

**UNIVERSIDAD DE TALCA
ESCUELA DE INGENIERÍA CIVIL EN BIOINFORMÁTICA
TALCA-CHILE**



**“IDENTIFICACIÓN DE INTEGRINAS COMO BLANCOS
FARMACOLÓGICOS PARA EL DISEÑO DE TERAPIAS
CONTRA EL CÁNCER COLORRECTAL”**

ISABEL MACARENA GARCÍA PEREIRA

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN BIOINFORMÁTICA**

Profesor tutor: Dra. Ariela Vergara-Jaque
Profesor Cotutor: Dr. Horacio Poblete Vilches
Profesor informante: Dr. Julio Caballero Ruiz

Marzo - 2022

CONSTANCIA

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



Talca, 2022

A mi madre, por brindarme mediante su esfuerzo todas las oportunidades que ella jamás soñó con tener, por inculcarme sus valores, mantenerme centrada y humilde. El esfuerzo puesto en este proceso es por y para ti.

AGRADECIMIENTOS

A mis profesores guía, Ariela Vergara y Horacio Poblete por su completa disposición y entrega de conocimiento, gracias por apoyarme con su excelente nivel académico y por además demostrar su calidez, humanidad, paciencia y bondad.

A mi madre, hermanos y abuelos, por su infinito cariño, apoyo y confianza.

A Marcela, Josefa, Valentina, Sofía, Camilo y Arturo por siempre estar distrayéndome y blindándome ante las dificultades de la vida, valoro cada palabra y momento invertido en escucharme y traspasarme su ánimo y tranquilidad.

A Rocío, Dámariz, Luis Gonzalo y Christian, por ser mis compañeros en las buenas y malas de todo este proceso universitario, agradezco eternamente su constancia, cariño y dedicación.

Finalmente, a mis mascotas Pankasán, Bucky, Samito y Yuri, por entregarme una salida de las personas cuando todo se siente como demasiado.

RESUMEN

El cáncer colorrectal (CRC) es el tercer cáncer más diagnosticado en el mundo, con probabilidades de supervivencia muy bajas si no es detectado en un estado temprano. En este contexto es que se buscan alternativas eficientes, y complementarias a las aproximaciones quirúrgicas y de quimioterapia que ya existen, para tratar este tipo de patologías. La identificación de proteínas que inciden en procesos celulares asociados a CRC resulta esencial para definir blancos farmacológicos que faciliten la unión de formulaciones médicas. Proteínas pertenecientes a la familia de integrinas, un tipo de estructuras heterodiméricas compuestas por una subunidad α y una subunidad β , participan en los pasos fundamentales para la reproducción y supervivencia celular. Variadas isoformas de estas proteínas están sobreexpresadas en diversos tipos de cáncer. Específicamente, las isoformas $\alpha5\beta1$, $\alphaV\beta6$ y $\alphaV\beta3$ están vinculadas directamente a cáncer colorrectal. En este contexto, el objetivo general del presente trabajo consistió en evaluar la unión de estas proteínas a fragmentos peptídicos que en su secuencia presentan los motivos RGD y NGR, los cuales han mostrado alta afinidad por integrinas. Para ello, se realizaron estudios bioinformáticos que permitieron identificar residuos claves en la interacción proteína-péptido mediante análisis de secuencia y estructura. Inicialmente, se modeló una serie de péptidos de interés, cuya afinidad con las integrinas fue evaluada a través de estudios de acoplamiento molecular. Posteriormente, se llevaron a cabo simulaciones de dinámica molecular y estimaciones de energía libre de unión de los complejos integrina-péptido mediante el método MM-GBSA (MM, mecánica molecular; GB, Born generalizado; SA, área de superficie). La finalidad de estos estudios fue identificar complejos de alta afinidad que puedan servir como blancos farmacológicos para diseñar adyuvantes contra el cáncer colorrectal.

ABSTRACT

Colorectal cancer (CRC) is the third most diagnosed cancer in the world, with very low suspicion of survival when is not detected at an early stage. In this context, efficient alternatives are sought, and complementary to the surgical and chemotherapy approaches that already exist, to treat this type of pathology. The identification of proteins that affect cellular processes associated with CRC is essential to define pharmacological targets that facilitate the binding of medical formulations. Proteins belonging to the family of integrins, a type of heterodimeric structure composed of an α subunit and a β subunit, participate in the fundamental steps for cell reproduction and survival. Various isoforms of these proteins are overexpressed in various types of cancer. Specifically, the $\alpha 5\beta 1$, $\alpha V\beta 6$, and $\alpha V\beta 3$ isoforms are directly linked to colorectal cancer. In this context, the general objective of this work was to evaluate the binding of these proteins to peptide fragments that have the RGD and NGR motifs in their sequence, which have shown a high affinity for integrins. For this, bioinformatic studies were carried out that allowed the identification of key residues in the protein-peptide interaction through sequence and structure analysis. Initially, a series of peptides of interest were modeled, whose affinity with integrins was evaluated through docking studies. Subsequently, molecular dynamics simulations and binding free energy estimates of integrin-peptide complexes were carried out using the MM-GBSA method (MM, molecular mechanics; GB, generalized Born; SA, surface area). The purpose of these studies was to identify high-affinity complexes that can serve as drug targets for designing adjuvants against colorectal.

INDICE DE CONTENIDO

1. INTRODUCCIÓN	11
1.1 Cáncer colorrectal	11
1.2 Integrinas	14
1.3 Integrinas y cáncer colorrectal	16
1.4 Motivo RGD y NGR	17
1.5 Péptidos contra el cáncer	20
1.6 Caracterización de asociaciones integrina-péptido	22
2. HIPÓTESIS	24
3. OBJETIVOS.....	24
3.1. Objetivo General	24
3.2. Objetivos Específicos.....	24
4.- MATERIALES Y MÉTODOS.....	25
4.1 Revisión de estructuras de integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$	25
4.2 Identificación de residuos que determine la unión de integrinas a péptidos RGD/NGR	26
4.2.1 Búsqueda de patrones a nivel de secuencia.	26
4.2.2 Búsqueda de patrones geométricos	26
4.3 Modelado de péptidos.....	27
4.4 Acoplamiento molecular Integrina-péptido.....	28
4.4.1 Preparación del acoplamiento molecular	29
4.4.2 Acoplamiento molecular con Auto Dock CrankPep	29
4.4.3 Criterio de término y clusterización de poses	30
4.5 Simulaciones moleculares de complejos proteína-péptido de alta afinidad.....	32
4.5.1 Elección de sistemas para simulaciones de dinámica molecular	32
4.5.2 Construcción de sistemas para simulaciones de dinámica molecular	33
4.5.2 Simulaciones de dinámica Molecular.....	33
4.6 Cálculos de energía libre de unión de los complejos.....	35
5.- RESULTADOS	37
5.1.- Identificación de patrones estructurales.....	37
5.2.- Identificación de patrones Geométricos	41
5.3.- Modelado de péptidos	43
5.4.- Acoplamiento Molecular	45
5.5.- Dinámica Molecular	52

5.6.- Energía Libre de Unión	55
6.- CONCLUSIÓN	57
7.- REFERENCIAS	59

INDICE DE FIGURAS

Figura 1.- Distribución de casos y muertes asociadas a los 10 cánceres más comunes en 2020 (Sung et al., 2021).....	11
Figura 2.- Anatomía del sistema gastrointestinal (American Cancer Society, 2020). El cáncer colorrectal aborda las cuatro secciones del colon además del recto.	12
Figura 3.- Esquema de la estructura y activación de las integrinas. El heterodímero es dinámico y su conformación pasa a un estado extendido ante la presencia de ligandos. Adaptada de (Mezu-Ndubuisi & Maheshwari, 2020).	15
Figura 4.- Integrinas expresadas en células epiteliales, en condiciones normales o patológicas, organizadas según sus ligandos extracelulares.	16
Figura 5.- Representación tridimensional de la interacción del péptido RGDFV cíclico (cilengitide) y la integrina $\alpha\beta3$	19
Figura 6.- Modo de unión del péptido Cilengitide a la integrina $\alpha\beta3$. Estructura con código PDB 1L5G obtenida a 3,2 Å de resolución (J.-P. Xiong et al., 2002).....	20
Figura 7.- Esquema de trabajo de investigación. En el recuadro morado se presenta el conjunto de péptidos identificados para este estudio incluyendo secuencias que contienen el motivo de unión NGR y RGD.....	23
Figura 8.- Esquema de flujo para elección de mejor pose en ADCP. Las 100 poses generadas en este paso son luego agrupadas en clústeres con ADFR.	31
Figura 9.- Alineamiento de secuencias entre subunidades α de integrinas de interés..	37
Figura 10.- Vista estructural del patrón conservado a nivel de secuencia entre las subunidades αV y $\alpha 5$	38
Figura 11.- Vista estructural del patrón ASP227 y ASP228 conservado a nivel de secuencia entre las subunidades αV y $\alpha 5$	38
Figura 12.- Alineamiento de secuencias entre subunidades β de integrinas de interés.....	39
Figura 13.- Vista estructural del patrón conservado a nivel de secuencia, con ASP-ASN-SER-PRO-GLU interactuando con GLU del péptido RGD en las 3 subunidades β estudiadas.....	39
Figura 14.- Vista estructural del patrón conservado a nivel de secuencia con SxS interactuando con el residuo Asp del péptido RGD en las 3 subunidades β estudiadas.	40
Figura 15.- Patrón Geométrico identificado entre subunidades α del conjunto en estudio.	41
Figura 16.- Patrón Geométrico identificado entre subunidades β del conjunto en estudio.	42
Figura 17.- Visualización 3D de modelos iniciales de péptidos de interés.	44
Figura 18.- A) Resultados de poses de docking de cada integrina en estudio asociada a los 6 péptidos de interés. B) Alineamiento de todos los complejos estudiados y sus poses resultantes del acoplamiento molecular.....	45
Figura 19.- A) Alineamiento de las 10 mejores poses para el péptido c(RGDFV) en 4WK0. B) Alineamiento de las 10 mejores poses para el péptido GACRGDCLGA en 4WK0.	47
Figura 20.- Residuos en un radio de 5Å interactuando con el péptido GACRGDCLGA para los sistemas conjugados con las integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$, $\alpha V\beta 6$	51
Figura 21.- Desviación cuadrática media (RMSD) de las proteínas interactuando con el péptido GACRGDCLGA en función del tiempo de equilibrado.....	52
Figura 22.- Distancia entre el motivo RGD del péptido GACRGDCLGA y los residuos que conforman las subunidades α y β de la integrina $\alpha V\beta 3$ (1L5G).	53
Figura 23.- Distancia entre el motivo RGD del péptido GACRGDCLGA y los residuos que conforman las subunidades α y β de la integrina $\alpha 5\beta 1$ (4WK0).....	54

Figura 24.- Distancia entre el motivo RGD del péptido GACRGDCLGA y los residuos que conforman las subunidades α y β de la integrina $\alpha\text{V}\beta\text{6}$ (4UM9). 55

INDICE DE TABLAS

Tabla 1.- Características generales de los péptidos seleccionados.	21
Tabla 2.- Puntuaciones asociadas a patrón geométrico entre subunidades α	42
Tabla 3.- Puntuaciones asociadas a patrón geométrico entre subunidades β	43
Tabla 4.- Puntuaciones asociadas a modelos generados para péptidos de interés.	44
Tabla 5.- Valores de RMSD para 10 poses con ranking superior por complejo.	46
Tabla 7.- Resumen Interacciones RGD/NGR-Integrina en radio de 5Å 48	48
Tabla 8.- Valores de energía obtenidos a partir de los cálculos de MM-GBSA para los 3 complejos integrina-péptido.	56

1. INTRODUCCIÓN

1.1 Cáncer colorrectal

El cáncer colorrectal (colorectal cancer en inglés, CRC), también conocido como adenocarcinoma colorrectal, generalmente surge de las células epiteliales glandulares del intestino grueso, cuando determinadas células del epitelio adquieren una serie de mutaciones genéticas o epigenéticas que les confieren una ventaja selectiva (Ewing et al., 2014). El cambio en el mecanismo de reproducción celular se asocia con tasas de replicación y supervivencia aumentadas, por lo que estas células hiperproliferativas dan lugar a un adenoma benigno, que luego puede evolucionar a carcinoma y metástasis por décadas (Vogelstein et al., 1988).

Este tipo de patología ocupa el tercer lugar dentro de las muertes asociadas a cáncer en el mundo y se proyecta un aumento en su incidencia para países en desarrollo (Rawla et al., 2019). En 2020, se observó una incidencia mundial del 10%, y casi el mismo porcentaje de las muertes anuales por cáncer de tipo colorrectal (Figura 1) (Sung et al., 2021). Además, muchos rasgos o hábitos personales se consideran factores de riesgo, ya que aumentan las posibilidades de desarrollar pólipos o CRC (Mármol et al., 2017).

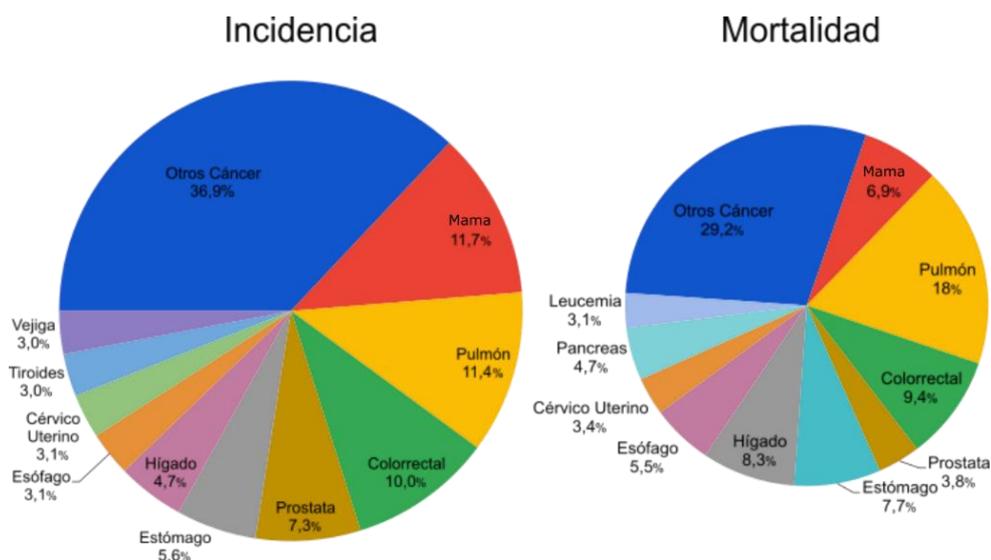


Figura 1.- Distribución de casos y muertes asociadas a los 10 cánceres más comunes en 2020 (Sung et al., 2021).

Se entiende por colorrecto el conjunto de colon y recto, los que además del ano, conforman el intestino grueso, correspondiente al segmento final en el sistema gastrointestinal (Figura 2). Las funciones del intestino grueso son: motilidad, secreción (endocrina y exocrina), absorción de agua, minerales, nutrientes y electrolitos de la materia alimentaria. Sumado a esto, el intestino también almacena y excreta las heces formadas (Fuentes Arderiu, 1998). La primera parte del intestino grueso es el colon y este músculo se divide en cuatro secciones (transverso, descendente, sigmoide y ascendente). El intestino grueso contiene una microflora diversa que puede descomponer los almidones y las proteínas restantes de las primeras etapas de digestión (Conlon & Bird, 2014).

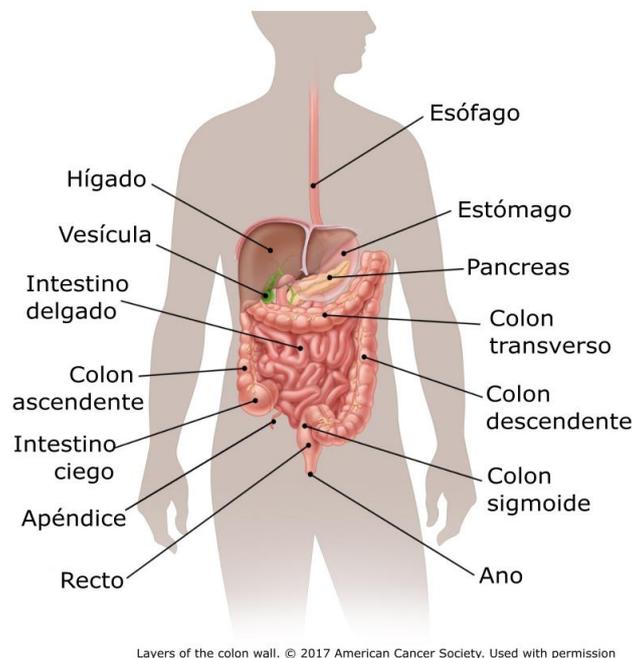


Figura 2.- Anatomía del sistema gastrointestinal (American Cancer Society, 2020). El cáncer colorrectal aborda las cuatro secciones del colon además del recto.

En una vista transversal, la pared intestinal se divide en cuatro capas. En la capa más externa se encuentra la serosa, una monocapa de células mesoteliales seguida de la muscularis externa o capa muscular gruesa. La submucosa es la siguiente capa con ubicación más central y contiene una rica mezcla de células, que incluye elementos estructurales como fibroblastos y tejido conectivo denso, sumado a células inmunológicamente importantes tales como: células plasmáticas, linfocitos, macrófagos, eosinófilos y mastocitos. Esta capa además comprende el tejido vascular y linfático. La muscularis mucosa, una capa delgada de músculo

liso, separa la submucosa más profunda de la mucosa, mientras que una lámina propia corre hacia el interior de esta capa y está compuesta de tejido conectivo y revestida por el epitelio luminal (Maqbool, 2013).

El CRC comienza en la capa más interna llamada mucosa, que agrupa al epitelio, tejido conectivo y submucosa. Esta capa puede crecer hacia afuera a través de algunas o todas las otras capas. Cuando las células cancerosas están en la pared, pueden crecer hasta convertirse en vasos sanguíneos o vasos linfáticos (Korphaisarn et al., 2019). Desde allí, pueden viajar a los ganglios linfáticos cercanos a partes distantes del cuerpo (American Cancer Society, 2020).

Como ocurre en otros tipos de cáncer, las mutaciones en oncogenes, genes supresores de tumores y genes relacionados con mecanismos de reparación del ADN, pueden provocar la aparición de CRC (Fearon & Vogelstein, 1990). Dependiendo del origen de la mutación, los carcinomas colorrectales pueden ser clasificados como esporádicos, heredados y familiares (Mármol et al., 2017). Cabe señalar que el CRC comprende a un grupo muy heterogéneo de enfermedades provocadas por una amplia gama de mutaciones y mutágenos que no comparten mutaciones impulsoras similares, lo que ha hecho difícil diseñar una terapia molecular genérica (Sideris & Papagrigoriadis, 2014).

El diagnóstico oportuno y la detección temprana del CRC pueden reducir la mortalidad de los pacientes (Singhal et al., 2014), aunque la tasa de supervivencia media con estados avanzados de esta enfermedad sigue siendo baja. Pese a los tratamientos quirúrgicos que existen para CRC cuando se diagnostica en sus etapas iniciales, en la mayoría de los casos es difícil eliminar por completo los tumores primarios o metastásicos (Siegel et al., 2020). Por lo tanto, la quimioterapia adyuvante se usa típicamente para suprimir el crecimiento tumoral y prolongar la supervivencia en pacientes, especialmente aquellos que albergan tumores en los que un tratamiento quirúrgico no es posible (Zhong et al., 2019). Sobre la base de estos antecedentes, encontrar mecanismos alternativos para el tratamiento del CRC en etapas más avanzadas, complementarios a la cirugía y quimioterapia, es de gran relevancia científica.

1.2 Integrinas

Las integrinas son una gran familia de glicoproteínas transmembranales, que forman heterodímeros de subunidades α y β unidas de manera no covalente (Li et al., 2020). Estas proteínas surgen como un blanco de estudio ya que participan en funciones reguladoras cruciales durante la adhesión celular, migración/invasión, proliferación, supervivencia y apoptosis (Nieberler et al., 2017). A nivel estructural la subunidad α cuenta con un gran dominio extracelular que comprende varias regiones con enlaces catiónicos y un dominio intracelular corto. Por su parte, la subunidad β tiene un gran dominio extracelular con secuencias repetidas de aminoácidos, conteniendo una gran cantidad de residuos de cisteína (Denadai et al., 2013). Cada subunidad tiene una única hélice que atraviesa la membrana y, por lo general, una cola citoplasmática corta (Figura 3). El tamaño varía, pero típicamente las subunidades α y β contienen alrededor de 1000 y 750 aminoácidos, respectivamente (Campbell & Humphries, 2011).

