to reduce nitrogen loss by inhibiting the catalytic activity of ureolytic enzymes distributed in the soil.

From another point of view, it is currently estimated that 50% of the world's population is infected by *Helicobacter pylori* (*H. pylori*). In the case of Chile, the infection rate is even higher, reaching around 70%. *H. pylori* is the most pathogenic ureolytic bacterium for humans, as it has a rapid adaptation to survive in different environments, even being able to survive in acidic ones such as the stomach. Once this bacterium has colonized the body, it can cause numerous pathologies, including cancer. Current treatment consists of broad-spectrum antibiotics. However, *H. pylori* has generated resistance to these drugs, so the efficacy of the therapy is drastically decreasing. In addition, this type of treatment generates side effects. In this way, new therapeutic strategies have emerged that consist in the use of UI. Numerous efforts are focused on the search for new *H. pylori* urease inhibitors, here called HpU, that are potent and selective, so that an effective therapy can be established to eradicate infections by this bacterium.

This thesis was focused in the implementation of a protocol based on the Virtual Screening method. Specifically in two of its variants: Structure-Based Virtual Screening and Pharmacophore-Based Virtual Screening. These techniques allowed

a search to be carried out in the ZINC15 compound database, which has a high dimensionality. The methodology consisted of applying successive filters, generating a filtering to the database using pharmacokinetic and physicochemical rules, creating pharmacophore hypotheses based on coumarin nuclei, performing molecular docking studies using different algorithms and finally, performing calculations of free energy of binding to protein-ligand complexes, with the objective of selecting a limited number of candidate compounds that have a better binding affinity to the HpU enzyme.

ÍNDICE

MARCO TEÓRICO.....	13
Aspectos generales:.....	14
Problemáticas ambientales producidas por ureasas:.....	15
Problemáticas clínicas producidas por ureasas:	16
Comparación de ureasas de diferentes organismos:	18
Inhibidores de ureasa reportados en la literatura:	20
Fundamentos teóricos:.....	21
HIPÓTESIS	24
OBJETIVOS	24
Objetivo general:	24
Objetivos específicos:	24
METODOLOGÍA.....	25
1) Antecedentes preliminares.	25
a) Revisión bibliográfica:.....	25
b) Preparación de inhibidores ya reportados en la literatura:.....	25
c) Preparación de la proteína:.....	26
d) Acoplamiento molecular:	27
2) Metodología del Trabajo de tesis de grado.....	29
1) Preparación de la base de datos de compuestos para ser utilizada como biblioteca virtual:	30
2) Filtrado a la base de datos a través de diferentes reglas que incluyen términos farmacocinéticos y fisicoquímicos:	30
3) Creación de hipótesis de farmacóforos para el conjunto de compuestos con núcleos de cumarinas.	32
4) Realización de estudios de acoplamiento molecular a los compuestos candidatos, mediante diferentes algoritmos, para reducir el número de candidatos a inhibidores de ureasa:	33
5) Realización de cálculos de energía libre de unión mediante el método MM-GBSA:.....	35
RESULTADOS	39

1) Preparación de la base de datos de compuestos para ser utilizada como biblioteca virtual:	39
2) Filtrado a la base de datos a través de diferentes reglas que incluyen términos farmacocinéticos y fisicoquímicos:	40
3) Hipótesis de farmacóforos para el conjunto de compuestos con núcleos de cumarinas y Cribado Virtual basado en farmacóforos.....	43
4) Estudios de acoplamiento molecular a los compuestos candidatos, mediante diferentes algoritmos para reducir el número de candidatos a inhibidores de ureasa:	47
a) Docking molecular para la variante LYN	48
b) Docking molecular para la variante KCX.....	52
5) Cálculos de energía libre de unión mediante el método MM-GBSA:.....	58
CONCLUSIÓN	73
REFERENCIAS.....	74
ANEXOS	82

ÍNDICE DE FIGURAS

Figura 1: Representación de la hidrólisis de urea por ureasas. Figura adaptada de [28].	17
Figura 2: Ensamblaje de la estructuras cuaternarias en ureasas de diferentes organismos. Una unidad funcional puede estar formada por a) una sola unidad (como en <i>Canavalia ensiformis</i> , PDB ID: 3LA4), b) un heterodímero (como <i>H. pylori</i> , PDB ID: 1E9Z) o c) un heterotrímero (como <i>Sporosarcina pasteurii</i> , PDB ID: 2UBP). Estas unidades funcionales (o monómeros) forman complejos como d) hexámeros, e) dodecámeros o f) trímeros. Figura adaptada [4].	19
Figura 3: Comparación de JbU y otras ureasas bacterianas. Comparación esquemática de una sola subunidad estructural de JbU con dos (<i>H. pylori</i> , HpU) y tres (<i>K. aerogenes</i> (KAU), <i>B. pasteurii</i> (BPU)) subunidades en ureasas bacterianas. Con referencia a JbU, el porcentaje de identidad de secuencia de las regiones correspondientes en ureasas bacterianas se indica encima del recuadro. Figura adaptada de [35].	20
Figura 4: Compuestos con actividad anti-ureasa.	21
Figura 5: Enfoque de Cribado Virtual empleado en la investigación de candidatos a fármacos. Adaptado de [69].	23
Figura 6: Estructuras tridimensional de HpU. (a) dodecámero, (b) unidad funcionalmente activa comprendiendo la subunidad β o accesoria, subunidad α o catalítica, dos iones de níquel y el inhibidor co-cristalizado DJM (2- $\{1-(3,5\text{-dimethylphenyl})-1\text{H-imidazol-2-yl}\}$ sulfanyl}-N-hydroxyacetamide).	27
Figura 7: Núcleo de estructuras 2D	28
Figura 8: Flujo de trabajo. Protocolo de Virtual Screening para la búsqueda de nuevos y potentes inhibidores de ureasa.	29
Figura 9: Grilla utilizada en el método de acoplamiento molecular. Grilla cúbica externa de 36 Å x 36 Å x 36 Å, cubriendo la subunidad catalítica (en color morado). El flap-loop que actúa como una compuerta al sitio catalítico se encuentra representada en color verde, en color naranja se muestran los dos iones de níquel. La superficie de color gris muestra los residuos del sitio catalítico.	34
Figura 10: Reglas de los 5 de Lipinski. Se observa la distribución de moléculas por cada característica asociada a la regla, en línea punteada se encuentra ubicado el cutoff que se aplicó para el número de moléculas seleccionadas.	40
Figura 11: Reglas de los 3 de Jorgensen Se observa la distribución de moléculas por cada característica asociada a la regla, en línea punteada se encuentra ubicado el cutoff que se aplicó para el número de moléculas seleccionadas.	41
Figura 12: Regla de Veber. Se observa la distribución de moléculas por cada	

característica asociada a la regla, en línea punteada se encuentra ubicada el cutoff que se aplicó para el número de moléculas seleccionadas.	42
Figura 13: Flujo de trabajo. Número de moléculas reportadas hasta la segunda capa metodológica.	43
Figura 14: Hipótesis de farmacóforos seleccionados basados en cumarinas. a) AANRR y b) AAANR.	45
Figura 15: Resultado de LBVS basada en las hipótesis de cumarinas. Distribución de moléculas con respecto al Phase Screen Score y en línea punteada se observa el cutoff seleccionado (0,65).	46
Figura 16: Flujo de trabajo. Número de moléculas reportadas en el filtro metodológico de “Hipótesis de farmacóforo” con LBVS basado en cumarinas.	47
Figura 17: Distribución de moléculas con respecto a la energía de docking para HTVS en LYN.	48
Figura 18: Conjunto de moléculas fuera del sitio de unión que fueron descartadas en el algoritmo de HTVS en LYN.	49
Figura 19: Distribución de moléculas con respecto a la energía de docking para SP en LYN.	50
Figura 20: Distribución de moléculas con respecto a la energía de docking para XP en LYN.	51
Figura 21: Distribución de moléculas con respecto a la energía de docking para HTVS en KCX.	53
Figura 22: Conjunto de moléculas fuera del sitio de unión que fueron descartadas en el algoritmo de HTVS en la variante KCX.	53
Figura 23: Distribución de moléculas con respecto a la energía de docking para SP en KCX.	54
Figura 24: Distribución de moléculas con respecto a la energía de docking molecular para XP en KCX.	55
Figura 25: Número de moléculas reportadas en el filtro metodológico de “Acoplamiento molecular”	56
Figura 26: Flujo de trabajo. Número de moléculas reportadas post filtro metodológico de “Acoplamiento molecular”	58
Figura 27: Energía de MM-GBSA para cada conjunto de moléculas y compuestos correspondientes al control de calidad interno.	60
Figura 28: Moléculas seleccionadas como candidatas.	66
Figura 29: Interacción de 13 moléculas candidatas con respecto a la hipótesis de farmacóforo AANRR3. Moléculas candidatas: ZINC000003316814 (magenta), ZINC000005810608 (verde claro), ZINC000006751398 (rosa claro), ZINC000008731861 (verde), ZINC000015106043 (morado), ZINC000072427629 (amarillo), ZINC001560403828 (cian), ZINC000002120368 (gris), ZINC00002584177 (azul), ZINC000007994969 (azulino), ZINC000008627328 (naranja claro), ZINC000008731862 (naranja) y ZINC000005810608 (verde limón),	

iones de níquel (naranja) y residuos claves (gris). 67

Figura 30: Interacción de 7 moléculas candidatas con respecto a la hipótesis de farmacóforo AAANR2. Moléculas candidatas: ZINC000008700584 (Verde limón), ZINC000072412074 (amarillo), ZINC000215300901 (magenta), ZINC000002522532 (rosa claro), ZINC000071769481 (naranja claro), ZINC000031169481 (naranja) y ZINC000091624618 (verde), iones de níquel (naranja) y residuos claves (gris). 68

Figura 31: Complejo proteína-ligando, correspondiente a molécula ZINC000008627328 (coloreada en cian) escogida como la mejor candidata de aquellas moléculas que interactúan con ambas variantes, en verde molécula KCX219. 69

Figura 32: Complejo proteína-ligando, correspondiente a molécula ZINC000072412074 (coloreada en cian) escogida como la mejor candidata que solo interactúa con KCX, en verde molécula KCX219. 70

Figura 33: Flujo de trabajo. Número de moléculas reportadas de la metodología completa. 71

ÍNDICE DE TABLAS

Tabla 1: Número de moléculas obtenidas en la descarga y preparación de la base de datos ZINC15.	39
Tabla 2: Número de moléculas obtenidas por cada regla aplicada.	43
Tabla 3: Resultados obtenidos del protocolo LBVS con hipótesis de farmacóforos basado en cumarinas.	44
Tabla 4: Número de moléculas obtenidas por cada set de farmacóforos: AAANR y AANRR.	46
Tabla 5: Número de moléculas obtenidas en el protocolo de docking molecular para LYN.	52
Tabla 6: Número de moléculas obtenidas en el protocolo de docking molecular para KCX.	55
Tabla 7: Número de moléculas post filtrado de docking molecular.	57
Tabla 8: Número de moléculas preseleccionadas.	59
Tabla 9: Energías de ΔG_{bind} para los compuestos DJM, HAE, BME y urea con proteína en 2 estados LYN y KCX.	60
Tabla 10: Números de moléculas con energía igual o más favorable a la de cada compuesto control.	61
Tabla 11: Números de moléculas seleccionadas como precandidatas en el conjunto de moléculas que interactúan con ambas variantes.	63
Tabla 12: Números de moléculas seleccionadas como precandidatas en el conjunto que solo interactúa con KCX.	65
Tabla 13: Tiempo de cálculos y análisis por cada procedimiento de LBVS y SBVS (Considerando una de las variantes).	72

MARCO TEÓRICO

Aspectos generales:

El metabolismo del nitrógeno es fundamental para desarrollar un adecuado estado de salud en los seres vivos, ya que forma parte de una variedad de procesos vitales. El nitrógeno es un elemento esencial presente en todos los aminoácidos; se deriva de la ingesta de proteínas en la dieta, siendo necesario para la síntesis de proteínas y el mantenimiento de la masa muscular, excretándose finalmente por los riñones. La excreción renal de nitrógeno residual consiste casi en su totalidad en urea y amoniaco, el cual puede encontrarse en dos formas moleculares distintas: NH_3 y NH_4^+ , las cuales están en equilibrio entre sí [2].

Las funciones y el metabolismo de la urea son importantes, porque la urea es la principal fuente circulante de compuestos que contienen nitrógeno y juegan un papel relevante en la regulación de la función renal. Los alimentos ricos en proteínas se convierten finalmente en los 9 aminoácidos esenciales y 11 no esenciales, donde la diferencia entre los dos grupos es que los aminoácidos esenciales no pueden ser sintetizados por el organismo y, por tanto, deben aportarse en la dieta. Los aminoácidos pueden ser empleados en dos tipos de rutas metabólicas distintas:

- Anabólico: se pueden usar para sintetizar proteínas.
- Catabólico: se degradan de manera constante en una reacción en la que el grupo α -amino se elimina y se convierte en urea en el hígado [3].

La urea es el sustrato natural de la enzima ureasa (urea amidohidrolasa, *EC 3.5.1.5*), la cual es una metaloenzima ubicua [4]. La ureasa se presenta como una enzima clave para el ciclo global del nitrógeno que se encuentra presente en plantas, hongos y bacterias [5], mas no en animales. Este tipo de enzima acelera 10^{14} veces la tasa de hidrólisis de la urea a amoniaco (NH_3), o amonio (NH_4^+), y a dióxido de carbono (CO_2) [6], en comparación con la descomposición de urea por reacción de eliminación [4].

Desde su descubrimiento en plantas [7], la cristalización de la enzima ureasa aislada de semillas de frijol rojo (*Canavalia ensiformis*) por James B. Sumner en 1926, demostró la naturaleza proteica de las enzimas [4], la cual fue investigada exhaustivamente y se convirtió en un hito en las ciencias bioquímicas como la primera enzima en ser cristalizada [8]. Además, se demostró que su actividad catalítica es estrictamente dependiente de iones de níquel (Ni^{2+}), y dicha

dependencia es una característica única de la ureasa entre las enzimas hidrolíticas [9]. La importancia de este elemento metálico para la función de las ureasas fue establecida en 1975, después de los estudios del grupo de Zerner que revelaron la presencia de iones de níquel en el sitio activo de la ureasa del frijol rojo (JbU, de sus siglas en inglés: *Jack bean urease*) como componente obligatorio para la actividad catalítica de esta enzima [4].

Problemáticas ambientales producidas por ureasas:

El nitrógeno es el segundo elemento más limitante después del carbono para la nutrición de plantas y hongos. Por lo tanto, escindir la urea para formar nitrógeno es esencial para el crecimiento de estas [10]. El uso de fertilizantes nitrogenados en la agricultura es el más alto en comparación con el uso de otros nutrientes. En el período 2018 – 2019, la industria de fertilizantes produjo más de 107 millones de toneladas de fertilizantes nitrogenados en todo el mundo y aproximadamente el 55% de estos están basados en urea [11]. La urea presenta distintas ventajas para la industria, como una alta concentración de nitrógeno (~46%) y menores costos de su producción en comparación con otras fuentes de este elemento químico. Sin embargo, la urea aplicada al suelo sufre una rápida hidrólisis debido a las ureasas de microorganismos que habitan estos nichos, produciendo amoníaco que puede perderse en la atmósfera. Estas pérdidas de amoníaco pueden ser tanto un problema económico (porque quedan menos nutrientes para que las plantas lo absorban, lo que afecta los rendimientos de los cultivos), como un problema ambiental [12].

La liberación descontrolada de NH_3 genera un desequilibrio en el ciclo del nitrógeno, y de este modo genera diferentes consecuencias medioambientales como acidificación o degradación de suelos [13] y lluvia ácida [7]. Además, al oxidarse el amonio a nitrato (NO_3^-), disminuye el potencial de lixiviación de nitratos [14], que es una vía importante de pérdida de nitrógeno en la agricultura. Los inhibidores de la nitrificación también reducen las emisiones de óxido nitroso (N_2O) de los fertilizantes [15], lo que tiene un impacto ambiental negativo, ya que el N_2O es un potente gas de efecto invernadero [16].

Asimismo, el amoníaco es uno de los componentes reactivos claves del nitrógeno y el principal gas alcalino de la atmósfera. El NH_3 contribuye directa y/o indirectamente a la eutrofización de los ecosistemas acuáticos, así como a la acidificación del suelo cuando se nitrifica o lixivía, disminuyendo la biodiversidad

vegetal a través de su deposición, y también daña la salud humana a través de la formación secundaria de aerosoles o partículas [17].

Por consiguiente, la pérdida de nitrógeno desde fertilizantes (entre un 10 y 40%) [16], se presenta como un severo problema económico y medioambiental. Una de las mejores estrategias de manejo para generar una reducción en la volatilización del amonio es incorporar inhibidores de ureasa (IU) mezclados con urea y de esta manera prevenir la actividad catalítica de las ureasas en los suelos [7], [18].

Problemáticas clínicas producidas por ureasas:

Se estima que el 50% de la población mundial está infectada por *Helicobacter pylori* (*H. pylori*) [18], la cual es una bacteria gram-negativa micro-aerofílica [19], cuyo nicho ecológico es el estómago humano. *H. pylori* fue aislado por primera vez de la mucosa gástrica de pacientes con gastritis crónica y cultivado *in vitro* por Warren y Marshall en 1983 [20]. El pH óptimo para el crecimiento de *H. pylori* es alrededor de 8,5 en medios líquidos, por lo tanto, el ambiente altamente ácido en el estómago no es adecuado para su desarrollo, a pesar que puede sobrevivir en estos ambientes. *H. pylori* tiene flagelos largos, que producen motilidad facilitando su movimiento [20]. La forma de espiral, junto con sus flagelos, permite que la célula cruce el gel mucoso viscoelástico pegajoso y se mueva hacia las células epiteliales gástricas [21]. La adhesión de *H. pylori* a la mucosa gástrica es sustancial en su colonización y persistencia a largo plazo en el estómago humano. Varias proteínas de la membrana externa codificadas por *H. pylori* actúan como adhesinas y establecen estrecho contacto con las células epiteliales gástricas [20].

Las consecuencias de la infección se han asociado con el desarrollo de diferentes enfermedades gastrointestinales, tales como úlceras gástricas, linfoma de tejido linfoide asociado a mucosas (MALT) y cáncer de vías biliares. Además, la infección por *H. pylori* también se ha relacionado con enfermedades extra-gástricas, como cardiopatías isquémicas [20], diabetes mellitus tipo 2 [22], anemia [20], rasgos metabólicos adversos en sujetos obesos [23] y resistencia a la insulina [24], entre otros. Asimismo en muchos países, se observó una correlación directa entre la prevalencia de cáncer gástrico e infección por *H. pylori* [25], cuya erradicación a nivel poblacional mostró una reducción de esta patología. *H. pylori* es la única bacteria clasificada en el grupo I de carcinógenos que afectan a humanos, la cual fue introducida en la lista por la Agencia Internacional de Investigación del Cáncer

[25]. El cáncer gástrico es el quinto y tercer tipo de cáncer más común en mujeres y hombres alrededor del mundo, causando en Chile el deceso de más de 3000 personas por año [26].

La ureasa actúa como una enzima secretada por *H. pylori* que contribuye a su protección, hidrolizando urea en NH_3 y CO_2 (Figura 1), lo cual ocasiona neutralización del ácido gástrico y propicia un ambiente más favorable para el desarrollo de esta bacteria [27]. Además, este mecanismo de adaptación de *H. pylori* genera altas concentraciones de amoníaco en el estómago, causando lesiones gástricas [27].

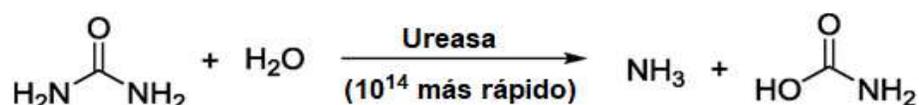


Figura 1: Representación de la hidrólisis de urea por ureasas. Figura adaptada de [28].

El actual tratamiento propuesto para suprimir la infección provocada por *H. pylori* es una combinación de antibióticos de amplio espectro, como amoxicilina y claritromicina, mezclados con inhibidores de la bomba de protones de las células parietales de la mucosa gástrica [29]. No obstante, estas terapias están siendo poco efectivas, ya que *H. pylori* está desarrollando una alta resistencia antibiótica debido al uso indiscriminado de estos fármacos en las últimas décadas. Sumado a esto, el tratamiento posee una serie de efectos secundarios, tales como náuseas, diarrea, angioedema, dolores de cabeza, entre otros síntomas, que propician el abandono de los tratamientos (sin completar los esquemas médicos indicados), lo cual contribuye cada vez más a la resistencia antibiótica. Se han propuesto nuevas estrategias farmacológicas para tratar la infección por *H. pylori*, las cuales están basadas en el uso de inhibidores de ureasa que interrumpen la ruta de patogenicidad y sobrevivencia de esta bacteria. Estos inhibidores deben cumplir una serie de requisitos farmacocinéticos, farmacodinámicos y toxicológicos para ser aprobados como un tratamiento efectivo. Los inhibidores clásicos de ureasa (fosforodiamidatos, derivados del ácido hidroxámico e imidazoles) son tóxicos para los seres humanos, lo que impide sus aplicaciones clínicas [30], [31].

Comparación de ureasas de diferentes organismos:

Las enzimas ureasas son miembros de la superfamilias de amidohidrolasas y fosfotriesterasas, que presentan metales en sus sitios catalíticos. Las ureasas poseen dos iones Ni^{2+} en su sitio activo [32]. En ureasa de *H. pylori* (aquí denominada HpU), estos están separados entre sí por una distancia que oscila entre 3,5 y 3,7 Å, siendo coordinados por H221 ($\text{N}^{\epsilon 2}$), H248 ($\text{N}^{\delta 1}$) y H274 ($\text{N}^{\epsilon 2}$) para un ion de níquel y H136 ($\text{N}^{\epsilon 2}$), H138 ($\text{N}^{\epsilon 2}$) y D362 ($\text{O}^{\delta 2}$) para el otro. Adicionalmente, los residuos C321 y R338 pueden interactuar con ligandos en el sitio activo mediante enlaces de hidrógeno [30]. La protonación de histidinas quelantes de níquel a pH ácido induce la liberación de estos iones con la consecuente pérdida de la actividad enzimática [33].

Estudios de cristalografía de rayos X revelaron que las ureasas vegetales y bacterianas pueden compartir una estructura trimérica común [9], [4]. El número de subunidades, es decir, las cadenas que se forman a partir del "monómero" o unidad funcional varían según el organismo. Para las ureasas de plantas y hongos, esta unidad funcional es una única cadena polipeptídica (α). En tanto que, las ureasas bacterianas están formadas por dos subunidades: α y β , que hasta ahora solo se encuentran presente en el género *H. pylori* o tres subunidades: α , β y γ en otros microorganismos como *Bacillus pasteurii*. La estructura supramolecular más abundante de las ureasas en plantas es un dímero de trímeros (α_3)₂, aunque se han descrito algunas ureasas de plantas diméricas, triméricas y tetraméricas. Muchas ureasas bacterianas son trímeros [$\alpha\beta\gamma$]₃; sin embargo, las ureasas de *H. pylori* se han encontrado principalmente en forma de un tetrámero de trímeros ([$\alpha\beta$]₃)₄ [4], como se puede observar en la Figura 2.

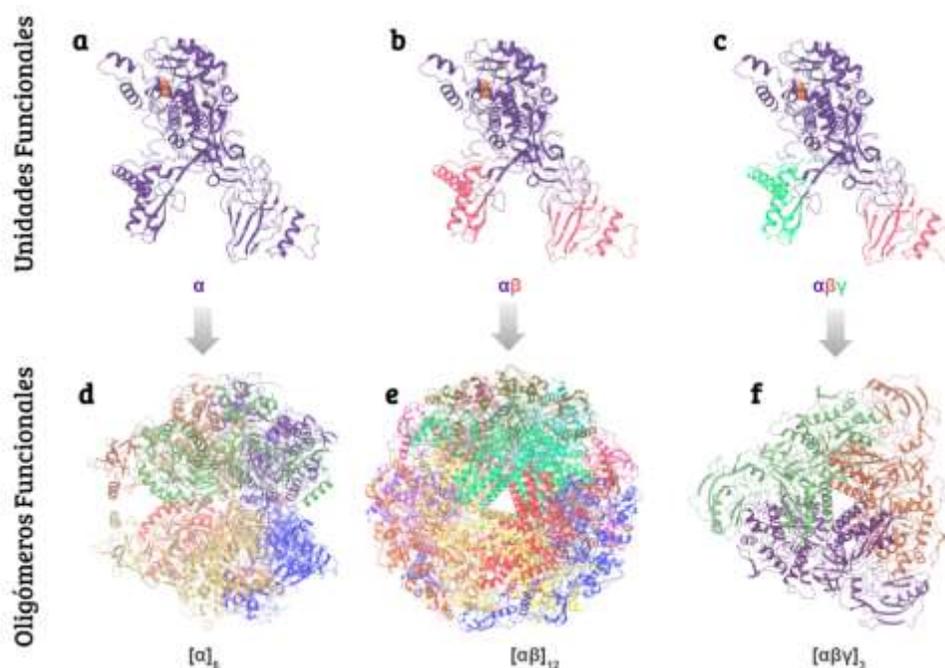


Figura 2: Ensamblaje de la estructuras cuaternarias en ureasas de diferentes organismos. Una unidad funcional puede estar formada por a) una sola unidad (como en *Canavalia ensiformis*, PDB ID: 3LA4), b) un heterodímero (como *H.pylori*, PDB ID: 1E9Z) o c) un heterotrímero (como *Sporosarcina pasteurii*, PDB ID: 2UBP). Estas unidades funcionales (o monómeros) forman complejos como d) hexámeros, e) dodecámeros o f) trímeros. Figura adaptada [4].

