

Universidad San Francisco de Quito USFQ

Colegio de Ciencias Biológicas y Ambientales

Proyecto de Investigación

“Descubrimiento de proteínas *off target* humanas para la proteasa NS3 del Virus del Dengue y sus implicaciones para el diseño de fármacos basados en estructura”

Daniel Gregorio Aguilera Pesantes

Ingeniería en Procesos Biotecnológicos

Trabajo de titulación presentado como requisito

para la obtención del título de

Ingeniero en Procesos Biotecnológicos

Quito, 22 de mayo de 2018

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN
DE TRABAJO DE TITULACIÓN

“Descubrimiento de proteínas *off target* humanas para la proteasa NS3 del Virus del Dengue y sus implicaciones para el diseño de fármacos basados en estructura”

Daniel Gregorio Aguilera Pesantes

Calificación:

Nombre del profesor, Título académico

Yovani Marrero Ponce, Ph.D.

Firma del profesor

Quito, 22 de mayo de 2018

Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _____

Nombres y apellidos: Daniel Gregorio Aguilera Pesantes

Código: 00117143

Cédula de Identidad: 0923501761

Lugar y fecha: Quito, 22 de mayo de 2018

Resumen:

El virus del Dengue (DENV) es probablemente uno de los agentes infecciosos más relevantes para los países en las áreas tropicales y subtropicales. En la actualidad se han incrementado los esfuerzos a nivel mundial para desarrollar moléculas que tengan la capacidad de inhibir el crecimiento de DENV y por ende su infección. Se ha encontrado que inhibir la actividad de la proteasa de DENV (NS2b/NS3 Pro) es una estrategia óptima para combatir la replicación viral debido a que su función es esencial para procesar las demás proteínas de DENV. Sin embargo, los compuestos descubiertos hasta el momento no han tenido utilidad clínica, ya sea por falta de potencia y/o toxicidad. Dicha toxicidad está relacionada fundamentalmente a la falta de selectividad (promiscuidad) de los quimiotipos identificados. Hasta el momento no se conocía cuáles son las proteínas humanas que tienen alta similitud estructural con la proteasa de DENV. Es por ello, que es importante identificar la existencia de proteínas en el hospedador sustancialmente similares a la proteína diana del virus. Esto tiene el propósito de determinar la posibilidad, en una etapa temprana, de que exista reactividad cruzada entre proteínas humanas y la proteína diana del virus. De esta manera, se realizó un escaneo de todo el proteoma humano en donde inicialmente se encontró que existen proteínas humanas estructuralmente similares a las proteínas NS3 Proteasa (8), NS3 Helicasa (289) y NS5 MTasa (86) de DENV. Adicionalmente se encontró que no existen proteínas humanas similares a las proteínas NS1 y NS5 Polimerasa de DENV. En cuanto a la NS3pro, se encontró que existen varias cavidades drogables en estas proteínas humanas que son similares al sitio catalítico de la proteasa. Mediante el empleo del descubrimiento de conocimiento no supervisado se establecieron las relaciones de similitud que existen entre las diferentes cavidades de la proteasa del DENV y sus respectivas cavidades de las proteínas *off target* humanas, estableciendo un panel de cavidades a tomar en consideración en el diseño de nuevos inhibidores de la proteasa del DENV que tengan selectividad y una menor toxicidad. Con el propósito de corroborar la similitud estructural encontrada, y de explorar el impacto del panel propuesto en el descubrimiento de fármacos basado en estructura, se desarrolló una simulación retrospectiva mediante un estudio de Docking molecular donde se evidenció que la mayor parte del universo químico propuesto como inhibidor de la proteasa de DENV pueden unirse con la misma intensidad tanto al sitio activo de la proteasa de DENV como a las cavidades identificadas en las proteínas humanas previamente reconocidas en esta investigación; es decir, que ningún compuesto propuesto hasta el momento muestra una selectividad deseada. De esta manera, se corroboró que las proteínas humanas (TYSND1, HTRA1-4 y FAM111A-B) tienen cavidades significativamente similares al sitio activo de la proteasa. Estas proteínas pueden ser incluidas como *off-targets* en paneles de prueba para estudios de compuestos con actividad inhibitoria contra la proteasa de DENV. Finalmente, la metodología descrita en el presente trabajo puede ser empleada para identificar *off-targets* para potencialmente cualquier proteína de interés.

ABSTRACT

Dengue Virus (DENV) is perhaps the most relevant infectious agent in the tropical and subtropical countries. Nowadays, worldwide efforts to develop new molecules capable to prevent DENV growth have increased. It has been observed that inhibiting DENV protease could be an optimum strategy to fight viral replication, since its function is essential for processing most of DENV proteins. However, all the discovered compounds until now are not clinically useful whether for a lack of activity or their toxicity. This toxicity is related to the low selectivity of the compounds. To date, it is unknown which human proteins have a high similarity with DENV Protease. Therefore, it is relevant to identify the existence of host proteins that are significantly similar to the virus target protein. Hence, the aim of this study is to determine preliminary cross reactivity between human and viral proteins. Thus, we scanned the entire human proteome in search of human proteins highly similar to the DENV non-structural proteins. We found no human proteins similar to DENV NS1 and NS5 Polymerase. However, we found several human proteins highly similar to DENV NS3 Protease (7), NS3 Helicase (289) and NS5 Methyl-Transferase (89). More specifically, we found that many of the human proteins have several druggable pockets which are similar to the DENV protease active site and binding sites. Using an unsupervised knowledge discovery scheme, similarity relationships were established between the DENV protease pockets and their respective off-target human pockets. Therefore, a test panel was created to allow for the assessment of new DENV protease inhibitors with high selectivity. Moreover, a molecular Docking study was performed to corroborate the impact of the proposed test panel in a structure-based drug discovery process. It was evident that the entire universe of Dengue protease inhibitors reported in the literature is capable of binding with the same intensity, the DENV protease and its human off targets. Therefore, it was confirmed that the human proteins (TYSND1, HTRA1-4 y FAM111A-B) have pockets that are significantly similar to the DENV protease active site and binding sites. Finally, the methodology described in this work could be employed to identify of targets for any protein of interest.

Tabla de Contenidos

Resumen:.....	4
Índice de tablas:	7
Índice de Figuras:.....	8
Introducción:.....	9
Justificación:	15
Materiales:.....	16
Metodología:	19
Resultados:	28
Discusión:	33
Conclusiones:.....	43
Recomendaciones:	44
Financiamiento:.....	¡Error! Marcador no definido.
Literatura citada:	45

Índice de tablas:

Tabla 1 Resumen de información a lo largo del flujo de trabajo de proteínas no estructurales de DENV para las que se encontraron proteínas humanas estructuralmente similares. BackPhyre-Refseq se refiere a la cantidad de proteínas humanas similares a la proteína no estructural de DENV (Score de confianza >90%). UniprotKP se refiere a la cantidad de proteínas humanas que pudieron ser rastreadas a un ID actual. Modeladas se refiere a las proteínas que pudieron ser procesadas por Phyre2 para generar un modelo de homología adecuado. AADS se refiere a la cantidad de estructuras generadas por homología que pudieron ser analizadas por AADS para la detección de cavidades y cavidades se refiere a la cantidad de cavidades detectadas y extraídas por el método de AADS en las proteínas humanas analizadas.	71
Tabla 2 identificación del porcentaje de proteínas humanas con función y sustrato similar a una proteína no estructural de DENV (basado en sus Gene Ontology) y de la proporción de sitios de unión al sustrato y sitios activos desconocidos explícitamente en UniprotKB.	71

Índice de Figuras:

Figure 1 Función Molecular de Proteasas. El diagrama de redes generado con BiNGO en Cytoscape muestra las funciones moleculares organizadas jerárquicamente según sus GOs. El tamaño del círculo depende del número de repeticiones del GO en la lista de proteínas y la intensidad del color incrementa según el nivel de enriquecimiento del término GO indicado. ...	77
Figure 2 Proceso Biológico y Función Molecular de MTasas con GO Slim. El diagrama de redes generado con BiNGO en Cytoscape muestra las funciones moleculares organizadas jerárquicamente según sus GOs. El tamaño del círculo depende del número de repeticiones del GO en la lista de proteínas y la intensidad del color incrementa según el nivel de enriquecimiento del término GO indicado.....	77
Figure 3 Proceso Biológico de Helicasas con GO Slim. El diagrama de redes generado con BiNGO en Cytoscape muestra las funciones moleculares organizadas jerárquicamente según sus GOs. El tamaño del círculo depende del número de repeticiones del GO en la lista de proteínas y la intensidad del color incrementa según el nivel de enriquecimiento del término GO indicado.	78
Figure 8 Superposición entre estructura NS3 hel de DENV PDB ID 2BMF (cian) y proteína humana UniprotKB ID A0A024QZ90 (dorado).....	78
Figure 9 Superposición entre estructura NS5 MTasa de DENV PDB ID 3P97 (dorado) y proteína humana UniprotKB ID Q7RTR0 (cian).....	79
Figure 10 Superposición entre estructura NS3 Pro de DENV PDB ID 3U1I (dorado) y proteína humana UniprotKB ID P83105 (cian).	79

Introducción:

El virus del Dengue (DENV) es probablemente uno de los agentes infecciosos más relevantes para los países en las áreas tropicales y sub-tropicales (Le, Thi Tuyet Van, Hoang Quan, & Tho Duoc, 2017). Un reporte del *World Health Organization* indica que la incidencia de infecciones relacionadas a DENV ha crecido 30 veces durante los últimos 50 años (World Health Organization, 2012). Se estima que existen entre 50 y 100 millones de personas son infectadas cada año en un área superior a 100 países endémicos. Además, se ha estimado que existen 2.5 billones de personas que habitan en un área riesgo de DENV (World Health Organization, 2012). Este hecho es bastante relevante ya que las infecciones relacionadas a DENV como la fiebre hemorrágica de Dengue o el Síndrome de Shock de Dengue tienen una mortalidad de hasta el 2.5% (Mahmood, Jameel, Aslam, & Tahir, 2009).

Desde el punto de vista patológico, las infecciones de DENV comienzan a ser evidentes después de un periodo de incubación de entre 3 y 15 días. La Fiebre clásica del dengue comienza con el apareamiento abrupto de una fiebre alta (Tuiskunen Bäck & Lundkvist, 2013). Durante esta fase febril, la deshidratación puede generar alteraciones y en niños pequeños inclusive convulsiones febriles. La condición es auto limitante y se caracteriza por una enfermedad debilitante que incluye jaquecas, dolor retro-orbital, mialgia, artralgia, erupciones de petequias y leucopenia (Tuiskunen Bäck & Lundkvist, 2013). Después de entre tres y cinco días del comienzo de la fiebre, aparecen un exantema macopapular. Este generalmente aparece inicialmente en el torso, antes de expandirse en sentido periférico. La fiebre de DENV se conoce popularmente como la fiebre de huesos rotos debido a sus síntomas incapacitantes junto con el

dolor severo en músculos y articulaciones. Usualmente la fiebre persiste por siete días (Tuiskunen Bäck & Lundkvist, 2013).

DENV pertenece al género *Flavivirus* y puede subdividirse en cuatro serotipos (DENV 1-4) que están distribuidos en todo el mundo (E. S. Gan et al., 2017). DENV posee un genoma de ARN de sentido positivo de 10.7 kb que codifica una poliproteína que es pos-transcripcionalmente procesada por la proteasa viral (NS3-NS2B) y proteasas del hospedador (Aguilera-Pesantes et al., 2017). Los productos de estos cortes son tres proteínas estructurales (cápside, membrana (prM/M) y envoltura) y siete proteínas no estructurales (NS1, NS2A, NS2B, NS3, NS4A, NS4B y NS5) (Gubler, 2002). Estas proteínas no estructurales tienen capacidad enzimática y son aquellas que llevan a cabo el ciclo infeccioso de DENV. Además, se ha encontrado que muchas son responsables de la evasión del sistema inmune (Castillo Ramirez & Urcuqui-Inchima, 2015; Morrison, Aguirre, & Fernandez-Sesma, 2012).

De esta manera, una vez que el virus ingresa y se pierde el recubrimiento de la nucleocápside, la molécula de ARN se traduce en una poliproteína. Durante este proceso, las secuencias de señalización y *stop* de transferencia en la poliproteína dirigen su translocación a través de la membrana del retículo endoplasmático (Rodenhuis-Zybert, Wilschut, & Smit, 2010). Luego, la poliproteína es procesada co y post-transcripcionalmente por las proteasas celulares y la proteasa derivada del virus (NS3 Pro) en las proteínas estructurales y no estructurales. Otros procesamientos incluyen la glicosilación de las proteínas de envoltura, prM y NS1. Una vez traducidas, procesadas y plegadas, las proteínas no estructurales comienzan la replicación del genoma viral (Rodenhuis-Zybert et al., 2010). Este nuevo ARN recién sintetizado es empacado por la proteína de cápside para formar la nucleocápside. Las proteínas prM y de envoltura luego forman heterodímeros que se orientan hacia el lumen del retículo endoplasmático. Luego estos se

asocian en trímeros y las interacciones oligoméricas inducen una curvatura en la superficie, guiando el proceso de exocitosis de las partículas virales (Rodenhuis-Zybert et al., 2010).

La NS2b-NS3 Pro es una serina proteasa relacionada a la tripsina. En su forma activa, el dominio proteasa de la NS3 es envuelto por la región hidrofílica de NS2b, interactuando fuertemente con esta. Los extremos de Ns2b son hidrofóbicos y se encuentran anclados a la membrana (Erbel et al., 2006). Este complejo Ns2b-Ns3 tiene un plegamiento similar a una quimotripsina, estando conformada por dos barriles beta, cada uno formado por seis β -strands. En la hendidura entre los barriles β se encuentra una triada catalítica, conformada por His51, Asp75 y Ser135 (Bollati et al., 2010). Cabe destacar que la presencia del cofactor NS2B es crucial para la función y para la conformación de la proteasa (Erbel et al., 2006). Por otra parte, en cuanto a la función catalítica se sabe que existe una red exhaustiva unida por puentes de hidrógeno alrededor de los residuos catalítico (Bollati et al., 2010). De esta manera existe un centro de un solo donador y dos aceptores de hidrogeno entre His51 y Asp75. Para el reconocimiento del sustrato existen tres cavidades alrededor de la triada catalítica denominados S1, S2 y S3 (Bollati et al., 2010).

NS3 Pro además tiene otras funciones no catalíticas que están relacionadas al sistema de evasión a la respuesta inmune del hospedador. Por ejemplo, se sabe que NS2b/NS3 interactúa con IKK ϵ , enmascarando su dominio cinasa y previniendo la fosforilación de IRF3 (Zeidler, Fernandes-Siqueira, Barbosa, & Da Poian, 2017). Por otra parte, se sabe que NS2b/NS3 Pro además degrada a DDX21, un miembro de la familia de helicasas DExD/Hbox (Zeidler et al., 2017). Además la proteasa inactiva a STING, inhibiendo la expresión la expresión de IFN mediada por TBK1 (Zeidler et al., 2017).

Se ha realizado un extenuante trabajo acerca del descubrimiento de fármacos para la proteasa de DENV. Sin embargo, una gran parte de los esfuerzos se han enfocado en el desarrollo de inhibidores que sean capaces de unirse al sitio activo (Yildiz, Ghosh, Bell, Sherman, & Hardy, 2013). Las estrategias han ido desde crear moléculas similares al sustrato natural o crear moléculas mimetizando sus propiedades. Otras estrategias computacionales se han enfocado en realizar escaneos de compuestos a gran escala enfocándose en el sitio activo de la proteasa. Sin embargo, hasta el momento los resultados no han sido completamente satisfactorios (Yildiz et al., 2013). El desarrollo de fármacos por esta vía ha sido desafiante debido a que el sitio activo de la proteasa es plano y que además la proteasa tiene mayor afinidad por aminoácidos cargados positivamente. Sin embargo, se sabe que las moléculas cargadas típicamente exhiben una biodisponibilidad oral limitada (Yildiz et al., 2013).

Además de las dificultades técnicas para el desarrollo de fármacos para DENV otras razones que justifican porque en la actualidad no se cuenta con un antiviral para el virus. A pesar de la relevancia de este agente, históricamente ha recibido poca atención hasta hace poco. En el pasado, los recursos financieros destinados para la investigación en el campo de DENV eran reducidos en comparación a otros patógenos (Perry, Buck, & Shresta, 2011). Por ejemplo, el *National Institute of Allergy and Infectious Diseases* (NIAID) destinó 4.5 millones de dólares para la investigación en DENV en el año 2000, versus los 231.1 millones de dólares para investigación en VIH. Como resultado de esto, en menos de 20 años ya se cuentan con más de 20 compuestos antivirales aprobados para el tratamiento de VIH (Perry et al., 2011) mientras que para DENV no existe ninguno (Low, Ooi, & Vasudevan, 2017). Por otra parte, Desde el 2016 se cuenta con una vacuna (DengVaxia) aprobada para combatir los cuatro serotipos de DENV. Sin embargo, recientemente se ha generado una fuerte resistencia contra DengVaxia debido a nueva

evidencia encontrada que indica que la vacuna puede ocasionar una sintomatología peor a lo esperado en pacientes que no han sido infectados anteriormente por DENV (BBC, 2018). Sin embargo, para este punto la vacuna ya había sido administrada a más de 800 mil niños en las Filipinas. Actualmente se ha cesado el programa de vacunación en este país (BBC, 2018). Estas últimas noticias sumadas a la idea que varios autores sacan a destacar el hecho de que a pesar de que se cuenta con una vacuna para el Virus de la Fiebre Amarilla desde hace mucho, todavía se siguen generando brotes cada cierto tiempo (Low et al., 2017), han generado un consenso de que un antiviral podría resultar aún más beneficioso para combatir la infección (Low et al., 2017; Vannice, Durbin, & Hombach, 2016). Sin embargo, se sabe que ningún antiviral ha llegado hasta ensayos clínicos hasta el momento (Koh-Stenta et al., 2015; Low et al., 2017).

En la actualidad, se están implementando técnicas computacionales para asistir el descubrimiento de compuestos con actividad biológica (Huang, 2007). En casos en donde se conocen las estructuras 3D a nivel atómico de las proteínas implicadas en una función de interés, la utilización de una estrategia de descubrimiento de fármacos basada en estructura podría ser bastante beneficiosa. Este tipo de estrategias pueden hacer posible que se realice un escaneo virtual de millones de compuestos contra un receptor determinado (Lytras & Papadopoulou, 2018). Esto implica que con un nivel de precisión bastante elevado se puede acelerar considerablemente el proceso de descubrimiento de compuestos activos. No obstante, para ejecutar exitosamente este tipo de estrategias es importante designar un receptor adecuado dentro de la proteína de interés que sea biológicamente relevante y que presente ciertas características específicas, por ejemplo, que el blanco sea drogable. Para obtener este tipo de información se necesita un conocimiento bastante detallado acerca del sistema en estudio (Huang, 2007).

De esta manera, desde hace poco han incrementado los esfuerzos en donde se han realizado varios estudios en donde se están utilizando estas técnicas computacionales para asistir el descubrimiento y desarrollo de moléculas que puedan tener una actividad para combatir la infección de DENV. Una buena parte de estos se han centrado en la realización de escaneos virtuales de grandes librerías de compuestos contra las proteínas no estructurales NS3-NS2B Pro, NS5 MTasa, NS5 Pol y la proteína E, siendo la proteasa NS3-NS2b el blanco más popular (Galiano, Garcia-Valtanen, Micol, & Encinar, 2016; Sahoo, Jena, Daf, & Kumar, 2016; Singh, Kumar, Mansuri, Sahoo, & Deep, 2016; Zhou et al., 2008).

Es importante también identificar si existen proteínas en el hospedador sustancialmente similares a la proteína diana. Esto tiene el fin de determinar la posibilidad de que exista reactividad cruzada al momento de diseñar una molécula con la capacidad de unirse a la proteína diana (Amineni et al., 2010). Se sabe que este tipo de uniones inespecíficas pueden tener consecuencias impredecibles sobre la salud del hospedador, ocasionando una diversidad de efectos secundarios o exhibiendo una toxicidad importante (Amineni et al., 2010). Sin embargo, hasta el momento no se ha realizado ningún estudio para determinar si existen proteínas humanas que son suficientemente similares para provocar una reacción cruzada al diseñarse un fármaco dirigido a una de las proteínas no estructurales de DENV.