Estas proteínas actúan en la superficie de las células endoteliales vasculares y desempeñan un papel importante en la angiogénesis, migración de leucocitos y metástasis tumoral, lo que hace de las integrinas un blanco adecuado para el tratamiento de diversas enfermedades inflamatorias y cáncer (Vhora et al., 2015).

Desde su descubrimiento, las integrinas se han estudiado con gran detalle. Sin embargo, los mecanismos exactos de su activación molecular y señalización bidireccional permanecieron poco claros hasta que Xiong y colaboradores (2001) resolvieron la primera estructura del dominio extracelular o ectodominio de integrinas, considerando las isoformas $\alpha V\beta 3$ en ausencia y presencia de Cilengitide, un pequeño péptido cíclico. Con la información obtenida en este estudio, se determinó que los dominios extracelulares de $\alpha V\beta 3$ existen en un estado de reposo, adoptando una conformación en donde las cabezas de las integrinas están dobladas hacia la membrana celular y varían a un estado activo, en donde se extienden, cuando un ligando se une en una pequeña hendidura entre α y β (ver Figura 3).

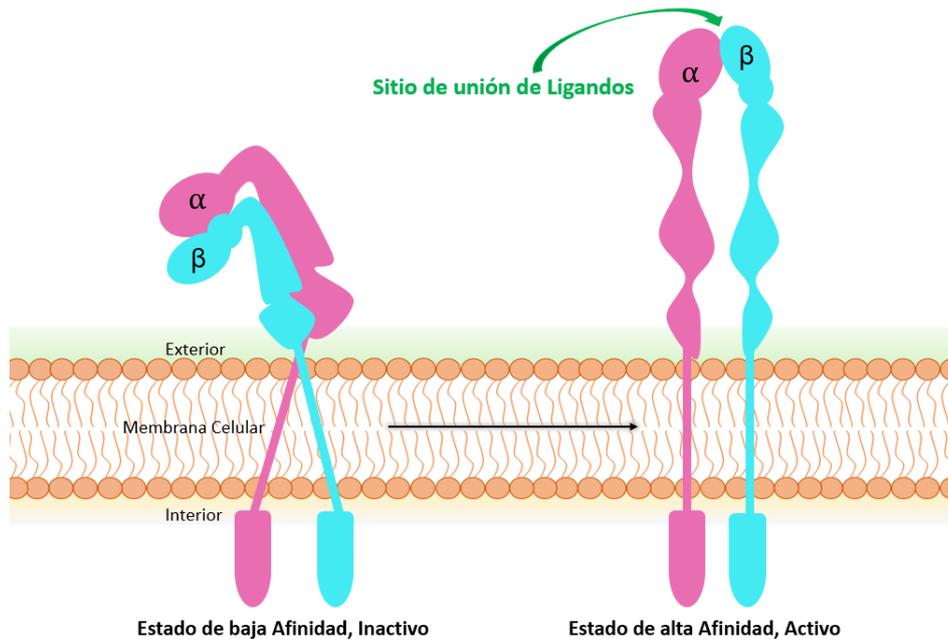


Figura 3.- Esquema de la estructura y activación de las integrinas. El heterodímero es dinámico y su conformación pasa a un estado extendido ante la presencia de ligandos. Adaptada de (Mezu-Ndubuisi & Maheshwari, 2020).

En integrinas de humanos, se han logrado identificar 18 variantes de subunidades α y 8 de subunidades β que pueden formar, en sus distintas combinaciones, 24 heterodímeros distintos con vínculo a diversos ligandos para mediar la adhesión entre las células y la membrana basal (Pinon & Wehrle-Haller, 2011). Las subunidades α y β no muestran homología entre sí, pero las diferentes variantes de subunidades α y β exhiben importantes similitudes (Barczyk et al., 2009). Cabe señalar, que el epitelio intestinal expresa una amplia variedad de integrinas con distintos tipos de ligandos (ver Figura 4), encontrándose complejos con claras preferencias de unión extracelular, tales como: lamininas para $\alpha3/\alpha6/\alpha7$ acopladas a $\beta1$ o $\beta4$, colágenos para $\alpha1/\alpha2$ acoplados a $\beta1$, tenascina para $\alpha9\beta1$ y ligandos que contienen un motivo RGD (fibronectina, osteopontina y vitronectina) para $\alpha5/\alpha8$ acoplado a $\beta1$ y αV acoplado a $\beta1/\beta3/\beta5/\beta6/\beta8$ (Beaulieu, 2019).

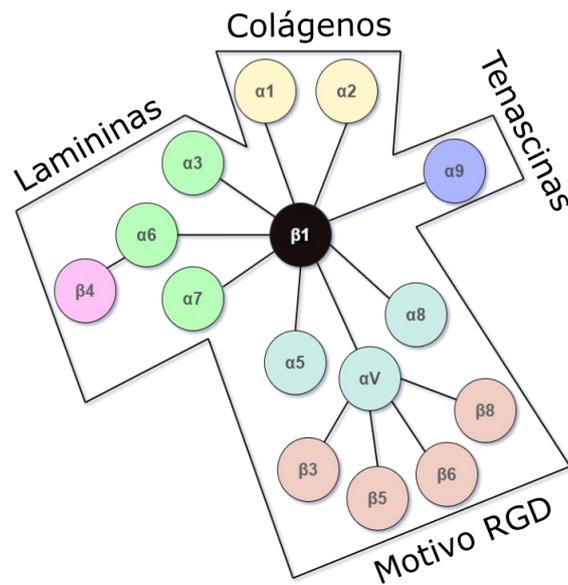


Figura 4.- Integrinas expresadas en células epiteliales, en condiciones normales o patológicas, organizadas según sus ligandos extracelulares. Las integrinas $\alpha\text{V}\beta 3$, $\alpha\text{V}\beta 5$, $\alpha\text{V}\beta 6$, $\alpha\text{V}\beta 8$ presentan como ligando extracelular a secuencias de proteínas que contengan el motivo RGD en su estructura.

1.3 Integrinas y cáncer colorrectal

Las integrinas juegan un papel importante en el inicio, la progresión y el desarrollo de metástasis en el CRC. Además, la sobreexpresión de integrinas puede estar asociada con un mal pronóstico en el desarrollo de la enfermedad (Le Tourneau et al., 2007). Particularmente, la integrina $\alpha\text{V}\beta 3$ se ha asociado a CRC luego de que en un modelo de ratón se implantaran células de cáncer de colon RKO y al ser tratadas con un antagonista de $\alpha\text{V}\beta 3$ se produjera una disminución del crecimiento tumoral y de la neovascularización. Adicionalmente, se observó un aumento del índice apoptótico del tumor *in vivo* (Kerr et al., 1999).

Por otro lado, tomando en consideración que el epitelio del colon reside sobre una membrana basal compuesta principalmente de lamininas, colágeno IV y proteoglicanos, la expresión de $\alpha\text{V}\beta 6$ facilitó el escape y la diseminación de células en presencia de un carcinoma invasor (Bates, 2005). Estudios adicionales muestran que $\alpha\text{V}\beta 6$ difícilmente puede detectarse en células epiteliales sanas (Breuss et al., 1993), mientras que un aumento significativo en su expresión se observó en respuesta a lesiones y/o inflamación, así como en tumores

epiteliales, como los que ocurren en CRC (Breuss et al., 1995). Por estos motivos, la integrina $\alpha\beta6$ ha resultado ser una diana terapéutica para el tratamiento del CRC (Peng et al., 2018).

Finalmente, la integrina $\alpha5\beta1$ se ha definido como un factor pro-angiogénico involucrado en la regulación de la angiogénesis tumoral al interactuar con el receptor del factor de crecimiento endotelial vascular (Li et al., 2020). Este caso se considera relevante ya que proteínas conteniendo el motivo RGD han demostrado interactuar con la integrina $\alpha5\beta1$ e inducir la adhesión celular, señalización intracelular y angiogénesis (Orecchia et al., 2003). Recientemente, Pelillo y colaboradores (2016) demostraron a través de un modelo experimental de células de cáncer colorrectal HCT-116, que el efecto anti-metastásico de NAMI-A, un fármaco a base de rutenio caracterizado por la actividad selectiva contra metástasis tumorales, depende de la modulación de la integrina $\alpha5\beta1$.

Cabe destacar, que las integrinas $\alpha\beta3$, $\alpha\beta6$ y $\alpha5\beta1$ no son las únicas que cumplen roles claves en CRC; sin embargo, la presente investigación considerará sólo este grupo.

1.4 Motivo RGD y NGR

En 1984 se describió por primera vez el motivo peptídico Arg-Gly-Asp (RGD) como una secuencia de unión a integrinas altamente conservada dentro de la fibronectina (Pierschbacher & Ruoslahti, 1984). Posteriormente, el motivo RGD se identificó en muchas otras proteínas de la matriz extracelular (ECM), incluidas la vitronectina (Suzuki et al., 1985), osteopontina (Oldberg et al., 1986), el factor Von Willebrand (Plow et al., 1985), y la laminina (Grant et al., 1989). Actualmente, se sabe que el motivo RGD no se encuentra restringido únicamente a interactuar con este tipo de conformaciones proteicas, sino que interactúa con más de 100 proteínas diferentes. En algunas proteínas, la actividad de adhesión celular se ha atribuido a la secuencia RGD, mientras que en otras la secuencia RGD parece estar funcionalmente silenciada (D'Souza et al., 1991).

El motivo RGD corresponde a un péptido con una alta especificidad en su actividad, ya que cambios tan pequeños como la sustitución del ácido aspártico por un ácido glutámico o de la glicina por una alanina han reducido significativamente la actividad del péptido en ensayos de unión celular (Hautanen et al., 1989). La conformación u orientación espacial de los

aminoácidos dentro de la secuencia también es relevante; por ejemplo, un péptido en el que el ácido aspártico está en forma D es inactivo, mientras que aquel en que la arginina es una D-arginina, forma un péptido activo (Pierschbacher & Ruoslahti, 1987).

Entre los 24 subtipos de integrina humana conocidos hasta la fecha, ocho dímeros de integrina ($\alpha V\beta 1$, $\alpha V\beta 3$, $\alpha V\beta 5$, $\alpha V\beta 6$, $\alpha V\beta 8$, $\alpha 5\beta 1$, $\alpha 8\beta 1$ y $\alpha IIb\beta 3$) reconocen el motivo RGD dentro de las proteínas de la matriz celular (Nieberler et al., 2017). Estos miembros constituyen la subfamilia de receptores de integrina más influyentes en cáncer y procesos de metástasis (Schittenhelm et al., 2013). Sobre la base de estos hallazgos, se han desarrollado péptidos sintéticos y péptido-miméticos que contienen el motivo RGD, para ser usados como eficaces ligandos y antagonistas de integrinas. Tales péptidos han logrado inhibir la adhesión de células dependientes de anclaje a proteínas de la matriz celular, controlando así las funciones biológicas que desarrollan tumores mediados por integrinas (Nieberler et al., 2017).

Dentro del mismo contexto, el motivo NGR, nombrado así por su secuencia Asn-Gly-Arg, fue identificado en la década de 1990 mediante la selección *in vivo* de bibliotecas de péptido-fagos en micrófagos portadores de tumores (Arap et al., 1998). Estudios dedicados a entender el proceso biológico que involucra a este motivo en la tumorigénesis, mostraron que NGR puede reconocer específicamente los vasos sanguíneos que expresan la aminopeptidasa N (CD13). Esta metaloproteinasa apenas se expresa en los vasos sanguíneos normales, pero está regulada al alza en los vasos sanguíneos angiogénicos (Pasqualini et al., 2000).

En los tejidos tumorales, CD13 se expresa en células endoteliales y pericitos y, en algunos casos, en células tumorales y fibroblastos (Corti et al., 2013). Debido a esta propiedad, péptido-miméticos que contienen el motivo NGR se han aprovechado para la administración dirigida de varias moléculas tipo fármaco a los vasos tumorales, en un intento por aumentar su actividad antitumoral (Corti et al., 2008).

Xiong y colaboradores (2002) revelaron la estructura cristalizada de $\alpha V\beta 3$ en complejo con un péptido conteniendo el motivo RGD. El sitio de interacción proteína-péptido se describió en una hendidura poco profunda ubicada entre las subunidades αV y $\beta 3$ de la cabeza de la integrina (ver Figura 5). Solo las cadenas laterales de la Arg y Asp del péptido RGD se

encontraron ligeramente sumergidas dentro de la cavidad, mientras que el resto del péptido fue expuesto al solvente. Además, se observó que cationes divalentes juegan un papel crítico en la mediación de la unión de integrinas a péptidos RGD (Craig et al., 2004).

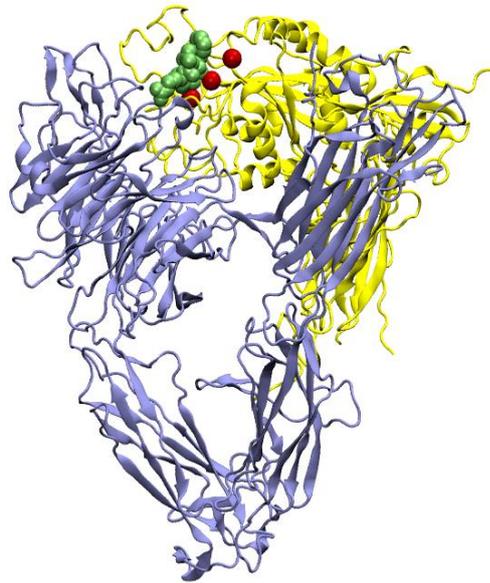


Figura 5.- Representación tridimensional de la interacción del péptido RGDFV cíclico (cilengitide) y la integrina $\alpha\beta3$. La integrina $\alpha\beta3$ se muestra dividida entre subunidad α (celeste) y subunidad $\beta3$ (amarilla). El péptido cíclico RGDFV se representa en esferas VDW verdes, mientras que los iones metálicos (Mn^{2+}) cercanos al péptido se representan en rojo.

Sumado a lo anterior, Yu y colaboradores (2014) describieron las bases moleculares de la interacción entre un tripéptido RGD y la integrina $\alpha\beta3$. El mecanismo de unión se estableció principalmente por interacciones electrostáticas, entre residuos de la integrina y iones metálicos que contribuyen a la unión del tripéptido RGD, e interacciones del tipo puente salino, los que se evaluaron según la distancia entre los átomos donadores y aceptores considerando un valor menor a 4Å. El modo de unión más detallado de este complejo se muestra en la Figura 6, donde la fracción Arg del péptido interactúa con los residuos Asp-218 y Asp-150 de αV . Asimismo, el carboxilato del residuo Asp del ligando se coordina con el sitio de adhesión dependiente de iones metálicos Mn^{2+} , que a su vez está coordinado por los residuos Glu-220, Ser-123, Ser-121 y presumiblemente por dos moléculas de agua (Zheng & Leftheris, 2020).

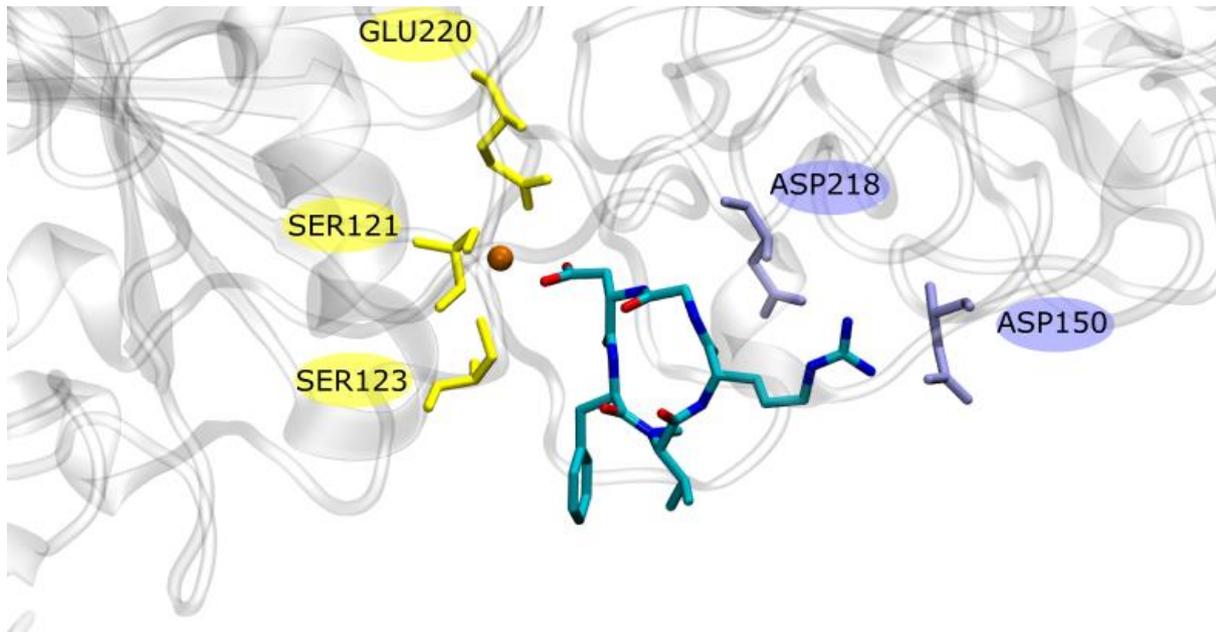


Figura 6.- Modo de unión del péptido Cilengitide a la integrina $\alpha V\beta 3$. Estructura con código PDB 1L5G obtenida a 3,2 Å de resolución (J.-P. Xiong et al., 2002).

1.5 Péptidos contra el cáncer

La idea de diseñar formulaciones médicas basadas en péptidos para el reconocimiento molecular se construye sobre el fundamento de mejorar la eficacia terapéutica de fármacos conjugados y reducir los efectos secundarios mediante la selección de receptores específicos en las células (Ruoslahti, 1996). En este sentido, identificar péptidos que se anclen a integrinas involucradas en CRC podría ser de gran ayuda en el diseño de nuevas terapias.

Existe evidencia de que la terapia sitio dirigida basada en el diseño de fármacos conjugados a péptidos es factible. Sin ir más lejos, terapias adicionales para tratar cáncer y otras enfermedades inflamatorias incluyen péptidos con motivos RGD que pueden usarse para sintetizar conjugados péptido-fármaco (Vhora et al., 2015). Además de péptidos conteniendo la secuencia RGD, la secuencia NGR también ha sido estudiada para la administración dirigida de fármacos al sistema vascular que expresa integrinas y células tumorales (Corti et al., 2008).

Por lo tanto, este motivo podría usarse también como una alternativa para el reconocimiento molecular.

Respecto al motivo NGR, se cuenta con una serie de fármacos contra cáncer conjugados a este motivo, siendo la doxorrubicina la primera droga unida a un péptido NGR (CNGRC) (Arap et al., 1998). Este primer conjugado mostró una toxicidad reducida y una eficacia mejorada contra los xenoinjertos de cáncer humano en ratones, en comparación con la doxorrubicina libre. Otros estudios basados en la selección *in vitro* de bibliotecas de péptidos con acción frente a $\alpha V\beta 3$ revelaron varios fragmentos que contienen el motivo NGR, tales como: NGRIPD, TNGRGP, NGRSFR, RSRNGR, y NGRNTV (Healy et al., 1995).

Tomando como referencia estos estudios, para el presente trabajo de memoria de título se seleccionó un conjunto de péptidos que contienen los motivos RGD y NGR para evaluar su interacción frente a las integrinas $\alpha V\beta 3$, $\alpha V\beta 6$ y $\alpha 5\beta 1$. Para RGD, se identificaron los péptidos GAC*RGDC*LGA (donde * indica cisteínas que forman un enlace disulfuro) (Koivunen et al., 1993), c(RGDFV) (Mas-Moruno et al., 2010) y CRGDRGPDC, también conocido como iRGD (Zhong et al., 2019). De la misma forma se seleccionó para NGR: CNGRC (Wei et al., 2013), NGRAHA y CVLNGRMEC (Corti & Curnis, 2011).

Tabla 1.- Características generales de los péptidos seleccionados.