Interesantemente, pese a que el monómero es catalíticamente funcional [34], se presume que los arreglos supramoleculares antes mencionados, le confieren capacidades particulares a la enzima. En el caso de HpU, se concibe como hipótesis que el dodecámero confiere protección a los sitios catalíticos del pH ácido del medio [33] mientras que en el sistema JbU, se le atribuye que proporciona la capacidad de aumentar la tolerancia a la desnaturalización producida por altas concentraciones del sustrato [35].

Balasubramanian y cols., compararon las estructuras cristalográficas monoméricas de JbU con ureasas de las bacterias *Klebsiella aerogenes* (KaU), *Bacillus pasteurii* (BpU) y HpU. Los autores identificaron características estructurales y geométricas comunes en la subunidad catalítica, como lo muestran los valores de la desviación cuadrática media (RMSD) obtenidos: 0,60 Å, 0,70 Å y 0,76 Å, para los pares: JbU-KaU, JbU-BpU y JbU-HpU, respectivamente. Entre las diferencias más significativas entre las distintas ureasas, se encuentra una región denominada “*flap loops*” que actúa como puerta de entrada del sustrato en el sitio catalítico. En HpU, este motivo estructural comprende a los residuos ubicados entre las posiciones 313

y 336 que forman un motivo del tipo hélice-loop-hélice, pero sólo las hélices alfa se conservan entre enzimas ureasas de otros microorganismos bacterianos [33]. Se ha demostrado que bloquear este segmento genera una inhibición catalítica sustancial en la enzima [31].

Esta conservación en términos de secuencia y de geometría del sitio catalítico en ureasas de diferentes fuentes sugiere que presentan un mecanismo catalítico común [33], [32]. En donde, la subunidad α de HpU tiene una identidad de secuencia de 59% frente a la subunidad monomérica de JbU [35], como se observa en la Figura 3.

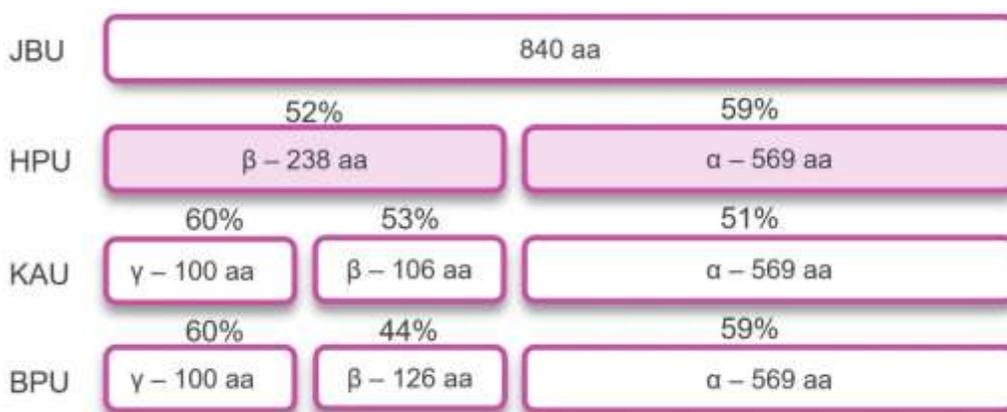


Figura 3: Comparación de JbU y otras ureasas bacterianas. Comparación esquemática de una sola subunidad estructural de JbU con dos (*H. pylori*, HpU) y tres (*K. aerogenes* (KAU), *B. pasteurii* (BPU)) subunidades en ureasas bacterianas. Con referencia a JbU, el porcentaje de identidad de secuencia de las regiones correspondientes en ureasas bacterianas se indica encima del recuadro. Figura adaptada de [35].

Inhibidores de ureasa reportados en la literatura:

Actualmente, los inhibidores de ureasa se presentan con una gran relevancia en los sectores críticos ya mencionados. Numerosos esfuerzos se están centrando en la búsqueda de nuevos inhibidores de ureasa que cumplan con las exigencias medioambientales y clínicas antes expuestas (potentes, no tóxicos y selectivos) [36].

Las fuentes naturales generan una inmensurable cantidad de compuestos orgánicos con demostrada actividad anti-ureasa que podrían ser usados como

punto de partida para el diseño o búsqueda de nuevos inhibidores más potentes. Actualmente, existen diferentes compuestos reportados con actividad anti-ureasa como es el caso de flavonoides [27], polifenoles [37], C-glicosilflavonoides [38], alcaloides, triazoles y tiazoles [40]. Sin embargo, derivados de cumarinas han emergido como una de las más nuevas y potentes familias de compuestos con actividad anti-ureolítica [39] [41], [42], [43], [20], [44], [45].



Figura 4: Compuestos con actividad anti-ureasa.

Fundamentos teóricos:

El cribado de alto rendimiento (HTS, de su nombre en inglés: *High Throughput Screening*) prueba decenas de miles a millones de compuestos para determinar la actividad contra un *target* molecular elegido, y es la estrategia más común en la industria farmacéutica para la identificación inicial de moléculas pequeñas que son activas contra un *target* (i.e. blanco o diana) proteico. Sin embargo, ni los proyectos de investigación más ambiciosos liderados por grandes farmacéuticas a nivel global, disponen de los recursos necesarios para probar de forma experimental tan alta dimensión de bases de datos de compuestos.

Es por ello que el *Virtual Screening* o Cribado Virtual (VS) ha demostrado repetidamente ser útil para enfrentar los desafíos del descubrimiento de fármacos mediante el desarrollo de las capacidades computacionales actuales, así como el desarrollo de diferentes algoritmos que permiten reducir el número de candidatos para ser probados en etapas experimentales posteriores. Esta técnica de *Virtual Screening* tiene por objetivo filtrar grandes bibliotecas de moléculas, reduciendo su

número en cantidades manejables de compuestos para ensayos biológicos y siendo la fuerza impulsora para encontrar nuevos inhibidores [48].

El VS utiliza modelos computacionales para describir las interacciones macromolécula-ligando, donde existe una variedad de métodos para este propósito, que se pueden categorizar principalmente en enfoques 2D (bidimensionales) y 3D (tridimensionales). Los métodos basados en descriptores (métodos 2D) se basan en el cálculo y la comparación de propiedades moleculares escalares y tienen como objetivo identificar moléculas que son similares con respecto a sus atributos moleculares calculados. La mayoría de los enfoques computacionales para el descubrimiento de compuestos se basan en métodos 3D, siendo la finalidad describir la complementariedad estérica y química de las conformaciones tridimensionales de un blanco macromolecular y el ligando de unión [69].

Existen dos categorías amplias que se basan en los métodos 3D, el Cribado Virtual Basado en Ligandos o LBVS (de su nombre en inglés: *Ligand-Based Virtual Screening*) y Cribado Virtual Basado en Estructura (SBVS, de su nombre en inglés: *Structure-Based Virtual Screening*) [50]. Las estrategias de LBVS comparan similitudes moleculares de compuestos activos e inactivos contra bases de datos con el fin de identificar compuestos candidatos para evaluación experimental [51]. Los métodos de LBVS incluyen, como una de sus variantes, al Cribado Virtual Basado en Farmacóforo (PBVS, de su nombre en inglés: *Pharmacophore-Based Virtual Screening*), que es la disposición tridimensional (geométrica y fisicoquímica) de características o patrones esenciales comunes obtenidos de determinadas moléculas activas. SBVS, por otro lado, utiliza la estructura 3D del blanco biológico (determinada experimentalmente a través de cristalografía de rayos X, resonancia magnética nuclear, microscopía electrónica criogénica, o computacionalmente mediante técnicas de modelado molecular). Este método permite acoplar las moléculas candidatas y clasificarlas, atendiendo a su afinidad de unión predicha o a su complementariedad con el sitio de unión (Figura 5) [49].

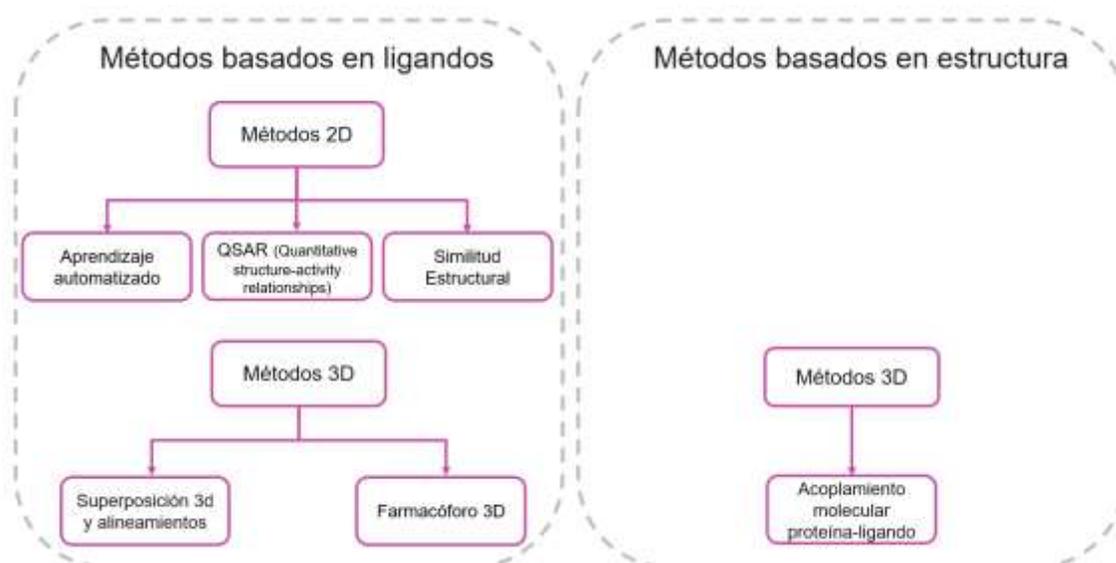


Figura 5: Enfoque de Cribado Virtual empleado en la investigación de candidatos a fármacos. Adaptado de [69].

Esta investigación se centra en la implementación de un protocolo basado en métodos teórico-computacionales como lo es el Cribado Virtual, específicamente la unión de ambas variantes como Cribado Virtual Basado en Farmacóforo y Cribado Virtual Basado en Estructura, mediante estudios de acoplamiento molecular. La finalidad de esta investigación es encontrar moléculas candidatas que puedan ser utilizadas para reducir la pérdida de fertilizantes en los suelos al mezclar inhibidores de ureasa con urea, y de esta manera prevenir la actividad catalítica de ureasa en los suelos. Por otro lado, se requiere que inhibidores de ureasa interrumpan la ruta de patogenicidad y sobrevivencia de *H. pylori* en el organismo evitando enfermedades asociadas a ella.

A través del método de Cribado Virtual se realiza una búsqueda en la base de datos ZINC15 la que proporciona una alta dimensionalidad de moléculas y utilizando como blanco molecular la enzima Ureasa de *H. pylori*. El método consistió en aplicar sucesivamente una serie de filtros como lo fueron: generar un filtrado a la base de datos mediante reglas farmacocinéticas y fisicoquímicas, crear hipótesis de farmacóforos basados en núcleos de cumarinas, realizar estudios de acoplamiento molecular mediante diferentes algoritmos y, por último, realizar cálculos de energía libre de unión (MM-GBSA) a los complejos proteína-ligando. El propósito de estos diferentes filtros es disminuir el número de moléculas de la base de datos, seleccionando un número acotado de compuestos candidatos que tengan una mejor afinidad de unión a la enzima HpU.

HIPÓTESIS

Es posible identificar nuevos inhibidores de ureasa de *H. pylori* mediante una metodología de cribado virtual basada en la comparación entre complejos enzima-inhibidor, considerando la información estructura-actividad de compuestos que inhiben a esta enzima.

OBJETIVOS

Objetivo general:

1. Implementar de una metodología de cribado virtual para identificar nuevos y potentes inhibidores de ureasa de *H. pylori*, utilizando información estructura-actividad disponible.

Objetivos específicos:

- 1.1. Preparar la base de datos de compuestos para ser utilizada como biblioteca virtual.
- 1.2. Filtrar la base de datos a través de diferentes reglas que incluyen términos farmacocinéticos y fisicoquímicos.
- 1.3. Generar Hipótesis de farmacóforos para el conjunto de compuestos con núcleos de cumarinas y Cribado Virtual basado en farmacóforos.
- 1.4. Aplicar métodos de acoplamiento molecular a los compuestos candidatos, mediante diferentes algoritmos para reducir el número de candidatos a inhibidores de ureasa.
- 1.5. Calcular la energía libre de unión mediante el método MM-GBSA.

METODOLOGÍA

Este trabajo de tesis de grado se centró en la búsqueda de nuevos inhibidores para la enzima ureasa de *H. pylori*. Es necesario contextualizar que, previo al inicio formal de dicha tesis en el período correspondiente al segundo semestre del año 2021, ya se había realizado una cantidad de trabajo durante mis prácticas profesionales y períodos no académicos. La información obtenida en ese período sirvió como punto de partida para la realización de esta tesis, por lo cual, es descrita a continuación como antecedentes preliminares.

1) Antecedentes preliminares.

a) Revisión bibliográfica:

Se realizó una búsqueda en revistas científicas indexadas en WOS (Web of Science) usando como criterio de búsqueda: “urease inhibitors” AND (como operador lógico) “coumarins”, en donde solo se analizaron los artículos desde el año 2010 en adelante. Posteriormente, se generó una clasificación de inhibidores basados en la subfamilia química a la cuál pertenecían (cumarinas ácidas de un solo núcleo, *bis*-cumarinas, entre otros), ponderando además, la información reportada en cada artículo: potencia de inhibición, sistema biológico en el cual fueron realizados los ensayos (*H. pylori*, *Canavalia ensiformis*, *Bacillus pasteurii*, entre otros microorganismos), presencia de estudios *in silico* que reporten el sitio y los modos de unión de dichos inhibidores y las condiciones iniciales sobre las que se realizaron estos ensayos computacionales. En aquellos artículos que reportaban una escasa o nula descripción a nivel computacional, se centraron esfuerzos para estudiar y reportar, usando métodos teórico-computacionales, los posibles sitios y modos de unión de las moléculas en la enzima de estudio.

b) Preparación de inhibidores ya reportados en la literatura:

Se utilizó el *software* de visualización Maestro, de la *suite* de Schrödinger, en su versión 2020-3. Cabe mencionar que Maestro es una interfaz gráfica de usuario multiplataforma, con una gran cantidad de herramientas de biología y química computacional. Mediante la herramienta “2D Sketcher” se dibujaron los inhibidores de ureasas reportados en la literatura, adicionando las correspondientes valencias. Los estados de protonación fueron asignados con la herramienta Epik [52] a pH 7,2 (pH en el cual usualmente son realizados los ensayos biológicos en esta enzima).

Finalmente, se generaron los tautómeros y todas las combinaciones de compuestos quirales encontrados.

c) Preparación de la proteína:

La estructura tridimensional de la enzima ureasa de *H. pylori* con mejor resolución a la fecha (2,0 Å, [53]) fue obtenida desde la página web de la Base de Datos de Proteínas [1] con el código PDB: 6zja. Esta estructura fue resuelta mediante microscopía electrónica criogénica en un arreglo dodecamérico, previamente descrito en la Sección “Introducción”. Se ha reportado que cada monómero es funcionalmente activo, por lo que, para disminuir costos computacionales, solo un monómero fue seleccionado para su uso en pasos posteriores. Dicho monómero comprende la subunidad catalítica α y la subunidad accesoria β . Además, cada subunidad catalítica posee dos iones de níquel en el sitio activo y también presenta un inhibidor co-cristalizado DJM (2-[[1-(3,5-dimethylphenyl)-1H-imidazol-2-yl]sulfonyl]-N-hydroxyacetamide), como se observa en la Figura 6. La proteína fue preparada con el *software* Maestro, usando la herramienta “*Protein Preparation Wizard*”, en donde los átomos de hidrógeno fueron adicionados a cada residuo. Los estados de protonación de cada residuo fueron asignados con Propka [54] considerando un valor de pH de 7,2. No fue necesario modelar residuos o segmentos faltantes, pues la estructura se encuentra completamente resuelta. Finalmente, una minimización energética fue realizada en la proteína ya preparada para relajar la estructura luego de las adiciones señaladas, usando como criterio de convergencia un valor RMSD de 0,3 Å entre los átomos pesados de la proteína (cualquier tipo de átomo menos los hidrógenos).

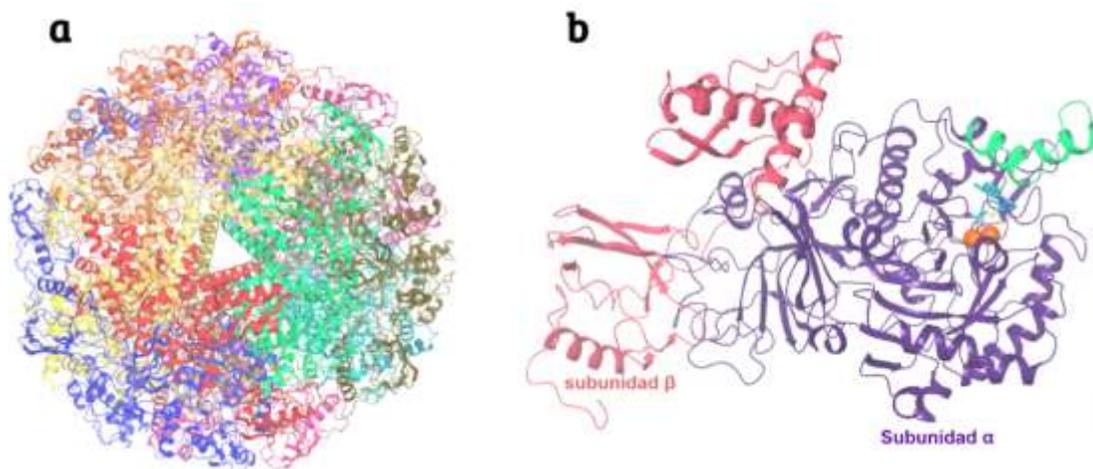


Figura 6: Estructuras tridimensional de HpU. (a) dodecámero, (b) unidad funcionalmente activa comprendiendo la subunidad β o accesoria, subunidad α o catalítica, dos iones de níquel y el inhibidor co-cristalizado DJM (2- $\{[1-(3,5\text{-dimethylphenyl})-1H\text{-imidazol-2-yl]sulfanyl}\}$ -N-hydroxyacetamide).

d) Acoplamiento molecular:

Se realizaron estudios de acoplamiento molecular (*i.e. docking*, por su denominación en inglés) mediante la herramienta *Glide* [55] de la suite de Schrödinger. Se ha reportado que algunas moléculas han generado inhibiciones no solo del tipo competitivo sino también no competitivo y mixto, proponiendo la unión en sitios alostéricos en la enzima ureasa [56]. Es por ello que se construyó una grilla cúbica cubriendo completamente la subunidad catalítica de la enzima ureasa, usando una grilla interna de 15 Å x 15 Å x 15 Å y una grilla externa de 36 Å x 36 Å x 36 Å. Luego, se realizaron experimentos de *docking* con el método acoplamiento por ajuste inducido (IFD, de sus siglas en inglés: *Induced Fit Docking*, [57]) o también conocido como acoplamiento flexible, el cual considera tanto las pequeñas relajaciones del esqueleto peptídico en la proteína como cambios significativos de las cadenas laterales de los aminoácidos, tras la unión de un ligando en el sitio de la proteína. El método consiste en 5 etapas consecutivas, 1) los ligandos son acoplados en la grilla externa que cubre a la enzima, buscando los modos de unión energéticamente más favorables en la grilla interna y permitiendo variar los grados de libertad (en términos de variaciones de conformación y orientación) de cada ligando, incrementándose así la exploración de la superficie de energía potencial. 2) Una vez encontrados los sitios y modos de unión, se genera una minimización energética que permite el refinamiento de las cadenas laterales de los aminoácidos de la proteína que están a cierta distancia del ligando (5 Å en este caso), usando el campo de fuerza OPLS3e [58]. 3) Posteriormente, se realiza una minimización energética del complejo proteína-ligando, haciendo que el sitio se ajuste, inducido por la unión del ligando. 4) Se genera una nueva etapa de acoplamiento molecular en cada complejo, usando el sitio de unión previamente ajustado por la

incorporación del ligando (ajuste inducido), en donde, aquellas moléculas con una energía de unión menor a 30 kcal/mol de la conformación más favorable son nuevamente reacopladas. 5) Finalmente, los puntajes de cada pose son calculados, y descritos principalmente por los términos: “*Prime energy*” y “*Glide Gscore*”, los cuales debiesen ser considerados en conjunto para la correcta selección y categorización de las poses obtenidas, a partir de estos parámetros energéticos [57]. Además, se realizó el mismo procedimiento ya mencionado, utilizando el inhibidor co-cristalizado DJM en la enzima ureasa, como control de calidad interno, para evaluar la fiabilidad de las funciones de puntuación que se emplearon, intentando reproducir el modo de unión del inhibidor DJM en la estructura 3D de ureasa. En esta etapa, se obtuvieron como resultado a lo más 10 poses por ligando.

Seguidamente, se realizaron cálculos energéticos de los complejos proteína-ligando con el método MM-GBSA (de su nombre en inglés: *Molecular Mechanics-Generalized Born Surface Area*), el cual es utilizado para la predicción de la energía libre de unión (ΔG_{bind}), entre ligandos y macromoléculas biológicas. MM-GBSA es descrito como un método preciso para la estimación de la energía libre de unión relativa para una serie congénica de compuestos [59], permitiendo obtener un valor estimado de la energía de unión para cada uno de los compuestos con demostrada actividad inhibitoria sobre ureasa. Estos valores fueron contrastados con los valores obtenidos a nivel experimental (IC_{50}), obteniendo así un coeficiente de determinación lineal (R^2) [60].

Finalmente, aquellas moléculas con núcleos de cumarinas (Figura 7), fueron usadas como puntos de partida para elaborar hipótesis de farmacóforos, siendo empleados en la realización de esta tesis, como se explica más adelante en este documento.

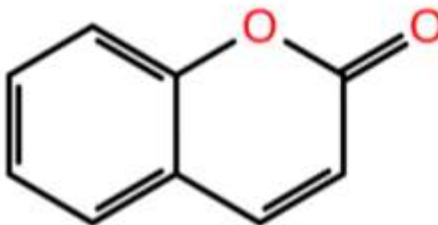


Figura 7: Núcleo de estructuras 2D

2) Metodología del Trabajo de tesis de grado

Para dar cumplimiento a los objetivos propuestos, se ha generado un flujo de trabajo presentado en la Figura 8, empleando el método de Cribado Virtual, el cual consiste en diferentes técnicas computacionales utilizadas para descubrir compuestos con actividad en un determinado blanco farmacológico, que son ampliamente utilizados en la etapa temprana del desarrollo de fármacos. Esto se realiza mediante la búsqueda en bases de datos de compuestos (bibliotecas virtuales) de alta dimensionalidad con la finalidad de seleccionar un número limitado de compuestos candidatos que tengan una mayor afinidad de unión a un receptor biológico (HpU en este caso) o que compartan características fisicoquímicas, geométricas o farmacocinéticas con compuestos cuya actividad sobre el blanco ya han sido probados.

Se aplicaron 6 filtros sucesivos que permitieron la reducción sistemática del número de compuestos para finalmente identificar algunos candidatos como potenciales nuevos inhibidores de ureasa. Esta tesis se enmarca en un proyecto mayor, el cual involucra la evaluación experimental de las moléculas candidatas como inhibidores de la enzima de *H. pylori* mediante posteriores ensayos *in-vitro*.