Justificación:

Recientemente se han publicado varios estudios en donde se han empleado técnicas computacionales para el descubrimiento de fármacos contra la proteasa de DENV. Por ejemplo, Mirza y sus colaboradores realizaron un escaneo de 18 millones de compuestos por *docking* contra la proteasa NS3-NS2B (Mirza, Salmas, Fatmi, & Durdagi, 2016). Una característica común muchos de estos estudios es que han empleado el sitio activo de estas proteínas como receptor para evaluar posibles ligandos (Behnam, Nitsche, Boldescu, & Klein, 2016; Nitsche, Holloway, Schirmeister, & Klein, 2014). No obstante, se sabe que la NS3-NS2Bpro pertenece a la familia de proteínas de las proteasas, las cuales existen naturalmente en el ser humano. De esta manera, existe la posibilidad de que algunas proteínas humanas sean estructuralmente similares a la proteasa de DENV, o al menos varias de sus cavidades (Lee, Redfern, & Orengo, 2007; Nitsche et al., 2014). Adicionalmente a esto, existe la posibilidad de que estas proteínas tengan sitios o cavidades que sean significativamente similares al sitio activo de la proteasa de DENV, el cual está siendo utilizado principalmente para guiar el desarrollo de fármacos (Amineni et al., 2010). De esta manera, de existir proteínas con cavidades sustancialmente similares al sitio activo de la proteasa de DENV, existe la posibilidad de que los compuestos que están siendo diseñados para unirse al sitio activo de DENV, tengan la capacidad de unirse también a estas cavidades en las proteínas humanas, teniendo una reactividad cruzada y causando un daño significativo (Amineni et al., 2010). Esta información puede ser vital para identificar *off-targets* humanos que pueden ser incluidos al momento de evaluar moléculas potencialmente inhibitorias para la proteasa de DENV. Esto tiene el fin de proveer de una herramienta adicional para optimizar el proceso de desarrollo de fármacos contra la proteasa de DENV, no solo buscando inhibidores potentes si no también con acción selectiva/específico con la diana deseada.

Materiales:

En el presente trabajo se emplearon principalmente programas informáticos y servidores de uso libre. A continuación, se encuentra una lista de cada una de las herramientas utilizadas y una breve descripción de su funcionalidad.

Base de datos Protein Data Bank RCSB *Research Collaboratory for Structural Bioinformatics* donde se encuentran estructuras tridimensionales determinadas experimentalmente de biomoléculas (Berman et al., 2002). Fue utilizada en esta investigación para recobrar las estructuras 3D de varias proteínas del DENV de los 4 serotipos y de las proteínas humanas de las que existe reporte (<http://www.rcsb.org>).

Phyre 2, BackPhyre, servidor web para la identificación de similaridad frente a una base pre-compilada de proteínas, basada en estructura secundaria (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). Fue utilizado para la búsqueda de proteínas humanas estructuralmente similares a las proteínas no estructurales de DENV (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>).

Phyre 2, Batch procesing, servidor web para la predicción de estructuras tridimensionales de proteínas basada en homología de secuencias (Kelley et al., 2015). Fue utilizada para construir las estructuras tridimensionales de las proteínas humanas con estructura desconocida.

Bases de datos *Reference Sequence (RefSeq)* de NCBI donde se encuentran secuencias curadas de proteínas (O'Leary et al., 2016).

Portal de Uniprot donde se encuentra recopilada y consolidada información acerca de anotaciones funcionales y estructurales de proteínas (Apweiler, 2004).

Cytoscape 3.6.0, software para la visualización y manipulación de redes (Shannon, 2003).

Extensión BiNGO de Cytoscape, aplicación para la creación de redes basadas en diferentes tipos y escalas de *Gene Ontologies* (Maere, Heymans, & Kuiper, 2005).

Extensión ClusterViz de Cytoscape, acoplada a algoritmo FAG-EC, aplicación para la generación de *clusters* a partir de una red previamente creada (J. Wang et al., 2015).

Aplicación COACH del servidor I-Tasser, servidor web para la predicción de los posibles sitios activos en proteínas (J. Yang, Roy, & Zhang, 2013).

AADS: servidor web para la extracción de cavidades y la predicción de los posibles sitios activos en proteínas (Singh, Biswas, & Jayaram, 2011)

F-pocket: servidor web para la detección y extracción de cavidades (Le Guilloux, Schmidtke, & Tuffery, 2009).

PockDrug: servidor web para calcular descriptores y predecir drogabilidad de cavidades (Hussein et al., 2015).

PyMOL 2.0, software para la visualización de biomoléculas (DeLano, 2002).

ProtDcal para calcular diferentes descriptores moleculares diseñados para proteínas de cada una de las cavidades (Ruiz-Blanco, Paz, Green, & Marrero-Ponce, 2015).

IMMAN para evaluar el set de atributos y seleccionar un subgrupo de estos (Urias et al., 2015).

Este software emplea estrategias de basadas en teoría de información que permiten seleccionar los atributos más idóneos (Urias et al., 2015). Se empleó el método no supervisado de *Brillouin Redundancy Index* acoplado a la entropía de Shannon (rSE), y se seleccionaron para todos los casos los 50 atributos más informativos (Todeschini & Consonni, 2009).

STATISTICA 8.0: software estadístico para realizar análisis de datos exploratorios y técnicas de minería de datos. Se emplearon las metodologías de agrupamiento de *Tree Joining Cluster*, *K-Means Cluster* y *Factor analysis* (PCA) (StatSoft, Inc., 2007).

Autodock: software para la realización de *Docking* molecular (Goodsell, Morris, & Olson, 1996).

Autodock Vina: software para la optimización del proceso de *Docking* molecular (Trott & Olson, 2009).

Autodock Tools: software para preparar proteínas y conocer coordenadas y dimensiones de área designada para realizar *Docking* (Morris et al., 2009).

Padel-ADV: software para la utilización de Autodock Vina de manera semi-automática en plataforma de Windows (Chun, 2009).

ToMoCoMD-CARDD QuBiLS-MIDAS: software empleado para construir estructuras 3D de ligandos a partir de estructuras 2D (sdf a pdb) (García-Jacas et al., 2014).

Marvin Sketch: software para representar moléculas en estructuras 2D (ChemAxon, 2014).

Metodología:

Se buscó analizar todas las proteínas no estructurales de DENV para las cuales se cuenta con una estructura deducida disponible. Para tener una buena representatividad y cobertura se tomaron en cuenta las variantes de las proteínas mencionadas para los cuatro serotipos reportados cuando fue posible. Se prefirió además las estructuras unidas a su sustrato natural o libres. Además, se consideró que la resolución de la estructura debe ser al menos de 2.5Å para tener resultados confiables (Huang, 2007). Tomando esto en cuenta se seleccionaron los PDB del RCSB de las proteínas NS1 (4OIG DENV1 y 4O6B DENV2), NS3 hel (2BMF DENV2 Y 2JLV DENV4), NS3-NS2B pro (3L6P DENV1, 2FOM DENV2, 3U1I DENV3 y 2WHX DENV4), NS5 MTasa (5IKM DENV1, 3EVG DENV2 y 3P97 DENV3) y NS5 Pol (5K5M DENV2 y 2J7W DENV3) (Akey et al., 2014; Chandramouli et al., 2010; Edeling, Diamond, & Fremont, 2014; Erbel et al., 2006; Geiss et al., 2009; Lim, Rahman, & Tejo, 2011; Luo et al., 2008, 2010; Noble, Seh, Chao, & Shi, 2012; Xu et al., 2005; Yap et al., 2007). Una vez seleccionadas las estructuras, estas son procesadas para remover cualquier tipo de molécula no relacionada a la proteína de interés que esté incluida en el PDB (por ejemplo aguas e iones), además dado que en varios casos se contó con estructuras duplicadas, se trabajó con una sola de ellas cuando fue relevante hacerlo (Humphrey, Dalke, & Schulten, 1996).

Para realizar la búsqueda de proteínas similares en el ser humano se ingresó la estructura procesada en el portal web de Phyre, en modo experto, en la opción de BackPhyre (Kelley et al., 2015). Esta función tiene la capacidad de convertir la secuencia y el perfil de nuestra estructura en un *Hidden Markov Model* (HMM). Luego, este es escaneado por medio de un sistema de pareo HMM-HMM contra una base de datos pre-compilada de HMMs (Kelley et al., 2015). El algoritmo empleado para realizar esto es el HHsearch, el cual se ha demostrado en pruebas

ciegas de CASP (*Critical Assessment of Techniques for Protein Structure Prediction*) que tiene uno de los mejores desempeños (Moult, Fidelis, Kryshtafovych, Schwede, & Tramontano, 2014; Soding, 2005). Además, esta metodología ha demostrado ser útil en diferentes tipos de investigaciones relevantes (Demircioglu, Sosa, Ingram, Ploegh, & Schwartz, 2016; Handa, Paul, Miller, Valentine, & Ghosh, 2016; MacNeill, 2014; Robb, Nano, & Boraston, 2012). Cabe recalcar que la base de datos empleada para este caso está constituida por todas las proteínas del genoma humano. Esto incluye todas las proteínas humanas depositadas en la base de datos *Reference Sequence (RefSeq)* de NCBI. Utilizar *RefSeq* representa una gran ventaja ya que esta es una base de datos confiable, ya que es curada manualmente y que posee una sola entrada por proteína (O'Leary et al., 2016). El método de Phyre2 empleado, tiene la ventaja de que no necesita que exista un gran nivel de homología a nivel de secuencia para encontrar proteínas similares, enfocándose más en los patrones de plegamiento, es decir en la similitud estructural. Esto nos permite realizar comparaciones entre organismos filogenéticamente no relacionados, los cuales van a compartir poca similitud a nivel de secuencia (Kelley et al., 2015). Una vez listo el *output* de BackPhyre se analizaron los scores de confianza y de porcentaje de identidad. El score de confianza nos indica la probabilidad porcentual entre 0 y 100% de que el par de proteínas realmente representa una estructura sustancialmente similar. Por otra parte, el porcentaje de identidad se refiere a la proporción de los residuos de la estructura ingresada que son equivalentes a la estructura encontrada (Kelley et al., 2015). Se seleccionaron para los pasos posteriores solo aquellas proteínas con un nivel de confianza del 90% o superior. Solo se encontraron proteínas humanas estructuralmente similares que cumplan este criterio para las proteínas NS3 pro, NS3 hel y NS5 MTasa, por lo cual los análisis posteriores solos se concentraron en estas tres proteínas. De esta manera, las proteínas NS1 y NS5 Polimerasa no

presentaron en el análisis un nivel de similitud significativo con ninguna proteína en todo el proteoma humano. De esta manera, se considera bajo este criterio que no existen proteínas off-target humanas para estas dos proteínas de DENV.

Los IDs de las proteínas humanas de *RefSeq* encontradas con BackPhyre fueron introducidas al portal de Uniprot para recuperar su secuencia de aminoácidos de una manera masiva (Apweiler, 2004). Además, se obtuvieron anotaciones relevantes como el nombre de la proteína, sitio activo, *binding site*, función, estado de desarrollo, especificidad de tejido, *gene ontology* GOs (proceso biológico y función molecular), localización subcelular y familia de proteínas. Se realizó minería de datos a nivel de texto de la información obtenida con el fin de describir el set de proteínas encontrado. Se realizó un análisis principalmente de los GOs, en los cuales se identificaron las funciones moleculares y los procesos biológicos más frecuentes.

La lista de proteínas humanas similares a las proteínas de DENV fueron analizadas mediante el software Cytoscape 3.6.0, empleando la extensión BiNGO (Maere et al., 2005; Shannon, 2003). Cytoscape está diseñado para crear *interaction networks* para sistemas moleculares biológicos utilizando un *framework* conceptual (Shannon, 2003). BiNGO específicamente construye las redes mediante los Gene Ontology al realizar un análisis de las funciones o términos que se encuentran sobre representados en la lista de genes o proteínas ingresadas, de manera jerárquica. Luego, realiza una prueba estadística para identificar el nivel enriquecimiento de estos términos al analizar la probabilidad de que los genes o proteínas ingresados forman parte de una categoría funcional determinada dentro de un grupo más grande, en este caso, el proteoma humano (Maere et al., 2005). Para todas las corridas de BiNGO se empleó un nivel de significancia de 0.01. Para los dos sets de proteínas más grandes (MTasas y Helicasas) las redes generadas fueron demasiado grandes haciendo difícil su análisis. De esta manera, se utilizó la extensión de

ClusterViz en Cytoscape para generar clústers a partir de la red principal (Wang et al., 2015). Para todos los casos se utilizó el algoritmo de clusterización FAG-EC en nivel débil y las demás opciones de default con el fin de promover la formación de grupos grandes (Jianxin Wang, Min Li, Chen, & Yi Pan, 2011). Las sub-redes generadas fueron analizadas individualmente.

La mayor parte de las proteínas humanas encontradas no poseen una estructura tri-dimensional conocida. Además, en varias ocasiones, cuando si existían, muchas estructuras estaban compuestas por solo fragmentos de la estructura completa, o en varias ocasiones existían múltiples versiones. De esta manera, las secuencias recuperadas de Uniprot fueron transformadas a formato FASTA empleando un *script in-house*, conservando su ID y fueron ingresadas en la opción de *Batch procesing* en el modo experto del portal de Phyre2 con el fin de obtener la estructura completa de cada una de las proteínas por medio de modelos de homología. Luego, estos fueron descargados y se analizó la información acerca de la calidad del modelo, correspondiente principalmente a la cobertura y a la confianza de los modelos obtenidos. El resumen de las métricas de calidad de los modelos de homologías para el set de las proteasas se encuentra en la Tabla S 1. A manera de comprobar de una manera más visual la calidad de los modelos obtenidos, cuando se conocía el PDB de la estructura (

Tabla S 2), se realizó una superposición entre la estructura conocida y el modelo de homología (Figura S 1).

Cabe destacar que, a partir de este momento, debido a la cantidad de datos observados, se prefirió centrar el estudio en la proteasa de DENV, en lugar de todas las proteínas no estructurales ya que se consideró era preferible realizar un análisis a mayor profundidad. Principalmente se tomó esta decisión ya que la proteasa es un blanco bastante estudiado como diana para guiar el desarrollo de fármacos para DENV. Sin embargo, cabe destacar también que para esta proteína existieron la menor cantidad de *off-targets* humanos identificados. De esta manera, el fin fue realizar una primera ronda exploratoria y desarrollar de una manera más manejable toda la metodología descrita. Posteriormente, la metodología descrita puede ser utilizada para los otros dos sets de datos. Por otra parte, se cuenta con todos los datos de las demás proteínas para ampliar el estudio en un futuro. De esta manera, debido a que la mayor parte de las proteínas humanas encontradas no poseen información acerca de su sitio activo o sitio de unión de manera explícita en una base de datos manejable y automática, se empleó el método de AADS para la predicción de los posibles sitios activos de las proteínas. El protocolo empleado para la detección de sitios activos se encuentra descrito en la literatura (Singh et al., 2016). El *output* de este programa resulta en una lista un poco mayor a 10 cavidades que pueden constituir el sitio activo. Sin embargo, el programa realiza una extracción de todas las cavidades antes de hacer el *ranking* (Singh et al., 2011). De esta manera, se decidió además trabajar con todas las cavidades extraídas ya que se considera que en realidad el sitio activo de DENV NS3pro, no es una cavidad, sino que se encuentra rodeado por tres cavidades principales que se encargan del reconocimiento del sustrato, de esta manera, estas cavidades pueden parecerse a cualquier otra potencialmente sin ser necesariamente catalíticas. Luego se seleccionaron los

residuos que se encuentran a 10 Å del centro de cada cavidad detectada y se extrajeron en un PDB separado. Por otra parte, si se conoce información acerca del sitio activo y del sustrato natural de las proteínas de DENV. De esta manera, se seleccionaron los PDB antes mencionados de estructuras unidas a su ligando natural y cuando no existió información al respecto, se empleó la herramienta COACH de I-Tasser (J. Yang, Roy, & Zhang, 2012; J. Yang et al., 2013). Del output de COACH seleccionó la predicción que constituye el ligando natural conocido con mayores scores. Luego, de esta predicción o de la estructura reportada en la literatura se seleccionó el área correspondiente a los residuos ubicados hasta 8 Å del ligando empleando la herramienta PyMOL 2.0 (DeLano, 2002). Las estructuras libres de ligandos fueron procesadas por AADS, pero a diferencia de la descripción anterior, se tomaron en cuenta solamente las cavidades detectadas que se ubicaban en la región alrededor de los a 8 Å del ligando, señalada anteriormente.

Una vez obtenidas todas las cavidades correspondientes a los sitios de unión de las proteínas de DENV y de las proteínas humanas similares se empleó el software ProtDcal para calcular diferentes descriptores moleculares diseñados para proteínas de cada una de las cavidades (Ruiz-Blanco et al., 2015). Alternativamente, se detectaron las cavidades de las proteínas humanas con f-pocket y se calculó una representación de descriptores diferentes empleando PockDrug (Hussein et al., 2015; Le Guilloux et al., 2009). Los pasos descritos a continuación se comparten se realizaron para ambas representaciones en paralelo, pero de manera idéntica.

Se seleccionaron los descriptores moleculares relacionados a la naturaleza de nuestras estructuras, de manera que se evitó usar descriptores relacionados a plegamiento de proteínas y aquellos basados en secuencia. Se seleccionaron principalmente descriptores relacionados a propiedades termodinámicas y otras propiedades químicas clásicas relevantes como

hidrofobicidad y carga. Para la computación de los descriptores se emplearon metodologías comunes como distancias euclidianas, media aritmética y desviación estándar. Todos los descriptores moleculares empleados se calcularon para la superficie de la cavidad y para la cavidad como una unidad proteica.

El software ProtDcal generó aproximadamente 300 diferentes descriptores moleculares para cada set de proteínas, mientras que PockDrug generó 16. Es común que, en la generación de descriptores moleculares, muchos de estos sean poco informativos y que muchos de ellos estén altamente correlacionados entre sí, lo cual crea redundancia entre los atributos (Guha & Willighagen, 2012). De esta manera, se empleó el software IMMAN para evaluar el set de atributos y seleccionar un subgrupo de estos (Urias et al., 2015). Este software emplea estrategias de basadas en teoría de información que permiten seleccionar los atributos más idóneos (Urias et al., 2015). Se empleó el método no supervisado de *Brillouin Redundancy Index* acoplado a la entropía de Shannon (rSE), y se seleccionaron para todos los casos los atributos más informativos (Todeschini & Consonni, 2009).

Para determinar cuáles son las cavidades de proteínas humanas que podrían ser significativamente similares a las cavidades de DENV NS3 Pro se optó por una estrategia de *data mining* no supervisada. Se utilizaron tres métodos de agrupamiento “todos contra todos”: análisis de componentes principales y clusterización por k-medias y *tree joining*. Para esto se utilizó como métrica para la definición de los clusters, el número más alto posible de partición de los grupos hasta antes de observar clusters formados por un solo caso. Luego, se analizó la concordancia entre los clusters y se reportaron las cavidades que fueron encontradas por más de un método que forma parte de un grupo consistente.

Finalmente, se realizó un experimento de *Docking* empleando el software autodock Vina, automatizado mediante Padel-ADV (Chun, 2009; Goodsell et al., 1996; Trott & Olson, 2009). Para esto inicialmente se construyó una base de datos de 80 compuestos reportados con actividad contra la proteasa de DENV en la literatura de manera manual empleando Marvin Sketch (ChemAxon, 2014) (Tabla S 3 y Tabla S 4). Cabe destacar que se realizó para esto una búsqueda literaria extensa para tomar una muestra representativa que represente todos los quimiotipos conocidos con actividad para el DENV que hayan sido reportados en publicaciones científicas. Además, se realizó la biblioteca de manera que los quimiotipos sean diversos y de que no exista sobre-representación de ningún quimiotipo.

Estas estructuras se transformaron a 3D empleando el software ToMoComD (García-Jacas et al., 2014). Luego se designaron las coordenadas y el tamaño del área correspondiente a cada cavidad sobre la cual se deseaba realizar el *Docking* con MGL Tools (Morris et al., 2009). Como parámetros para la realización del *Docking* se empleó un rango de energía de 3, un nivel de exhaustividad de 50 y número de modos 10 por molécula. Además, cada molécula fue dockeada por triplicado.

Para realizar las comparaciones a detalle entre los resultados del *Docking* se empleó la métrica de selectividad reportada por Sandeep y colaboradores, la cual se encuentra en a continuación (Sandeep, Nagasree, Hanisha, & Kumar, 2011). Sin, embargo, dado que el criterio descrito por Sandeep ($V > m + 3sd$) fue demasiado fuerte y ninguna molécula lo cumplió, se decidió aliviarlo con menores números de desviaciones estándar ($V > m + 2sd$, $V > m + 1sd$). Además, se emplearon tres métricas que consisten en el valor máximo, mínimo y promedio del score de *Docking* de las tres primeras poses de cada experimento incluyendo sus réplicas. Estos se seleccionaron debido a su grado de correlación previamente analizado.

$$V = \frac{V_0}{\left[\frac{(M_L + M_R)}{2} \right]}$$

Donde V = score de selectividad

V_0 = Valor de energía de unión obtenida en la simulación de *docking* según su respectiva métrica.

M_L = *Score* promedio obtenido para todos los ligandos de una proteína específica.

M_R = *Score* promedio obtenido para todas las proteínas de un ligando específico.

Resultados:

En la Tabla 1 se encuentra un resumen de la cantidad de proteínas humanas estructuralmente similares a las proteínas no estructurales de DENV y su flujo de información hasta la extracción de sus cavidades por AADS. No se incluyen las proteínas NS1 y NS5pol ya que no se encontraron proteínas humanas similares a estas.