Péptido	Motivo Unión	Características
GAC*RGDC*LGA	RGD	Péptido sintético cíclico, en donde cisteínas flanquean el motivo RGD, con 10 veces más poder de inhibición que su contraparte lineal GRGDSP. Su acción se evaluó contra integrina $\alpha 5\beta 1$ y αV (Koivunen et al., 1993).
c(RGDFV)	RGD	Pentapéptido cíclico descubierto en 1995, con actividad antagonista subnanomolar para el receptor $\alpha V\beta 3$. También se observaron afinidades nanomolares frente a integrinas $\alpha V\beta 5$ y $\alpha 5\beta 1$.
CRGDRGPDC	RGD	Facilita la acumulación activa de fármacos en los tumores, por la sobreexpresión en los vasos tumorales. En

		experimentos <i>in vivo</i> conjugado con nanopartículas, este péptido promovió significativamente la acumulación de las mismas con preferencia por xenoinjertos de tumores colorrectales (Zhong et al., 2019).
CNGRC	NGR	Péptido corto con la capacidad de unirse específicamente a CD13/aminopeptidasa N <i>in vivo</i> (Luan & Xu, 2007).
NGRAHA y CVLNGRMEC	NGR	Experimentalmente estos péptidos lograron inhibir la adhesión celular, especialmente la adhesión mediada por las integrinas $\alpha V\beta 5$ y $\alpha V\beta 3$ (Koivunen et al., 1993). En modelos de ratón mostraron ser selectivos por células tumorales (Arap et al., 1998).

1.6 Caracterización de asociaciones integrina-péptido

En base a los antecedentes antes expuestos, el presente trabajo de investigación se orienta al estudio de tres isoformas de integrinas ($\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$), las cuales han sido identificadas como proteínas relevantes en la proliferación y supervivencia de células cancerígenas en colon y recto. La asociación de estas integrinas fue evaluada frente a un conjunto de seis péptidos conteniendo los motivos RGD: GAC*RGDC*LGA, c(RGDFV) y CRGDRGPDC; y NGR: CNGRC, NGRAHA y CVLNGRMEC. Las asociaciones integrina-péptido fueron estudiadas a nivel estructural empleando cálculos de acoplamiento y dinámica molecular, con el fin de determinar conformaciones favorecidas de los complejos. Posteriormente, las conformaciones más afines fueron sometidas a una caracterización energética a través de estimaciones de energía libre usando el método MM-GBSA. El objetivo a largo plazo, como muestra la Figura 7, es identificar complejos de alta afinidad que permitan sitio-dirigir formulaciones médicas conjugadas con péptidos que se anclen a integrinas involucradas en cáncer.

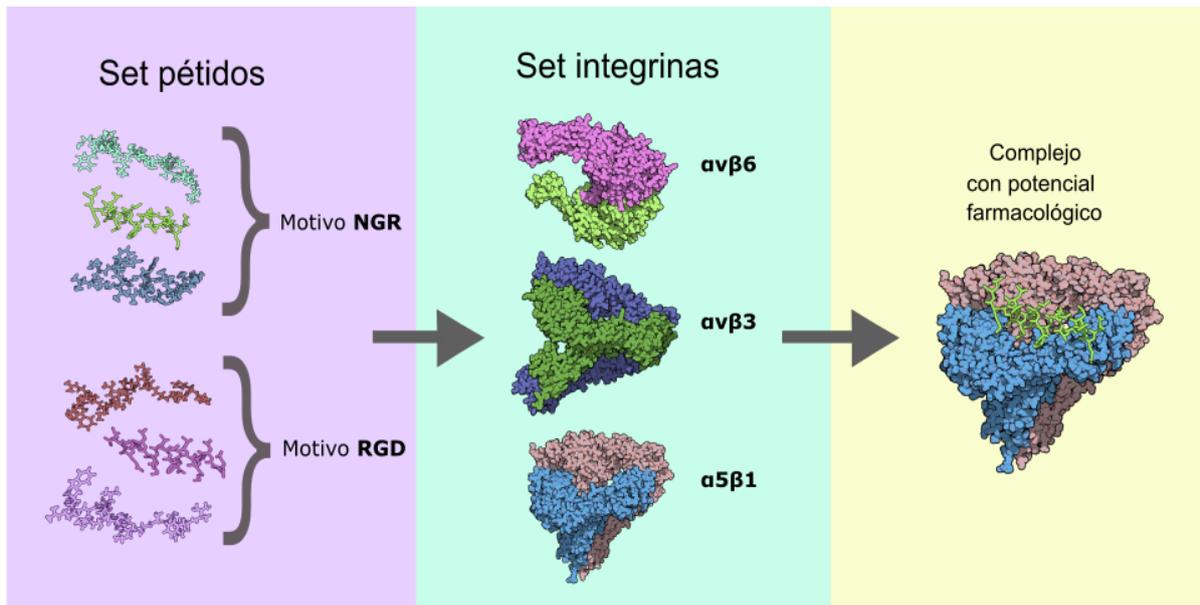


Figura 7.- Esquema de trabajo de investigación. En el recuadro morado se presenta el conjunto de péptidos identificados para este estudio incluyendo secuencias que contienen el motivo de unión NGR y RGD. La asociación de estas secuencias se estudió frente a tres integrinas relevantes en CRC, representadas en el recuadro color verde agua. El objetivo principal es encontrar complejos con potencial farmacológico frente a CRC (recuadro amarillo).

2. HIPÓTESIS

Las integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$, implicadas en cáncer colorrectal, comparten dominios conservados de secuencia y estructura que permiten el anclaje de fragmentos peptídicos conteniendo los motivos RGD y NGR, lo cual puede ser evaluado a través de técnicas de modelado y simulación molecular.

3. OBJETIVOS

3.1. Objetivo General

Determinar características conservadas en las integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$, implicadas en cáncer colorrectal, y evaluar su unión a péptidos conteniendo los dominios de reconocimiento molecular RGD y NGR.

3.2. Objetivos Específicos

1. Evaluar elementos conservados a nivel de estructura y secuencia en las integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$, involucradas en cáncer colorrectal.
2. Identificar por técnicas *ab initio* el plegamiento de péptidos conteniendo los motivos RGD y NGR de reconocimiento a integrinas.
3. Determinar conformaciones favorecidas entre los péptidos que contienen los motivos RGD y NGR y las tres integrinas consideradas en el estudio.
4. Caracterizar interacciones claves y la energía de unión de un conjunto de complejos integrina-péptido de variada afinidad.

4.- MATERIALES Y MÉTODOS

El estudio estructural es un punto clave en la comprensión de la biología de las proteínas. La función de complejos biológicos se basa en asociaciones moleculares, y estas son consecuencia de interacciones enlazantes y no enlazantes (Hospital et al., 2015). Las integrinas son una familia de heterodímeros α/β transmembrana que forman interacciones no covalentes y existen en un equilibrio entre conformaciones en reposo- baja afinidad- y activas- alta afinidad (Bennett, 2005). Tales complejos en interacción con proteínas y moléculas circundantes se encuentran regulando la adhesión, crecimiento, migración y diferenciación de células epiteliales intestinales (Sordat et al., 1998). Por lo tanto, considerando la importancia biológica de las integrinas es que se proponen los siguientes métodos para evaluar la interacción entre un conjunto de péptidos de interés en complejo con 3 integrinas involucradas en cáncer colorrectal. De manera general, los estudios planteados tienen como fin caracterizar la unión de las integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$ con péptidos que contengan los motivos RGD o NGR.

4.1 Revisión de estructuras de integrinas $\alpha 5\beta 1$, $\alpha V\beta 3$ y $\alpha V\beta 6$

Las estructuras bases utilizadas en este trabajo corresponden a los cristales 4UM9 (Dong et al., 2014), 1L5G (J.-P. Xiong et al., 2002) y 4WK0 (Xia & Springer, 2014) para las integrinas $\alpha V\beta 6$, $\alpha V\beta 3$ y $\alpha 5\beta 1$, respectivamente. La información de cada proteína fue obtenida a partir de la base de datos PDB (Berman et al., 2000) y Uniprot (The UniProt Consortium, 2021). Para la selección de las estructuras se consideró que estén expresadas en humanos, sean de alta resolución y contengan la parte exterior o *headpiece* de la integrina, ya que esta región es la que interactúa con ligandos y péptidos. Adicionalmente, considerando el comportamiento dinámico de la integrina, se tomaron cristales conjugados con péptidos RGD, ya que como se muestra en la Figura 3, en presencia de un ligando la estructura se encuentra en su estado extendido y en una conformación óptima para interactuar con moléculas externas (Springer & Dustin, 2012).

Para determinar patrones claves en la unión de integrinas a péptidos conteniendo los motivos RGD/NGR, se analizaron las estructuras de las integrinas con péptidos co-cristalizados

mediante el uso del programa Pymol (Schrödinger, LLC, 2015). Específicamente, se evaluaron los residuos a una distancia de 6Å de cada uno de los aminoácidos del motivo RGD y su nivel de conservación entre las 3 integrinas analizadas.

4.2 Identificación de residuos que determine la unión de integrinas a péptidos RGD/NGR

El análisis de la secuencia de aminoácidos en proteínas es capaz de evaluar características que determinen su función. En algunos casos, la secuencia de una proteína se relaciona de manera distante con cualquier otra proteína conocida limitando detectar su semejanza por métodos de alineación. Sin embargo, las relaciones pueden revelarse por la aparición en su secuencia de un grupo particular de residuos que desempeñan una función específica. En el caso de las integrinas, este análisis se realizó con el fin de encontrar regiones o residuos con significancia biológica en la unión de dos motivos específicos, NGR y RGD.

4.2.1 Búsqueda de patrones a nivel de secuencia

Como los residuos biológicamente significativos de las proteínas se conservan evolutivamente (Kimura, 1983), una alineación múltiple de secuencias puede reducirse a una expresión de consenso denominada expresión o patrón regular. Para determinar residuos de interés, se realizó un alineamiento múltiple de las secuencias de las integrinas $\alpha V\beta 6$, $\alpha V\beta 3$ y $\alpha 5\beta 1$ usando el servidor ClustalW (Sievers et al., 2011). Las secuencias se ingresaron en formato FASTA, obteniendo como output los alineamientos y un archivo de visualización para analizar con Jalview (Waterhouse et al., 2009). A partir de este análisis, integrado con el paso previo, que permitió identificar la esfera de residuos de interés, se determinaron los patrones de consenso más relevantes que determinan la unión de las 3 integrinas en este estudio a fragmentos peptídicos conteniendo los motivos de unión RGD y NGR.

4.2.2 Búsqueda de patrones geométricos

Las superficies de las proteínas son relativamente grandes y existe la posibilidad de interacciones múltiples con ligandos pequeños y otras proteínas. Algunas interacciones pueden resultar en una modificación alostérica de la función originalmente definida de la

proteína, mientras que otras pueden revelar una nueva función (Wong et al., 2015). Encontrar patrones a nivel de estructura puede implicar descubrimientos relevantes a niveles evolutivos, ya que usualmente las proteínas se encuentran más conservadas a nivel estructural que de secuencia (Marchler-Bauer et al., 2002). Para detectar patrones geométricos en la superficie de las integrinas se utilizó el servidor web GeomFinder (Núñez-Vivanco et al., 2016). GeomFinder realiza una estimación de similitudes comparando todos los pares de patrones 3D en un conjunto de proteínas en estudio. Los patrones relevantes en las integrinas se analizaron considerando residuos cercanos a los sitios de unión de péptidos reportados en los cristales.

La versión más actual de GeomFinder, versión 2, permite únicamente comparación de pares de estructuras, y para los análisis del set de integrinas se consideraron sus subunidades de manera independiente. Por lo que en el caso de las subunidades β , en donde se debía analizar las subunidades β_6 , β_3 y β_1 se estableció una comparación en 2 pasos, comparando primeramente β_6 contra β_3 , y luego β_3 contra β_1 , los cuales por transitividad corresponderían a patrones entre las 3 subunidades. Seguido a lo anterior se contrastaron los patrones geométricos identificados contra los residuos considerados como claves en el punto 4.1.

Este paso no debió ser replicado en las subunidades α , ya que α_V se encuentra en dos de nuestras 3 integrinas de interés, teniendo que ser analizado únicamente frente a α_5 . La búsqueda de patrones entre ambas subunidades, siendo analizadas en conjunto, no indica la existencia de hits que existan en la superficie que se crea entre ambas, por lo que no se añaden patrones combinando ambas subunidades a los análisis posteriores.

4.3 Modelado de péptidos

Las secuencias de todos los péptidos seleccionados para evaluar su interacción con integrinas (ver Tabla 1), no exceden los 20 aminoácidos. Actualmente, es posible obtener una aproximación de su estructura utilizando herramientas de predicción y modelado de estructura secundaria. Para obtener la estructura de los péptidos se utilizó el servidor PEP-FOLD V3 (Lamiable et al., 2016), a excepción del péptido 1, el cual no fue modelado pues se

tomó como referencia el péptido co-cristalizado en 1L5G.

PEP-FOLD V3 es un enfoque *de novo* destinado a predecir estructuras de péptidos a partir de secuencias de aminoácidos, generando modelos en base a cadenas ocultas de Markov. El método emplea un campo de fuerza *coarse-grained* que permite el modelado de péptidos mediante ciclos de simulaciones moleculares, generando conformaciones 3D de péptidos de hasta 80 aminoácidos. El servidor entrega una lista de todos los modelos generados, el detalle de los clústeres y la mejor conformación de cada uno de los 5 mejores clústeres. Los mejores modelos fueron seleccionados tomando como criterio el menor valor de energía sOPEP, el cual representa la energía *coarse-grained* de PEP-FOLD. Este valor es obtenido desde el potencial OPEP (Potencial optimizado para una predicción de estructuras eficiente, por sus siglas en inglés), que permite limitar la rugosidad del paisaje energético de los péptidos simplificando la representación de las cadenas laterales, las que son agrupada en cuentas únicas o *beads*. Los parámetros de OPEPv3 se optimizaron mediante un procedimiento de algoritmo genético, que se expresa como función objetivo con una suma de términos locales, no enlazados y de enlace de hidrógeno (Maupetit et al., 2007).

Parámetros especiales fueron seleccionados para inducir la formación de puentes disulfuro, en el caso del péptido GAC*RGDC*LGA, en donde los asteriscos indican los dos residuos de cisteína que conformarán la interacción disulfuro.

4.4 Acoplamiento molecular Integrina-péptido

Los péptidos se pliegan al unirse a otra macromolécula mediante una gran cantidad de interacciones reguladoras y pueden proporcionar módulos de reconocimiento altamente específicos. Para tener una aproximación de la afinidad entre los péptidos escogidos y las integrinas seleccionadas para este estudio, se utilizó la técnica de acoplamiento molecular o *docking*. Esta metodología es clave en la biología molecular estructural y ayuda considerablemente al diseño de fármacos asistido por computadora. La técnica de *docking* tiene como objetivo predecir el/los modo/s de unión predominantes de un ligando con una proteína de estructura tridimensional conocida (Morris & Lim-Wilby, 2008). AutoDock CrankPep (Zhang & Sanner, 2019) o ADCP, por sus siglas en inglés, es un motor de acoplamiento AutoDock especializado para péptidos. Se ha demostrado que vuelve a acoplar

con éxito péptidos con hasta 20 aminoácidos de longitud (Zhang & Sanner, 2019).

Este protocolo de acoplamiento *de novo* combina el plegamiento del péptido y el modelado del receptor respecto al mismo, para lo anterior utiliza una búsqueda de Monte-Carlo para plegar el péptido y al mismo tiempo optimizar la interacción entre el péptido y la molécula receptora, produciendo estructuras de péptidos acoplados.

4.4.1 Preparación del acoplamiento molecular

Dentro de los pasos previos al acoplamiento se considera la construcción de mapas de afinidad, los que fueron calculados para describir de manera eficiente a los receptores en el acoplamiento automático. Estos mapas aceleran enormemente el proceso de acoplamiento y permiten a los usuarios experimentar con el campo de fuerza (Zhang et al., 2019). El cálculo de mapas de afinidad se realizó utilizando AutoDock4, mediante la interfaz gráfica AutoGridFR (Ravindranath et al., 2015), considerando un relleno de 4 Å en cada lado del péptido de mayor longitud, siguiendo la metodología sugerida por Zhang & Sanner (2019). La caja de *docking* además fue centrada en los ligandos RGD co-cristalizados en cada una de las 3 integrinas de interés, con un espaciado de cuadrícula de 0.375 Å. Otro factor considerado en la construcción de los mapas fue la búsqueda de sitios de unión dentro de la caja de interés mediante Auto Site 1.1 (Zhang et al., 2019). Esta última herramienta permite la descripción de los bolsillos de unión detectados automáticamente en los receptores y contiene un algoritmo para agregar gradientes de afinidad que ayudan a los métodos de búsqueda a optimizar la solución utilizando menos evaluaciones de las funciones de puntuación. Finalmente, los mapas fueron generados para todos los tipos de átomos presentes y se almacenaron en un solo archivo junto con los metadatos que respaldan la procedencia de los datos y su reproducibilidad para tener una gestión facilitada.

4.4.2 Acoplamiento molecular con Auto Dock CrankPep

La ejecución de ADCP se realizó generando 100 confórmeros y 10.000.000 réplicas para cada uno de los 18 complejos integrina-péptido. En todos los casos se entregó como referencia el péptido modelado en el paso anterior con PEPFOLD, indicando también la secuencia de aminoácidos que componen al mismo. Dentro del conjunto de péptidos, 2 de éstos necesitaron parámetros especiales de ejecución: 1.- En el caso de c(RGDFV) se agregó el parámetro -c para indicar la ciclicidad de la cadena principal en este, reportada en bibliografía.

2.- En el caso del péptido GAC*RGDC*LGA se agregó el parámetro -C para indicar la formación de puentes disulfuro en su estructura. Los 4 péptidos restantes no requirieron de parámetros extras en la ejecución.

4.4.3 Criterio de término y clusterización de poses

Para los complejos en estudio, se realizó una evaluación inicial de todas las combinaciones de pares posibles entre proteína-péptidos, teniendo un total de 18 complejos. Para estudios posteriores, se escogieron sólo 3 complejos.

El flujo que ADCP implementado para obtener un muestreo de poses se puede ver en la Figura 8, en donde se aplica un movimiento Monte Carlo (MC) seleccionado al azar para alterar la pose obtenida. Seguido a esto, se puntúa la pose modificada y el movimiento se acepta o se rechaza según un criterio de MC. Si se rechaza el movimiento, se restaura la pose antes del movimiento y se intenta otro movimiento. Si se acepta, la pose alterada se convierte en la actual y se usa para actualizar el caché de las poses de acoplamiento. Este procedimiento se repite hasta que se cumple uno de los criterios de terminación: si se alcanza el número máximo de pasos o no se ha mejorado la mejor puntuación para 10 millones de pasos.

Una vez que todas las réplicas terminan la búsqueda de nuevos conformeros, las poses de acoplamiento se agrupan en clústeres, para elegir las poses de acoplamiento óptimas. La agrupación se realizó utilizando AutoDockFR (ADFR), con el fin de eliminar las soluciones duplicadas de la población, asegurando así la diversidad dentro del muestreo, sumado a la implementación del elitismo adaptativo que agrega automáticamente a la próxima generación la mejor solución de cada grupo.

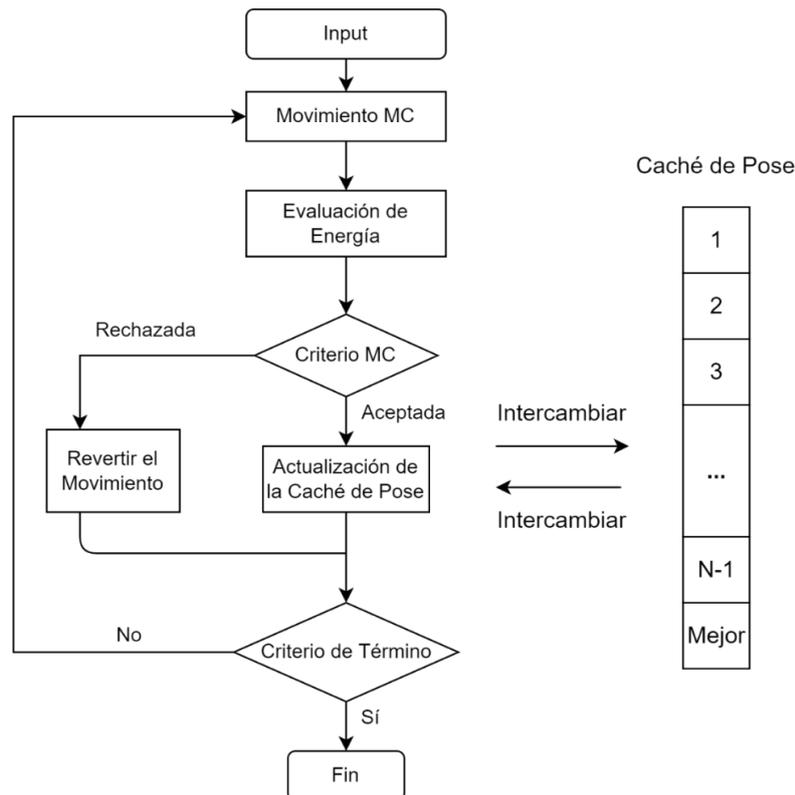


Figura 8.- Esquema de flujo para elección de mejor pose en ADCP. Las 100 poses generadas en este paso son luego agrupadas en clústeres con ADFR.