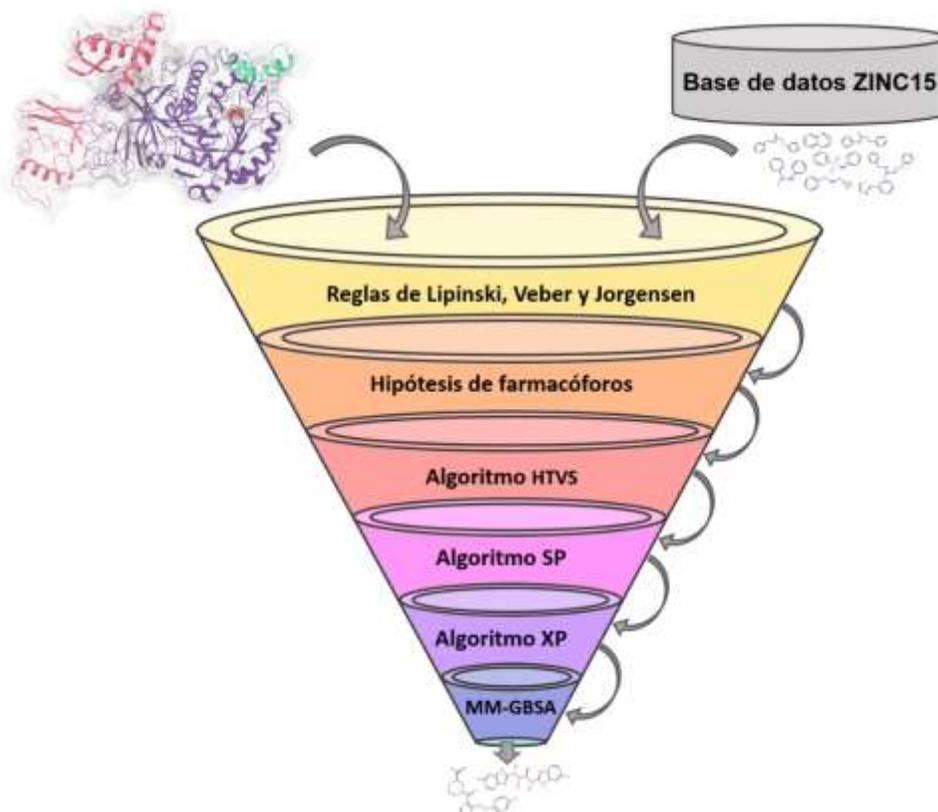


Figura 8: Flujo de trabajo. Protocolo de Virtual Screening para la búsqueda de nuevos y potentes inhibidores de ureasa.

1) Preparación de la base de datos de compuestos para ser utilizada como biblioteca virtual:

a) **Base de datos de compuestos ZINC15** [61]. ZINC15 es una base de datos de libre acceso que presenta más de 230 millones de compuestos comercialmente disponibles para ser comprados en diferentes empresas farmacéuticas y distribuidores alrededor del mundo. La información estructural de los compuestos pueden ser descargada en varios formatos como: .sdf, .mol2, .json, por mencionar algunos, lo que facilita tareas relacionadas al diseño o búsqueda de nuevos fármacos. Aquí es necesario enfatizar que pese a que la molécula no será nueva, si lo será su función como ligando con una probable afinidad por ureasa. Por otro lado, ZINC15 también presenta información química de los compuestos que almacena, como: peso molecular (PM), coeficiente de partición (LogP), carga total neta a un determinado pH, área de superficie polar (PSA, de su nombre en inglés: *Polar Surface Area*), número de donadores y aceptores de enlaces de hidrógeno y número de enlaces rotables. Todas estas características pueden ser descargadas junto con la información de cada compuesto en plantillas de datos (.csv).

Para este trabajo, se descargó y fue usada localmente la base de datos ZINC15 [61] en formato mol2, donde se utilizaron los criterios de: venta en stock (*now*), coeficiente de partición ≤ 5 y peso molecular ≤ 500 g/mol. Lo cual será explicado en el punto 2 de esta sección.

b) **Preparación de la base de datos.** Se eliminaron todas las moléculas con nombre repetido, y a través de la herramienta *LigPrep* de la *suite* de Schrödinger se preparó la base de datos con el campo de fuerza OPLS3e, a un pH 7,2, en donde además se determinaron las quiralidades generando a lo más 2 poses por ligando.

2) Filtrado a la base de datos a través de diferentes reglas que incluyen términos farmacocinéticos y fisicoquímicos:

Filtros basados en las características farmacocinéticas de los compuestos que forman la base de datos ZINC15 descargada. Existen métodos teóricos que permiten evaluar una variedad de características

fisicoquímicas y farmacocinéticas (FC) para cada compuesto con el propósito de determinar si se cumple con algunos criterios ADME (Administración, Distribución, Metabolismo y Excreción) que son considerados necesarios para clasificar una molécula determinada como fármacos (usualmente denominado en inglés como *drug-like*). Estas características intentan asegurar una buena biodisponibilidad y baja toxicidad (farmacodinámica o FD) del compuesto en el cuerpo humano. Para ello, y tras el estudio de estas características FC y FD sobre grandes poblaciones de compuestos aprobados por la Administración de Drogas y Alimentos (FDA, de sus siglas en inglés: *Food and Drug Administration*), se han ido identificando algunas reglas útiles que permiten discriminar aquellas moléculas que presentan características deseadas:

- a) **“Regla de los 5”** (RO5, del inglés: Rule of Five) reportada por Christopher A. Lipinski. La regla describe las propiedades moleculares que tienen importancia FC en el cuerpo humano. [62]. Los requisitos son los siguientes:
 - i) Peso molecular ≤ 500 g/mol.
 - ii) $\text{LogP} \leq 5$.
 - iii) Aceptores de enlaces de hidrógeno ≤ 10 .
 - iv) Donadores de enlaces de hidrógeno ≤ 5 .

- b) **“La regla de tres de Jorgensen”** (ROT, del inglés: Rule of Three), se encuentran relacionadas con farmacocinética (ADME), las cuales son determinantes para conocer la biodisponibilidad oral de las moléculas candidatas a fármacos [64]. Los requisitos son:
 - i) Factor de solubilidad acuosa (logS) $> -5,7$.
 - ii) Permeabilidad aparente en las células Caco-2 (Caco) > 22 nm/s.
 - iii) Número de metabolitos primarios < 7 .

- c) **“Regla de Veber”**. Se describe como una regla para evaluar la biodisponibilidad que tendrá el candidato a fármaco, considerando una reducida flexibilidad molecular, medida en función del número de enlaces rotables, el área de superficie polar y el recuento total de enlaces de hidrógeno (suma de donadores y aceptores). Por lo tanto, es importante conocer las propiedades moleculares que limitan la biodisponibilidad oral para facilitar el diseño de nuevos candidatos a fármacos viables [63]. Los requisitos son los siguientes:
 - i) Enlaces rotables ≤ 10 .
 - ii) Área de superficie polar (PSA) ≤ 140 Å.

iii) Total de enlaces de hidrógeno < 12 (entre donadores y aceptores).

La base de datos completa se sometió a un análisis con el *software Qikprop* de la *suite* de Schrödinger. Esta herramienta permitió obtener las 10 propiedades antes mencionadas en las 3 reglas descritas, así como otros 39 descriptores.

De los resultados obtenidos se seleccionó el 59% de los compuestos, donde aquellos compuestos que violaron dos o más reglas de Lipinski, o una o más reglas de Jorgensen y Veber fueron descartados en los siguientes pasos metodológicos.

3) Creación de hipótesis de farmacóforos para el conjunto de compuestos con núcleos de cumarinas.

Hipótesis de farmacóforos. Considerando que un protocolo PBVS es un subtipo de Cribado Virtual basado en ligandos y que utiliza un conjunto de características estéricas y electrónicas que son necesarias para asegurar la óptima interacción, el farmacóforo sirve como plantilla para buscar en una base de datos aquellos compuestos con características químicas similares. Se utilizó el conjunto de inhibidores basados en cumarinas descritos en los antecedentes preliminares de esta sección metodológica, que fue empleado para generar hasta 10 hipótesis de farmacóforos con la herramienta *Phase* [65] de la *suite* Schrödinger. Se consideraron de cuatro a seis características farmacofóricas: donador de enlaces de hidrógeno, aceptor de enlaces de hidrógeno, grupos con carga negativa, grupos con carga positiva, grupos hidrofóbicos y anillos aromáticos.

Se escogieron las 2 mejores hipótesis con respecto a la medición de BEDROC la cual está acotada por el intervalo [0, 1] y este se puede interpretar como la probabilidad de que un compuesto activo se clasifique con mayor rapidez que un compuesto seleccionado de forma aleatoria, de esta manera midiendo el reconocimiento temprano de activos de la base de datos, y estas hipótesis de farmacóforos fueron aplicadas a la biblioteca de compuestos obtenidos del paso 2), con la finalidad de encontrar compuestos que cumplan con las condiciones que presentan las diferentes hipótesis de farmacóforos.

En relación con el número de compuestos obtenidos del LBVS basado en cumarinas en el procedimiento realizado, se extrajo un 16% de los

compuestos con mejor “*PhaseScreenScore*”, el cual corresponde a un puntaje asignado al alineamiento de los compuestos con las características de ambos farmacóforos.

4) Realización de estudios de acoplamiento molecular a los compuestos candidatos, mediante diferentes algoritmos, para reducir el número de candidatos a inhibidores de ureasa:

a) **Se realizó un cribado virtual de alto rendimiento** (HTVS, de su nombre en inglés: *High-Throughput Virtual Screening*), que fue aplicado sobre el conjunto de compuestos obtenidos del paso 3). HTVS consiste en un acoplamiento molecular para predecir principalmente cómo interactúa una proteína con moléculas pequeñas, a fin de identificar la energía de unión del complejo proteína-ligando. Se utilizó la herramienta *Glide* [55] de la *suite* Schrödinger y el algoritmo HTVS con el campo de fuerza OPLS en su versión OPLS3e [58]. Se realizó este procedimiento tanto en la proteína que está en presencia de una lisina 219 carbamilada (KCX) y también se realizó el mismo procedimiento en la proteína cuya lisina 219 se encuentra en estado neutro (LYN). Desde este punto, los procedimientos aplicados a cada proteína resultan denotados como KCX y LYN de acuerdo a como se encuentra modificada el residuo lisina 219. Se consideró una grilla interna lo suficientemente grande para permitir que los compuestos roten sin restricciones (15 Å x 15 Å x 15 Å). Por otro lado, la grilla externa cubre completamente la subunidad catalítica, similar a la usada en el paso 1d) de antecedentes preliminares (36 Å x 36 Å x 36 Å) (Figura 9), y el centro de la grilla se determinó por un centroide creado a partir de las coordenadas de los residuos claves de HpU que forman el sitio catalítico (H136, H138, KCX219/LYN219, H221, H248, H274, C322, R338, D362, en HpU). Todos los estudios de acoplamiento molecular fueron realizados en presencia de ambos iones Ni²⁺ en el sitio catalítico de la enzima. Tres poses fueron obtenidas por cada compuesto y, en relación al número de compuestos obtenidos en el procedimiento realizado, se extrajo un 26% en LYN y 30% en KCX, considerando los compuestos con mejor energía de *docking* “*Glide Gscore*”, además los compuestos que se encontraron posicionados fuera del sitio catalítico fueron descartados, independiente de la energía de unión que reportaron.

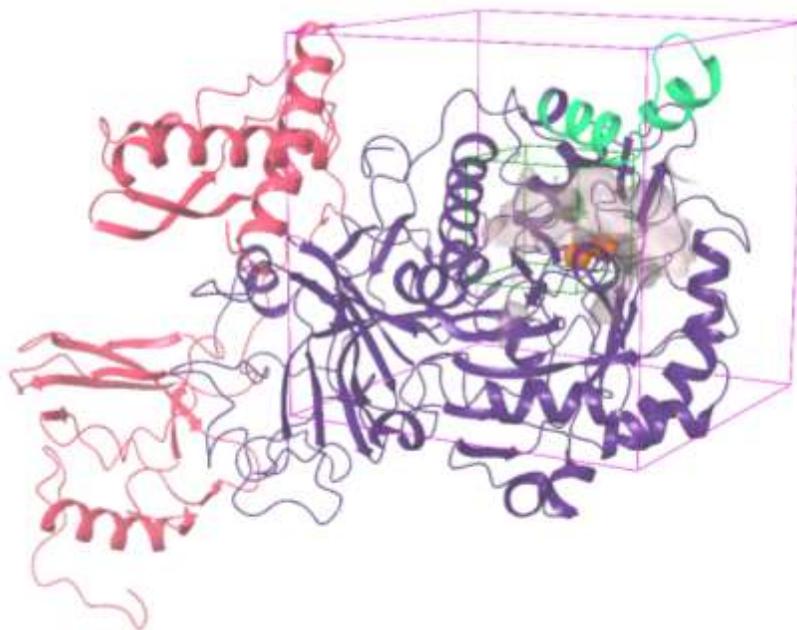


Figura 9: Grilla utilizada en el método de acoplamiento molecular. Grilla cúbica externa de 36 Å x 36 Å x 36 Å, cubriendo la subunidad catalítica (en color morado). El flap-loop que actúa como una compuerta al sitio catalítico se encuentra representada en color verde, en color naranja se muestran los dos iones de níquel. La superficie de color gris muestra los residuos del sitio catalítico.

b) Docking molecular con los algoritmos SP y XP. Desde el conjunto de compuestos obtenidos en el punto anterior, se volvieron a realizar dos etapas sucesivas de acoplamiento molecular, siguiendo el mismo protocolo descrito en el punto anterior pero mejorando sucesivamente la función de puntuación empleada. Estos estudios de *docking* se realizaron mediante la herramienta *Glide* [55] de la *suite* Schrödinger y las funciones de puntuación de precisión estándar (SP, del inglés: *Standard Precision*) y de extra precisión (XP, del inglés: *Extra Precision*). Tanto HTVS como SP comparten la misma función de puntuación, pero HTVS reduce la minuciosidad con que son refinados los grados de libertad torsionales de cada ligando, haciendo que el muestreo conformacional en SP sea más exhaustivo que en HTVS [66]. En la etapa correspondiente a la aplicación del algoritmo SP, se seleccionó un 22% para LYN y 19% para KCX, seleccionando las mejores energías de unión reportadas, generando dos nuevos conjuntos de compuestos más acotados, donde los compuestos que se encontraron posicionados fuera del sitio catalítico fueron descartados, independiente de la energía de unión que reportaron.

En el caso de XP, se seleccionó un 4% para LYN y 5% para KCX de las mejores energías de unión reportadas considerando que los compuestos que se encontraron posicionados fuera del sitio catalítico fueron descartados, independiente de la energía de unión que reportaron. A este nuevo conjunto de compuestos, se le aplicó un filtro que consiste en eliminar moléculas con el mismo nombre, dejando solo una de ellas (la que presenta mejor energía de *docking*). Seguidamente, al conjunto de datos obtenidos del filtro anterior, se volvió a aplicar una nueva condición que descartó los compuestos que se encontraron alejados de los iones de níquel a más de 3,5 Å. Finalmente, se realizó una última selección que permitió eliminar las moléculas duplicadas, considerando un puntaje de corte (*cutoff*) de 0,0 Å de RMSD (dejando la pose de mejor energía de *docking*) mientras que las moléculas resultantes por cada conjunto de moléculas (KCX y LYN) fueron utilizadas para el siguiente paso metodológico.

5) Realización de cálculos de energía libre de unión mediante el método MM-GBSA:

Cálculos de la energía de unión proteína-ligando mediante el método MM-GBSA. El método de MM-GBSA, como fue descrito anteriormente, es utilizado para la predicción de la energía libre de unión, entre ligandos y macromoléculas biológicas. Los valores de ΔG_{bind} , usando el método MM-GBSA, fueron obtenidos mediante la herramienta *Prime* [67] de la *suite* de Schrödinger, en donde todos los átomos de la proteína que se localizaron en una esfera de 5 Å del ligando en todas direcciones, se refinaron como una región flexible, permitiendo adaptar el sitio de unión al ligando previamente posicionado en etapas anteriores. Para esta tarea, se utilizó un modelo de solvente implícito optimizado denominado VSGB (del inglés, *Variable Dielectric Surface Generalized Born*), que se basa en el modelo SGB (en inglés *Surface Generalized Born*) y el tratamiento dieléctrico variable (VD) de la polarización de las cadenas laterales de proteínas, siendo un modelo de energía para el ajuste de estructuras proteicas de alta resolución, presentando correcciones basadas en la física para enlaces de hidrógeno, interacciones π - π , e interacciones hidrofóbicas [68]. Los parámetros para los átomos de cada complejo proteína-ligando, así como los iones de níquel presentes en el sitio catalítico fueron obtenidos del campo de fuerza OPLS3e [58].

Luego de haber obtenido los resultados de ΔG_{bind} con MM-GBSA, se formaron 3 nuevos conjuntos de datos, el primero de ellos corresponde a

aquellas moléculas que se encuentran en ambos grupos (LYN y KCX), es decir, moléculas capaces de unirse al sitio de unión de ureasa independiente del estado en que se encuentre la lisina 219, el segundo corresponde a las moléculas que solo se encontraron en LYN y el tercero las moléculas presentes solo en KCX. Seguidamente en cada uno de los conjuntos de datos se descartaron aquellas moléculas que presentan una energía mayor o igual a cero kcal/mol. Posteriormente se escogieron 80 moléculas con valores de ΔG_{bind} obtenidos con MM-GBSA inferiores al ΔG_{bind} del compuesto DJM.

Mediante métodos teóricos se evaluaron las 80 moléculas seleccionadas como pre-candidatas con la herramienta web *SwissADME* que brinda acceso gratuito a un grupo de modelos predictivos rápidos pero robustos para propiedades fisicoquímicas, farmacocinéticas, semejanza con las drogas y compatibilidad con la química médica [73], enfocándose principalmente en características fisicoquímicas y farmacocinéticas con la intención de que cumplan con algunos criterios ADME. Estos criterios corresponden a:

a) Filtro de Lipinski [62]:

- i) Peso molecular ≤ 500 gr/mol.
- ii) Coeficiente de partición $\leq 4,15$.
- iii) Aceptores de enlaces de hidrógeno ≤ 10 .
- iv) Donadores de enlaces de hidrógeno ≤ 5 .

b) Filtro de Ghose [74]:

- i) Peso molecular entre 160 y 480 gr/mol.
- ii) Coeficiente de partición entre -0,4 y 5,6.
- iii) Recuento de átomos entre 20 y 70.
- iv) Refractividad molar entre 40 y 130.

c) Filtro de Veber [63]:

- i) Número de enlaces rotables ≤ 10 .
- ii) Área de superficie polar $\leq 140 \text{ \AA}$.
- iii) Total de enlaces de hidrógeno < 12 (donadores y aceptores).

d) Filtro de Egan [75]:

- i) Coeficiente de partición $\leq 5,88$.
- ii) Área de superficie polar $\leq 131,6$.

e) Filtro de Muegge [76]:

- i) Peso molecular entre 200 y 600 gr/mol.
- ii) Coeficiente de partición entre -2 y 5.
- iii) Área de superficie polar ≤ 150 .
- iv) Número de anillos ≤ 7 .
- v) Número de carbonos > 4 .
- vi) Número de heteroátomos > 1 .
- vii) Número de enlaces rotables ≤ 15 .
- viii) Aceptores de enlaces de hidrógeno ≤ 10 .
- ix) Donadores de enlaces de hidrógeno ≤ 5 .

f) Alerta estructural BRENK [77]: consiste en una lista de 105 fragmentos identificados por como putativamente tóxico, químicamente reactivo, metabólicamente inestable o tener propiedades responsables de una mala farmacocinética.

g) PAINS [78]: (o compuestos promiscuos) son moléculas que contienen subestructuras que muestran una respuesta potente en los ensayos

independientemente de la proteína *target*. Dichos fragmentos, que producen resultados biológicos falsos positivos, han sido identificados al analizar seis ensayos ortogonales y descomponer las moléculas activas en 2 o más ensayos en 481 fragmentos recurrentes, considerados como potencialmente conducentes a compuestos promiscuos.

h) Leadlikeness [76]:

- i) Peso molecular entre 250 y 350 gr/mol.
- ii) Coeficiente de partición $\leq 3,5$.
- iii) Número de enlaces rotables ≤ 7 .

De este conjunto de moléculas recientemente filtradas con *SwissADME* se seleccionaron como candidatas las 20 mejores, siendo las que presentan la menor o nula cantidad de violaciones a las reglas que fueron evaluadas.

RESULTADOS

A continuación, se presentan los resultados obtenidos en la tesis de grado, donde la metodología de *Virtual Screening* fue utilizada para la identificación de moléculas candidatas para la inhibición de HpU.

1) Preparación de la base de datos de compuestos para ser utilizada como biblioteca virtual:

- a) La base de datos ZINC15 fue descargada en formato mol2 bajo los criterios venta en stock (*now*), acotando la descarga a aquellas moléculas con un coeficiente de partición ≤ 5 y un peso molecular ≤ 500 g/mol, siendo un total de 7.153.060 moléculas. Tras un análisis exploratorio a los archivos descargados fue posible encontrar un gran número de moléculas repetidas (con el mismo identificador/nombre) dentro de la base de datos ZINC15, correspondiente a tautómeros las que abarcan alrededor del 7,5% de la base de datos descargada, estas moléculas fueron eliminadas dejando solo una de cada molécula repetida, reduciendo la biblioteca de compuestos a 6.619.086 moléculas.

Luego al ser preparadas con la herramienta *LigPrep* de la *suite* de Schrödinger, se generaron 2 poses, como máximo, por cada molécula teniendo un total de 8.271.505 moléculas (Tabla 1).

Filtro	Número de moléculas
Base de datos Zinc15 descargada	7.153.060
Base de datos filtrada (eliminación de repetidas)	6.619.086
Base de datos preparada	8.271.505

Tabla 1: Número de moléculas obtenidas en la descarga y preparación de la base de datos ZINC15.

2) Filtrado a la base de datos a través de diferentes reglas que incluyen términos farmacocinéticos y fisicoquímicos:

- a) Al conjunto de moléculas preparadas con *LigPrep* recién mencionadas se aplicaron un conjunto de filtros pertenecientes a las reglas que incluyen términos farmacocinéticos y fisicoquímicos, los que se encuentran detallados en la Sección 2a) de “Metodología”. Al aplicar la “Regla de los 5”, se redujo la base de datos en un 20%, lo que corresponde a 6.591.236 moléculas considerando que solo se permitía una violación a cualquiera de los cuatro términos. En la Figura 10, se observa la distribución de moléculas en relación a cada característica perteneciente a la regla de Lipinski, donde en el caso de peso molecular, se consideraron todas las moléculas con un *cutoff* igual o menor a 500 g/mol; para coeficiente de partición, todas las moléculas con *cutoff* menor o igual a 5. En el caso de aceptores de enlaces de hidrógeno, se seleccionaron las moléculas con una cantidad de hasta 10 aceptores y donores de enlaces de hidrogeno, seleccionándose aquellas que se encuentran entre 0 y 5.

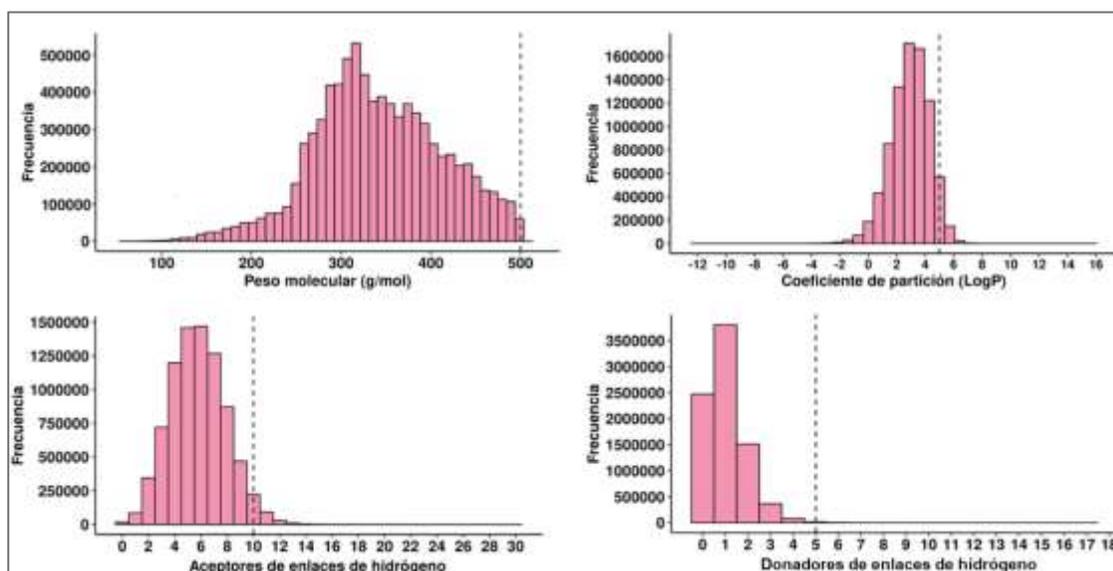


Figura 10: Reglas de los 5 de Lipinski. Se observa la distribución de moléculas por cada característica asociada a la regla, en línea punteada se encuentra ubicado el *cutoff* que se aplicó para el número de moléculas seleccionadas.