En general se observó para las búsquedas de BackPhyre que casi todos los hits obtuvieron un 90% de confianza y un porcentaje de identidad menor al 30%. Esto implica, como era esperable un bajo nivel de homología entre los organismos. No se pudo recuperar todas las secuencias ya que en ocasiones el código de acceso de RefSeq indicado en el Output de BackPhyre había sido actualizado para el momento en que se realizó la búsqueda o debido a que en algunas pocas ocasiones varios IDs fueron mapeados al mismo ID de UniprotKB. Las estructuras generadas por homologías mediante Batch Processing de Phyre obtuvieron un valor de confianza mayor al 99% para el 99% o más de las estructuras y la cobertura del alineamiento se mantuvo la encima del 80%. Ciertas estructuras no pudieron ser generadas ya que las secuencias eran muy cortas (menor a 50 residuos) o muy largas (mayor a 3500 residuos) o no se contaba con un *template* adecuado. Se realizó una búsqueda de los PDB de las estructuras reportadas para las proteínas que existían

Tabla S 2 y se superpusieron con los modelos de homologías generados. En todos los casos, se observó un RMSD menor a 2Å, teniendo algunas con 0Å inclusive. Un ejemplo de esto se encuentra en la Figura S 1. En el caso de AADS, ciertas estructuras no pudieron ser analizadas debido a que su estructura no poseía cavidades evidentes, o eran demasiado grande para analizarse (mayor a 3000 residuos). A manera de una validación secundaria, se superpusieron además las estructuras de las proteínas no estructurales del virus con los modelos construidos de sus pares humanos Figure 4, Figure 5 y Figure 6.

A pesar de que se observó que la identidad de secuencia compartida entre las proteínas de DENV y las proteínas humanas similares detectas fue bastante reducida, se encontró que una buena parte de las proteínas humanas tienen una actividad similar a la proteína de DENV a la cual se parecen. En la

Tabla 2 se encuentra un resumen del porcentaje de proteínas humanas identificadas con función y sustrato similar a una proteína no estructural de DENV. Además, la proporción de sitios de unión al sustrato y sitios activos desconocidos explícitamente en UniprotKB.

Sin embargo, no toda la información funcional de los diferentes sets de proteína fue evidente y se encontraron otras actividades no necesariamente relacionadas a la proteína de DENV en cuestión. De esta manera se analizó mediante BiNGO en Cytoscape cada lista de proteínas. Para el caso de las proteasas se identificaron principalmente dos grupos funcionales enriquecidos. El primer grupo está constituido por proteínas con actividad endopeptidasas del tipo serina y el segundo grupo está constituido por proteínas asociadas a factores de crecimiento.

En el set de MTasas se observó según sus GOs, que las proteínas estaban asociadas a diferentes procesos biológicos (Figure 2). Algunos de estos estaban relacionados a metabolismo de pequeñas moléculas como SAM u otras aminos biogénicas, creatina. Otras funciones estaban relacionadas al procesamiento y unión a diferentes tipos de ARN como ARNnc, ARNr, ARNr, ARNsno y ARNt. También se encontraron aquellas relacionadas al procesamiento metabólicos

de ribonucleósidos tipo purina. Otras funciones se relacionaban a la regulación de la gluconeogénesis y del proceso metabólico de homocisteína, dopamina, catecolamina, catecol y diol. Otra función se centró en la metilación de histonas, de ADN y de ARN (tRNA y rRNA). Finalmente, un grupo se centró su función en la síntesis de poli-aminas como espermidina y espermina.

En el set de MTasas se observó que las proteínas estaban asociadas a diferentes procesos biológicos (Figure 3). Algunos de estos estaban relacionados a procesos metabólicos relacionados a ácidos nucleicos. Por ejemplo, actividad ligasa de regulación positiva y negativa, reparación de doble cadena de ADN vía recombinación homóloga, biosíntesis de macromoléculas, transcripción acoplada a reparación por escisión de par de nucleótidos (Escisión y rellenado de gaps), regulación de replicación de ADN dependiente de ADN y su iniciación y ensamblaje de complejos proteicos. Otro grupo de actividades se relaciona a la modificación o ensamblaje de la cromatina y organización de los telómeros. Otro grupo de proteínas está relacionado al procesamiento dependiente de ubiquitinas del proteosoma. Un grupo pequeño se relaciona a la regulación de proteínas virales y su transporte. Otro grupo se relaciona a la regulación del ciclo celular mediante procesos metabólicos de ADN y de la recombinación mitótica. Otro grupo se relaciona a las respuestas a stress en cuanto a proteínas que median el plegamiento mediante chaperonas. Además, otro grupo de proteínas está relacionado a funciones del ciclo celular en general como división celular, fase M de meiosis, transporte transmembrana, replicación de ADN, plegamiento de ADN y organización de organelos. Finalmente, un grupo representativo se relaciona a la expresión génica mediante el procesamiento de ARN, específicamente en funciones relacionadas al splicing de ARN.

En cuanto a la comparación de cavidades se destaca que para la representación de AADS-ProtDCal se seleccionaron 12 descriptores. De esta manera, dentro del análisis de agrupamiento se observó bastante consistencia entre los resultados de las tres metodologías empleadas. En general, se formaron 6 clusters de cavidades en donde están distribuidas las cavidades de DENV de los diferentes serotipos. Para armar estos clusters se seleccionaron los grupos de cavidades que consistentemente se agrupaban en al menos dos de los métodos descritos. De esta manera, realmente se obtuvieron estos 6 clusters para el análisis jerárquico y de k-medias mientras que en ACP formaron tres factores (2,3 y 8) en donde se encontraban cavidades de DENV con un alto *loading*. El resumen de esta representación se encuentra en la Tabla 3. Se destaca que las cavidades detectadas corresponden a siete de las ocho proteínas utilizadas en el análisis.

En el caso del análisis empleando la representación de Fpocket-PockDrug se seleccionaron 8 descriptores según el análisis de IMMAN (Urias et al., 2015) (Figura S 2). Dentro del análisis comparativo de la representación de Fpocket-PockDrug se encontró que la mayoría de los métodos fueron complementarios, obteniéndose tres clusters consistentes. En general, se compartieron más coincidencias entre el cluster jerárquico y de k medias que en comparación al ACP. Los factores con alto *loading* del PCA que contenían las cavidades de DENV no contenían las mismas cavidades humanas que en los otros métodos. Un resumen del resultado de agrupamiento empleando esta representación se encuentra en la Tabla 4. Se destaca que mediante esta representación las cavidades humanas solo correspondieron a dos de las ocho proteínas utilizadas en el análisis.

Cabe destacar que existieron diferencias marcadas entre la detección de cavidades inclusive desde los métodos de extracción (F-pocket y AADS). Por ejemplo, las cavidades de AADS son más grandes que las de F-pocket. Además, las cavidades como tales son diferentes. De esta

manera, inclusive las cavidades detectadas en DENV NS3 Pro no fueron exactamente iguales. Un ejemplo de un caso en el cual si coincidieron las cavidades por ambos métodos se encuentra en la Figure 7. Desde esta diferencia en el punto de referencia para el análisis de agrupamiento van a producir diferencias notorias en el agrupamiento de datos.

Al analizar más detalladamente, encontramos que, en ambas representaciones, se detectaron cavidades en dos de las ocho proteínas humanas similares a la proteasa. Sin embargo, solamente en la proteína humana *Trypsin Domain-containing1(Tysnd1)* o *peroxisomal leader peptide-processing protease*, se encontraron cavidades que fueron detectadas en ambas representaciones.

Se realizó un ejercicio de *docking* de 80 moléculas con actividad confirmada contra la proteasa de DENV. Se emplearon los parámetros de selectividad descritos a cada grupo de datos y se reportaron aquellas moléculas para las que se cumplió el parámetro en al menos dos de las métricas (Tabla 7). Además, la correlación entre las diferentes métricas de *docking* de los diferentes clusters fue analizado para reducir dimensionalidad. Se encontró que, en varios casos, las cavidades reportadas dentro de un mismo cluster tuvieron una alta correlación en sus scores de Docking molecular. Esto hace que los datos sean redundantes. Por lo tanto, en la Tabla 5 y la Tabla 6, se encuentra una lista de cavidades reducidas por clusters según su representación que se propone usar como referencia para reducir la extensión de experimentos de *Docking* en un futuro, en donde se quieran analizar estas cavidades de proteínas humanas para verificar reactividad cruzada. Además, fueron eliminadas aquellas cavidades humanas que mostraron una correlación demasiado reducida en referencia a la cavidad de la proteasa de DENV con la cual fue comparada.

Discusión:

La metodología propuesta en la presente investigación consiste en una alternativa exploratoria de amplio alcance. De manera clásica, la metodología computacionalmente asistida generalmente empleada en la comparación de sitios activos o cavidades se realiza de una manera manual y cualitativa (Sarafianos et al., 1999); (Russell et al., 2006). De esta manera, muchos de estos trabajos se basan en la superposición de dos proteínas o de sus sitios activos conocidos en los que se sospecha similitud (Sarafianos et al., 1999); (Russell et al., 2006); (Janez Konc & Janežič, 2014). Luego, se realiza usualmente una comparación de la calidad del alineamiento estructural y de la naturaleza de los residuos. De esta manera, se discute de manera cualitativa las implicaciones de las características de ciertos residuos como su hidrofobicidad, carga, etc., para indicar el grado de similitud (Sarafianos et al., 1999); (Russell et al., 2006). Otras metodologías clásicas incluyen la descripción de la actividad enzimática a nivel experimental, en donde se estudia la cinética de las enzimas para catalizar los mismos sustratos (Küçükgoze, Terao, Garattini, & Leimkühler, 2017).

Una limitante en la metodología de los trabajos anteriores se encuentra en que necesariamente se basan en comparaciones entre proteínas que mediante conocimiento previo se sabe que son similares, lo cual puede ser difícil de deducir (C.-K. Chen et al., 2010; Russell et al., 2006; Sarafianos et al., 1999). Para superar esta limitante, se han empleado técnicas de *screening* a nivel de secuencia para detectar genes potencialmente similares (Amineni et al., 2010; C.-K. Chen et al., 2010; Hasan, Daugelat, Rao, & Schreiber, 2006; Roemer et al., 2003). Esto se está realizando especialmente en trabajos en los cuales buscan definir genes o proteínas relevantes en patógenos para ser estudiados como blancos para el desarrollo de fármacos o terapias. Estas metodologías emplean una estrategia con enfoque de genómica comparativa, genómica funcional

y genómica sustractiva (Roemer et al., 2003; Hasan et al., 2006; Amineni et al., 2010). Sin embargo, estas estrategias se basan en similitud global a nivel de secuencia de genes o proteínas, lo cual es útil cuando se está realizando una comparación entre organismos filogenéticamente relacionados (Doolittle, Feng, Tsang, Cho, & Little, 1996).

No obstante, en el presente trabajo y en varios más se ha demostrado que en ocasiones las proteínas no necesariamente requieren que se comparta un nivel significativo de similitud a nivel de secuencia para que sean estructuralmente similares. La mayor parte de las proteínas encontradas presentan un nivel menor al 30% de identidad de secuencia, lo cual corresponde al porcentaje de identidad encontrado usualmente en proteínas no homólogas, y sin embargo son estructuralmente muy similares (Zhang, Jaroszewski, Rychlewski, & Godzik, 1997). Es más, en el presente trabajo se encontró evidencia de que la similitud estructural está relacionada a la función de una proteína (Lee et al., 2007; Mills, Beuning, & Ondrechen, 2015; Orengo, Jones, & Thornton, 1994). Considerando que muchas de estas proteínas tienen funciones similares y que inclusive catalizan sustratos similares, se puede intuir que pueden existir similitudes importantes a nivel de su sitio catalítico o sitios de unión al sustrato.

Emplear técnicas asistidas por métodos computacionales tiene un gran valor en la investigación biológica y en el campo del desarrollo de fármacos. Estas nos permiten escanear, generar, procesar y analizar una gran cantidad de información de una manera mucho más rápida a los métodos manuales. Además, nos permiten realizar metodologías de carácter predictivo de gran utilidad (Lahti, Tang, Capriotti, Liu, & Altman, 2012). Específicamente en el campo de la comparación de cavidades, existen algunos pocos trabajos que exploran técnicas masivas computacionales como el trabajo de Weskamp y colaboradores (Weskamp, Hüllermeier, & Klebe, 2009a). Este se basó en el análisis de todas las cavidades de la base de datos CavBase,

realizar un mapeo y agrupamiento con el fin de encontrar familias de cavidades similares (Leinweber et al., 2016; Weskamp, Hüllermeier, & Klebe, 2009). Adicionalmente, es necesario indicar existen trabajos publicados que han generado servidores o programas libres para realizar comparaciones masivas entre sitios activos. Estos emplean diferentes métricas y metodologías para establecer el nivel de similitud de las cavidades. Es más, en instancias iniciales se intentó utilizar estos servidores para realizar el presente trabajo. Sin embargo, cada uno de estos tuvo diferentes limitantes que hicieron impráctica su utilización. La primera limitante observada es que todos estos servidores realizan la comparación con cavidades extraídas de proteínas del PDB Bank. De esta manera, solo podría ser detectada la cavidad si la estructura de la proteína se encuentra elucidada y por ende no permiten realizar un escaneo completo del proteoma humano completo (Anand, Yeturu, & Chandra, 2012; Das, Krein, & Breneman, 2010; Gao & Skolnick, 2013; Kinoshita, Murakami, & Nakamura, 2007; J. Konc & Janezic, 2010; Powers et al., 2011). Por otra parte, en muchos de estos para definir el sitio que se desea comparar es necesario tener un ligando unido al sitio de interés y además los sitios con los que son comparados, son aquellos que también estuvieron unidos a un ligando en una estructura reportada (Anand et al., 2012; Das et al., 2010; Gao & Skolnick, 2013; Kinoshita et al., 2007; J. Konc & Janezic, 2010; Powers et al., 2011). Esto limita la posibilidad de que se pueda detectar una estructura o una cavidad reportada no unida a un ligando. Además, existieron servidores que no admitían un ligando peptídico para reconocer el sitio de comparación, lo cual es crucial para la proteasa de DENV (Anand et al., 2012; Kinoshita et al., 2007). Finalmente, otros servidores o páginas web no se encontraban en servicio al momento de realizar la experimentación del presente trabajo (Anand et al., 2012; Powers et al., 2011).

Otros trabajos en los cuales que han estudiado cavidades y sitios activos han sido destinados principalmente para entrenar “algoritmos” que luego sean capaces de predecir sitios activos en proteínas o analizar su drogabilidad (An, Totrov, & Abagyan, 2005; Kellenberger, Schalon, & Rognan, 2008; Kinnings & Jackson, 2009; Singh et al., 2011; Desaphy, Azdimousa, Kellenberger, & Rognan, 2012; Konc & Janežič, 2014; Leinweber et al., 2016).

El presente trabajo representa un aporte en el desarrollo de una metodología que permite analizar el proteoma humano completo en busca de proteínas con cavidades potencialmente similares a cavidades específicas de una proteína que deseamos usar como blanco. Debido a que se evitó que la búsqueda de proteínas se realice en base a secuencia, y se realizó en base a estructura, fue posible identificar proteínas similares en dos organismos filogenéticamente no relacionados (DENV y humanos) (Kelley et al., 2015). Además, dado que se utilizaron las técnicas de la *suite* de Phyre2, se pudo tomar en cuenta muchas proteínas para las cuales no existía una estructura elucidada disponible abiertamente (Kelley et al., 2015). De esta manera fue posible mediante los análisis posteriores analizar rápidamente un set grande de cavidades y evaluar la similitud entre estas de manera bastante precisa y con una gran velocidad.

Por otra parte, la utilización de descriptores moleculares es una estrategia bastante deseable en estas ocasiones, ya que se ha demostrado que tienen una elevada eficiencia en sistemas predictivos y en sistemas donde se busca clasificar o caracterizar proteínas (Ruiz-Blanco et al., 2015). Además, este es un método que no depende de alineamiento, por lo cual es posible comparar potenciales sitios activos sin la necesidad de que estos sean exactamente iguales (Richter, Wenzel, Stein, Gabdoulline, & Wade, 2008; La et al., 2009). Además, el hecho de que se utilizaron técnicas de minería de datos no supervisado hizo posible realizar un análisis en un

escenario en donde no existe una métrica específica a la cual se desea llegar como una variable de respuesta o un punto de comparación previamente conocido (La et al., 2009).

En el sentido práctico del caso de estudio, los resultados obtenidos tienen implicaciones importantes. En el diseño de fármacos, la reactividad cruzada de ligandos diseñados entre un blanco específico de un patógeno y un *off-target* del hospedador es algo que se busca evitar (Depta & Pichler, 2003; Hasan et al., 2006; Roemer et al., 2003). Varios autores han señalado la necesidad de desarrollar mejores técnicas para la predicción de reactividad cruzada (Depta & Pichler, 2003). La unión de un fármaco con una proteína del hospedador puede tener efectos adversos para el metabolismo o la salud en general del paciente cuando se le está administrando el medicamento. Estas consecuencias no deseadas pueden traducirse en toxicidad o en una serie de efectos secundarios (Depta & Pichler, 2003). Desde el punto de vista económico, muchas veces esta toxicidad o producción de efectos adversos generados por las reacciones cruzadas no son evidenciadas en los ensayos preclínicos del proceso de desarrollo de fármacos (Alomar, 2014). De esta manera, es frecuente que estos salgan a la luz en las fases de ensayos clínicos posteriores (Fase 1, 2 o 3), siendo una de las razones por las cuales una farmacéutica se vea forzada a abandonar el desarrollo de un fármaco, generando pérdidas significativas para la empresa (Alomar, 2014). Desde el punto de vista social, este fracaso además va a traducirse en que la población tendría que esperar más tiempo en tener un fármaco disponible para tratar un mal específico (Alomar, 2014).

De esta manera, en el presente trabajo se identificó una serie de proteínas humanas que son estructuralmente similares a tres proteínas no estructurales de DENV (NS3 Pro, NS3 Hel y NS5 MTasa) y se identificaron cavidades dentro de estas que son similares específicamente al sitio activo de NS3 Pro. Esto significa que se identificaron proteínas humanas con las cuales podría

existir una reacción cruzada al momento de diseñar un fármaco dirigido al sitio activo la NS3Pro de DENV. De esta manera, si se desea realizar un fármaco para esta diana, al momento de realizar los ensayos *in silico* de *screening* o *in vitro* de inhibición, debería además agregarse un panel de prueba que incluya las proteínas resultantes (*off-targets*) en este análisis para confirmar si puede ocurrir una reacción cruzada en algún punto y analizar sus consecuencias. De esta manera, puede abandonarse tempranamente el desarrollo de un fármaco que potencialmente va a ver problemas en etapas posteriores y puede incurrir en una pérdida para la compañía en caso de existir o en caso contrario se puede proceder con la investigación del fármaco con mayor seguridad.

Estos *off-targets* humanos de la NS3 Pro de DENV pueden ser considerados además *anti-targets*. Un *antitarget* se refiere a cualquier tipo de diana biológica, como un receptor o una enzima que, al ser afectado por una droga, ocasiona efectos indeseables en un paciente. Estas proteínas usualmente son de gran importancia para diferentes funciones celulares o moleculares. Algunos de los *antitargets* más estudiados incluyen a la familia de las cinasas, las proteasas y canales de HERG (Braga et al., 2014; Edwards, Høyer-Hansen, Blasi, & Sloane, 2008; Urbán, 2015).

Se encontró que todas las proteínas identificadas en el presente trabajo tienen funciones celulares relevantes y aunque en algunos casos su función exacta no es completamente conocida, se sabe que están fuertemente asociadas a diferentes enfermedades y complicaciones.

Uno de los hits más relevantes, *TrypsinDomain-containing1* (Tysnd1) o *peroxisomal leader peptide-processing protease* es una serina proteasa parecida a la tripsina que se ubica principalmente en la matriz del proteosoma. Esta se encarga junto a PsLon (*proteosomal Lon protease*) de procesar precursores de proteínas que contienen PTS1 y PTS2 (*Peroxisomal targeting signal*, motivo que identifica proteínas del peroxisoma) (Kurochkin et al., 2007). Se

sabe que sus sustratos están involucrados en la β -oxidación de ácidos grasos de cadena larga y ramificados. Se ha identificado que su inactivación provoca que disminuya significativamente la β -oxidación de ácidos grasos de cadena larga (Okumoto, Kametani, & Fujiki, 2011). En estudios realizados en ratones, se encontró que noquear el gen *Tysnd1* tiene consecuencias negativas. En primer lugar, los ratones machos mostraron infertilidad. Se encontró que los espermatozoides del epidídimo carecían del recubrimiento del acrosoma. Además, se observó que los ratones desarrollaron deficiencia hepática al administrarse fitol (precursor de ácido fitánico). Finalmente, la pérdida de este gen interfirió con la localización de las proteínas *Acaa1*, *Phyh* y *Agps* las cuales podrían causar un fenotipo similar al síndrome Zellweger (Mizuno et al., 2013).

Por otra parte, se detectaron las cuatro proteínas de la familia HTRA en humanos, un grupo de serinas proteasas oligoméricas que combinan un dominio proteasa similar a tripsina con al menos un dominio de interacción PDZ (Vande Walle, Lamkanfi, & Vandenabeele, 2008). Entre estos HTRA2 o Omi, es una serina proteasa principalmente ubicada en espacio intermembrana de la mitocondria (Bulteau & Bayot, 2011). Se sugiere que está relacionada a mecanismos de respuesta a estrés (Gray et al., 2000). HTRA2 es importante para procesos de muerte celular programada. Para participar en el proceso apoptótico esta proteína es liberada al citosol, promoviendo la activación de las caspasas dependientes del citocromo c, al neutralizar (escindir) proteínas inhibidoras de apoptosis (IAPs) con su motivo *IAP-binding* (Suzuki, Takahashi-Niki, Akagi, Hashikawa, & Takahashi, 2004). Inicialmente en estudios de *knock-out* en ratones se observó que tiene un efecto neuro-protectivo. Aunque no se observó diferencia en apoptosis en su fenotipo se observó una pérdida importante en la población de neuronas en el *stratum*, resultando en un síndrome neurodegenerativo con un fenotipo similar al Parkinson que llevó a los individuos a morir a los 30 días (Martins et al., 2004). Empleando esta información, en un

estudio de asociación se relacionaron ciertos alelos de HTRA2 a la enfermedad Parkinson Alemana en humanos (Strauss et al., 2005). Por otra parte, HTRA2 ha sido relacionada a procesos de oncogénesis (Bulteau & Bayot, 2011). Además, ciertas variantes están asociadas al desarrollo temprano de un síndrome mitocondrial, relacionado con aciduria 3-metilglutaconico (Oláhová et al., 2017).