La agrupación se realiza utilizando la solución de menor energía, aún no agrupada, como semilla de grupo, y agregando al grupo todas las soluciones con RMSD menor que 2Å , para átomos del péptido, con respecto a la semilla de grupo. El procedimiento se repite hasta que todas las soluciones que se van a agrupar pertenecen a un grupo. A continuación, se crea una población de apareamiento, que contiene el mejor individuo de cada grupo junto con todos los individuos no agrupados. El mejor individuo de cada grupo se copia automáticamente en la próxima generación (elitismo adaptativo). Luego, el algoritmo genético (GA, por sus siglas en inglés) selecciona a los padres para cruzar, mutar y minimizar para generar descendencia, que compite con sus padres para agregarse a la población de la próxima generación. Todos los individuos creados se someten a un rápido paso de minimización. Si el individuo minimizado tiene un puntaje que es mejor que el puntaje de referencia (el mejor puntaje visto hasta ahora), sufre una minimización más agresiva y su puntaje se convierte en el puntaje de referencia. Los dos mejores individuos identificados durante este procedimiento de apareamiento se agregan a la siguiente generación si aún no están presentes en esa población. Una vez que se completa la población para la próxima generación, se convierte en

la población entrante para la próxima generación en el ciclo de optimización de GA.

El experimento de acoplamiento ADFR realiza 50 evoluciones GA independientes, cada una de las cuales produce una solución. Luego, estas soluciones se agrupan para eliminar las soluciones duplicadas y se informa el individuo con la mejor puntuación de cada grupo, lo que da como resultado una lista clasificada por “ranking” de soluciones para el acoplamiento. La pose escogida en cada caso corresponde a la solución con menor energía después de la clusterización tomando la agrupación con ranking 1.

4.5 Simulaciones moleculares de complejos proteína-péptido de alta afinidad

De los 18 complejos que fueron modelados en base a acoplamiento molecular proteína-péptido, sólo los 3 complejos de mayor afinidad y proyección o poses favorecidas se utilizaron como punto de partida para realizar simulaciones de dinámica molecular.

4.5.1 Elección de sistemas para simulaciones de dinámica molecular

Tomando la pose generada con ranking 1 de la clusterización previa se realizó una comparación al *score* de *docking* entregado por ADCP para cada complejo. La función de puntuación de ADCP consta de dos componentes: una puntuación para la conformación del péptido; y una puntuación para la interacción entre el péptido y el receptor, considerándose más favorables el valor mientras más negativo sea.

Sumado al análisis energético, y visualizando de interacciones con el ligando se realizó un análisis de los residuos que podrían interactuar dentro de un *cutoff* de 6Å con el mismo método empleado en la identificación de residuos conservados. Esto con el fin de considerar una posible esfera de interacción del péptido, destacando los motivos conservados en el análisis original de estructuras co-cristalizadas con péptidos RGD. En base a lo anterior, se escogió al péptido GAC*RGDC*LGA para caracterizar más en detalle su interacción con las integrinas. Esto por ser aquel con mayor afinidad promedio frente a las 3 integrinas y adoptar una ubicación más cercana a residuos identificados como clave en el anclaje del motivo RGD con las subunidades. Además, este péptido se posicionó entre ambas subunidades sin

establecer una preferencia en la orientación, lo que permitirá evaluar el comportamiento del heterodímero en conjunto.

4.5.2 Construcción de sistemas para simulaciones de dinámica molecular

Al realizar las simulaciones de dinámica molecular, se busca replicar el ambiente biológico en que se encuentran expresadas las integrinas, imitando las condiciones celulares originales. Para este fin, solo la región extracelular de las integrinas fue considerada en las simulaciones, dado que los péptidos interactúan únicamente con este segmento. Los sistemas fueron solvatados con una caja de agua tipo TIP3P utilizando condiciones periódicas de borde. Se agregó además al sistema una concentración 0.15 M de NaCl con el fin de neutralizar cargas e imitar las condiciones fisiológicas de las proteínas. La temperatura del sistema se ajustó en 310 K y la presión se mantuvo constante a 1 atm. Se utilizaron condiciones isobáricas-isotérmicas (NPT) en las simulaciones con el propósito de determinar cómo cambia la energía interna del sistema a una temperatura y presión constante.

4.5.2 Simulaciones de dinámica Molecular

La dinámica molecular (MD) simula el movimiento natural de los átomos que componen un sistema. La energía proporcionada en un procedimiento de dinámica molecular permite que los átomos se muevan e incluso colisionen con las partículas cercanas (Pitman & Menz, 2006). Esta técnica se utiliza habitualmente para modelar los movimientos o trayectorias dependientes del tiempo de macromoléculas biológicas. En implementaciones convencionales de MD, las fuerzas que actúan sobre los átomos se calculan como derivadas de potenciales. Estas fuerzas se sustituyen en las ecuaciones de movimiento de Newton, que se resuelven mediante mecánica molecular iterativamente para cada partícula del sistema (Gunsteren & Berendsen, 1990), siendo un proceso que si bien es preciso, también es computacionalmente costoso (Krokhotin & Dokholyan, 2015).

En el contexto de las simulaciones de MD, el término campo de fuerza se refiere a la combinación de una fórmula matemática y parámetros asociados que se utilizan para describir la energía de un sistema en función de sus coordenadas atómicas. Un campo de fuerza consta tanto del conjunto de ecuaciones utilizadas para calcular la energía potencial y las fuerzas a partir de las coordenadas de las partículas, como de una colección de parámetros

utilizados en las ecuaciones. Para la mayoría de los propósitos, estas aproximaciones funcionan bien, pero no pueden reproducir efectos cuánticos como la formación o ruptura de enlaces. Todos los campos de fuerza comunes subdividen las funciones potenciales en dos clases. Las interacciones enlazantes cubren estiramiento de enlaces covalentes, potenciales de torsión, diedros, flexión de ángulo y potenciales de impropios o fuera del plano, todos los cuales normalmente se fijan a lo largo de una simulación. Las interacciones no enlazantes restantes consisten en repulsión y atracción de Lennard-Jones, así como fuerzas electrostáticas de Coulomb (Kukol, 2008).

Las simulaciones de MD para los complejos integrina-péptido se realizarán utilizando el programa NAMD v2.13 (Kalé et al., 1999), incluyendo los archivos de topología y parámetros del campo de fuerza CHARMM v36 (MacKerell et al., 1998). El campo de fuerza de CHARMM consiste en una colección de términos que describen la energía potencial de un sistema. Los parámetros en el campo de fuerza, como las cargas parciales y los valores de equilibrio de las longitudes de los enlaces, ángulos y ángulos diedros se desarrollan en función de tipos de átomos únicos.

Los sistemas en estudios fueron minimizados usando 30,000 pasos. Seguido a esto, los complejos integrina-péptido fueron equilibrados utilizando restricciones armónicas a los carbonos α de la proteína. La temperatura constante se impuso aplicando el termostato Langevin con un coeficiente de amortiguación de 1 ps^{-1} . Además, el método de pistón de Langevin (Feller et al., 1995) se utilizó para mantener una presión constante. Las interacciones electrostáticas de largo alcance se calcularon utilizando el método de adición Particle-Mesh Ewald (Essmann et al., 1995), con un cutoff aplicado entre 8 y 9 Å. Todos los enlaces covalentes que involucran hidrógeno, así como las geometrías intramoleculares de las moléculas de agua, se restringieron utilizando el algoritmo SETTLE (Miyamoto & Kollman, 1992). Se utilizó el integrador de múltiples time-steps Verlet-I / r-RESPA (Tuckerman et al., 1992) con un time step de 2 femtosegundos. Por último, se generaron simulaciones de 100 ns para caracterizar las principales interacciones de los complejos.

Los análisis estructurales se llevaron a cabo mediante el uso del software VMD (Humphrey et al., 1996), con complementos para el cálculo aproximado de RMSD, que mide la variación de los grupos de átomos a lo largo de la trayectoria. También se analizó la formación de

enlaces de hidrógeno y otras interacciones no enlazantes entre las integrinas y los péptidos, además de medir los residuos interactuando con el motivo de interés.

4.6 Cálculos de energía libre de unión de los complejos

La energía libre es la cantidad clave en la descripción de la termodinámica de los sistemas biológicos y, por tanto, un objetivo importante de las simulaciones biomoleculares. Los cálculos computacionales de energía libre sirven para proporcionar una estimación precisa de la afinidad de unión de una proteína y su ligando (Jespers et al., 2021). Para estimar la energía libre de enlace relativa en los complejos integrina-péptido se utilizó el método MM-GBSA (por sus siglas en inglés: MM, mecánica molecular; GB, Born generalizado; SA, área de superficie).

En preparación para los cálculos de MM-GBSA, se extrajeron *frames* representativos de las trayectorias de equilibrio de los sistemas. Se eliminaron las moléculas de agua y los iones explícitos. El análisis de MM-GBSA se realizó en tres subconjuntos de cada sistema: la integrina sola, el péptido solo y el complejo integrina-péptido. Para cada uno de estos subconjuntos, la energía libre se calculó de la siguiente manera:

Ecuación 1.

$$G_{\text{tot}} = HMM + G_{\text{solv-pol}} + G_{\text{solv-np}} - T\Delta S_{\text{conf}}$$

donde HMM, $G_{\text{solv-pol}}$, $G_{\text{solv-np}}$ y $T\Delta S_{\text{conf}}$ corresponden a la suma de los términos de energía enlazada y de Lennard-Jones, la contribución polar de la energía de solvatación, la contribución no polar a la energía de solvatación y la entropía conformacional, respectivamente. Tanto HMM como $G_{\text{solv-pol}}$ se calcularon utilizando NAMD 2.8 (Phillips et al., 2005) con el modelo de solvente implícito generalizado de Born (Tanner et al., 2011) y los parámetros descritos anteriormente. La constante dieléctrica del disolvente se estableció en 78,5. $G_{\text{solv-np}}$ se calculó como una función lineal del área superficial accesible al solvente (SASA), que se estimó con un radio de sonda de 1.4 Å (Abroshan et al., 2010). La energía libre de unión de los tres heterodímeros y los péptidos estudiados fue calculada por la diferencia:

Ecuación 2.

$$\Delta G_{\text{bind}} = G_{\text{tot}}(\text{integrina-péptido}) - G_{\text{tot}}(\text{integrina}) - G_{\text{tot}}(\text{péptido})$$

donde los valores de G_{tot} fueron los promedios de la simulación. Cabe destacar que las porciones iniciales de las simulaciones se excluyeron del cálculo, debido a su inestabilidad

energética. El cambio en la entropía conformacional debería, en principio, incluirse; sin embargo, no se calculó debido al gran costo computacional.

5.- RESULTADOS

5.1.- Identificación de patrones estructurales

Para evaluar la existencia de patrones conservados que modulen la unión del motivo RGD a las integrinas de interés, se analizaron los posibles residuos y estructuras conservadas y compartidas entre las distintas subunidades α y β , respectivamente. Para ello, se alinearon las estructuras y se midieron los residuos en una esfera de 6 Å. En base a este análisis, se esperaba identificar los residuos conservados que establecieran interacciones químicas con los motivos NGR y RGD; en el último caso, semejantes a las reportadas en la Figura 6.



Figura 9.- Alineamiento de secuencias entre subunidades α de integrinas de interés. En negro se destacan los residuos conservados en un radio de 6 Å del motivo RGD del péptido co-cristalizado en cada estructura, encontrando los patrones Phe-Tyr-Trp-Gln y Asp-Asp.

La Figura 9 muestra destacados en rectángulos negros los dos patrones identificados en el alineamiento de las dos subunidades α V y α 5, consideradas en el estudio. El primer patrón considera la secuencia FYWQ, la cual interactúa principalmente con el residuo Arg del motivo RGD, como se puede ver en la Figura 10. El segundo patrón se compone de dos aspartatos e interactúa con la misma porción R del motivo, como se puede ver en la Figura 11. Cabe recalcar que este alineamiento se realiza únicamente entre 2 subunidades, α V y α 5, ya que α V se comparte entre las integrinas de cristal 1L5G (α V β 3) y 4UM9 (α V β 6). A nivel global, ambas subunidades presentan un 53,98% de identidad en sus secuencias.

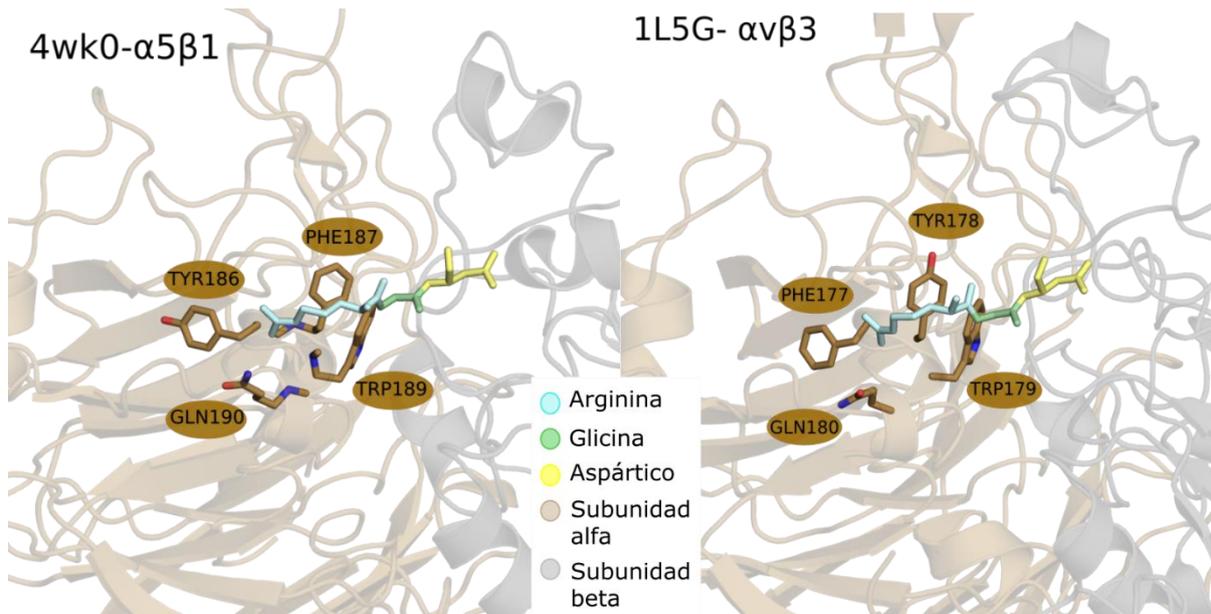


Figura 10.- Vista estructural del patrón conservado a nivel de secuencia entre las subunidades αv y $\alpha 5$. TYR186, PHE187, TRP189 y GLN190 interactúan con el residuo ARG del péptido RGD.

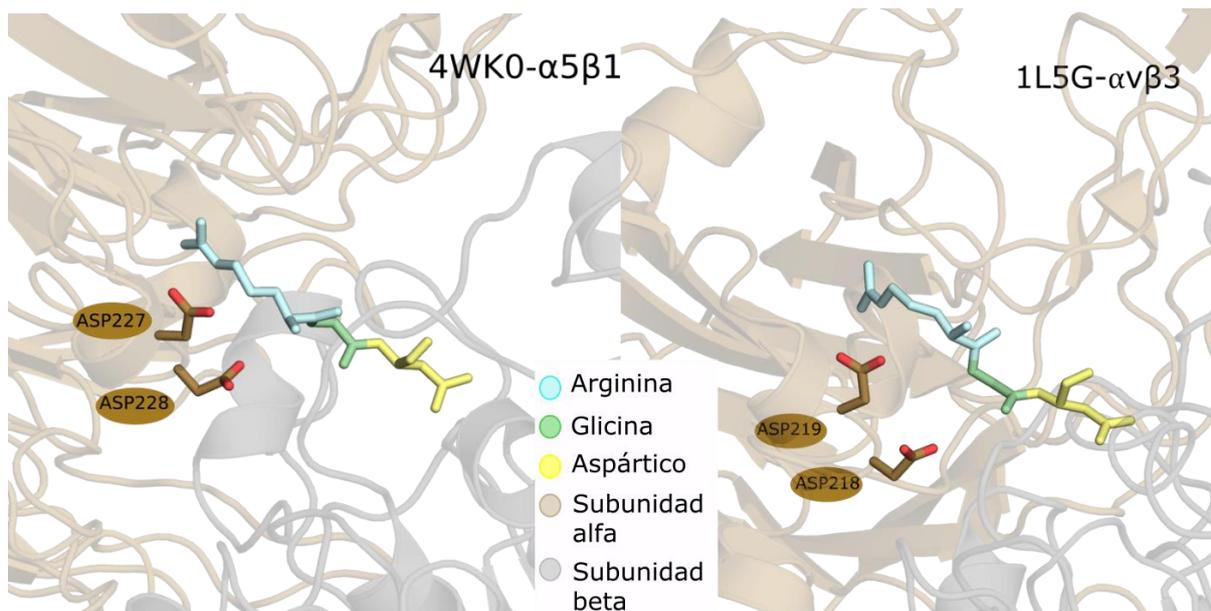


Figura 11.- Vista estructural del patrón ASP227 y ASP228 conservado a nivel de secuencia entre las subunidades αv y $\alpha 5$. Ambos residuos interactúan con la ARG del péptido RGD.

La Figura 12 muestra el alineamiento realizado entre las subunidades $\beta 1$ y $\beta 3$ y $\beta 6$, donde se identificaron 2 patrones conservados entre las estructuras. Las subunidades $\beta 1$ y $\beta 3$ presentan 45.43% de identidad en sus secuencias, entre $\beta 3$ y $\beta 6$ existe 50.72% de identidad y

finalmente entre $\beta 1$ y $\beta 6$ se observa un 45.81% de identidad en sus secuencias. En este caso, se identificó un patrón de doble serina y un segundo un patrón en la secuencia SxNxDxPE.

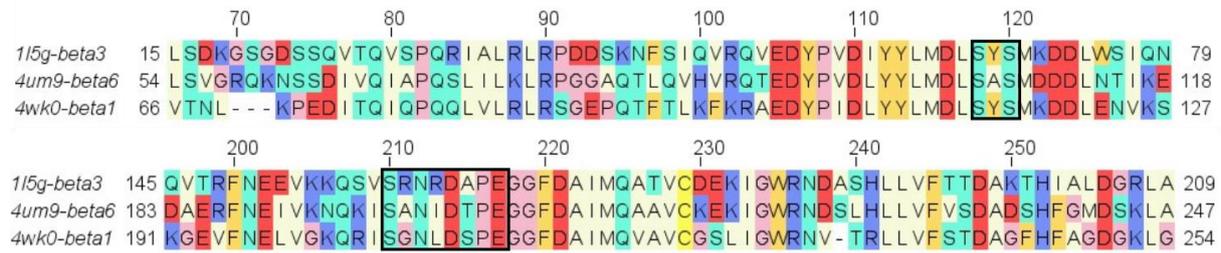


Figura 12.- Alineamiento de secuencias entre subunidades β de integrinas de interés. En rectángulos negro se destacan las zonas que contienen los residuos conservados en un radio de 6 Å del motivo RGD del péptido co-cristalizado en cada estructura. Se encontraron los patrones SxS y SxNxDxPE.

Para evaluar la orientación espacial de los patrones se realizó una visualización de las estructuras y los residuos identificados se muestran en la Figura 13, en donde se aprecia que el patrón identificado en la subunidad β se orienta hacia el residuo D del motivo RGD.