Cabe mencionar que si bien se realizó la descarga de la base de datos ZINC15, considerando dentro de los filtros que están moléculas tuvieran un peso molecular menor o igual a 500 g/mol y el coeficiente de partición fuera menor o igual a 5, es necesario corroborar la información. Específicamente porque al aplicar *Qikprop* a la base de datos, se logró reconocer que este

filtro aplicado por ZINC15 no es completamente fiable pues existen moléculas que violaron estas condiciones al momento de descargar el conjunto de moléculas que se utilizó como biblioteca de compuestos.

- b) Seguidamente, se filtró acorde a la “Regla de los 3” de Jorgensen, en donde el número de la base de datos fue reducida a 5.018.7683 moléculas, disminuyendo la biblioteca de compuestos obtenida respecto del filtro anterior en un 24%. En el caso de esta regla, se eliminaron todas aquellas moléculas que violasen uno o más criterios. En la Figura 11, se observa la distribución de moléculas en relación a cada característica perteneciente a la regla. En relación al factor de solubilidad acuosa, se consideran las moléculas que tienen un *cutoff* mayor a -5,7. Mientras que para la permeabilidad en células Caco-2, se consideraron solo las moléculas tienen un *cutoff* mayor a 22 nm/s. Por último, para el caso del número de metabolitos primarios, se consideran las moléculas que presentan entre 0 a 7 metabolitos.

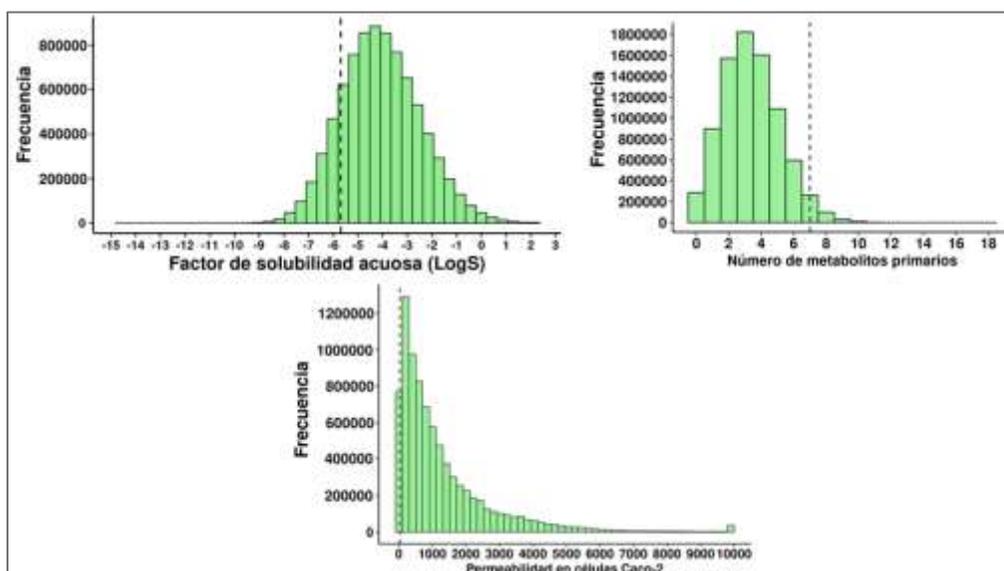


Figura 11: Reglas de los 3 de Jorgensen Se observa la distribución de moléculas por cada característica asociada a la regla, en línea punteada se encuentra ubicado el *cutoff* que se aplicó para el número de moléculas seleccionadas.

- c) Como tercera regla, se aplicó la “Regla de Veber”, en la cual no se aceptaron infracciones, donde se obtuvo un total de 4.903.299 moléculas, acotando la base de datos con respecto a la regla anterior en un 2%. En la Figura 12, se observa la distribución de moléculas en relación a cada característica perteneciente a esta regla, siendo el número de enlaces rotables

considerando como *cutoff* menor o igual a 10, en el caso de área de superficie polar todas las moléculas que presentan un PSA menor o igual a 140 Å y por último la suma de aceptores y donadores de enlaces de hidrógeno menor a 12.

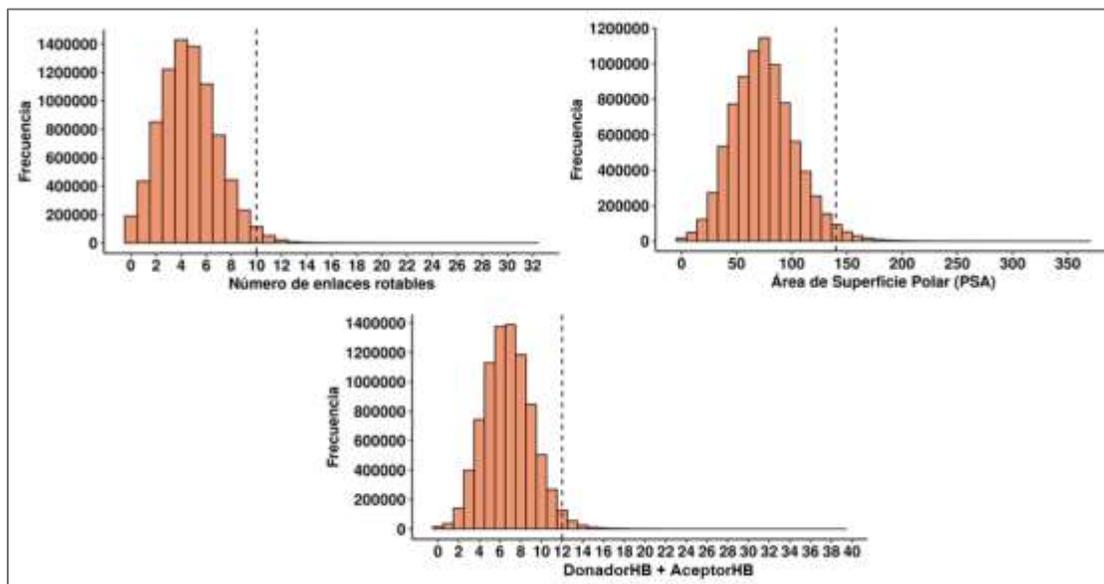


Figura 12: Regla de Veber. Se observa la distribución de moléculas por cada característica asociada a la regla, en línea punteada se encuentra ubicada el *cutoff* que se aplicó para el número de moléculas seleccionadas.

Al aplicar las tres reglas antes mencionadas, se obtiene como número de moléculas final de esta capa de VS un total de 4.903.299 moléculas, las que corresponden al 59% de la base de datos con las moléculas preparadas del objetivo anterior. En la Tabla 2, se encuentra el número de moléculas obtenidas post filtro para cada una de ellas. Este objetivo nos permite asegurar, al menos teóricamente, que cada molécula que cumplió con los criterios de las reglas de Lipinski, Jorgensen y Veber poseen características aceptables de biodisponibilidad oral, uno de los criterios farmacocinéticos claves en el desarrollo de compuestos bioactivos y terapéuticos [63][70].

Reglas	Número de moléculas post filtro
Regla de los 5 – Lipinski	6.591.236
Regla de los 3 – Jorgensen	5.018.783

Regla de Veber	4.903.299
----------------	-----------

Tabla 2: Número de moléculas obtenidas por cada regla aplicada.

En la Figura 13, se observa el número de moléculas obtenida por cada capa de VS aplicado.

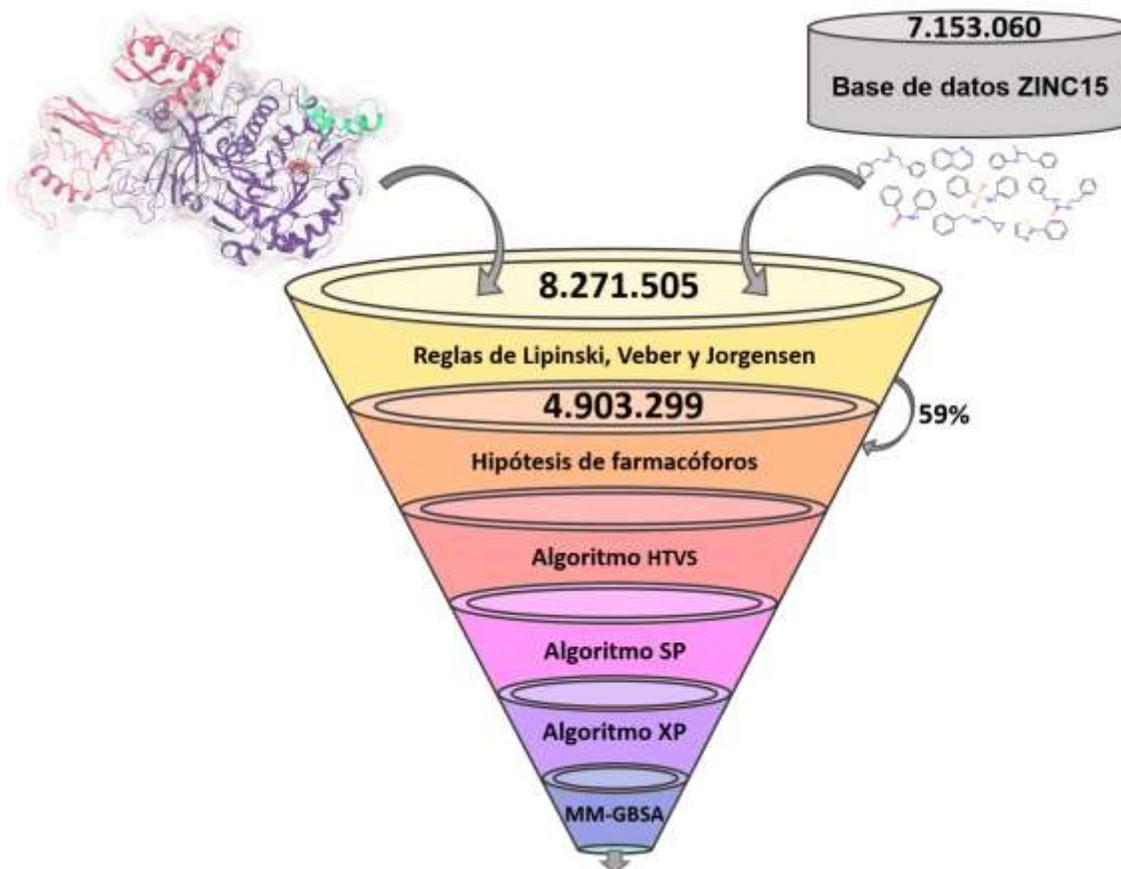


Figura 13: Flujo de trabajo. Número de moléculas reportadas hasta la segunda capa metodológica.

3) Hipótesis de farmacóforos para el conjunto de compuestos con núcleos de cumarinas y Cribado Virtual basado en farmacóforos.

Mediante el protocolo de LBVS y la herramienta *Phase* de la suite de Schrödinger, se utilizó el conjunto de inhibidores de cumarinas como plantilla para la creación de 10 hipótesis de farmacóforos. En la Tabla 3, se puede observar las

características de las 10 hipótesis con núcleo de cumarinas que se generaron, siendo escogidas AANRR3 con un puntaje de 0,985 y AAANR2 con un puntaje de 0,980 (Figura 14), ya que se consideró dentro de la elección el puntaje de BEBROC, además del número de características (priorizando aquellas con más características) y que estas no tuvieran: distancia, ángulos y características, iguales o con una gran similitud entre las hipótesis. El puntaje de BEDROC (Curvas ROC mejoradas por discriminación de Boltzmann) corresponde a un valor estadístico utilizado en VS, el cual se encuentra acotado por un intervalo de [0, 1] y se interpreta como la probabilidad de que el compuesto activo se clasifique con mayor rapidez que un compuesto seleccionado de forma aleatoria [71]. Otro de los criterios empleados fue seleccionar las hipótesis que impliquen a la mayor cantidad de características fisicoquímicas, debido a que mientras mayor sea el número de características, mayor también será la discriminación de moléculas que se ajuste a dichas características. Finalmente, también se consideró la geometría de estas hipótesis, procurando que los dos farmacóforos seleccionados no sean espacialmente similares en distancias y ángulos.

Característica de la hipótesis	Puntaje de BEDROC
AANRR3	0,985
AANR1	0,981
AAANR2	0,980
AANRR2	0,979
AAANR1	0,979
ANRR1	0,979
AANRR1	0,974
AANR3	0,954
AANR2	0,954
ANRR2	0,939

Tabla 3: Resultados obtenidos del protocolo LBVS con hipótesis de farmacóforos basado en cumarinas.

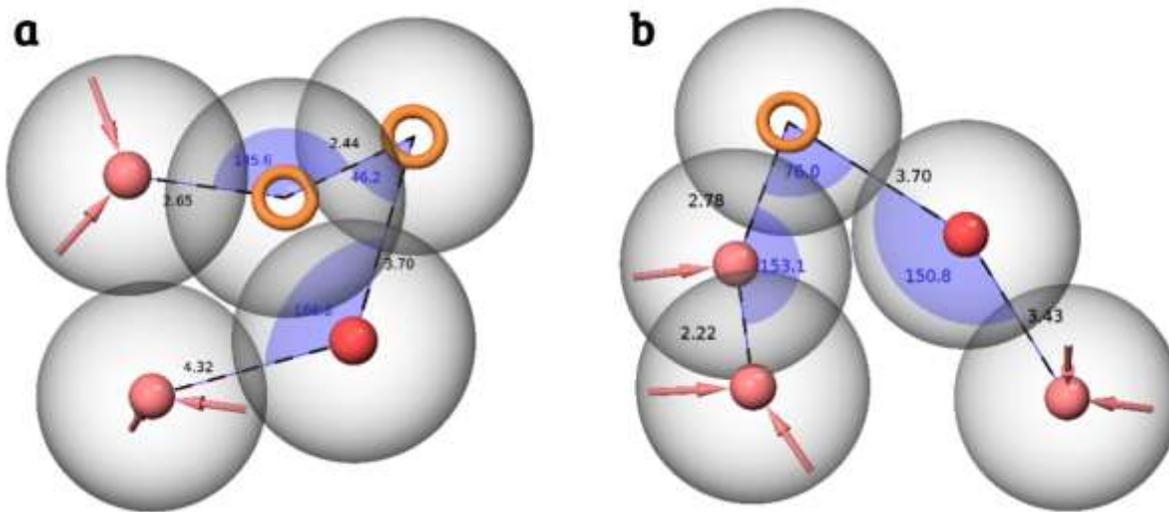


Figura 14: Hipótesis de farmacóforos seleccionados basados en cumarinas. a) AANRR y b) AAANR.

Con respecto a la realización de la hipótesis de farmacóforo, solo se utilizaron moléculas con núcleos cumarínicos, debido a que en ellas existe una diversidad química que está dada por 142 grupos radicales dentro de los que se encuentran cumarinas y bis-cumarinas. Cada una de estas hipótesis seleccionadas se encuentran construidos por distintas características, donde para la primera de ellas (Figura 14a) está dada por dos anillos aromáticos (anillo naranja), dos aceptores de enlaces de hidrógeno (esfera rojo suave) y un grupo con carga negativa (esfera roja) y para la segunda (Figura 14b) está definida por un anillo aromático, tres aceptores de enlaces de hidrógeno y un grupo con carga negativa.

Además mediante las dos hipótesis recientemente mencionadas se espera que el conjunto de moléculas de la base de datos seleccionadas genere interacciones con al menos uno de los iones de níquel y los residuos claves del sitio catalíticos: H136, H138, KCX219/LYN219, H221, H248, H274, C322, R338 y D362.

Luego al aplicar ambas hipótesis de farmacóforo a la biblioteca de compuestos que presentaba 4.903.299 moléculas, se encontraron 1.429.460 moléculas con una similitud a los farmacóforos empleados, lo que corresponde al 29% del total de moléculas alineadas. Con respecto al número de moléculas obtenidas a cada hipótesis de forma individual, las hipótesis AAANR y AANRR tienen 772.351 y 657.109 moléculas, respectivamente.

En la Figura 15, se puede apreciar que a cada molécula es asignado un puntaje “Phase Screen Score”, el cual corresponde a la similitud que existe entre la hipótesis farmacóforica y cada compuesto, donde los puntajes se distribuyen entre -0,45 y

2,00 (mientras más positivo el valor, más parecida es la molécula al farmacóforo). Por lo que se asigna un *cutoff* para seleccionar moléculas con un *Phase Screen Score* mayor a 0,65, debido a que se buscaba que estas moléculas tuvieran la mayor semejanza posible con las hipótesis de farmacóforos.

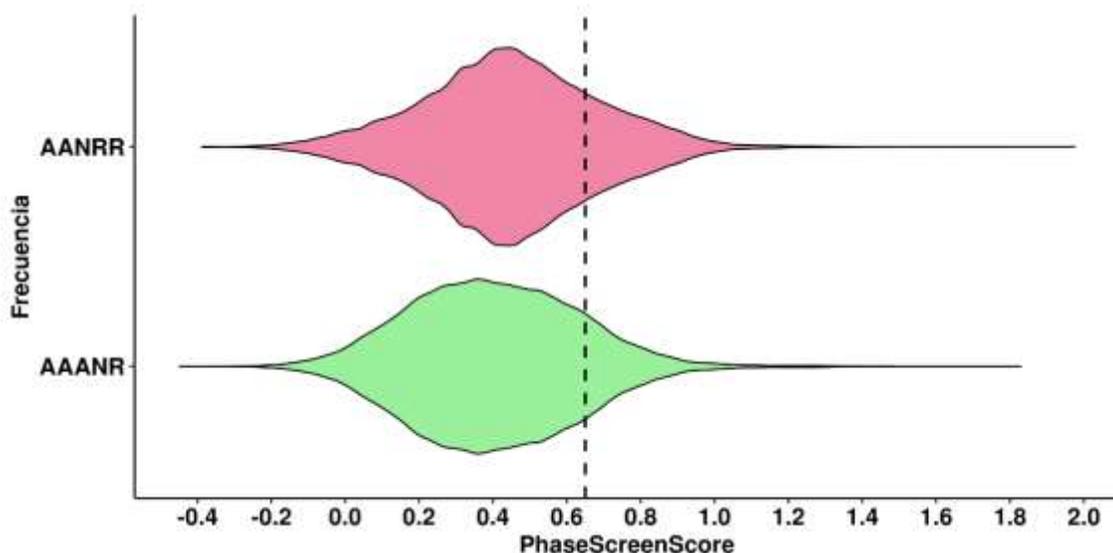


Figura 15: Resultado de LBVS basada en las hipótesis de cumarinas. Distribución de moléculas con respecto al *Phase Screen Score* y en línea punteada se observa el *cutoff* seleccionado (0,65).

En la Tabla 4, se puede observar cómo disminuye el número de moléculas de la biblioteca de compuestos para cada conjunto, donde para los farmacóforos AAANR y AANRR hay 107.930 y 126.677 moléculas con un *PhaseScreenScore* mayor al valor umbral empleado. El total de moléculas correspondiente a la nueva biblioteca, considerando ambos conjuntos es de 234.607 para aproximadamente el 16% de los compuestos evaluados en esta etapa (Figura 16).

LBVS basado en cumarinas	N° de moléculas inicial	<i>Cutoff</i>	N° de moléculas seleccionadas	Porcentaje
AANRR	657.109	> 0,65	126.677	19
AAANR	772.351		107.930	14

Tabla 4: Número de moléculas obtenidas por cada set de farmacóforos: AAANR y AANRR.

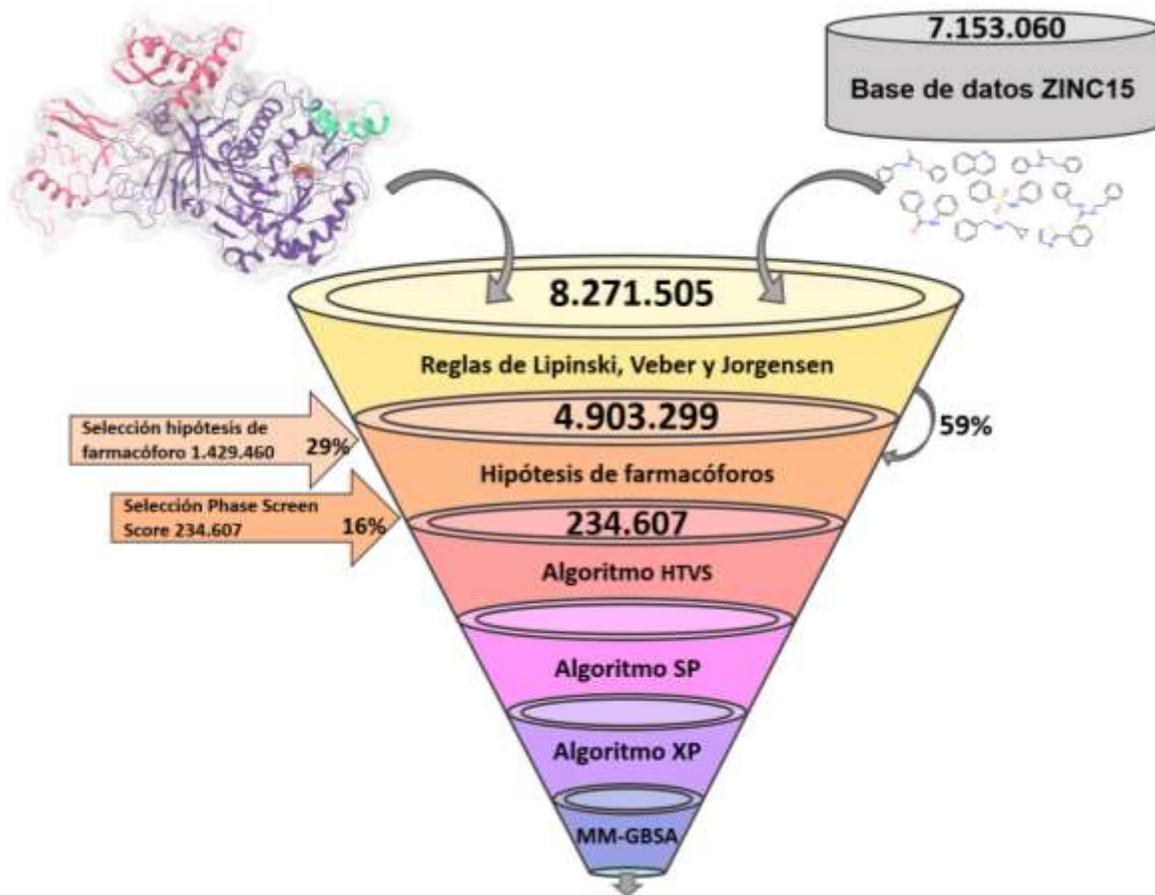


Figura 16: Flujo de trabajo. Número de moléculas reportadas en el filtro metodológico de “Hipótesis de farmacóforo” con LBVS basado en cumarinas.

4) Estudios de acoplamiento molecular a los compuestos candidatos, mediante diferentes algoritmos para reducir el número de candidatos a inhibidores de ureasa:

Como fue descrito en la Sección “Metodología”, en esta etapa y las posteriores, se utilizó la proteína en 2 estados: en presencia de una lisina 219 neutra (sistema aquí llamado LYN) y lisina 219 carbamilada (sistema aquí llamado KCX). Los resultados serán expuestos en el mismo orden, primero para la variante LYN y seguidamente KCX.

a) Docking molecular para la variante LYN

i) Algoritmo HTVS

Se utilizaron tres algoritmos distintos para *docking* molecular, donde el primero de ellos es HTVS. Se comenzaron los estudios con 234.607 moléculas obtenidas de LBVS basadas en cumarinas y mediante el algoritmo de HTVS, con el cual se predijo principalmente el sitio de unión y la interacción proteína-ligando, se obtuvo como resultado de este algoritmo 552.716 moléculas, ya que se generaron a los más 3 poses por cada molécula.

Como se observa en la Figura 17, la distribución de moléculas con respecto a la energía de docking se encuentra entre -10,9 kcal/mol y 10,1 kcal/mol, la mayor cantidad de moléculas se encuentra ubicada alrededor de los -4 kcal/mol.

Con respecto al número de moléculas seleccionadas, se basó en el número de moléculas obtenidas por cada cuartil, siempre considerando que el valor de energía entre más negativo corresponde a una mejor unión entre proteína-ligando, es por ello que la selección de moléculas se llevó a cabo considerando el primer cuartil correspondiente a moléculas con energía menor o igual -4,63. Escogiendo como *cutoff* todas las moléculas que presentaron una energía inferior a -4,5 kcal/mol seleccionando 156.814 moléculas.