HTRA1 es una serina proteasa conservada de 51 kDa. (Baldi et al., 2002). La mayor parte es secretada en el espacio extracelular mientras que el remanente se localiza en el citoplasma unido a microtúbulos. En el espacio extracelular actúa como una enzima que digiere proteínas y otras moléculas de la matriz extracelular. Intracelularmente ha sido implicada como un supresor tumoral y se la ha relacionado con procesos de proliferación y migración celular (Fex Svenningsen et al., 2017). Se ha propuesto que influye en el crecimiento al regular la disponibilidad de IGFs al escindir proteínas que se unen a IGFs (Nie, Hampton, Li, Findlay, & Salamonsen, 2003). Por otra parte, se ha reportado que HTRA1 puede unirse a MIF (Fex Svenningsen et al., 2017). En cuanto a su relación a enfermedad se ha encontrado que su *downregulation* ha sido detectada en diferentes tipos de cáncer, en donde su sobreexpresión fomenta la apoptosis celular y la reducción de la proliferación celular en líneas de cáncer. Por otra parte, su *upregulation* ha sido reportada en placentación, preeclampsia, artritis, Alzheimer, distrofia muscular de Duchenne y cardiomiopatía dilatada idiopática (Ng et al., 2017; L.-J. Wang, Cheong, Lee, Lee, & Chen, 2012). Se destaca además que HTRA1 está involucrada en el desarrollo placentar (Tossetta et al., 2016).

HTRA3 es una proteína que forma trímeros y que tiene cuatro dominos: un *insulin-like growth factor binding*, un inhibidor de serina proteasa tipo Kazal, una serina proteasa y un dominio PDZ (Nie et al., 2003). Se sabe que induce la apoptosis mediada por mitocondria y que puede actuar

como supresor tumoral (Glaza et al., 2015; Zhao, Zhang, Zhang, Feng, & Qu, 2016). Por ejemplo, se ha encontrado que el nivel de expresión de HTRA3 decrece con el grado de cáncer de endometrio (Bowden, Di Nezza-Cossens, Jobling, Salamonsen, & Nie, 2006). Además, su expresión se encuentra asociada a la adquisición de un fenotipo invasivo de cáncer oral (Moriya et al., 2015). Por otra parte, su expresión está asociada al desarrollo placentar y a la implantación de embriones (Bowden et al., 2006; Nie et al., 2006).

HTRA4 es una proteína secretada que tiene un dominio serina proteasa, un dominio PDZ y un dominio N terminal de unión a IGF (Chen et al., 2014; Clausen, Kaiser, Huber, & Ehrmann, 2011). Su expresión está prácticamente restringida a la placenta y se ha encontrado que esta se sobre-expresa en la placenta y el suero de pacientes durante el desarrollo temprano de preeclampsia (Chen et al., 2014; Inagaki et al., 2012). Además, se sabe que es crucial para la diferenciación del EVT y que promueve el proceso de invasión del trofoblasto (Chen et al., 2014; Wang et al., 2012).

Además, se detectaron ambos genes de la familia FAM111 “*Family with sequence similarity III*”, los cuales no tienen una función clara definida. Una de estas, FAM111A es una proteasa similar a la tripsina de 611 aminoácidos. Se sabe que una mutación en FAM111A está relacionada al síndrome de Kenny-Caffey y osteocraniostenosis (OCS), condiciones genéticas caracterizadas por hipoparatiroidismo y desarrollo esquelético deficiente (Unger et al., 2013). FAM111A además es un factor de replicación para el PCNA (*Proliferating cell nuclear antigen loading*) (Alabert et al., 2014)

FAM111B hasta ahora tiene una función desconocida. Tiene un dominio serina proteasa similar a tripsina conservado. Ciertas mutaciones de FAM111B producen un fenotipo complejo afectando varios órganos como son el páncreas exocrino, la piel, el hígado, el musculo

Esquelético y los pulmones. Además, ciertas mutaciones en FAM111B están asociadas a disfunción exocrina del páncreas heredable y dominante (Seo et al., 2016). Además, se ha encontrado una mutación causal en FAM111B para fibrosis poikiloderma hereditaria (Mercier et al., 2013).

Adicionalmente, cabe destacar que en el presente trabajo se encontró que para las proteínas NS1 y NS5pol de DENV no existen proteínas humanas que sean significativamente similares estructuralmente. Esto implica que es bastante improbable que existan reacciones cruzadas con proteínas humanas al diseñar un fármaco para una de estas proteínas. De esta manera, basado en este parámetro, se recomienda que estas dos proteínas sean más estudiadas para este fin.

Finalmente, se encontró que en la base de datos de moléculas activas evaluadas, ninguno de los inhibidores actuales cumplió con el parámetro de selectividad propuesto por Sandeep y colaboradores (Sandeep et al., 2011). Dado que se consideró que este parámetro fue demasiado exigente, se disminuyó el número de desviaciones estándar a tomar en consideración en este parámetro. Una vez reducido este parámetro solo pocas moléculas resaltaron con una selectividad considerable. Esto implica que prácticamente ninguna de las moléculas reportadas en la literatura en la actualidad va a ser selectiva para la proteasa de DENV y por ende va a tener problemas asociados a reactividad cruzada con las proteínas anteriormente descritas. De todas maneras, los quimiotipos reportados en las moléculas 9, 56, 62, 63 y 65 pueden ser considerados como los más prometedores para ser utilizados como punto de partida para crear derivaciones que puedan generar inhibidores de la proteasa con mejor selectividad.

Conclusiones:

Se detectó que existen proteínas humanas estructuralmente similares a las proteínas no estructurales de DENV: NS3 Pro, NS3 Hel y NS5 MTasa.

Se encontró que no existen proteínas humanas estructuralmente similares a las proteínas no estructurales de DENV: NS1 y NS5 Pol.

Se encontró que algunas de las proteínas humanas estructuralmente similares a la NS3 Pro de DENV tienen cavidades que son sustancialmente similares al sitio activo de NS3 Pro.

Se detectó mediante *docking* que varios compuestos para los que existe una actividad reportada contra la NS3 Pro, tienen la capacidad de unirse de manera similar tanto al sitio activo de NS3 Pro como a las cavidades similares detectadas en las proteínas humanas.

Se encontró que varias de las proteínas estudiadas tienen funciones importantes celulares y que su disfunción se relaciona fuertemente a complicaciones médicas relevantes.

Se describió una metodología que hace posible detectar proteínas similares en todo el proteoma humano y que permite analizar selectivamente si ciertas cavidades dentro de estas son sustancialmente similares entre ellas.

Se encontró que los quimio-tipos reportados en la literatura con actividad contra la proteasa de DENV tienen un nivel bajo de selectividad.

Se propone que los quimio-tipos descritos en las moléculas 9, 56, 62, 63 y 65 sean empleados como punto de partida para descubrir nuevos derivados con mejor selectividad contra la proteasa de DENV.

Se propone además una lista simplificada de cavidades de proteínas humanas para ser empleadas en experimentos de cribado virtual para el descubrimiento de nuevos quimio-tipos con alta selectividad para la proteasa de DENV.

Recomendaciones:

Se recomienda extender el presente trabajo al estudio de cavidades dentro de la NS3 Pro diferentes a su sitio activo.

Se recomienda continuar con la presente investigación, aplicando la metodología descrita en el presente trabajo con las demás proteínas no estructurales de DENV para las cuales se encontraron proteínas humanas estructuralmente similares (NS3 Hel y NS5 MTasa).

Literatura citada:

- Aguilera-Pesantes, D., Robayo, L. E., Méndez, P. E., Mollocana, D., Marrero-Ponce, Y., Torres, F. J., & Méndez, M. A. (2017). Discovering key residues of dengue virus NS2b-NS3-protease: New binding sites for antiviral inhibitors design. *Biochemical and Biophysical Research Communications*, 492(4), 631–642. <https://doi.org/10.1016/j.bbrc.2017.03.107>
- Akey, D. L., Brown, W. C., Dutta, S., Konwerski, J., Jose, J., Jurkiw, T. J., ... Smith, J. L. (2014). Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science (New York, N.Y.)*, 343(6173), 881–885. <https://doi.org/10.1126/science.1247749>
- Alabert, C., Bukowski-Wills, J.-C., Lee, S.-B., Kustatscher, G., Nakamura, K., de Lima Alves, F., ... Groth, A. (2014). Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nature Cell Biology*, 16(3), 281–291. <https://doi.org/10.1038/ncb2918>
- Alomar, M. J. (2014). Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharmaceutical Journal*, 22(2), 83–94. <https://doi.org/10.1016/j.jsps.2013.02.003>
- Amineni, U., Pradhan, D., & Marisetty, H. (2010). In silico identification of common putative drug targets in *Leptospira interrogans*. *Journal of Chemical Biology*, 3(4), 165–173. <https://doi.org/10.1007/s12154-010-0039-1>
- An, J., Totrov, M., & Abagyan, R. (2005). Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Molecular & Cellular Proteomics*, 4(6), 752–761. <https://doi.org/10.1074/mcp.M400159-MCP200>

- Anand, P., Yeturu, K., & Chandra, N. (2012). PocketAnnotate: towards site-based function annotation. *Nucleic Acids Research*, *40*(W1), W400–W408.
<https://doi.org/10.1093/nar/gks421>
- Apweiler, R. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*, *32*(90001), 115D – 119. <https://doi.org/10.1093/nar/gkh131>
- Aravapalli, S., Lai, H., Teramoto, T., Alliston, K. R., Lushington, G. H., Ferguson, E. L., ... Groutas, W. C. (2012). Inhibitors of Dengue virus and West Nile virus proteases based on the aminobenzamide scaffold. *Bioorganic & Medicinal Chemistry*, *20*(13), 4140–4148.
<https://doi.org/10.1016/j.bmc.2012.04.055>
- Balasubramanian, A., Manzano, M., Teramoto, T., Pilankatta, R., & Padmanabhan, R. (2016). High-throughput screening for the identification of small-molecule inhibitors of the flaviviral protease. *Antiviral Research*, *134*, 6–16.
<https://doi.org/10.1016/j.antiviral.2016.08.014>
- Baldi, A., Luca, A. D., Morini, M., Battista, T., Felsani, A., Baldi, F., ... Paggi, M. G. (2002). The HtrA1 serine protease is down-regulated during human melanoma progression and represses growth of metastatic melanoma cells. *Oncogene*, *21*(43), 6684–6688.
<https://doi.org/10.1038/sj.onc.1205911>
- BBC. (2018). Philippines gripped by dengue vaccine fears.
- Behnam, M. A. M., Nitsche, C., Boldescu, V., & Klein, C. D. (2016). The Medicinal Chemistry of Dengue Virus. *Journal of Medicinal Chemistry*, *59*(12), 5622–5649.
<https://doi.org/10.1021/acs.jmedchem.5b01653>
- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., ... Zardecki, C. (2002). The Protein Data Bank. *Acta Crystallographica Section D*

- Biological Crystallography*, 58(6), 899–907.
<https://doi.org/10.1107/S0907444902003451>
- Bodenreider, C., Beer, D., Keller, T. H., Sonntag, S., Wen, D., Yap, L., ... Lim, S. P. (2009). A fluorescence quenching assay to discriminate between specific and nonspecific inhibitors of dengue virus protease. *Analytical Biochemistry*, 395(2), 195–204.
<https://doi.org/10.1016/j.ab.2009.08.013>
- Bollati, M., Alvarez, K., Assenberg, R., Baronti, C., Canard, B., Cook, S., ... Bolognesi, M. (2010). Structure and functionality in flavivirus NS-proteins: Perspectives for drug design. *Antiviral Research*, 87(2), 125–148.
<https://doi.org/10.1016/j.antiviral.2009.11.009>
- Boonyasuppayakorn, S., Reichert, E. D., Manzano, M., Nagarajan, K., & Padmanabhan, R. (2014). Amodiaquine, an antimalarial drug, inhibits dengue virus type 2 replication and infectivity. *Antiviral Research*, 106, 125–134.
<https://doi.org/10.1016/j.antiviral.2014.03.014>
- Bowden, M. A., Di Nezza-Cossens, L. A., Jobling, T., Salamonsen, L. A., & Nie, G. (2006). Serine proteases HTRA1 and HTRA3 are down-regulated with increasing grades of human endometrial cancer. *Gynecologic Oncology*, 103(1), 253–260.
<https://doi.org/10.1016/j.ygyno.2006.03.006>
- Braga, R. C., Alves, V. M., Silva, M. F. B., Muratov, E., Fourches, D., Tropsha, A., & Andrade, C. H. (2014). Tuning HERG out: antitarget QSAR models for drug development. *Current Topics in Medicinal Chemistry*, 14(11), 1399–1415.
- Brecher, M., Li, Z., Liu, B., Zhang, J., Koetzner, C. A., Alifarag, A., ... Li, H. (2017). A conformational switch high-throughput screening assay and allosteric inhibition of the

- flavivirus NS2B-NS3 protease. *PLOS Pathogens*, 13(5), e1006411.
<https://doi.org/10.1371/journal.ppat.1006411>
- Bulteau, A.-L., & Bayot, A. (2011). Mitochondrial proteases and cancer. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1807(6), 595–601.
<https://doi.org/10.1016/j.bbabi.2010.12.011>
- Cabarcas-Montalvo, M., Maldonado-Rojas, W., Montes-Grajales, D., Bertel-Sevilla, A., Wagner-Döbler, I., Sztajer, H., ... Olivero-Verbel, J. (2016). Discovery of antiviral molecules for dengue: In silico search and biological evaluation. *European Journal of Medicinal Chemistry*, 110, 87–97. <https://doi.org/10.1016/j.ejmech.2015.12.030>
- Castillo Ramirez, J. A., & Urcuqui-Inchima, S. (2015). Dengue Virus Control of Type I IFN Responses: A History of Manipulation and Control. *Journal of Interferon & Cytokine Research*, 35(6), 421–430. <https://doi.org/10.1089/jir.2014.0129>
- Chandramouli, S., Joseph, J. S., Daudenarde, S., Gatchalian, J., Cornillez-Ty, C., & Kuhn, P. (2010). Serotype-Specific Structural Differences in the Protease-Cofactor Complexes of the Dengue Virus Family. *Journal of Virology*, 84(6), 3059–3067.
<https://doi.org/10.1128/JVI.02044-09>
- Chao, B., Tong, X.-K., Tang, W., Li, D.-W., He, P.-L., Garcia, J.-M., ... Hu, Y.-H. (2012). Discovery and Optimization of 2,4-Diaminoquinazoline Derivatives as a New Class of Potent Dengue Virus Inhibitors. *Journal of Medicinal Chemistry*, 55(7), 3135–3143.
<https://doi.org/10.1021/jm2015952>
- ChemAxon. (2014). MarvinSketch (Version 6.2.2).
- Chen, C.-K., Leung, S. S. F., Guilbert, C., Jacobson, M. P., McKerrow, J. H., & Podust, L. M. (2010). Structural Characterization of CYP51 from *Trypanosoma cruzi* and *Trypanosoma*

- brucei Bound to the Antifungal Drugs Posaconazole and Fluconazole. *PLoS Neglected Tropical Diseases*, 4(4), e651. <https://doi.org/10.1371/journal.pntd.0000651>
- Chen, Y.-Y., Chuang, P.-Y., Chen, C.-P., Chiu, Y.-H., Lo, H.-F., Cheong, M.-L., ... Chen, H. (2014). Functional Antagonism between High Temperature Requirement Protein A (HtrA) Family Members Regulates Trophoblast Invasion. *Journal of Biological Chemistry*, 289(33), 22958–22968. <https://doi.org/10.1074/jbc.M114.576744>
- Chun, Y. (2009). *PaDEL-ADV*.
- Clausen, T., Kaiser, M., Huber, R., & Ehrmann, M. (2011). HTRA proteases: regulated proteolysis in protein quality control. *Nature Reviews Molecular Cell Biology*, 12(3), 152–162. <https://doi.org/10.1038/nrm3065>
- Cregar-Hernandez, L., Jiao, G.-S., Johnson, A. T., Lehrer, A. T., Wong, T. A. S., & Margosiak, S. A. (2011). Small molecule pan-dengue and West Nile virus NS3 protease inhibitors. *Antiviral Chemistry and Chemotherapy*, 21(5), 209–218. <https://doi.org/10.3851/IMP1767>
- Das, S., Krein, M. P., & Breneman, C. M. (2010). PESDserv: a server for high-throughput comparison of protein binding site surfaces. *Bioinformatics*, 26(15), 1913–1914. <https://doi.org/10.1093/bioinformatics/btq288>
- de la Cruz, L., Nguyen, T. H. D., Ozawa, K., Shin, J., Graham, B., Huber, T., & Otting, G. (2011). Binding of Low Molecular Weight Inhibitors Promotes Large Conformational Changes in the Dengue Virus NS2B-NS3 Protease: Fold Analysis by Pseudocontact Shifts. *Journal of the American Chemical Society*, 133(47), 19205–19215. <https://doi.org/10.1021/ja208435s>
- DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool, 40(1), 82–92.

- Demircioglu, F. E., Sosa, B. A., Ingram, J., Ploegh, H. L., & Schwartz, T. U. (2016). Structures of TorsinA and its disease-mutant complexed with an activator reveal the molecular basis for primary dystonia. *ELife*, 5. <https://doi.org/10.7554/eLife.17983>
- Deng, J., Li, N., Liu, H., Zuo, Z., Liew, O. W., Xu, W., ... Zhu, W. (2012). Discovery of Novel Small Molecule Inhibitors of Dengue Viral NS2B-NS3 Protease Using Virtual Screening and Scaffold Hopping. *Journal of Medicinal Chemistry*, 55(14), 6278–6293. <https://doi.org/10.1021/jm300146f>
- Depta, J. P. H., & Pichler, W. J. (2003). Cross-reactivity with drugs at the T cell level. *Current Opinion in Allergy and Clinical Immunology*, 3(4), 261–267. <https://doi.org/10.1097/01.all.0000083951.99396.30>
- Desaphy, J., Azdimousa, K., Kellenberger, E., & Rognan, D. (2012). Comparison and Druggability Prediction of Protein–Ligand Binding Sites from Pharmacophore-Annotated Cavity Shapes. *Journal of Chemical Information and Modeling*, 52(8), 2287–2299. <https://doi.org/10.1021/ci300184x>
- Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G., & Little, E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science (New York, N.Y.)*, 271(5248), 470–477.
- Edeling, M. A., Diamond, M. S., & Fremont, D. H. (2014). Structural basis of Flavivirus NS1 assembly and antibody recognition. *Proceedings of the National Academy of Sciences*, 111(11), 4285–4290. <https://doi.org/10.1073/pnas.1322036111>
- Edwards, D., Høyer-Hansen, G., Blasi, F., & Sloane, B. F. (Eds.). (2008). *The Cancer Degradome*. New York, NY: Springer New York. <https://doi.org/10.1007/978-0-387-69057-5>

- Erbel, P., Schiering, N., D'Arcy, A., Renatus, M., Kroemer, M., Lim, S. P., ... Hommel, U. (2006). Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nature Structural & Molecular Biology*, *13*(4), 372–373. <https://doi.org/10.1038/nsmb1073>
- Fex Svenningsen, Å., Löring, S., Sørensen, A. L., Huynh, H. U. B., Hjärresen, S., Martin, N., ... Benedikz, E. (2017). Macrophage migration inhibitory factor (MIF) modulates trophic signaling through interaction with serine protease HTRA1. *Cellular and Molecular Life Sciences*, *74*(24), 4561–4572. <https://doi.org/10.1007/s00018-017-2592-z>
- Galiano, V., Garcia-Valtanen, P., Micol, V., & Encinar, J. A. (2016). Looking for inhibitors of the dengue virus NS5 RNA-dependent RNA-polymerase using a molecular docking approach. *Drug Design, Development and Therapy, Volume 10*, 3163–3181. <https://doi.org/10.2147/DDDT.S117369>
- Gan, C. S., Lee, T. K., Heh, C. H., Rahman, N., Noor, R., & Othman, S. (2017). The synthetic molecules YK51 and YK73 attenuate replication of dengue virus serotype 2, *34*(1), 270–283.
- Gan, E. S., Cheong, W. F., Chan, K. R., Ong, E. Z., Chai, X., Tan, H. C., ... Ooi, E. E. (2017). Hypoxia enhances antibody-dependent dengue virus infection. *The EMBO Journal*, *36*(10), 1348–1363. <https://doi.org/10.15252/emj.201695642>
- Ganesh, V. K., Muller, N., Judge, K., Luan, C.-H., Padmanabhan, R., & Murthy, K. H. M. (2005). Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. *Bioorganic & Medicinal Chemistry*, *13*(1), 257–264. <https://doi.org/10.1016/j.bmc.2004.09.036>

- Gao, M., & Skolnick, J. (2013). APoc: large-scale identification of similar protein pockets. *Bioinformatics*, 29(5), 597–604. <https://doi.org/10.1093/bioinformatics/btt024>
- García-Jacas, C. R., Marrero-Ponce, Y., Acevedo-Martínez, L., Barigye, S. J., Valdés-Martín, J. R., & Contreras-Torres, E. (2014). QuBiLS-MIDAS: A parallel free-software for molecular descriptors computation based on multilinear algebraic maps: Software News and Updates. *Journal of Computational Chemistry*, 35(18), 1395–1409. <https://doi.org/10.1002/jcc.23640>
- Geiss, B. J., Thompson, A. A., Andrews, A. J., Sons, R. L., Gari, H. H., Keenan, S. M., & Peersen, O. B. (2009). Analysis of Flavivirus NS5 Methyltransferase Cap Binding. *Journal of Molecular Biology*, 385(5), 1643–1654. <https://doi.org/10.1016/j.jmb.2008.11.058>
- Glaza, P., Osipiuk, J., Wenta, T., Zurawa-Janicka, D., Jarzab, M., Lesner, A., ... Lipinska, B. (2015). Structural and Functional Analysis of Human HtrA3 Protease and Its Subdomains. *PLOS ONE*, 10(6), e0131142. <https://doi.org/10.1371/journal.pone.0131142>
- Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: applications of AutoDock. *Journal of Molecular Recognition: JMR*, 9(1), 1–5. [https://doi.org/10.1002/\(SICI\)1099-1352\(199601\)9:1<::AID-JMR241>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1099-1352(199601)9:1<::AID-JMR241>3.0.CO;2-6)
- Gray, C. W., Ward, R. V., Karran, E., Turconi, S., Rowles, A., Viglienghi, D., ... Creasy, C. L. (2000). Characterization of human HtrA2, a novel serine protease involved in the mammalian cellular stress response. *European Journal of Biochemistry*, 267(18), 5699–5710.