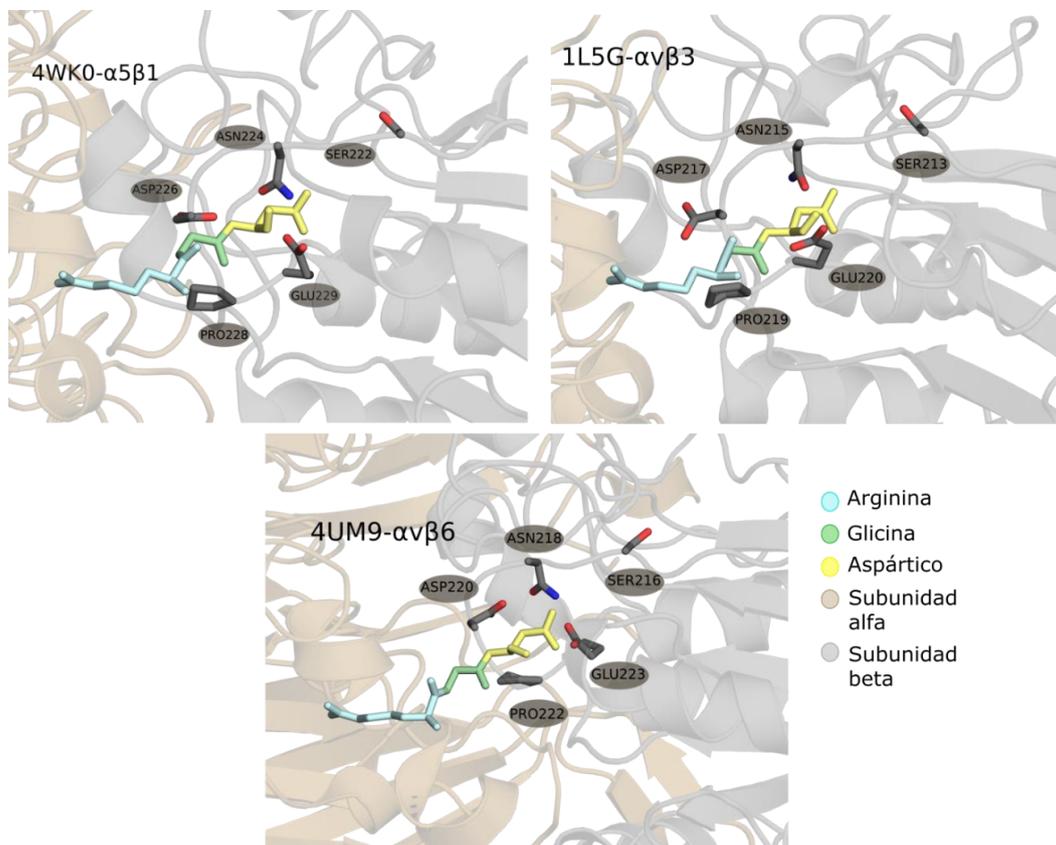


Figura 13.- Vista estructural del patrón conservado a nivel de secuencia, con ASP-ASN-SER-

PRO-GLU interactuando con GLU del péptido RGD en las 3 subunidades β estudiadas.

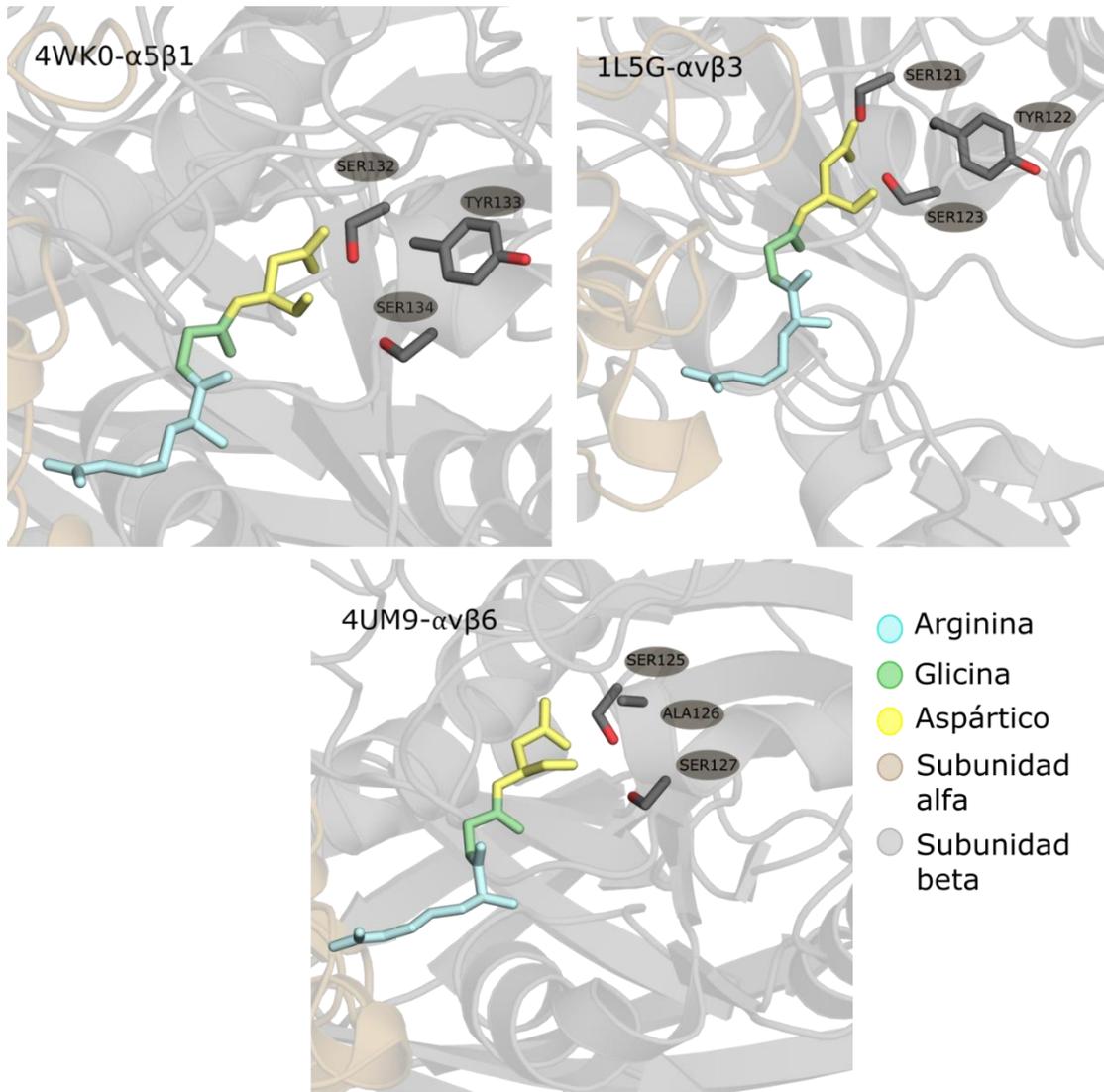


Figura 14.- Vista estructural del patrón conservado a nivel de secuencia con SxS interactuando con el residuo Asp del péptido RGD en las 3 subunidades β estudiadas.

Estos resultados nos permiten concluir que los patrones identificados en las unidades α se encuentran orientados a establecer un contacto o interacción con el residuo de arginina del motivo RGD, mientras que la subunidad β interactúa mayormente con el residuo Asp del motivo.

5.2.- Identificación de patrones Geométricos

Tomando el resultado anterior como base y asumiendo que los patrones identificados mediante la interacción del motivo RGD con las integrinas se conservan entre las diversas subunidades, es que se buscó identificar si los patrones existen solamente a nivel estructural o si la orientación, distancia y geometría de los residuos en el espacio también forma un patrón en la superficie 3D de la proteína.

Para hacer este análisis se utilizó el programa GeomFinder, realizando 3 búsquedas: (1) primero una búsqueda entre las subunidades α , luego (2) entre las subunidades β y finalmente (3) la búsqueda con la proteína completa para identificar posibles patrones en la superficie 3D de ambas subunidades interactuando. Los resultados iniciales fueron filtrados según los aminoácidos de interés identificados en el punto anterior obteniéndose los valores que se muestran en las Tablas 2 y 3, las cuales muestran puntuaciones de 100% para los parámetros de perímetro secuencia, energía no ligada y distancia.

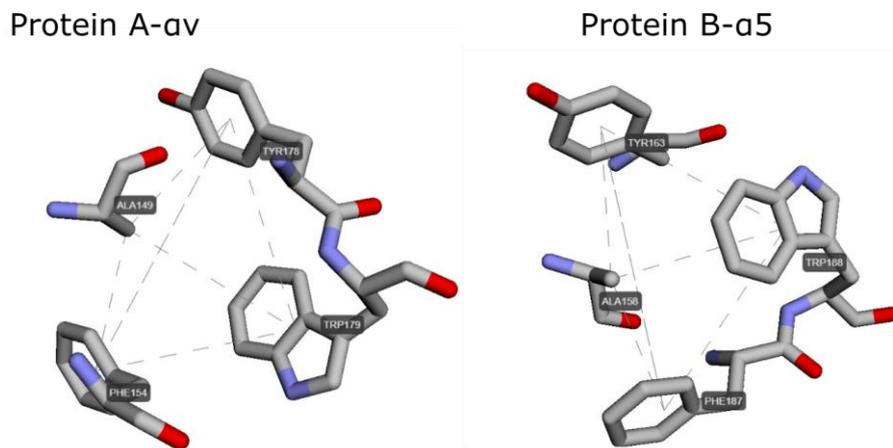


Figura 15.- Patrón Geométrico identificado entre subunidades α del conjunto en estudio.

Estos resultados nos permiten concluir que los que aminoácidos identificados mediante alineamientos se encuentran conformando patrones en la superficie geométrica de la proteína por lo que la unión del motivo RGD no sería únicamente afectada por la existencia de los aminoácidos como tal, sino que también la orientación espacial de los mismos. La Tabla 2 muestra los aminoácidos que componen un patrón geométrico 100% idéntico en los cuatro aspectos evaluados por GeomFinder (secuencia-Sc, distancia-Dist, energía no enlazante-Nbe y perímetro-Tsp) entre ambas subunidades α . Es importante destacar que este patrón está

compuesto por residuos (Tyr y Phe) identificados en el alineamiento mostrado anteriormente, y además anexa 2 residuos Ala y Trp.

Tabla 2.- Puntuaciones asociadas a patrón geométrico entre subunidades α .

Site A (115g- α V)	Site B (4wk0- α 5)	GScore	Snbe	Sdist	STsp	SSc
Ala149 Phe154 Tyr178 Trp179	Ala158 Tyr163 Phe187 Trp188	100	100	100	100	100

En la tabla 2 se presenta un Gscore similar para cada par de proteínas evaluadas. Este puntaje es definido como la combinación de similitudes parciales asociadas (S) de los cuatro descriptores previamente mencionados (Sc, Dist, Nbe y Tsp), considerando una ponderación igual de 25% para cada uno de ellos.

Sumados a los resultados anteriores, a través del análisis de a pares entre las subunidades β 3 y β 6 y luego las subunidades β 6 y β 1, fue posible identificar un patrón geométrico, mediante la misma lógica establecida para la subunidad α . Este patrón correspondiente a doble serina fue identificado en el alineamiento mostrado anteriormente, y además se le anexaron los residuos de metionina y aspartato para la conformación de un patrón geométrico 100% idéntico entre las 3 unidades. Dicho patrón también estaría mediando la interacción del motivo RGD y las integrinas.

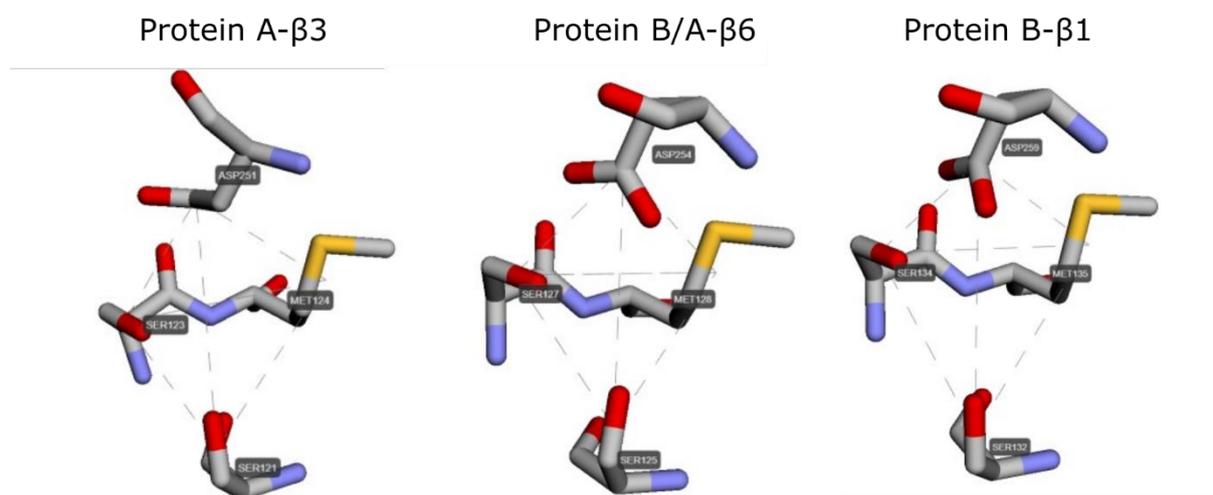


Figura 16.- Patrón Geométrico identificado entre subunidades β del conjunto en estudio.

Tabla 3.- Puntuaciones asociadas a patrón geométrico entre subunidades β .

Site A (115g- β 3)	Site B (4um9- β 6)	Gscore	Snbe	Sdist	STsp	SSc
Ser121 Ser123 Met124 Asp251	Ser125 Ser127 Met128 Asp254	100	100	100	100	100
Site A (4um9- β 6)	Site B (4wk0- β 1)	Gscore	Snbe	Sdist	STsp	SSc
Ser125 Ser127 Met128 Asp254	Ser132 Ser134 Met135 Asp259	100	100	100	100	100

En base a los análisis anteriores es posible concluir que por la homología que existe entre las diferentes subunidades, dentro de cada familia α y β , existen residuos conservados tanto a nivel de secuencia como estructura. Tales residuos estarían directamente involucrados en la afinidad que presenta el motivo RGD por una pequeña hendidura que existe entre la unión de las unidades α y β en las integrinas. De estos patrones conservados se observó que:

- La orientación del motivo RGD fue, en los 3 casos, con el residuo ácido interactuando con la subunidad β , mientras que el residuo básico interactúa con la subunidad α .
- Arg, del patrón RGD, modula su unión a la integrina mediante aminoácidos aromáticos y cargados, esperándose encontrar interacciones con residuos de características semejantes al inducir la unión del conjunto de péptidos e integrinas.
- Asp, del patrón RGD, modula su unión a la integrina mediante interacciones electroestáticas con regiones polares de residuos circundantes. Además de establecer un patrón geométrico con metionina, residuo que contiene un azufre en su cadena lateral.

5.3.- Modelado de péptidos

El modelado de péptidos con estructura desconocida entregó como resultado una serie de conformaciones con distintos scores según estimaciones energéticas. La estructura óptima de cada péptido fue identificada como aquella con menor valor de score Sopep. La Tabla 4 revela la secuencia de los péptidos que se modelaron, la cantidad de clústeres que se generaron en el proceso de modelado, el tamaño de los clústeres y el score del mejor modelo en cada caso.

Tabla 4.- Puntuaciones asociadas a modelos generados para péptidos de interés.

Péptido	Número clústeres generados	Tamaño de clúster	Score Sopep
CNGRC	2	20	-1.2
CVLNGRMEC	21	5	-6.6
NGRAHA	6	17	-2.5
CRGDRGPDC	35	12	-5.0
GACRGDCLGA	40	3	-6.4

La visualización de los péptidos modelados no permite identificar claramente la formación de estructuras secundarias tipo láminas o hélices. Cabe destacar que la distancia entre los carbonos α de ambas cisteínas que componen el péptido GACRGDCLGA fue de 5.3 Å, lo que se encuentra en el rango para la posible formación de un enlace disulfuro.

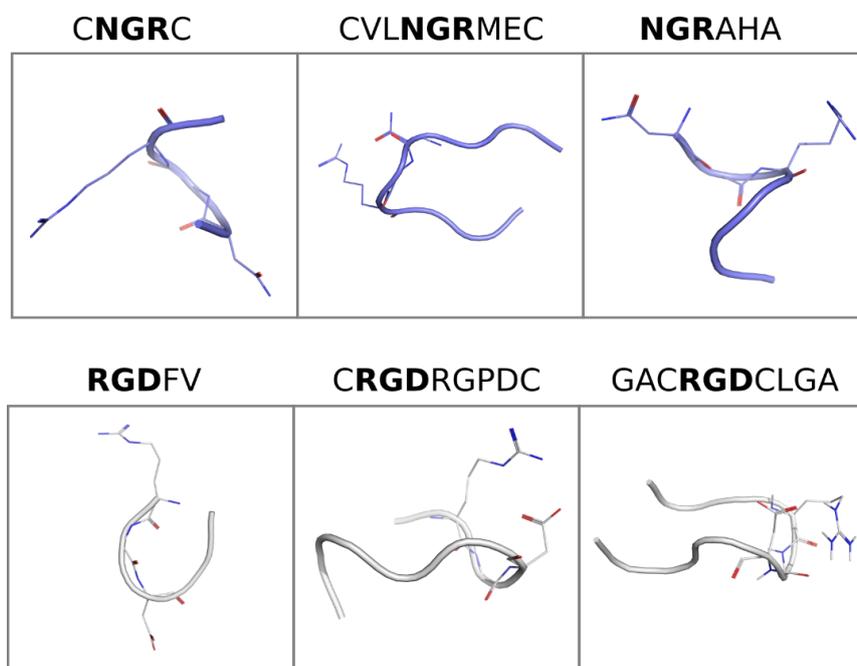


Figura 17.- Visualización 3D de modelos iniciales de péptidos de interés. En color azul se representan los péptidos que contienen el motivo NGR en su secuencia. En color gris se representan los péptidos que contienen el motivo RGD en su secuencia. Ambos motivos han sido representados con líneas y coloreado en azul los átomos de nitrógeno y rojo los átomos de oxígeno.

5.4.- Acoplamiento Molecular

Los estudios de acoplamiento molecular fueron realizados para 18 complejos, compuestos por las 3 integrinas en estudio (1L5G, 4UM9 y 4WK0), cada una interactuando con 6 péptidos. Como sitio de interacción de los péptidos en las integrinas se definió una hendidura entre las subunidades α y β , centrado el área de interacción en el ligando original presente en cada proteína. La figura 18 muestra una comparación de los resultados finales de acoplamiento para: A) las 3 integrinas con cada uno de los 6 péptidos, y B) un alineamiento de todos los complejos destacando el lugar de acoplamiento de los péptidos. En esta última figura es posible apreciar que la región de unión de los péptidos se mantuvo en un sitio intermonomérico entre las subunidades α y β .