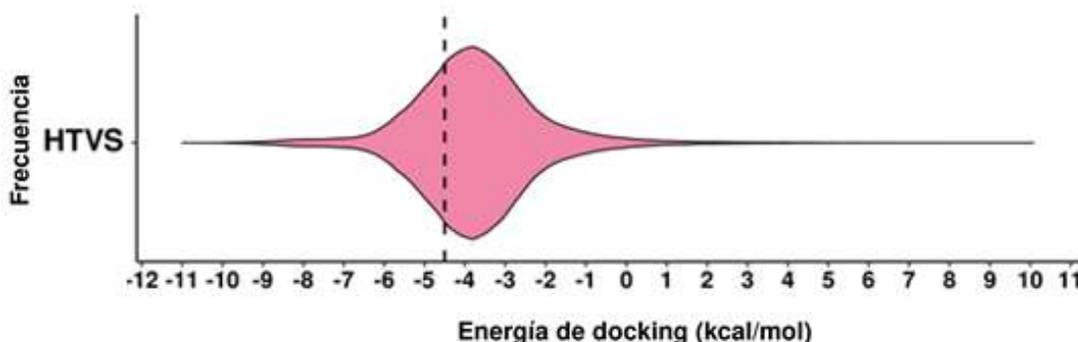


Figura 17: Distribución de moléculas con respecto a la energía de docking para HTVS en LYN.

Del conjunto de moléculas seleccionadas, hay una proporción que se encuentran fuera del sitio de unión (Figura 18), las que corresponden a 13.345 moléculas, quedando 143.469 moléculas finales después de este pre filtrado, lo cual corresponde al 26% de las 552.716 moléculas obtenidas mediante HTVS.

Todos estos compuestos que se encuentran ubicados fuera del sitio de unión fueron descartados, ya que no generan interacciones con los iones de níquel y residuos del sitio catalítico que han sido descritos en la literatura como residuos claves para la inhibición de ureasa.

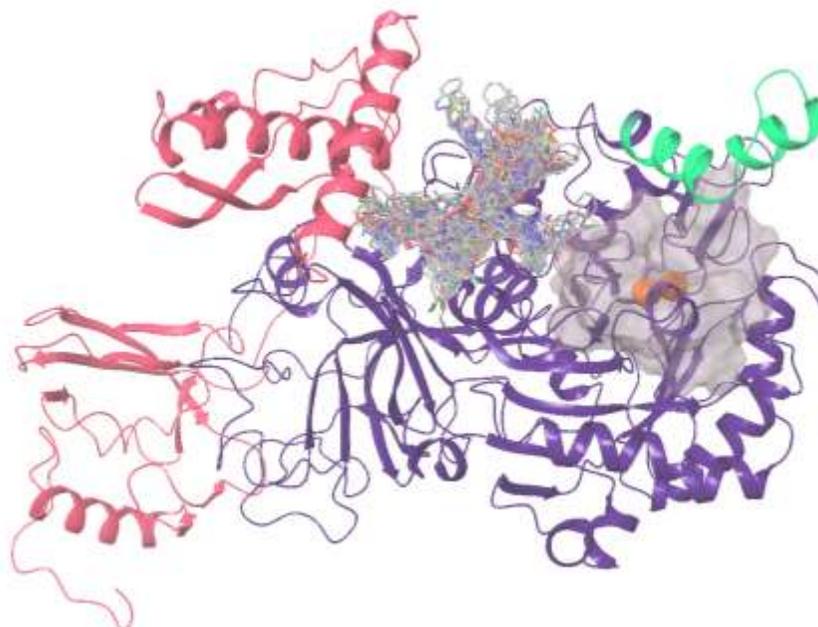


Figura 18: Conjunto de moléculas fuera del sitio de unión que fueron descartadas en el algoritmo de HTVS en LYN.

ii) Algoritmo SP

Se comenzaron los estudios con el algoritmo de precisión estándar (SP) con 143.469 moléculas obtenidas con el algoritmo HTVS, donde se obtuvo como resultado de este algoritmo 429.972 poses, ya que se generaron a lo más 3 poses por cada molécula.

Como se observa en la Figura 19, la distribución de moléculas con respecto a la energía de docking, se encuentra entre -11,3 kcal/mol y 4,1 kcal/mol, ubicándose la mayor cantidad de moléculas alrededor del valor -6 kcal/mol.

El número de moléculas seleccionadas, se basó en la estadística del número de moléculas obtenidas por cada cuartil, siempre considerando que el valor de energía más negativo corresponde a una mejor unión entre proteína-ligando. Por lo que se consideró el primer cuartil para la selección de moléculas que correspondieron a energías menores o iguales a -6,37 kcal/mol. El puntaje de corte seleccionado permitió seleccionar todas las moléculas con energías inferiores a -6,5 kcal/mol, seleccionándose así 93.829 moléculas.

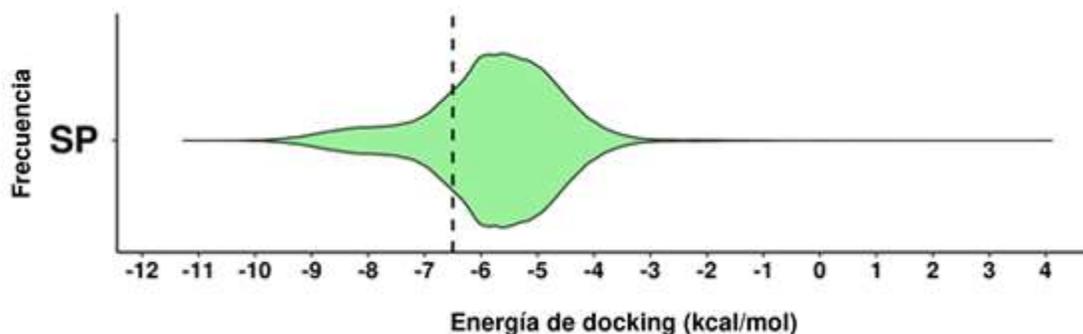


Figura 19: Distribución de moléculas con respecto a la energía de docking para SP en LYN.

Del conjunto de moléculas seleccionadas, una proporción se encuentran fuera del sitio de unión, siendo descartadas 67 moléculas, lo cual resultó en 93.762 moléculas finales a partir este filtrado para el 22% de las 429.972 poses obtenidas del resultado de SP.

iii) Algoritmo XP

Se comenzaron los estudios con el algoritmo “extra precisión”, o XP, como el tercer paso de acoplamiento molecular. Este es un método con mayor rigurosidad que HTVS y SP, dado que clasifica semi-cuantitativamente la capacidad de los ligandos candidatos para unirse a una conformación específica de la proteína. Se iniciaron los estudios con 93.762 moléculas obtenidas del algoritmo SP, donde se obtuvo como resultado de este algoritmo 267.641 poses ya que se generaron a los más 3 poses por cada una.

Como se observa en la Figura 20, la distribución de moléculas con respecto a la energía de docking se encuentra entre -14,0 kcal/mol y -6,4 kcal/mol, la mayor cantidad de moléculas se encuentra ubicada alrededor del valor -7 kcal/mol.

El número de moléculas seleccionadas estuvo basado en las estadísticas del número de moléculas obtenidas por cada cuartil, siempre considerando que el valor de energía más negativo corresponde a una mejor unión entre proteína-ligando. Por lo que la selección de moléculas se llevó a cabo considerando el primer cuartil correspondiente a moléculas con valores de energías menores o iguales a -7,32 kcal/mol. Además, se consideró dentro de los parámetros para la elección de *cutoff*, un número de moléculas bastante acotado ya que estos serían de importancia para el siguiente paso metodológico. Escogiendo como puntaje de corte todas las moléculas que tengan una energía inferior a -8,5 kcal/mol, seleccionando 11.368 moléculas.

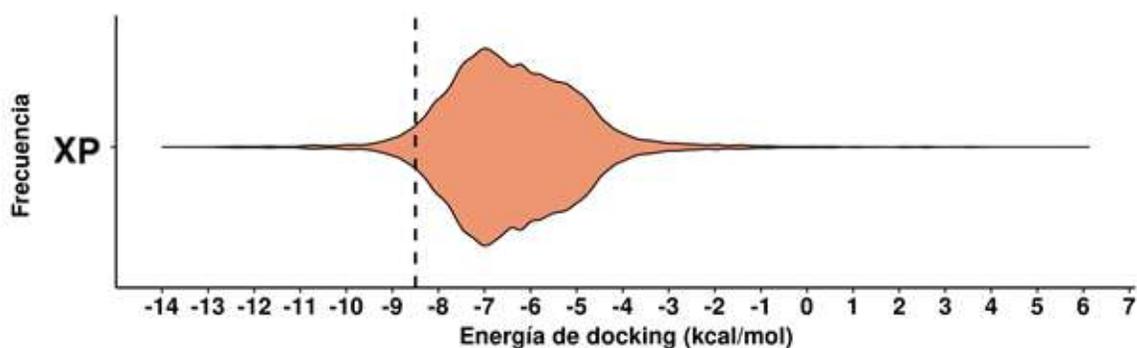


Figura 20: Distribución de moléculas con respecto a la energía de docking para XP en LYN.

En el conjunto de moléculas seleccionadas no existieron moléculas fuera del sitio, por lo que las 11.368 moléculas finales de este pre filtrado corresponden al 4% de las 267.641 moléculas que se obtuvieron del resultado de XP.

A continuación, en la Tabla 5, se encuentra información detallada con respecto al número de moléculas que se generaron por cada algoritmo utilizado y el número de moléculas finales seleccionadas por cada algoritmo de *docking* molecular.

LYN							
Algoritmo de docking	N° moléculas de entrada	N° moléculas de salida	Cutoff	N° moléculas seleccionadas	N° moléculas fuera del sitio	N° moléculas final	Porcentaje
HTVS	234.607	552.716	< -4,5	156.814	13.345	143.469	26
SP	143.469	429.972	< -6,5	93.829	67	93762	22
XP	93.762	267.641	< -8,5	11.368	0	11.368	4

Tabla 5: Número de moléculas obtenidas en el protocolo de docking molecular para LYN.

b) Docking molecular para la variante KCX

i) Algoritmo HTVS

Se utilizaron los mismos tres algoritmos para *docking* molecular empleados en la variante anterior, en donde el primero de ellos es HTVS. Se comenzaron los estudios en KCX con 234.607 moléculas al igual que en la variante LYN obtenidas de LBVS basadas en cumarinas. Mediante el algoritmo HTVS se obtuvo como resultado 600.588 poses, ya que se generaron a lo más 3 poses por cada una.

Como se observa en la Figura 21, la distribución de moléculas con respecto a la energía de docking molecular, se encuentra entre -10,9 kcal/mol y 11,5 kcal/mol, estando ubicada la mayor cantidad de moléculas alrededor del valor -4 kcal/mol.

Aquí el primer cuartil corresponde a -4,77 kcal/mol, por lo que se escogió como *cutoff* -4,5 kcal/mol seleccionando 196.349 moléculas.

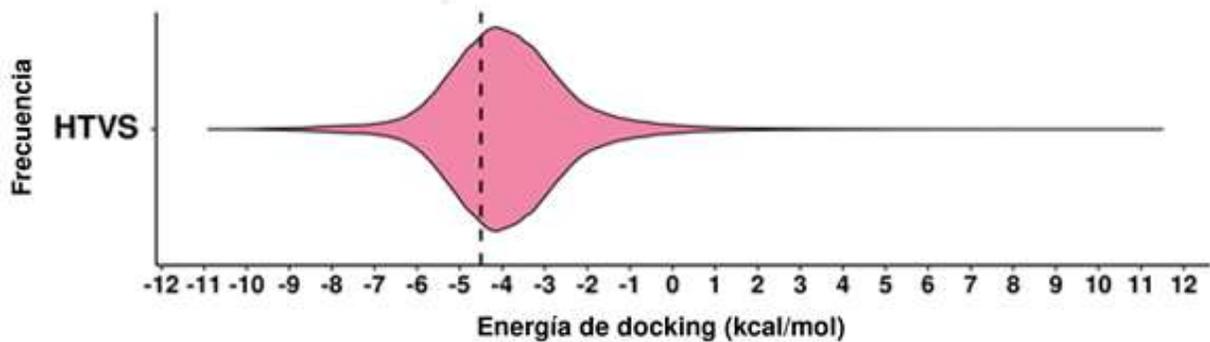


Figura 21: Distribución de moléculas con respecto a la energía de docking para HTVS en KCX.

Del conjunto de moléculas seleccionadas, una proporción de moléculas que se encuentran fuera del sitio de unión (Figura 22), las que corresponden a 9.901 moléculas, quedando 186.448 moléculas finales de este pre-filtrado, lo cual corresponde al 31% de las 600.588 moléculas que se obtuvieron del resultado de HTVS.

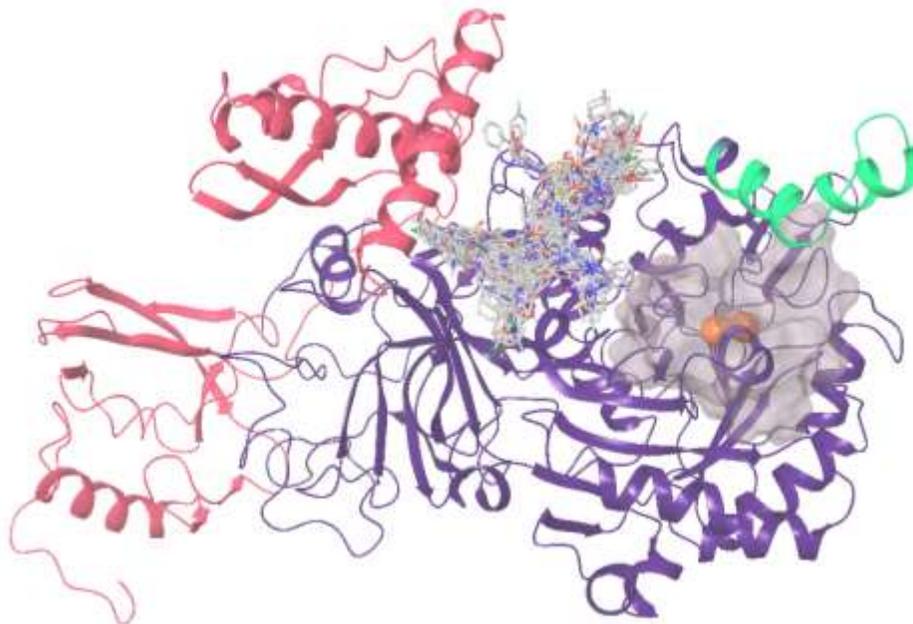


Figura 22: Conjunto de moléculas fuera del sitio de unión que fueron descartadas en el algoritmo de HTVS en la variante KCX.

ii) Algoritmo SP

Se comenzaron los estudios con el algoritmo de precisión estándar o SP con 186.448 moléculas obtenidas del algoritmo HTVS, identificándose 557.590 poses, dado que se generaron 3 poses, como máximo, por cada una.

Como se observa en la Figura 23, la distribución de moléculas con respecto a la energía de docking se encuentra entre -11,7 kcal/mol y 3,5 kcal/mol, la mayor cantidad de moléculas se encuentra ubicada alrededor de los -5.5 kcal/mol. Aquí el primer cuartil corresponde a -6,24 kcal/mol, de modo que se seleccionó un valor de corte de -6.5 kcal/mol, lo que permitió seleccionar 105.364 moléculas.

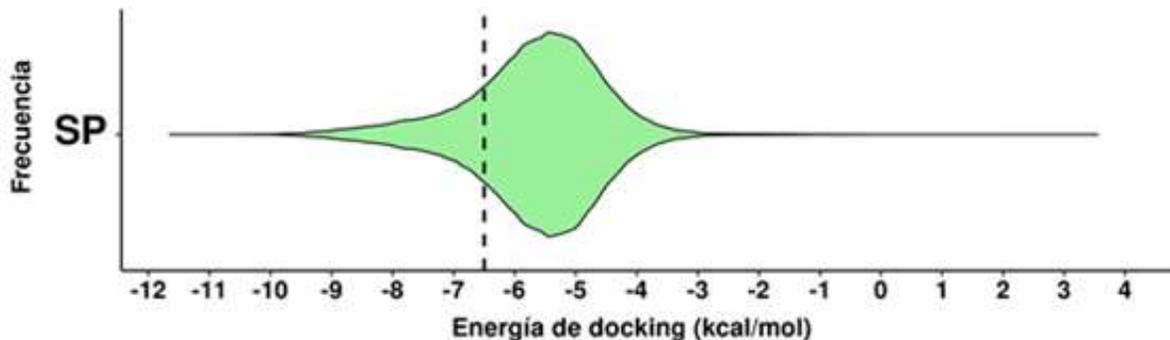


Figura 23: Distribución de moléculas con respecto a la energía de docking para SP en KCX

Del conjunto de moléculas seleccionadas, hay una proporción que se encuentran fuera del sitio de unión y que corresponden a 138 compuestos descartados, quedando 105.226 moléculas finales después de este pre-filtrado la que corresponde al 19% de las 557.590 moléculas que se obtuvieron del resultado de SP.

iii) Algoritmo XP

Se comenzaron los estudios con el algoritmo “extra precisión” o XP como el tercer método de *docking* molecular utilizado, con 105.226 moléculas obtenidas del algoritmo SP, donde se obtuvo como resultado de este algoritmo, 297.494 moléculas, ya que se generaron a los más 3 poses por cada una. Las energías de docking en esta etapa oscilan entre -13,8 kcal/mol y -6,4 kcal/mol, la mayor cantidad de moléculas se encuentra ubicada alrededor de los -7 kcal/mol, cuyo primer cuartil finaliza

en -7,37 kcal/mol. Por lo que el puntaje de corte empleado fue -8,5 kcal/mol, seleccionando 16.160 moléculas (Figura 24).

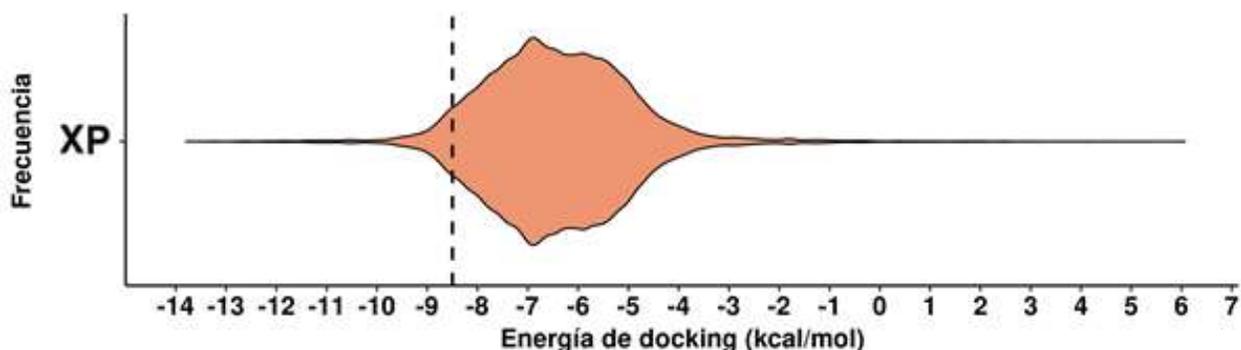


Figura 24: Distribución de moléculas con respecto a la energía de docking molecular para XP en KCX.

En el conjunto de moléculas seleccionadas no existieron moléculas fuera del sitio, por lo que las 16.160 moléculas finales de este filtrado corresponden al 5% de las 297.494 moléculas que se obtuvieron del resultado de XP.

En la Tabla 6, se encuentra detallada la información con respecto a número de moléculas que se generaron por cada algoritmo utilizado y el número de moléculas finales seleccionadas por cada procedimiento de *docking* molecular en la variante KCX.

KCX							
Algoritmo de docking	N° moléculas de entrada	N° moléculas de salida	Cutoff	N° moléculas seleccionadas	N° moléculas fuera del sitio	N° moléculas final	Porcentaje
HTVS	234.607	600.588	< -4,5	196.349	9901	186.448	31
SP	186.448	557.590	< -6,5	105.364	138	105.226	19
XP	105.226	297.494	< -8,5	16.160	0	16.160	5

Tabla 6: Número de moléculas obtenidas en el protocolo de docking molecular para KCX.

En la Figura 25, se puede observar el número de moléculas obtenidas por cada conjunto de moléculas para las variantes LYN y KCX hasta el filtrado de *docking* molecular con algoritmo XP. Donde, además, se observa que el número de moléculas corresponde a la variante KCX es mayor que para la variante LYN, lo que genera ciertos indicios que la lisina carbamilada ayuda a la coordinación de los iones de níquel y las moléculas precandidatas. Por otro lado, es destacable que el número de moléculas con valores de energía de interacción intermolecular no favorables (valores positivos) es mayor para el método HTVS, seguido por SP y finalmente XP, lo cual advierte la rigurosidad de los resultados que entrega este último algoritmo, penalizando aquellas moléculas que no se unen favorablemente en los sitios de unión de HpU.

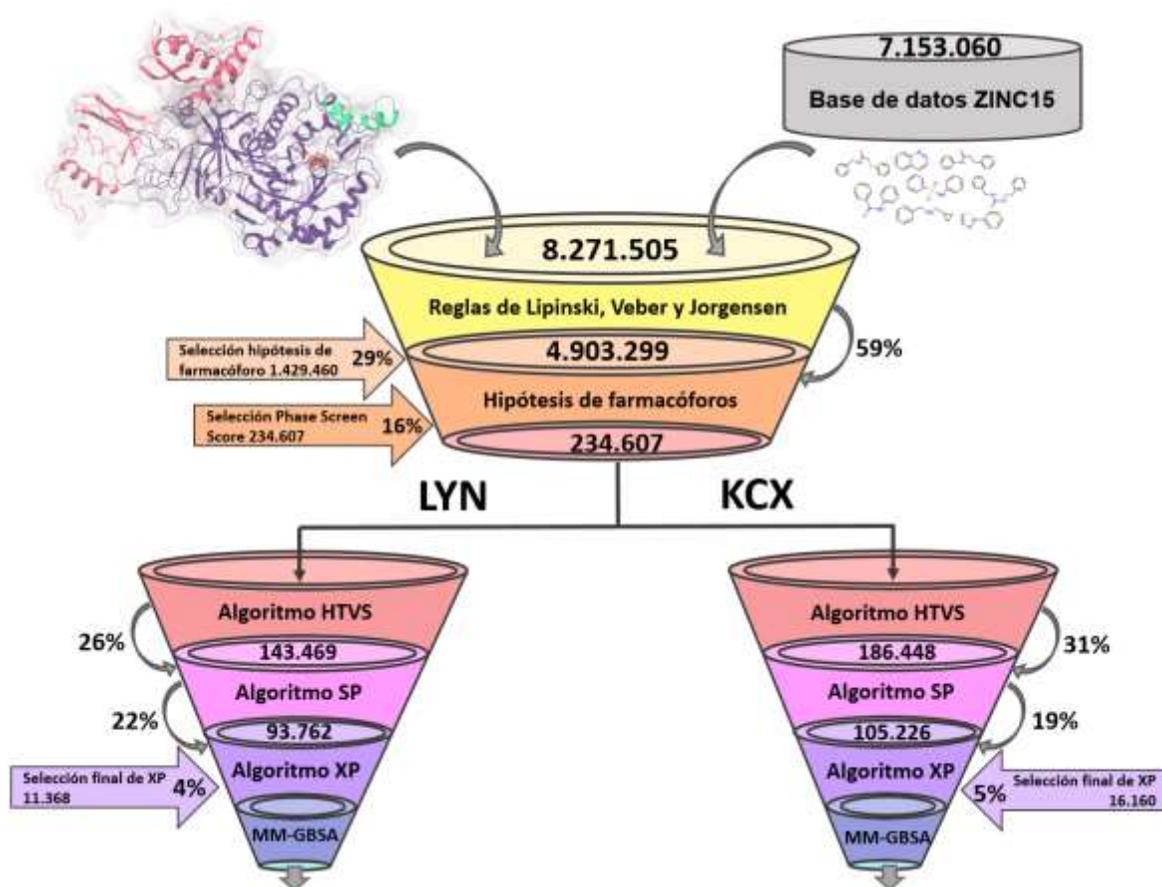


Figura 25: Número de moléculas reportadas en el filtro metodológico de "Acoplamiento molecular".

Ya terminados los estudios de *docking* molecular, se tienen 11.368 moléculas para la variante LYN y 16.160 para la variante KCX, en donde para cada una de ellas se realizaron diferentes filtrados.

El primer filtrado correspondió a la eliminación de las moléculas que presentaban el mismo nombre, de las cuales se eliminaron 10.031 moléculas para la variante LYN y 14.551 para la variante KCX. Estos compuestos se eliminaron debido a que, por cada procedimiento de *docking* molecular para cada algoritmo, se generaban hasta 3 poses. Por lo que el nombre de cada molécula podría estar repetido hasta 27 veces. No obstante, fue necesario evaluar todas las posibles variaciones (rotacionales, traslacionales y/o tautoméricas) en HTVS, SP y XP hasta encontrar aquella con la energía más favorable.

A los compuestos que se obtuvieron del filtrado recién mencionado, se les aplicó un nuevo filtro mediante el cual fueron descartadas las moléculas que se encuentran a una distancia mayor a 3,5 Å de alguno de los dos iones de níquel, quedando 1.188 moléculas para la variante LYN y 1.471 moléculas para la variante KCX. Cabe mencionar que los iones de níquel en la enzima ureasa han sido descritos como fundamentales para el proceso de hidrólisis. Asimismo, la presencia de inhibidores con una acción quelante es una de las características más deseadas en este campo de investigación [72], es por ello que es necesario retener aquellas moléculas que se encuentran a una distancia cercana (menor a 3,5 Å) de alguno de los iones de níquel.