- Gubler, D. J. (2002). The Global Emergence/Resurgence of Arboviral Diseases As Public Health Problems. *Archives of Medical Research*, 33(4), 330–342. [https://doi.org/10.1016/S0188-4409\(02\)00378-8](https://doi.org/10.1016/S0188-4409(02)00378-8)
- Guha, R., & Willighagen, E. (2012). A survey of quantitative descriptions of molecular structure. *Current Topics in Medicinal Chemistry*, 12(18), 1946–1956.
- Handa, S., Paul, B. G., Miller, J. F., Valentine, D. L., & Ghosh, P. (2016). Conservation of the C-type lectin fold for accommodating massive sequence variation in archaeal diversity-generating retroelements. *BMC Structural Biology*, 16(1). <https://doi.org/10.1186/s12900-016-0064-6>
- Hasan, S., Daugelat, S., Rao, P. S. S., & Schreiber, M. (2006). Prioritizing Genomic Drug Targets in Pathogens: Application to Mycobacterium tuberculosis. *PLoS Computational Biology*, 2(6), e61. <https://doi.org/10.1371/journal.pcbi.0020061>
- Heh, C. H., Othman, R., Buckle, M. J. C., Sharifuddin, Y., Yusof, R., & Rahman, N. A. (2013). Rational Discovery of Dengue Type 2 Non-Competitive Inhibitors. *Chemical Biology & Drug Design*, 82(1), 1–11. <https://doi.org/10.1111/cbdd.12122>
- Huang, Z. (Ed.). (2007). *Drug discovery research: new frontiers in the post-genomic era*. Hoboken, N.J: Wiley-Interscience.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Hussein, H. A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., & Camproux, A.-C. (2015). PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Research*, 43(W1), W436–W442. <https://doi.org/10.1093/nar/gkv462>

- Inagaki, A., Nishizawa, H., Ota, S., Suzuki, M., Inuzuka, H., Miyamura, H., ... Udagawa, Y. (2012). Upregulation of HtrA4 in the placentas of patients with severe pre-eclampsia. *Placenta*, 33(11), 919–926. <https://doi.org/10.1016/j.placenta.2012.08.003>
- Jianxin Wang, Min Li, Chen, J., & Yi Pan. (2011). A Fast Hierarchical Clustering Algorithm for Functional Modules Discovery in Protein Interaction Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(3), 607–620. <https://doi.org/10.1109/TCBB.2010.75>
- Kellenberger, E., Schalon, C., & Rognan, D. (2008). How to Measure the Similarity Between Protein Ligand-Binding Sites?, (4), 1.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Kiat, T. S., Pippen, R., Yusof, R., Ibrahim, H., Khalid, N., & Rahman, N. A. (2006). Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. *Bioorganic & Medicinal Chemistry Letters*, 16(12), 3337–3340. <https://doi.org/10.1016/j.bmcl.2005.12.075>
- Kinnings, S. L., & Jackson, R. M. (2009). Binding Site Similarity Analysis for the Functional Classification of the Protein Kinase Family. *Journal of Chemical Information and Modeling*, 49(2), 318–329. <https://doi.org/10.1021/ci800289y>
- Kinoshita, K., Murakami, Y., & Nakamura, H. (2007). eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape. *Nucleic Acids Research*, 35(Web Server), W398–W402. <https://doi.org/10.1093/nar/gkm351>

- Knehans, T., Schüller, A., Doan, D. N., Nacro, K., Hill, J., Güntert, P., ... Vasudevan, S. G. (2011). Structure-guided fragment-based in silico drug design of dengue protease inhibitors. *Journal of Computer-Aided Molecular Design*, 25(3), 263–274.
<https://doi.org/10.1007/s10822-011-9418-0>
- Koh-Stenta, X., Joy, J., Wang, S. F., Kwek, P., Wee, J., Wan, K. F., ... Nacro, K. (2015). Identification of covalent active site inhibitors of dengue virus protease. *Drug Design, Development and Therapy*, 6389. <https://doi.org/10.2147/DDDT.S94207>
- Konc, J., & Janezic, D. (2010). ProBiS: a web server for detection of structurally similar protein binding sites. *Nucleic Acids Research*, 38(Web Server), W436–W440.
<https://doi.org/10.1093/nar/gkq479>
- Konc, Janez, & Janežič, D. (2014). Binding site comparison for function prediction and pharmaceutical discovery. *Current Opinion in Structural Biology*, 25, 34–39.
<https://doi.org/10.1016/j.sbi.2013.11.012>
- Küçükgoze, G., Terao, M., Garattini, E., & Leimkühler, S. (2017). Direct Comparison of the Enzymatic Characteristics and Superoxide Production of the Four Aldehyde Oxidase Enzymes Present in Mouse. *Drug Metabolism and Disposition*, 45(8), 947–955.
<https://doi.org/10.1124/dmd.117.075937>
- Kurochkin, I. V., Mizuno, Y., Konagaya, A., Sakaki, Y., Schönbach, C., & Okazaki, Y. (2007). Novel peroxisomal protease Tysnd1 processes PTS1- and PTS2-containing enzymes involved in β -oxidation of fatty acids. *The EMBO Journal*, 26(3), 835–845.
<https://doi.org/10.1038/sj.emboj.7601525>

- La, D., Esquivel-Rodríguez, J., Venkatraman, V., Li, B., Sael, L., Ueng, S., ... Kihara, D. (2009). 3D-SURFER: software for high-throughput protein surface comparison and analysis. *Bioinformatics*, 25(21), 2843–2844. <https://doi.org/10.1093/bioinformatics/btp542>
- Lahti, J. L., Tang, G. W., Capriotti, E., Liu, T., & Altman, R. B. (2012). Bioinformatics and variability in drug response: a protein structural perspective. *Journal of The Royal Society Interface*, 9(72), 1409–1437. <https://doi.org/10.1098/rsif.2011.0843>
- Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1), 168. <https://doi.org/10.1186/1471-2105-10-168>
- Le, T. V., Thi Tuyet Van, N., Hoang Quan, N., & Tho Duoc, P. (2017). Phylogeny of Dengue virus type 2 isolated in the Central Highlands, Vietnam. *Revista de Biología Tropical*, 65(2). <https://doi.org/10.15517/rbt.v65i2.23535>
- Lee, D., Redfern, O., & Orengo, C. (2007). Predicting protein function from sequence and structure. *Nature Reviews Molecular Cell Biology*, 8(12), 995–1005. <https://doi.org/10.1038/nrm2281>
- Leinweber, M., Fober, T., Strickert, M., Baumgartner, L., Klebe, G., Freisleben, B., & Hullermeier, E. (2016a). CavSimBase: A Database for Large Scale Comparison of Protein Binding Sites. *IEEE Transactions on Knowledge and Data Engineering*, 28(6), 1423–1434. <https://doi.org/10.1109/TKDE.2016.2520484>
- Leinweber, M., Fober, T., Strickert, M., Baumgartner, L., Klebe, G., Freisleben, B., & Hullermeier, E. (2016b). CavSimBase: A Database for Large Scale Comparison of Protein Binding Sites. *IEEE Transactions on Knowledge and Data Engineering*, 28(6), 1423–1434. <https://doi.org/10.1109/TKDE.2016.2520484>

- Li, L., Basavannacharya, C., Chan, K. W. K., Shang, L., Vasudevan, S. G., & Yin, Z. (2015). Structure-guided Discovery of a Novel Non-peptide Inhibitor of Dengue Virus NS2B-NS3 Protease. *Chemical Biology & Drug Design*, 86(3), 255–264. <https://doi.org/10.1111/cbdd.12500>
- Li, Z., Sakamuru, S., Huang, R., Brecher, M., Koetzner, C. A., Zhang, J., ... Li, H. (2018). Erythrosin B is a potent and broad-spectrum orthosteric inhibitor of the flavivirus NS2B-NS3 protease. *Antiviral Research*, 150, 217–225. <https://doi.org/10.1016/j.antiviral.2017.12.018>
- Lim, S. V., Rahman, M. B. A., & Tejo, B. A. (2011). Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. *BMC Bioinformatics*, 12(Suppl 13), S24. <https://doi.org/10.1186/1471-2105-12-S13-S24>
- Liu, H., Wu, R., Sun, Y., Ye, Y., Chen, J., Luo, X., ... Liu, H. (2014). Identification of novel thiadiazoloacrylamide analogues as inhibitors of dengue-2 virus NS2B/NS3 protease. *Bioorganic & Medicinal Chemistry*, 22(22), 6344–6352. <https://doi.org/10.1016/j.bmc.2014.09.057>
- Low, J. G. H., Ooi, E. E., & Vasudevan, S. G. (2017). Current Status of Dengue Therapeutics Research and Development. *The Journal of Infectious Diseases*, 215(suppl_2), S96–S102. <https://doi.org/10.1093/infdis/jiw423>
- Luo, D., Wei, N., Doan, D. N., Paradkar, P. N., Chong, Y., Davidson, A. D., ... Vasudevan, S. G. (2010). Flexibility between the Protease and Helicase Domains of the Dengue Virus NS3 Protein Conferred by the Linker Region and Its Functional Implications. *Journal of Biological Chemistry*, 285(24), 18817–18827. <https://doi.org/10.1074/jbc.M109.090936>

- Luo, D., Xu, T., Watson, R. P., Scherer-Becker, D., Sampath, A., Jahnke, W., ... Lescar, J. (2008). Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *The EMBO Journal*, 27(23), 3209–3219. <https://doi.org/10.1038/emboj.2008.232>
- Lytras, M. D., & Papadopoulou, P. (Eds.). (2018). *Applying Big Data Analytics in Bioinformatics and Medicine*: IGI Global. <https://doi.org/10.4018/978-1-5225-2607-0>
- MacNeill, S. (2014). Identification of a Candidate Rad1 Subunit for the Kinetoplastid 9-1-1 (Rad9-Hus1-Rad1) Complex. *Biology*, 3(4), 922–927. <https://doi.org/10.3390/biology3040922>
- Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics*, 21(16), 3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Mahmood, K., Jameel, Y., Aslam, H., & Tahir, M. (2009). INCIDENCE OF DENGUE HAEMORRHAGIC FEVER IN LOCAL POPULATION OF LAHORE, PAKISTAN, 25(3), 93–96.
- Martins, L. M., Morrison, A., Klupsch, K., Fedele, V., Moiso, N., Teismann, P., ... Downward, J. (2004). Neuroprotective Role of the Reaper-Related Serine Protease HtrA2/Omi Revealed by Targeted Deletion in Mice. *Molecular and Cellular Biology*, 24(22), 9848–9862. <https://doi.org/10.1128/MCB.24.22.9848-9862.2004>
- Mendgen, T., Steuer, C., & Klein, C. D. (2012). Privileged Scaffolds or Promiscuous Binders: A Comparative Study on Rhodanines and Related Heterocycles in Medicinal Chemistry. *Journal of Medicinal Chemistry*, 55(2), 743–753. <https://doi.org/10.1021/jm201243p>
- Mercier, S., Küry, S., Shaboodien, G., Houniet, D. T., Khumalo, N. P., Bou-Hanna, C., ... Mayosi, B. M. (2013). Mutations in FAM111B Cause Hereditary Fibrosing Poikiloderma

- with Tendon Contracture, Myopathy, and Pulmonary Fibrosis. *The American Journal of Human Genetics*, 93(6), 1100–1107. <https://doi.org/10.1016/j.ajhg.2013.10.013>
- Mills, C. L., Beuning, P. J., & Ondrechen, M. J. (2015). Biochemical functional predictions for protein structures of unknown or uncertain function. *Computational and Structural Biotechnology Journal*, 13, 182–191. <https://doi.org/10.1016/j.csbj.2015.02.003>
- Mirza, S. B., Salmas, R. E., Fatmi, M. Q., & Durdagi, S. (2016). Virtual screening of eighteen million compounds against dengue virus: Combined molecular docking and molecular dynamics simulations study. *Journal of Molecular Graphics and Modelling*, 66, 99–107. <https://doi.org/10.1016/j.jmgm.2016.03.008>
- Mizuno, Y., Ninomiya, Y., Nakachi, Y., Iseki, M., Iwasa, H., Akita, M., ... Okazaki, Y. (2013). Tysnd1 Deficiency in Mice Interferes with the Peroxisomal Localization of PTS2 Enzymes, Causing Lipid Metabolic Abnormalities and Male Infertility. *PLoS Genetics*, 9(2), e1003286. <https://doi.org/10.1371/journal.pgen.1003286>
- Moriya, Y., Uzawa, N., Morita, T., Mogushi, K., Miyaguchi, K., Takahashi, K.-I., ... Harada, K. (2015). The high-temperature requirement factor A3 (HtrA3) is associated with acquisition of the invasive phenotype in oral squamous cell carcinoma cells. *Oral Oncology*, 51(1), 84–89. <https://doi.org/10.1016/j.oraloncology.2014.10.001>
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791. <https://doi.org/10.1002/jcc.21256>
- Morrison, J., Aguirre, S., & Fernandez-Sesma, A. (2012). Innate Immunity Evasion by Dengue Virus. *Viruses*, 4(12), 397–413. <https://doi.org/10.3390/v4030397>

- Moult, J., Fidelis, K., Kryshchak, A., Schwede, T., & Tramontano, A. (2014). Critical assessment of methods of protein structure prediction (CASP) - round x: Critical Assessment of Structure Prediction. *Proteins: Structure, Function, and Bioinformatics*, 82, 1–6. <https://doi.org/10.1002/prot.24452>
- Mueller, N. H., Pattabiraman, N., Ansarah-Sobrinho, C., Viswanathan, P., Pierson, T. C., & Padmanabhan, R. (2008). Identification and Biochemical Characterization of Small-Molecule Inhibitors of West Nile Virus Serine Protease by a High-Throughput Screen. *Antimicrobial Agents and Chemotherapy*, 52(9), 3385–3393. <https://doi.org/10.1128/AAC.01508-07>
- Ng, T. K., Liang, X. Y., Lu, F., Liu, D. T., Yam, G. H., Ma, L., ... Pang, C. P. (2017). Protective effects of an HTRA1 insertion–deletion variant against age-related macular degeneration in the Chinese populations. *Laboratory Investigation*, 97(1), 43–52. <https://doi.org/10.1038/labinvest.2016.117>
- Nguyen, T., Lee, S., Wang, H.-K., Chen, H.-Y., Wu, Y.-T., Lin, S., ... Kim, D. (2013). In Vitro Evaluation of Novel Inhibitors against the NS2B-NS3 Protease of Dengue Fever Virus Type 4. *Molecules*, 18(12), 15600–15612. <https://doi.org/10.3390/molecules181215600>
- Nie, G., Li, Y., Hale, K., Okada, H., Manuelpillai, U., Wallace, E. M., & Salamonsen, L. A. (2006). Serine Peptidase HTRA3 Is Closely Associated with Human Placental Development and Is Elevated in Pregnancy Serum. *Biology of Reproduction*, 74(2), 366–374. <https://doi.org/10.1095/biolreprod.105.047324>
- Nie, G.-Y., Hampton, A., Li, Y., Findlay, J. K., & Salamonsen, L. A. (2003). Identification and cloning of two isoforms of human high-temperature requirement factor A3 (HtrA3), characterization of its genomic structure and comparison of its tissue distribution with

- HtrA1 and HtrA2. *Biochemical Journal*, 371(1), 39–48.
<https://doi.org/10.1042/bj20021569>
- Nitsche, C., Holloway, S., Schirmeister, T., & Klein, C. D. (2014). Biochemistry and Medicinal Chemistry of the Dengue Virus Protease. *Chemical Reviews*, 114(22), 11348–11381.
<https://doi.org/10.1021/cr500233q>
- Nitsche, C., Steuer, C., & Klein, C. D. (2011). Arylcianoacrylamides as inhibitors of the Dengue and West Nile virus proteases. *Bioorganic & Medicinal Chemistry*, 19(24), 7318–7337.
<https://doi.org/10.1016/j.bmc.2011.10.061>
- Noble, C. G., Seh, C. C., Chao, A. T., & Shi, P. Y. (2012). Ligand-Bound Structures of the Dengue Virus Protease Reveal the Active Conformation. *Journal of Virology*, 86(1), 438–446. <https://doi.org/10.1128/JVI.06225-11>
- Okumoto, K., Kametani, Y., & Fujiki, Y. (2011). Two Proteases, Trypsin Domain-containing 1 (Tysnd1) and Peroxisomal Lon Protease (PsLon), Cooperatively Regulate Fatty Acid β -Oxidation in Peroxisomal Matrix. *Journal of Biological Chemistry*, 286(52), 44367–44379. <https://doi.org/10.1074/jbc.M111.285197>
- Oláhová, M., Thompson, K., Hardy, S. A., Barbosa, I. A., Besse, A., Anagnostou, M.-E., ... Taylor, R. W. (2017). Pathogenic variants in HTRA2 cause an early-onset mitochondrial syndrome associated with 3-methylglutaconic aciduria. *Journal of Inherited Metabolic Disease*, 40(1), 121–130. <https://doi.org/10.1007/s10545-016-9977-2>
- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733–D745.
<https://doi.org/10.1093/nar/gkv1189>

- Orengo, C. A., Jones, D. T., & Thornton, J. M. (1994). Protein superfamilies and domain superfolds. *Nature*, *372*(6507), 631–634. <https://doi.org/10.1038/372631a0>
- Othman, R., Othman, R., Baharuddin, A., Ramakrishnan, N. R., Abd Rahman, N., Yusof, R., & Karsani, S. A. (2017). Molecular Docking Studies of Selected Medicinal Drugs as Dengue Virus-2 Protease Inhibitors. *Sains Malaysiana*, *46*(10), 1865–1875. <https://doi.org/10.17576/jsm-2017-4610-25>
- Padmanabhan, R., & Groutas, W. C. (2010). Patent WO 2010/039534 A2.
- Pambudi, S., Kawashita, N., Phanthanawiboon, S., Omokoko, M. D., Masrinoul, P., Yamashita, A., ... Kurosu, T. (2013). A small compound targeting the interaction between nonstructural proteins 2B and 3 inhibits dengue virus replication. *Biochemical and Biophysical Research Communications*, *440*(3), 393–398. <https://doi.org/10.1016/j.bbrc.2013.09.078>
- Pathak, N., Lai, M.-L., Chen, W.-Y., Hsieh, B.-W., Yu, G.-Y., & Yang, J.-M. (2017). Pharmacophore anchor models of flaviviral NS3 proteases lead to drug repurposing for DENV infection. *BMC Bioinformatics*, *18*(S16). <https://doi.org/10.1186/s12859-017-1957-5>
- Perry, S. T., Buck, M. D., & Shresta, S. (2011). Better late than never: antivirals for dengue. *Expert Review of Anti-Infective Therapy*, *9*(7), 755–757. <https://doi.org/10.1586/eri.11.68>
- Powers, R., Copeland, J. C., Stark, J. L., Caprez, A., Guru, A., & Swanson, D. (2011). Searching the protein structure database for ligand-binding site similarities using CPASS v.2. *BMC Research Notes*, *4*(1), 17. <https://doi.org/10.1186/1756-0500-4-17>
- Rahman, A., Yusof, R., & Tat, T. K. (n.d.). Patent PI 20060021.