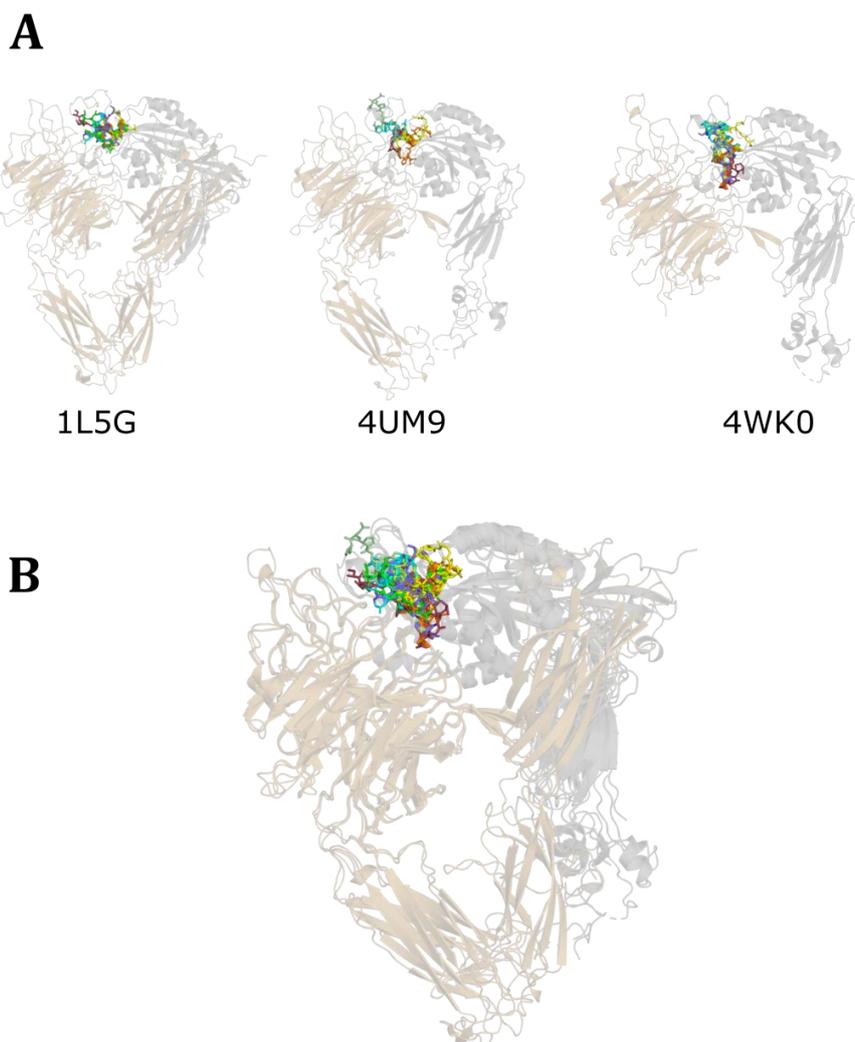


Figura 18.- A) Resultados de poses de docking de cada integrina en estudio asociada a los 6 péptidos de interés. B) Alineamiento de todos los complejos estudiados y sus poses resultantes del acoplamiento molecular.

Con el fin de estimar las diferencias estructurales entre todas las conformaciones obtenidas por *docking*, se realizó un alineamiento con Pymol de las 10 mejores poses para cada complejo. La tabla 5 presenta los valores de RMSD asociados a los alineamientos estructurales de los complejos. La mayor diferencia estructural o discrepancia entre las poses se observó para el péptido GACRGDCLGA para las 3 integrinas. A su vez, el resultado de *docking* con poses más cercanas a un consenso se dio en el caso del péptido c(RGDFV).

Tabla 5.- Valores de RMSD para 10 poses con ranking superior por complejo.

Péptido	RMSD $\alpha v\beta 3$ (1L5G)	RMSD $\alpha v\beta 6$ (4UM9)	RMSD $\alpha 5\beta 1$ (4WK0)
<u>c(RGDFV)</u>	<u>0.71</u>	<u>0.41</u>	<u>0.34</u>
NGRAHA	1.39	1.87	2.85
<u>GACRGDCLGA</u>	<u>2.06</u>	<u>2.83</u>	<u>3.36</u>
CVLNGRMEC	3.06	*	0.61
CRGDRGPDC	3.8	2.86	2.01
CNGRC	1.51	1.71	1.46

La figura 19 muestra el alineamiento en los casos de mayor y menor valor de RMSD para los péptidos identificados, destacando los resultados para el sistema conjugado con la integrina $\alpha 5\beta 1$, por ser la que presentaba los valores más extremos en el análisis. En la figura 19A se logra identificar una clara orientación del motivo RGD del péptido c(RGDFV), observándose una mayor discrepancia en la orientación del aminoácido aromático que compone su secuencia. En B) podemos apreciar el resultado del péptido con mayores diferencias entre sus 10 mejores poses, donde ambos extremos que flanquean el motivo RGD tienen orientaciones no concluyentes.

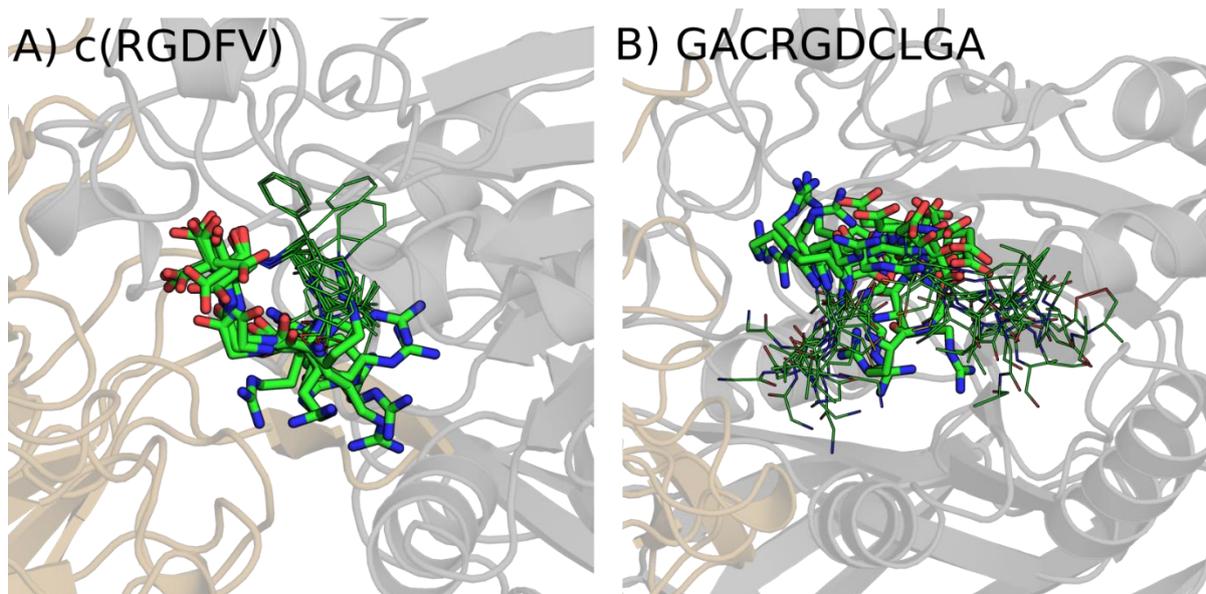


Figura 19.- A) Alineamiento de las 10 mejores poses para el péptido c(RGDFV) en 4WK0. B) Alineamiento de las 10 mejores poses para el péptido GACRGDCLGA en 4WK0. En ambos casos el motivo RGD se encuentra representado en licorice, mientras que el resto del péptido se representa en líneas. La estructura coloreada café representa la subunidad α de la integrina, mientras que coloreada en gris se encuentra la subunidad β .

Tabla 6.- Score de docking o afinidad (kcal/mol) estimada entre Péptidos e Integrinas.

Péptido	$\alpha v\beta 3$ (1L5G)	$\alpha v\beta 6$ (4UM9)	$\alpha 5\beta 1$ (4WK0)
c(RGDFV)	-15.7	-8.2	-10.4
NGRAHA	-16.3	-16.9	-15.6
<u>GACRGDCLGA</u>	<u>-26.3</u>	<u>-24.3</u>	<u>-25.2</u>
CVLNGRMEC	-18.7	*	-20.2
CRGDRGPDC	-19.8	-19.4	-18.8
CNGRC	-16.4	-14.7	-14.3

La tabla 6 indica el score de *docking* para las mejores conformaciones obtenidas en cada complejo estudiado. Se aprecia que el péptido GACRGDCLGA tiene la mayor afinidad con las integrinas, destacando la asociación con $\alpha v\beta 3$ (1L5G). Cabe destacar que en el caso del péptido CVLNGRMEC, en el sistema con 4UM9, no se reporta un valor de afinidad pues el criterio de clusterización no fue alcanzado y no se logró agrupar ni generar un ranking de las poses.

Tabla 6.- Resumen Interacciones RGD/NGR-Integrina en radio de 5Å

Péptido	$\alpha v\beta 3$ (1L5G)	$\alpha v\beta 6$ (4UM9)	$\alpha 5\beta 1$ (4WK0)
RGDFV	Asp150, <u>Phe177</u> , <u>Trp178</u> , <u>Gln179</u> , Thr212, Ala213, Gln214, Ala215, <u>Asp218</u> , Tyr122, Tyr166, Met180, Arg214, <u>Asn215</u> , Arg216, <u>Asp217</u> , Ala218	<u>Ser125</u> , Ala126, <u>Ser127</u> , Met128, Val169, Lys170, Glu175, Pro179, Cys180, Ser182, Ile183, <u>Ser216</u> , Ala217, <u>Asn218</u> , Ile219, <u>Asp220</u> , Thr221, <u>Glu223</u>	Trp157, <u>Phe187</u> , <u>Ser132</u> , Tyr133, <u>Ser134</u> , Lys182m Pro186, Cys187, Thr188, Gly223, <u>Asn224</u> , Leu225, <u>Asp226</u> , Ser227, <u>Glu229</u> , Asp259, Ala260, Phe321
CRGDRGPDC	Ala215, Ile216, <u>Asp218</u> , <u>Asp219</u> , Arg248, Asp119, <u>Ser121</u> , <u>Ser123</u> , Met124, Phe156, <u>Asp217</u> , Ala218, <u>Pro219</u> , <u>Glu220</u> , Thr250, Asp251, Ala252, Lys253, Asn313	Tyr178, Asp218, <u>Ser125</u> , Ala126, Phe168, Val169, Lys170, Glu175, Pro179, Cys180, Ser182, Ile183, Ala217, <u>Asn218</u> , Ile219, <u>Asp220</u> , Thr221, <u>Glu223</u>	<u>Phe187</u> , Ser224, <u>Asp227</u> , <u>Asp228</u> , Gly242, Leu225, Ser227, <u>Pro228</u> , Glu320, Phe321
GA- CRGDCLGA	<u>Tyr178</u> , <u>Trp179</u> , <u>Asp218</u> , Arg248, Tyr166, <u>Asn215</u> Arg216, <u>Asp217</u> , Ala218, Asp251, Ala252, Lys253, Asn313, Val314	Ala215, Ile216, <u>Asp218</u> , <u>Asp219</u> , Thr221, <u>Pro222</u> , Glu316, Lys338	<u>Ser132</u> , Tyr133, <u>Ser134</u> , Lys182, Pro186, Cys187, Thr188, Gln191, <u>Ser222</u> , Gly223, <u>Asn224</u> , Leu225, <u>Asp226</u> , Ser227, <u>Glu229</u> , Asp259, Ala260, Thr318, Glu320, Phe321
NGRAHA	<u>Ser121</u> , Tyr122, <u>Ser123</u> , Asp125, Arg214, <u>Asn215</u> , Arg216, <u>Asp217</u> , Ala218, <u>Pro219</u> , <u>Glu220</u> , Asp251, Ala252, Lys253, Asn313	Asp123, <u>Ser125</u> , <u>Ser127</u> , Met128, Asp130, Asp131, <u>Asp220</u> , Thr221, <u>Pro222</u> , <u>Glu223</u> , Ser253, Asp254, Ala255, Thr314, Glu316, Gln317, Leu336, Gln337, Lys338, Asp339, Ser340	<u>Phe187</u> , Ser224, <u>Asp227</u> , <u>Asp228</u> , <u>Ser132</u> , Tyr133, <u>Ser134</u> , Met135, Val168, <u>Ser222</u> , Gly223, <u>Asn224</u> , Leu225, <u>Asp226</u> , Ser227, <u>Glu229</u> , Asp259, Ala260, Gly261, Glu320, Phe321

CNGRC	Asp148, <u>Ala149</u> , Asp150, Gly151, Phe154, <u>Tyr178</u> , <u>Trp179</u> , Ala215, <u>Asp218</u> , <u>Tyr122</u> , Ser123, Tyr166, Arg214, <u>Asn215</u> , Arg216, <u>Asp217</u> , <u>Ala218</u> , <u>Glu220</u>	<u>Tyr178</u> , <u>Trp179</u> , Ala215, <u>Asp218</u> , <u>Asp219</u> , <u>Ala126</u> , <u>Ser127</u> , <u>Ile219</u> , <u>Asp220</u> , Thr221, <u>Pro222</u> , Asp254, Ala255, Asp256, Gln317	Ile225, Tyr226, <u>Asp227</u> , <u>Asp228</u> , Ser229, Lys254, Gly255, Asn256, Leu257, Thr258, Ser227, Glu320, Phe321, Pro323
CVLNGR-MEC	Asp150, <u>Phe177</u> , <u>Tyr178</u> , <u>Trp179</u> , <u>Gln180</u> , Ala215, Ile216, Phe217, <u>Asp218</u> , <u>Asp219</u> , Arg248, Arg216, <u>Asp217</u> , Ala218, Ala252, Lys253	-	<u>Phe187</u> , <u>Trp188</u> , Ser224, <u>Asp227</u> , <u>Asp228</u> , <u>Leu225</u> , <u>Asp226</u> , Ser227, <u>Pro228</u> , Asp259, Glu320, Phe321

En la tabla 7 se describen todas las interacciones establecidas en un radio de 5 Å del motivo RGD y NGR de cada péptido en su interacción con las integrinas. En color negro se presentan los aminoácidos pertenecientes a la subunidad α de cada integrina, mientras que en azul se muestran aquellos que componen la subunidad β ; al mismo tiempo se encuentran subrayados los aminoácidos identificados como claves en la interacción con el motivo RGD en el primer punto de la metodología.

De los análisis sobre las poses adoptadas en los péptidos que contienen el motivo NGR se observó que:

- NGRAHA muestra una mayor tendencia a interactuar con la subunidad β .
- Los péptidos con NGR dentro de su secuencia y no en el extremo, CNGRC y CVLNGRMEC, se orientan entre ambas subunidades, y comparten residuos cercanos al motivo de interés con RGD, destacándose el motivo de doble aspartato de la subunidad α .
- El residuo Arg se orienta hacia la subunidad α y el residuo Asn hacia la subunidad β . Esto coincide con lo observado para el motivo RGD, donde la Arg se orienta hacia la misma subunidad.

Respecto a las poses adoptadas por los péptidos que contienen el motivo RGD se observó que:

- Combinaciones específicas (RGDFV- α V β 6, CRGDRGPDC- α V β 6, GACRGDCLGA- α 5 β 1) de todos los péptidos con motivo RGD tendieron a establecer mayores interacciones con la subunidad β .

- Gran parte de los residuos que se encuentran dentro del radio de 5 Å son repetidos entre los distintos complejos y se encuentran formando parte de los aminoácidos identificados como conservados en las estructuras cristalizadas bases de esta investigación.

De manera general, y compartido entre ambos motivos, se observa que:

- De las tres integrinas, $\alpha V\beta 6$ (4UM9) tiene la mayor tendencia a ubicar los péptidos orientados hacia la subunidad β .
- Dentro del conjunto de complejos estudiados no se presentan diferencias mayores en los valores de afinidad asociados al score de *docking* que permitan seleccionar como más afines a péptidos conteniendo el motivo NGR con respecto a péptidos de motivo RGD.

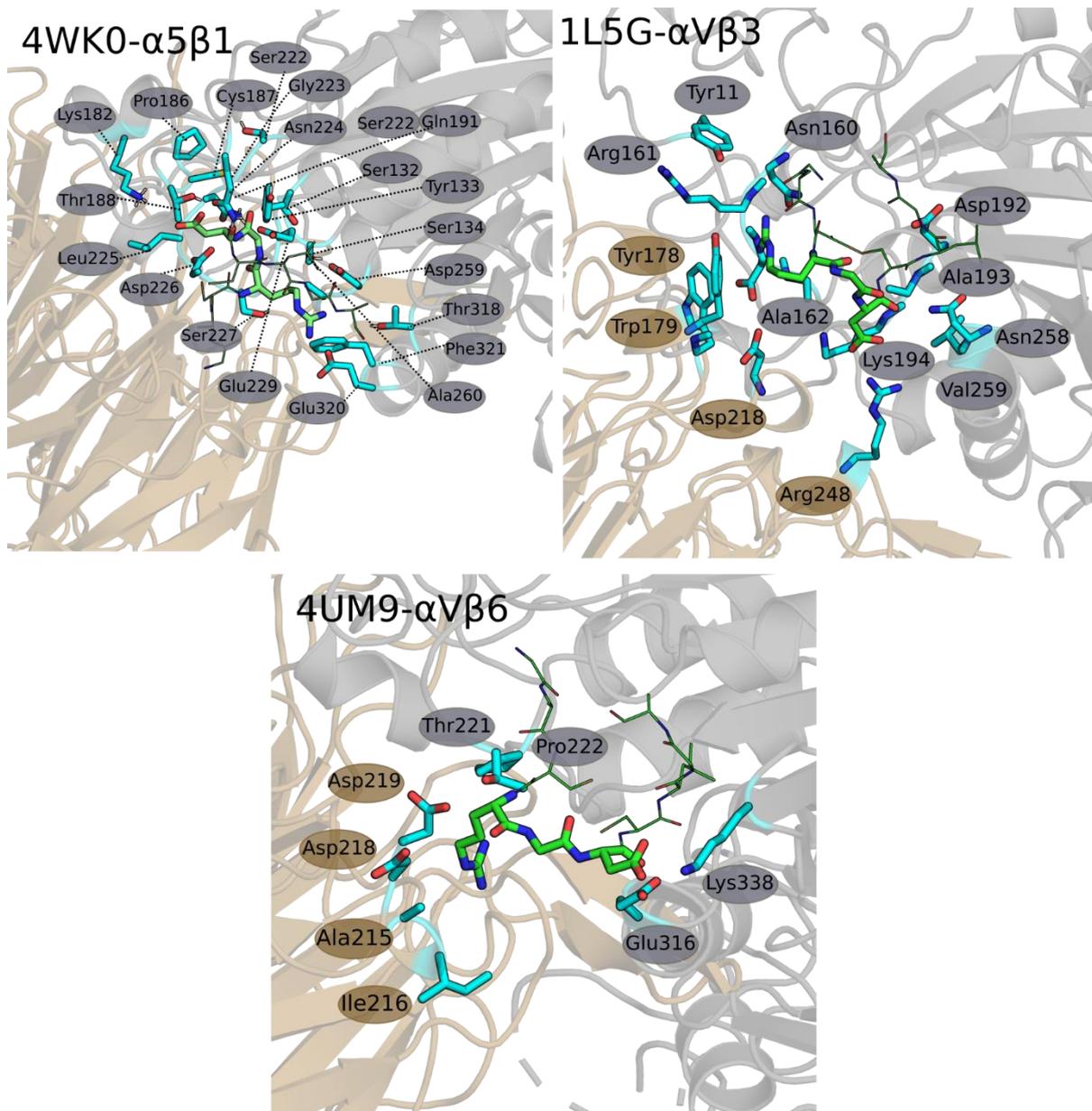


Figura 20.- Residuos en un radio de 5Å interactuando con el péptido GACRGDCLGA para los sistemas conjugados con las integrinas $\alpha5\beta1$, $\alphaV\beta3$, $\alphaV\beta6$.

A partir de estos estudios, se seleccionaron 3 sistemas de interés incorporando como criterio de selección: las afinidades energéticas, estabilidad del péptido y factores estructurales de relevancia. Así, se consideró como objeto de estudio los complejos incluyendo un mismo péptido asociado a 3 diferentes integrinas. Esto con el fin de realizar cálculos energéticos de afinidad mediante el método MM-GBSA. Es importante mencionar que los métodos evaluando únicamente acoplamiento molecular entre complejos muchas veces no son concluyentes y deben ser complementadas con simulaciones moleculares de los sistemas. De

lo anterior se deriva la decisión de evaluar a través de simulaciones de dinámica molecular y MM-GBSA al péptido GACRGDCLGA conjugado con las 3 integrinas de interés.

5.5.- Dinámica Molecular

Las trayectorias obtenidas de las simulaciones de dinámica molecular indican las posiciones cartesianas de cada átomo del sistema en cada paso de tiempo. Para evaluar los estados de equilibrio de cada complejo se realizaron análisis de RMSD, o desviación de raíz cuadrática media, una medida estándar de distancia estructural entre coordenadas. La Figura 21 recopila los resultados de este análisis para el péptido en los 3 sistemas en estudio. Durante los primeros 68 nanosegundos los complejos muestran importantes desviaciones, para luego alcanzar una estructura relativamente estable. La integrina $\alpha 5\beta 1$ (4WK0) asociada al péptido GACRGDCLGA presenta las menores fluctuaciones, lo que permite inferir que es el complejo más estable.