Y como último filtro, se evaluó entre moléculas que presentaban un RMSD igual a cero, donde si bien no tenían el mismo nombre, eran idénticas estructuralmente, ya que están correspondían a tautómeros y de esta forma son coincidentes, resultando en 461 moléculas, lo cual corresponde a un 4% de las 11.368 para la variante LYN y 479 moléculas correspondientes al 3% de las 16.160 analizadas para la variante KCX (Tabla 7).

Tipo	N° moléculas de entrada	N° moléculas sin nombre repetido	N° moléculas con distancia $\leq 3,5$ Å de Ni	N° moléculas RMSD > 0	Porcentaje
LYN	11.368	1.337	1.188	461	4
KCX	16.160	1.609	1.471	479	3

Tabla 7: Número de moléculas post filtrado de docking molecular

En la Figura 26, se puede observar el número de moléculas seleccionadas por cada conjunto de moléculas para LYN y KCX considerando los nuevos filtros aplicados. Donde el conjunto de moléculas para la variante KCX sigue siendo mayor que el número de moléculas para la variante LYN.

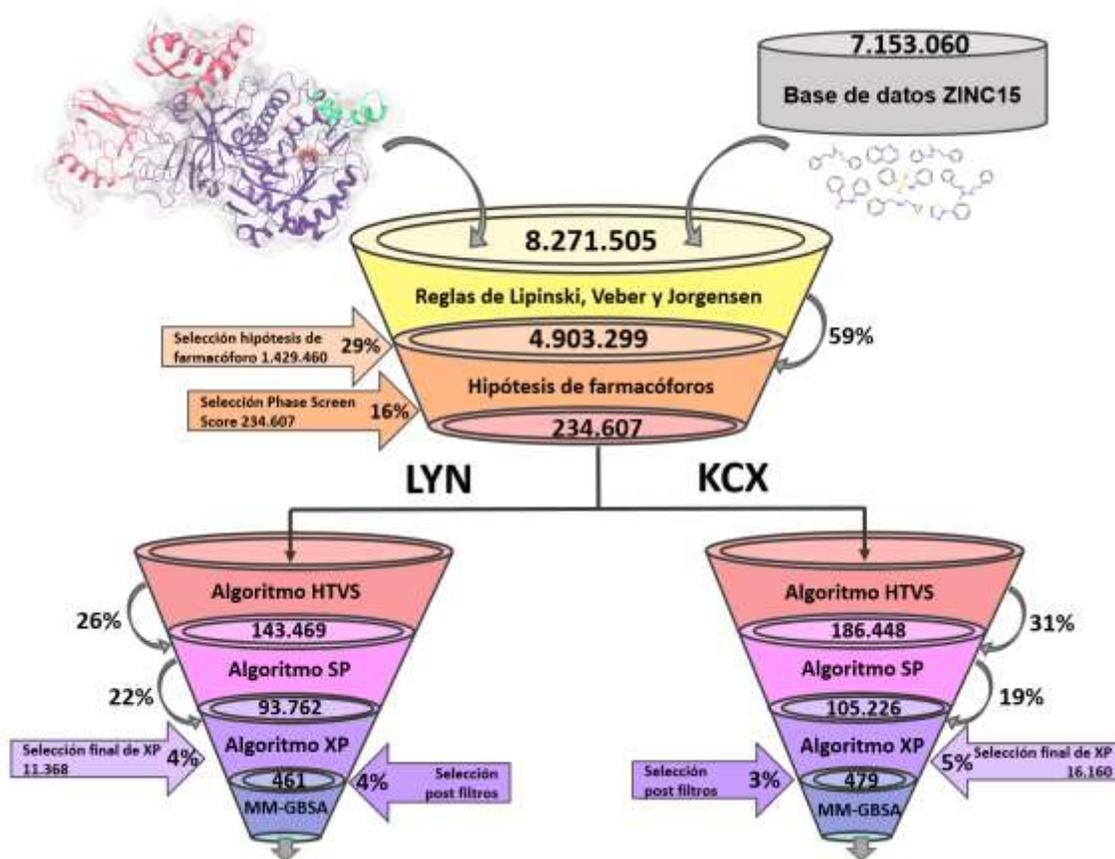


Figura 26: Flujo de trabajo. Número de moléculas reportadas post filtro metodológico de "Acoplamiento molecular".

5) Cálculos de energía libre de unión mediante el método MM-GBSA:

Se realizaron cálculos energéticos para cada conjunto de moléculas obtenidos del filtrado anterior tanto para LYN como KCX, en donde todos los átomos de la proteína que se localizaban a 5 Å del ligando en todas direcciones, se refinaron como una región flexible. En el caso de LYN, las energías de MM-GBSA oscilan entre -55,61 y 88,22 kcal/mol, mientras que en KCX oscilan entre -101,10 y 57,44 kcal/mol. Donde claramente se observa que las energías generadas por la variante KCX son más favorables a las obtenidas para la variante LYN.

Luego, se generaron 3 conjuntos de moléculas, el primero de ellos corresponde a moléculas que superando todos los filtros aplicados, se encuentran interactuando tanto con LYN (LYN_Ambos) como con KCX (KCX_Ambos). Las energías de interacción intermolecular de estas moléculas con la variante KCX se distribuyen entre -101,10 y 50,41 kcal/mol, y con la variante LYN oscilan entre -52,16 y 74,31 kcal/mol, el siguiente conjunto de moléculas corresponde solo a moléculas que se encuentran en la variante KCX siendo 197 moléculas y sus energías varían entre -94 y 57,44 kcal/mol y por último las que se encuentran solo en la variante LYN siendo 179 moléculas y sus energías fluctúan entre -55,61 y 85,22 kcal/mol.

Considerando que los valores energéticos positivos son considerados desfavorables, se procedió a descartar todas aquellas moléculas con energías de unión positivas para los 3 grupos, donde para las moléculas que interactúan con LYN y KCX a la vez, se reduce a 139 moléculas, en LYN se reduce a 81 moléculas y KCX se reduce a 181 moléculas (Tabla 8).

Conjunto de moléculas	N° moléculas iniciales	N° moléculas con energía negativa	Porcentaje
LYN ambos	282	139	49
KCX ambos	282	139	49
LYN	179	81	45
KCX	197	181	92

Tabla 8: Número de moléculas preseleccionadas.

Como fue mencionado en la sección “Metodología”, se realizaron cálculos energéticos con el método MM-GBSA para el compuesto DJM, el cual es un inhibidor co-cristalizado con HpU, pero a su vez también se utilizaron los inhibidores HAE (ácido acetohidroxámico) y BME (beta-mercaptoetanol), además del sustrato natural de la enzima: urea. Dichos compuestos se utilizaron como control; y sus energías de interacción con HpU (LYN y KCX) se muestran en la Tabla 9. Del mismo modo, cabe mencionar que estas moléculas utilizadas de control presentan una actividad biológica de 0,02 mM para DJM, 0,04 mM para HAE y 13,5 mM para BME [53], lo que es concordante con los valores energéticos obtenidos.

Compuestos	Proteína con LYN 219	Proteína con KCX 219
DJM	-60,4 kcal/mol	-64,3 kcal/mol
HAE	-15,6 kcal/mol	-17,4 kcal/mol
BME	-11,5 kcal/mol	-10,8 kcal/mol
Urea	-17,6 kcal/mol	-39,6 kcal/mol

Tabla 9: Energías de ΔG_{bind} para los compuestos DJM, HAE, BME y urea con proteína en 2 estados LYN y KCX.

Se puede observar en la Figura 27 como se ubican las energías de los compuestos control en cada conjunto de moléculas.

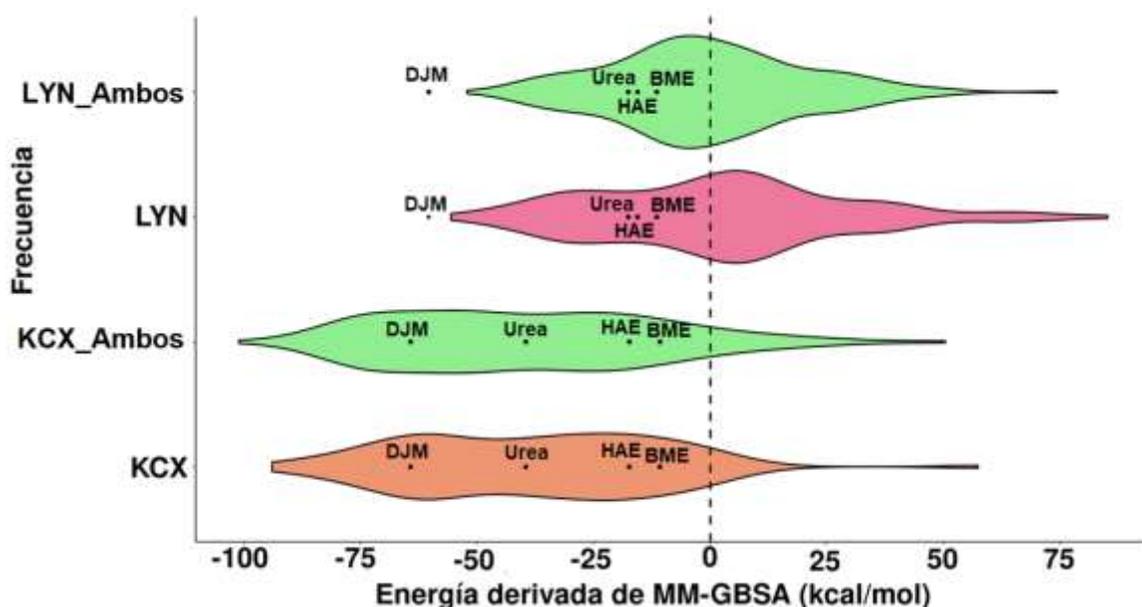


Figura 27: Energía de MM-GBSA para cada conjunto de moléculas y compuestos correspondientes al control de calidad interno.

Como se observa en la Figura 27, DJM que es el inhibidor co-cristalizado que presenta la estructura 3D aquí empleada (PDB id: 6zja), es el control que presenta la mejor energía de interacción con la proteína en cualquiera de sus variantes. De modo que la energía de afinidad de DMJ fue usada como valor de corte en cada grupo. Para los grupos de compuestos de KCX y KCX_Ambos hay 39 y 37

moléculas con mejor energía de afinidad que DMJ. Por otro lado, DMJ presenta una energía más favorable que cualquiera de las moléculas interactuando con LYN.

Para la molécula de urea, que es el sustrato natural de ureasa, nuestra metodología reporta una gran cantidad de compuestos que tienen mejor energía de unión en los 3 conjuntos de moléculas, siendo en LYN_Ambos 50 moléculas, LYN 46 moléculas, KCX_Ambos 80 moléculas y KCX 90 moléculas. Y en el caso de HAE y BME también existe un número de moléculas considerables que presentan mejor energía de unión proteína-ligando (Tabla 10).

Compuestos	LYN_Ambos	KCX_Ambos	LYN	KCX
DJM	0	39	0	37
HAE	50	80	46	90
BME	55	121	50	145
Urea	76	127	75	158

Tabla 10: *Números de moléculas con energía igual o más favorable a la de cada compuesto control.*

Es por ello que se escogieron como moléculas precandidatas, las mejores 40 moléculas encontradas en ambos grupos a la vez (Tabla 11) y aquellas que solo interactúan con KCX (Tabla 12). Por lo que en esta capa de VS quedan descartadas todas las moléculas que sólo interactúan con la variante LYN, y se puede comenzar a hipotetizar que la lisina carbamilada (KCX219) genera una mejor coordinación en el sitio de unión con las moléculas aquí reportadas. Dado que el objetivo final de esta tesis de grado es obtener inhibidores de HpU que permitan tanto mitigar la infección producida por *H. pylori* en humanos como detener la volatilización de fertilizantes del suelo producido por ureasas. Es por ello que las características farmacocinéticas, o dicho de otro modo, la interacción que estas moléculas encontradas puedan tener con el organismo humano son un factor clave a la hora de llevar estas moléculas a estudios posteriores. En consecuencia se evaluaron mediante diferentes métodos teóricos distintas características farmacocinéticas y fisicoquímicas con la intención de conocer si cumplían con la mayor cantidad de criterios ADME. En las Tablas 11 y 12 se muestran un conjunto de características o reglas ampliamente empleadas en la literatura para evaluar diferentes parámetros que aseguren características ADME deseables dentro de las que se encuentran Filtro de Lipinski: peso molecular ≤ 500 gr/mol, coeficiente de partición $\leq 4,15$, aceptores de enlaces de hidrógeno ≤ 10 y donadores de enlaces de hidrógeno \leq

5. Filtro de Ghose: peso molecular entre 160 y 480 gr/mol, coeficiente de partición entre -0,4 y 5,6, recuento de átomos entre 20 y 70 y refractividad molar entre 40 y 130. Filtro de Veber: número de enlaces rotables ≤ 10 , área de superficie polar $\leq 140 \text{ \AA}$ y total de enlaces de hidrógeno < 12 (donadores y aceptores). Filtro de Egan: coeficiente de partición $\leq 5,88$ y área de superficie polar $\leq 131,6$. Filtro de Muegge: peso molecular entre 200 y 600 gr/mol, coeficiente de partición entre -2 y 5, área de superficie polar ≤ 150 , número de anillos ≤ 7 , número de carbonos > 4 , número de heteroátomos > 1 , número de enlaces rotables ≤ 15 , sceptores de entonces de hidrógeno ≤ 10 y donadores de enlaces de hidrógeno ≤ 5 . Leadlikeness: peso molecular entre 250 y 350 gr/mol, coeficiente de partición $\leq 3,5$ y número de enlaces rotables ≤ 7 . Además de la alerta estructural BRENK y PAINS. Se aprecia en dichas tablas el número de violaciones para cada regla o característica evaluada, siendo aquellas moléculas que presenten el menor número de violaciones (última columna) aquellas más idóneas.

N°	Nombre	Energía MM-GBSA	ROF (4)	ROT (3)	Ghose (4)	Egan (2)	Veber (3)	Muegge (9)	PAINS	BRENK	Leadlikeness (3)	Suma del N° de violaciones
1	ZINC000008627328	-101,10	0	0	0	0	0	0	0	0	0	0
2	ZINC000002522532	-83,13	0	0	0	0	0	0	0	0	0	0
3	ZINC000031169481	-81,98	0	0	0	0	0	0	0	0	0	0
4	ZINC000091624618	-71,38	0	0	0	0	0	0	0	0	0	0
5	ZINC000007994969	-71,34	0	0	0	0	0	0	0	0	0	0
6	ZINC000008731862	-69,53	0	0	0	0	0	0	0	0	0	0
7	ZINC000002584177	-68,47	0	0	0	0	0	0	0	0	0	0
8	ZINC000100437608	-81,85	0	0	0	0	0	0	0	0	1	1
9	ZINC000071769481	-81,57	0	0	0	0	0	0	0	0	1	1
10	ZINC000002120368	-80,61	0	0	0	0	0	0	0	0	1	1
11	ZINC000004858708	-80,49	0	0	1	0	0	0	0	0	0	1
12	ZINC000089286421	-79,39	0	0	1	0	0	0	0	0	0	1
13	ZINC000002479300	-78,92	0	0	0	0	0	0	0	0	1	1
14	ZINC000013545426	-77,84	0	0	0	0	0	0	0	0	1	1
15	ZINC000001117084	-74,25	0	0	0	0	0	0	0	0	1	1
16	ZINC000004320287	-74,00	0	0	0	0	0	0	0	0	1	1
17	ZINC000002479297	-73,70	0	1	0	0	0	0	0	0	0	1
18	ZINC000095095717	-72,63	0	0	0	0	0	0	0	0	1	1

19	ZINC000004839614	-71,80	0	0	0	0	0	0	0	0	1	1
20	ZINC000051325904	-71,74	0	0	0	0	0	0	0	0	1	1
21	ZINC000004320258	-69,54	0	0	0	0	0	0	0	0	1	1
22	ZINC000012360390	-68,79	0	0	0	0	0	0	0	0	1	1
23	ZINC000009668624	-67,80	0	0	0	0	0	0	0	0	1	1
24	ZINC000040286706	-67,67	0	1	0	0	0	0	0	0	0	1
25	ZINC000089991200	-77,77	0	0	0	0	0	0	0	1	1	2
26	ZINC000199798219	-77,32	0	0	0	0	0	0	0	1	1	2
27	ZINC000095532132	-76,76	0	0	0	0	0	0	0	1	1	2
28	ZINC000033795588	-70,50	0	0	0	0	0	0	0	1	1	2
29	ZINC000262738161	-69,95	0	1	1	0	0	0	0	0	0	2
30	ZINC000222498800	-68,01	0	0	0	0	0	0	0	0	2	2
21	ZINC000222659948	-67,20	0	0	0	0	0	0	0	0	2	2
32	ZINC000006760357	-62,82	0	0	1	0	0	0	0	0	1	2
33	ZINC000019798822	-61,49	0	1	0	0	0	0	0	0	1	2
34	ZINC000257233422	-72,34	0	0	0	0	0	0	0	0	3	3
35	ZINC000001038956	-71,18	0	0	0	1	0	1	0	0	1	3
36	ZINC000005000999	-70,80	1	0	0	1	1	0	0	0	0	3
37	ZINC000012706865	-70,12	0	0	0	0	0	0	0	0	3	3
38	ZINC000004386690	-76,48	1	0	0	1	0	1	0	0	1	4
39	ZINC000001000353	-75,87	0	1	0	0	0	0	0	1	2	4
40	ZINC000004809091	-73,29	0	0	1	1	1	1	0	1	0	5

Tabla 11: *Números de moléculas seleccionadas como precandidatas en el conjunto de moléculas que interactúan con ambas variantes.*

N°	Nombre	Energía MM-GBSA	ROF (4)	ROT (3)	Ghose (4)	Egan (2)	Veber (3)	Muegge (9)	PAINS	BRENK	Leadlikeness (3)	N° de violaciones
1	ZINC000072412074	-77,99	0	0	0	0	0	0	0	0	0	0
2	ZINC000072427629	-69,30	0	0	0	0	0	0	0	0	0	0
3	ZINC000006751398	-68,42	0	0	0	0	0	0	0	0	0	0
4	ZINC000215300901	-68,10	0	0	0	0	0	0	0	0	0	0

5	ZINC000003316814	-66,53	0	0	0	0	0	0	0	0	0	0
6	ZINC000008731861	-64,53	0	0	0	0	0	0	0	0	0	0
7	ZINC000008700584	-63,97	0	0	0	0	0	0	0	0	0	0
8	ZINC001560403828	-63,42	0	0	0	0	0	0	0	0	0	0
9	ZINC000005810608	-94,00	0	0	0	0	0	0	0	0	1	1
10	ZINC000015106043	-79,70	0	0	0	0	0	0	0	0	1	1
11	ZINC000216429060	-78,71	0	0	0	0	0	0	0	0	1	1
12	ZINC000012360391	-73,06	0	0	0	0	0	0	0	0	1	1
13	ZINC000003298752	-72,43	0	0	0	0	0	0	0	0	1	1
14	ZINC000219014368	-71,60	0	0	0	0	0	0	0	0	1	1
15	ZINC000095387364	-69,38	0	0	0	0	0	0	0	0	1	1
16	ZINC000011867201	-68,88	0	0	0	0	0	0	0	0	1	1
17	ZINC000072423992	-67,64	0	0	1	0	0	0	0	0	0	1
18	ZINC000072409234	-65,86	0	0	1	0	0	0	0	0	0	1
19	ZINC000000614859	-65,42	0	0	0	0	0	0	0	1	0	1
20	ZINC000003863640	-64,06	0	0	0	0	0	0	0	1	0	1
21	ZINC000080316223	-89,34	0	0	0	0	0	0	1	0	1	2
22	ZINC000022289134	-88,14	0	0	0	1	0	0	0	0	1	2
23	ZINC000080316289	-86,90	0	0	0	0	0	0	1	0	1	2
24	ZINC000032934908	-83,13	0	0	0	0	0	0	0	1	1	2
25	ZINC000000204307	-82,53	0	0	0	1	0	0	0	0	1	2
26	ZINC000004602529	-81,16	0	0	0	0	0	0	0	0	2	2
27	ZINC000089991199	-71,34	0	0	0	0	0	0	0	1	1	2
28	ZINC000012360392	-69,87	0	1	0	0	0	0	0	0	1	2
29	ZINC000023115099	-68,47	0	0	0	0	0	0	0	0	2	2
30	ZINC000004387901	-63,06	0	0	0	0	0	0	0	2	0	2
21	ZINC000006758619	-93,55	0	1	1	0	0	0	0	0	1	3
32	ZINC000001418849	-74,68	0	1	1	0	0	0	0	0	1	3
33	ZINC000005710763	-68,77	0	0	0	0	0	0	0	1	2	3
34	ZINC000257314628	-81,48	0	0	0	0	1	0	0	0	3	4
35	ZINC000001155565	-76,50	0	0	0	1	1	0	0	0	2	4

36	ZINC000017080370	-75,58	0	0	0	1	0	0	0	0	3	4
37	ZINC000252693382	-75,47	0	0	0	0	0	0	0	2	2	4
38	ZINC000020501865	-72,05	0	1	0	0	0	0	0	1	2	4
39	ZINC000009060557	-67,67	0	0	0	1	1	0	0	0	2	4
40	ZINC000008892723	-66,44	0	0	0	1	1	0	0	1	2	5

Tabla 12: *Números de moléculas seleccionadas como precandidatas en el conjunto que solo interactúa con KCX.*

Como siguiente paso, se escogieron las 10 moléculas candidatas de cada conjunto de moléculas representado en la Figura 28, considerando aquellas que infringieron 0 o 1 violación a todas las reglas evaluadas. Finalmente, para cada grupo de moléculas candidatas, se evaluó la interacción proteína-ligando a una distancia de 4.0 Å. En donde se reportan los residuos proteicos que interactúan con cada molécula y el tipo de interacción encontrado.

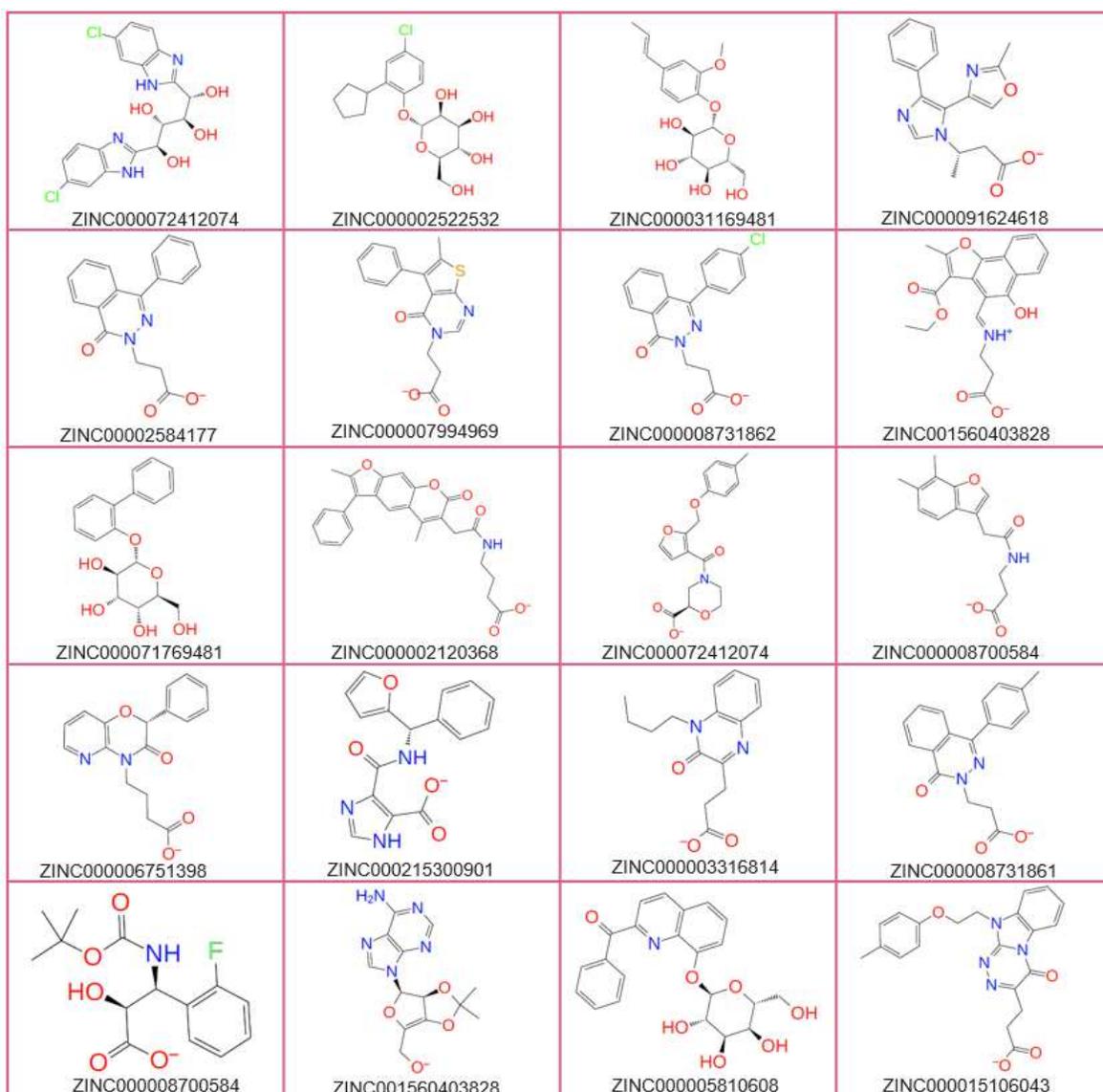


Figura 28: Moléculas seleccionadas como candidatas.