- Raut, R., Beesetti, H., Tyagi, P., Khanna, I., Jain, S. K., Jeankumar, V. U., ... Swaminathan, S. (2015). A small molecule inhibitor of dengue virus type 2 protease inhibits the replication of all four dengue virus serotypes in cell culture. *Virology Journal*, *12*(1), 16. <https://doi.org/10.1186/s12985-015-0248-x>
- Richter, S., Wenzel, A., Stein, M., Gabdoulline, R. R., & Wade, R. C. (2008). webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Research*, *36*(Web Server), W276–W280. <https://doi.org/10.1093/nar/gkn181>
- Robb, C. S., Nano, F. E., & Boraston, A. B. (2012). The Structure of the Conserved Type Six Secretion Protein TssL (DotU) from *Francisella novicida*. *Journal of Molecular Biology*, *419*(5), 277–283. <https://doi.org/10.1016/j.jmb.2012.04.003>
- Rodenhuis-Zybert, I. A., Wilschut, J., & Smit, J. M. (2010). Dengue virus life cycle: viral and host factors modulating infectivity. *Cellular and Molecular Life Sciences*, *67*(16), 2773–2786. <https://doi.org/10.1007/s00018-010-0357-z>
- Roemer, T., Jiang, B., Davison, J., Ketela, T., Veillette, K., Breton, A., ... Bussey, H. (2003). Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. *Molecular Microbiology*, *50*(1), 167–181.
- Rothan, H. A., Buckle, M. J., Ammar, Y. A., Mohammadjavad, P., Shatrah, O., Noorsaadah, A. R., & Rohana, Y. (2013). Study the antiviral activity of some derivatives of tetracycline and non-steroid anti inflammatory drugs towards dengue virus. *Tropical Biomedicine*, *30*(4), 681–690.
- Ruiz-Blanco, Y. B., Paz, W., Green, J., & Marrero-Ponce, Y. (2015). ProtDCal: A program to compute general-purpose-numerical descriptors for sequences and 3D-structures of proteins. *BMC Bioinformatics*, *16*(1). <https://doi.org/10.1186/s12859-015-0586-0>

- Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., ... Skehel, J. J. (2006). The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*, *443*(7107), 45–49.
<https://doi.org/10.1038/nature05114>
- Sahoo, M., Jena, L., Daf, S., & Kumar, S. (2016). Virtual Screening for Potential Inhibitors of NS3 Protein of Zika Virus. *Genomics & Informatics*, *14*(3), 104.
<https://doi.org/10.5808/GI.2016.14.3.104>
- Sandeep, G., Nagasree, K. P., Hanisha, M., & Kumar, M. M. K. (2011). AUDocker LE: A GUI for virtual screening with AUTODOCK Vina. *BMC Research Notes*, *4*(1), 445.
<https://doi.org/10.1186/1756-0500-4-445>
- Sarafianos, S. G., Das, K., Ding, J., Boyer, P. L., Hughes, S. H., & Arnold, E. (1999). Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *Chemistry & Biology*, *6*(5), R137-146.
- Seo, A., Walsh, T., Lee, M. K., Ho, P. A., Hsu, E. K., Sidbury, R., ... Shimamura, A. (2016). FAM111B Mutation Is Associated With Inherited Exocrine Pancreatic Dysfunction: *Pancreas*, *45*(6), 858–862. <https://doi.org/10.1097/MPA.0000000000000529>
- Shannon, P. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504.
<https://doi.org/10.1101/gr.1239303>
- Shiryaev, S. A., Cheltsov, A. V., Gawlik, K., Ratnikov, B. I., & Strongin, A. Y. (2011). Virtual Ligand Screening of the National Cancer Institute (NCI) Compound Library Leads to the Allosteric Inhibitory Scaffolds of the West Nile Virus NS3 Proteinase. *ASSAY and Drug Development Technologies*, *9*(1), 69–78. <https://doi.org/10.1089/adt.2010.0309>

- Singh, Biswas, D., & Jayaram, B. (2011). AADS - An Automated Active Site Identification, Docking, and Scoring Protocol for Protein Targets Based on Physicochemical Descriptors. *Journal of Chemical Information and Modeling*, 51(10), 2515–2527. <https://doi.org/10.1021/ci200193z>
- Singh, J., Kumar, M., Mansuri, R., Sahoo, G., & Deep, A. (2016). Inhibitor designing, virtual screening, and docking studies for methyltransferase: A potential target against dengue virus. *Journal of Pharmacy and Bioallied Sciences*, 8(3), 188. <https://doi.org/10.4103/0975-7406.171682>
- Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*, 21(7), 951–960. <https://doi.org/10.1093/bioinformatics/bti125>
- StatSoft, Inc. (2007). STATISTICA (data analysis software system) (Version 8.0). Retrieved from www.statsoft.com
- Steuer, C., Gege, C., Fischl, W., Heinonen, K. H., Bartenschlager, R., & Klein, C. D. (2011). Synthesis and biological evaluation of α -ketoamides as inhibitors of the Dengue virus protease with antiviral activity in cell-culture. *Bioorganic & Medicinal Chemistry*, 19(13), 4067–4074. <https://doi.org/10.1016/j.bmc.2011.05.015>
- Strauss, K. M., Martins, L. M., Plun-Favreau, H., Marx, F. P., Kautzmann, S., Berg, D., ... Krüger, R. (2005). Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Human Molecular Genetics*, 14(15), 2099–2111. <https://doi.org/10.1093/hmg/ddi215>
- Suzuki, Y., Takahashi-Niki, K., Akagi, T., Hashikawa, T., & Takahashi, R. (2004). Mitochondrial protease Omi/HtrA2 enhances caspase activation through multiple

- pathways. *Cell Death & Differentiation*, 11(2), 208–216.
<https://doi.org/10.1038/sj.cdd.4401343>
- Tiew, K.-C., Dou, D., Teramoto, T., Lai, H., Alliston, K. R., Lushington, G. H., ... Groutas, W. C. (2012). Inhibition of Dengue virus and West Nile virus proteases by click chemistry-derived benz[d]isothiazol-3(2H)-one derivatives. *Bioorganic & Medicinal Chemistry*, 20(3), 1213–1221. <https://doi.org/10.1016/j.bmc.2011.12.047>
- Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics*. Weinheim: Wiley-VCH.
- Tomlinson, S. M., Malmstrom, R. D., Russo, A., Mueller, N., Pang, Y.-P., & Watowich, S. J. (2009). Structure-based discovery of dengue virus protease inhibitors. *Antiviral Research*, 82(3), 110–114. <https://doi.org/10.1016/j.antiviral.2009.02.190>
- Tomlinson, S. M., & Watowich, S. J. (2011). Anthracene-based inhibitors of dengue virus NS2B–NS3 protease☆. *Antiviral Research*, 89(2), 127–135.
<https://doi.org/10.1016/j.antiviral.2010.12.006>
- Tossetta, G., Avellini, C., Licini, C., Giannubilo, S. R., Castellucci, M., & Marzioni, D. (2016). High temperature requirement A1 and fibronectin: two possible players in placental tissue remodelling. *European Journal of Histochemistry*, 60(4).
<https://doi.org/10.4081/ejh.2016.2724>
- Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. <https://doi.org/10.1002/jcc.21334>
- Tuiskunen Bäck, A., & Lundkvist, Å. (2013). Dengue viruses – an overview. *Infection Ecology & Epidemiology*, 3(1), 19839. <https://doi.org/10.3402/iee.v3i0.19839>

- Unger, S., Gónna, M. W., Le Béhec, A., Do Vale-Pereira, S., Bedeschi, M. F., Geiberger, S., ...
Superti-Furga, A. (2013). FAM111A Mutations Result in Hypoparathyroidism and
Impaired Skeletal Development. *The American Journal of Human Genetics*, 92(6), 990–
995. <https://doi.org/10.1016/j.ajhg.2013.04.020>
- Urbàn, L. (Ed.). (2015). *Antitargets and drug safety*. Weinheim: Wiley-VCH-Verl.
- Urias, R. W. P., Barigye, S. J., Marrero-Ponce, Y., García-Jacas, C. R., Valdes-Martín, J. R., &
Perez-Gimenez, F. (2015). IMMAN: free software for information theory-based
chemometric analysis. *Molecular Diversity*, 19(2), 305–319.
<https://doi.org/10.1007/s11030-014-9565-z>
- Vande Walle, L., Lamkanfi, M., & Vandenabeele, P. (2008). The mitochondrial serine protease
HtrA2/Omi: an overview. *Cell Death & Differentiation*, 15(3), 453–460.
<https://doi.org/10.1038/sj.cdd.4402291>
- Vannice, K. S., Durbin, A., & Hombach, J. (2016). Status of vaccine research and development
of vaccines for dengue. *Vaccine*, 34(26), 2934–2938.
<https://doi.org/10.1016/j.vaccine.2015.12.073>
- Viswanathan, U., Tomlinson, S. M., Fonner, J. M., Mock, S. A., & Watowich, S. J. (2014).
Identification of a Novel Inhibitor of Dengue Virus Protease through Use of a Virtual
Screening Drug Discovery Web Portal. *Journal of Chemical Information and Modeling*,
54(10), 2816–2825. <https://doi.org/10.1021/ci500531r>
- Wang, J., Zhong, J., Chen, G., Li, M., Wu, F., & Pan, Y. (2015). ClusterViz: A Cytoscape APP
for Cluster Analysis of Biological Network. *IEEE/ACM Transactions on Computational
Biology and Bioinformatics*, 12(4), 815–822.
<https://doi.org/10.1109/TCBB.2014.2361348>

- Wang, L.-J., Cheong, M.-L., Lee, Y.-S., Lee, M.-T., & Chen, H. (2012). High-Temperature Requirement Protein A4 (HtrA4) Suppresses the Fusogenic Activity of Syncytin-1 and Promotes Trophoblast Invasion. *Molecular and Cellular Biology*, 32(18), 3707–3717. <https://doi.org/10.1128/MCB.00223-12>
- Weskamp, N., Hüllermeier, E., & Klebe, G. (2009a). Merging chemical and biological space: Structural mapping of enzyme binding pocket space. *Proteins: Structure, Function, and Bioinformatics*, 76(2), 317–330. <https://doi.org/10.1002/prot.22345>
- Weskamp, N., Hüllermeier, E., & Klebe, G. (2009b). Merging chemical and biological space: Structural mapping of enzyme binding pocket space. *Proteins: Structure, Function, and Bioinformatics*, 76(2), 317–330. <https://doi.org/10.1002/prot.22345>
- World Health Organization. (2012). *Global strategy for dengue prevention and control, 2012-2020*. Geneva, Switzerland: World Health Organization. Retrieved from http://apps.who.int/iris/bitstream/10665/75303/1/9789241504034_eng.pdf
- Wu, H., Bock, S., Snitko, M., Berger, T., Weidner, T., Holloway, S., ... Bodem, J. (2015). Novel Dengue Virus NS2B/NS3 Protease Inhibitors. *Antimicrobial Agents and Chemotherapy*, 59(2), 1100–1109. <https://doi.org/10.1128/AAC.03543-14>
- Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Chene, P., ... Lescar, J. (2005). Structure of the Dengue Virus Helicase/Nucleoside Triphosphatase Catalytic Domain at a Resolution of 2.4 Å. *Journal of Virology*, 79(16), 10278–10288. <https://doi.org/10.1128/JVI.79.16.10278-10288.2005>
- Yang, C., Hu, H., Wu, R., Wu, S., Lee, S., Jiaang, W.-T., ... Yueh, A. (2014). A Novel Dengue Virus Inhibitor, BP13944, Discovered by High-Throughput Screening with Dengue Virus

- Replicon Cells Selects for Resistance in the Viral NS2B/NS3 Protease. *Antimicrobial Agents and Chemotherapy*, 58(1), 110–119. <https://doi.org/10.1128/AAC.01281-13>
- Yang, J., Roy, A., & Zhang, Y. (2012). BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, 41(D1), D1096–D1103. <https://doi.org/10.1093/nar/gks966>
- Yang, J., Roy, A., & Zhang, Y. (2013). Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29(20), 2588–2595. <https://doi.org/10.1093/bioinformatics/btt447>
- Yap, T. L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., ... Lescar, J. (2007). Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution. *Journal of Virology*, 81(9), 4753–4765. <https://doi.org/10.1128/JVI.02283-06>
- Yildiz, M., Ghosh, S., Bell, J. A., Sherman, W., & Hardy, J. A. (2013). Allosteric Inhibition of the NS2B-NS3 Protease from Dengue Virus. *ACS Chemical Biology*, 8(12), 2744–2752. <https://doi.org/10.1021/cb400612h>
- Zeidler, J., Fernandes-Siqueira, L., Barbosa, G., & Da Poian, A. (2017). Non-Canonical Roles of Dengue Virus Non-Structural Proteins. *Viruses*, 9(12), 42. <https://doi.org/10.3390/v9030042>
- Zhang, B., Jaroszewski, L., Rychlewski, L., & Godzik, A. (1997). Similarities and differences between nonhomologous proteins with similar folds: evaluation of threading strategies. *Folding and Design*, 2(5), 307–317. [https://doi.org/10.1016/S1359-0278\(97\)00042-4](https://doi.org/10.1016/S1359-0278(97)00042-4)

- Zhao, J., Zhang, J., Zhang, X., Feng, M., & Qu, J. (2016). High temperature requirement A3 (HTRA3) expression predicts postoperative recurrence and survival in patients with non-small-cell lung cancer. *Oncotarget*, 7(26). <https://doi.org/10.18632/oncotarget.9173>
- Zhou, Z., Khaliq, M., Suk, J.-E., Patkar, C., Li, L., Kuhn, R. J., & Post, C. B. (2008). Antiviral Compounds Discovered by Virtual Screening of Small-Molecule Libraries against Dengue Virus E Protein. *ACS Chemical Biology*, 3(12), 765–775. <https://doi.org/10.1021/cb800176t>
- Zou, B., Chan, W. L., Ding, M., Leong, S. Y., Nilar, S., Seah, P. G., ... Smith, P. W. (2015). Lead Optimization of Spiropyrazolopyridones: A New and Potent Class of Dengue Virus Inhibitors. *ACS Medicinal Chemistry Letters*, 6(3), 344–348. <https://doi.org/10.1021/ml500521r>

Tablas:

Tabla 1 Resumen de información a lo largo del flujo de trabajo de proteínas no estructurales de DENV para las que se encontraron proteínas humanas estructuralmente similares. BackPhyre-Refseq se refiere a la cantidad de proteínas humanas similares a la proteína no estructural de DENV (Score de confianza > 90%). UniprotKB se refiere a la cantidad de proteínas humanas que pudieron ser rastreadas a un ID actual. Modeladas se refiere a las proteínas que pudieron ser procesadas por Phyre2 para generar un modelo de homología adecuado. AADS se refiere a la cantidad de estructuras generadas por homología que pudieron ser analizadas por AADS para la detección de cavidades y cavidades se refiere a la cantidad de cavidades detectadas y extraídas por el método de AADS en las proteínas humanas analizadas.

	NS3 Pro	NS3 Hel	NS5 Mtasa
BackPhyre-Refseq	10	410	104
UniprotKB	7	289	86
Modeladas	7	285	86
AADS	7	254	74
Cavidades	84	5500	1673

Tabla 2 identificación del porcentaje de proteínas humanas con función y sustrato similar a una proteína no estructural de DENV (basado en sus Gene Ontology) y de la proporción de sitios de unión al sustrato y sitios activos desconocidos explícitamente en UniprotKB.

	NS3 Pro	NS3 hel	NS5 Mtasa
Función general similar	100%	39%	94%
Sustrato similar	100%	43%	43%
Sitio de unión desconocido	100%	89%	40%
Sitio activo desconocido	38%	98%	67%

Tabla 3 Clusters de cavidades detectadas por más de un método de agrupamiento en representación AADS – ProtDCal. Se muestran las cavidades que forman los clusters consenso entre los métodos de agrupamiento. Se muestra además la proteína de la cual proviene cada cavidad detectada.

Cluster	Cavity	Joining	K-means	PCA	Protein
1	DENV1_NS3Pro_3u1i_CAV2	x	x	x	NS3pro
	DENV2_NS3Pro_2fom_CAV3	x	x	x	NS3pro
	Cavity_1_VTPALIGRWEFDS	x	x	x	TYSND1
	Cavity_2_TVCGWRAELFSI	x	x	x	TYSND1
2	DENV3_NS3Pro_3u1i_CAV6	x	x	x	NS3pro
	Cavity_7_HSGRVCWETLF	x	x	x	TYSND1
	Cavity_7_WRFQVLPSHEYG	x	x	x	HTRA4
	Cavity_6_SWTAHFVLGYQ	x	x	x	FAM111A
3	DENV1_NS3pro_3l6p_CAV3	x	x	x	NS3pro
	DENV4_NS3Pro_2whx_CAV5	x	x	x	NS3pro
	Cavity_14_FLIDNEVRGS	x	x	x	HTRA2
	Cavity_2_LFPAYENGVDSIKHMT	x	x		HTRA2
	Cavity_2_VYINSFCLDTQHPAG	x	x	x	FAM111B
	Cavity_4_TQVNGPMFELDISRH		x	x	TYSND1
Cavity_5_TGQFYSHKVAIPDEC	x	x	x	FAM111A	
4	DENV4_NS3Pro_2whx_CAV2	x	x	x	NS3pro
	Cavity_7_ETFKHYRSLAVIN	x	x	x	HTRA1
	Cavity_8_GFSLDPRTYAIVME	x	x	x	HTRA2
5	DENV3_NS3Pro_3u1i_CAV8	x	x		NS3pro
	Cavity_14_TAGDISVQ	x	x		HTRA4
	Cavity_15_GIAPLVT	x	x		HTRA1
	Cavity_15_SALITVPM	x	x		HTRA4
	Cavity_16_LAMIVQ	x	x		HTRA2
6	DENV4_NS3Pro_2whx_CAV3	x	x	x	NS3pro
	Cavity_10_YLFKPSVDETG	x	x		HTRA4
	Cavity_11_LAVTSQIYDPFR	x	x	x	HTRA3
	Cavity_15_NQPLTFIVASGMD	x	x	x	HTRA2

Tabla 4 Clusters de cavidades detectadas por más de un método de agrupamiento en representación F-Pocket – PockDrug. Se muestran las cavidades que forman los clusters consenso entre los métodos de agrupamiento. Además, se filtraron las cavidades humanas que se predijo su drogabilidad ($Druggability > 0.5$). Se muestra además la proteína de la cual proviene cada cavidad detectada.

Cluster	Pocket	Tree joining	K-Means	PCA	Druggability	Protein
1	DENV2_2FOM_Pocket6	x	x	x	0.73	NS3pro
	Pocket4	x	x		0.55	HTRA2
	Pocket2	x	x		0.72	Tysnd1
	Pocket0	x	x	x	0.61	Tysnd1
2	DENV3_3U1I_Pocket5	x	x	x	0.73	NS3pro
	Pocket3	x		x	0.92	Tysnd1
	Pocket3	x		x	0.67	HTRA2
3	DENV1_3L6P_Pocket2	x	x	x	0.71	NS3pro
	Pocket5	x	x	x	0.50	Tysnd1

Tabla 5 Clusters de cavidades detectadas por más de un método de agrupamiento en representación AADS – ProtDCal. Se filtraron las cavidades que presentaron patrones de correlación alta en los scores de inhibidores probados mediante docking. Estas cavidades se recomiendan utilizar para futuros experimentos de screening de compuestos activos para buscar reactividad cruzada.

Cluster	Cavity	Protein
1	DENV2_2fom_CAV3	NS3pro
	Cavity_1	TYSND1
2	DENV3_3uli_CAV6	NS3pro
	Cavity_7	TYSND1
	Cavity_7	HTRA4
	DENV4_2whx_CAV5	NS3pro
	Cavity_14	HTRA2
	Cavity_2	FAM111B
4	DENV4_2whx_CAV2	NS3pro
	Cavity_8	HTRA2
5	DENV3_3uli_CAV8	NS3pro
	Cavity_15	HTRA1
	Cavity_15	HTRA4
	Cavity_16	HTRA2
6	DENV4_2whx_CAV3	NS3pro
	Cavity_10	HTRA4

Tabla 6 Clusters de cavidades detectadas por más de un método de agrupamiento en representación F-Pocket – PockDrug. Se filtraron las cavidades que presentaron patrones de correlación alta en los scores de inhibidores probados mediante docking. Estas cavidades se recomiendan utilizar para futuros experimentos de screening de compuestos activos para buscar reactividad cruzada.

Cluster	Cavity	Protein
1	DENV2_2FOM_pocket6	NS3pro
	Pocket4	HTRA2
	Pocket0	Tysnd1
2	DENV3_3U1I_pocket5	NS3pro
	Pocket3	Tysnd1
	Pocket3	HTRA2
3	DENV1_3L6P_pocket2	NS3pro
	Pocket5	Tysnd1

Tabla 7 Ligandos consenso que cumplen con dos criterios de selectividad para cada cavidad de la Proteasa NS3 de DENV. Se tomaron en cuenta solo aquellos ligandos que cumplieran con el criterio de selectividad propuesto en al menos dos de las tres métricas evaluadas durante el experimento de Docking.