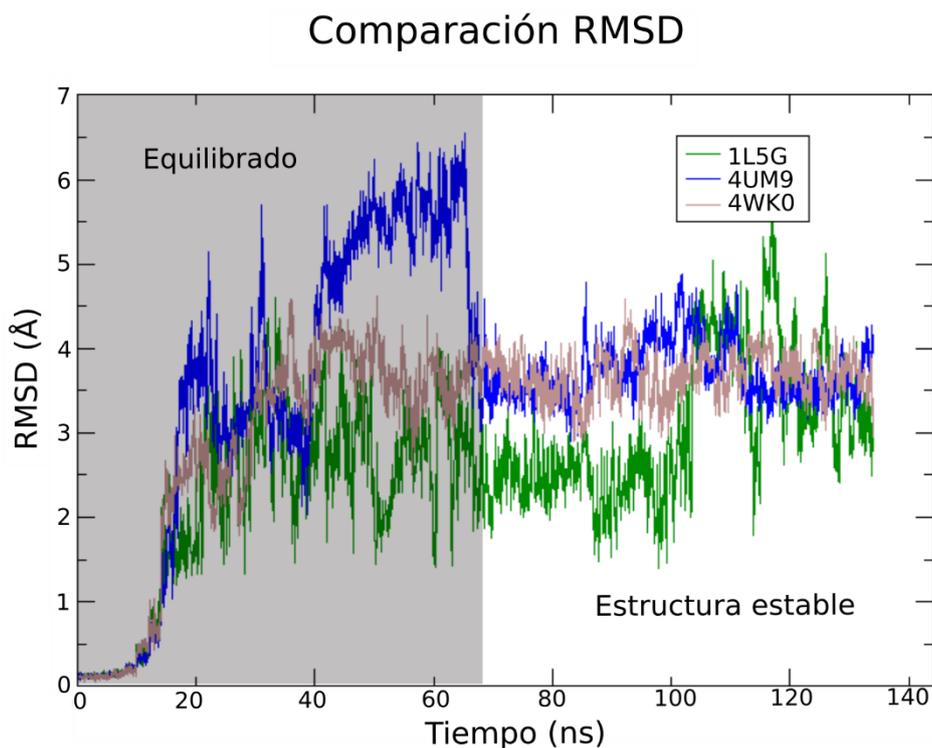


Figura 21.- Desviación cuadrática media (RMSD) de las proteínas interactuando con el péptido GACRGDCLGA en función del tiempo de equilibrado: Los primeros 40 nanosegundos representan al primer equilibrado que posee restricciones en el péptido. Los nanosegundos restantes muestran los valores de RMSD de una segunda dinámica molecular sin restricciones.

Sumado a los análisis de la estabilidad de los sistemas, se evaluaron las distancias a lo largo del tiempo de los principales residuos que establecen interacciones directas con el motivo RGD del péptido GACRGDCLGA. Como se muestra en la figura 22, las interacciones más prevalentes para la integrina $\alpha\beta3(1L5G)$ ocurren con los residuos Glu121 y Asp150 de la subunidad α . Al final de la simulación se observa contacto con D148.

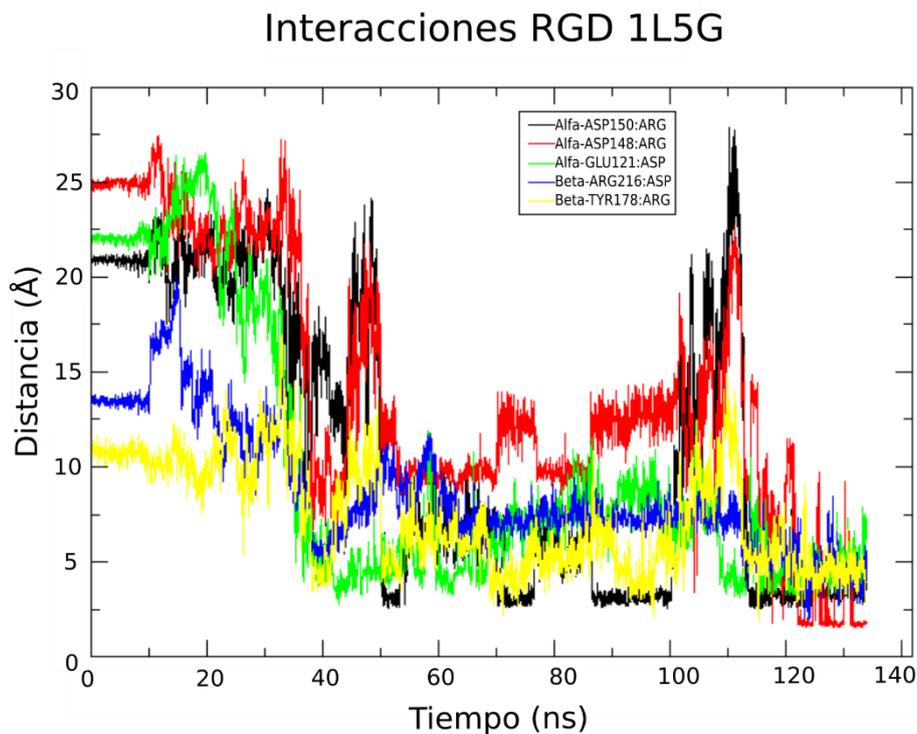


Figura 22.- Distancia entre el motivo RGD del péptido GACRGDCLGA y los residuos que conforman las subunidades α y β de la integrina $\alpha\beta3(1L5G)$.

Para el caso de la integrina $\alpha5\beta1$ interactuando con el péptido GACRGDCLG (figura 23) se observan contactos más definidos. En este complejo destaca la interacción del motivo RGD con los residuos Asp130, Thr188 y Asp 259 de la subunidad β . La interacción con el residuo Tyr133 muestra también una considerable estabilidad.

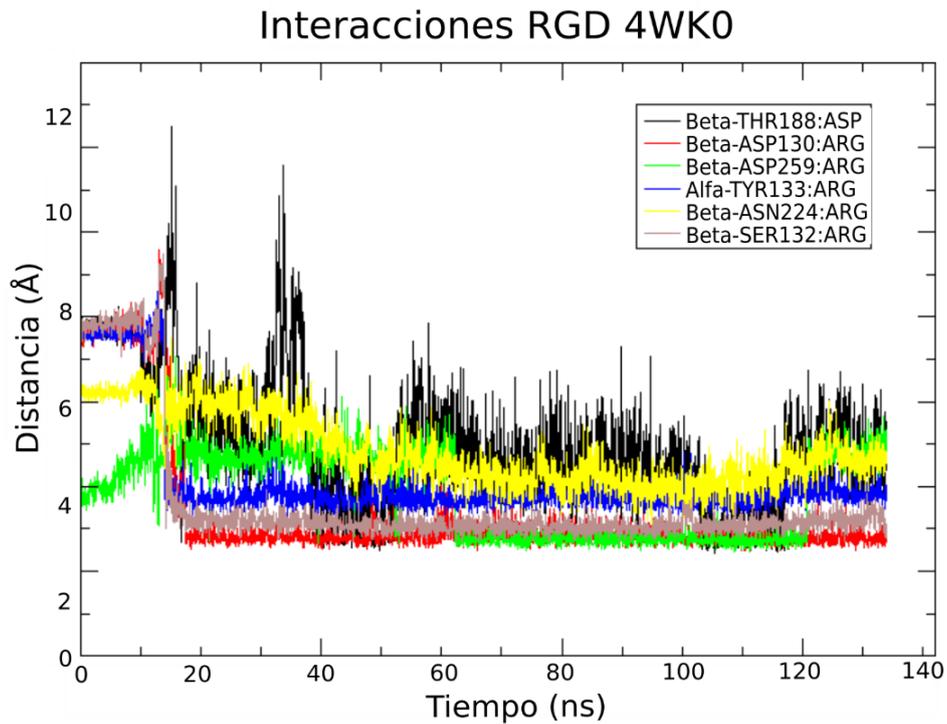


Figura 23.- Distancia entre el motivo RGD del péptido GACRGDCLGA y los residuos que conforman las subunidades α y β de la integrina $\alpha 5\beta 1$ (4WK0).

Por último, para el caso de la integrina $\alpha V\beta 6$ interactuando con el péptido GACRGDCLGA destaca la asociación del motivo RGD con los residuos Asp256 y Lys338 de la subunidad β y Asp219 de la subunidad α (figura 24).

Interacciones RGD 4UM9

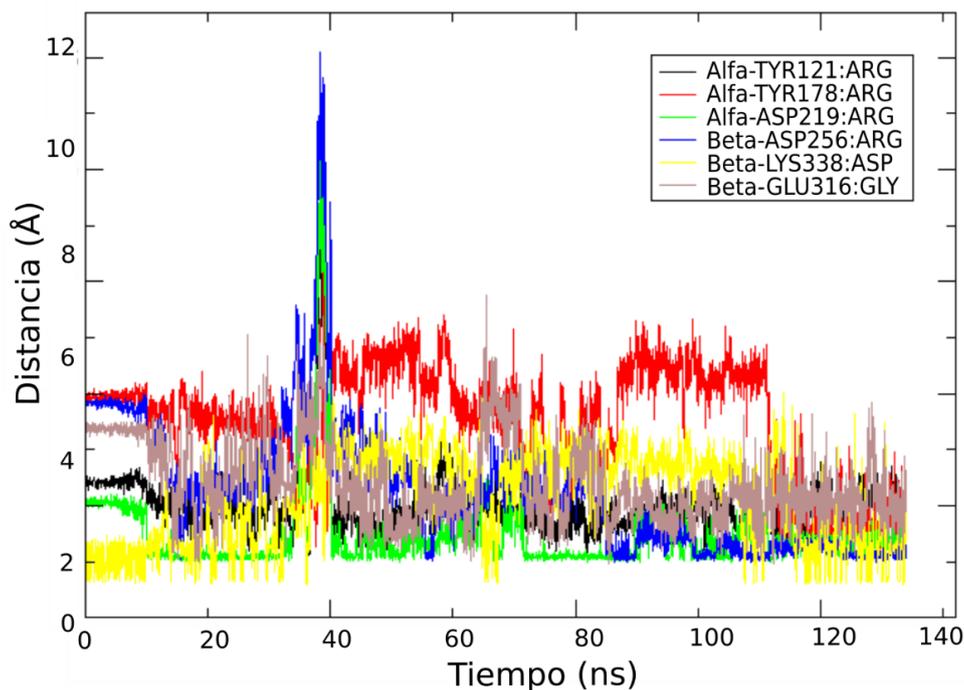


Figura 24.- Distancia entre el motivo RGD del péptido GACRGDCLGA y los residuos que conforman las subunidades α y β de la integrina α V β 6 (4UM9).

A partir de estos análisis es posible destacar la importancia de residuos electronegativos para coordinar la unión de la arginina del motivo RGD en los tres sistemas en estudio.

5.6.- Energía Libre de Unión

MM-GBSA permite la estimación de la energía libre de unión de un complejo considerando los estados libres y enlazados, a diferencia de los métodos más robustos que generalmente requieren la consideración de estados intermedios definidos de diversas maneras. Además, MM-GBSA tiene la ventaja de permitir una fácil separación de varias contribuciones a la energía libre, como las contribuciones de Van der Waals, electrostáticas y de solvatación.

La Tabla 8 presenta los valores obtenidos en la estimación de energía ΔG_{BSA} para cada uno de los 3 complejos considerados en este estudio. De acuerdo con los valores obtenidos en el desglose de energía, el aporte más significativo a la energía de unión proviene del componente electrostático de los sistemas. El complejo más estable resultó ser α 5 β 1-

GACRGDCLGA, lo que coincide con estudios experimentales que indican a este péptido como inhibidor de la integrina $\alpha 5\beta 1$ (Koivunen et al., 1993). Sin embargo, en base a las energías de unión obtenidas con MM-GBSA este péptido también podría tener un efecto modulador en las integrinas $\alpha V\beta 3$ y $\alpha V\beta 6$.

Tabla 7.- Valores de energía obtenidos a partir de los cálculos de MM-GBSA para los 3 complejos integrina-péptido.

Complejo	ΔH_{MM}^{vdW}	ΔH_{elecMM}	SASA	$\Delta G_{BSA} kcal/mol$
Integrina ($\alpha V\beta 6$)	-53.35 +/- 0.17	7.12 +/- 0.26	-9.10 +/- 0.01	-55.32 +/- 0.24
Integrina ($\alpha 5\beta 1$)	-40.71 +/- 0.15	-23.71 +/- 0.33	-6.11 +/- 0.01	-70.54 +/- 0.27
Integrina ($\alpha V\beta 3$)	-34.30 +/- 0.25	3.83 +/- 0.42	-6.37 +/- 0.03	-36.84 +/- 0.53

6.- CONCLUSIÓN

El descubrimiento de que la pequeña secuencia tripeptídica RGD es suficiente para inhibir la interacción de las proteínas de la MEC con la integrina (Pierschbacher & Ruoslahti, 1984) ha estimulado la búsqueda de pequeñas moléculas peptídicas o peptidomiméticas para abordar selectivamente varios subtipos de integrinas y, por lo tanto, sus características fisiopatológicas con alta afinidad. Dentro de la búsqueda de variantes de péptidos conteniendo el motivo RGD se ha identificado también la importancia de péptidos conformados por la secuencia NGR, donde se comparten 2 de sus 3 residuos cambiando únicamente el largo de la cadena del residuo ácido. La orientación de un motivo es “al revés” de la otra, entendiéndose este punto como que la lectura de NGR es lo mismo que RGD de atrás hacia adelante, sin considerar el ácido que los compone.

A partir del estudio de péptidos conteniendo los motivos RGD y NGR en su interacción con las integrinas $\alpha V\beta 6$, $\alpha 5\beta 1$ y $\alpha V\beta 6$ es posible alcanzar las siguientes conclusiones:

- En los análisis de secuencias se pudo identificar 4 patrones coordinando la interacción del motivo RGD y las 3 integrinas consideradas. Dos de los patrones se identificaron en la subunidad α (Phe-Tyr-Trp-Gln y Asp-Asp) y 2 en la subunidad β (SxS y SxNxPxPE).
- Estudios sobre la geometría de las proteínas demostraron que el patrón de secuencia Phe-Tyr-Trp-Gln (subunidades α), es parte de un patrón estructural localizado en la superficie proteica, compuesto por Ala-Phe-Tyr-Trp. Lo mismo ocurre en las subunidades β para el patrón de secuencia SxS, localizado en la superficie e incluyendo el patrón Ser-Ser-Met-Asp.
- Mediante estudios de *docking* se pudo establecer el modo de contacto entre los complejos, definiéndose como área de interés un bolsillo entre las subunidades α y β . La mayoría de los sistemas con péptidos de motivo RGD mostraron las cadenas laterales de la Arg y Asp del péptido interactuando con residuos que conforman la interfaz de unión de las subunidades α y β . El resto de los aminoácidos que componen los péptidos varían en su orientación entre residuos expuestos al solvente y residuos que se profundizan en la hendidura de las proteínas. En el caso de los sistemas con péptidos de motivo NGR las cadenas laterales de Arg y Asn reflejan una orientación similar a los residuos Arg y Asp en RGD.

- Las simulaciones de dinámica molecular permitieron la caracterización de interacciones específicas que coordinan los sistemas compuestos por las 3 integrinas de interés y el péptido GACRGDCLGA. La fracción Arg del péptido interactúa principalmente con 2 residuos de aspartato en las 3 proteínas. Por su parte, el carboxilato del residuo Asp varía en sus interacciones entre aminoácidos con carga positiva (arginina y lisina) y polar sin carga (treonina).
- El método MMGBSA permitió determinar una energía relativa de unión de los complejos, identificando un importante componente electrostático. La mayor estabilidad fue observada para el complejo $\alpha 5\beta 1$: GACRGDCLGA, aunque energías de unión favorables fueron estimadas para el péptido interactuando tanto con la integrina $\alpha V\beta 3$ como $\alpha V\beta 6$.

7.- REFERENCIAS

- Abroshan, H., Akbarzadeh, H., & Parsafar, G. A. (2010). Molecular dynamics simulation and MM–PBSA calculations of sickle cell hemoglobin in dimer form with Val, Trp, or Phe at the lateral contact. *Journal of Physical Organic Chemistry*, 23(9), 866-877. <https://doi.org/10.1002/poc.1679>
- American Cancer Society. (2020). *Colorectal Cancer Facts & Figures 2020-2022*. 48.
- Arap, W., Pasqualini, R., & Ruoslahti, E. (1998). Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. *Science (New York, N. Y.)*, 279(5349), 377-380. <https://doi.org/10.1126/science.279.5349.377>
- Barczyk, M., Carracedo, S., & Gullberg, D. (2009). Integrins. *Cell and Tissue Research*, 339(1), 269. <https://doi.org/10.1007/s00441-009-0834-6>
- Bates, R. C. (2005, noviembre 24). *The $\alpha v \beta 6$ integrin as a novel molecular target for colorectal cancer* (London, UK). [Http://Dx.Doi.Org/10.2217/14796694.1.6.821](http://Dx.Doi.Org/10.2217/14796694.1.6.821); Future Medicine Ltd London, UK. <https://www.futuremedicine.com/doi/abs/10.2217/14796694.1.6.821>
- Beaulieu, J.-F. (2019). Integrin $\alpha 6 \beta 4$ in Colorectal Cancer: Expression, Regulation, Functional Alterations and Use as a Biomarker. *Cancers*, 12(1). <https://doi.org/10.3390/cancers12010041>
- Bennett, J. S. (2005). Structure and function of the platelet integrin $\alpha IIb \beta 3$. *The Journal of Clinical Investigation*, 115(12), 3363-3369. <https://doi.org/10.1172/JCI26989>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242. <https://doi.org/10.1093/nar/28.1.235>
- Breuss, J. M., Gallo, J., DeLisser, H. M., Klimanskaya, I. V., Folkesson, H. G., Pittet, J. F., Nishimura, S. L., Aldape, K., Landers, D. V., & Carpenter, W. (1995). Expression of the beta 6 integrin subunit in development, neoplasia and tissue repair suggests a role in epithelial remodeling. *Journal of Cell Science*, 108 (Pt 6), 2241-2251.
- Breuss, J. M., Gillett, N., Lu, L., Sheppard, D., & Pytela, R. (1993). Restricted distribution of integrin beta 6 mRNA in primate epithelial tissues. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, 41(10), 1521-1527. <https://doi.org/10.1177/41.10.8245410>
- Campbell, I. D., & Humphries, M. J. (2011). Integrin Structure, Activation, and Interactions. *Cold Spring Harbor Perspectives in Biology*, 3(3). <https://doi.org/10.1101/cshperspect.a004994>
- Conlon, M. A., & Bird, A. R. (2014). The Impact of Diet and Lifestyle on Gut Microbiota and Human Health. *Nutrients*, 7(1), 17-44. <https://doi.org/10.3390/nu7010017>
- Corti, A., & Curnis, F. (2011). Tumor vasculature targeting through NGR peptide-based drug delivery systems. *Current Pharmaceutical Biotechnology*, 12(8), 1128-1134. <https://doi.org/10.2174/1389201111796117373>
- Corti, A., Curnis, F., Arap, W., & Pasqualini, R. (2008). The neovasculature homing motif NGR: More than meets the eye. *Blood*, 112(7), 2628-2635. <https://doi.org/10.1182/blood-2008-04-150862>
- Corti, A., Curnis, F., Rossoni, G., Marcucci, F., & Gregorc, V. (2013). Peptide-Mediated Targeting of Cytokines to Tumor Vasculature: The NGR-hTNF Example. *BioDrugs*, 27(6), 591-603. <https://doi.org/10.1007/s40259-013-0048-z>