En la Figura 29 se observan 13 moléculas candidatas con respecto a la hipótesis de farmacóforo AANRR3 donde todas presentan al igual que el farmacóforo: dos anillos aromáticos, un grupo con carga negativa y gran parte de ellas uno o dos aceptores de enlaces de hidrógeno. Las interacciones que se presentan son: enlaces de hidrógeno (línea punteada azul), enlace de hidrógeno con un átomo de anillo aromático (línea punteada naranja) y puentes salinos (línea punteada magenta). La mayoría de las interacciones se encuentran asociadas al residuo H221 y R338. A su vez existen interacciones con: A169, K219, D223, H248,

G279, M317, H322 y D362, además se encuentran a distancia inferior a 3 Å de los iones de Níquel.

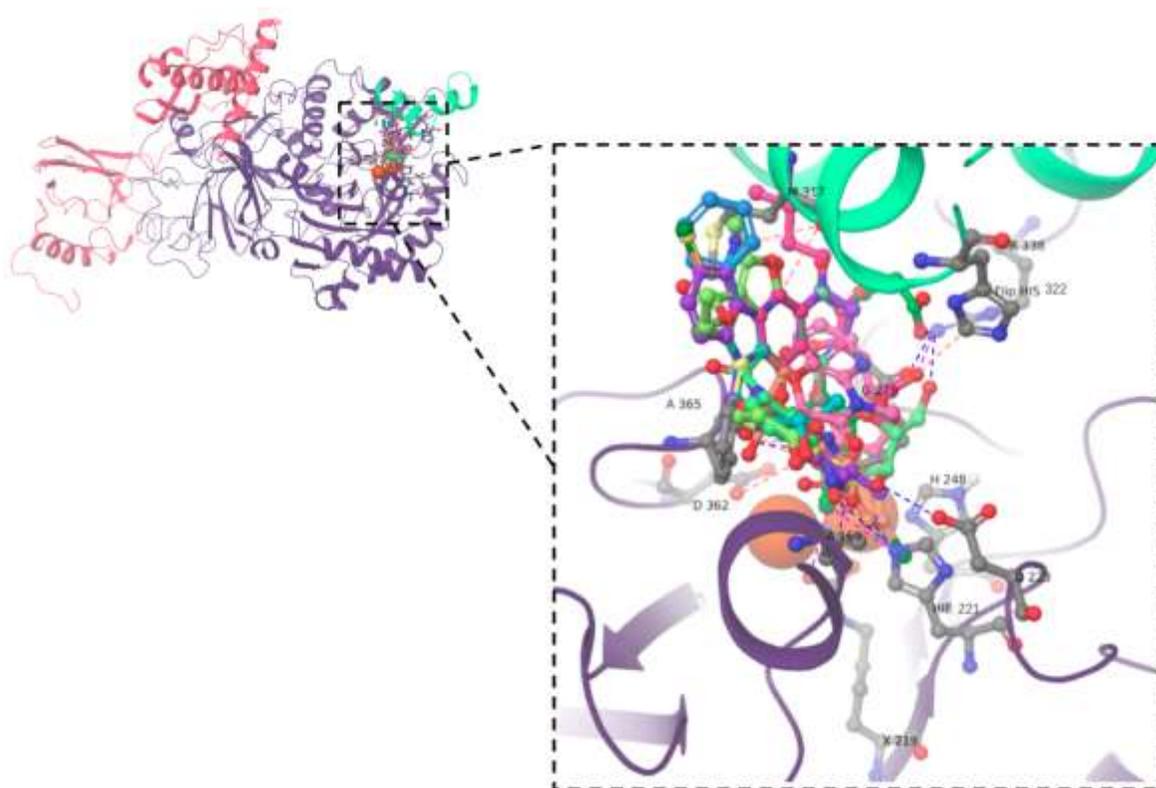


Figura 29: Interacción de 13 moléculas candidatas con respecto a la hipótesis de farmacóforo AANRR3. Moléculas candidatas: ZINC000003316814 (magenta), ZINC000005810608 (verde claro), ZINC000006751398 (rosa claro), ZINC000008731861 (verde), ZINC000015106043 (morado), ZINC000072427629 (amarillo), ZINC001560403828 (cian), ZINC000002120368 (gris), ZINC00002584177 (azul), ZINC000007994969 (azulino), ZINC000008627328 (naranja claro), ZINC000008731862 (naranja) y ZINC000005810608 (verde limón), iones de níquel (naranja) y residuos claves (gris).

En la Figura 30 se observan 7 moléculas candidatas con respecto a la hipótesis de farmacóforo AAANR2 donde todas presentan al igual que el farmacóforo: un anillo aromático, un grupo con carga negativa y gran parte de ellas dos a tres aceptores de enlaces de hidrógeno. Las interacciones que se presentan son: enlaces de hidrógeno (línea punteada azul), enlace de hidrógeno con un átomo de anillo aromático (línea punteada naranja) e interacciones π - π Staking (línea punteada gris). La mayoría de las interacciones se encuentran asociadas al residuo

H221, H322 y R338. A su vez existen interacciones con: H248, D362, A365, además se encuentran a distancia inferior a 3 Å de los iones de Níquel.

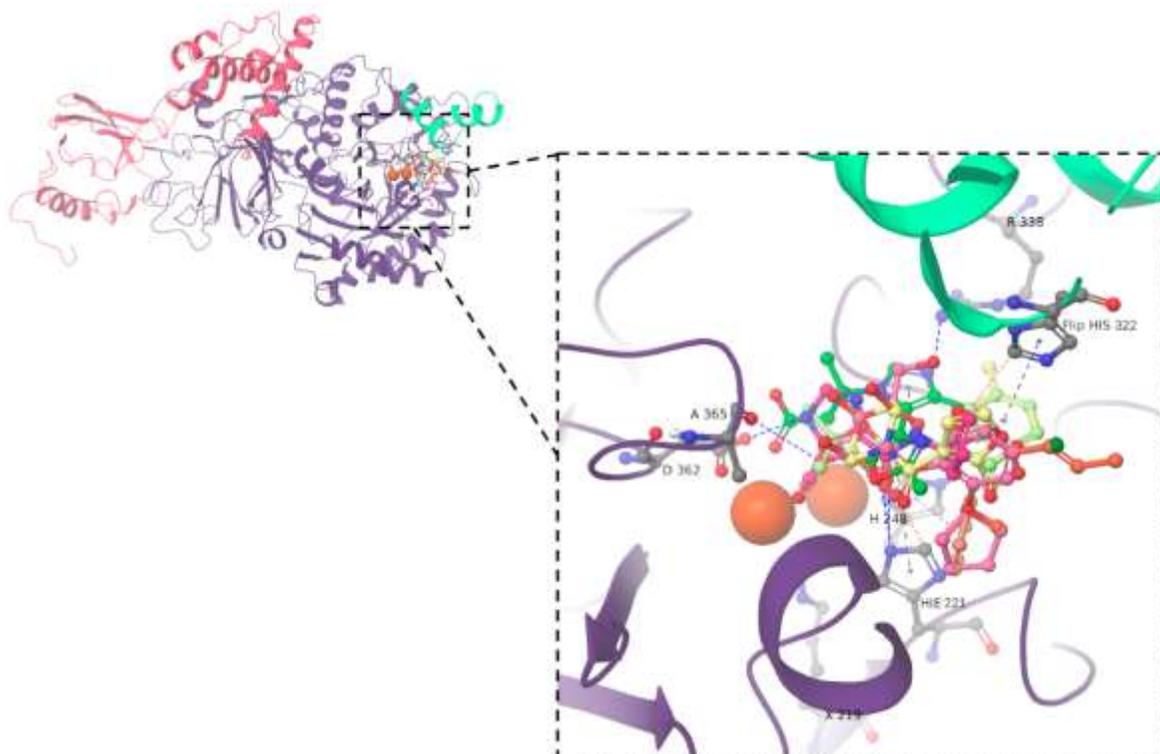


Figura 30: Interacción de 7 moléculas candidatas con respecto a la hipótesis de farmacóforo AAANR2. Moléculas candidatas: ZINC000008700584 (Verde limón), ZINC000072412074 (amarillo), ZINC000215300901 (magenta), ZINC000002522532 (rosa claro), ZINC000071769481 (naranja claro), ZINC000031169481 (naranja) y ZINC000091624618 (verde), iones de níquel (naranja) y residuos claves (gris).

A continuación, en la Figura 31 la mejor molécula seleccionada a partir del conjunto de compuestos que interactúan con ambas variantes. Mientras que en la Figura 32, se exhibe la mejor molécula del conjunto de compuestos que solo interaccionan con KCX. Ambas moléculas seleccionadas presentan la mejor energía de unión sin incumplirse con las reglas farmacocinéticas de cada conjunto de compuestos.

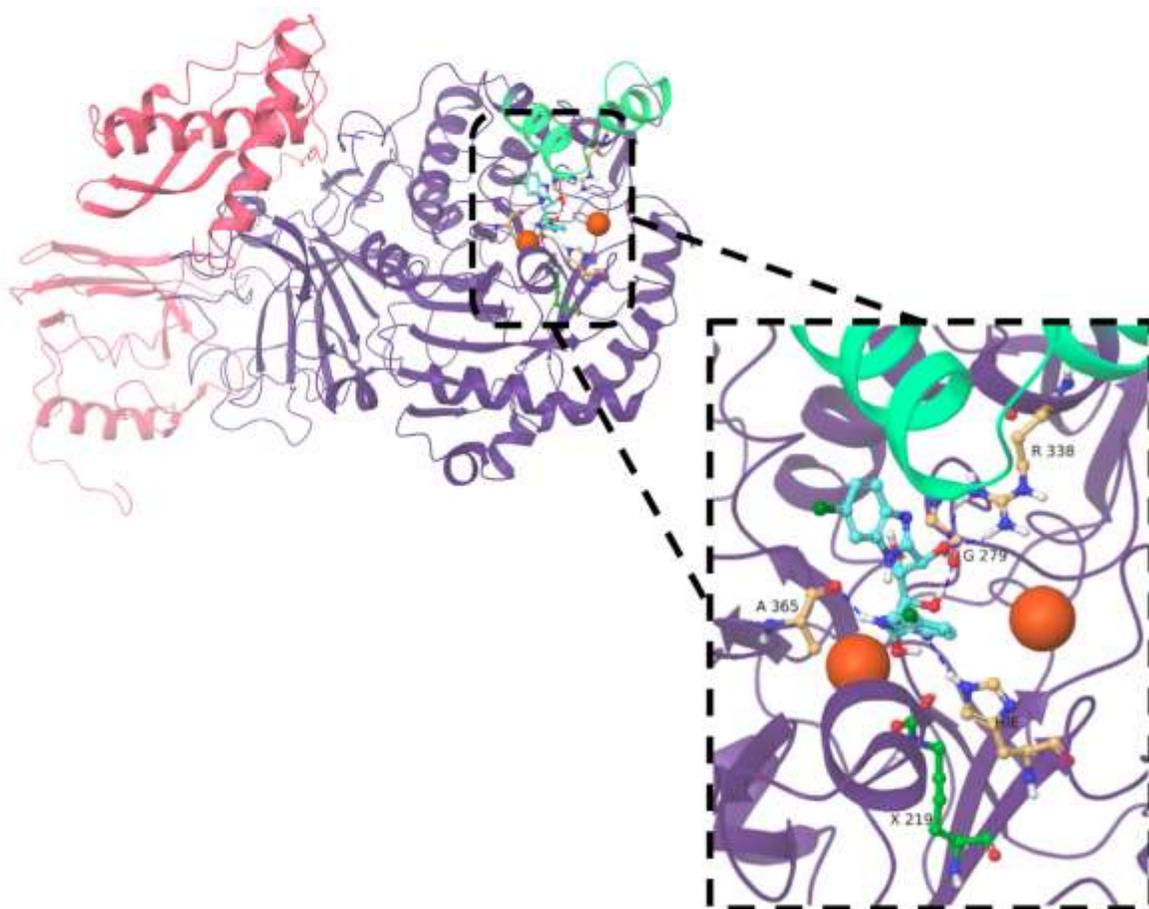


Figura 31: Complejo proteína-ligando, correspondiente a molécula ZINC000008627328 (coloreada en cyan) escogida como la mejor candidata de aquellas moléculas que interactúan con ambas variantes, en verde molécula KCX219.

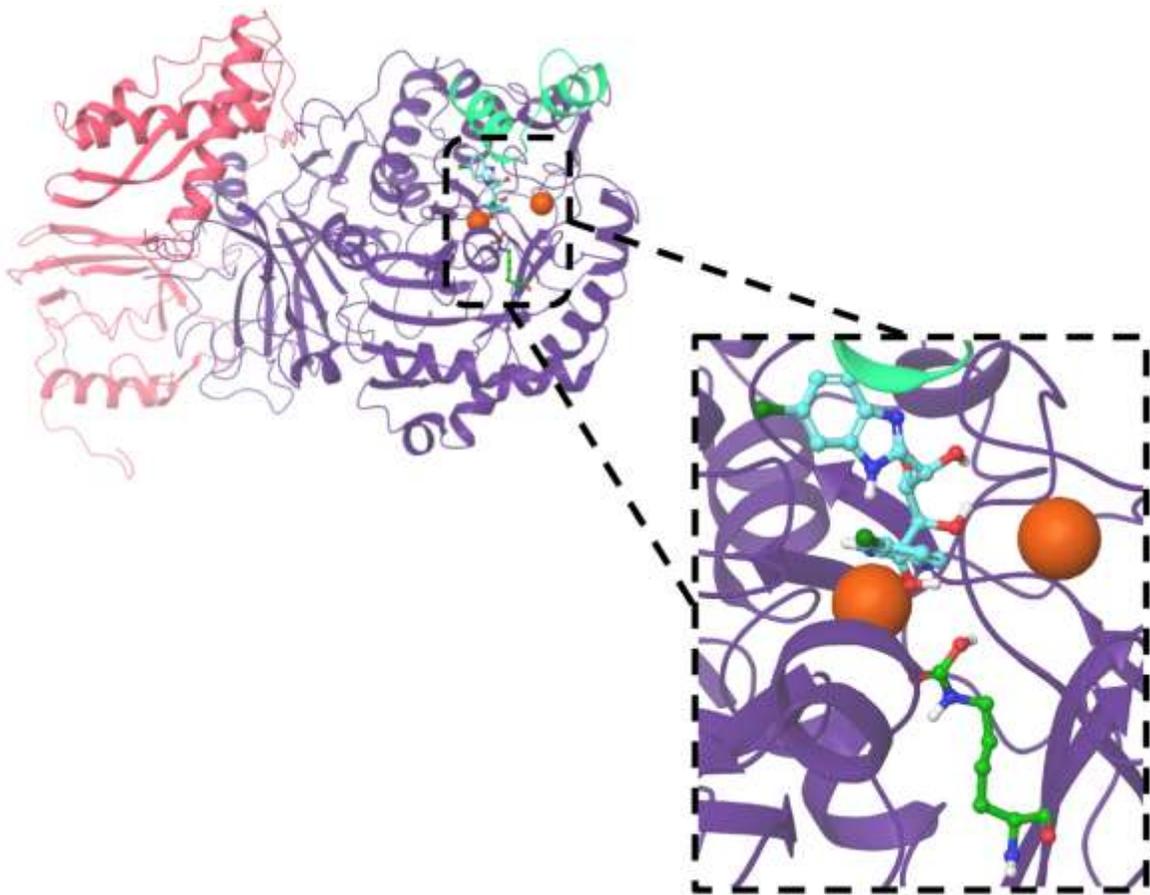


Figura 32: Complejo proteína-ligando, correspondiente a molécula ZINC000072412074 (coloreada en cian) escogida como la mejor candidata que solo interactúa con KCX, en verde molécula KCX219.

En la Figura 33, se puede observar el flujo de trabajo finalizado, con el número de moléculas que se obtuvo de cada procedimiento realizado a lo largo del trabajo de la tesis de grado.

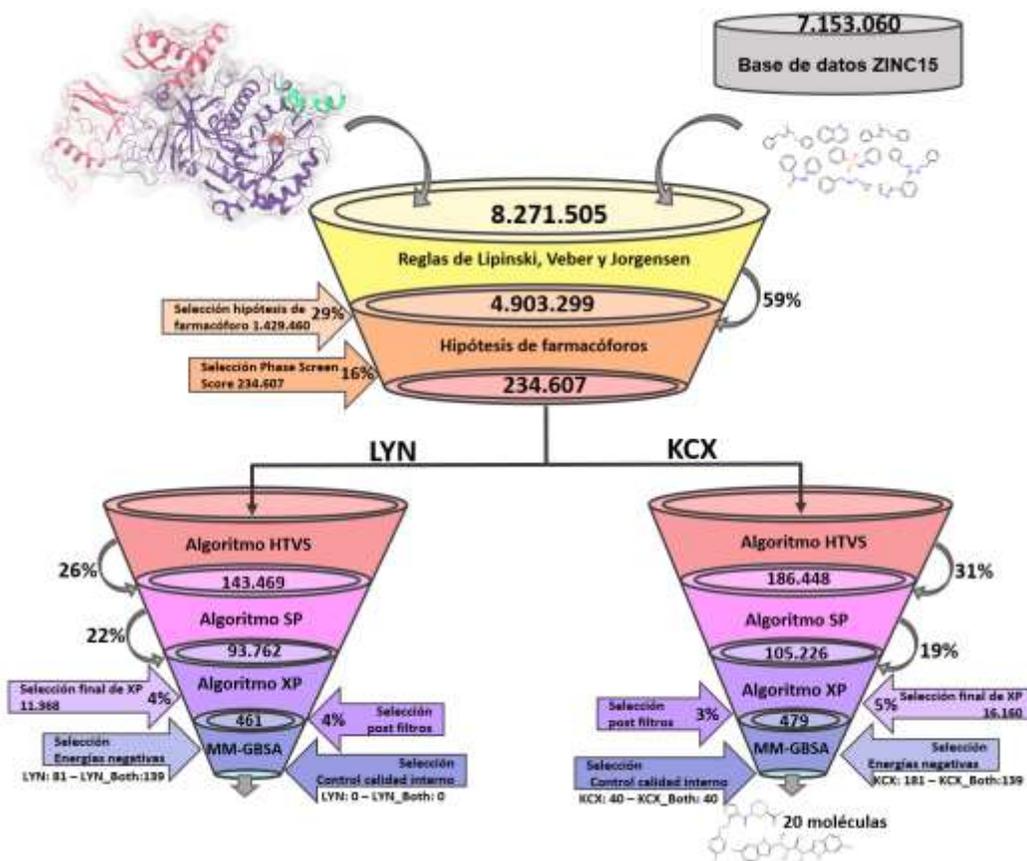


Figura 33: Flujo de trabajo. Número de moléculas reportadas de la metodología completa.

Finalmente en la Tabla 13, se observan los días de cálculo y análisis que tardaron cada uno de los procedimientos. En el caso de hipótesis de farmacóforo y creación de la base de datos para aplicar la hipótesis (LBVS), se ocuparon 9 días debido a que el número de moléculas fue de 4.903.299. Subsiguientemente, el análisis y filtrado demoró 8 días más, por lo que fue un proceso largo para acotar la biblioteca de moléculas. En el caso de los protocolos de *docking* molecular HTVS y SP, fueron necesarios entre 1 y 2 días aproximadamente; mientras que los análisis tardaron 2 ó 3 días, debido a que hubo filtrados que se realizaron manualmente. Por último, la ejecución del algoritmo de XP conllevó a 11 días de cálculo porque este método es mucho más riguroso que los otros 2 algoritmos recién mencionados, dada su capacidad de clasificar los ligandos candidatos para que se puedan unir a una conformación específica del receptor (HpU). Por lo que requiere una mayor cantidad de tiempo, minimizando la cantidad de falsos positivos, y por efecto de la puntuación separa los ligandos activos de los inactivos, siendo especialmente útil para la optimización de compuestos potenciales. Es por ello que existe la posibilidad de realizar SBVS solo con el algoritmo de XP ya que es mucho más drástico y aplica

una mayor cantidad de restricciones, reduciéndose el número de días de cálculo y análisis al no ejecutar los métodos de HTVS y SP.

Métodos	Días de cálculo	Días de análisis	Total días
LBVS	9	8	17
HTVS	1	2	3
SP	2	3	5
XP	11	4	15

Tabla 13: Tiempo de cálculos y análisis por cada procedimiento de LBVS y SBVS (Considerando una de las variantes).

CONCLUSIÓN

Las herramientas *Phase*, *Glide* y *Prime* de la suite de *Schrödinger* han demostrado ser una aplicación exitosa para la detección virtual de moléculas contra una diana biológica, en donde, a través de *Phase* se logró identificar hipótesis de farmacóforos que fueron útiles para la búsqueda de nuevos inhibidores. Seguidamente, *Glide* fue de utilidad para la realización de estudios de acoplamiento molecular y por último con *Prime* se realizaron los cálculos energéticos mediante MM-GBSA, de esta manera por medio de la metodología de *Virtual Screening* fue posible encontrar 20 moléculas candidatas.

Estas moléculas candidatas encontradas presentan energías libres de unión incluso más favorables que los inhibidores clásicos de esta enzima o que el sustrato natural (urea) y que además cumplen rigurosamente la mayoría de las reglas farmacocinéticas y farmacodinámicas usadas en la literatura (Reglas de: Lipinski, Veber, Jorgensen, Egan, Ghose, Muegge y Leadlikeness y las alertas estructurales PAINS y BRENK).

Específicamente, este conjunto de 20 compuestos seleccionados como candidatos, muestra una mejor afinidad de unión que la molécula de urea (-39,6 kcal/mol). También, se encontraron 18 moléculas que tienen una afinidad de unión más favorable que DJM (-64,3 kcal/mol) el inhibidor co-cristalizado de la enzima en estudio.

Por otro lado, la ejecución completa de este protocolo de descubrimiento de drogas puede ser ampliamente abordada por la comunidad sin la necesidad de grandes recursos computacionales. Dicho protocolo fue ejecutado sobre dos computadores Ryzen con 15 y 12 núcleos, sin necesidad de utilizar la tecnología CUDA sobre sus tarjetas gráficas.

REFERENCIAS

- [1] H. M. Berman *et al.*, “The protein data bank,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 58, no. 6 I, pp. 899–907, 2002, doi: 10.1107/S0907444902003451.
- [2] I. D. Weiner, W. E. Mitch, and J. M. Sands, “Urea and ammonia metabolism and the control of renal nitrogen excretion,” *Clin. J. Am. Soc. Nephrol.*, vol. 10, no. 8, pp. 1444–1458, 2015, doi: 10.2215/CJN.10311013.
- [3] H. A. Franch and W. E. Mitch, “Navigating between the scylla and charybdis of prescribing dietary protein for chronic kidney diseases,” *Annu. Rev. Nutr.*, vol. 29, pp. 341–364, 2009, doi: 10.1146/annurev-nutr-080508-141051.
- [4] K. Kappaun, A. R. Piovesan, C. R. Carlini, and R. Ligabue-Braun, “Ureases: Historical aspects, catalytic, and non-catalytic properties – A review,” *J. Adv. Res.*, vol. 13, pp. 3–17, 2018, doi: 10.1016/j.jare.2018.05.010.
- [5] C. Follmer, “Ureases as a target for the treatment of gastric and urinary infections,” *J. Clin. Pathol.*, vol. 63, no. 5, pp. 424–430, 2010, doi: 10.1136/jcp.2009.072595.
- [6] M. J. Maroney and S. Ciurli, “Nonredox nickel enzymes,” *Chem. Rev.*, vol. 114, no. 8, pp. 4206–4228, 2014, doi: 10.1021/cr4004488.
- [7] L. P. Horta *et al.*, “Urease inhibitors of agricultural interest inspired by structures of plant phenolic aldehydes,” *J. Braz. Chem. Soc.*, vol. 27, no. 8, pp. 1512–1519, 2016, doi: 10.21577/0103-5053.20160208.
- [8] J. B. Sumner, “the Isolation and Crystallization of the Enzyme Urease,” *J. Biol. Chem.*, vol. 69, no. 2, pp. 435–441, 1926, doi: 10.1016/s0021-9258(18)84560-4.
- [9] B. Krajewska, “Ureases I. Functional, catalytic and kinetic properties: A review,” *J. Mol. Catal. B Enzym.*, vol. 59, no. 1–3, pp. 9–21, 2009, doi: 10.1016/j.molcatb.2009.01.003.
- [10] Y. Nasu and R. E. Campbell, “Unnaturally aglow with a bright inner light,” *Science (80-90)*, vol. 359, no. 6378, pp. 868–869, 2018, doi: 10.1126/science.aas9159.
- [11] H. Cantarella, R. Otto, J. R. Soares, and A. G. de B. Silva, “Agronomic efficiency of NBPT as a urease inhibitor: A review,” *J. Adv. Res.*, vol. 13, pp. 19–27, 2018, doi: 10.1016/j.jare.2018.05.008.