Cluster	NS3 Protease Cavity	Ligands (Selectivity > 1sd)	Ligands (Selectivity > 2sd)
AP1	DV2_c3	9, 30, 44, 56, 65, 66, 74	9, 56, 65, 66
	DV3_c2	9, 12, 29, 36, 37, 38, 39, 41, 43, 62, 64, 65	65
AP2	DV3_c6	9, 33, 36, 37, 38, 43, 52, 53, 56, 62, 63, 64, 65, 66	65
AP3	DV1_c3	12, 14, 25, 29, 40, 52, 53, 56, 62, 63, 64, 65, 66, 76	-
	DV4_c5	12, 14, 29, 36, 44, 53, 56, 62, 63, 65	65
AP4	DV4_c2	12, 39, 40, 57, 62, 63, 64, 65	63
AP5	DV3_c8	9, 12, 29, 33, 36, 37, 38, 39, 40, 41, 62, 63, 64, 76	63
AP6	DV4_c03	12, 14, 27, 36, 38, 62, 63, 64, 65	-
FP1	DV2_p06	9, 12, 14, 25, 29, 40, 41, 43, 53, 54, 65, 66, 76, 77	-
FP2	DV3_p05	12, 27, 28, 29, 39, 41, 52, 62, 63, 64, 77	62
FP3	DV1_p02	12, 13, 28, 29, 32, 43, 44, 52, 53, 56, 63, 65, 76, 77	-

Figuras:

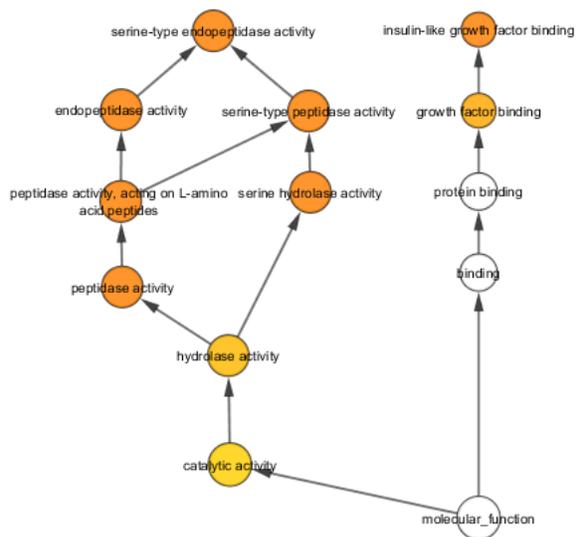


Figure 1 Función Molecular de Proteasas. El diagrama de redes generado con BiNGO en Cytoscape muestra las funciones moleculares organizadas jerárquicamente según sus GOs. El tamaño del círculo depende del número de repeticiones del GO en la lista de proteínas y la intensidad del color incrementa según el nivel de enriquecimiento del término GO indicado.

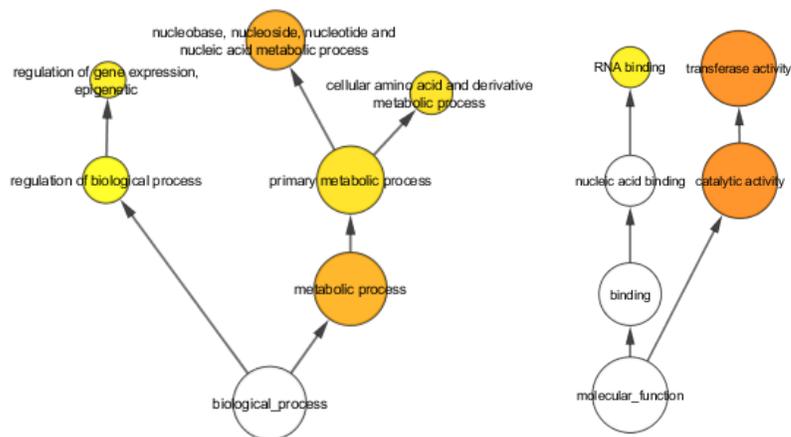


Figure 2 Proceso Biológico y Función Molecular de MTasas con GO Slim. El diagrama de redes generado con BiNGO en Cytoscape muestra las funciones moleculares organizadas jerárquicamente según sus GOs. El tamaño del círculo depende del número de repeticiones del GO en la lista de proteínas y la intensidad del color incrementa según el nivel de enriquecimiento del término GO indicado.



Figure 3 Proceso Biológico de Helicasas con GO Slim. El diagrama de redes generado con BiNGO en Cytoscape muestra las funciones moleculares organizadas jerárquicamente según sus GOs. El tamaño del círculo depende del número de repeticiones del GO en la lista de proteínas y la intensidad del color incrementa según el nivel de enriquecimiento del término GO indicado.



Figure 4 Superposición entre estructura NS3 hel de DENV PDB ID 2BMF (cian) y proteína humana HCG1998531 (dorado).



Figure 5 Superposición entre estructura NS5 MTasa de DENV PDB ID 3P97 (dorado) y proteína humana NLRP9 (cian).



Figure 6 Superposición entre estructura NS3 Pro de DENV PDB ID 3U11 (dorado) y proteína humana HTRA4 (cian).

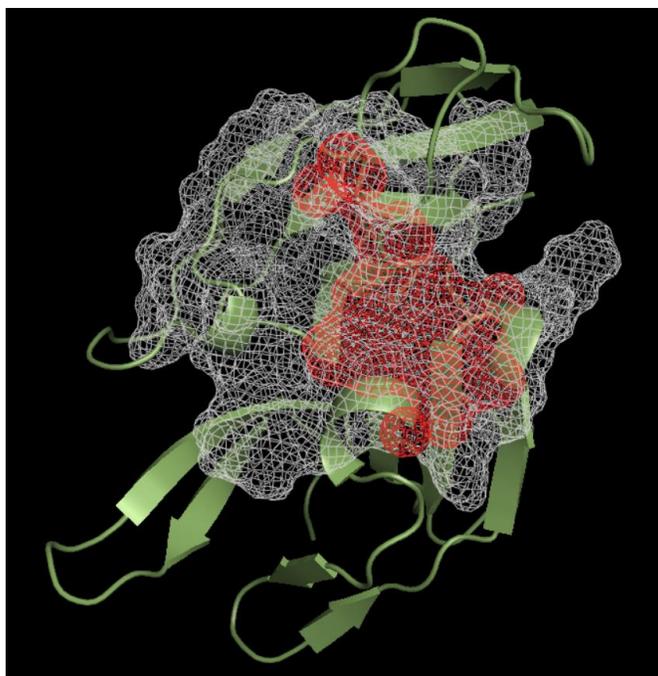


Figure 7 Estructura de proteasa 2FOM y ejemplo de cavidades coincidentes en ambas representaciones. Se muestra en mesh blanco la cavidad detectada por representación AADS y en rojo la detectada por F-pocket.

ANEXOS:

Tabla S 1 Resumen de métricas de calidad de modelos de homología para proteínas humanas similares a la NS3pro

Description	Job id	Hit	Confidence (%)	Sequence identity (%)	Alignment coverage (%)	Hit info 1	Hit info 2
O43464	d0b673ed5b78b0f5	c1lcyA_	100	100	69	2	PDB header:hydrolase
Q6SJ93	a0f300cf0feda344	c5c2zA_	99.9	23	28	1.96	PDB header:toxin
A0A024R4Z3	c7b05efa091ed239	c5c2zA_	99.9	23	33	1.96	PDB header:toxin
P83105	4fdf057e3072c655	c1lcyA_	100	50	66	2	PDB header:hydrolase
P83110	6c4ed4da7a6a7e6c	c4ri0A_	100	100	70	3.27	PDB header:hydrolase
Q2T9J0	ecb176347d51b96a	c5il9A_	100	25	39	2.2	PDB header:hydrolase
Q96PZ2	02b9961afa78825b	c5c2zA_	99.9	23	33	1.96	PDB header:toxin
Q92743	13f00088671db771	c1lcyA_	100	56	66	2	PDB header:hydrolase

Tabla S 2 Listado de estructuras existentes en el PDB Bank de proteínas humanas modeladas para el grupo relacionado a NS3 Pro

Gene	PDB ID	Method	Resolution (Å)	Chain	Position	Length	Name
HTRA2	1LCY	X-ray	2	A	134-458	324	Crystal Structure of the Mitochondrial Serine Protease HtrA2
HTRA2	2PZD	X-ray	2.75	A/B	359-458	99	Crystal Structure of the HtrA2/Omi PDZ Domain Bound to a Phage-Derived Ligand (WTMFWV)
HTRA2	5FHT	X-ray	1.95	A	134-458	324	HtrA2 protease mutant V226K
HTRA2	5M3N	X-ray	1.65	A	134-458	324	HTRA2 wild-type structure
HTRA2	5M3O	X-ray	1.7	A	134-458	324	HTRA2 A141S mutant structure
HTRA2	5TNY	X-ray	1.7	A	134-458	324	HTRA2 G399S mutant
HTRA2	5TNZ	X-ray	1.75	A	134-458	324	HtrA2 S142D mutant
HTRA2	5TO0	X-ray	1.9	A	134-458	324	HTRA2 S276C mutant
HTRA2	5TO1	X-ray	1.69	A	134-458	324	HtrA2 exposed (L266R, F303A) mutant
HTRA3	2P3W	X-ray	1.7	A/B	354-453	99	Crystal Structure of the HtrA3 PDZ Domain Bound to a Phage-Derived Ligand (FGRWV)
HTRA3	4RI0	X-ray	3.27	A/B/C	130-453	323	Serine Protease HtrA3, mutationally inactivated
HTRA1	2JOA	NMR	-	A	380-480	100	HtrA1 bound to an optimized peptide: NMR assignment of PDZ domain and ligand resonances
HTRA1	2YTW	NMR	-	A	370-480	110	Solution structure of the PDZ-domain of human protease HTRA 1 precursor
HTRA1	3NUM	X-ray	2.75	A	158-480	322	Substrate induced remodeling of the active site regulates HtrA1 activity
HTRA1	3NWU	X-ray	3.2	A/B/C	158-375	217	Substrate induced remodeling of the active site regulates HtrA1 activity
HTRA1	3NZI	X-ray	2.75	A	158-480	322	Substrate induced remodeling of the active site regulates HtrA1 activity
HTRA1	3TJN	X-ray	3	A/B/D	161-367	206	HtrA1 catalytic domain, apo form
HTRA1	3TJO	X-ray	2.3	A/B/D	161-370	209	HtrA1 catalytic domain, mutationally inactivated
HTRA1	3TJQ	X-ray	2	A	35-156	121	N-domain of HtrA1

Tabla S 3 Datos de compuestos con actividad reportada en la literatura contra la Proteasa NS3 de Dengue empleados en el ejercicio demostrativo de Docking.

N	Original ID	Source	Year	Type	Ki (um) DV1	Ki (um) DV2	Ki (um) DV3	Ki (um) DV4	Ki1	Ki2	KIS	Ki' DV1	Ki' DV2	Ki' DV3	Ki' DV4	IC50 DV1	IC50 DV2	IC50 DV3	IC50 DV4	EC50 DV1	EC50 DV2	EC50 DV3	EC50 DV4	CC50	Inhibition	D
1	1	(Ganesh et al., 2005)	2005	DV2		44																			Competitive	
2	4	(Ganesh et al., 2005)	2005	DV2		23																			Competitive	
3	5	(Ganesh et al., 2005)	2005	DV2		14.00																			Competitive	
4	4- hydroxypandurat in A	(Kiat et al., 2006)	2006	DV2		21.00																			Competitive	
5	Panduratin A	(Kiat et al., 2006)	2006	DV2		25.00																			Competitive	
6	Piperine	(Rahman, Yusof, & Tat, n.d.)	2006	NI	22.20																				NI	
7	A	(Mueller et al., 2008)	2008	DV2		28.60																			Competitive	
8	B	(Mueller et al., 2008)	2008	DV2		30.20																			Competitive	
9	C	(Mueller et al., 2008)	2008	DV2		17.00																			Competitive	
10	ARDP0006	(Tomlinson et al., 2009)	2009	DV2		432.00														4.20				69.00	Competitive	
11	1	(Bodenreider et al., 2009)	2009	DV1-4											36.40	6.00	17.5	32.80							Competitive	
12	3	(Bodenreider et al., 2009)	2009	DV1-4											10.70	1.1	8.50	13.60							Competitive	
13	6	(Bodenreider et al., 2009)	2009	DV1-4											170.00	55.00	86.00	43.00							NI	
14	NA	(Padmanabhan & Groutas, 2010)	2010	NI	2.36																				NI	
15	NA	(Padmanabhan & Groutas, 2010)	2010	NI																10.60					NI	
16	NA	(Padmanabhan & Groutas, 2010)	2010	NI																2.35					NI	
17	6A49	(Tomlinson & Watowich, 2011)	2011	DV2		15.00						10.00													Mixed	
18	6A60	(Tomlinson & Watowich, 2011)	2011	DV2		7.00						3.00													Mixed	
19	BP2109	(C. Yang et al., 2014)	2011	DV2												15.40				0.17					Competitive	
20	36	(Steuer et al., 2011)	2011	NI	50.00																				Competitive	
21	21	(Nitsche, Steuer, & Klein, 2011)	2011	NI	35.70																				Competitive	
22	1	(Knehans et al., 2011)	2011	DV2		2.00										7.70									Competitive	
23	2	(Knehans et al., 2011)	2011	DV2		31.1										37.90									Competitive	
24	1	(de la Cruz et al., 2011)	2011	NI	NI																				NI	
25	2	(Shiryaev, Cheltsov, Gawlik, Ratnikov, & Strongin, 2011)	2011	DV2												2.75				39.90				213.00	Non Competitive	
26	14	(Shiryaev et al., 2011)	2011	DV2												2.04				59.50				117.00	Non Competitive	

27	4 (166347)	(Cregar-Hernandez et al., 2011)	2012	DV1-4			3.00	5.00	5.00	11.00								Competitive
28	9 (166550)	(Cregar-Hernandez et al., 2011)	2012	DV1-4			5.30	10.00	4.00	12.40								Competitive
29	7n	(Aravapalli et al., 2012)	2012	DV2		8.77												Competitive
30	7n	(Tiew et al., 2012)	2012	DV2				3.48										NI
31	23	(Mendgen, Steuer, & Klein, 2012)	2012	NI	50.00													NI
32	1	(Deng et al., 2012)	2012	DV2				13.1				35.90		12.60				NI
33	23i	(Deng et al., 2012)	2012	DV2				9.45				24.70		100.00				NI
34	4a	(Chao et al., 2012)	2012	DV2								0.15						NI
35	4o	(Chao et al., 2012)	2012	DV2								0.00						NI
36	4l	(Tiew et al., 2012)	2012	DV2				4.87										NI
37	7j	(Tiew et al., 2012)	2012	DV2				6.22										NI
38	7k	(Tiew et al., 2012)	2012	DV2				13.40										NI
39	7l	(Tiew et al., 2012)	2012	DV2				4.45										NI
40	7m	(Tiew et al., 2012)	2012	DV2				4.59										NI
41	7n	(Tiew et al., 2012)	2012	DV2				3.48										NI
42	2	(Nguyen et al., 2013)	2013	DV4		4.00						15.60						Competitive
43	14	(Nguyen et al., 2013)	2013	DV4		4.90						12.50						Competitive
44	22	(Nguyen et al., 2013)	2013	DV4		3.40						3.90						Competitive
45	1	(Heh et al., 2013)	2013	DV2		69												Non Competitive
46	SK-12	(Pambudi et al., 2013)	2013	DV1-4								0.97	0.98	2.43	0.74			Non Competitive
47	Doxycycline	(Rothan et al., 2013)	2013	DV2		55.60						40.00						Non Competitive
48	Mefenamic acid	(Rothan et al., 2013)	2013	DV2		32.00						32.00						Non Competitive
49	Ribavirin	(Rothan et al., 2013)	2013	DV2								30.00						Non Competitive
50	BP13944	(C. Yang et al., 2014)	2014	DV2					22.60			1.03						Uncompetitive
51	Amodiaquine	(Boonyasuppayakorn, Reichert, Manzano, Nagarajan, & Padmanabhan, 2014)	2014	DV2								7.41						NI
52	1	(Liu et al., 2014)	2014	DV2					6.09									NI
53	8b	(Liu et al., 2014)	2014	DV2					2.24									NI
54		(Viswanathan, Tomlinson, Fonner, Mock, & Watowich, 2014)	2014	DV2			7.00	15.00	1130									Non Competitive
55	Comp1	(Wu et al., 2015)	2015	DV1-2				98	31.8									Non Competitive
56	Comp14	(L. Li et al., 2015)	2015	DV2								5						Competitive

Tabla S 4 Estructuras en formato SMILE de compuestos con actividad reportada en la literatura contra la Proteasa NS3 de Dengue empleados en el ejercicio demostrativo de Docking.. DV: Dengue Virus.

N	Original ID	Source	Year	Type	SMILE
1	1	(Ganesh et al., 2005)	2005	DV2	<chem>N/C(N)=N/N=C/C1=CC=C(/C=N/N=C(N)/N)C=C1</chem>
2	4	(Ganesh et al., 2005)	2005	DV2	<chem>N/C(N)=N/N=C(C(N1)=O)/C2=C1C=CC=C2</chem>
3	5	(Ganesh et al., 2005)	2005	DV2	<chem>CC1=CC=C(C(N/N=C(N)/N)P([H])(O)=O)C=C1</chem>
4	4- hydroxypand uratin A	(Kiat et al., 2006)	2006	DV2	<chem>OC1=C(C([C@@H]2[C@H](C/C=C(C)/C)C(C)=CC[C@H]2C3=CC=CC=C3)=O)C(O)=CC(O)=C1</chem>
5	Panduratin A	(Kiat et al., 2006)	2006	DV2	<chem>OC1=C(C([C@@H]2[C@H](C/C=C(C)/C)C(C)=CC[C@H]2C3=CC=CC=C3)=O)C(O)=CC(OC)=C1</chem>
6	Piperine	(Rahman et al., n.d.)	2006	NI	<chem>O=C(N1CCCCC1)/C=C/C=C/C2=CC=C(OCO3)C3=C2</chem>
7	A	(Mueller et al., 2008)	2008	DV2	<chem>COC1=CC=CC(C(NC(C2=CC=CC=C2)=O)C3=C(O)C4=C(C=C3)C=CC=N4)=C1</chem>
8	B	(Mueller et al., 2008)	2008	DV2	<chem>OC1=C(C(NC2=NC=CS2)C3=C(O)C=CC=C3)C=CC4=C1N=CC=C4</chem>
9	C	(Mueller et al., 2008)	2008	DV2	<chem>CN(C1=O)C(SCC(NC2=CC(OCO3)=C3C=C2)=O)=NC4=C1NC5=C4C=CC=C5</chem>
10	ARDP0006	(Tomlinson et al., 2009)	2009	DV2	<chem>OC1=C2C(C(C(C([N+])([O-])=O)=CC=C3O)=C3C2=O)=O)=C([N+])([O-])=O)C=C1</chem>
11	1	(Bodenreider et al., 2009)	2009	DV1-4	<chem>CC(CN1)N=C1C2=CC=C(C=C2)NC3=C(C=CC=C4)C4=C(NC5=CC=C(C6=NC(CN6)C)C=C5)N=N3</chem>
12	3	(Bodenreider et al., 2009)	2009	DV1-4	<chem>C1(NC2=C3C(C=CC=C3)=C(NC4=CC=C(C5=NC(CCCC6)C6N5)C=C4)N=N2)=CC=C(C7=NC(CCCC8)C8N7)C=C1</chem>
13	6	(Bodenreider et al., 2009)	2009	DV1-4	<chem>C1(C=CC(NC2=C3C(C=CC=C3)=C(NC4=CC=C(C5=NCCN5)C=C4)N=N2)=C6)=C6CCNC1</chem>
14	NA	(Padmanabhan & Groutas, 2010)	2010	NI	<chem>OC1=C(C(NC2=NC(C=CC=C3)=C3S2)C4=CC(OCC5=CC=CC=C5)=CC=C4)C=CC6=C1N=CC=C6</chem>
15	NA	(Padmanabhan & Groutas, 2010)	2010	NI	<chem>OC1=C(CN2CCOCC2)C=CC3=C1N=CC=C3</chem>
16	NA	(Padmanabhan & Groutas, 2010)	2010	NI	<chem>CCN(CC)CC1=C(O)C2=C(C(C1)=C1)C=CC=N2</chem>
17	6A49	(Tomlinson & Watowich, 2011)	2011	DV2	<chem>OC(C1=C2C=CC=C1)=CC=C2/N=N/C3=CC=C([N+])([O-])=O)C=C3[N+])([O-])=O</chem>
18	6A60	(Tomlinson & Watowich, 2011)	2011	DV2	<chem>OC1=C2C(C=CC(O)=C2)=CC3=CC=C(O)C=C31</chem>
19	BP2109	(C. Yang et al., 2014)	2011	DV2	<chem>O=C1C(C=C(OCC[N+](CCCCCCCC)(CC)CC)C=C2)C2C3=C1C=C(OCC[N+](CCCCCCCC)(CC)CC)C=C3</chem>
20	36	(Steuer et al., 2011)	2011	NI	<chem>[H]N(C(C)(C)C(C/C=C/C1=CNC2=C1C=CC=C2)=O)=O</chem>
21	21	(Nitsche et al., 2011)	2011	NI	<chem>[H]N([H])C(/C(C#N)=C/C1=CC=C(O)C=C1)=O</chem>
22	1	(Knehans et al., 2011)	2011	DV2	<chem>NC(NC(C=C1)=CC=C1C(OC2=CC=C(NC(N)=N)C=C2)=O)=N</chem>
23	2	(Knehans et al., 2011)	2011	DV2	<chem>NC(NC(C=C1)=CC=C1C(OC2=CC=C(C#N)C=C2)=O)=N</chem>
24	1	(de la Cruz et al., 2011)	2011	NI	<chem>NC(NC(C=C1)=CC=C1C(OC2=CC=C([N+])([O-])=O)C=C2)=O)=N</chem>
25	2	(Shiryaev et al., 2011)	2011	DV2	<chem>NC1CC(N)C(/N=N/C2=CC=C(S(NC3=NC=CS3)(=O)=O)C=C2)CC1/N=N/C4=CC=C(S(=O)(NC5=NC=CS5)=O)C=C4</chem>
26	14	(Shiryaev et al., 2011)	2011	DV2	<chem>OC(C(CCCCC)=C/1)=CC(C1=N)NC2=CC=C(S(NC3=NC=CS3)(=O)=O)C=C2)=O</chem>
27	4 (166347)	(Cregar-Hernandez et al., 2011)	2012	DV1-4	<chem>NC(N[C@H]1[C@H](OC2=CC=C(NC(N)=N)C=C2NC(N)=N)C[C@H](OC3=CC=C(NC(N)=N)C=C3NC(N)=N)[C@@H](NC(N)=N)C1)=N</chem>
28	9 (166550)	(Cregar-Hernandez et al., 2011)	2012	DV1-4	<chem>NC(N[C@H]1[C@H](OC2=CC=C(NC(N)=N)C=C2)C[C@H](OC3=CC=C(NC(N)=N)C=C3)[C@@H](NC4=CC=C(NC(N)=N)C=C4)C1)=N</chem>
29	7n	(Aravapalli et al., 2012)	2012	DV2	<chem>O=C(NC1=CC=C(OC2=CC=CC=C2)C=C1)C3=CC=C(NC(NCC4=CN(CC5=CC=C(F)C=C5)N=N4)=O)C=C3</chem>
30	7n	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N([C@@H](CC2=CC=CC=C2)C3=NN=C(C4=CC=C(OC)C=C4)O3)SC5=CC=CC=C51</chem>