- Craig, D., Gao, M., Schulten, K., & Vogel, V. (2004). Structural Insights into How the MIDAS Ion Stabilizes Integrin Binding to an RGD Peptide under Force. *Structure*, 12(11), 2049-2058. <https://doi.org/10.1016/j.str.2004.09.009>
- Denadai, M. V., Viana, L. S., Affonso Jr, R. J., Silva, S. R., Oliveira, I. D., Toledo, S. R., & Matos, D. (2013). Expression of integrin genes and proteins in progression and dissemination of colorectal adenocarcinoma. *BMC Clinical Pathology*, 13(1), 16. <https://doi.org/10.1186/1472-6890-13-16>
- Dong, X., Hudson, N. E., Lu, C., & Springer, T. A. (2014). Structural determinants of integrin β -subunit specificity for latent TGF- β . *Nature Structural & Molecular Biology*, 21(12), 1091-1096. <https://doi.org/10.1038/nsmb.2905>
- D'Souza, S. E., Ginsberg, M. H., & Plow, E. F. (1991). Arginyl-glycyl-aspartic acid (RGD): A cell adhesion motif. *Trends in Biochemical Sciences*, 16, 246-250. [https://doi.org/10.1016/0968-0004\(91\)90096-E](https://doi.org/10.1016/0968-0004(91)90096-E)
- Dunbrack, R. L., & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science: A Publication of the Protein Society*, 6(8), 1661-1681. <https://doi.org/10.1002/pro.5560060807>
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19), 8577-8593. <https://doi.org/10.1063/1.470117>
- Ewing, I., Hurley, J. J., Josephides, E., & Millar, A. (2014). The molecular genetics of colorectal cancer. *Frontline Gastroenterology*, 5(1), 26-30. <https://doi.org/10.1136/flgastro-2013-100329>
- Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61(5), 759-767. [https://doi.org/10.1016/0092-8674\(90\)90186-I](https://doi.org/10.1016/0092-8674(90)90186-I)
- Feller, S. E., Zhang, Y., Pastor, R. W., & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, 103(11), 4613-4621. <https://doi.org/10.1063/1.470648>
- Fuentes Arderiu. (1998). *Bioquímica clínica y patología molecular. II*. Reverte.
- Gao, X., Dong, X., Li, X., Liu, Z., & Liu, H. (2020). Prediction of disulfide bond engineering sites using a machine learning method. *Scientific Reports*, 10(1), 10330. <https://doi.org/10.1038/s41598-020-67230-z>
- Grant, D. S., Tashiro, K.-I., Segui-Real, B., Yamada, Y., Martin, G. R., & Kleinman, H. K. (1989). Two different laminin domains mediate the differentiation of human endothelial cells into capillary-like structures in vitro. *Cell*, 58(5), 933-943. [https://doi.org/10.1016/0092-8674\(89\)90945-8](https://doi.org/10.1016/0092-8674(89)90945-8)
- Gunsteren, W. F. van, & Berendsen, H. J. C. (1990). Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English*, 29(9), 992-1023. <https://doi.org/10.1002/anie.199009921>
- Hautanen, A., Gailit, J., Mann, D. M., & Ruoslahti, E. (1989). Effects of modifications of the RGD sequence and its context on recognition by the fibronectin receptor. *The Journal of Biological Chemistry*, 264(3), 1437-1442.
- Healy, J. M., Murayama, O., Maeda, T., Yoshino, K., Sekiguchi, K., & Kikuchi, M. (1995). Peptide ligands for integrin α v β 3 selected from random phage display libraries. *Biochemistry*, 34(12), 3948-3955. <https://doi.org/10.1021/bi00012a012>
- Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: Advances and applications. *Advances and Applications in Bioinformatics and Chemistry: AABC*, 8, 37-47. <https://doi.org/10.2147/AABC.S70333>

- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38, 27-28. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Jespers, W., Åqvist, J., & Gutiérrez-de-Terán, H. (2021). Free Energy Calculations for Protein-Ligand Binding Prediction. *Methods in Molecular Biology (Clifton, N.J.)*, 2266, 203-226. https://doi.org/10.1007/978-1-0716-1209-5_12
- Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., & Schulten, K. (1999). NAMD2: Greater Scalability for Parallel Molecular Dynamics. *Journal of Computational Physics*, 151(1), 283-312. <https://doi.org/10.1006/jcph.1999.6201>
- Kerr, J. S., Wexler, R. S., Mousa, S. A., Robinson, C. S., Wexler, E. J., Mohamed, S., Voss, M. E., Devenny, J. J., Czerniak, P. M., Gudzelak, A., & Slee, A. M. (1999). Novel small molecule alpha v integrin antagonists: Comparative anti-cancer efficacy with known angiogenesis inhibitors. *Anticancer Research*, 19(2A), 959-968.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511623486>
- Koivunen, E., Gay, D. A., & Ruoslahti, E. (1993). Selection of peptides binding to the alpha 5 beta 1 integrin from phage display library. *The Journal of Biological Chemistry*, 268(27), 20205-20210.
- Korphaisarn, K., Morris, V., Davis, J. S., Overman, M. J., Fogelman, D. R., Kee, B. K., Dasari, A., Raghav, K. P. S., Shureiqi, I., Trupti, M., Wolff, R. A., Eng, C., Menter, D. G., Hamilton, S., & Kopetz, S. (2019). Signet ring cell colorectal cancer: Genomic insights into a rare subpopulation of colorectal adenocarcinoma. *British Journal of Cancer*, 121(6), 505-510. <https://doi.org/10.1038/s41416-019-0548-9>
- Krokhotin, A., & Dokholyan, N. V. (2015). Chapter Three—Computational Methods Toward Accurate RNA Structure Prediction Using Coarse-Grained and All-Atom Models. En S.-J. Chen & D. H. Burke-Aguero (Eds.), *Methods in Enzymology* (Vol. 553, pp. 65-89). Academic Press. <https://doi.org/10.1016/bs.mie.2014.10.052>
- Kukul, A. (Ed.). (2008). *Molecular Modeling of Proteins* (Vol. 443). Humana Press. <https://doi.org/10.1007/978-1-59745-177-2>
- Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., & Tufféry, P. (2016). PEP-FOLD3: Faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, 44(W1), W449-454. <https://doi.org/10.1093/nar/gkw329>
- Le Tourneau, C., Faivre, S., & Raymond, E. (2007). The role of integrins in colorectal cancer. *Oncology (Williston Park, N.Y.)*, 21(9 Suppl 3), 21-24.
- Li, H., Wang, Y., Rong, S., Li, L., Chen, T., Fan, Y., Wang, Y., Yang, C., Yang, C., Cho, W. C., & Yang, J. (2020). Integrin $\alpha 1$ promotes tumorigenicity and progressive capacity of colorectal cancer. *International Journal of Biological Sciences*, 16(5), 815-826. <https://doi.org/10.7150/ijbs.37275>
- Liu, D., & Cowburn, D. (2016). Combining biophysical methods to analyze the disulfide bond in SH2 domain of C-terminal Src kinase. *Biophysics Reports*, 2(1), 33-43. <https://doi.org/10.1007/s41048-016-0025-4>
- Luan, Y., & Xu, W. (2007). The structure and main functions of aminopeptidase N. *Current Medicinal Chemistry*, 14(6), 639-647. <https://doi.org/10.2174/092986707780059571>
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., ... Karplus,

- M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry. B*, *102*(18), 3586-3616.
<https://doi.org/10.1021/jp973084f>
- Maqbool, A. (2013). Colon: Structure, Function, and Disorders. En B. Caballero (Ed.), *Encyclopedia of Human Nutrition (Third Edition)* (pp. 378-396). Academic Press.
<https://doi.org/10.1016/B978-0-12-375083-9.00059-3>
- Marchler-Bauer, A., Panchenko, A. R., Ariel, N., & Bryant, S. H. (2002). Comparison of sequence and structure alignments for protein domains. *Proteins*, *48*(3), 439-446.
<https://doi.org/10.1002/prot.10163>
- Mármol, I., Sánchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E., & Rodríguez Yoldi, M. J. (2017). Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *International Journal of Molecular Sciences*, *18*(1).
<https://doi.org/10.3390/ijms18010197>
- Mas-Moruno, C., Rechenmacher, F., & Kessler, H. (2010). Cilengitide: The First Anti-Angiogenic Small Molecule Drug Candidate. Design, Synthesis and Clinical Evaluation. *Anti-Cancer Agents in Medicinal Chemistry*, *10*(10), 753-768.
<https://doi.org/10.2174/187152010794728639>
- Maupetit, J., Tuffery, P., & Derreumaux, P. (2007). A coarse-grained protein force field for folding and structure prediction. *Proteins*, *69*(2), 394-408.
<https://doi.org/10.1002/prot.21505>
- Mezu-Ndubuisi, O. J., & Maheshwari, A. (2020). The role of integrins in inflammation and angiogenesis. *Pediatric Research*, 1-8. <https://doi.org/10.1038/s41390-020-01177-9>
- Miyamoto, S., & Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, *13*(8), 952-962. <https://doi.org/10.1002/jcc.540130805>
- Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in Molecular Biology (Clifton, N.J.)*, *443*, 365-382. https://doi.org/10.1007/978-1-59745-177-2_19
- Nieberler, M., Reuning, U., Reichart, F., Notni, J., Wester, H.-J., Schwaiger, M., Weinmüller, M., Räder, A., Steiger, K., & Kessler, H. (2017). Exploring the Role of RGD-Recognizing Integrins in Cancer. *Cancers*, *9*(9), 116.
<https://doi.org/10.3390/cancers9090116>
- Núñez-Vivanco, G., Valdés-Jiménez, A., Besoain, F., & Reyes-Parada, M. (2016). Geomfinder: A multi-feature identifier of similar three-dimensional protein patterns: a ligand-independent approach. *Journal of Cheminformatics*, *8*(1), 19.
<https://doi.org/10.1186/s13321-016-0131-9>
- Oldberg, A., Franzén, A., & Heinegård, D. (1986). Cloning and sequence analysis of rat bone sialoprotein (osteopontin) cDNA reveals an Arg-Gly-Asp cell-binding sequence. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(23), 8819-8823. <https://doi.org/10.1073/pnas.83.23.8819>
- Orecchia, A., Lacal, P. M., Schietroma, C., Morea, V., Zambruno, G., & Failla, C. M. (2003). Vascular endothelial growth factor receptor-1 is deposited in the extracellular matrix by endothelial cells and is a ligand for the alpha 5 beta 1 integrin. *Journal of Cell Science*, *116*(Pt 17), 3479-3489. <https://doi.org/10.1242/jcs.00673>
- Pasqualini, R., Koivunen, E., Kain, R., Lahdenranta, J., Sakamoto, M., Stryhn, A., Ashmun, R. A., Shapiro, L. H., Arap, W., & Ruoslahti, E. (2000). Aminopeptidase N is a receptor for tumor-homing peptides and a target for inhibiting angiogenesis. *Cancer Research*, *60*(3), 722-727.

- Pelillo, C., Mollica, H., Eble, J. A., Grosche, J., Herzog, L., Codan, B., Sava, G., & Bergamo, A. (2016). Inhibition of adhesion, migration and of $\alpha 5 \beta 1$ integrin in the HCT-116 colorectal cancer cells treated with the ruthenium drug NAMI-A. *Journal of Inorganic Biochemistry*, 160, 225-235. <https://doi.org/10.1016/j.jinorgbio.2016.02.025>
- Peng, C., Zou, X., Xia, W., Gao, H., Li, Z., Liu, N., Xu, Z., Gao, C., He, Z., Niu, W., Fang, R., Biswas, S., Agrez, M., Zhi, X., & Niu, J. (2018). Integrin $\alpha v \beta 6$ plays a bi-directional regulation role between colon cancer cells and cancer-associated fibroblasts. *Bioscience Reports*, 38(6). <https://doi.org/10.1042/BSR20180243>
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781-1802. <https://doi.org/10.1002/jcc.20289>
- Pierschbacher, M. D., & Ruoslahti, E. (1984). Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature*, 309(5963), 30-33. <https://doi.org/10.1038/309030a0>
- Pierschbacher, M. D., & Ruoslahti, E. (1987). Influence of stereochemistry of the sequence Arg-Gly-Asp-Xaa on binding specificity in cell adhesion. *The Journal of Biological Chemistry*, 262(36), 17294-17298.
- Pinon, P., & Wehrle-Haller, B. (2011). Integrins: Versatile receptors controlling melanocyte adhesion, migration and proliferation. *Pigment Cell & Melanoma Research*, 24(2), 282-294. <https://doi.org/10.1111/j.1755-148X.2010.00806.x>
- Pitman, M. R., & Menz, R. I. (2006). 2—Methods for Protein Homology Modelling. En D. K. Arora, R. M. Berka, & G. B. Singh (Eds.), *Applied Mycology and Biotechnology* (Vol. 6, pp. 37-59). Elsevier. [https://doi.org/10.1016/S1874-5334\(06\)80005-5](https://doi.org/10.1016/S1874-5334(06)80005-5)
- Plow, E. F., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. A., & Ginsberg, M. H. (1985). The effect of Arg-Gly-Asp-containing peptides on fibrinogen and von Willebrand factor binding to platelets. *Proceedings of the National Academy of Sciences of the United States of America*, 82(23), 8057-8061. <https://doi.org/10.1073/pnas.82.23.8057>
- Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., & Sanner, M. F. (2015). AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Computational Biology*, 11(12), e1004586. <https://doi.org/10.1371/journal.pcbi.1004586>
- Rawla, P., Sunkara, T., & Barsouk, A. (2019). Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przegląd Gastroenterologiczny*, 14(2), 89-103. <https://doi.org/10.5114/pg.2018.81072>
- Ruoslahti, E. (1996). RGD and other recognition sequences for integrins. *Annual Review of Cell and Developmental Biology*, 12, 697-715. <https://doi.org/10.1146/annurev.cellbio.12.1.697>
- Schittenhelm, J., Klein, A., Tatagiba, M. S., Meyermann, R., Fend, F., Goodman, S. L., & Sipos, B. (2013). Comparing the expression of integrins $\alpha v \beta 3$, $\alpha v \beta 5$, $\alpha v \beta 6$, $\alpha v \beta 8$, fibronectin and fibrinogen in human brain metastases and their corresponding primary tumors. *International Journal of Clinical and Experimental Pathology*, 6(12), 2719-2732.
- Schrödinger. (2021). *Maestro* (Schrödinger Release 2021-4) [Computer software].
- Schrödinger, LLC. (2015). *The PyMOL Molecular Graphics System, Version 1.8*.

- Sideris, M., & Papagrigroriadis, S. (2014). Molecular biomarkers and classification models in the evaluation of the prognosis of colorectal cancer. *Anticancer Research*, 34(5), 2061-2068.
- Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1), 7-30. <https://doi.org/10.3322/caac.21590>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. <https://doi.org/10.1038/msb.2011.75>
- Singhal, S., Changela, K., Basi, P., Mathur, S., Reddy, S., Momeni, M., Krishnaiah, M., & Anand, S. (2014). Prescreening with FOBT Improves Yield and Is Cost-Effective in Colorectal Screening in the Elderly. *ISRN Gastroenterology*, 2014, 179291. <https://doi.org/10.1155/2014/179291>
- Sordat, I., Bosman, F. T., Dorta, G., Rousselle, P., Aberdam, D., Blum, A. L., & Sordat, B. (1998). Differential expression of laminin-5 subunits and integrin receptors in human colorectal neoplasia. *The Journal of Pathology*, 185(1), 44-52. [https://doi.org/10.1002/\(SICI\)1096-9896\(199805\)185:1<44::AID-PATH46>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-9896(199805)185:1<44::AID-PATH46>3.0.CO;2-A)
- Springer, T. A., & Dustin, M. L. (2012). Integrin inside-out signaling and the immunological synapse. *Current Opinion in Cell Biology*, 24(1), 107-115. <https://doi.org/10.1016/j.ceb.2011.10.004>
- Stupp, R., Hegi, M. E., Gorlia, T., Erridge, S. C., Perry, J., Hong, Y.-K., Aldape, K. D., Lhermitte, B., Pietsch, T., Grujicic, D., Steinbach, J. P., Wick, W., Tarnawski, R., Nam, D.-H., Hau, P., Weyerbrock, A., Taphoorn, M. J. B., Shen, C.-C., Rao, N., ... CENTRIC study team. (2014). Cilengitide combined with standard treatment for patients with newly diagnosed glioblastoma with methylated MGMT promoter (CENTRIC EORTC 26071-22072 study): A multicentre, randomised, open-label, phase 3 trial. *The Lancet. Oncology*, 15(10), 1100-1108. [https://doi.org/10.1016/S1470-2045\(14\)70379-1](https://doi.org/10.1016/S1470-2045(14)70379-1)
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, n/a(n/a). <https://doi.org/10.3322/caac.21660>
- Suzuki, S., Oldberg, A., Hayman, E. G., Pierschbacher, M. D., & Ruoslahti, E. (1985). Complete amino acid sequence of human vitronectin deduced from cDNA. Similarity of cell attachment sites in vitronectin and fibronectin. *The EMBO Journal*, 4(10), 2519-2524.
- Tanner, D. E., Chan, K.-Y., Phillips, J. C., & Schulten, K. (2011). Parallel Generalized Born Implicit Solvent Calculations with NAMD. *Journal of Chemical Theory and Computation*, 7(11), 3635-3642. <https://doi.org/10.1021/ct200563j>
- The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480-D489. <https://doi.org/10.1093/nar/gkaa1100>
- Tuckerman, M., Berne, B. J., & Martyna, G. J. (1992). Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3), 1990-2001. <https://doi.org/10.1063/1.463137>
- Vhora, I., Patil, S., Bhatt, P., & Misra, A. (2015). Chapter One - Protein– and Peptide–Drug Conjugates: An Emerging Drug Delivery Technology. En R. Donev (Ed.), *Advances in Protein Chemistry and Structural Biology* (Vol. 98, pp. 1-55). Academic Press. <https://doi.org/10.1016/bs.apcsb.2014.11.001>

- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., & Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. *The New England Journal of Medicine*, *319*(9), 525-532. <https://doi.org/10.1056/NEJM198809013190901>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wei, Y., Yin, G., Yin, H., Yan, D., Ma, C., Huang, Z., Liao, X., Yao, Y., & Chen, X. (2013). Inhibiting Effects of a Cyclic Peptide CNGRC on Proliferation and Migration of Tumor Cells In Vitro. *International Journal of Peptide Research and Therapeutics*, *19*(2), 163-173. <https://doi.org/10.1007/s10989-012-9327-7>
- Wong, A., Gehring, C., & Irving, H. R. (2015). Conserved Functional Motifs and Homology Modeling to Predict Hidden Moonlighting Functional Sites. *Frontiers in Bioengineering and Biotechnology*, *3*. <https://doi.org/10.3389/fbioe.2015.00082>
- Xia, W., & Springer, T. A. (2014). Metal ion and ligand binding of integrin $\alpha 5\beta 1$. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(50), 17863-17868. <https://doi.org/10.1073/pnas.1420645111>
- Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L., & Arnaout, M. A. (2001). Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science (New York, N. Y.)*, *294*(5541), 339-345. <https://doi.org/10.1126/science.1064535>
- Xiong, J.-P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L., & Arnaout, M. A. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science (New York, N. Y.)*, *296*(5565), 151-155. <https://doi.org/10.1126/science.1069040>
- Yu, Y.-P., Wang, Q., Liu, Y.-C., & Xie, Y. (2014). Molecular basis for the targeted binding of RGD-containing peptide to integrin $\alpha V\beta 3$. *Biomaterials*, *35*(5), 1667-1675. <https://doi.org/10.1016/j.biomaterials.2013.10.072>
- Zhang, Y., Forli, S., Omelchenko, A., & Sanner, M. F. (2019). AutoGridFR: Improvements on AutoDock Affinity Maps and Associated Software Tools. *Journal of Computational Chemistry*, *40*(32), 2882-2886. <https://doi.org/10.1002/jcc.26054>
- Zhang, Y., & Sanner, M. F. (2019a). Docking Flexible Cyclic Peptides with AutoDock CrankPep. *Journal of Chemical Theory and Computation*, *15*(10), 5161-5168. <https://doi.org/10.1021/acs.jctc.9b00557>
- Zhang, Y., & Sanner, M. F. (2019b). AutoDock CrankPep: Combining folding and docking to predict protein-peptide complexes. *Bioinformatics (Oxford, England)*, *35*(24), 5121-5127. <https://doi.org/10.1093/bioinformatics/btz459>
- Zheng, Y., & Leftheris, K. (2020). Insights into Protein–Ligand Interactions in Integrin Complexes: Advances in Structure Determinations. *Journal of Medicinal Chemistry*, *63*(11), 5675-5696. <https://doi.org/10.1021/acs.jmedchem.9b01869>
- Zhong, Y., Su, T., Shi, Q., Feng, Y., Tao, Z., Huang, Q., Li, L., Hu, L., Li, S., Tan, H., Liu, S., & Yang, H. (2019). Co-Administration Of iRGD Enhances Tumor-Targeted Delivery And Anti-Tumor Effects Of Paclitaxel-Loaded PLGA Nanoparticles For Colorectal Cancer Treatment. *International Journal of Nanomedicine*, *14*, 8543-8560. <https://doi.org/10.2147/IJN.S219820>