- [12] M. A. Sutton, J. W. Erisman, F. Dentener, and D. Möller, "Ammonia in the environment: From ancient times to the present," *Environ. Pollut.*, vol. 156, no. 3, pp. 583–604, 2008, doi: 10.1016/j.envpol.2008.03.013.
- [13] R. Tao *et al.*, "Effects of urease and nitrification inhibitors on the soil mineral nitrogen dynamics and nitrous oxide (N₂O) emissions on calcareous soil," *Environ. Sci. Pollut. Res.*, vol. 25, no. 9, pp. 9155–9164, 2018, doi: 10.1007/s11356-018-1226-9.
- [14] G. Subbarao *et al.*, "Scope and strategies for regulation of nitrification in agricultural systems - Challenges and opportunities," *CRC. Crit. Rev. Plant Sci.*, vol. 25, no. 4, pp. 303–335, 2006, doi: 10.1080/07352680600794232.
- [15] C. S. Snyder, E. A. Davidson, P. Smith, and R. T. Venterea, "Agriculture: Sustainable crop and animal production to help mitigate nitrous oxide emissions," *Curr. Opin. Environ. Sustain.*, vol. 9–10, pp. 46–54, 2014, doi: 10.1016/j.cosust.2014.07.005.
- [16] L. V. Modolo, A. X. de Souza, L. P. Horta, D. P. Araujo, and Â. de Fátima, "An overview on the potential of natural products as ureases inhibitors: A review," *J. Adv. Res.*, vol. 6, no. 1, pp. 35–44, 2015, doi: 10.1016/j.jare.2014.09.001.
- [17] Q. Li *et al.*, "A new urease-inhibiting formulation decreases ammonia volatilization and improves maize nitrogen utilization in North China Plain," *Sci. Rep.*, vol. 7, no. March, pp. 1–9, 2017, doi: 10.1038/srep43853.
- [18] H. Suter, S. K. Lam, C. Walker, and D. Chen, "Enhanced efficiency fertilisers reduce nitrous oxide emissions and improve fertiliser 15N recovery in a Southern Australian pasture," *Sci. Total Environ.*, vol. 699, p. 134147, 2020, doi: 10.1016/j.scitotenv.2019.134147.
- [19] D. R. Scott, D. Weeks, C. Hong, S. Postius, K. Melchers, and G. Sachs, "The role of internal urease in acid resistance of *Helicobacter pylori*," *Gastroenterology*, vol. 114, no. 1, pp. 58–70, 1998, doi: 10.1016/S0016-5085(98)70633-X.
- [20] Q. Liu *et al.*, "Natural Products for the Prevention and Management of *Helicobacter pylori* Infection," *Compr. Rev. Food Sci. Food Saf.*, vol. 17, no. 4, pp. 937–952, 2018, doi: 10.1111/1541-4337.12355.
- [21] P. Parreira, M. Fátima Duarte, C. A. Reis, and M. C. L. Martins, "*Helicobacter pylori* infection: A brief overview on alternative natural treatments to conventional therapy," *Crit. Rev. Microbiol.*, vol. 42, no. 1, pp. 94–105, 2016, doi: 10.3109/1040841X.2014.892055.

- [22] J. Z. Li *et al.*, "Helicobacter pylori infection is associated with type 2 diabetes, not type 1 diabetes: An updated meta-analysis," *Gastroenterol. Res. Pract.*, vol. 2017, 2017, doi: 10.1155/2017/5715403.
- [23] C. X. Chen, Y. S. Mao, P. Foster, Z. W. Zhu, J. Du, and C. Y. Guo, "Possible association between helicobacter pylori infection and nonalcoholic fatty liver disease," *Appl. Physiol. Nutr. Metab.*, vol. 42, no. 3, pp. 295–301, 2017, doi: 10.1139/apnm-2016-0499.
- [24] S. Upala, A. Sanguankeo, S. A. Saleem, and V. Jaruvongvanich, "Effects of Helicobacter pylori eradication on insulin resistance and metabolic parameters: A systematic review and meta-analysis," *Eur. J. Gastroenterol. Hepatol.*, vol. 29, no. 2, pp. 153–159, 2017, doi: 10.1097/MEG.0000000000000774.
- [25] C. Silvia Pop, "Helicobacter pylori infection: old and new."
- [26] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA. Cancer J. Clin.*, vol. 65, no. 2, pp. 87–108, 2015, doi: 10.3322/caac.21262.
- [27] R. Kataria and A. Khatkar, "Molecular docking, synthesis, kinetics study, structure-activity relationship and ADMET analysis of morin analogous as Helicobacter pylori urease inhibitors," *BMC Chem.*, vol. 13, no. 3, pp. 1–17, 2019, doi: 10.1186/s13065-019-0562-2.
- [28] Y. F. Rego *et al.*, "A review on the development of urease inhibitors as antimicrobial agents against pathogenic bacteria," *J. Adv. Res.*, vol. 13, pp. 69–100, 2018, doi: 10.1016/j.jare.2018.05.003.
- [29] J. M. Liou *et al.*, "Concomitant, bismuth quadruple, and 14-day triple therapy in the first-line treatment of Helicobacter pylori: a multicentre, open-label, randomised trial," *Lancet*, vol. 388, no. 10058, pp. 2355–2365, 2016, doi: 10.1016/S0140-6736(16)31409-X.
- [30] H. Azizian, F. Nabati, A. Sharifi, F. Siavoshi, M. Mahdavi, and M. Amanlou, "Large-scale virtual screening for the identification of new Helicobacter pylori urease inhibitor scaffolds," *J. Mol. Model.*, vol. 18, no. 7, pp. 2917–2927, 2012, doi: 10.1007/s00894-011-1310-2.
- [31] Á. G. Díaz-Sánchez *et al.*, "Inhibition of urease by disulfiram, an FDA-approved thiol reagent used in humans," *Molecules*, vol. 21, no. 12, 2016, doi: 10.3390/molecules21121628.

- [32] B. Zambelli, F. Musiani, S. Benini, and S. Ciurli, "Chemistry of Ni²⁺ in urease: Sensing, trafficking, and catalysis," *Acc. Chem. Res.*, vol. 44, no. 7, pp. 520–530, 2011, doi: 10.1021/ar200041k.
- [33] N. C. Ha, S. T. Oh, J. Y. Sung, K. A. Cha, M. H. Lee, and B. H. Oh, "Supramolecular assembly and acid resistance of *Helicobacter pylori* urease," *Nat. Struct. Biol.*, vol. 8, no. 6, pp. 505–509, 2001, doi: 10.1038/88563.
- [34] M. S. Asgari *et al.*, "New 1,2,3-triazole–(thio)barbituric acid hybrids as urease inhibitors: Design, synthesis, in vitro urease inhibition, docking study, and molecular dynamic simulation," *Arch. Pharm. (Weinheim)*, vol. 353, no. 9, pp. 1–14, 2020, doi: 10.1002/ardp.202000023.
- [35] A. Balasubramanian and K. Ponnuraj, "Crystal Structure of the First Plant Urease from Jack Bean: 83 Years of Journey from Its First Crystal to Molecular Structure," *J. Mol. Biol.*, vol. 400, no. 3, pp. 274–283, 2010, doi: 10.1016/j.jmb.2010.05.009.
- [36] T. M. Chandra Babu *et al.*, "Molecular docking, molecular dynamics simulation, biological evaluation and 2D QSAR analysis of flavonoids from *Syzygium alternifolium* as potent anti-*Helicobacter pylori* agents," *RSC Adv.*, vol. 7, no. 30, pp. 18277–18292, 2017, doi: 10.1039/c6ra27872h.
- [37] L. Tan *et al.*, "Kinetics and mechanism study of competitive inhibition of Jack-Bean urease by baicalin," *Sci. World J.*, vol. 2013, 2013, doi: 10.1155/2013/879501.
- [38] S. Perveen *et al.*, "Antioxidant and urease inhibitory C-glycosylflavonoids from *Celtis africana*," *J. Asian Nat. Prod. Res.*, vol. 13, no. 9, pp. 799–804, 2011, doi: 10.1080/10286020.2011.593171.
- [39] R. Salehi Ashani *et al.*, "Synthesis, Biological Evaluation and Molecular Docking of Deferasirox and Substituted 1,2,4-Triazole Derivatives as Novel Potent Urease Inhibitors: Proposing Repositioning Candidate," *Chem. Biodivers.*, vol. 17, no. 5, 2020, doi: 10.1002/cbdv.201900710.
- [40] R. L. Shrestha, A. Adhikari, B. P. Marasini, R. N. Jha, and M. I. Choudhary, "Novel inhibitors of urease from *Corydalis govaniana* Wall.," *Phytochem. Lett.*, vol. 6, no. 2, pp. 228–231, 2013, doi: 10.1016/j.phytol.2013.02.002.
- [41] M. Alomari *et al.*, "Design, synthesis, in vitro evaluation, molecular docking and ADME properties studies of hybrid bis-coumarin with thiadiazole as a new inhibitor of Urease," *Bioorg. Chem.*, vol. 92, no. August, p. 103235, 2019, doi: 10.1016/j.bioorg.2019.103235.

- [42] U. Salar *et al.*, "Bis-coumarins; non-cytotoxic selective urease inhibitors and antiglycation agents," *Bioorg. Chem.*, vol. 91, p. 103170, 2019, doi: 10.1016/j.bioorg.2019.103170.
- [43] U. Rashid *et al.*, "Synthesis of 2-acylated and sulfonated 4-hydroxycoumarins: In vitro urease inhibition and molecular docking studies," *Bioorg. Chem.*, vol. 66, pp. 111–116, 2016, doi: 10.1016/j.bioorg.2016.04.005.
- [44] K. M. Khan *et al.*, "Biscoumarin: New class of urease inhibitors; Economical synthesis and activity," *Bioorganic Med. Chem.*, vol. 12, no. 8, pp. 1963–1968, 2004, doi: 10.1016/j.bmc.2004.01.010.
- [45] F. Naz *et al.*, "4-Oxycoumarinyl linked acetohydrazide Schiff bases as potent urease inhibitors," *Bioorg. Chem.*, vol. 105, p. 104365, 2020, doi: 10.1016/j.bioorg.2020.104365.
- [46] I. Khan *et al.*, "Exploring biological efficacy of coumarin clubbed thiazolo[3,2-b][1,2,4]triazoles as efficient inhibitors of urease: A biochemical and in silico approach," *Int. J. Biol. Macromol.*, vol. 142, pp. 345–354, 2020, doi: 10.1016/j.ijbiomac.2019.09.105.
- [47] M. Faisal *et al.*, "Enzyme inhibitory activities an insight into the structure–Activity relationship of biscoumarin derivatives," *Eur. J. Med. Chem.*, vol. 141, pp. 386–403, 2017, doi: 10.1016/j.ejmech.2017.10.009.
- [48] A. Manuscript, "Europe PMC Funders Group Biophysical Screening for the Discovery of Small-Molecule Ligands," vol. 1008, pp. 357–388, 2015, doi: 10.1007/978-1-62703-398-5.
- [49] C. McInnes, "Virtual screening strategies in drug discovery," *Curr. Opin. Chem. Biol.*, vol. 11, no. 5, pp. 494–502, 2007, doi: 10.1016/j.cbpa.2007.08.033.
- [50] O. Dror, A. Shulman-peleg, R. Nussinov, and H. J. Wolfson, "Predicting Molecular Interactions," pp. 71–90, 2004.
- [51] A. Jahn, G. Hinselmann, N. Fechner, and A. Zell, "Optimal assignment methods for ligand-based virtual screening," *J. Cheminform.*, vol. 1, no. 1, pp. 1–23, 2009, doi: 10.1186/1758-2946-1-14.
- [52] J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin, and M. Uchimaya, "Epik: A software program for pKa prediction and protonation state generation for drug-like molecules," *J. Comput. Aided. Mol. Des.*, vol. 21, no. 12, pp. 681–691, 2007, doi: 10.1007/s10822-007-9133-z.
- [53] E. S. Cunha, X. Chen, M. Sanz-Gaitero, D. J. Mills, and H. Luecke, "Cryo-EM structure of *Helicobacter pylori* urease with an inhibitor in the active site at 2.0

Å resolution,” *Nat. Commun.*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-020-20485-6.

- [54] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen, “Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values,” *J. Chem. Theory Comput.*, vol. 7, no. 7, pp. 2284–2295, 2011, doi: 10.1021/ct200133y.
- [55] E. B. Miller *et al.*, “A Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding,” *ChemRxiv*. ChemRxiv, Mar. 13, 2020, doi: 10.26434/chemrxiv.11983845.v1.
- [56] G. Palareti *et al.*, “Comparison between different D-Dimer cutoff values to assess the individual risk of recurrent venous thromboembolism: Analysis of results obtained in the DULCIS study,” *Int. J. Lab. Hematol.*, vol. 38, no. 1, pp. 42–49, 2016, doi: 10.1111/ijlh.12426.
- [57] R. A. Friesner *et al.*, “Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes,” *J. Med. Chem.*, vol. 49, no. 21, pp. 6177–6196, Oct. 2006, doi: 10.1021/jm051256o.
- [58] G. C. P. van Zundert, N. W. Moriarty, O. V. Sobolev, P. D. Adams, and K. W. Borrelli, “Macromolecular refinement of X-ray and cryoelectron microscopy structures with Phenix/OPLS3e for improved structure and ligand quality,” *Structure*, Apr. 2021, doi: 10.1016/j.str.2021.03.011.
- [59] C. Mulakala and V. N. Viswanadhan, “Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies?,” *J. Mol. Graph. Model.*, vol. 46, pp. 41–51, 2013, doi: 10.1016/j.jmglm.2013.09.005.
- [60] H. Sun, Y. Li, S. Tian, L. Xu, and T. Hou, “Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set,” *Phys. Chem. Chem. Phys.*, vol. 16, no. 31, pp. 16719–16729, 2014, doi: 10.1039/c4cp01388c.
- [61] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman, “Zinc12,” *Chem. Inf. Model.*, vol. 52, no. 7, pp. 1757–1768, 2012, [Online]. Available: <http://zinc.docking.org/>.
- [62] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings,” *Adv. Drug Deliv. Rev.*, vol. 64, no. SUPPL., pp. 4–17, 2012, doi: 10.1016/j.addr.2012.09.019.

- [63] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, "Molecular properties that influence the oral bioavailability of drug candidates," *J. Med. Chem.*, vol. 45, no. 12, pp. 2615–2623, 2002, doi: 10.1021/jm020017n.
- [64] F. Ntie-Kang, L. L. Lifongo, P. N. Judson, W. Sippl, and S. M. N. Efange, "How 'drug-like' are naturally occurring anti-cancer compounds?," *J. Mol. Model.*, vol. 20, no. 1, 2014, doi: 10.1007/s00894-014-2069-z.
- [65] G. Chemi, S. Gemma, G. Campiani, S. Brogi, S. Butini, and M. Brindisi, "Computational tool for fast in silico evaluation of hERG K⁺ channel affinity," *Front. Chem.*, vol. 5, no. Feb, pp. 1–9, 2017, doi: 10.3389/fchem.2017.00007.
- [66] R. A. Friesner *et al.*, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, vol. 47, no. 7, pp. 1739–1749, 2004, doi: 10.1021/jm0306430.
- [67] N. K. Salam, M. Adzhigirey, W. Sherman, D. A. Pearlman, and D. Thirumalai, "Structure-based approach to the prediction of disulfide bonds in proteins," *Protein Eng. Des. Sel.*, vol. 27, no. 10, pp. 365–374, 2014, doi: 10.1093/protein/gzu017.
- [68] J. Li, R. Abel, K. Zhu, Y. Cao, S. Zhao, and R. A. Friesner, "The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling," *Proteins Struct. Funct. Bioinforma.*, vol. 79, no. 10, pp. 2794–2812, 2011, doi: 10.1002/prot.23106.
- [69] M. S. Murgueitio, M. Bermudez, J. Mortier, and G. Wolber, "In silico virtual screening approaches for anti-viral drug discovery," *Drug Discov. Today Technol.*, vol. 9, no. 3, pp. 219–225, 2012, doi: 10.1016/j.ddtec.2012.07.009.
- [70] M. T. Kim, A. Sedykh, S. K. Chakravarti, R. D. Saiakhov, and H. Zhu, "Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches," *Pharm. Res.*, vol. 31, no. 4, pp. 1002–1014, 2014, doi: 10.1007/s11095-013-1222-1.
- [71] S. Lätti, S. Niinivehmas, and O. T. Pentikäinen, "Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization," *J. Cheminform.*, vol. 8, no. 1, pp. 1–5, 2016, doi: 10.1186/s13321-016-0158-y.
- [72] E. L. Carter, D. E. Tronrud, S. R. Taber, P. A. Karplus, and R. P. Hausinger, "Iron-containing urease in a pathogenic bacterium," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 32, pp. 13095–13099, 2011, doi: 10.1073/pnas.1106915108.

- [73] A. Daina, O. Michielin, and V. Zoete, "SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci. Rep.*, vol. 7, no. October 2016, pp. 1–13, 2017, doi: 10.1038/srep42717.
- [74] X. Feng, "Study on numbers of multi-tooth meshing teeth pairs for involute internal gear pairs with small tooth number difference," *Adv. Mater. Res.*, vol. 655–657, no. Cmc, pp. 578–585, 2013, doi: 10.4028.
- [75] W. J. Egan, K. M. Merz, and J. J. Baldwin, "Prediction of drug absorption using multivariate statistics," *J. Med. Chem.*, vol. 43, no. 21, pp. 3867–3877, 2000, doi: 10.1021/jm000292e.
- [76] I. Muegge, S. L. Heald, and D. Brittelli, "Simple selection criteria for drug-like chemical matter," *J. Med. Chem.*, vol. 44, no. 12, pp. 1841–1846, 2001, doi: 10.1021/jm015507e.
- [77] R. Brenk et al., "Lessons learnt from assembling screening libraries for drug discovery for neglected diseases," *ChemMedChem*, vol. 3, no. 3, pp. 435–444, 2008, doi: 10.1002/cmdc.200700139.
- [78] A. Daina, O. Michielin, and V. Zoete, "SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci. Rep.*, vol. 7, no. October 2016, pp. 1–13, 2017, doi: 10.1038/srep42717.
- [79] Verheij, H. J. (2006). Leadlikeness and structural diversity of synthetic screening libraries. *Molecular Diversity*, 10(3), 377–388. <https://doi.org/10.1007/s11030-006-9040-6>

ANEXOS

- 1) Script en lenguaje Perl utilizado para separar archivos ZINC mol2 multimoleculares en archivos individuales y eliminando los repetidos. Obtenido de www.biostars.org/p/9011/

```
#!/usr/bin/perl
use strict;
use warnings;
use Cwd;

my $file="";
my @ligands=();
my $mol2="";
my $workdir=getcwd;

for my $i (0..@ARGV-1){
    if ($ARGV[$i] eq "-f") {++$i;$file = $ARGV[$i];}
    if ($ARGV[$i] eq "-w") {++$i;$workdir .= $ARGV[$i];}
}

open IN, "$file";
while (<IN>){
    my $line = $_;
    chomp($line);
    if ($line =~ m/@<TRIPOS>MOLECULE/ && $mol2 eq ""){
        $mol2.=$line."\n";
    }
    elsif ($line =~ m/@<TRIPOS>MOLECULE/ && $mol2 ne "" ){
        my $outfile.=pop(@ligands)."mol2";
        open OUT, ">$outfile";
```

```

    print OUT "$mol2";
    close OUT;
    $mol2=$line."\n";
}
else{
    if($line =~ m/ZINC/) {push(@ligands,$line);}
    $mol2.=$line."\n";
    chomp($line);
}
}
close IN;
exit;

```

2) Script en bash utilizado para eliminar moléculas con el mismo nombre dejando solo 1 pose con la mejor energía de *docking*.

```

#!/bin/bash

a=1
aux3=0
rm control.txt 2.log

for (( aux1=1; aux1 < 10000; aux1++ ))
do

    if [ "$aux1" -eq "$aux3" ]
    then
        continue
    fi

    name_aux1=$(cat xp_kcx_${aux1}.mol2 | grep "ZINC") poses=1

```

```
for (( aux2=$aux1+1; aux2 < 10000; aux2++ ))
do
  name_aux2=$(cat xp_kcx_${aux2}.mol2 | grep "ZINC")

  printf "$aux1\t$aux2\n" >> control.txt

  if [[ "$name_aux1" == "$name_aux2" ]]
  then
    poses=`expr $poses + 1`

    if [ "$poses" -gt "$a" ]
    then
      rm xp_kcx_${aux2}.mol2
    fi
  fi

done
done
```

3) Script en bash utilizado para eliminar las moléculas que se encuentran a una distancia mayor a 3,5 Å. Este script llama a “dist_niquel.tcl”

```
# dist_niquel.tcl
set 1 ../a.mol2

set outfile [open dist.dat w];
mol load mol2 $1

set Num_mol [atomselect top "resname UNL1 UNL2 and same residue as within 3.5
of resname UNL3"]
set list_mol [$Num_mol get {resname}]
set count_mol [llength $list_mol]
puts $outfile "$count_mol"

close $outfile
exit

#Eliminar moléculas con distancias mayores a 3,5 A
#!/bin/bash

# sudo apt install babel, vmd
# obabel prot_1.mol2 XP_KCX_1.mol2 -j -O a.mol2
# resname UNL1 UNL2 and same residue as within 3.5 of resname UNL3

a=0
COUNTER1=1
DIR=1

cd $DIR/
rm dist.dat
rm -r ../moles
mkdir ../moles

for i in $(ls -v *.mol2)
do

    rm ../a.mol2
    obabel ../prot.mol2 $i -j -O ../a.mol2
    sed -i 's/UNK0/UNL3/' ../a.mol2
```

```
cp ../a.mol2 ../moles
mv ../moles/a.mol2 ../moles/a_{$i}.mol2
```

```
vmd -dispdev text -e ../dist_niquel.tcl
$mol=$(cat dist.dat | awk '{print $1}')
```

```
if [ "$mol" == "$a" ]
then
    rm $i
    printf "ELIMINANDO $i\n"
```

```
fi
```

- 4) Script en bash eliminar moléculas que tienen RMSD igual a cero entre ellas, dejando la que tiene mejor energía de docking, este script llama a "Alineacion.tcl".**

```
#Alineación.tcl
set 1_name kcx_17.mae
set 2_name kcx_18.mae

exec rm -r Alig/ rmsd.dat
exec mkdir Alig/

puts "$1_name\t$2_name\n"

set outfile [open Alig/rmsd.dat w];

mol load mae $1_name
set PDBref [atomselect 0 "all"]

# set TotalPDB

# for {set i 2 } {$i <= $TotalPDB } { incr i } {

mol load mae $2_name
set sel [atomselect 1 "all"]
```

```

$sel move [measure fit $sel $PDBref]
puts $outfile "[measure rmsd $sel $PDBref]"
puts "RMSD [measure rmsd $sel $PDBref]"

#$sel writepdb Alig/$i.mae
}

close $outfile

exit

#Eliminar moléculas con RMSD igual a cero
#!/bin/bash
a=0.0
COUNTER1=1

for i in $(ls -v)
do
    aux1=$(cat $i | grep "ZINC")
#    echo "$aux1"

    COUNTER2=1
    for j in $(ls -v)
    do
        aux2=$(cat $j | grep "ZINC")

        printf "molecule: $COUNTER1\t$COUNTER2\n" >> 1.log
        if [[ "$aux1" == "$aux2" ]]
        then
            #printf "$aux1\t$aux2\n"
            sed -i "1c\set 1_name ".$i" ../Alineacion.tcl
            sed -i "2c\set 2_name ".$j" ../Alineacion.tcl
            vmd -dispdev text -e ../Alineacion.tcl $i $j

            rmsds=$(cat Alig/rmsd.dat | awk '{print $1}')
            printf "RMSD: $rmsds\n"

            if (( $(echo "$rmsds < $a" |bc -l) ))
            then
                printf "Borrando $j $aux2"

```

```
        rm $j
    fi
fi
    COUNTER2=`expr $COUNTER2 + 1`
done
    COUNTER1=`expr $COUNTER1 + 1`
```

```
done
echo "terminado"
```