31	23	(Mendgen et al., 2012)	2012	NI	<chem>S=C(NC/I=O)SC1=C/C2=C(Br)C=CS2</chem>
32	1	(Deng et al., 2012)	2012	DV2	<chem>OC1=C(/C=N/NC2=NC(C=CC=C3)=C3N2)C=C(C(OCC)=O)C=C1/C=N/NC4=NC5=C(C=CC=C5)N4</chem>
33	23i	(Deng et al., 2012)	2012	DV2	<chem>O=C(N/N=C/CC1=CC=CC=C1)C2=CC(/C=N/NC3=CC=CC(Br)=C3)=NC4=C2C=CC=C4</chem>
34	4a	(Chao et al., 2012)	2012	DV2	<chem>COC1=C2C(N)=NC(N)=NC2=CC=C1</chem>
35	4o	(Chao et al., 2012)	2012	DV2	<chem>CC(C)(C)OC1=C2C(N)=NC(N)=NC2=CC=C1</chem>
36	4l	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N(CC(NCC2=CN(CC(NC3=CC=C(OC4=CC=CC=C4)C=C3)=O)N=N2)=O)SC5=CC=CC=C51</chem>
37	7j	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N(CC(NC2=CC(OCC3=CN(CC(C4=CC=CC=C4)=O)N=N3)=CC=C2)=O)SC5=CC=CC=C51</chem>
38	7k	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N(CC(NC2=CC(OCC3=CN(CC(NC4=CC=CC=C4)=O)N=N3)=CC=C2)=O)SC5=CC=CC=C51</chem>
39	7l	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N(CC(NC2=CC(OCC3=CN(CC(NC4=CC=C(OC5=CC=CC=C5)C=C4)=O)N=N3)=CC=C2)=O)SC6=CC=CC=C61</chem>
40	7m	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N(CC(NC2=CC(OCC3=CN(CC(NC4=CC=CC(OC5=CC=CC=C5)=C4)=O)N=N3)=CC=C2)=O)SC6=CC=CC=C61</chem>
41	7n	(Tiew et al., 2012)	2012	DV2	<chem>O=C1N(CC(NC2=CC(OCC3=CN(CC(NC4=CC=CC=C4OC5=CC=CC=C5)=O)N=N3)=CC=C2)=O)SC6=CC=CC=C61</chem>
42	2	(Nguyen et al., 2013)	2013	DV4	<chem>O=C(C1=CC=C(CN2N=C([N+])([O-])=O)C(C1)C2C)O1)NC3=NC4=C(N3CCN(CC)CC)C=CC=C4</chem>
43	14	(Nguyen et al., 2013)	2013	DV4	<chem>CC(N1C(C2=CC=C(C(O)=O)C=C2)OC(C3=CC(C4=CC=CC=C4)=NC5=C3C=CC=C5)=N1)=O</chem>
44	22	(Nguyen et al., 2013)	2013	DV4	<chem>O=C1N(C2=C(C)C=CC=C2C)C(SCC3=CC(N(C=C(C1)C=C4)C4=N3)=O)=NC5=C1SC(N5C6=C(C)C=CC=C6C)=S</chem>
45	1	(Heh et al., 2013)	2013	DV2	<chem>O=C(C=C(C1=CC=CC=C1)O2)C3=C2C=CC(C4=NN=N[N-]4)=C3</chem>
46	SK-12	(Pambudi et al., 2013)	2013	DV1-4	<chem>OC1=C(SCC(O)=O)C=C(CS(C2=CC([N+])([O-])=O)C(C)C=C2)(=O)=O)C3=CC=CC=C31</chem>
47	Doxycycline	(Rothan et al., 2013)	2013	DV2	<chem>NC(C1=C(O)[C@@H](N(C)C)[C@]2([H])[C@@H](O)[C@](C3=C(O)[C@]2(O)C1=O)([H])[C@@H](C)C4=CC=CC(O)=C4C3=O)=O</chem>
48	Mefenamic acid	(Rothan et al., 2013)	2013	DV2	<chem>CC1=CC=CC(NC2=C(C(O)=O)C=CC=C2)=C1C</chem>
49	Ribavirin	(Rothan et al., 2013)	2013	DV2	<chem>O[C@@H]1[C@@H](CO)O[C@@H](N2N=C(C(N)=O)N=C2)[C@@H]1O</chem>
50	BP13944	(C. Yang et al., 2014)	2014	DV2	<chem>CC[N+](CCCCCCCCCCCCCCCC)(CO)CO</chem>
51	Amodiaquine	(Boonyasuppayakorn et al., 2014)	2014	DV2	<chem>C1C1=CC=C2C(N=CC=C2NC3=CC(CN(CC)CC)=C(O)C=C3)=C1</chem>
52	1	(Liu et al., 2014)	2014	DV2	<chem>N=C1N2N=C(SCC3=CC=CC=C3)SC2=NC(=O)C1=C/C1=CN(CC2=CC=CC=C2)C2=CC=CC=C12</chem>
53	8b	(Liu et al., 2014)	2014	DV2	<chem>FC1=CC=C(CN2C=C(\C=C(/C#N)C=O)NC3=NN=C(SCC4=CC=CC=C4)S3)C3=C2C=CC=C3)C=C1</chem>
54		(Viswanathan et al., 2014)	2014	DV2	<chem>FC1=CC2=C(C=C1)C=C(CC2)C1CC2=CC(F)=CC=C2CC1=O</chem>
55	Comp1	(Wu et al., 2015)	2015	DV1-2	<chem>O=C(NC1=CC=CS1)C1=C(SC2=CC=C(C=C2)N(=O)=O)C=CC=C1</chem>
56	Comp14	(L. Li et al., 2015)	2015	DV2	<chem>CN1CCN(CC1)C1=CC(C)=C(NC(=O)C2=C3C=CC=CC3=NC3=C2CN(CC2=CC=CC=C2)CC3)C=C1</chem>
57	MB21	(Raut et al., 2015)	2015	DV2	<chem>OC(=O)C1=CC=C(C=C1)C1=CC=C(C1)\C=C(/C#N)C1=NC2=C(N1)C=CC(C1)=C2</chem>
58	14	(Zou et al., 2015)	2015	DV2	<chem>C1C1=CC=C(CN2N=CC3=C2NC(=O)CC32C(=O)NC3=CC=C(C1)C=C23)N=C1</chem>
59	1	(Zou et al., 2015)	2015	DV2	<chem>C1C1=CC=C(CN2N=CC3=C2NC(=O)CC32C(=O)NC3=CC=C(C1)C=C23)C=C1</chem>
60	Comp8	(Wu et al., 2015)	2015	DV1-2	<chem>OC1=CC2=C(C=C1O)N=C(NC(=O)C1=C(SC3=CC=C(C=C3)C(F)(F)F)C=CC=C1)S2</chem>
61	Compound1	(Koh-Stenta et al., 2015)	2015	DV2	<chem>CCC1=CC=C(C=C1)S(=O)(=O)N1N=C(OC(=O)C2=CC=CC=C2)C=C1N</chem>
62	Compound2	(Koh-Stenta et al., 2015)	2015	DV2	<chem>CCC1=CC=C(C=C1)S(=O)(=O)N1N=C(OC(=O)C2=CC=CC(=C2)C2=CC=CC=C2)C=C1N</chem>
63	Compound3	(Koh-Stenta et al., 2015)	2015	DV2	<chem>NC1=CC(CC2=CC=C(C=C2)C2=CC=CC=C2)=NN1S(=O)(=O)C1=CC=C(C=C1)C1=CC=CC=C1</chem>
64	54692801	(Cabarcas-Montalvo et al., 2016)	2016	DV2	<chem>CC1(C)CC(O)C2=C(C1)C1=C(NC2C2=CC=C(O2)C2=CC=C(O)C(=C2)C(O)=O)C=CC2=CC=CC=C12</chem>
65	54715399	(Cabarcas-Montalvo et al., 2016)	2016	DV2	<chem>COC1=C(CSC2=NC3=CC=CC=C3S2)C=C(C=C1)C(C1=C(O)C2=CC=CC=C2OC1=O)C1=C(O)C2=C(OC1=O)C=CC=C2</chem>
66	54681617	(Cabarcas-Montalvo et al., 2016)	2016	DV2	<chem>COC1=C(COC2=CC=C(NC(C)=O)C=C2)C=C(C=C1)C(C1=C(O)C2=CC=CC=C2OC1=O)C1=C(O)C2=C(OC1=O)C=CC=C2</chem>
67	Compound	(Balasubramanian et	2016	DV1-4	<chem>CC1=CC=C(C=C1)C1=NC(=O)\C(S1)=C\C1=CC(Br)=C(O)C(O)=C1</chem>

Tabla S 5 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster AP1

	DV2_c3- DV3_c2	DV2_c3- Tys_c2	DV2_c3- Tys_c1	DV3_c2- Tys_c2	DV3_c2- Tys_c1	Tys_c2- Tys_c1
T3_Average	0.93	0.85	0.84	0.88	0.89	0.99
T3_Max	0.93	0.84	0.83	0.89	0.90	0.98
T3_Min	0.91	0.86	0.84	0.88	0.88	0.99

Tabla S 6 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster AP2

	DV3_c6- FAMA_c6	DV3_c6- Tys_c7	DV3_c6- HT4_c7	FAMA_c6- Tys_c7	FAMA_c6- HT4_c7	Tys_c7- HT4_c7
T3_Average	0.92	0.88	0.93	0.88	0.92	0.86
T3_Max	0.91	0.85	0.91	0.88	0.90	0.84
T3_Min	0.90	0.88	0.92	0.87	0.92	0.86

Tabla S 7 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster AP3

	DV1_c3- DV4_c5	DV1_c3- FAMA_c5	DV1_c3- FAM111B_c2	DV1_c3- HTRA2_c2	DV1_c3- HTRA2_c14	DV1_c3- Tysnd1_c4
T3_Average	0.94	0.94	0.94	0.80	0.85	0.89
T3_Max	0.94	0.92	0.92	0.82	0.83	0.85
T3_Min	0.92	0.92	0.92	0.73	0.83	0.88
	DV4_c5- FAMA_c5	DV4_c5- FAM111B_c2	DV4_c5- HTRA2_c2	DV4_c5- HTRA2_c14	DV4_c5- Tysnd1_c4	FAMA_c5- FAM111B_c2
T3_Average	0.93	0.95	0.79	0.85	0.88	0.97
T3_Max	0.91	0.93	0.82	0.84	0.85	0.95
T3_Min	0.91	0.93	0.70	0.81	0.87	0.93
	FAMA_c5- HTRA2_c2	FAMA_c5- HTRA2_c14	FAMA_c5- Tysnd1_c4	FAM111B_c2- HTRA2_c2	FAM111B_c2- HTRA2_c14	FAM111B_c2- Tysnd1_c4
T3_Average	0.83	0.89	0.91	0.81	0.87	0.90
T3_Max	0.86	0.89	0.91	0.84	0.85	0.89
T3_Min	0.75	0.84	0.89	0.72	0.84	0.89
	HTRA2_c2- HTRA2_c14	HTRA2_c2- Tysnd1_c4	HTRA2_c14- Tysnd1_c4			
T3_Average	0.89	0.85	0.88			
T3_Max	0.90	0.86	0.88			
T3_Min	0.79	0.77	0.83			

Tabla S 8 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster AP4

	DV4_c2- DV4_c2	DV4_c2- HTRA1_c7	HTRA1_c7- HTRA2_c08
T3_Average	0.88	0.93	0.85
T3_Max	0.89	0.92	0.85
T3_Min	0.84	0.91	0.83

Tabla S 9 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster AP5

	DV3_c8- HT1_c15	DV3_c8- HT2c16	DV3_c8- HT4_c15	DV3_c8- HT4_c14	HT1_c15- HT2_c16	HT1_c15- HT4_c15
T3_Average	0.80	0.76	0.89	0.89	0.76	0.89
T3_Max	0.78	0.78	0.87	0.87	0.78	0.88
T3_Min	0.82	0.72	0.89	0.87	0.70	0.88

	HT1_c15- HT4_c14	HT2_c16- HT4_c15	HT2_c16- HT4_c14	HT4_c15- HT4_c14
T3_Average	0.89	0.80	0.78	0.97
T3_Max	0.87	0.81	0.80	0.97
T3_Min	0.88	0.77	0.75	0.96

Tabla S 10 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster AP6

	DV4_c03- HT2_c15	DV4_c03- HT4_c10	DV4_c03- HT3_c11	HT2_c15- HT4_c10	HT2_c15- HT3_c11	HT4_c10- HT3_c11
T3_Average	0.69	0.91	0.86	0.73	0.79	0.88
T3_Max	0.77	0.91	0.85	0.78	0.80	0.86
T3_Min	0.56	0.89	0.83	0.62	0.70	0.86

Tabla S 11 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster FP1

	DV2_p06- HT2_p4	DV2_p06- TYS_p0	DV2_p06- TYS_P2	HT2_p4- TYS_p0	HT2_p4- TYS_P2	TYS_p0- TYS_P2
T3_Average	0.88	0.90	0.87	0.89	0.88	0.98
T3_Max	0.87	0.91	0.87	0.90	0.89	0.97
T3_Min	0.79	0.85	0.82	0.83	0.83	0.98

Tabla S 12 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster FP2

	DV3_p05- HT2_p3	DV3_p05- Tys_p3	HT2_p3- Tys_p3
T3_Average	0.88	0.89	0.88
T3_Max	0.90	0.87	0.90
T3_Min	0.84	0.86	0.83

Tabla S 13 Correlación entre resultados de las métricas de docking entre receptores (pares) del cluster FP3

	DV1_p02- TYS_p5
T3_Average	0.78
T3_Max	0.69
T3_Min	0.80

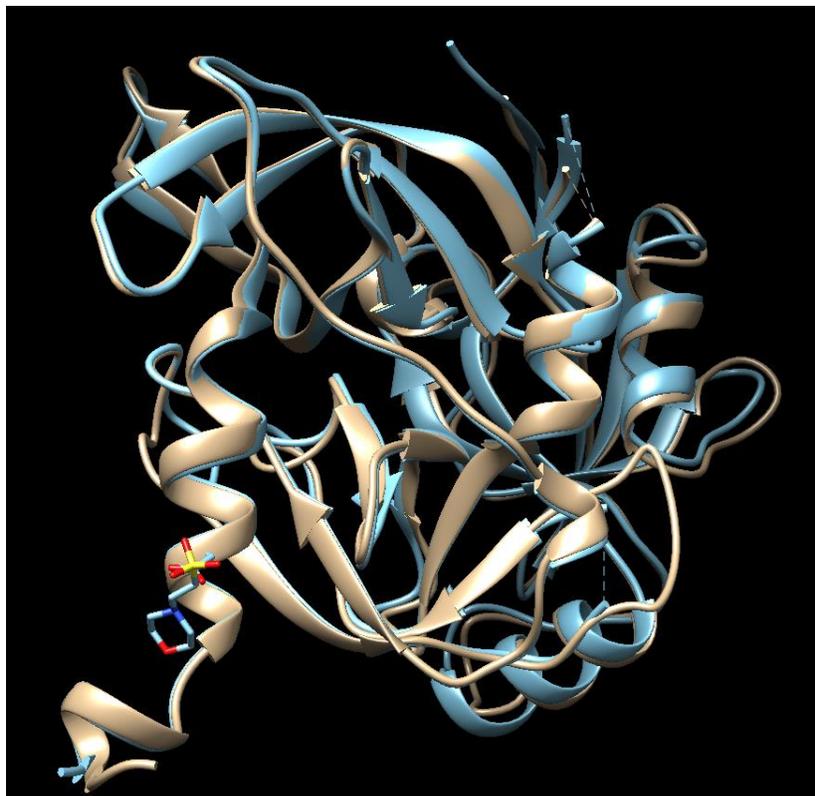


Figura S 1 Superposición entre estructura de homólogas de HTRA2 (café) y PDB ID 5M3N (celeste). RMSD=0.755

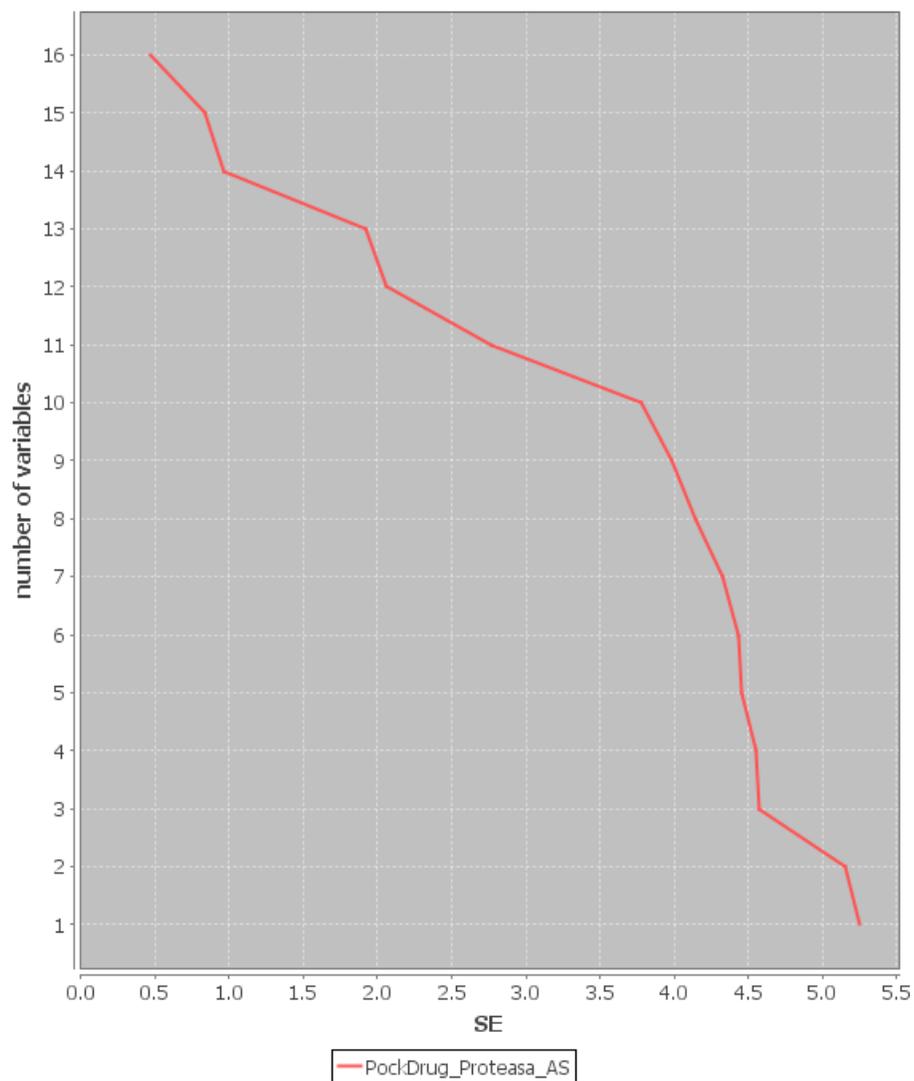


Figura S 2 Distribución de Shannon según el número de variables en representación Fpocket-PockDrug